Publications du **La**boratoire de **C**ombinatoire et d' **I**nformatique **M**athématique

11

Édité par P. Leroux et C. Reutenauer

Séries formelles et combinatoire algébrique

4e colloque
15-19 juin 1992
Université du Québec à Montréal, Québec

Actes

Département de mathématiques et d'informatique

Université du Québec à Montréal

Le 4e Colloque " Séries Formelles et Combinatoire Algébrique " a été organisé par le LACIM (Université du Québec à Montréal) du 15 au 19 juin 1992.

Comité de programme

C. Reutenauer, UQAM, président; M. Delest, Un. Bordeaux I; G. Duchamp, Un. de Rouen; P. Flajolet, INRIA; G. Hahn, Un. de Montréal; D. Jackson, Un. of Waterloo; C. Lam, Un. Concordia; P. Leroux, UQAM; G.-C. Rota, MIT; M.-P. Schützenberger, LITP, Un. Paris VII; R. Stanley, MIT; D. Thérien, Un. McGill; D. Zeilberger, Temple Un.

Comité D'organisation

P. Leroux, UQAM, président; S. Brlek, UQAM, coprésident;
M. Blais, LACIM, secrétaire; R. Bédard, UQAM; J. Labelle, UQAM;
G. Duchamp, Un. de Rouen.

Avec le soutien financier de

PAFACC, UQAM

CRSNG, Canada

FCAR, Québec

FONDATION DE L' UQAM

En coopération avec

Association for Computing Machinery

Dépôt légal, premier semestre 1992, Bibliothèque nationale du Québec.

ISBN 2-89276-103-4 LACIM Montréal

Laboratoire de combinatoire et d'informatique mathématique
Département de mathématiques et d'informatique
Université du Québec à Montréal
C.P. 8888, Succ. A
Montréal, Qc.
Canada H3C 3P8

AVANT-PROPOS

Ce volume contient les textes de six des neuf conférences plénières ainsi que des trente et une communications présentées au 4e colloque sur les Séries formelles et la combinatoire algébrique, à l'Université du Québec à Montréal (UQAM), du 15 au 19 juin 1992. Ce colloque a pour but d'approfondir les liens entre la combinatoire et l'informatique, et leurs applications à d'autres domaines des mathématiques et des sciences. On trouvera dans une publication séparée les treize communications ayant fait l'objet d'une séance d'affichage et les deux présentations de logiciels également au programme.

Les thèmes représentés au colloque sont variés et témoignent d'une interaction fructueuse entre la combinatoire et l'informatique, l'algèbre et l'analyse. De façon plus précise on peut les regrouper à l'intérieur des domaines suivants:

— les séries formelles,
— la combinatoire des mots et les langages formels,
— la combinatoire énumérative, l'étude et la décomposition des structures,
— la combinatoire algébrique, les groupes de Coxeter,
— les fonctions symétriques et la théorie des représentations,
— l'algorithmique et le calcul formel algébrique,
— aspects géométriques, topologiques ou probabilistes.

Nous désirons remercier chaleureusement les membres du Comité de programme et les arbitres qui ont participé à la sélection des communications parmi les soixante et onze soumises. La liste des arbitres est donnée en annexe. Nos remerciements vont également à:

— nos commanditaires, le Conseil de recherche en sciences naturelles et en génie (CRSNG) du Canada, le Fonds pour la formation de chercheurs et l'aide à la recherche (FCAR) du Québec, la Fondation de l'UQAM, le Programme d'aide financière (PAFACC) de l'UQAM, ainsi que l'Association for Computing Machinery (ACM) qui a parrainé le colloque,
— la secrétaire du LACIM, Madame Manon Blais, pour son excellent travail et son implication,
— les conférenciers et les participants pour leur contribution au succès de ce colloque.

Pierre Leroux et Christophe Reutenauer
pour le Comité de programme

FOREWORD

This volume contains the texts of six of the nine invited talks and of the thirty one communications given at the 4th Conference on Formal Power Series and Algebraic Combinatorics, held at the Université du Québec à Montréal (UQAM), June 15 to 19, 1992. The purpose of the conference is to thoroughly explore the relationships between combinatorics and computer science, and their applications in other parts of mathematics and science. A separate publication contains the thirteen papers of the poster session and resumes of the two software demonstrations in the program.

The topics covered by the conference are diverse and illustrate very well the fruitful interactions between combinatorics and computer science, algebra and analysis. They can be regrouped within the following fields:

— formal power series,
— combinatorics on words and formal languages,
— enumerative combinatorics, the study and decomposition of structures,
— algebraic combinatorics, Coxeter groups,
— symmetric functions and representation theory,
— algorithms and computer algebra,
— geometric, topological and probabilistic aspects.

We would like to thank the members of the program committee and the numerous referees who participated in the selection of papers among the seventy one that were submitted. The list of referees is given in the appendix. Our sincere thanks are also adressed to:

— our sponsors, the Natural Sciences and Engeneering Research Council (NSERC) of Canada, the FCAR funding agency of Quebec, the UQAM Foundation and the PAFACC of UQAM, together with the Association for Computing Machinery (ACM), as a cooperating organization,
— the secretary of LACIM, Mrs Manon Blais, for her excellent work and her involvement,
— the speakers and all participants for their contribution in making this conference a success.

Pierre Leroux and Christophe Reutenauer
for the Program Committee

TABLE DES MATIÈRES

YOUNG STRAIGHTENING IN A QUOTIENT S_N-MODULE

Hélène Barcelo †
Department of Mathematics, Arizona State University, Tempe, Az 85287-1804.

ABSTRACT. We describe a straightening algorithm for the action of S_n on a certain graded ring $\mathcal{R}_\mu$. The ring $\mathcal{R}_\mu$ appears in the work of C. de Concini and C. Procesi [1], and T.Tanisaki [7] and more recently in the work of A. Garsia and C. Procesi[3]. This ring is a graded version of the permutation representation resulting from the action of S_n on the left cosets of a Young subgroup. As a corollary of our straightening algorithm we obtain a combinatorial proof of the fact that the top degree component of $\mathcal{R}_\mu$ affords the irreducible representation of S_n indexed by μ.

Introduction

This paper is concerned with certain graded S_n-modules $\mathcal{R}_\mu$ studied by H. Kraft in [5], C. de Concini and C. Procesi in [1] and more recently by A. Garsia and C. Procesi in [3].

Given $\mu = (\mu_1, \ldots \mu_k)$ a partition of n let p^μ denote the character of the permutation representation resulting from the action of S_n on the left cosets of the Young subgroup

$$S_{\mu_1} \times S_{\mu_2} \times \ldots \times S_{\mu_k}$$

It is well known [8] that p^μ contains the irreducible character χ^μ (in the Young-Frobenius indexation) with multiplicity 1. It follows by combining the results of [5] and [1] that $\mathcal{R}_\mu$ yields a graded version of this representation. In particular its graded character is given by a polynomial $p^\mu(q)$ which reduces to p^μ for $q = 1$. We refer to [3] for an elementary presentation of the background material. It was conjectured by H. Kraft in [5] that χ^μ is given by the top degree component of $p^\mu(q)$. This conjecture was first proved by de Concini and Procesi in their 1981 paper [1]. Another proof can be found in the 1991 paper of Garsia and Procesi [3]. In this paper the authors describe the graded character $p^\mu(q)$ as a linear combination of irreducibles χ^λ with coefficient equal to the cocharge version of the Kostka-Foulkes polynomials $K_{\lambda\mu}(q)$ [6]. Garsia and Procesi also construct a homogeneous basis $\mathcal{B}_\mu$ for the ring $\mathcal{R}_\mu$, and one of their questions was to know if the action of S_n on the homogeneous component of highest degree of $\mathcal{B}_\mu$ *naturally* yields the irreducible representation of S_n indexed by μ. Our objective in this paper is to answer this question by developing a straightening law for the action of S_n on the basis $\mathcal{B}_\mu^{top}$ of $\mathcal{R}_\mu^{top}$, the homogeneous component of highest degree of $\mathcal{R}_\mu$. As a corollary we obtain an alternative, direct combinatorial proof of Kraft's conjecture.

One way to obtain the irreducible representation of S_n indexed by μ is through the action of S_n on Young's natural set of units $\{E_{T_i,T_i}\}_{i=1}^{\eta_\mu}$ (where T_i for $i = 1, \ldots n$ are the standard tableaux

† Work partially supported by ASU Grant (FGIA-076-92).

of shape μ). More precisely, one obtains the matrices of the representation by finding the images $\sigma E_{T_i,T_1}$ for each of the permutations σ of S_n. For a given standard tableau T_i let

$$\sigma E_{T_i,T_1} = \sum_{j=1}^{\eta_\mu} a_j(T_i) E_{T_j,T_1}.$$

We shall show that the action of S_n on a basis $\mathcal{B}_\mu^{top}$ of $\mathcal{R}_\mu^{top}$ (the homogeneous component of highest degree of $\mathcal{R}_\mu$) is an action identical to the one on Young's natural units. For this we shall calculate the images of the elements of $\mathcal{B}_\mu^{top}$ under the action of the permutations σ of S_n. As we shall see in section 1 the elements $m(T_i)$ of $\mathcal{B}_\mu^{top}$ are monomials indexed by the set of standard tableaux of shape μ. Thus if we let

$$\sigma m(T_i) = \sum_{j=1}^{\eta_\mu} b_j(T_i) m(T_j)$$

our result will be that

$$a_j(T_i) = b_j(T_i)$$

for all i and j.

This paper is divided in four sections. In section 1 we give the description of the ring $\mathcal{R}_\mu$ as a quotient of the polynomial ring $Q[x_1, x_2, \ldots, x_n]$. We also describe the basis $\mathcal{B}_\mu^{top}$. These results are due to T. Tanisaki, A. Garsia, and C. Procesi. In section 2 we develop some of the congruence relations that will be needed for working on the quotient ring $\mathcal{R}_\mu$. Section 3 is devoted to the construction of a straightening algorithm. More precisely we give a rule for finding the coefficients $b_j(T_i)$ in the expansion $\sigma m(T_i) = \sum_{j=1}^{\eta_\mu} b_j(T_i) m(T_j)$. In section 4 we show that this is the same as Young's straightening law.

1. The ring $\mathcal{R}_\mu$.

Some of the results we need have been recently described in a paper of A. Garsia and C. Procesi [3]. We shall adopt here their notation. The partitions of n will be represented by $n-$vectors:

$$\mu = (\mu_1, \mu_2, \ldots, \mu_n) \quad (0 \le \mu_1 \le \mu_2 \ldots \le \mu_n)$$

or by their corresponding Ferrers' diagrams drawn according to the French notation. Figure 1(a) gives the Ferrers' diagram associated to the partition $\mu = (0,0,1,2,2)$.

$\mu = (0,0,1,2,2)$ (a) $\quad$ $\mu = (0,0,0,2,3)$ (b)

Figure 1

The number of positive components of μ is the height of μ and is denoted by $h(\mu)$. The conjugate of a partition μ is the partition μ' whose Ferrers' diagram is the transpose of the diagram of μ. In our previous example $h(\mu) = 3$, and $\mu' = (0,0,0,2,3)$ (see figure 1(b)). Let $X_n = \{x_1, x_2, \ldots, x_n\}$ be an ordered set of commuting variables. For any integer r ($\le n$) and for any subset $S \subseteq X_n$ such that $r \leq |S|$, let $e_r(S)$ be the r^{th} elementary symmetric function of the variables in S, that is

$$e_r(S) = \sum_{\substack{i_1 < i_2 < \ldots < i_r \\ x_{i_j} \in S}} x_{i_1} x_{i_2} \ldots x_{i_r}$$

Define $d_k(\mu) = \sum_{i=1}^{k} \mu_i{}'$ for all $k = 1, \ldots, n$. Let $\mathbf{Q}[x_1, x_2, \ldots, x_n]$ be the ring of polynomials in the variables $x_1, \ldots x_n$ with rational coefficients. Let $\mathcal{I}_\mu$ be the ideal (in $\mathbf{Q}[x_1, x_2, \ldots, x_n]$) generated by the collection of partial elementary symmetric functions:

$$\mathcal{C}_\mu = \{e_r(S) \mid k - d_k(\mu) < r \leq k, \ |S| = k, \ S \subseteq X_n\} \ . \tag{1.1}$$

The following presentation of the rings $\mathcal{R}_\mu$ is due to T. Tanisaki (see [7]). For a given μ, the ring $\mathcal{R}_\mu$ is given by the quotient

$$\mathcal{R}_\mu = \mathbf{Q}[x_1, x_2, \ldots, x_n]/\mathcal{I}_\mu \ .$$

For example when $\mu = (0,0,1,2,2)$ we have $\mu' = (0,0,0,2,3)$ while $(d_1(\mu), \ldots, d_n(\mu)) = (0,0,0,2,5)$ and $(1 - d_1(\mu), \ldots, 5 - d_5(\mu)) = (1,2,3,2,0)$. Schematically we can represent the pairs (k, r) satisfying the condition $k - d_k(\mu) < r \leq k$ by the diagram of figure 2.

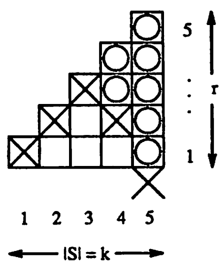


Figure 2

The squares with coordinates given by $(k, k - d_k(\mu))$ for $k = 1, \ldots n$ are marked with an X. From equation (1.1) it is easy to see that the partial elementary symmetric functions $e_r(S)$ belonging to $\mathcal{C}_\mu$ are the ones for which the points $(|S|, r)$ are given by the coordinates of all the squares above the one marked with an X. Thus in our example, $\mathcal{I}_{122}$ is generated by the following 15 partial elementary symmetric functions:

$$e_1(X_5), \ e_2(X_5), \ e_3(X_5), \ e_4(X_5), \ e_5(X_5),$$
$$e_3(x_1, x_2, x_3, x_4), \ e_3(x_1, x_2, x_3, x_5), \ e_3(x_1, x_2, x_4, x_5),$$
$$e_3(x_1, x_3, x_4, x_5), \ e_3(x_2, x_3, x_4, x_5), \ x_1x_2x_3x_4,$$
$$x_1x_2x_3x_5, \ x_1x_2x_4x_5, \ x_1x_3x_4x_5, \ x_2x_3x_4x_5 \ .$$

It is easy to see that there is, in general, redundancy among the generators of $\mathcal{I}_\mu$. For example, consider

$$e_4(X_5) = x_1x_2x_3x_4 + x_1x_2x_3x_5 + x_1x_2x_4x_5 + x_1x_3x_4x_5 + x_2x_3x_4x_5 \ . \tag{1.2}$$

Note that each of the monomials of the right hand side of equation (1.2) is also included in $\mathcal{I}_\mu$. Thus it is unnecessary to add $e_4(X_5)$. We have studied which partial elementary symmetric functions are sufficient to generate the ideals $\mathcal{I}_\mu$. But before we summarize these results in the next two lemmas, observe first that:

Remark : for any partition μ of n it is always true that $n - d_n(\mu) = 0$; thus for all $i = 1,\ldots,n$, $e_i(X_n)$ will always belong to $\mathcal{C}_\mu$.

Lemma 1.1

Let μ be a partition of n and let S be a proper subset of X_n . Let $1 \le k < n$ $(k \ne n)$ and $r+1 \le k$. If for all subsets S of cardinality k we have that $e_r(S) \in \mathcal{I}_\mu$ then

$$e_{r+1}(S) \in \mathcal{I}_\mu$$

for all subsets S of cardinality k .

Proof

Let $S = \{x_{s_1}, x_{s_2}, \ldots x_{s_k}\}$. Since k is strictly less than n there exists a variable say x_* that belongs to X_n but not to S . Define S_i to be the set obtained from S by removing the variable x_{s_i}, and define S_i^* to be the one obtained from S by replacing the variable x_{s_i} by the variable x_*. We shall prove that for any S satisfying the conditions of the lemma the following equation holds

$$(r+1)e_{r+1}(S) \;=\; \sum_{i=1}^{k} x_{s_i} e_r(S_i^*) \;-\; (r)x_* e_r(S)$$

Note that the left hand side of equation (1.3) is by definition

$$(r+1)e_{r+1}(S) = (r+1) \sum_{\substack{i_1<\cdots<i_{r+1} \\ x_{s_{i_j}} \in S}} x_{s_{i_1}} \cdots x_{s_{i_{r+1}}}$$

On the other hand the right hand side of equation (1.3) is equal to:

$$= \sum_{i=1}^{k} x_{s_i} \Big[x_* e_{r-1}(S_i) + e_r(S_i) \Big] - (r)x_* e_r(S)$$

$$= \sum_{i=1}^{k} x_{s_i} x_* e_{r-1}(S_i) + \sum_{i=1}^{k} x_{s_i} e_r(S_i) - (r)x_* e_r(S) \qquad (1.4)$$

A moment of thought reveals that in equation (1.4) the terms containing x_* cancel out, leaving

$$\sum_{i=1}^{k} x_{s_i} \sum_{\substack{i_1<\cdots<i_r, \\ x_{s_{i_j}} \in S_i}} x_{s_{i_1}} \cdots x_{s_{i_r}}$$

which is precisely $(r+1)e_{r+1}(S)$. This completes the proof of the lemma since for all $i = 1,\ldots k$ we have that $e_r(S_i^*)$ belongs to $\mathcal{I}_\mu$. ♣

Lemma 1.2

Let μ be a partition of n, and let k be a fixed integer such that $k \le n$. If for every subset S of X_n, of cardinality $k-1$, there exists an r such that $e_r(S)$ belongs to $\mathcal{I}_\mu$, then for all subsets S^+ of X_n of cardinality k we have:

$$e_r(S^+) \in \mathcal{I}_\mu .$$

Proof

For any given set S^+, let $\{S_j\}_{j=1}^k$ be the collection of all subsets of S^+ of cardinality $k-1$. The proof follows from the fact that:

$$(k-r)e_r(S^+) = \sum_{j=1}^{k} e_r(S_j) \tag{1.5}$$

Indeed it is easy to see that every monomial $x_{s_{i_1}} \dots x_{s_{i_r}}$ of $e_r(S^+)$ appears exactly $(k-r)$ times on the right hand side of equation (1.5). Thus since each $e_r(S_j)$ belongs to $\mathcal{I}_\mu$ so does $e_r(S^+)$. ♣

We remark that a similar argument shows that if for all $S \subseteq X_n$ of cardinality k, $e_r(S) \in \mathcal{I}_\mu$ then $e_{r+1}(S') \in \mathcal{I}_\mu$ for all $S' \subseteq X_n$ of cardinality $k+1$. Note that this also follows from the two previous lemmas, except for the case where $r = k = n-1$.

Applying these two lemmas to our previous example yields that $\mathcal{I}_{122}$ is generated by the following seven partial elementary symmetric functions:

$$\mathcal{C}^*_\mu = \{e_1(X_5),\ e_2(X_5),\ e_3(x_1,x_2,x_3,x_4),\ e_3(x_1,x_2,x_3,x_5),\ e_3(x_1,x_2,x_4,x_5),$$
$$e_3(x_1,x_3,x_4,x_5),\ e_3(x_2,x_3,x_4,x_5)\}\ .$$

The ideals $\mathcal{I}_\mu$ are generated by homogeneous polynomials, thereby inducing a natural grading on each ring $\mathcal{R}_\mu$. The symmetric group S_n acts naturally on $\mathbf{Q}[x_1,\dots,x_n]$ by simply permuting the variables. That is:

$$\sigma P(x_1,\dots,x_n) = P(x_{\sigma_1},x_{\sigma_2},\dots,x_{\sigma_n})$$

for each $\sigma \in S_n$ and each polynomial $P(x_1,\dots,x_n) \in \mathbf{Q}[x_1,x_2,\dots,x_n]$. Clearly this action leaves each of the ideals $\mathcal{I}_\mu$ invariant. Thus one can define an action of S_n on each of the quotient rings $\mathcal{R}_\mu$. One also observes that the natural grading of $\mathcal{R}_\mu$ is preserved by this action of S_n. As we mentioned in the introduction we shall be concerned here with the action of S_n on the homogeneous component of highest degree of a given $\mathcal{R}_\mu$. A. Garsia and C. Procesi showed in [3, proposition 4.2] that each ring $\mathcal{R}_\mu$ has a basis of homogeneous monomials that can be constructed as follow. First recall that there is a Ferrers diagram associated to any partition μ of n. If one fills the diagram using all the integers 1 to n the resulting diagram is called an *injective tableau*. If the integers of each row and column of an injective tableau are in increasing order (from left to right, and from bottom to top) the tableau is said to be standard . Since we are only concerned here with injective tableaux we shall (by abuse of language) use the term tableau for injective tableau. Given a partition μ, for each tableau T of shape μ we shall associate the following monomial :

$$m(T) = \prod_{i \in T} x_i^{h(i,T)-1}$$

where $h(i,T)$ denotes the height of the letter i in T. (Recall that we draw the tableaux according to the French notation.) Let $\mathcal{B}_\mu$ be the lower order ideal of monomials whose maximal elements are the monomials of the standard tableaux of shape μ. Proposition 4.2 of [3] states that indeed the monomials of $\mathcal{B}_\mu$ form a basis for the corresponding ring $\mathcal{R}_\mu$. In particular the standard tableau monomials $\{m(T)|T$ *is a standard tableau of shape* $\mu\}$ are a basis for the homogeneous component of highest degree:

$$n(\mu) = \sum_{i=1}^{h(\mu)} (i-1)\mu_i$$

where for convenience we let $\mu = (\mu_1, \ldots, \mu_{h(\mu)})$ with $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_{h(\mu)} > 0$. Denote the highest homogeneous component of $\mathcal{R}_\mu$ by $\mathcal{R}_\mu^{top}$. Therefore in our previous example a basis for $\mathcal{R}_{122}^{top}$ is given by

$$\mathcal{B}_{122}^{top} = \{x_2x_5x_3^2,\ x_2x_5x_4^2,\ x_2x_4x_5^2,\ x_3x_5x_4^2,\ x_3x_4x_5^2\}\ .$$

Observe that the number of elements of $\mathcal{B}_\mu^{top}$ is equal to the number of standard tableaux of shape μ denoted by η_μ. We claim (theorem 4.1) that the matrices obtained by acting with the permutations of S_n on $\mathcal{B}_\mu^{top}$, are the same as the matrices of Young's natural representation. In order to prove this result we have to express any element of the module $\mathcal{R}_\mu^{top}$ as a linear combination of the elements of the basis $\mathcal{B}_\mu^{top}$. The next section is devoted to the construction of a straightening algorithm that will (step by step) express any monomials of $\mathcal{R}_\mu^{top}$ as a linear combination of the monomials in $\mathcal{B}_\mu^{top}$.

2. Congruence relations in $\mathcal{R}_\mu$.

We first need to determine which monomials of a given ring $\mathcal{R}_\mu$ are congruent to zero modulo $\mathcal{I}_\mu$. The following result can be found in a paper by A. Garsia and N. Bergeron [2]. Let μ be a partition of n, and let $X = \{x_1, \ldots x_n\}$. For any sequence of n integers $p_1, \ldots, p_n$ let $x^p = x_1^{p_1} \ldots x_n^{p_n}$. Define $x^p << x^q$ if and only if $p_i \leq q_i$, for all $i = 1, \ldots, n$.

Lemma 2.1 [2, proposition 4.5]

In $\mathcal{R}_\mu$,

$$x^p \not\equiv 0 \Longleftrightarrow x^p << m(T)$$

We mention here that this is an equivalent form of proposition 4.5 of [2]. Indeed their proposition states that a monomial of $\mathcal{R}_\mu$ is not congruent to zero if and only if it is an S_n–image of an element of $\mathcal{B}_\mu$. Here is a schematic interpretation of this result that will be useful later on. For any tableau T of shape $\mu = (\mu_1, \ldots, \mu_k)$ with $\mu_1 \geq \ldots \geq \mu_k > 0$, $(\mu \vdash n,\ k = h(\mu))$, we shall represent the sequence of exponents of $m(T)$ by the diagram of figure 3.

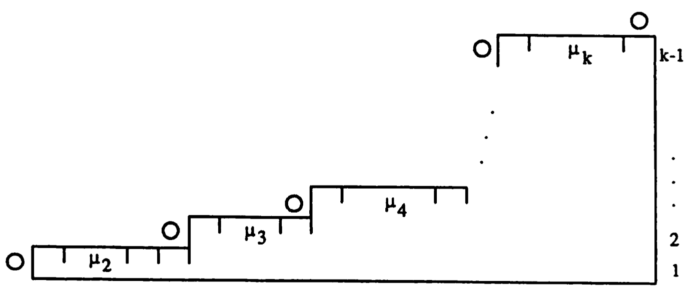


Figure 3

With this in mind, one can reformulate lemma 2.1 as follows: any monomial x^p of $\mathcal{R}_\mu$ whose sequence of powers $(p_1, \ldots, p_n)$ fills at least one of the corner marked with a *circle*, is congruent to zero modulo $\mathcal{I}_\mu$. (In other words if a monomial of $\mathcal{R}_\mu$ has μ_2 variables of degree 1, μ_3 variables of degree $2, \ldots, \mu_k$ variables of degree $h(\mu) - 1$, it is not congruent to zero (and indeed belongs to $\mathcal{R}_{\mu}^{top}$).) Note that two distinct tableaux T_1 and T_2 can yield the same monomial $m(T)$. For example if

Figure 4

then $m(T_1) = m(T_2) = x_3 x_4 x_5^2$. Therefore we shall only consider row-increasing tableaux. (That is, the ones for which the rows are in increasing order from left to right). A few more observations are needed before we can give a description of our straightening algorithm. For this we shall look at an example. Let T be the following tableau of shape $\mu = (3,3,3)$:

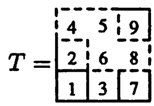


Figure 5

Since T is not a standard tableau, its corresponding monomial $m(T) = x_2 x_6 x_8 x_4^2 x_5^2 x_9^2$ belongs to $\mathcal{R}_{333}^{top}$ but not to the basis $\mathcal{B}_{333}^{top}$. Thus we shall expand $m(T)$ as a linear combination of monomials of $\mathcal{B}_{333}^{top}$. For this notice that the first break of standardness occurs in position $(2,3)$. We shall draw a ribbon as in figure 5, delimiting the position $(2,3)$. Let $X = \{1,\ldots,9\}$, $A = \{x_4, x_5\}$, and $B = \{x_6, x_8\}$. The first step of our algorithm will be to transform the monomial $m(T)$ into a new monomial $m'(T)$ as follows:

1. subtract one from all the exponents of the variables of $m(T)$ belonging to the set A.

This yields the new monomial $m'(T) = x_2 x_6 x_8 x_4 x_5 x_9^2$. The second step consits of multiplying $m'(T)$ by $e_2(A \cup B)$, to get a polynomial :

2. $p(T) = m'(T) e_2(A \cup B)$.

If $e_2(A \cup B)$ beolnged to $\mathcal{C}_{333}$ we would be done. Indeed we would then have that

$$x_2 x_6 x_8 x_4 x_5 x_9^2 e_2(A \cup B) \equiv 0 \ in \ \mathcal{R}_{333} \qquad (*)$$

But $(*)$ simply means that in $\mathcal{R}_{333}$ we have:

$$\begin{aligned} x_2 x_6 x_8 x_4^2 x_5^2 x_9^2 = -\big(& x_2 x_6^2 x_8 x_4^2 x_5 x_9^2 + x_2 x_6 x_8^2 x_4^2 x_5 x_9^2 + x_2 x_6^2 x_8 x_4 x_5^2 x_9^2 \\ & + x_2 x_6 x_8^2 x_4 x_5^2 x_9^2 + x_2 x_6^2 x_8^2 x_4 x_5 x_9^2 \big) \ . \end{aligned} \qquad (**)$$

Now looking at the right hand side of equation $(**)$ one sees that each of the monomials corresponds to a standard tableau of shape $(3,3,3)$. Thus we have expressed $x_2 x_6 x_8 x_4^2 x_5^2 x_9^2$ as a linear combination of the elements of the basis $\mathcal{B}^{top}_{333}$. Unfortunately $e_2(A \cup B)$ does not belong to $\mathcal{C}_{333}$, as one can easily see from figure 6, which depicts the admissible pairs $(|S|, r)$.

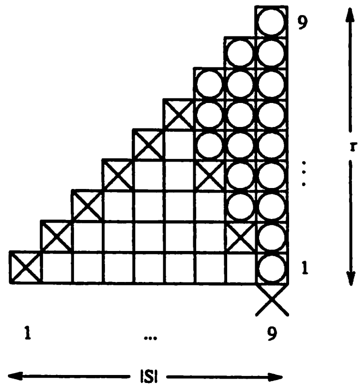


Figure 6

On the other hand according to our remark of section 1, $e_2(X)$ and $e_1(X)$ are certainly present in $\mathcal{C}_{333}$. We claim here that :

$$x_2x_6x_8x_4x_5x_9^2(e_2(X) - x_2e_1(X)) \equiv x_2x_6x_8x_4x_5x_9^2e_2(A \cup B) \; in \; \mathcal{R}_{333} \qquad (2.1).$$

But the left hand side of equation 2.1 is certainly congruent to zero in $\mathcal{R}_{333}$, therefore yielding $(*)$ and consequently $(**)$. We shall prove this assertion in two steps. First observe that the sequence of exponents of a monomial corresponding to any tableau of shape $(3,3,3)$ is given by figure 7(a).

(a) (b)

Figure 7

Thus for any $i \in \{1,3,7,9\}$ we have that

$$x_2x_6x_8x_4x_5x_9^2x_i \equiv 0. \qquad (***)$$

Indeed, it is clear that the corresponding sequence of exponents for these monomials will fill at least one of the *forbidden* squares introduced in lemma 2.1 (see figure 7 b, for the case $x_2x_6x_8x_4x_5x_9^2x_3$). Thus we now have that

$$x_2x_6x_8x_4x_5x_9^2e_2(X) \equiv x_2x_6x_8x_4x_5x_9^2e_2(S)$$

where $S = \{x_2, x_4, x_5, x_6, x_8\}$. One realizes that S differs from the set $A \cup B$ only in the variable x_2. Therefore our next step is to eliminate x_2 of S. For this notice that $e_2(S) - x_2e_1(S)$ has the double effect of removing all pairs x_2x_j (for $j = 4,5,6,8$) from $e_2(S)$ and of adding to it the monomial $-x_2^2$. Thus

$$e_2(S) - x_2e_1(S) = e_2(4,5,6,8) - x_2^2$$

But clearly

$$x_2x_6x_8x_4x_5x_9^2x_2^2 \equiv 0 \text{ in } \mathcal{R}_{333}.$$

We are therefore left with

$$x_2x_6x_8x_4x_5x_9^2e_2(4,5,6,8),$$

which yields equation 2.1 as desired. Observe that it is not always the case that one application of steps 1 and 2 yields a set of *standard monomials* (monomials corresponding to standard tableaux). But as we shall show in theorem 3.1 a recursive application of these two steps will eventually lead to a set of standard monomials. We are now ready to generalize this construction.

Let T be a tableau where the first break of standardness occurs in position k of its $(i+1)^{th}$ row. We shall depict the elements of rows i and $i+1$ as in figure 8.

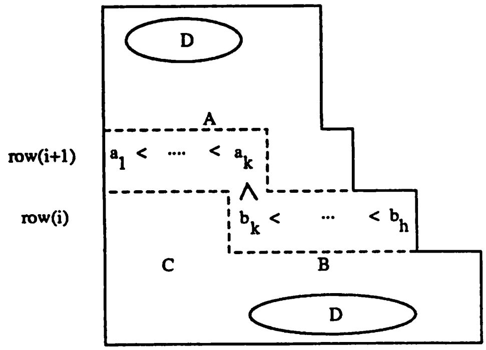


Figure 8

Let D be the set of variables corresponding to the entries of the first $(i-1)$ rows, together with $a_{k+1},\ldots,a_l$ and together with rows $(i+2)$, $(i+3)$ $,\ldots,h(\mu)$. Let $m(D)$ be the monomial obtained from $m(T)$ by removing all the variables not belonging to D. Let $A=\{a_1,\ldots,a_k\}$, $B=\{b_k,\ldots,b_m\}$ and $C=\{b_1,\ldots,b_{k-1}\}$. (See figure 8).

To avoid confusion with indices denote the height of the i^{th} row by h. With this notation $m(T)$ is given by:

$$m(T)=m(D)\Big(\prod_{i\in A}x_i\Big)^h\Big(\prod_{i\in B}x_i\Big)^{h-1}\Big(\prod_{i\in C}x_i\Big)^{h-1}.$$

As in the previous example define a new monomial $m'(T)$ by subtracting one from all the exponents of the variables of $m(T)$ belonging to A. More precisely:

$$1. \qquad m'(T)=m(D)\Big(\prod_{i\in A}x_i\Big)^{h-1}\Big(\prod_{i\in B}x_i\Big)^{h-1}\Big(\prod_{i\in C}x_i\Big)^{h-1}.$$

Now using the principle of inclusion-exclusion we define the polynomial $p(T)$ as follows:

$$2. \qquad p(T)=m'(T)\left(\sum_{i=0}^{k-1}(-1)^i\Big[\Big(\sum_{\substack{j_1<\cdots<j_i\\ j_l\in C}}x_{j_1}\cdots x_{j_i}\Big)e_{k-i}(X)\Big]\right)$$

where $x_{j_1}\ldots x_{j_i}=1$ when $i=0$. Observe that $p(T)$ is congruent to zero in $\mathcal{R}_\mu$. The next proposition is the last step we shall need to prove theorem (3.1)

Proposition 2.2.

$$p(T) \equiv m'(T)e_k(A \cup B) \ in \ \mathcal{R}_\mu.$$

Proof

Observe first that

$$\sum_{\substack{j_1<\dots<j_i \\ j_l \in C}} x_{j_1} \dots x_{j_i} = e_i(C).$$

Thus $p(T)$ is nothing more than :

$$p(T) = m'(T)\sum_{i=0}^{k-1}(-1)^i e_i(C)e_{k-i}(X)$$

where $e_0(C) = 1$. On the other hand we have that

$$\sum_{i=0}^{k-1}(-1)^i e_i(C)e_{k-i}(X) = \sum_{i=0}^{k-1}\left(\prod_{j\in C}(1-tx_j)\Big|_{t^i} \prod_{j\in A\cup B\cup C\cup D}(1+tx_j)\Big|_{t^{k-i}}\right)$$

$$= \left(\prod_{j\in C}(1-tx_j) \prod_{j\in A\cup B\cup C\cup D}(1+tx_j)\right)\Big|_{t^k}$$

$$= \left(\prod_{j\in C}(1-t^2x_j^2) \prod_{j\in A\cup B\cup D}(1+tx_j)\right)\Big|_{t^k}$$

Therefore we have that,

$$p(T) = m(D)\Big(\prod_{i\in A} x_i\Big)^{h-1}\Big(\prod_{i\in B} x_i\Big)^{h-1}\Big(\prod_{i\in C} x_i\Big)^{h-1}\left(\prod_{j\in C}(1-t^2x_j^2) \prod_{j\in A\cup B\cup D}(1+tx_j)\right)\Big|_{t^k}. \qquad (2.2)$$

It is not too difficult to realize that in the right-hand side of equation (2.2) the coefficients of t^k involving variables in D or C will give rise to monomials congruent to zero in $\mathcal{R}_\mu$. Indeed the corresponding sequence of exponents will be outside the permissible squares of the diagram of lemma 2.1. See figure (9), where A' corresponds to the set of variables which had their exponent diminished by one.

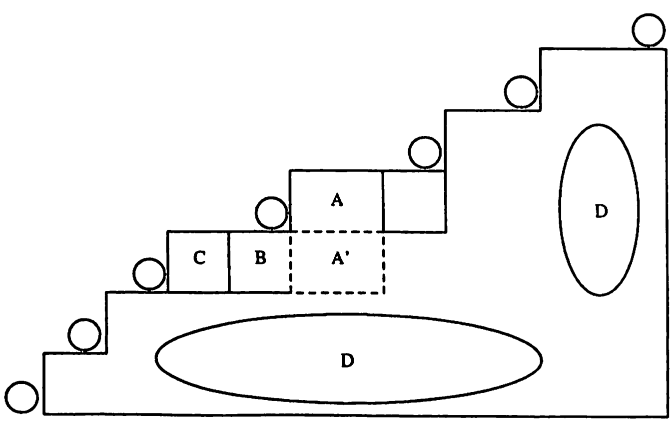


Figure 9

Thus we are left with

$$p(T) \equiv m(D)(\prod_{i\in A} x_i)^{h-1}(\prod_{i\in B} x_i)^{h-1}(\prod_{i\in C} x_i)^{h-1}\Big(\prod_{j\in A\cup B}(1+tx_j)\Big|_{t^k}\Big)$$
$$\equiv m'(T)e_k(A\cup B). \quad \clubsuit$$

As we observed earlier on, $p(T)\equiv 0$ in $\mathcal{R}_\mu$. Thus

$$m'(T)e_k(A\cup B)\equiv 0 \ in \ \mathcal{R}_\mu \ . \tag{2.3}$$

We can now finalize the straightening algorithm.

3. Straightening algorithm in $\mathcal{R}_\mu^{top}$.

We first need a total order on the set of row-increasing tableaux. Define the row-word $w(T)$ of a tableau T, to be the word obtained by reading the successive rows of T from left to right and from bottom to top. This given, order all the row-increasing tableaux of shape μ according to the lexicographic order of their corresponding row-words. For a given partition μ of n let $\{T_1,\dots,T_{n_\mu}\}$ be the set of standard tableaux of that shape, and let $\{m(T_i)\}_{i=1}^{n_\mu}$ be the set of corresponding monomials. As we saw in section 1, a basis for $\mathcal{R}_\mu^{top}$ is given by:

$$\mathcal{B}_\mu^{top}=\{m(T_1),\dots m(T_{n_\mu})\}$$

Let T be a non-standard injective tableau, filled with the integers 1 to n. The monomial $m(T)$ associated to this tableau is an element of $\mathcal{R}_\mu^{top}$, and we want to express it as a linear combination of the elements of $\mathcal{B}_\mu^{top}$. For this we shall give an algorithm (*straightening algorithm*) that will explicitly produce the coefficients $a_i(T)$ in the expansion of $m(T)$:

$$m(T) \ \equiv \sum_{i=1}^{n_\mu} a_i(T)m(T_i) \tag{3.1}$$

This algorithm is, in its essence, similar to Young's straightening law. For more details the reader can consult the original work of A. Young, [8, QSA II p. 95] and [8,QSA III p. 356] or the more recent work of A. Garsia and M. Wachs [GW, proposition 2.2] . We shall adopt here the same notation as in [GW]. Consider the set of row-words of all the row-increasing tableaux of shape μ, ordered according to the lexicographic order of their corresponding row-words. Clearly the first row-word in this ordered set is the one corresponding to the super standard tableau of shape μ. A super standard tableau is a tableau obtained by filling its rows (from bottom to top and from left to right), with the consecutive integers $1,2,\dots n$. For this super-standard tableau T^* equation 3.1 reduces to

$$m(T^*) \ = \ m(T_1)$$

i.e.: $a_i(T) = 0$, for all $i\neq 1$ and $a_1 = 1$. Therefore we shall proceed by induction on this order, and assume that T is the first row-increasing tableau for which equation 3.1 has not yet been established . Assume that the first break of standardness in T occurs in position k of the $(i+1)^{th}$ row where the element in this position is smaller (rather than larger) than the element directly below it. As in section 2, let $A=\{a_1,\dots,a_k\}$ and $B=\{b_k,\dots,b_h\}$. Observe that all the letters of A are smaller than any letter of B. We claim that:

$$m(T)=-\sum_{\tau\neq\epsilon} m(\tau T) \tag{3.2}$$

where the sum is over all permutations τ (distinct from the identity) in $S_{A\cup B}$ where the first $|A|$ letters and the next $|B|$ letters are in increasing order. In two line notation, this simply means that τ has the form:

$$\tau = \begin{bmatrix} A & B \\ \tau_A & \tau_B \end{bmatrix}$$

with τ_A, τ_B a pair of subsets (such that $|A| = |\tau_A|$, and $|B| = |\tau_B|$) partioning $A \cup B$. Observe first that upon proving equation (3.2) our argument will be completed. Indeed each monomial on the right hand side of equation (3.2) corresponds to a tableau τT whose row word is lexicographically smaller than the row word of T. To see this, one should observe the following two facts:

1. the elements of rows $1, 2, \ldots i-1$ have not changed.
2. Acting with any of the permutations τ ($\tau \neq \epsilon$) on T will bring at least one of the $a's$ down to the i^{th} row. But as we observed earlier, all of the $a's$ are smaller than all of the $b's$. This makes the row word $w(\tau T)$ of τT, lexicographically smaller than $w(T)$.

We shall now prove equation (3.2) . Using the construction introduced in section 2, let

$$p(T) = m'(T) \sum_{i=0}^{k-1} (-1)^i e_i(C) e_{k-i}(X)$$

where $m'(T)$ is the monomial obtained from $m(T)$ by subtracting one from the exponents of all the variables of $m(T)$belonging to A. By proposition 2.2 we have that :

$$p(T) \equiv m'(T) e_k(a \cup B),$$

Observe that $p(T)$ is a sum of monomials, among which we find $m(T)$. Indeed in

$$e_k(A \cup B) = \sum_{\substack{j_1 < \cdots < j_i \\ j_k \in A \cup B}} x_{j_1} \ldots x_{j_i},$$

choosing the k elements $a_1, \ldots a_k$ of A, yields that the monomial $m'(T) x_{a_1} \ldots x_{a_k} \in p(T)$. But this monomial is precisely $m(T)$. Thus equation 2.3 yields that

$$m(T) = -\big(m'(T) e'_k(A \cup B)\big)$$

where $e'_k(A \cup B) = e_k(A \cup B) - x_{a_1} \ldots x_{a_k}$. We claim that

$$-\big(m'(T) e'_k(A \cup B)\big) = -\sum_{\tau \neq \epsilon} m(\tau T) \,. \tag{3.3}$$

Indeed a moment of thought reveals that for each choice of k variables $x_{i_1} < \ldots < x_{i_k}$ among $A \cup B$, the corresponding monomial $m'(T) x_{i_1} \ldots x_{i_k}$ of $m'(T) e_k(A \cup B)$ is precisely $m(\tau T)$ where $\tau_A = \{x_{i_1}, \ldots, x_{i_k}\}$ and $\tau_B = \{A \cup B\} \backslash \{x_{i_1}, \ldots x_{i_k}\}$. ♣

We have now shown:

Theorem 3.1 (Straightening algorithm)

For any (injective) tableau T of shape μ the coefficients $a_i(T)$, described in the algorithm above, satisfy :

$$m(T) \equiv \sum_{i=1}^{n_\mu} a_i(T) m(T_i)$$

in $\mathcal{R}_\mu^{top}$ where the sum is over all standard tableaux of shape μ, ordered according to the lexicographic order of their corresponding row-words.

4. Action of S_n on $\mathcal{R}_\mu^{top}$.

Let $\mu = (\mu_1, \ldots, \mu_k)$ be a partition of n. As we mentioned earlier (section 1), there is a natural action of the symmetric group S_n on $\mathcal{R}_\mu$ which presrves the natural grading of $\mathcal{R}_\mu$. Briefly, for any polynomial $P(x_1, \ldots x_n)$ of $\mathcal{R}_\mu$ and for any permutation σ of S_n

$$\sigma P(x_1, \ldots, x_n) = P(x_{\sigma_1}, \ldots, x_{\sigma_n}).$$

We want to discuss here the action of S_n on $\mathcal{R}_\mu^{top}$ via its basis $\mathcal{B}_\mu^{top}$. More precisely for any monomials $m(T)$ of $\mathcal{B}_\mu^{top}$ we are interested in finding its image under the action of a permutation σ of S_n. It is not difficult to see that the images $\sigma m(T)$ can be obtained by acting directly with σ on the standard tableaux T themselves: replace every entry i of T by its image $\sigma(i)$. In this manner for any given standard tableau T of shape μ,

$$\sigma m(T) = m(\sigma T).$$

We shall show that the matrices obtained from the action of S_n on the elements of $\mathcal{B}_\mu^{top}$ are exactly the same ones resulting from the action of S_n on the so called Young's natural units.

We shall first recall the definition of Young's natural units. Again we will use the notation found in [GW]. For a set A of integers define $[A]$ to be the formal sum of all permutations of A. That is,

$$[A] = \sum_{\sigma \in S_A} \sigma$$

if S_A denotes the symmetric group of A. Also let $[A']$ be given by:

$$[A'] = \sum_{\sigma \in S_A} sign(\sigma)\sigma.$$

Note that $[A]$ and $[A']$ can be interpreted as elements of the group algebra of S_n, denoted here by $A(S_n)$. For a given tableau T of shape μ, let $R_1, \ldots, R_k$ denote the rows of T and $C_1, \ldots, C_h$ denote the columns of T. In $A(S_n)$, define the row group of T to be

$$P(T) = [R_1][R_2] \ldots [R_k]$$

and the signed column group of T :

$$N(T) = [C_1][C_2] \ldots [C_h].$$

Next, let $E(T) = P(T)N(T)$ and for any two tableaux T_1 and T_2 of the same shape, define

$$E_{T_1,T_2} = P(T_1)\sigma_{T_1,T_2}N(T_2)$$

where σ_{T_1,T_2} is the permutation which sends T_2 to T_1. Let $T_1, \ldots, T_{n_\mu}$ be the standard tableaux of shape μ ordered according to the lexicographic order of their row-words (see section 3). A. Young showed that for any given partition μ of n, and for each $s = 1, 2, \ldots, n_\mu$ the elements (now called Young's natural units)

$$E_{T_1,T_s},\ E_{T_2,T_s},\ \ldots, E_{T_{n_\mu},T_s}$$

span a subspace of the group algebra $A(S_n)$ which is invariant under left multiplication. Moreover the matrices expressing this action in terms of this basis are the *same* for each s and they give the irreducible representaiton of S_n usually indexed by μ. From now on set $s = 1$. We plan to show that for any given shape μ, the matrices resulting from the action of S_n on $\mathcal{B}_\mu^{top}$ are exactly the same as the one resulting from the action of S_n on

$$E_{T_1,T_1},\ E_{T_2,T_1},\ \ldots, E_{T_{n_\mu},T_1}$$

Thus for any $\sigma \in S_n$, we shall express $\sigma E_{T_i,T_1}$ as a linear combination of the elements of the set $\{E_{T_1,T_1},\ E_{T_2,T_1},\ \ldots, E_{T_{n_\mu},T_1}\}$. For this we shall need a few relations between E_{T_i,T_1} and $E(T)$ (for more details see [GW, eq. 2.3]). Given any two tableaux T_1, T_2 of shape μ, one has

$$\begin{aligned}\sigma_{T_1,T_2}P(T_2) &= P(T_1)\sigma_{T_1,T_2}\\ \sigma_{T_1,T_2}N(T_2) &= N(T_1)\sigma_{T_1,T_2}\\ E_{T_1,T_2} &= E(T_1)\sigma_{T_1,T_2} = \sigma_{T_1,T_2}E(T_2)\end{aligned}$$

Therefore $\sigma E_{T_i,T_1}$ becomes:

$$\begin{aligned}\sigma\sigma_{T_i,T_1}E(T_1) &= \sigma_{T_*,T_1}E(T_1)\\ &= \sigma_{T_*,T_1}P(T_1)N(T_1) = P(T_*)\sigma_{T_*,T_1}N(T_1)\\ &= P(T_*)N(T_*)\sigma_{T_*,T_1} = E(T_*)\sigma_{T_*,T_1}\end{aligned}$$

where T_* is the tableau σT_i. The reason for replacing $\sigma E_{T_i,T_1}$ with $E(T_*)\sigma_{T_*,T_1}$, is that Young also give an algorithm for expanding any element $E(T)$ (for any injective tableau T) as a linear combination of elements of $\{E_{T_i,T},\ E_{T_2,T},\ldots, E_{T_{n_\mu},T}\}$. More precisely, this procedure, called Young's straightening formula, is stated as follows in [GW, prop. 2.2]

Lemma 4.1 Young's straightening algorithm

For any (injective) tableau T of shape μ there are some coefficients $a_i(T)$ giving

$$E(T) = \sum_{i=1}^{n_\mu} a_i(T)E_{T_i,T}. \tag{4.1}$$

First observe that using equation 4.1 together with equation 4.2 we get

$$\begin{aligned}\sigma E_{T_i,T_1} &= E(T_*)\sigma_{T_*,T_1}\\ &= \sum_{i=1}^{n_\mu} a_i(T_*)E_{T_i,T_*}\sigma_{T_*,T_1}\\ &= \sum_{i=1}^{n_\mu} a_i(T_*)E_{T_i,T_1}.\end{aligned}$$

Therefore in order to expand $\sigma E_{T_i,T_1}$ as a linear combination of Young's natural units $\{E_{T_i,T_1}\}_{i=1}^{n_\mu}$ we need only to find the coefficients $a_i(T_*)$ in the expansion of $E(T_*)$ given in equation (4.1). The remarkable fact is that Young's proof of lemma 4.1 is algorithmic and analogous to our own straightening algorithm given in the proof of theorem 3.1. But thanks to A. Garsia and M. Wachs, Young's straightening algorithm has been eloquently reproduced (as well as generalized to skew-shaped tableaux) in [GW, prop. 2.2]. Thus a glance at the proof of proposition 2.2 of [GW] will convince the reader that the coefficients $a_i(T)$ of equation 4.1 are precisely the same ones appearing in our previous equation 3.1. Thus keeping in mind the remark following the definitions of Young's natural units we have proved:

Theorem 4.2

The action of S_n on the basis $\mathcal{B}_\mu^{top}$ of $\mathcal{R}_\mu^{top}$ yields the same matrices as the action of S_n on Young's basis of natural units for the irreducible representation indexed by μ.

The author would like to thank A. Garsia for suggesting this problem and for helpful discussions as well, as S. Sundaram for her invaluable comments.

Bibliography

[1] C. de Concini, C. Procesi : Symmetric Functions, Conjugacy Classes and the Flag Variety, Inv. Math. 64, 203–279, (1981).

[2] A.M. Garsia, N. Bergeron: On Certain Spaces of Harmonic polynomials, preprint (1991).

[3] A.M. Garsia, C. Procesi: On Certain Graded S_n-Modules and the q-Kostka Polynomials, to appear in Advances in Math.

[4] A.M. Garsia, M. Wachs: Combinatorial aspects of Skew Representations of the Symmetric Group, J. Comb. Theory, A, 50, 47–81 (1989).

[5] H. Kraft: Conjugacy Classes and Weyl Group Representations, Proc. 1980 Torun Conf. Poland, Astérisque, 87–88, 191–205 (1981).

[6] I.G. Macdonald,: Symmetric functions and Hall polynomials, Clarendon Press, Oxford (1979).

[7] T. Tanisaki: Defining ideals of the closures of the conjugacy classes and representations of the Weyl groups, Tôhoku Math Journ., 34, 575–585 (1982).

[8] A. Young: On quantitative substitutional analysis (sixth paper), The collected papers of A. Young, University of Toronto Press, 434–435 (1977).

THE RANDOM GENERATION OF UNDERDIAGONAL WALKS

E. Barcucci, R. Pinzani, R. Sprugnoli
Dipartimento di Sistemi e Informatica
Università di Firenze, Firenze (Italy)

ABSTRACT

In this paper, we propose an algorithm for the random generation of underdiagonal walks. We consider the plane walks which are made up of different kinds of east, north-east and north steps and which start from the origin and remain under the main diagonal. The algorithm is very simple: it randomly generates plane walks and refuses the walks crossing the diagonal. We prove that the algorithm works in linear time with respect to the walks' length when the number of different kinds of east steps is greater than, or equal to, the number of different kinds of north steps. Finally, some results of our experiments are reported in order to support our theoretical results with empirical evidence.

1. INTRODUCTION

Various studies (see, for instance, [5, 6, 7]) have been made on one-dimensional walks made up of three kinds of unitary steps: right (r), left (l) and in-place (s). Particular attention has been given to walks that begin at the origin O and never go to its left. This means that in each subwalk which begins at the origin O, the number of right-hand steps must be greater than, or equal to, the number of left-hand steps.

These walks can be codified by means of some words defined on an alphabet having three symbols, such as $\{r, l, s\}$. If we denote the number of letters r (l) in word w by $|w|_r$ ($|w|_l$), then the language of the words codifying the walks is:

$$\mathcal{L} = \{w \in \{r, l, s\}^* \mid \ |w'|_r \geq |w'|_l \text{ for each } w' \text{ prefix of } w\}$$

The words $w \in \mathcal{L}$ are also called Motzkin left factors because they are the prefixes of the Motzkin words, i.e., words $w \in \{r, l, s\}^*$, such that $|w|_r = |w|_l$ and $|w'|_r \geq |w'|_l$ for each w' prefix of w. This kind of walk has been very thoroughly studied because these walks correspond biunivocally to single-rooted directed animals [6].

The walks on the line can also be represented two-dimensionally by recording the time on the abscissa and the moves on the ordinates. As a result, the elementary steps corresponding to the right, left and in-place steps are north-east, south-east and east, respectively. For example, the *rrslsrl* walk can be represented as in Figure 1. In this case, the condition that a one-dimensional walk mustn't go to the left of the origin means that it mustn't go under the x axis.

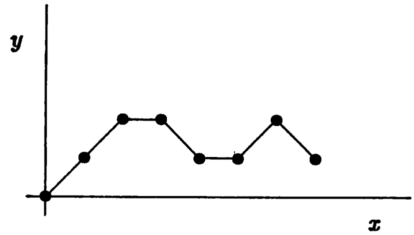


Fig. 1 - The 7-walk corresponding to *rrslsrl*

Another way of representing this kind of walk on a plane is to consider the east, north, and north-east steps as elementary steps corresponding to r, l, s. In this case, the walk corresponding to the preceding example is represented in Figure 2.

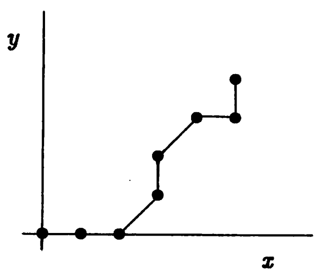


Fig. 2 - The underdiagonal walk corresponding to *rrslsrl*

In this case, the condition that the one dimensional walk mustn't go to the left of O means that the walk mustn't go over the $y = x$ diagonal. This is why these walks are also called ***underdiagonal*** walks.

In [1] we introduced a linear algorithm for the random generation of single-rooted directed animals whose first step is to generate a word in the $\mathcal{L}$ language. In the present study, we examine a more general class of underdiagonal walks made up of several kinds of east, north and north-east steps and we introduce an algorithm that generates them randomly. This algorithm is shown to be linear when the number of east steps is greater than, or equal to, the number of north steps.

2. UNDERDIAGONAL WALKS

We examine underdiagonal walks in which a kinds of east steps, b kinds of north-east steps and c kinds of north steps can be used. This type of walk is usually called *coloured walk*. Walks made up of n steps, (n-long walks or n-walks), can be codified by n-long words (n-words) on the alphabet $\mathcal{A} = \{x_1, x_2, \ldots, x_a, z_1, \ldots, z_b, y_1, \ldots, y_c\}$, in which the x_i, z_i and y_i represent the east, north-east and north steps.

For a walk to be underdiagonal, the number of north steps in each subwalk beginning at the origin must be less than, or equal to, the number of east steps. Consequently, if $w \in \mathcal{A}^*$ is a word that codifies an underdiagonal walk, we obtain

$$\sum_{k=1}^{c} |w'|_{y_k} \leq \sum_{k=1}^{a} |w'|_{x_k}$$

for each w' prefix of w.

It is then relatively simple to provide a non-ambiguous context-free grammar that generates the language $\mathcal{S}$ of the words codifying the underdiagonal walks. The non-terminal symbol D corresponds to the underdiagonal walks ending on the main diagonal; as noted previously, the x and y symbols should balance. The non-terminal symbol F corresponds to the walks not ending on the main diagonal; each of these walks can be uniquely decomposed into a sequence of D-walks joined together by some positive number of x steps. The total number of these x steps represents the distance of the ending point of the walk from the main diagonal. Finally, the non-terminal symbol S corresponds to all the underdiagonal walks, each of which may end on the main diagonal (a D-walk) or not (an F-walk).

$$
\begin{aligned}
S &::= D \mid F \\
F &::= DAD \mid DAF \\
D &::= \epsilon \mid BD \mid ADCD \\
A &::= x_1 \mid x_2 \mid \ldots \mid x_a \\
B &::= z_1 \mid z_2 \mid \ldots \mid z_b \\
C &::= y_1 \mid y_2 \mid \ldots \mid y_c
\end{aligned}
$$

By applying Schutzenberger's method [8], we obtain the following system:

$$
\begin{aligned}
S(t) &= D(t) + F(t) \\
F(t) &= A(t)D^2(t) + A(t)D(t)F(t) \\
D(t) &= 1 + B(t)D(t) + A(t)C(t)D^2(t)
\end{aligned}
$$

$$A(t) = at$$
$$B(t) = bt$$
$$C(t) = ct$$

from which we can get $D(t)$ and $S(t)$. The expression for $D(t)$ is the following:

$$D(t) = \frac{1-bt-\sqrt{\Delta}}{2act^2}$$

in which $\Delta = (1-(b-2\sqrt{ac})t)(1-(b+2\sqrt{ac})t)$.

$D(t) = \sum_n d_n t^n$ is the generating function of the words that codify the underdiagonal n-walks that end on the diagonal, that is, the w words, such that:

$$\sum_{k=1}^{c} |w|_{y_k} = \sum_{k=1}^{a} |w|_{x_k} \quad \text{and}$$

$$\sum_{k=1}^{c} |w'|_{y_k} \leq \sum_{k=1}^{a} |w'|_{x_k} \text{ for each } w' \text{ prefix of } w$$

It is worth noting that in the $D(t)$ function, a and c only appear as the product ac. This implies that the number of n-walks with a east steps, b north-east steps and c north steps is equal to the number of n-walks with c east steps, b north-east steps and a north steps. In order to verify this, it is sufficient to run the walks backwards.

For the generating function of the underdiagonal n-walks we obtain the following expression:

$$S(t) = \frac{1}{2at}\frac{1-(b+2a)t-\sqrt{\Delta}}{(a+b+c)t-1}$$

Because of this generating function, when $a, b, c > 0$ there is no closed form for $s_n = [t^n]\, S(t)$. Therefore, we now determine an asymptotic expression for s_n which will be important for our remarks on the random generation of underdiagonal walks in Section 3. Since:

$$s_n = \frac{1}{2a}\,[t^{n+1}]\,\frac{1-(b+2a)t-\sqrt{\Delta}}{(a+b+c)t-1} \tag{2.1}$$

let us consider this last function. It presents three singularities: a pole at $t = (a+b+c)^{-1}$ and two algebraic singularities at $t = (b+2\sqrt{ac})^{-1}$ and $t = (b-2\sqrt{ac})^{-1}$. By our hypothesis on a, b, c, the last singularity cannot have minimum modulus; besides, since $a+c-2\sqrt{ac} = (\sqrt{a}-\sqrt{c})^2 \geq 0$, we always have $a+b+c \geq b+2\sqrt{ac}$, and the equality holds if and only if $a = c$. So $t = (a+b+c)^{-1}$ is the singularity of minimum modulus unless $a = c$. We have:

$$\sqrt{\Delta}\,\Big|_{t=1/(a+b+c)} = \frac{|c-a|}{a+b+c}$$

and when $c > a$ the numerator in (2.1) is annulled for $t = (a+b+c)^{-1}$. Hence, for $c > a$, this quantity is not a pole for the function $S(t)$. We thus have three different cases, according to whether a is greater than, equal to or less than c.

case $a > c$

Since in this case $t = (a+b+c)^{-1}$ is a first order pole, we can apply Darboux's method and we immediately obtain:

$$s_n \sim \frac{1}{2a}\Big(1-\frac{b+2a}{a+b+c}-\frac{a-c}{a+b+c}\Big)(a+b+c)^{n+1} = \frac{a-c}{a}(a+b+c)^n \qquad (2.2)$$

We could now remove the pole from the generating function and obtain a more precise approximation for s_n by the method of subtracted singularities. However, as we shall see in Section 3, formula (2.2) is usually sufficient for our purposes.

case $a = c$

In this case, $a+b+c = b+2\sqrt{ac} = b+2a$ and $\Delta = (1-(b-2a)t)(1-(b+2a)t)$. The function $S(t)$ simplifies to:

$$S(t) = \frac{1}{2at}\Big(\sqrt{\frac{1-(b-2a)t}{1-(b+2a)t}}-1\Big)$$

The minimum modulus singularity is now the algebraic singularity at $t=(b+2a)^{-1}$. In this case, we need a more precise approximation of s_n and, therefore, instead of applying the Darboux's method, we develop $S(t)$ around the dominating singularity and obtain an asymptotic development:

$$tS(t) \sim \frac{1}{2a}\sqrt{\frac{1-(b-2a)t}{1-(b+2a)t}} \sim$$

$$\sim \frac{1}{a}\sqrt{\frac{a}{b+2a}}\Big((1-(b+2a)t)^{-1/2}+\frac{b-2a}{8a}(1-(b+2a)t)^{1/2}-\frac{(b-2a)^2}{128a^2}(1-(b+2a)t)^{3/2}\Big)$$

We can now extract the coefficient of t^{n+1} from this expression:

$$s_n \sim \frac{1}{\sqrt{a(b+2a)}}\frac{(b+2a)^{n+1}}{\sqrt{\pi(n+1)}}\Big(1-\frac{b}{16a(n+1)}+\frac{16a^2-3b^2}{512a^2(n+1)^2}\Big) \qquad (2.3)$$

When $a = b = c = 1$, this formula coincides with the one found in [1], which approximates the number of single-rooted directed animals having $n+1$ nodes.

<u>case $a < c$</u>

This is the most difficult case. As observed before, the dominating singularity is the algebraic singularity at $t = (b+2\sqrt{ac})^{-1}$. As in the case $a = c$, the best way to obtain an asymptotic approximation for s_n is to develop $S(t)$ around this singularity. The use of a formal system, such as MAPLE [3], may help in performing the necessary computations. In any case, we obtain the following asymptotic development:

$$tS(t) \sim -\frac{4\sqrt{ac}\sqrt{b+2\sqrt{ac}}}{a(\sqrt{c}-\sqrt{a})^2}(1-\gamma t)^{1/2}\Bigg(1+\Big(\frac{b-2\sqrt{ac}}{8\sqrt{ac}}+\frac{a+b+c}{(\sqrt{c}-\sqrt{a})^2}\Big)(1-\gamma t)+$$

$$+\left(\frac{(a+b+c)^2}{(\sqrt{c}-\sqrt{a})^4}+\frac{b-2\sqrt{ac}}{8\sqrt{ac}}\frac{a+b+c}{(\sqrt{c}-\sqrt{a})^2}-\frac{(b-2\sqrt{ac})^2}{128ac}\right)(1-\gamma t)^2+\ldots\Bigg)$$

where γ denotes the reciprocal of the dominating singularity, i.e., $\gamma = b+2\sqrt{ac}$. In order to simplify our notations, let us set:

$$K = \frac{4\sqrt{ac}\sqrt{b+2\sqrt{ac}}}{a(\sqrt{c}-\sqrt{a})^2} \qquad C_1 = \frac{b-2\sqrt{ac}}{8\sqrt{ac}}+\frac{a+b+c}{(\sqrt{c}-\sqrt{a})^2}$$

$$C_2 = \frac{(a+b+c)^2}{(\sqrt{c}-\sqrt{a})^4}+\frac{b-2\sqrt{ac}}{8\sqrt{ac}}\frac{a+b+c}{(\sqrt{c}-\sqrt{a})^2}-\frac{(b-2\sqrt{ac})^2}{128ac}$$

We then find the asymptotic approximation:

$$s_n = \frac{2K}{n+1}\binom{2n}{n}\left(\frac{\gamma}{4}\right)^{n+1}\left(1-\frac{3C_1}{2n-1}+\frac{15C_2}{(2n-1)(2n-3)}+O(n^{-3})\right)=$$

$$=\frac{K}{2}\frac{(b+2\sqrt{ac})^{n+1}}{(n+1)\sqrt{\pi n}}\left(1-\frac{1+12C_1}{8n}+O(n^{-2})\right) \tag{2.4}$$

The main value will be used in Section 3, whilst the corrections will be useful in the remarks of Section 4.

3. RANDOM GENERATION OF UNDERDIAGONAL WALKS

In order to generate a random underdiagonal n-walk, we only have to generate a random n-word in $\mathcal{S}$. We propose the following algorithm:

1) The letters of the word are generated one after another by taking them out of $\{x_1, x_2, \ldots, x_a, z_1, \ldots, z_b, y_1, \ldots, y_c\}$ (for example, by generating a random integer number k included between 1 and $a+b+c$ and then by taking x_k if $k \leq a$, z_{k-a} if $a < k \leq a+b$ and y_{k-a-b} if $k > a+b$);

2) The difference $\sum_{i=1}^{a} |w|_{x_i} - \sum_{j=1}^{c} |w|_{y_j}$ is taken into account;

3) If this difference becomes negative before generating n letters, the prefix generated up to then is discarded and we start from the beginning again;

4) If n letters are generated before the difference is less than zero, they constitute the word desidered.

We want to calculate the average number of generated letters necessary for obtaining an n-word (i.e., the expected number of calls to a random number generation routine *random*). It will be shown that if $a \geq c$, the algorithm is linear and that it is not so if $a < c$.

Generally speaking, if we want to obtain an n-word in $\mathcal{S}$, some words shorter than n not belonging to $\mathcal{S}$ are generated first and then an n-word in $\mathcal{S}$ is generated.

Let $p_{n,k}$ be the probability that k ($k \geq n$) calls to *random* must be made to generate an n-word in $\mathcal{S}$. If we determine the probability generating function $P_n(t) = \sum_k p_{n,k} t^k$ for each $n \in \mathbb{N}$, we are able to evaluate the average number avg_n of characters generated in order to obtain an n-word in $\mathcal{S}$ because $avg_n = P_n'(1)$. By using the same method, we could also calculate the variance because $var_n = P_n''(1) + P_n'(1) - (P_n'(1))^2$.

We use D_k for denoting the language of the words that codify the k-walks ending on the diagonal, that is the k-words whose number of x letters is equal to its number of y letters. We now go on to examine the languages $N_1, N_2, N_3, \ldots$ of the negative words, that is, the k-words ($k \leq i$ for N_i) that interrupt the generation of a word in $\mathcal{S}$. We have:

$$
\begin{aligned}
C &::= y_1 \mid y_2 \mid \ldots \mid y_c \\
N_1 &::= C \\
N_2 &::= C \mid D_1 C \\
N_3 &::= C \mid D_1 C \mid D_2 C \\
&\ldots \\
N_n &::= C \mid D_1 C \mid D_2 C \mid \ldots \mid D_{n-1} C
\end{aligned}
$$

from which we get

$$N_n(t) = c \sum_{k=0}^{n-1} d_k t^{k+1} \qquad \text{where } d_k = [t^k]\, D(t)$$

When we want to generate a word in $\mathcal{S}$ by means of the algorithm, we generate a (possibly empty) sequence of words in N_n followed by a word in $\mathcal{S}$. As a result, the language L_n of the sequence of letters generating the codifications of underdiagonal n-walks is defined by:

$$R_n ::= \epsilon \mid N_n R_n$$
$$L_n ::= R_n S_n$$

in which S_n is the language of n-words in $\mathcal{S}$. We thus get:

$$R_n(t) = \frac{1}{1-N_n(t)} = \frac{1}{1-c\sum_{k=0}^{n-1} d_k t^{k+1}}$$

and

$$L_n(t) = \frac{s_n t^n}{1-c\sum_{k=0}^{n-1} d_k t^{k+1}}$$

Since every letter of $\mathcal{A}$ is generated with a probability of $1/(a+b+c)$, we can obtain the probability generating function as follows:

$$P_n(t) = L_n(t/(a+b+c)) = \frac{s_n t^n}{(a+b+c)^n - c\sum_{k=0}^{n-1} d_k (a+b+c)^{n-k-1} t^{k+1}}$$

We now prove that $P_n(1) = 1$. We can write

$$S(t) = \frac{1}{2at}\,\frac{1-(b+2a)t-\sqrt{\Delta}}{(a+b+c)t-1} = \frac{1}{1-(a+b+c)t} - ct\,\frac{1-bt-\sqrt{\Delta}}{2act^2}\,\frac{1}{(a+b+c)t-1}$$

From this it follows that:

$$s_n = (a+b+c)^n - c\sum_{k=0}^{n-1} d_k (a+b+c)^{n-k-1}$$

and therefore $P_n(1) = 1$.

If we indicate $Q_n(t) = (a+b+c)^n - c\sum_{k=0}^{n-1} d_k (a+b+c)^{n-k-1} t^{k+1}$, we can write $P_n(t) = s_n t^n / Q_n(t)$. Therefore $Q_n(1) = s_n$ and:

$$avg_n = P_n'(1) = \frac{n s_n Q_n(1) - s_n Q_n'(1)}{Q_n^2(1)} = \frac{n s_n^2 - s_n Q_n'(1)}{s_n^2} = n - \frac{Q_n'(1)}{s_n}$$

From the preceding formula for $Q_n(t)$ we immediately find:

$$Q_n'(1) = -c\sum_{k=0}^{n-1} (k+1)\, d_k (a+b+c)^{n-k-1} = -q_n$$

This allows us to go on to the generating function $q(t) = \sum_{k=0}^{\infty} q_n t^n$, which is the convolution of the derivative of $D(t)$ with the geometric series $(1-(a+b+c)t)^{-1}$. In fact:

$$q(t) = ct\frac{\mathrm{d}}{\mathrm{d}t}(tD(t))\frac{1}{1-(a+b+c)t} = \frac{1-bt-\sqrt{\Delta}}{2at\sqrt{\Delta}\,(1-(a+b+c)t)} \tag{3.1}$$

Here we used the explicit formula for $D(t)$ found in Section 2 and a great deal of routine computations performed by computer.

As for $S(t)$, here again we have three singularities and since the numerator of $q(t)$ is not annulled for $t = (a+b+c)^{-1}$, it may seem that we only have two cases. Actually, the remarks to be done and the exact results obtained for $a > c$ and $a < c$ differ from each other very significantly and we are now going to examine the three different cases, as we did in Section 2.

case $a > c$

The dominating singularity for $q(t)$ is a first order pole at $t = (a+b+c)^{-1}$. The residual at this point is easily found and we have:

$$q_n \sim \frac{c}{a(a-c)}(a+b+c)^{n+1}$$

This value is to be divided by s_n, as computed for the corresponding case in Section 2. We immediately find:

$$avg_n \sim n + \frac{c(a+b+c)}{(a-c)^2} \tag{3.2}$$

that is, the expected number of calls to *random* is n plus a constant per generation.

case $a = c$

In this case, the dominating singularity is the algebraic singularity at $t = (b+2a)^{-1}$, and the formula for $q(t)$ simplifies as follows:

$$2atq(t) = \frac{1-bt}{(1-(b+2a)t)^{3/2}(1-(b-2a)t)^{1/2}} - \frac{1}{1-(a+b+c)t}$$

Again, we develop the first term around the singularity and find:

$$\frac{1-bt}{(1-(b+2a)t)^{3/2}(1-(b-2a)t)^{1/2}} = \sqrt{\frac{a}{b+2a}}\,(1-(b+2a)t)^{3/2}\Big(1+$$

$$+\frac{3b+2a}{8a}(1-(b+2a)t) - \frac{(6a+5b)(b-2a)}{128a^2}(1-(b+2a)t)^2 + \ldots\Big)$$

We can now extract the coefficient of t^{n+1} and obtain:

$$q_n \sim \frac{2n+3}{2\sqrt{a(b+2a)}}\binom{2n+2}{n+1}\left(\frac{b+2a}{4}\right)^{n+1}\left(1+\frac{3b+2a}{8a(2n+3)}+\frac{(6a+5b)(b-2a)}{128a^2(2n+3)(2n+1)}+\ldots\right)+$$

$$-\frac{(b+2a)^{n+1}}{2a}$$

At this point, obtaining avg_n by formula (3.1) is routine procedure:

$$avg_n = 2n - \frac{1}{2}\sqrt{\frac{b+2a}{a}}\sqrt{\pi(n+1)} + \frac{6a+b}{4a} + o(1) \tag{3.3}$$

This means that the expected number of calls to *random* is something less than $2n$ per generation.

<u>case $a < c$</u>

As in the case of $a > c$, the dominating singularity is the first order pole at $t = (a+b+c)^{-1}$, but the condition $a < c$ implies some difference in the constant. In fact, we have:

$$q_n \sim \frac{(a+b+c)^{n+1}}{c-a}$$

By using the value of s_n found in formula (2.4), we have:

$$avg_n \sim n+(n+1)\sqrt{n}\left(\frac{a+b+c}{b+2\sqrt{ac}}\right)^{n+1}\frac{\sqrt{\pi}\,2a(\sqrt{c}-\sqrt{a})^2}{\sqrt[4]{ac}\,\sqrt{b+2\sqrt{ac}}\,(c-a)}\left(1+\frac{1+12C_1}{8n}\right) \tag{3.4}$$

Consequently, the expected number of calls to *random* grows exponentially with n, and our method cannot be used to randomly generate this kind of underdiagonal walks in an efficient way.

It may be interesting to observe that by using the method of subtracted singularities it is possible to obtain a more accurate approximation of q_n. This explicitly shows that the error introduced by the value of q_n also introduces an error in the order of $O((b+2\sqrt{ac})^n)$, which is exponentially smaller than the main value:

$$q_n \sim \frac{(a+b+c)^{n+1}}{c-a} - \frac{\sqrt{b\sqrt{ac}+2ac}}{2a(\sqrt{c}-\sqrt{a})^2}\left(\frac{b+2\sqrt{ac}}{4}\right)^{n+1}\binom{2n+2}{n+1}+$$

$$+\frac{(c+a)(b+2\sqrt{ac})^{3/2}}{4a\,\sqrt[4]{ac}\,(\sqrt{c}-\sqrt{a})^4}\left(\frac{b+2\sqrt{ac}}{4}\right)^{n+1}\frac{1}{2n+1}\binom{2n+2}{n+1}$$

4. EXPERIMENTAL RESULTS

We performed a series of experiments both to check the validity of our approach to the random generation of underdiagonal walks and to give empirical evidence that we can actually generate random walks in linear time when $a \geq c$. In our experiments, we use the average value of the difference:

$$\delta(w) = \sum_{i=1}^{a} |w|_{x_i} - \sum_{j=1}^{c} |w|_{y_j}$$

as an indicator to prove the randomness of the generated words; $\delta(w)$ corresponds to the distance from the main diagonal (measured on the x axis) of the end point of the walk codified by w. We computed the average value of $\delta(w)$ for the n-words in $\mathcal{S}$, that is

$$\delta_n = \sum_{w \in S_n} \delta(w)/s_n$$

by using another application of Schutzenberger's method to the grammar of Section 2. We introduce a new indeterminate u, which counts every x symbol positively and every y symbol negatively and we obtain:

$$\begin{aligned}
S(t,u) &= D(t,u) + F(t,u)\\
F(t,u) &= A(t,u)D^2(t,u) + A(t,u)D(t,u)F(t,u)\\
D(t,u) &= 1 + B(t)D(t,u) + A(t,u)C(t,u)D^2(t,u)\\
A(t,u) &= atu\\
B(t) &= bt\\
C(t,u) &= ctu^{-1}
\end{aligned}$$

In this way, the coefficient $s_{n,k}$ of $t^n u^k$ in $S(t,u)$ represents the number of n-walks ending at a distance k from the main diagonal and it is easy to evaluate $S(t,u)$:

$$S(t,u) = \frac{1-(b+2au)t-\sqrt{(1-bt)^2-4act^2}}{2at(atu^2+btu+ct-u)}$$

What we need is the sum $\sum_{k=0}^{n} ks_{n,k}$, representing the total distance from the main diagonal of all the n-long walks. The simplest way to obtain this quantity is to note that:

$$\sum_{k=0}^{n} ks_{n,k} = [t^n] \left.\frac{\partial}{\partial u}S(t,u)\right|_{u=1}$$

By performing the necessary computations, we eventually find:

$$\sum_{k=0}^{n} ks_{n,k} = [t^n]\left(\frac{1}{1-(a+b+c)t} - \frac{(1-(b+2a)t)\left(1-(b+2a)t-\sqrt{\Delta}\right)}{2at(1-(a+b+c)t)^2}\right) \qquad (4.1)$$

At this point, we only need to extract the coefficient of t^n from this generating function. The singularities are the same as for $S(t)$, and, therefore, we have three different cases according to whether a is less than, equal to, or greater than c. We discuss these three cases separately.

case $a > c$

The dominating singularity of (4.1) is $t = (a+b+c)^{-1}$, a second order pole. By using Darboux's method, we easily obtain:

$$\sum_{k=0}^{n} ks_{n,k} \sim \frac{a-c}{a}(a+b+c)^{n-1}(n+1) + (a+b+c)^n$$

and, therefore, by dividing by s_n:

$$\delta_n = \frac{a-c}{a+b+c}(n+1) + \frac{a}{a-c} + o(1)$$

a quantity linearly increasing with n.

We performed a series of experiments by generating 100,000 n-words for $n = 1000$ and various combinations of a, b, c ($a > c$). The results are summarized in Figure 3, and the experimental data agree with the expected values to a large extent.

			calls to *random*		distance from the diagonal	
a	b	c	expected	experim.	expected	experim.
2	1	1	1004	1004.38	252.25	252.89
3	1	1	1001.25	1001.25	401.9	401.07
3	1	2	1012	1012.07	169.83	170.70
3	2	2	1014	1014.18	146	146.80
4	1	1	1000.67	1000.67	501.83	500.73
4	2	3	1027.00	1026.88	115.22	117.27
4	3	2	1004.50	1004.52	224.44	224.17
5	1	1	1000.44	1000.44	573.25	571.96
5	3	4	1048	1047.91	88.42	91.38
5	4	3	1009	1009.04	169.33	169.78

Fig. 3 - Experimental results for $a > c$

We wish to point out that, by formula (3.2), when the constant $K = c(a+b+c)(a-c)^{-2}$ is high, and we limit ourselves to experiments with $n \approx K$, we obtain seemingly erroneous results. In Figure 4 we show some examples of this.

				calls to *random*		distance from the diagonal	
a	b	c	n	expected	experim.	expected	experim.
20	20	17	1000	1107.67	1101.81	59.35	63.71
			10000	10107.67	10107.61	533.04	537.88
20	10	18	1000	1216.00	1180.06	51.71	58.18
			10000	10216.00	10213.46	426.81	436.62
20	20	18	1000	1261.00	1207.61	44.52	50.29
			10000	10261.00	10259.92	354.86	363.03
20	1	19	1000	1760.00	1369.73	45.03	50.94
			10000	10760.00	10729.70	270.03	290.13
20	10	19	1000	1931.00	1401.69	40.43	44.75
			10000	10931.00	10880.27	224.10	243.65
20	20	19	1000	2121.00	1433.90	36.97	39.63
			10000	11121.00	11072.43	189.51	206.81

Fig. 4 - Experimental results for $a > c$

case $a = c$

As we already know, when $a = c$ we have $a+b+c = b+2\sqrt{ac} = b+2a$, and (4.1) has an algebraic, dominating singularity at $t = (b+2a)^{-1}$. The same formula (4.1) simplifies and we obtain:

$$\sum_{k=0}^{n} ks_{n,k} = [t^n]\left(\frac{1}{1-(b+2a)t} - S(t)\right) = (b+2a)^n - s_n$$

Then we divide by s_n (formula (3.3)):

$$\delta_n = \frac{(b+2a)^n}{s_n} - 1 \sim \sqrt{\frac{a}{b+2a}}\sqrt{\pi(n+1)}\left(1 + \frac{b}{16a(n+1)} + \frac{5b^2-16a^2}{512a^2(n+1)^2}\right) - 1$$

and this time the average distance from the main diagonal only increases as $\sqrt{n}$ does. This is also the case of Motzkin words, and the previous formula for δ_n coincides with the one given in [1] when we set $a = b = c = 1$.

In this case, too, the experimental results agree with the expected values, as shown in Figure 5. It is worth noting that the words in the first two lines in the table correspond to the directed animals over the square and triangular lattices [7] respectively.

			calls to *random*		distance from the diagonal	
a	b	c	expected	experim.	expected	experim.
1	1	1	1953.19	1956.65	31.38	31.37
1	2	1	1945.92	1846.10	27.04	26.15
1	4	1	1933.82	1926.40	21.90	22.12
1	8	1	1914.83	1914.24	16.74	16.84
10	1	10	1960.90	1959.30	37.70	37.84
10	5	10	1957.30	1945.50	34.47	34.52
10	10	10	1953.19	1955.86	31.38	31.44
10	15	10	1949.42	1948.94	28.98	29.04
20	15	20	1955.20	1946.96	32.82	32.87
20	20	20	1953.19	1948.79	31.38	31.47

Fig. 5 - Experimental results for $a = c$

<u>case $a < c$</u>

As usual, this is the most difficult case. Formula (4.1) can be written as:

$$\frac{2at(1-(a+b+c)t) - (1-(b+2a)t)\left(1-(b+2a)t-\sqrt{\Delta}\right)}{2at(1-(a+b+c)t)^2}$$

where the numerator is annulled for $t = (a+b+c)^{-1}$. By using de l'Hospital's rule, we discover that $t = (a+b+c)^{-1}$ is not a simple pole either, and, therefore, the dominating singularity is $t = (b+2\sqrt{ac})^{-1}$. We can now either expand everything around this point or use the implicit function method as described by Bender (see [2], theorem 5). This only gives the main term but it also somewhat reduces the number of necessary computations, which could be performed by computer. In any case, we find:

$$\sum_{k=0}^{n} ks_{n,k} \sim \frac{4\sqrt{ac}\,\sqrt{b+2\sqrt{ac}}}{\sqrt{a}(\sqrt{c}-\sqrt{a})^3}\,\frac{(b+2\sqrt{ac})^{n+1}}{(n+1)\sqrt{\pi(n+1)}}$$

Therefore, we divide by s_n and obtain:

$$\delta_n = \frac{2\sqrt{a}}{\sqrt{c}-\sqrt{a}} + o(1)$$

This means that the average distance from the main diagonal tends to become constant. In Figure 6, we summarize some of our experiments. Obviously, avg_n increases exponentially with n and the generation time may become rather prohibitive. Moreover, we can see in Figure 6 that experimental data are significantly different from expected values but this however is not particularly surprising.

			calls to *random*		distance from the diagonal	
a	b	c	expected	experim.	expected	experim.
5	6	6	39916.4	56133.3	20.95	14.99
7	8	8	9557.0	14736.6	28.97	17.79
9	1	10	8227.9	12767.5	36.97	22.28
9	10	10	4984.7	7978.0	36.97	19.91
14	1	15	3312.5	5370.5	56.98	26.54
14	15	15	2589.7	4210.6	56.98	23.09
19	1	20	2350.4	3833.0	76.99	29.22
19	10	20	2150.2	3502.9	76.98	26.90
19	20	20	2002.7	3315.8	76.98	24.86
19	30	20	1898.4	3140.6	76.99	23.36

Fig. 6 - Experimental results for $a < c$

Let us consider the correction introduced in formula (3.4); it gives us an idea of the values of n, for which it is possible to obtain some agreement between experimental data and both expected values.

We can claim that experimental data and expected values agree when, for example, the relative error is within 12.5% or 1/8. So we should have:

$$\frac{12C_1+1}{8n} < \frac{1}{8}$$

and we can define n_{min} as the minimum value of n for which this inequality holds. So n_{min} represents the minimal length of the words to be generated in order to achieve the agreement required. By feeding n_{min} into formula (3.4), we find the average number of calls to *random* necessary for generating a word of length n_{min}.

Unfortunately, $n_{min} = \lceil 12C_1+1 \rceil$ is minimized when c is much larger than a, but in this case the quantity $(a+b+c)/(b+2\sqrt{ac})$ is maximized and we have to perform a large number of calls to *random*. For example, let us consider the case $a = b = 1$. When $c = 2$, we find $n_{min} = 279$ but we also have $avg_n > 262{,}000{,}000$. Consequently, it is practically impossible to generate a single word, not to say 10,000 words. If we increase c to 9, we find $n_{min} = 32$ (a very small quantity!) but $avg_n > 216{,}000{,}000$. The (relatively) best situation is achieved for $c = 4$ when we have $n_{min} = 71$ and avg_n

something more than 112,000,000 calls to *random*. In any case, we must conclude that it is actually impossible to provide empirical evidence for our formula (3.4).

5. CONCLUSIONS

We have shown that a very simple algorithm allows us to generate a random underdiagonal walk with a kinds of east steps, b kinds of north-east steps and c kinds of north steps in expected linear time, whenever $a \geq c$. This actually means that m generations of n-walks are performed in time $O(mn)$. Besides, no extra space is required by the generation routine, except for the space required for the string to be produced.

When $a < c$, the same algorithm works in exponential time and hence it is of no practical interest. This is so because there are some general methods for randomly generating n-strings in context-free languages, which perform in polynomial time (see Cohen and Hickey [4]). Therefore, finding out a linear time algorithm for generating random underdiagonal walks with $a < c$ is still an open question.

REFERENCES

[1] E. Barcucci, R. Pinzani, R. Sprugnoli, "Génération aléatoire des animaux dirigés", Atélier Franco-Québecois de Combinatoire, Bordeaux, 1991.

[2] E. A. Bender, "Asymptotic methods in enumeration", SIAM Rev., 16 (1974), 485-514.

[3] B. W. Char, K. O. Geddes, G. H. Gonnet, B. L. Leong, M. B. Monagan, S. M. Watt, Maple V: Language Reference Manual, Springer-Verlag, New York, 1991.

[4] J. Cohen, T. Hickey, "Uniform random generation of strings in context-free languages", SIAM J. Comput., 12 (1983), 645-655.

[5] R. Donaghey, L. W. Shapiro, "Motzkin numbers", Journal of Combinatorial Theory A, 23 (1977), 291-301.

[6] D. Gouyou-Beauchamps, G. Viennot, "Equivalence of the two-dimensional directed animals problem to a one-dimensional path problem", Advances in Applied Mathematics, 9 (1988), 334-357.

[7] J. G. Penaud, "Une nouvelle bijection pour les animaux dirigés", Rapport LaBRI 89-45, Bordeaux.

[8] M. P. Schutzenberger, "Context-free languages and pushdown automata", Information and Control, 6 (1963), 246-264.

Fourier Transform Over Semisimple Algebras and Harmonic Analysis for Probabilistic Algorithms[1]

François Bergeron and Luc Favreau

1. Introduction

Given a finite group G (with 1 as unit) we study in this paper some aspects of probabilistic algorithms of the form

```
r := 1;
repeat
    choose g ∈ G with probability p(g);
    r := r · g
until satisfied
```

where $p : G \longrightarrow [0,1]$ is a probability distribution on G. We are also interested in questions such as the explicit computation of the probability of obtaining some element $g \in G$ after n iteration of the loop. This study is clearly equivalent to the computation of successive powers, in the group algebra $\mathcal{A}(G)$ of G, of

$$\alpha = \sum_{g \in G} p(g)\, g.$$

However, these powers are often hard to calculate in a nice closed form. For this reason we shall consider their Fourier transforms in some suitable context. The idea being that intricate computation of products in the group algebra are replaced by simple point wise products. More precisely, we shall derive new extended versions of a formula of Garsia (see (1) below) and of similar formulas obtained in [4], as well as generalizations of those. All these formulas can be considered to give, in explicit form, a Fourier transform such as defined below.

Let $\mathcal{A}$ be a semisimple algebra, and let $B = v_1, v_2, \ldots, v_n$ be some fixed (linear) basis for the subalgebra $\mathcal{B}$, of $\mathcal{A}$, spanned by the complete set of primitive idempotents $e_1, e_2, \ldots, e_n$ of $\mathcal{A}$. Recall that these idempotents are such that

$$e_k e_j = \begin{cases} e_k & \text{if } k = j \\ 0 & \text{otherwise,} \end{cases}$$

and

$$\sum_{k=1}^{n} e_k = 1.$$

[1] LACIM, UQAM, Montréal H3P 3P8, Canada. Email: bergeron@lacim.uqam.ca. With support from NSERC.

Moreover, none of them can be written as the sum of two orthogonal idempotents. The *Fourier transform* $\widehat{f}$ (with respect to B) of $f = \sum_k f_k v_k \in \mathcal{B}$ is defined to be the vector $(\widehat{f}_1, \widehat{f}_2, \ldots, \widehat{f}_n)$ of coordinates of f in this canonical basis $(e_k)_{1\leq k\leq n}$

$$f = \sum_{k=1}^{n} \widehat{f}_k e_k.$$

This transform clearly enjoys the usual nice property, of the Fourier transform, of sending convolution to component wise product. For a finite group G, if $\mathcal{B}$ is the center $C(G)$ of the group algebra of G, then the canonical idempotents are essentially given by the characters χ_ρ of irreducible representations ρ of G, considered as elements of the group algebra

$$e_\rho = \frac{\chi_\rho(1)}{|G|} \sum_{g\in G} \chi_\rho(g^{-1})\, g.$$

If the basis B is chosen to be the set of conjugacy classes in G

$$c_\rho = \sum_{g\in c(\rho)} g,$$

where $c(\rho) = \{h^{-1}\rho h \mid h \in G\}$ and $\rho \in G$, the Fourier transform with respect to this basis is the usual Fourier transform over G. Consider for instance the cyclic group $\langle x\rangle$ of order n generated by x. Since the group is abelian, any element of the group algebra of $\langle x\rangle$

$$f(x) = \sum_{k=0}^{n-1} f_k x^k,$$

is an element of the center $C(\langle x\rangle)$. Moreover, the irreducible characters give in this case, the following idempotents

$$e_j = \frac{1}{n}\sum_{k=0}^{n-1} q^{-kj} x^k,$$

where $q = e^{2i\pi/n}$, for $0 \leq j \leq n-1$. One concludes that

$$f(x) = \sum_{j=0}^{n-1} \widehat{f}_j e_j,$$

where

$$\widehat{f}_j = f(x)e_j|_1 = \frac{1}{n}\sum_{k=0}^{n-1} f_k q^{-(n-k)j}$$
$$= \frac{1}{n} f(q^j).$$

This is the traditional definition of the discrete Fourier Transform.

2. THE DESCENT ALGEBRA OF A_n

The semisimple algebras considered in this paper are subalgebras of the group algebra of finite Coxeter groups. For the Fourier transform, we give explicit expression, with respect to given basis for these algebras. As a guiding example, let us consider the semisimple subalgebra $\Gamma[A_{n-1}] = \Gamma[S_n]$ of the symmetric group spanned by the linearly independent descent classes

$$D_k = \sum_{d(\sigma)=k} \sigma,$$

where $0 \leq k \leq n-1$ and $d(\sigma) = Card\{1 \leq i \leq n-1 \mid \sigma(i) > \sigma(i+1)\}$. The fact that this is a subalgebra and that it is semisimple is shown in [10]. Now, in [9] A. Garsia gives a beautiful explicit formula

$$\sum_{k=1}^{n} t^k e_k = \frac{1}{n!} \sum_{k=0}^{n-1} (t-k)^{(n)} D_k, \tag{1}$$

relating the basis D_k and the canonical idempotents e_k of $\Gamma(S_n)$. Here, $(t)^{(n)}$ stands for the rising factorial

$$(t)^{(n)} = t(t+1)(t+2)\cdots(t+n-1).$$

For $t \geq 1$,

$$\frac{1}{t^n} \sum_{k=1}^{n} t^k e_k = \sum_{k=0}^{n-1} \frac{(t-k)^{(n)}}{n!t^n} D_k,$$

is a probability distribution on S_n. It follows immediately from the orthogonality of the e_k's that

$$\begin{aligned}
\left(\sum_{k=0}^{n-1} \frac{(t-k)^{(n)}}{n!t^n} D_k \right)^j &= \left(\frac{1}{t^n} \sum_{k=1}^{n} t^k e_k \right)^j \\
&= \frac{1}{t^{jn}} \sum_{k=1}^{n} t^{jk} e_k \\
&= \sum_{k=0}^{n-1} \frac{(t^j-k)^{(n)}}{n!t^{jn}} D_k.
\end{aligned}$$

This gives an answer to a problem of the type considered at the beginning of this text. In order to generalize this last computation, we extend formula (1) using an umbral argument. Thus we obtain

$$\sum_{k=1}^{n} t_k e_k = \sum_{k=0}^{n-1} \left(\sum_{j=1}^{n} \Psi_n(k,j) t_j \right) D_k,$$

where the $\Psi_n(k,j)$'s are the coefficients appearing in the expansion of the polynomial

$$\psi_k^n(t) = \frac{1}{n!}(t-k)^{(n)} = \sum_{j=1}^{n} \Psi_n(k,j) t^j. \tag{2}$$

Hence if Φ_n stands for the inverse of the matrix Ψ_n, then the Fourier transform, with respect to the basis $(D_k)_{0 \le k \le n-1}$, is

$$\widehat{s}_k = \sum_{j=1}^{n} \Phi_n(k,j) s_j. \tag{3}$$

Thus Φ_n is the matrix for the Fourier transform and Ψ_n is the inverse Fourier transform. For example

$$\Phi_5 = \begin{bmatrix} 1 & -4 & 6 & -4 & 1 \\ 1 & -2 & 0 & 2 & -1 \\ 1 & 2 & -6 & 2 & 1 \\ 1 & 10 & 0 & -10 & -1 \\ 1 & 26 & 66 & 26 & 1 \end{bmatrix} \qquad \Phi_6 = \begin{bmatrix} 1 & -5 & 10 & -10 & 5 & -1 \\ 1 & -3 & 2 & 2 & -3 & 1 \\ 1 & 1 & -8 & 8 & -1 & -1 \\ 1 & 9 & -10 & -10 & 9 & 1 \\ 1 & 25 & 40 & -40 & -25 & -1 \\ 1 & 57 & 302 & 302 & 57 & 1 \end{bmatrix}$$

If $\varphi_k^n(t) = \sum_{j=1}^{n} \Phi_n(k,j) t^j$ stands for the enumerating polynomial of the k^{th} row of the matrix Φ_n, then it is readily observed on the previous examples that $\varphi_n^n(t)$ appear to be the well known *eulerian polynomials*

$$\mathbf{A}_n(t) = \sum_{\sigma \in S_n} t^{d(\sigma)+1},$$

whose generating function is

$$\mathbf{A}(x,t) = \frac{1-t}{1-te^{x(1-t)}}. \tag{4}$$

In general, we can characterized the entries of Φ_n as follows

Proposition 1. *The enumerating polynomial of the k^{th} row of the matrix $\Phi_n = \Psi_n^{-1}$, for the Fourier transform from the basis of D_k's to the basis of orthogonal idempotents e_k of the descent algebra $\Gamma(A_{n-1})$, is*

$$\varphi_k^n(t) = (1-t)^{(n-k)}\mathbf{A}_k(t). \tag{5}$$

Proof. Using expression (3) for the generating function $\mathbf{A}(x,t)$, one readily verifies that

$$(1-xt)\frac{\partial}{\partial t}\mathbf{A}(x,t) = \ t\mathbf{A} + t(1-t)\frac{\partial}{\partial x}\mathbf{A}(x,t).$$

Comparing respective coefficients of $x^k/k!$ in this last equation, we obtain

$$\mathbf{A}_{k+1}(x) = xk\mathbf{A}_k + \mathbf{A}_k + x(1-x)\mathbf{A}_k', \tag{6}$$

for $k \geq 2$, and $\mathbf{A}_1(x) = 1$. Multiplying both sides of (6) by $(1-t)^{n-k}$, and using the definition of $\varphi_k^n(t)$, it follows that

$$\varphi_{k+1}^{n+1}(t) = t(1+k)\varphi_k^n(t) + t(1-t)\frac{d}{dt}\varphi_k^n(t). \tag{7}$$

A direct translation of (7) in term of Φ_n gives

$$\Phi_n = \begin{bmatrix} \Phi_n(1,1) & \dots & \Phi_n(1,n) \\ & \Phi_{n-1}M_{n-1} & \\ & & \end{bmatrix} \tag{8}$$

where M_{n-1} has the following expression

$$M_{n-1} = \begin{bmatrix} 1 & n-1 & 0 & \dots & 0 & 0 \\ 0 & 2 & n-2 & \dots & 0 & 0 \\ 0 & 0 & 3 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & n-1 & 1 \end{bmatrix}$$

Now, let us set

$$[\Phi_n : 0] = \begin{bmatrix} \Phi_n(1,1) & \Phi_n(1,2) & \dots & \Phi_n(1,n) & 0 \\ \Phi_n(2,1) & \Phi_n(2,2) & \dots & \Phi_n(2,n) & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \Phi_n(n,1) & \Phi_n(n,2) & \dots & \Phi_n(n,n) & 0 \end{bmatrix}$$

and the analogous convention for $[0 : \Phi_n]$. The definition of $\varphi_k^n(t)$ implies that $\varphi_k^n(t) = (1-t)\varphi_k^{n-1}(t)$, whenever $1 \le k \le n-1$. It follows that

$$\Phi_n = \begin{bmatrix} & [\Phi_{n-1} : 0] \; - \; [0 : \Phi_{n-1}] & \\ \Phi_n(n,1) & \dots & \Phi_n(n,n) \end{bmatrix} \tag{9}$$

We want to show that $n!\Psi_n\Phi_n = n!\mathrm{I}_n$. Denoting by ψ_{kj}^n the entries of the k^{th} row of $n!\Psi_n$, that is $\psi_{kj}^n = n!\Psi_n(k,j)$, it is easy to check that

$$\begin{aligned}(\psi_{k1}^n, \psi_{k2}^n, \dots, \psi_{kn}^n) = (0, \psi_{k1}^{n-1}, \psi_{k2}^{n-1}, \dots, \psi_{k(n-1)}^{n-1}) \\ + (n-k)(\psi_{k1}^{n-1}, \psi_{k2}^{n-1}, \dots, \psi_{k(n-1)}^{n-1}, 0),\end{aligned} \tag{10}$$

for $1 \le k \le n-1$, and that

$$\begin{aligned}(\psi_{n1}^n, \psi_{n2}^n, \dots, \psi_{nn}^n) = (0, \psi_{(n-1)1}^{n-1}, \psi_{(n-1)2}^{n-1}, \dots, \psi_{(n-1)(n-1)}^{n-1}) \\ + (1-k)(\psi_{(n-1)1}^{n-1}, \psi_{(n-1)2}^{n-1}, \dots, \psi_{(n-1)(n-1)}^{n-1}, 0).\end{aligned} \tag{11}$$

The proof of the proposition is by a straightforward induction on n using an adequate mixture of (8) and (9) with (10) and (11). The crucial portion of the argument is to use in the decomposition given in (10) (or (11)) the first part together with (8) and the second part with (9). ∎

As we shall see, this is an instance of a general expression for the rows of the Fourier transform of the descent algebras of many Coxeter groups. But before going on with these other cases, let us derive other properties of $\Gamma[A_{n-1}]$ using (1). The following theorem appears (up to a small variation) in [11] as theorem 1.1. However we give a different proof that will allow us to derive new similar results for descent algebras of other Coxeter groups.

Proposition 2. *The descent class D_1 is an algebraic generator for $\Gamma[A_{n-1}]$ with minimal polynomial*

$$p(x) = \prod_{i=1}^{n} (x - 2^i + n + 1)$$

Proof. First, observe that

$$\begin{aligned}\sum_{i=1}^{n} 2^i e_i &= \frac{1}{n!}\sum_{k=0}^{n} (2-k)^{(n)} D_k \\ &= \frac{1}{n!}((2)^{(n)} D_0 + (1)^{(n)} D_1) \\ &= (n+1)D_0 + D_1,\end{aligned} \tag{12}$$

since $(t)^{(n)} = 0$ whenever $t < 0$. Recall also that $D_0 = \sum e_i$. Now using (3), one sees that the successive powers D_1^k, $k = 0, 1, \ldots, n-1$, form a basis of $\Gamma[A_{n-1}]$ because the transition matrix between the basis of the e_k's and that of the D_1^k's is the (invertible) Vandermonde

$$V = \begin{bmatrix} 1 & 2-(n+1) & \cdots & (2-(n+1))^{n-1} \\ 1 & 4-(n+1) & \cdots & (4-(n+1))^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 2^i-(n+1) & \cdots & (2^i-(n+1))^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 2^n-(n+1) & \cdots & (2^n-(n+1))^{n-1} \end{bmatrix}$$

Moreover, the coefficients $\alpha_0, \alpha_1, \ldots, \alpha_{n-1}$ of the linear combination of the columns of V giving the expression of D_1^n in term of the D_1^k, $0 \le k \le n-1$, are such that the polynomial

$$p(x) = x^n - (\alpha_0 + \alpha_1 x + \ldots + \alpha_{n-1}x^{n-1})$$

has zeros at $2-(n+1), 4-(n+1), \ldots, 2^n-(n+1)$, hence p(x) can only be

$$\prod_{i=1}^{n}(x - 2^i + n + 1)$$

as announced. ∎

This makes it clear why $\Gamma[A_{n-1}] \simeq \mathbb{Q}[x]/\langle p(x)\rangle$ is a commutative semisimple algebra, because we are led back to a close relative of the usual Fourier transform for which the underlying semisimple algebra is isomorphic to $\mathbb{Q}[x]/\langle x^n - 1\rangle$. The point of this last remark being that in both cases the modulo is taken with respect to a polynomial with distinct roots in which case we have the following (see theorem 1.2 in [11])

Proposition 3. *For a polynomial $p(x)$ with distinct roots $r_1, r_2, \ldots, r_n$, the algebra $\mathbb{Q}[x]/\langle p(x)\rangle$ is commutative and semisimple, and its primitive idempotents are given by the interpolating Lagrange polynomials*

$$e_i = \frac{\prod_{j\neq i}(x - r_j)}{\prod_{j\neq i}(r_i - r_j)}$$

Proof. Straightforward. ∎

It follows directly from this last proposition that the idempotents for the algebra $\Gamma[A_{n-1}]$ admit the following expression

$$e_i = \frac{\prod_{j\neq i}(D_1 - (2^j - n - 1)D_0)}{\prod_{j\neq i}(2^i - 2^j)}.$$

Using (1) twice, we can also express the successive powers D_1^k, for $0 \leq k \leq n-1$, in term of the descent classes D_m, for $0 \leq m \leq n-1$, as follows

$$\begin{aligned} D_1^k &= \sum_{i=1}^{n} (2^i - (n+1))^k e_i \\ &= \sum_{i=1}^{n} \sum_{j=0}^{k} \binom{k}{j} (-n-1)^{k-j} 2^{ij} e_i \\ &= \sum_{j=0}^{k} \binom{k}{j} (-n-1)^{k-j} \left(\sum_{i=1}^{n} 2^{ji} e_i \right) \\ &= \sum_{j=0}^{k} \binom{k}{j} (-n-1)^{k-j} \left(\frac{1}{n!} \sum_{m=0}^{n-1} (2^j - m)^{(n)} D_m \right) \\ &= \sum_{m=0}^{n-1} \left(\sum_{j=0}^{k} \binom{k}{j} \binom{2^j - m + n}{n} (-n-1)^{k-j} \right) D_m. \end{aligned}$$

Hence we have explicit forms of the various relations between the three basis of $\Gamma(A_{n-1})$.

3. B_n AND OTHER COXETER GROUPS

Similar consideration can be made in the context of the group algebra of the hyperoctahedral group B_n (see [4]), if one considers the semisimple subalgebra $\Gamma[B_n]$ of the symmetric group spanned by the linearly independent descent classes

$$D_k = \sum_{d(\sigma)=k} \sigma,$$

where $0 \leq k \leq n$ and $d(\sigma) = Card\{0 \leq i \leq n-1 \mid \sigma(i) > \sigma(i+1)\}$. Recall that elements of B_n are signed permutations and that for convenience sake one can set $\sigma(0) = 0$. It was shown in [4] that there is in this context a Garsia like formula

$$\sum_{k=0}^{n} t^k e_k = \frac{1}{2^n n!} \sum_{k=0}^{n} (t-2k)^{((n))} D_k, \tag{13}$$

relating the basis D_k and the canonical basis of idempotents e_k of $\Gamma(B_n)$. Here, $(t)^{((n))}$ stands for the double rising factorial

$$(t)^{((n))} = (t+1)(t+3)\cdots(t+2n-1).$$

Once again umbral considerations on (12) imply that

$$\sum_{k=0}^{n} t_k e_k = \sum_{k=0}^{n} \left(\sum_{j=0}^{n} \Psi_n(k,j) t_j \right) D_k,$$

where the $\Psi_n(k,j)$'s are the coefficients appearing in the expansion of the polynomial

$$\frac{1}{2^n n!}(t-2k)^{((n))} = \sum_{j=0}^{n} \Psi_n(k,j) t^j.$$

The matrices for $n = 4, 5$ are

$$\Phi_4 = \begin{bmatrix} 1 & -4 & 6 & -4 & 1 \\ 1 & -2 & 0 & 2 & -1 \\ 1 & 4 & -10 & 4 & 1 \\ 1 & 22 & 0 & -22 & -1 \\ 1 & 76 & 230 & 76 & 1 \end{bmatrix} \qquad \Phi_5 = \begin{bmatrix} 1 & -5 & 10 & -10 & 5 & -1 \\ 1 & -3 & 2 & 2 & -3 & 1 \\ 1 & 3 & -14 & 14 & -3 & -1 \\ 1 & 21 & -22 & -22 & 21 & 1 \\ 1 & 75 & 154 & -154 & -75 & -1 \\ 1 & 237 & 1682 & 1682 & 237 & 1 \end{bmatrix}$$

As in the A_n case, one can observe that the generating polynomials for the last row of these matrices appear to be corresponding hyperoctahedral descent polynomials

$$\mathbf{B}_n(x) = \sum_{\sigma \in B_n} x^{d(\sigma)},$$

In general, one is led to deduce the following proposition in a manner similar to the proof of proposition 1.

Proposition 4. *The enumerating polynomial of the k^{th} row of the matrix $\Phi_n = \Psi_n^{-1}$, for the Fourier transform from the basis of D_k's to the basis of orthogonal idempotents e_k of the descent algebra $\Gamma(B_n)$, is*

$$\varphi_k^n(x) = (1-x)^{(n-j)} \mathbf{B}_j(x). \tag{14}$$

■

In order to unfold the proof of (6), one needs an expression for the generating function of the polynomials $\mathbf{B}_n(x)$, but it is easy to verify that they satisfy the following recurrence

$$\mathbf{B}_{n+1}(x) = (1+x)\mathbf{B}_n(x) + 2xn\mathbf{B}_n(x) + (2x - 2x^2)\frac{d}{dx}\mathbf{B}_n(x),$$

hence that their exponential generating function is

$$\sum_{n\geq 0} \mathrm{B}_n(x)\frac{u^n}{n!} = \frac{(1-x)e^{u(1-x)}}{1-xe^{2u(1-x)}}.$$

The proof of the following proposition is also similar to that of proposition 2.

Proposition 5. *The descent class D_1 is an algebraic generator for $\Gamma[B_n]$ with minimal polynomial*

$$p(x) = \prod_{i=0}^{n}(x - 3^i + n + 1)$$

Hence the idempotents of $\Gamma(B_n)$ admit the expression

$$e_i = \frac{\prod_{j\neq i}(D_1 - (3^j - n - 1)D_0)}{\prod_{j\neq i}(3^i - 3^j)},$$

by proposition 2. Moreover, for $0 \leq k \leq n$ and $0 \leq m \leq n$, we have

$$D_1^k = \sum_{m=0}^{n-1}\left(\sum_{j=0}^{k}\binom{k}{j}\binom{3^j - m + n}{n}(-n-1)^{k-j}\right)D_m.$$

For all Coxeter groups W of type A_n, B_n, H_3 or $I_2(p)$, the following Garsia like formula has been derived in [4]

$$\sum_{i=0}^{n} e_i t^i = \frac{1}{|W|}\sum_{k=0}^{n}\pi_k^W(t)D_k,$$

where the e_i's are a basis of orthogonal idempotents for the subalgebra of $\Sigma(W)$ spanned by the descent classes

$$D_k = \{w \in W \mid d(w) = k\},$$

and where the polynomials $\pi_k^W(t)$ are defined in term of the *exponents* (ϵ_k) of the group W as

$$\pi_k^W(t) = (t-\epsilon_k)(t-\epsilon_{k-1})\cdots(t-\epsilon_1)(t+\epsilon_1)\cdots(t+\epsilon_{n-k}).$$

Recall that the descent set of an element $w \in W$ is defined to be

$$desc(w) = \{s \in S \mid \ell(ws) < \ell(w)\},$$

where S is the set of Coxeter generators of W, and $\ell(w)$ is the so called *length* of W, that is the length of a reduced expression for w in term of the generators (see [12]). With this definition of descent set, one sets $d(w) = \#desc(w)$. For the descent algebra of the dihedral groups $I_2(p)$, the Fourier transform matrices are simply

$$\begin{bmatrix} 1 & -2 & 1 \\ 1 & 0 & -1 \\ 1 & 2p-2 & 1 \end{bmatrix}.$$

Observe that the last row of the matrix is again given by the coefficients of the polynomial enumerating elements of $I_2(p)$ with respect to number of descents. The same kind of results also hold for finite Coxeter groups that are direct products of groups of type A_n, B_n, H_3 or the dihedral groups since the corresponding descent algebras are simply tensor products of the descent algebras of the respective components. The Fourier transform matrices are the tensor product of the matrices corresponding to the transforms of individual terms in the product.

4. The shuffle algebra

Another family of closely related problems is obtained in the following manner. Consider the anti-automorphism θ of any group algebra, defined on the elements of the group by

$$\theta(g) = g^{-1}.$$

This anti-automorphism sends the descent algebras into other nice algebras who evidently share the properties of the original algebras. Thus for the symmetric group case, $\theta(\Gamma(S_n))$ is the algebra generated by

$$\Xi_k = \theta(D_k).$$

Let us denote $\sqcup\!\sqcup$ the usual *shuffle product*, that is

$$12 \sqcup\!\sqcup 34 = 1234 + 1324 + 1342 + 3124 + 3142 + 3412,$$

and observe, using (3), that

$$\Xi_1 + (n+1)\Xi_0 = 1\,2\cdots j \sqcup\!\sqcup (j+1)\,(j+2)\cdots n$$

Thus a problem considered by Diaconis in [2 and 3] is equivalent to the computation of powers of

$$\frac{1}{2^n}(\Xi_1 + (n+1)\Xi_0) = \frac{1}{2^n}\sum_{j=0}^{n} 1\,2\cdots j \sqcup\!\sqcup (j+1)\,(j+2)\cdots n.$$

Applying (back and forth) θ on Garsia's formula with $t = 2$ gives

$$\left[\frac{1}{2^n}(\Xi_1 + (n+1)\Xi_0)\right]^N = \frac{1}{2^N n!}\sum_{k=0}^{n-1}(2^N - k)_{(n)}\,\Xi_k,$$

as discussed in [4]. This last expression is essentially what is used in [3] to study the number of shuffles needed in order to really mix a deck of n cards. A similar expression can be obtained for the B_n case involving a mixture of shuffles and the following operation on words $w = a_1 a_2 \dots a_k$ on the alphabet $\{1, \overline{1}, 2, \overline{2}, \dots, n, \overline{n}\}$

$$\overline{a_1 a_2 \dots a_k} = \overline{a_k} \dots \overline{a_2}\, \overline{a_1},$$

where

$$\overline{a} = \begin{cases} i & \text{if } a = \overline{i}, \\ \overline{i} & \text{if } a = i, \end{cases}$$

for $1 \leq i \leq n$.

Clearly, considering any automorphisms or antiautomorphisms would lead to other isomorphic algebras for which these cconsideration are interesting. One nice example is the linear extension of

$$\eta(g) = \text{sign}(g) g.$$

5. Various related problems

Another nice example (see [7]) is the case $\mathbb{Z}_2^n$ for which we consider the algebra $\mathcal{B}$ spanned by

$$D_k = [0]^{n-k} \sqcup\!\sqcup [1]^k,$$

where [0] and [1] stand for the elements of $\mathbb{Z}_2$, with $0 \leq k \leq n$. This algebra is semisimple with canonical idempotents e_i characterized by the formula

$$\sum_{k=0}^{n} t^k e_k = \frac{1}{2^n} \sum_{k=0}^{n} (1-t)^{(n-k)} (1+t)^k D_k.$$

From this, one readily derives the expression for the Fourier transform with respect to the basis corresponding to the D_k's. In this case, the descent polynomial is simply $(1+t)^n$, and the rest of the Fourier transform is obtained in the same manner as for the algebras $\Gamma(A_n)$ and $\Gamma(B_n)$. The study of the powers of

$$\alpha = \frac{1}{n+1}(D_0 + D_1),$$

corresponds to the study of successive random changing of one bit in an n-bit word. Computing the Fourier transform of α, we obtain

$$\alpha = \sum_{k=0}^{n} \left(1 - \frac{2k}{n+1}\right) e_k,$$

hence the j^{th} power of α is the inverse Fourier transform of

$$\sum_{k=0}^{n}\left(1-\frac{2k}{n+1}\right)^{j} e_k.$$

For n large enough and taking $j = s\,\frac{(n+1)}{2}$, this last expression becomes

$$\begin{aligned}\sum_{k=0}^{n}\left(1-\frac{2k}{n+1}\right)^{s\,\frac{(n+1)}{2}} e_k &\approx \sum_{k=0}^{n}(t^s)^k e_k \\ &\approx \frac{1}{2^n}\sum_{k=0}^{n}(1-t^s)^{(n-k)}(1+t^s)^k D_k,\end{aligned}$$

where $t = \frac{1}{\mathrm{e}}$.

This example can readily be generalized to G^n for any finite group G. However, for simplicity's sake, we shall only outline what happens in the case $\mathbb{Z}_3^n$. We consider the algebra $\mathcal{B}$ generated by

$$D_{jkl} = [0]^j \sqcup\!\sqcup [1]^k \sqcup\!\sqcup [2]^l,$$

where [0], [1] and [2] stand for the elements of $\mathbb{Z}_3$, with $j+k+l=n$. This algebra is semisimple with canonical idempotents e_{jkl} characterized by the formula

$$\sum_{j+k+l=n} s^j t^k r^l e_{jkl} = \frac{1}{3^n}\sum_{j+k+l=n}(s+t+r)^j(s+\xi t+\xi^2 r)^k(s+\xi^2 t+\xi r)^k D_{jkl}, \tag{15}$$

where ξ is a primitive third root of unity. This last expression is easily derived by taking the n^{th} tensor power of both sides of

$$se_0 + te_1 + re_2 = \frac{1}{3}\left((s+t+r)[0] + (s+\xi t+\xi^2 r)[1] + (s+\xi^2 t+\xi r)[2]\right),$$

where the idempotents e_i's are the characters of the irreducible representations of $\mathbb{Z}_3$. The Fourier transform with respect to the basis corresponding to the D_{jkl}'s is easily computed using (15).

6. Conclusion

Many other problems of the form treated in this text can be studied using the Fourier transform formulas outlined above. Some have been considered by various authors such as: Aldous [1], Diaconis [6], Letac and Takacs [14], Flatto, Odlyzko and Wales [8]; and range from walks on graphs, to problems appearing in coding theory. Another interesting source of semisimple algebras is through *Hecke algebras*

$$\mathcal{H}(G,e) = e\mathcal{A}(G)e$$

where e is any idempotent of the group algebra $\mathcal{A}(G)$ of G. *Gelfand pairs* (G,H), for H a subgroup of G, correspond to the special case when the semisimple algebra $\mathcal{H}(G,e_H)$ is commutative, with

$$e_H = \frac{1}{|H|}\sum_{h\in H} h.$$

Natural extensions of this work would include Fourier transform formulas for the subalgebra, of the complete descent algebra $\Sigma(S_n)$ considered by Garsia and Reutenauer in [10], spanned by the idempotents E_λ (λ partition of n); as well as the subalgebra spanned by the corresponding idempotents (constructed in [5]) for the complete descent algebras of any other finite Coxeter groups.

REFERENCES

[1] D. Aldous, *Minimization algorithms and random walks on the d-cube*, Ann. Prob. **11**, 1983, 403–413.

[2] D. Aldous and P. Diaconis, *Shuffling Cards and Stopping Times*, Amer. Math. Monthly **93**, 1986, 333–348.

[3] D. Bayer and P. Diaconis, *Trailing the Dovetail Shuffle to its Lair*, Technical Report No. 2, Dept. Statistics, Stanford University, 1990.

[4] F. Bergeron and N. Bergeron, *Orthogonal Idempotents in the Descent Algebra of B_n*, J. of Pure and Appl. Algebra, (accepted 1991).

[5] F. Bergeron, N. Bergeron, R. B. Howlett and D. E. Taylor, *A Decomposition of the Descent Algebra of Finite Coxeter Groups*, Journal of Algebraic Combinatorics, Vol.1, No.1, 1992, 23–42.

[6] P. Diaconis, *Group Representations in Probability and Statistics*, Institute of Mathematical Statistics, Hayward, CA, 1988.

[7] P. Diaconis, R. L. Graham and J. A. Morisson, *Asymptotic Analysis of a Random Walk on a Hypercube with Many Dimensions*, Random Structures and Algorithms, Vol. 1, No. 1, 1990.

[8] L. Flatto, A. Odlyzko and D. Wales, *Random shuffles and group representations*, Ann. Prob. **13**, 154–178.

[9] A. Garsia, *Combinatorics of the Free Lie Algebra and the Symmetric Group*, Analysis, Et Cetera, Research papers published in honor of Jurgen Moser's 60th birthday, ed: Paul H. Rabinowitz and Eduard Zehnder, Academic Press, 1990.

[10] A. Garsia and C. Reutenauer, *A Decomposition of Solomon's Descent Algebras*, Adv. in Math. **77**, No 2, 1989, 189–262.

[11] M. Gerstenhaber and S. D. Schack, *A Hodge-Type Decomposition for Commutative Algebra Cohomology*, J. Pure Appl. Algebra **48**, 1987, 229-247.

[12] J. E. Humphreys, *Reflection Groups and Coxeter Groups*, Cambridge University Press, 1990.

[13] G. Letac, *Problèmes classiques de probabilité sur un couple de Gelfand*, SLN in Math. **861**, Springer-Verlag.

[14] G. Letac and L. Takacs, *Random Walks on the m-dimensional Cube* , J. Reine Angew. Math. **310**, 187–195.

Décomposition hyperoctahédrale de l'homologie de Hochschild

Nantel Bergeron
Harvard University, Department of Mathematics
Cambridge, Massachusetts 02138

Résumé

Nous étudions l'homologie Hyperoctahédrale. Plus précisément, nous étudions les familles d'opérateurs de la catégorie hyperoctahédrale qui commutent avec les opérateurs de bord de Hochschild

$$b_n = \sum_{i=0}^{n} (-1)^i d_i$$

où $d_i:[n] \to [n-1]$ est l'application de face dans cette catégorie. Cette étude est intimement reliée aux projections dans les espaces de puissances symétriques et hyperoctahédrales d'algèbres de Lie libres. Nous discuterons aussi de l'homologie de Harisson dans un tel contexte et de certains autres problèmes connexes.

I. Introduction.

Le problème de décomposer l'homologie de Hochschild pour les algèbres commutatives sur les corps de caractéristique 0, a été étudié par Gerstenhaber et Schack dans [9] et par Barr dans [1]. Gerstenhaber et Schack donnent une caractérisation complète des familles d'éléments $f_n \in \mathbb{Q}[S_n]$ qui commutent avec les bords de Hochschild; $b_n f_n = f_{n-1} b_n$. Plus précisément, pour chaque entier n, ils introduisent une famille d'idempotents orthogonaux $\{e_n^k \in \mathbb{Q}[S_n]: 1 \leq k \leq n\}$ telle que $b_n e_n^k = e_{n-1}^k b_n$ et que pour toute famille $f_n \in \mathbb{Q}[S_n]$ où $b_n f_n = f_{n-1} b_n$, nous avons

$$f_n = \sum_{k=1}^{n} \mathrm{sgn}(f_k)\ e_n^k. \tag{I.1}$$

Ici, sgn: $\mathbb{Q}[S_n] \to \mathbb{Q}$ est l'homomorphisme signé. Une autre conséquence de ces résultats, pour A une algèbre commutative sur un corps de caractéristique 0 et M un A-bimodule, est la décomposition suivante de l'homologie de Hochschild

$$H_n(A, M) = \bigoplus_{k=1}^{n} H_n^k(A, M) \tag{I.2}$$

où $H_n^k(A, M) = e_n^k H_n(A, M)$. Dans [10], Loday donne une version de ces résultats pour les algèbres sur les corps de caractéristique non nulle. De plus, il donne aussi une décomposition similaire de l'homologie cyclique. L'équation (I.1) nous donne que la décomposition (I.2) est la plus fine dans l'algèbre $\mathfrak{L}' = \mathbb{Q}[\mathbf{Fin}']$ définie par Loday dans [10] où $\mathbf{Fin}'$ est la catégorie des ensembles $[n] = \{0, 1, ..., n\}$ pointés. L'algèbre $\mathfrak{L}'$ est un cadre naturel dans lequel les équations de commutation ci-haut ont un sens.

D'autre part, les idempotents de $\{e_n^k\}$ sont étudiés dans les travaux de Garsia [7], et avec Reutenauer dans [8]. Dans [7], la série génératrice des e_n^k y est donnée sous la forme

$$\sum_{k=1}^{n} e_n^k x^k = \frac{1}{n!} \sum_{\sigma \in S_n} (x - d(\sigma)) \uparrow_{A_n} \operatorname{sgn}(\sigma)\sigma \tag{I.3}$$

où $d(\sigma) = \operatorname{Card}\{i: \sigma_i > \sigma_{i+1}\}$, $(x) \uparrow_{A_n} = x(x+1) \cdots (x+n-1)$. En fait, l'équation (I.3) est l'image de l'équation dans [7] sous l'homomorphisme défini par $\sigma \mapsto \operatorname{sgn}(\sigma)\sigma$. Dans nos travaux avec F. Bergeron [5], nous avons développé une formule similaire à (I.3) pour le groupe hyperoctahédral B_n. Plus précisément, nous avons que les éléments $\rho_n^k \in \mathbb{Q}[B_n]$ définis par

$$\rho_n(x) = \sum_{k=0}^{n} \rho_n^k x^k = \frac{1}{|B_n|} \sum_{\pi \in B_n} (x - 2d(\pi)) \uparrow_{B_n} \operatorname{sgn}(\pi)\pi \tag{I.4}$$

sont des idempotents orthogonaux. Ici $d(\pi)$ dénote le nombre de descentes de π que nous définirons plus bas et $(x) \uparrow_{B_n} = (x+1)(x+3) \cdots (x+2n-1)$. Une fois de plus (I.4) n'est pas l'expression exacte de [5] mais elle est obtenue sous l'homomorphisme défini par $\pi \mapsto \operatorname{sgn}(\pi)\pi$. Parmi les propriétés remarquables des idempotents ρ_n^k est la commutation $b_n \rho_n^k = \rho_{n-1}^k b_n$. Cette nouvelle famille d'idempotents, nous permettra de donner une décomposition de l'homologie $H(A, M)$ différente de (I.2) pour les algèbres A munies d'un automorphisme involutif.

Dans la Section 1, en imitant Loday, nous construirons une algèbre $\mathfrak{B}'$ dans laquelle toutes les équations de commutation ci-haut auront un sens. Dans la Section 2, nous traiterons la décomposition de $H(A, M)$ donnée par les idempotents ρ_n^k. Cette dernière n'est pas la décomposition la plus fine possible pour l'algèbre $\mathfrak{B}'$. En particulier, si l'on munit une algèbre commutative A de l'involution triviale, la décomposition devient triviale! Nous verrons, dans la Section 3, certains raffinements de la décomposition de la Section 2. Pour pousser plus loin la décomposition, nous rappelons à la Section 4 la théorie des algèbres de Lie Hyperoctahédrales libres. La Section 5 est dédiée à l'étude de l'homologie hyperoctahédrale de Harisson. En conclusion, nos travaux convergent vers la décomposition la plus fine de $H(A, M)$ pour A une algèbre commutative munie d'un automorphisme involutif. Nous présenterons certains autres résultats partiels dans cette voie.

Avant de débuter cette discussion nous répondons à une question de C. Procesi [communication personnelle]. Pour l'énoncer nous devons nous placer dans un contexte plus général. D'abord, quelques définitions. Un groupe fini qui admet une présentation de la forme

$$\langle s_i, 1 \leq i \leq n; (s_i s_j)^{m_{ij}} = 1, m_{ii} = 1 \rangle$$

est appelé un *groupe de Coxeter.* Pour W un groupe de Coxeter et $w \in W$, nous posons $\ell(w) = \min\{p : w = s_{i_1} s_{i_2} \cdots s_{i_p}\}$ et $d(w) = \mathrm{Card}\{s_i : \ell(w) > \ell(ws_i)\}$. Dans l'algèbre de groupe $\mathbb{Q}[W]$, Procesi s'est demandé si les éléments $\ell_n^k = \sum_{d(w)=k} \mathrm{sgn}(w)w$ engendrent une sous-algèbre commutative de dimension $n+1$. Cette question est motivée du fait que pour les groupes de type A (groupes symétriques) Loday [10] a démontré ce fait en utilisant les formules de commutation avec les bords de Hochschild. Par la suite, nous avons démontré [5] ce même résultat pour les groupes de type B; ceci découle de la formule (I.4). En effet, si l'on pose

$$\lambda_n^k = (-1)^{k-1} \rho_n(2k+1) \tag{I.5}$$

on observe que

$$\lambda_n^k = \sum_{i=0}^{k} (-1)^i \binom{n+i}{i} \ell_n^{k-i}.$$

D'une part la sous-algèbre engendrée par les $\{\ell_n^k\}$ est égale à celle engendrée par les $\{\lambda_n^k\}$, d'autre part cette dernière est commutative et de dimension $n+1$ puisque

$$\lambda_n^k \lambda_n^{k'} = (-1)^{(k-1)(k'-1)} \rho_n(2k+1) \rho_n(2k'+1) = \lambda_n^{k'} \lambda_n^k.$$

Nous avons donc répondu affirmativement à la question de Procesi pour ces deux grandes familles de groupes de Coxeter. Cependant, ceci est faux en général! En effet, on vérifie à l'aide de Maple que pour le groupe D_5 la sous-algèbre engendrée par les $\{\ell_n^k\}$ n'est pas commutative, sa dimension étant plus grande que 6, et elle n'est même pas semi-simple! Il semble qu'une formule de type (I.3) ou (I.4) soit la clef nécessaire pour avoir une telle sous-algèbre semi-simple.

1. Catégorie hyperoctahédrale

Soit $\mathbf{Fin}'_B$ la catégorie ayant pour objets les ensembles $[n] = \{0, 1, ..., n\}$ et pour morphismes les fonctions signées $f_\epsilon : [n] \to [m]$. Un tel morphisme est la donnée d'une fonction $f : [n] \to [m]$ telle que $f(0) = 0$ et d'une suite de signes $\epsilon = (\epsilon_1, \epsilon_2, ..., \epsilon_n) \in \mathbb{Z}_2^n$. Ici, $\mathbb{Z}_2 = \{-1, 1\}$ est le groupe multiplicatif à deux éléments. Dans cette catégorie, deux morphismes $f_\epsilon\colon [n] \to [m]$ et $g_\gamma\colon [s] \to [r]$ sont composables seulement si $m = s$ et dans ce cas

$$f_\epsilon \circ g_\gamma = (f \circ g)_{\epsilon\gamma_f}$$

où $\gamma_f = (\gamma_{f(1)}, \gamma_{f(2)}, ..., \gamma_{f(n)}) \in \mathbb{Z}_2^n$. Nous appelons $\mathbf{Fin}'_B$ la *catégorie hyperoctahédrale pointée.* Il est parfois plus pratique d'écrire $f = (f_1, f_2, ..., f_n)$ pour désigner un morphisme f_ϵ de la catégorie $\mathbf{Fin}'_B$, où $f_i = \epsilon_i f(i)$. De cette façon, le groupe hyperoctahédral B_n peut être representé dans $\mathbf{Fin}'_B$ par les morphismes $\pi = \sigma_\epsilon\colon [n] \to [n]$ où σ est une permutation de $[n]$ laissant fixe l'élément 0.

D'autre part, soit Δ la catégorie simpliciale. On sait que cette catégorie ayant pour objets les ensembles $[n]$ et pour morphismes les fonctions faiblement croissantes, est engendrée par les applications faces $d_i\colon [n] \to [n-1]$ et dégénérescences $s_i\colon [n] \to [n+1]$ pour $0 \le i \le n$. Nous définissons un foncteur $\Delta^{\text{op}} \to \mathbf{Fin}'_B$ en envoyant d_i et s_i sur les morphismes, dénotés d_i et s_i par abus de langage, définis par: $d_i = (1, 2, ..., i-1, i, i, i+1, ..., n-1)$ si $i < n$, $d_n = (1, 2, ..., n-1, 0)$ et $s_i = (1, 2, ..., i-1, i+1, ..., n+1)$. Nous avons une théorie de l'homologie hyperoctahédrale chaque fois qu'un foncteur $\Delta^{\text{op}} \to K$**-module** se factorise à travers la catégorie $\mathbf{Fin}'_B$. Ici nous nous intéressons plus particulièrement au bord de Hochschild

$$b_n = \sum_{i=o}^{n} (-1)^i d_i\colon [n] \to [n-1]. \tag{1.1}$$

Un calcul simple montre que $b_{n-1} b_n = 0$.

Nous posons maintenant $\mathfrak{B}' = \mathbb{Q}[\mathbf{Fin}'_B]$ l'algèbre des morphismes de $\mathbf{Fin}'_B$. Plus précisément, un élément de $\mathfrak{B}'$ est une somme formelle finie $\sum c_i f_i$ telle que pour tout i, $c_i \in \mathbb{Q}$ et f_i est un homomorphisme de $\mathbf{Fin}'_B$. Par convention nous posons $f \circ g = 0$ si f et g ne sont pas composables dans $\mathbf{Fin}'_B$. Nous remarquons que $b_n \in \mathfrak{B}'$ et que $\mathbb{Q}[B_n]$ est une sous-algèbre de $\mathfrak{B}'$. Nous avons donc un cadre de travail dans lequel les produits tels que $f_{n-1} b_n$ et $b_n f_n$ ont un sens pour $f_n \in \mathbb{Q}[n]$. Le but de ce travail est de caractériser les familles $\{f_n \in \mathbb{Q}[B_n]\}$ telles que

$$f_{n-1} b_n = b_n f_n. \tag{1.2}$$

Exemple 1.1 Si $\mathcal{A}$ est une algèbre commutative sur un corps de caractéristique 0 muni d'un automorphisme involutif

$$\begin{aligned} \mathcal{A} &\longrightarrow \mathcal{A} \\ a &\longmapsto \overline{a} \end{aligned}$$

et si M est un bimodule symétrique ($a.m = m.a$), nous posons $C_n(\mathcal{A}; M) = M \otimes \mathcal{A}^{\otimes n}$. Pour simplifier la notation, nous écrivons $(m, a_1, ..., a_n) \in C_n(\mathcal{A}; M)$ pour désigner le tenseur $m \otimes a_1 \otimes \cdots \otimes a_n$. Nous avons alors le foncteur $\mathbf{Fin}'_B \to K$**-module** défini par $[n] \mapsto C_n(\mathcal{A}; M)$ et $f_\epsilon \in \text{Hom}_{\mathbf{Fin}'_B}([n], [m]) \longmapsto f_\epsilon(a_0, a_1, ..., a_n) = (a_{f_\epsilon^{-1}(0)}, a_{f_\epsilon^{-1}(1)}, ..., a_{f_\epsilon^{-1}(m)}) \in C_m(\mathcal{A}; M)$

où pour $f^{-1}(i) = \{j_1, j_2, ..., j_k\}$ et $\epsilon_0 = 1$ nous posons $f_\epsilon^{-1}(i) = \{\epsilon_{j_1}j_1, \epsilon_{j_2}j_2, ..., \epsilon_{j_k}j_k\}$ et $a_{\{\epsilon_{j_1}j_1, \epsilon_{j_2}j_2, ..., \epsilon_{j_k}j_k\}} = a_{\epsilon_{j_1}j_1}a_{\epsilon_{j_2}j_2}\cdots a_{\epsilon_{j_k}j_k}$ avec la convention que $a_\emptyset = 1$ et que $a_{-i} = \overline{a}_i$. Nous avons maintenant un foncteur $\Delta^{\mathrm{op}} \to \mathbf{Fin}'_B \to K$-**module** qui se factorise à travers $\mathbf{Fin}'_B$. Il est aisé de vérifier que ce foncteur correspond à la théorie classique de l'homologie de Hochschild. En d'autres termes, nous avons le complexe

$$C_*(\mathcal{A}; M) = \cdots \xrightarrow{b_{n+1}} C_n(\mathcal{A}; M) \xrightarrow{b_n} C_{n-1}(\mathcal{A}; M) \xrightarrow{b_{n-1}} \cdots \xrightarrow{b_1} C_0(\mathcal{A}; M)$$

et l'homologie de Hochschild classique est donnée par

$$H_n(\mathcal{A}; M) = \frac{\ker(b_n)}{\mathrm{Im}(b_{n+1})}. \tag{1.3}$$

2. Décomposition donnée par l'agèbre de descentes

Dans cette section, nous rappelons certains résultats de [5]. Ceux-ci sont basés sur nos travaux [3][4] sur les algèbres de descentes des groupes hyperoctahédraux, transformées par l'homomorphisme défini par $\pi \mapsto \mathrm{sgn}(\pi)\pi$. Nous en rappelons ici les grandes lignes. Nous terminons cette section par l'étude de ces résultats à travers le foncteur $\mathbf{Fin}'_B \to \mathbf{Fin}'$ qui oublie la structure de signes.

Pour $\pi = (\pi_1, \pi_2, ..., \pi_n) \in B_n$, nous définissons l'ensemble $D(\pi) \subseteq [n-1]$ des *descentes* de π comme étant l'ensemble des positions $0 \le i \le n-1$ pour lesquelles $\pi_i > \pi_{i+1}$ avec, bien sur, $\pi_0 = 0$. Rappelons que $\pi = \sigma_\epsilon \in \mathrm{Hom}_{\mathbf{Fin}'_B}([n],[n])$ et que $\pi_i = \epsilon_i\sigma(i)$. Etant donné un ensemble de descentes $D = \{i_1 < i_2 < \cdots < i_k\} \subseteq [n-1]$, nous construisons une composition $p(D) = (i_2 - i_1, i_3 - i_2, ..., i_k - i_{k-1}, n - i_k)$ de l'entier $n - i_1$. Par convention, $p(\{n\}) = \emptyset$ est la seule composition de 0. Ceci est une correspondance bijective entre les sous-ensembles de $[n-1]$ et les compositions d'entiers m, $0 \le m \le n$. Pour simplifier l'écriture, nous posons $p(\pi) = p(D(\pi))$. Dans le contexte plus général des groupes de Coxeter, Solomon [12] a montré que l'espace vectoriel engendré par les éléments $X_p = \sum_{p(\pi)=p} \mathrm{sgn}(\pi)\pi$ pour p une composition de $0 \le m \le n$, est fermé sous la multiplication dans $\mathbb{Q}[B_n]$. Autrement dit, nous avons une sous-algèbre de $\mathbb{Q}[B_n]$. C'est cette sous-algèbre que nous appelons l'*algèbre des descentes* de B_n. Nous avons montré que les éléments

$$I_p = \sum_{\substack{q_0 \models r \le n-m \\ q_i \models p_i}} Z_{k(q_0)} N_{k(q_1)} N_{k(q_2)} \cdots N_{k(q_k)}\ X_{q_0 q_1 q_2 \cdots q_k} \tag{2.1}$$

pour $p = (p_1, p_2, ..., p_k) \models m \le n$, forment une base d'idempotents de l'algèbre des descentes de B_n. Ici, $p \models m$ signifie que p est une composition de l'entier m, $k(p)$ dénote le nombre de parts de p, $Z_i = (-1)^i \frac{1\cdot 3\cdot 5\cdot\cdots\cdot(2i-1)}{2^i i!}$, $N_i = \frac{(-1)^{i-1}}{i}$ et $q_0 q_1 q_2 \cdots q_k$ est la composition de $n-r$ obtenue en concaténant les compositions $q_0, q_1, ..., q_k$ bout à bout. A partir des éléments (2.1), nous avons construit une famille complète d'idempotents orthogonaux pour la partie semi-simple de l'algèbre des descentes de B_n. Pour la définir, posons $\lambda(p)$ le partage de $m \leq n$ obtenu de $p \models m \leq n$ en réordonnant les parts de p. Les éléments

$$E_\lambda = \frac{1}{2^k k!} \sum_{\lambda(p)=\lambda} I_p \tag{2.2}$$

pour λ un partage de $m \leq n$ forment une famille d'idempotents minimaux orthogonaux. Pour la suite, nous devons rappeler une autre expression des E_λ. Soit $S = i_1 < i_2 < \cdots < i_r \subseteq \{1, 2, ..., n\}$, pour $\pi = \sigma_\epsilon \in B_r$ nous posons $S\pi = (\epsilon_1 i_{\sigma(1)}, \epsilon_2 i_{\sigma(2)}, ..., \epsilon_r i_{\sigma(r)})$. Par extension linéaire, nous posons $I_{[S]} = SI_{(r)}$ où $I_{(r)}$ est élément de l'algèbre des descentes de B_r. Similairement, nous posons $\tilde{I}_{[S]} = SI_\emptyset$ où $\emptyset \models 0 \leq r$. Pour $\lambda = (\lambda_1, \lambda_2, ..., \lambda_k)$ un partage de $m \leq n$, nous avons

$$E_\lambda = \frac{1}{2^k s(\lambda)} \sum_{\substack{S_0+S_1+\cdots+S_k=\{1,2,...,n\} \\ |S_0|=n-m,\ |S_i|=\lambda_i}} \sum_{\sigma\in\mathcal{S}_k} \pm \tilde{I}_{[S_0]} \cdot I_{[S_{\sigma(1)}]} \cdot I_{[S_{\sigma(2)}]} \cdots I_{[S_{\sigma(k)}]} \tag{2.3}$$

où la première somme est sur toutes les décompositions de l'ensemble $\{1, 2, ..., n\}$ en $k+1$ sous-ensembles disjoints non vides, $\mathcal{S}_k$ dénote le groupe symétrique sur k éléments et $s(\lambda)$ dénote la cardinalité du stabilisateur de la suite λ sous l'action du groupe symétrique $\mathcal{S}_k$. De plus, l'opération "$\cdot$" dans les termes de cette somme est le produit de concaténation étendu linéairement aux sommes formelles de suites. Le signe $\pm$ dépend de la décomposition de $\{1, 2, ..., n\}$ et de la permutation $\sigma \in \mathcal{S}_k$.

Nous avons montré [5] que les éléments ρ_n^k apparaissant dans l'équation (I.4) sont obtenus par

$$\rho_n^k = \sum_{\substack{\lambda\vdash m\leq n \\ k(\lambda)=k}} E_\lambda \tag{2.4}$$

où $\lambda \vdash m$ dénote que λ est un partage de m. Le théorème suivant est démontré dans la Section 5 de [5].

Théorème 2.1 *Dans l'algèbre $\mathfrak{B}'$ nous avons pour $0 \leq k \leq n$*

$$b_n \rho_n^k = \rho_{n-1}^k b_n \tag{2.5}$$

et $\{\rho_n^k\}$ forment une famille d'idempotents orthogonaux de $\mathbb{Q}[B_n]$ telle que

$$\sum_{k=0}^{n} \rho_n^k = \text{Identité} \tag{2.6}$$

Ce théorème nous permet de décomposer le complexe de l'exemple 1.1:

$$C_*(\mathcal{A}; M) = \bigoplus_{k \geq 0} C_*^{(k)}(\mathcal{A}; M)$$

et par le fait même nous avons le théorème suivant

$$H_n(\mathcal{A}; M) = \bigoplus_{k \geq 0} H_n^{(k)}(\mathcal{A}; M) \tag{2.7}$$

où $C_n^{(k)}(\mathcal{A}; M) = \rho_n^k C_n(\mathcal{A}; M)$ et $H_n^{(k)}(\mathcal{A}; M) = \rho_n^k H_n(\mathcal{A}; M)$.

Pour conclure cette section, considérons l'homomorphisme d'algèbre $F: \mathfrak{B}' \to \mathfrak{L}'$ induit par le foncteur $\mathbf{Fin}'_B \to \mathbf{Fin}'$ défini par $[n] \mapsto [n]$ et $f_\epsilon \mapsto f$.

Proposition 2.2

$$F\rho_n(x) = (1, 2, ..., n) \in \mathcal{S}_n \tag{2.8}$$

Preuve Pour ceci, il suffit de montrer que $F\rho_n^k = 0$ pour $k > 0$. En effet, puisque $\sum E_\lambda = (1, 2, ..., n) \in B_n$ nous avons $\sum_{k=0}^n \rho_n^k = (1, 2, ..., n)$. En appliquant F de chaque côté de cette dernière égalité, en assumant que $F\rho_n^k = 0$ pour $k > 0$, nous obtenons $F\rho_n^0 = (1, 2, ..., n)$. Considérons les équations (2.3) et (2.4), nous avons

$$F\rho_n^k = \frac{1}{2^k s(\lambda)} \sum_{\substack{S_0+S_1+\cdots+S_k=\{1,2,\ldots,n\} \\ \sigma \in \mathcal{S}_k}} \pm F\tilde{I}_{[S_0]} \cdot FI_{[S_{\sigma(1)}]} \cdots FI_{[S_{\sigma(k)}]}.$$

Il est donc suffisant de montrer que $FI_{(r)} = 0$ pour $I_{(r)}$ donné par (2.1) avec $p = (r) \models r \geq 1$. D'autre part, pour $p \models r$, nous avons montré [5] que

$$X_p = X_p^A X_{(r)} \tag{2.9}$$

où X_p^A est un élément de la base de Solomon dans l'algèbre des descentes du groupe symétrique $\mathcal{S}_r$. En comparant la définition (2.1) avec la définition analogue de e_n^1 dans [7][8], nous pouvons montrer, en utilisant (2.9), que $I_{(r)} = e_n^1 X_{(r)}$. Donc $FI_{(r)} = e_n^1 FX_{(r)}$. Maintenant, un terme typique de $X_{(r)}$ est de la forme $x = \text{sgn}(\pi)\pi = \text{sgn}(\pi)(-i_k, ..., -i_1, j_1, ..., j_{r-k})$ où

$S = \{i_1 < i_2 < \cdots < i_k\} \subseteq \{1,2,...,r\}$ et $\{j_1 < j_2 < \cdots < j_{r-k}\} = \{1,2,...,r\} - S$. Nous avons $F\text{sgn}(\pi)\pi = (-1)^k \text{sgn}(\sigma)\sigma$ où $\sigma = (i_k, i_{k-1}, ..., i_1, j_1, ..., j_{r-k})$. Soit π' l'élément de $X_{(r)}$ obtenu de π en changeant le signe de 1 (soit $i_1 = 1$ ou $j_1 = 1$). Nous remarquons que $F\text{sgn}(\pi')\pi' = (-1)^{k\pm 1}\text{sgn}(\sigma)\sigma = -F\text{sgn}(\pi)\pi$. Ceci construit une involution sans point fixe pour laquelle tous les termes de $FX_{(r)}$ s'annullent. Ce qui conclue la preuve du théorème.

L'équation (2.8) nous donne que la relation (2.5) est triviale lorsque l'on oublie la structure hyperoctahédrale. En particulier la décomposition (2.7) de l'homologie de Hochschild est triviale si l'on munit $\mathcal{A}$ de l'involution triviale.

3. Raffinement de la décomposition

Dans cette section nous allons raffiner le résultat donné par le théorème 2.1. Pour cela nous débutons par une proposition.

Proposition 3.1 *Pour $X_n = X_{(n)}$ dans l'algèbre des descentes de B_n, nous avons*

$$b_n X_n = X_{n-1} b_n \tag{3.1}$$

Preuve Rappelons que $X_n = \sum_{\pi_1<\pi_2<\cdots<\pi_n} \text{sgn}(\pi)\pi$. Fixons $1 \leq i < n$ et supposons que $\pi_{i+1}^{-1} \neq \pi_i^{-1} \pm 1$. Dans ce cas, nous devons avoir $\pi_i \pi_{i+1} < 0$. D'autre part, pour $(i, i+1)$ la transposition simple qui échange i et $i+1$, posons $\pi' = (i, i+1)\pi$. Nous avons que π' est élément du support de X_n, donc les termes $\text{sgn}(\pi)\pi$ et $\text{sgn}(\pi')\pi'$ de X_n s'annullent l'un et l'autre lorsqu'ils sont multipliés par d_i à gauche puisque $d_i\pi = d_i\pi'$ et $\text{sgn}(\pi) = -\text{sgn}(\pi')$. Ceci montre que lorsque l'on considère le produit $d_i X_n$, seuls les termes pour lesquels $\pi_{i+1}^{-1} = \pi_i^{-1} \pm 1$ survivent. Mais si $\pi_{i+1}^{-1} = \pi_i^{-1} - 1$ alors $\pi_i > \pi_{i+1}$ et donc π n'est pas dans le support de X_n. Il nous reste deux possibilités.

1. Supposons que $\pi_{i+1}^{-1} = \pi_i^{-1} + 1 > 0$. Pour $\pi = \sigma_\epsilon = (\pi_1, \pi_2, ..., \pi_n)$, nous posons $j = \pi_i^{-1}$ et $\pi' = (\pi'_1, ..., \pi'_j, \pi'_{j+2}, ..., \pi'_n) \in B_{n-1}$ avec $\pi'_k = \pi_k$ si $\sigma_k < i+1$ et $\pi'_k = \epsilon_k(\sigma_k - 1)$ si $\sigma_k > i+1$. Nous avons alors que $\text{sgn}(\pi')\pi'$ est un terme de X_{n-1}, que $\text{sgn}(\pi') = (-1)^{i-j}\text{sgn}(\pi)$ et que $d_i\pi = \pi' d_j$.
2. Supposons que $\pi_{i+1}^{-1} = \pi_i^{-1} + 1 < 0$. Pour $\pi = \sigma_\epsilon = (\pi_1, \pi_2, ..., \pi_n)$, nous posons $j = -\pi_{i+1}^{-1}$ et $\pi' = (\pi'_1, ..., \pi'_j, \pi'_{j+2}, ..., \pi'_n) \in B_{n-1}$ avec $\pi'_k = \pi_k$ si $\sigma_k < i$ et $\pi'_k = \epsilon_k(\sigma_k - 1)$ si $\sigma_k > i$. Nous avons alors que $\text{sgn}(\pi')\pi'$ est un terme de X_{n-1}, que $\text{sgn}(\pi') = -(-1)^{i-1-j}\text{sgn}(\pi) = (-1)^{i-j}\text{sgn}(\pi)$ et que $d_i\pi = \pi' d_j$.

Donc,

$$\sum_{i=1}^{n-1} \sum_{\pi_1<\pi_2<\cdots<\pi_n} (-1)^i \text{sgn}(\pi) d_i \pi = \sum_{j=1}^{n-1} \sum_{\pi'_1<\pi'_2<\cdots<\pi'_n} (-1)^j \text{sgn}(\pi')\pi' d_j. \tag{3.2}$$

Dans le cas où $i = 0$, posons π' l'élément obtenu en changeant le signe de 1 dans π. Les deux termes $\mathrm{sgn}(\pi)\pi$ et $\mathrm{sgn}(\pi')\pi'$ de X_n s'annullent l'un et l'autre lorsqu'ils sont multipliés par d_0 à gauche puisque $d_0\pi = d_0\pi'$ et $\mathrm{sgn}(\pi) = -\mathrm{sgn}(\pi')$, donc $d_0X_n = 0$. Finalement dans le cas où $i = n$ nous notons que pour $\mathrm{sgn}(\pi)\pi$ un terme de X_n, nous avons deux possibilités.

1. Soit que $\pi_1 = -n$; dans ce cas nous posons $\pi' = (\pi_2, \pi_3, \ldots, \pi_n) \in B_{n-1}$ et nous obtenons $d_n\pi = \pi' d_0$ et $\mathrm{sgn}(\pi) = (-1)^n\mathrm{sgn}(\pi')$.

2. Soit que $\pi_n = n$; dans ce cas nous posons $\pi' = (\pi_1, \pi_2, \ldots, \pi_{n-1}) \in B_{n-1}$ et nous obtenons $d_n\pi = \pi' d_n$ et $\mathrm{sgn}(\pi) = \mathrm{sgn}(\pi')$.

Ceci donne, $(-1)^n d_n X_n = X_{n-1}d_0 + (-1)^n X_{n-1}d_n$, ce qui, avec (3.2) complète la preuve de (3.1).

Cette proposition nous permet de raffiner le résultat du théorème 2.1. En effet, nous remarquons que la famille d'éléments $\rho_n^{0,k} = e_n^k X_n$ commutent avec les bords de Hochschild. D'autre part, nous avons

$$\rho_n^{0,k} = \sum_{\substack{\lambda \vdash n \\ k(\lambda)=k}} E_\lambda. \tag{3.3}$$

Ceci est obtenu en combinant (2.9), (2.1), (2.2) et un résultat de Garsia et Reutenauer [8] qui décrit e_n^k par une formule analogue à (2.4). Maintenant, en rappelant le fait que les idempotents E_λ sont orthogonaux, grâce à (2.4), nous avons que les idempotents $\rho_n^{+,k} = \rho_n^k - \rho_n^{0,k}$ avec les idempotents $\rho_n^{0,k}$ forment une famille d'idempotents orthogonaux qui commutent avec les bords de Hochschild. Remarquons que $\rho_n^{0,0} = \rho_n^{+,n} = 0$ sont les seuls éléments triviaux de cette famille.

Théorème 3.2 *Dans l'algèbre $\mathfrak{B}'$ nous avons pour $0 \leq k \leq n$*

$$\begin{aligned} b_n\rho_n^{0,k} &= \rho_{n-1}^{0,k} b_n \\ b_n\rho_n^{+,k} &= \rho_{n-1}^{+,k} b_n \end{aligned} \tag{3.4}$$

et $\{\rho_n^{0,k}, \rho_n^{+,k}\}$ forment une famille d'idempotents orthogonaux de $\mathbb{Q}[B_n]$ telle que

$$\sum_{k=0}^{n} (\rho_n^{0,k} + \rho_n^{+,k}) = \text{Identité} \tag{3.5}$$

Dans le cas de l'exemple 1.1, ceci nous donne une décomposition plus fine de l'homologie $H_n(\mathcal{A}; M)$.

Pour terminer cette section, nous démontrons une généralisation triviale du lemme de Barr [1]. Ce lemme sugère que la décomposition donnée par le théorème 3.2 est encore loin d'être minimale.

Lemme 3.3 *Pour $f \in \mathbb{Q}[B_n]$ nous avons $b_n f = 0$ si et seulement si*

$$f = \sum_{\epsilon \in \mathbb{Z}_n^2} c_\epsilon \, \varepsilon_n 1_\epsilon. \tag{3.6}$$

où $c_\epsilon \in \mathbb{Q}$, $\varepsilon_n = \sum_{\sigma \in \mathcal{S}_n} sgn(\sigma)\sigma$ et $1_\epsilon = (\epsilon_1, \epsilon_2, ..., \epsilon_n) \in B_n$.

Preuve Le lemme de Barr original soutient que pour $u \in \mathbb{Q}[\mathcal{S}_n]$, $b_n u = 0$ si et seulement si $u = c\varepsilon_n$ où $c \in \mathbb{Q}$. D'autre part, si $f \in \mathbb{Q}[B_n]$ alors nous avons la décomposition suivante $f = \sum_{\epsilon \in \mathbb{Z}_n^2} f_{(\epsilon)} 1_\epsilon$ avec $f_{(\epsilon)} \in \mathbb{Q}[\mathcal{S}_n]$. En appliquant b_n de chaque côté de cette décomposition et utilisant le lemme de Barr, on obtient (3.6).

4. Algèbre de Lie hyperoctahédrale libre

Pour raffiner notre décomposition, nous nous devons d'introduire ici certaines notions de la théorie des algèbres de Lie hyperoctahédrales libres. Ceci nous donnera une meilleure compréhension des espaces dans lesquels les idempotent introduits ci-haut projettent. La première partie de cette section peut se retrouver dans [3].

Soit $A = \{\overline{a}_f < \cdots < \overline{a}_1 < a_1 < \cdots < a_f\}$ un alphabet avec deux copies (*négative* et *positive*) de f lettres. Soit A^* l'ensemble de tous les mots dans cet alphabet et A^n l'ensemble de tous les mots de n lettres. Nous définissons l'algèbre $\mathbb{Q}[A^*]$ l'ensemble de toutes les somme formelles finies de mots $\sum c_w w$ avec comme produit la linéarisation du produit de concaténation. L'algèbre $\mathbb{Q}[A^*]$ est graduée par la longueur des mots et nous dénotons les espaces homogènes de degré n par $\mathbb{Q}[A^n]$. Nous définissons dans $\mathbb{Q}[A^*]$ deux nouvelles opérations, la première définie algébriquement par $a_i \mapsto \overline{a}_i$ et $\overline{a}_i \mapsto \overline{\overline{a}}_i = a_i$, et la seconde définie linéairement par $w = b_1 b_2 \cdots b_n \mapsto \overleftarrow{b_1 b_2 \cdots b_n} = \overline{b}_n \overline{b}_{n-1} \cdots \overline{b}_1$. Nous définissons aussi une action de l'algèbre $\mathbb{Q}[B_n]$ sur l'espace homogène $\mathbb{Q}[A^n]$ en étendant l'action $\pi : b_1 b_2 \cdots b_n \mapsto b_{\pi_1} b_{\pi_2} \cdots b_{\pi_n}$ avec la convention que $b_{-i} = \overline{b}_i$.

L'algèbre de *Lie hyperoctahédrale libre* est l'algèbre de Lie engendrée, dans $\mathbb{Q}[A^*]$, par les lettres de A en utilisant le crochet de Lie $[f, g] = fg - gf$. Nous dénotons par Lie(A) cette algèbre et généralement, nous appelons les éléments de Lie(A) des *polynômes de Lie.* Un polynôme de Lie P est dit *positif* si $\overleftarrow{P} = -P$ et *négatif* si $\overleftarrow{P} = P$. Pour ce qui suit, nous utiliserons la lettre "P", respectivement "Q", pour désigner un polynôme de Lie homogène positif, respectivement négatif. Le lecteur vérifiera que les polynômes de Lie positifs forment une sous-algèbre de Lie. Il est possible de construire une base ordonnée $\mathcal{B} = \{Q_1, Q_2, ..., P_1, P_2, ...\}$ de Lie(A) telle que les polynômes Q_i, respectivement P_i, sont des polynômes de Lie homogènes négatifs, respectivement positifs. Nous posons

$$(P_1\cdots P_lQ_1\cdots Q_k)^{S_l\times S_k}=\frac{1}{l!k!}\sum_{\substack{\sigma\in S_l\\ \tau\in S_k}}P_{\sigma_1}\cdots P_{\sigma_l}Q_{\tau_1}\cdots Q_{\tau_k}.$$

Nous appelons $(P_1\cdots P_lQ_1\cdots Q_k)^{S_l\times S_k}$ un *monôme de Lie hyperoctahédral* de degré (l,k). Une variation sur un thème du théorème de Poincaré-Birkoff-Witt donne [3][7] que

$$\{(P_{i_1}\cdots P_{i_l}Q_{j_1}\cdots Q_{j_k})^{S_l\times S_k}\colon P_{i_r},Q_{j_r}\in\mathcal{B}, i_1\geq\cdots\geq i_l \text{ et } j_1\geq\cdots\geq j_k\} \tag{4.1}$$

est une base de $\mathbb{Q}[A^*]$. Nous posons maintenant $\mathrm{Lie}^{l,k}(A)=\mathbb{Q}[(P_1\cdots P_lQ_1\cdots Q_k)^{S_l\times S_k}]$; l'espace linéaire engendré par les monômes de Lie hyperoctahédraux de degré (l,k). La base (4.1) nous donne que

$$\mathbb{Q}[A^*]=\bigoplus_{\substack{l\geq 0\\ k\geq 0}}\mathrm{Lie}^{l,k}(A). \tag{4.2}$$

Pour la suite, nous posons $\mathrm{Lie}_n^{l,k}(A)=\mathrm{Lie}^{l,k}(A)\cap\mathbb{Q}[A^n]$, $\mathrm{Lie}_n^{*,k}(A)=\bigoplus_{l\geq 0}\mathrm{Lie}_n^{l,k}(A)$ et $\mathrm{Lie}_n^{+,k}(A)=\bigoplus_{l>0}\mathrm{Lie}_n^{l,k}(A)$. Nous avons montré dans [3] que $\mathbb{Q}[A^n]\theta\rho_n^k=\mathrm{Lie}_n^{*,k}(A)$ où $\theta\colon\mathbb{Q}[B_n]\to\mathbb{Q}[B_n]$ est l'automorphisme défini par $\pi\mapsto\mathrm{sgn}(\pi)\pi$. En fait, suite à la discussion de la Section 3 les résultats de [3] montrent que

$$\begin{aligned}\mathbb{Q}[A^n]\theta\rho_n^{0,k}&=\mathrm{Lie}_n^{0,k}(A)\\ \mathbb{Q}[A^n]\theta\rho_n^{+,k}&=\mathrm{Lie}_n^{+,k}(A)\end{aligned} \tag{4.3}$$

Rappelons d'autre part que Garsia [7] avait d'abord décomposé l'algèbre $\mathbb{Q}[A^*]$ en utilisant les idempotents e_n^k. Pour cela nous posons

$$(R_1R_2\cdots R_k)^{S_k}=\frac{1}{k!}\sum_{\sigma\in S_k}R_{\sigma_1}R_{\sigma_2}\cdots R_{\sigma_k}$$

où $R_i\in\mathrm{Lie}(A)$; nous oublions la scission positif/négatif. Nous appelons $(R_1R_2\cdots R_k)^{S_k}$ un *monôme de Lie symétrique* de degré k. Garsia a montré que la décomposition

$$\mathbb{Q}[A^*]=\bigoplus_{k\geq 0}\mathrm{Lie}^k(A) \tag{4.4}$$

où $\mathrm{Lie}^k(A)$ est l'espace engendré par les monômes de Lie symétriques de degré k, est obtenu par

$$\mathbb{Q}[A^n]\theta e_n^k=\mathrm{Lie}_n^k(A) \tag{4.5}$$

où $\mathrm{Lie}_n^k(A)=\mathrm{Lie}^k(A)\cap\mathbb{Q}[A^n]$.

Remarque 4.1 La décomposition (4.4) montre que l'algèbre enveloppante $\mathbb{Q}[A^*]$ de $\mathrm{Lie}(A)$ est aussi obtenue comme l'algèbre des puissances symétriques de $\mathrm{Lie}(A)$. Les dimensions des espaces $\mathrm{Lie}^k(A)$ sont intimement reliées aux classes de conjugaisons du groupe symétrique par le biais des nombres de Stirling de seconde espèce. Le lecteur ne s'étonnera pas de voir un lien similaire entre la décomposition (4.2) et les classes de conjugaisons du groupe hyperoctahédral [2][5].

Conjecture 4.2 *Il existe des idempotents orthogonaux $\rho_n^{l,k} \in \mathbb{Q}[B_n]$ tels que*

$$b_n\rho_n^{l,k} = \rho_{n-1}^{l,k}b_n \quad \text{et} \quad \mathbb{Q}[A^n]\theta\rho_n^{l,k} = \mathrm{Lie}_n^{l,k}(A). \tag{4.6}$$

Le théorème 3.2 est un pas dans cette voie. Pour finir cette section nous allons raffiner davantage (3.4) et (3.5) dans la direction désirée par (4.6). Pour cela posons $\rho_n^{l,0} = \rho_n^0 e_n^l \rho_n^0$. Nous avons la proposition suivante qui montre que les éléments $\rho_n^{l,0}$ décomposent ρ_n^0 en idempotents plus fins, orthogonaux à $\rho_n^{*,k}$ pour $k \neq 0$.

Proposition 4.3 *Les éléments $\rho_n^{l,0}$ sont des idempotents orthogonaux tels que*

$$b_n\rho_n^{l,0} = \rho_{n-1}^{l,0}b_n \quad \text{et} \quad \mathbb{Q}[A^n]\theta\rho_n^{l,0} = \mathrm{Lie}_n^{l,0}(A). \tag{4.7}$$

Preuve Nous avons que ρ_n^0 s'exprime simultanément en termes de la décomposition (4.2) et (4.4) par une expression de la forme

$$\rho_n^0 = \sum_{\substack{l\geq 1\\ i_1\geq\cdots\geq i_l}} (P_{i_1}P_{i_2}\cdots P_{i_l})^{S_l}.$$

Donc

$$\rho_n^{l,0} = \sum_{i_1\geq\cdots\geq i_l} (P_{i_1}P_{i_2}\cdots P_{i_l})^{S_l}.$$

De ceci, il est clair que $\rho_n^{l,0}$ est un idempotent. Celui-ci s'annulle sur les espaces $\mathrm{Lie}_n^{l',0}(A)$ pour $l' \neq l$ et est l'identité sur $\mathrm{Lie}_n^{l,0}(A)$, ce qui complète la preuve de la proposition.

5. Homologie de Harisson Hyperoctahédrale $H_*^{1,0}(\mathcal{A}, M)$

Dans l'exemple 1.1, étant donné la conjecture 4.2, nous aurions la décomposition suivante de l'homologie de Hochschild:

$$H_n(\mathcal{A}, M) = \bigoplus_{\substack{k\geq 0\\ l\geq 0}} H_n^{l,k}(\mathcal{A}, M). \tag{5.1}$$

Ici nous allons discuter les composantes suivantes qui sont obtenues par (3.4) et (4.7):

$$H_n^1(\mathcal{A}, M) = H_n^{0,1}(\mathcal{A}, M) \oplus H_n^{1,0}(\mathcal{A}, M). \tag{5.2}$$

Gerstennhaber et Schack ont montré que $H_n^1(\mathcal{A}, M) = e_n^1 H_n(\mathcal{A}, M)$ correspond à l'homologie de Harisson. Plus précisément, $H_n^1(\mathcal{A}, M)$ peut être obtenue des chaînes du complexe $C_*(\mathcal{A}, M)$ qui s'annulent sur les mélanges signés. Ceci est équivalent à dire que θe_n^1 projette dans un espace orthogonal aux produits de mélanges non triviaux. Nous montrons dans cette section une interprétation du même type pour l'homologie $H_n^{1,0}(\mathcal{A}, M)$.

Le *produit de mélange* de deux mots u, v dans A^* est dénoté ici par $u \sqcup\!\sqcup v$. Rappelons que le produit de mélange de deux mots u et v est la somme formelle de tous les mots qui sont obtenus en *mélangeant* u et v. Par exemple $ab \sqcup\!\sqcup cd = abcd + acbd + cabd + cadb + cdab + acdb$. Etant donné le produit scalaire $\langle -, - \rangle$ sur $\mathbb{Q}[A^*]$ pour lequel les mots de A^* sont orthogonaux, Ree [11] a montré la proposition suivante.

Proposition 5.1 [11] *R est un polynôme de Lie si et seulement si*

$$\langle R, u \sqcup\!\sqcup v \rangle = 0 \tag{5.3}$$

pour tous u, v non vides.

Cette proposition démontre que $H_n^1(\mathcal{A}, M)$ est bien l'homologie de Harrison.

Nous introduisons maintenant le produit de mélange hyperoctahédral. Dans ce contexte, nous définissons $\widehat{u} = \sum_{xy=u} \overleftarrow{x} \sqcup\!\sqcup y$. Un produit de mélange hyperoctahédral est soit $u \sqcup\!\sqcup v$, soit $\widehat{u} \sqcup\!\sqcup v$ ou soit $\widehat{u} \sqcup\!\sqcup \widehat{v}$. Nous avons la proposition suivante.

Proposition 5.2 *R est un polynôme de Lie positif si et seulement si*

$$\langle R, u \sqcup\!\sqcup v \rangle = 0 \quad et \quad \langle R, \widehat{u} \rangle = 0 \tag{5.4}$$

pour tous u, v non vides.

Preuve Supposons d'abord que R soit un polynôme de Lie positif. La proposition 5.1, nous donne la première égalité de (5.4) pour tous u, v non vides. D'autre part, nous avons $\langle R, \widehat{u} \rangle = \langle R, u \rangle + \langle R, \overleftarrow{u} \rangle$ puisque tous les autre termes de la somme $\widehat{u}$ sont des produits de mélange non triviaux. Par dualité et positivité de R, nous avons $\langle R, \widehat{u} \rangle = \langle R + \overleftarrow{R}, u \rangle = 0$. Supposons maintenant que (5.4) est satisfaite pour tous u, v non vides. La proposition 5.1 nous donne que R est un polynôme de Lie. Donc $0 = \langle R, \widehat{u} \rangle = \langle R + \overleftarrow{R}, u \rangle$ pour tout u non vide. Nous concluons que $R + \overleftarrow{R} = \sum_{u \neq \emptyset} \langle R + \overleftarrow{R}, u \rangle u = 0$ et donc $R = -\overleftarrow{R}$ est positif.

Les équations (5.4) sont équivalentes à dire que R est orthogonal à tous produits de mélanges hyperoctahédraux non triviaux. Nous avons donc le corollaire suivant.

Corollaire 5.3 *l'homologie $H_n^{1,0}(\mathcal{A}, M)$ est l'homologie de Harisson hyperoctahédrale obtenue des chaînes du complexe $C_*(\mathcal{A}, M)$ qui s'annulent sur les mélanges hyperoctahédraux signés.*

6. Conclusion

Nous avons présentement une preuve de la conjecture 4.2. L'idée est une utilisation judicieuse des familles d'idempotents e_n^l et ρ_n^k pour produire des projections orthogonales dans $\text{Lie}_n^{l,k}(A)$ qui satisfont immédiatement les relations de commutation avec b_n. Nous réservons ce résultat pour un article ultérieur à celui-ci.

D'autre part, plusieurs questions restent encore ouvertes. Est-ce que la décomposition (4.6) est minimale? Le lemme 3.3 semble dire non à cela. Que dire des autres groupes de Coxeter? Pour chaque famille de groupes de Coxeter W_n nous pouvons construire une catégorie $\mathbf{Fin}'_W$ et demander quelle famille d'éléments f_n commutent avec b_n? Mais plus proche de nous, dans un contexte hyperoctahédral, peut-on trouver des règles de commutation avec le bord de Conne B_n qui sont compatibles avec (4.6), pour décomposer l'homologie cyclique? Ceci se ferait à l'intérieur de $\mathbb{Q}[\mathbf{Fin}_B]$ où les morphismes de $\mathbf{Fin}_B$ ne sont pas contraints à $f(0) = 0$.

Bibliographie

[1] M. Barr, *Harrison homology, Hochschild homology and triples*, J. Algebra **8**, (1963), 314-323.

[2] N. Bergeron, *A Hyperoctahedral Analogue of the Free Lie Algebra*, J. Comb. Theory Ser. A, **58** No. 2, (1991), 256-278.

[3] N. Bergeron, *A Decomposition of the Descent Algebra of the Hyperoctahedral Group II*, J. Algebra **147**, (1992).

[4] F. Bergeron et N. Bergeron, *A Decomposition of the Descent Algebra of the Hyperoctahedral Group I*, J. Algebra **147**, (1992).

[5] F. Bergeron et N. Bergeron, *Orthogonal Idempotents in the Descent Algebra of B_n and Applications*, JPAA (accepté 1990).

[6] F. Bergeron et N. Bergeron, R. B. Howlett et D. E. Taylor, *A Decomposition of the Descent Algebra of Finite Coxeter Groups*, J. of Alg. Comb. (accepté 1991).

[7] A. Garsia, *Combinatorics of the Free Lie Algebra and the Symmetric Group*, Analysis, Et Cetera, Research papers published in honor of Jurgen Moser's 60th birthday, ed: Paul H. Rabinowitz and Eduard Zehnder, Academic Press, 1990.

[8] A. Garsia et C. Reutenauer, *A Decomposition of Solomon's Descent Algebras*, Adv. in Math. **77**, No 2, 1989, 189-262.

[9] M. Gerstenhaber et S. D. Schack, *A Hodge-type Decomposition for commutative algebra cohomology*, J. Pure Appl. Algebra **48**, 1987, 229-247.

[10] J.-L. Loday, *Opérations sur l'homologie cyclique des algèbres commutatives*, Inv. Math. **96**, 1989, 205-230.

[11] R. Ree, *Lie elements and the algebra associated with shuffles*, Ann. of Math. **68**, No 2, 1958, 210-219.

[12] L. Solomon, *A Mackey Formula in the Group Ring of a Coxeter Group*, J. Algebra **41**, 1976, 255–264.

Axel Thue's work on repetitions in words*

J. Berstel
L.I.T.P.
Institut Blaise Pascal
Université Pierre et Marie Curie
Paris, France

Abstract

The purpose of this survey is to present, in contemporary terminology, the fundamental contributions of Axel Thue to the study of combinatorial properties of sequences of symbols, insofar as repetitions are concerned. The present state of the art is also sketched.

1 Introduction

In a series of four papers which appeared during the period 1906–1914, Axel Thue considered several combinatorial problems which arise in the study of sequences of symbols. Two of these papers [44, 46] deal with word problems for finitely presented semigroups (these papers contain the definition of what is now called a "Thue system"). He was able to solve the word problem in special cases. It was only in 1947 that the general case was shown to be unsolvable independently by E. L. Post [28] and A. A. Markov [24].

The other two papers [43, 45] deal with repetitions in finite and infinite words. Perhaps because these papers were published in a journal with restricted availability (this is guessed by G. A. Hedlund [20]), this work of Thue was widely ignored during a long time, and consequently some of his results have been rediscovered again and again. Axel Thue's papers on sequences are now more easily accessible since they are included in the "Selected Papers" [47] which were edited in 1977.

It is the purpose of the present paper to give an account of Axel Thue's work on repetitions in sequences, both in more recent terminology and in relation with new results and directions of research. It appears that there is a noticeable difference, both in style and in amount of results, between the 1906 paper (22 pages) and the 1912 paper (67 pages). The first of these papers mainly contains the

*Partially supported by the PRC "Mathématique et Informatique"

construction of an infinite square-free word over three letters. Thue gives also an infinite square-free word over four letters obtained by what is now called an iterated morphism, whilst the three letter word is constructed in a slighly more complicated way (a uniform tag-system, in the terminology of Cobham [12]).

The second paper attacks the more general problem of what Thue calls *irreducible* words. He devotes special attention to the case of two and three letters. In particular, he introduces what is now called the *Thue-Morse sequence*, and shows that all twosided infinite overlap-free words are derived from this sequence. There are several aspects he did not consider: first, many combinatorial properties of the Thue-Morse sequence (such as the number of factors, the recurrence index, and so on) were only investigated by M. Morse [25] or later; next, the characterization of all onesided infinite overlap-free words — which is much more difficult than that of twosided words — was only given later by Fife [15]. However, Thue gives a complete description of circular overlap-free words. We will also mention the problem of counting the number of overlap-free words over two letters.

Axel Thue's investigation of square-free words over three letters is even more detailed. He gives, in this paper, another construction of an infinite square-free word, by iterated morphism, and then initiates, in a 30 pages development, a tentative to describe all square-free words over three letters. He observes that every infinite square-free word is an infinite product of words chosen in a set of six words, and classifies those infinite square-free words that are products of four among these six words. His classification, he observes, is similar both in statement and in proof technique to what is found in diophantine equations: the solutions are parametrized by some variables which are easier to manage.

The paper is organized as follows: after some preliminary definitions, we introduce the so-called Thue-Morse sequence. We next describe Thue's results on this word, and give a short account of other developments about overlap-free words. The next section contains a presentation of Thue's constructions of square-free words, and a comparison with other methods. Then, Thue's classification — which has been ignored for large parts — is described. We end with a short description of *avoidable* patterns, which is the main stream of actual research.

An *alphabet* is a finite set (of *symbols* or *letters*). A *word* over some alphabet A is a (finite) sequence of elements in A. The length of a word w is denoted by $|w|$. The *empty word* of length 0 is denoted by ε. An *infinite* word is a mapping form $\mathbf{N}$ into A, and a twosided infinite word is a mapping from $\mathbf{Z}$ into A. A *circular* word or *necklace* is the equivalence class of a finite word under circular permutation. It can also be considered as a mapping of $\mathbf{Z}/n\mathbf{Z}$ into A for some positive integer n.

A *factor* of a word w is any word u that occurs in w, i. e. such that there exist word x, y with $w = xuy$. A *square* is a nonempty word of the form uu. A word is *square-free* if none of its factors is a square. Similarly, an *overlap* is a word of the form $xuxux$, where x is nonempty. The terminology is justified by the fact that xux has two occurrences in $xuxux$, one as a *prefix* (initial factor) one as a *suffix* (final factor) and that these occurrences have a common part (the central x). As before, a word is *overlap-free* if none of its factors is an overlap.

The set of words over A is the free monoid generated by A. It is denoted by A^*. A function $h : A^* \to B^*$ is a *morphism* if $h(uv) = h(u)h(v)$ for all words u, v. If there is a letter a such that $h(a)$ starts with the letter a, then $h^n(a)$ starts with the letter a for all $n \geq 0$. If the set of words $\{h^n(a) \mid n \geq 0\}$ is infinite, the morphism is *prolongeable* in a and defines a unique infinite word say $\mathbf{x}$ by the requirement that all $h^n(a)$ are prefixes of $\mathbf{x}$. The word $\mathbf{x}$ is said to be obtained by iterating h on a, and $\mathbf{x}$ is also denoted by $h^\omega(a)$. Clearly, $\mathbf{x}$ is a fixed point of h. This construction is frequently used by Axel Thue.

For Axel Thue, a word w over an alphabet of size n is *irreducible* if any two occurrences of the same word as a factor in w are always separated by at least $n-2$ letters. This means that an irreducible word over two letters is *overlap-free* and that an irreducible word over three letters is *square-free.*

2 The Thue-Morse sequence

In this section, we recall some basic properties concerning the Thue-Morse sequence. Other properties and proofs can be found in Lothaire [22] and Salomaa [34].

Let $A = \{a, b\}$ be a two letter alphabet. Consider the morphism μ from the free monoid A^* into itself defined by

$$\mu(a) = ab, \qquad \mu(b) = ba$$

Setting, for $n \geq 0$,

$$u_n = \mu^n(a), \qquad v_n = \mu^n(b)$$

one gets

$$\begin{array}{ll} u_0 = a & v_0 = b \\ u_1 = ab & v_1 = ba \\ u_2 = abba & v_2 = baab \\ u_3 = abbabaab & v_3 = baababba \\ \quad\cdots & \end{array}$$

and more generally

$$u_{n+1} = u_n v_n, \qquad v_{n+1} = v_n u_n$$

and

$$u_n = \overline{v}_n, \qquad v_n = \overline{u}_n$$

where $\overline{w}$ is obtained from w by exchanging a and b. Words u_n and v_n are frequently called *Morse blocks*. It is easily seen that u_{2n} and v_{2n} are palindroms, and that $u_{2n+1} = v^{\sim}_{2n+1}$, where $w^{\sim}$ is the reversal of w. The morphism μ can be extended to infinite words; it has two fixed points

$$\mathbf{t} = abbabaabbaababbabaab \cdots = \mu(\mathbf{t})$$

$$\overline{\mathbf{t}} = baababbaabbabaababba \cdots = \mu(\overline{\mathbf{t}})$$

and u_n (resp. v_n) is the prefix of length 2^n of $\mathbf{t}$ (resp. of $\overline{\mathbf{t}}$). It is equivalent to say that $\mathbf{t}$ is the limit of the sequence $(u_n)_{n \geq 0}$ (for the usual topology on finite and infinite words), obtained by iterating the morphism μ.

The *Thue-Morse sequence* is the word $\mathbf{t}$. There are several other characterizations of this word. Let t_n be the n-th symbol in $\mathbf{t}$, starting with $n = 0$. Then it is easily shown by induction that

$$t_n = \begin{cases} a & \text{if } d_1(n) \equiv 0 \pmod 2 \\ b & \text{if } d_1(n) \equiv 1 \pmod 2 \end{cases}$$

where $d_1(n)$ is the number of bits equal to 1 in the binary expansion of n. For instance, $bin(19) = 10011$, consequently $d_1(19) = 3$, and indeed $t_{19} = a$.

As a consequence, there is a finite automaton computing the values t_n as a function of $bin(n)$. This automaton has two states 0 and 1. It reads the string $bin(n)$ from left to right, starting in state 0. At the end, the state reached is 0 or 1 according to $t_n = b$ or $t_n = a$. In fact, the automaton computes $d_1(n)$ modulo 2. Another description is given by Christol, Kamae, Mendès-France, Rauzy in [11]. There are many generalizations of the Thue-Morse sequence, motivated by its simplicity, and by its numerous properties. The first definition of the sequence, by iterating a given morphism, is of course strongly related to Lindenmayer systems (see e.g. [32]). In the case where the morphism is *uniform*, that is when the lengths of the images of the letters are equal, a general theorem of Cobham [12] shows that the sequence $\mathbf{x}$ obtained by iterating the morphism can also be generated by a finite automaton working on expansions of natural integers in some base k. An equivalent way to state this is to say that there are only finitely many distinct subsequences $(x_{k^r n+s})_{n \geq 0}$ for $r \geq 0$ and $0 \leq s \leq k^r - 1$. Let us call such a sequence *automatic* (more precisely k-automatic). Another extension is by arithmetics. Consider a k letter alphabet $\{0, 1, \ldots, k-1\}$ and define an infinite word $\mathbf{x}$ by taking x_n to be the sum, modulo k, of all the digits in the expression for n in base k. The Thue-Morse sequence is then just the case $k = 2$. Since there is an automaton for computing x_n from the k-ary expansion of n, there is also a uniform morphism generating $\mathbf{x}$. For instance, if $k = 3$, the morphism, say μ_3, is given by $0 \rightarrow 012$, $1 \rightarrow 120$, $2 \rightarrow 201$ (this general definition was in fact already given by Prouhet, in 1851. Several authors, such as Adler, Li [1] and Brlek [6], discuss the fact that Prouhet was the first to mention what perhaps should be called the Prouhet sequence).

Other sequences related to the Thue-Morse sequence are obtained by counting factors in the binary expansion, instead of bits. The *Rudin-Shapiro* [33, 38] sequence is the infinite word $\mathbf{x}$ over $\{a, b\}$ defined by

$$x_n = \begin{cases} a & \text{if } d_{11}(n) \equiv 0 \pmod 2 \\ b & \text{if } d_{11}(n) \equiv 1 \pmod 2 \end{cases}$$

where $d_{11}(n)$ is the number of factors 11 in the binary expansion of n. Similarly, in the sequence of Baum and Sweet [3], the n-th symbol is a or b according to whether there exists a factor of odd length containing only the bit 0 in the binary expansion of n. Again, this sequence is automatic. Many number-theoretic results have been given for automatic sequences. Let us just mention the following, due to Loxton and van der Poorten [23]:

THEOREM *For any automatic infinite word* $\mathbf{x}$ *over the alphabet* $\{0, \ldots, p-1\}$, *the real number*

$$\sum x_n p^{-n}$$

is transcendental.

As an example, the real number whose binary expansion is $0.011010011\cdots$ (associated to the Thue-Morse sequence) is trancendental (this was already known before).

3 Overlap-free words

As already mentioned, the Thue-Morse sequence is overlap-free. Indeed, A. Thue proved[1]

THEOREM (Satz 6) *The sequence* $\mathbf{t}$ *is overlap-free.*

What Thue actually shows, is that a word w over the two letter alphabet $A = \{a, b\}$ is overlap-free iff $\mu(w)$ is overlap-free. Thue observes that the same result holds for circular words. More precisely, he gives the following complete characterization of circular overlap-free words:

THEOREM (Satz 13) *Every circular overlap-free word over the two letter alphabet* $A = \{a, b\}$ *is of the form* $\mu^n(ab)$, $\mu^n(aab)$ *or* $\mu^n(abb)$ *for some* $n \geq 0$.

As a consequence, a circular overlap-free word has length 2^n or $3 \cdot 2^n$ for some $n \geq 0$. These results are interesting because they are related to overlap-free squares. It is indeed easy to show that a circular word w is overlap-free iff the (ordinary) word ww is overlap-free. Thus, Thue characterizes overlap-free squares, a result that was discovered later also by [42]. T. Harju [19] gives a result which is similar, but different.

THEOREM (Satz 9) *For every twosided infinite overlap-free word* $\mathbf{x}$, *there exists a unique infinite overlap-free word* $\mathbf{y}$ *such that* $\mathbf{x} = \mu(\mathbf{y})$.

[1] The mention Satz n refers to theorem n in [45]

This gives, in some sense, a complete description of the set of overlap-free twosided infinite words; indeed, it means that this set is a *minimal* set. More precisely, recall that a *dynamical system* is a set X of infinite words that is closed for the shift operator, defined by $T(\mathbf{x})(n) = \mathbf{x}(n+1)$, and that is closed for the usual topology on infinite words. It is not difficult to show that $\mathbf{x}$ is in X iff $\mathrm{Fact}(\mathbf{x}) \subset \mathrm{Fact}(X)$, where $\mathrm{Fact}(X)$ is the set of finite words that are factors of some element in X. A dynamical system X is minimal if it does not contain strictly any other dynamical system. This means that X is equal to the dynamical system generated by any of its elements, and also that $\mathrm{Fact}(\mathbf{x}) = \mathrm{Fact}(X)$ for any $\mathbf{x} \in X$.

The property that the dynamical system generated by the (twosided) Thue-Morse sequence is minimal was explicitly proved by Gottschalk and Hedlund [16]. As a consequence, every factor appears with bounded gaps (is *recurrent*, in the terminology of M. Morse [25]). Axel Thue (Satz 11) only mentions that every factor appears infinitely many often.

The structure of onesided infinite overlap-free words is more complicated. Axel Thue was interested in the tree of infinite overlap-free words and tried to characterize those overlap-free words which can be extended into infinite overlap-free words. His main result in this direction is

THEOREM (Satz 15) *Let w be an overlap-free word of w length n such that there exist words u and v of length $8n$ with the property that uwv is still overlap-free. Then any overlap-free word x of length $26n$ contains w as a factor.*

In the proof of this result, he shows that the word x contains a Morse block which contains w, and he concludes that w is indefinitely extensible in both directions. An explicit description of the tree of infinite overlap-free words by means of a finite automaton was given by E. D. Fife and deserves a mention.

Fife defines three operators on words, say α, β, γ, and he shows that every overlap-free infinite words is the "value" of some infinite word $\mathbf{f}$ in the three operators, provided the word $\mathbf{f}$ is in some rational set he gives explicitely. To be more precise, let $X_n = \{u_n, v_n\}$ be the set of Morse blocks of index n and let $X = \bigcup_{n\geq 0} X_n$. Any word $w \in A^*X_1$ admits a *canonical decomposition* $(z, y, \bar{y})$ where y is the longest word in X such that $w = zy\bar{y}$. It is equivalent to say that $(z, y, \bar{y})$ is the canonical decomposition of w if $\bar{y}y$ is not a suffix of z. As an example, the canonical decomposition of $aabaabbabaab$ is

$$(aaba, abba, baab)$$

and the decomposition of $abaabbaababbaabbabaab$ is

$$(abaab, baababba, abbabaab)$$

The three functions $\alpha, \beta, \gamma : A^*X_1 \longrightarrow A^*X_1$, acting on the right, are defined as follows for a word $w \in A^*X_1$ with canonical decomposition $(z, y, \bar{y})$:

$$w \cdot \alpha \quad = zy\bar{y} \cdot \alpha = zy\bar{y}y\bar{y}y\bar{y} = wy\bar{y}y\bar{y}$$

$$w \cdot \beta \;= zy\bar{y} \cdot \beta = zy\bar{y}y\bar{y}\bar{y}y = wy\bar{y}\bar{y}y$$
$$w \cdot \gamma \;= zy\bar{y} \cdot \gamma = zy\bar{y}\bar{y}y = w\bar{y}y$$

Since w is a prefix of $w \cdot \alpha$, $w \cdot \beta$, and of $w \cdot \gamma$, it makes sense to define $w \cdot f$ by induction for all "words" f in B^*, with $B = \{\alpha, \beta, \gamma\}$. By continuity, $w \cdot \mathbf{f}$ is definde also for infinite words $\mathbf{f}$. Here are some examples:

$$ab \cdot \alpha = abaab$$
$$ab \cdot \beta = ababba$$
$$ab \cdot \gamma = abba$$
$$ab \cdot \gamma^\omega = \mathbf{t}$$
$$aab \cdot \alpha = aabaab = a(ab \cdot \alpha)$$
$$ab \cdot \alpha\beta\gamma = abaababbabaababbaabbabaab$$

Observe that the last word contains an overlap. Note also that, for $w \in A^*X_1$ and $f \in B^*$, one has $\mu(w \cdot f) = \mu(w) \cdot f = w \cdot \gamma f$. A *description* of an infinite word $\mathbf{x}$ starting with ab or aab is an infinite word $\mathbf{f}$ over B such that $\mathbf{x} = ab \cdot \mathbf{f}$ or $\mathbf{x} = aab \cdot \mathbf{f}$, according to $\mathbf{x}$ starts with ab or aab.

PROPOSITION *Every infinite overlap-free word starting with the letter a admits a unique description.*

Let
$$F = B^\omega - B^*IB^\omega$$
be the (rational) set of infinite words over B having no factor in the set
$$I = \{\alpha, \beta\}(\gamma^2)^*\{\beta\alpha, \gamma\beta, \alpha\gamma\}$$
and let G bet the set of words $\mathbf{f}$ such that $\beta\mathbf{f}$ is in F. Then:

THEOREM (Fife's Theorem) *Let $\mathbf{x}$ be an infinite word over $A = \{a, b\}$.*

(*i*) *if $\mathbf{x}$ starts with ab, then $\mathbf{x}$ is overlap-free iff its description is in F;*

(*ii*) *if $\mathbf{x}$ starts with aab, then $\mathbf{x}$ is overlap-free iff its description is in G.*

A direct consequence is the following

COROLLARY *An overlap-free word w is the prefix of an infinite overlap- free word iff w is a prefix of a word $ab \cdot f$ with $f \in W$ or of a word $aab \cdot f$ with $\beta f \in W$, where $W = B^* - B^*IB^*$.*

This implies in particular a result of Restivo and Salemi [30], namely that it is decidable whether an overlap-free word is extensible into an infinite overlap-free word. Another consequence of Fife's description is the following

COROLLARY *The Thue-Morse word $\mathbf{t}$ is the greatest infinite overlap-free word, in lexicographical order, that start with the letter a.*

Indeed, the choice of the letters α, β, et γ implies that if $\mathbf{f} \le \mathbf{f}'$, then $ab \cdot \mathbf{f} \le ab \cdot \mathbf{f}'$. The greatest word in F is γ^ω, and this shows the corollary. A. Carpi [8] has developed a description for finite overlap-free words by means of a finite automaton. Unfortunately, his automaton is rather big (more than 300 states).

There is another property that singles out the Thue-Morse word (and which was rediscovered and generalized by P. Séébold [37]). Call a morphism *overlap-free* if the image of an overlap-free word is always overlap-free.

THEOREM (Satz 16) *Let h be an overlap-free morphism. Then there is an integer n such that $h = \mu^n$ or $h = \pi \circ \mu^n$, where π is the morphism that exchanges the two letters of the alphabet.*

Thus, the infinite words $\mathbf{t}$ and $\bar{\mathbf{t}}$ are the only infinite overlap-free words generated by iterated morphisms.

Since overlap-free words have a strong structure, it seems natural to count them. The first result is due to Restivo and Salemi [30]. They prove that the number γ_n of overlap-free words over two letters grows polynomially in n (in fact slower than n^4). Kobayashi [21] has used Fife's theorem to derive the lower of the more precise bounds for γ_n :

THEOREM *There are constants C_1 and C_2 such that*

$$C_1 n^\alpha < \gamma_n < C_2 n^\beta$$

where $\alpha = 1.155\ldots$ and $\beta = 1.5866\ldots$.

One might ask what is the "real" limit. In fact, a recent and surprising result by J. Cassaigne [10] shows that there is no limit. More precisely, set

$$\alpha' = \sup\{r \mid \exists C > 0, \forall n, \gamma_n \geq Cn^r\}$$

and

$$\beta' = \inf\{r \mid \exists C > 0, \forall n, \gamma_n \leq Cn^r\}$$

Then

THEOREM *One has $1.155 < \alpha' < 1.276 < 1.332 < \beta' < 1.587$.*

This is to be compared with the situation for square-free words. Indeed, Brandenburg [5] proved that for the number $c(n)$ of square-free words of length n over three letters, there are constants $c_1 \geq 1.032$ and $c_2 \leq 1.38$ such that $6c_1^n < c(n) < 6c_2^n$. Brandenburg also proved that the number of cube-free words over two letters grows exponentially.

4 Square-free words

4.1 First examples

It is easily seen that the only square-free words over two letters are a, b, ab, ba, aba, bab. However, there exist arbitrarily long square-free words over three letters, and by a simple argument, there exist infinite square-free words over three letters. Historically the first infinite square-free word was given by Thue in his 1906 paper. It is over four letters, and it is obtained by iterating the following morphism h, starting with the letter a:

$$\begin{array}{rcl} a & \mapsto & adbcb \\ b & \mapsto & abdcb \\ c & \mapsto & abcdb \\ c & \mapsto & abcbd \end{array}$$

Thue explains his construction as follows : take a square-free word over three letters, here *abcb*, and interleave it with the letter d. This gives the morphism. The proof is not very difficult.

In the same paper, Thue gives another infinite square-free word, over three letters. The word is by iterating the following construction: given a square-free word w over $A = \{a, b, c\}$, build $\alpha(w)$ by replacing each letter a by *abac*, each b by *babc*, and each c either by *bcac* or by *acbc*, according to the letter preceding c in w is a or b. Starting with a, one gets an infinite word

$$abacbabcabacbcacbabcabacbabcacbcabacbabc\cdots$$

which he shows to be square-free. Although the definition is not by a morphism, the construction is very close to it. There exist several ways to formulate it differently: in fact, one has a fourth letter hidden in the description, which appears when we note differently a letter c preceded by an a and a letter c preceded by a b. The four letter word thus obtained is generated by a morphism, and at the end, the two variants of the letter c are identified.

In the 1912 paper, Axel Thue gives a morphism for generating an infinite square-free word over three letters. The morphism is the following (Satz 18):

$$\begin{array}{rcl} a & \mapsto & abcab \\ b & \mapsto & acabcb \\ c & \mapsto & acbcacb \end{array}$$

This morphism seems to be rather complicated. Its *size*, i. e. the sum of the length of the images, is 18. It has been shown by A. Carpi [7] that this is the best bound : every morphism over three letters that preserves square-free words has size at least 18. (See also the discussion in [4].) However, there is a simpler morphism that generates a square-free word (starting with a) given e.g. by Hall [18], namely

$$\begin{array}{rcl} a & \mapsto & abc \\ b & \mapsto & ac \\ c & \mapsto & b \end{array}$$

This mophism does not preserve square-free words, because the image of *aba* is *abcacabc*.

4.2 A Classification

Since every twosided infinite square-free word $\mathbf{x}$ over three letters a, b and c is some product of the six words in the set

$$X = \{ab, abc, abcb, ac, acb, acbc\}$$

Thue studies a classification according to words of X that appear in $\mathbf{x}$. It is quite remarkable that he achieves a classification of those square-free infinite words that contain exactly four of the six words in X. After some discussion, he reduces the 15 cases (two words lacking among six) to the following three cases :

$$aca \text{ and } bcb \tag{I}$$

$$aba \text{ and } aca \tag{II}$$

$$aba \text{ and } bab \tag{III}$$

are missing in the infinite word under consideration. In order to describe these three families, he gives some "parametrization", and as we will see, reduces them to minimal dynamical systems.

Consider first square-free words of type (I), i. e. without occurrences of aca or bcb. Define a morphism h from $A = \{a, b, c\}$ into $B = \{\alpha, \beta\}$ by

$$\begin{array}{rcl} a & \mapsto & \alpha \\ b & \mapsto & \alpha\beta\beta \\ c & \mapsto & \alpha\beta \end{array}$$

Then the following holds

THEOREM (Satz 20, 21) *If* $\mathbf{x}$ *is a square-free infinite word of type* (I), *then* $h(\mathbf{x})$ *is overlap-free. Conversely, for every overlap-free word* $\mathbf{y}$, *there exists a unique word* $\mathbf{x}$ *such that* $h(\mathbf{x}) = \mathbf{y}$, *and* $\mathbf{x}$ *is square-free of type* (I).

Thus, the square-free words of type (I) are described by the (minimal) set of overlap-free words over two letters. For the two other types, the situation is slightly more involved (and the proofs are more difficult). First, Thue observes that the cases (II) and (III) reduce one to each other. Any word $\mathbf{x}$ of type (II) is uniquely decomposable as a product of words in the set $\{ca, cb, cab, cba\}$. Let s be the substitution defined by

$$\begin{array}{rcl} ca & \mapsto & abc \\ cb & \mapsto & acb \\ cab & \mapsto & abcb \\ cba & \mapsto & acbc \end{array}$$

For a word $\mathbf{x}$ of type (II), the word $y = s(\mathbf{x})$ is of type (III), and conversely, every word $\mathbf{y}$ of type (III) is of this form. Thus, it suffices to describe square-free words of type (II). For this, Thue introduces a new, five letter alphabet $\{A, B, C, D, E\}$, and a morphism $h : \{A, B, C, D, E\}^* \rightarrow \{a, b, c\}^*$ defined by

$$\begin{aligned} A &\mapsto abcbacbcacbabcacbc \\ B &\mapsto abcbacbc \\ C &\mapsto abcbacbcacb \\ D &\mapsto abcbabcacbc \\ E &\mapsto abcbabcacb \end{aligned}$$

Finally, he defines a set of words

$$\begin{aligned} W = \; & \{AB, AD, BA, BC, CA, CD, CE, DB, DE, EC, ED, \\ & BEB, EBE, DAC, DCBD, CBDC\} \end{aligned}$$

In order to state simply the next theorem, let us denote by $\mathcal{Y}$ the set of twosided infinite square-free words over the five letter alphabet $\{A, B, C, D, E\}$ that have no factors in W. Then Thue proves

THEOREM (Satz 26) *The set of twosided infinite square-free words of type (II) is the set of words of the form* $h(\mathbf{y})$ *for* $\mathbf{y}$ *in* $\mathcal{Y}$.

This theorem seems to be a little disappointing, since a rather simple description of three letter square-free words is replaced by a cumbersome and complicated family $\mathcal{Y}$ of words over five letters. However, this family has an important property: let α be the morphism from $\{A, B, C, D, E\}^*$ into itself defined by

$$\begin{aligned} A &\mapsto BDAEAC \\ B &\mapsto BDC \\ C &\mapsto BDAE \\ D &\mapsto BEAC \\ E &\mapsto BEAE \end{aligned}$$

THEOREM (Satz 23,24) *The morphism* α *is a bijection of the set* $\mathcal{Y}$ *onto itself.*

Thus, as before, the set $\mathcal{Y}$ is a minimal dynamical system.

A full description of the tree of square-free words, like Fife's description for overlap-free words, is not yet available. Shelton and Soni have investigated this tree [39, 40, 41]. They have shown in particular that the set of infinite square-free words over three letters is perfect. Roughly speaking, this means if $\mathbf{x}$ is any square-free infinite word, then for any prefix p of $\mathbf{x}$, there are infinitely many infinite square-free words that have p as prefix. They show also that it is decidable whether a square-free word p of length n is a prefix of some infinite square-free word, and their procedure is "uniform": There is a constant K such that if there exists a word q of length $n + Kn^{3/2}$ such that pq is square-free, then p is the prefix of some infinite square-free word.

4.3 Repetitions

As already mentioned, Thue calls a word on n letters *irreducible* if every factor xyx verifies $|y| \geq n - 2$. A more general concept, first considered by F. Dejean [14], is to require that the length of the word y separating the occurrences of x is bounded from below by the length of x (times some factor). More precisely, we call *repetition* a word xyx with x non empty, and *index* of this repetition the quotient $|y|/|x|$. We are looking for words where all repetitions have high index. F. Dejean has proved that there exists an infinite word over 3 letters that has only repetitions of index greater or equal to 1/3, and she also shows that this bound is the best possible. Call *repetition threshold* the smallest number ρ_k such that there exists an infinite word over k letters that has only repetitions of index greater or equal to ρ_k. Thus, Dejean's result may be stated as : $\rho_3 = 1/3$. She conjectured that $\rho_4 = 3/2$, a result proved by Pansiot [27], and that $\rho_k = k - 2$ for $k \geq 5$. The conjecture was proved up to 9 by Moulin-Ollagnier [26]

5 Avoidable patterns

The overlap-freeness of the Thue-Morse sequence, and the square- freeness of the other words we have presented can be expressed in the more general framework of avoidable and unavoidable patterns in strings. This concept has been introduced in the context of equations defining algebras. Certain unavoidable words have been used e.g. in [35] to characterize those finite semigroups S that are inherently nonfinitely based, in the sense that S is not a member of any locally finite semigroup variety definable by finitely many equations. It may be noticed that Axel Thue replaces his research on repetitions in strings in an even slightly more general context, since he considers avoiding patterns with constants. However, he has not stated results in this specific framework.

Consider an alphabet E of "pattern symbols". A word e over E is a pattern. A pattern e is said to *occur* in some word $w \in A^*$ if there is a nonerasing morphism $h : E^* \to A^*$ such that $h(e)$ is a factor of w. A pattern e is *avoidable* over k letters, or is k-avoidable, if there is an infinite word $\mathbf{x}$ over k letters such that e does not occur in $\mathbf{x}$. The Thue-Morse sequence shows that the patterns aaa and $ababa$ are 2-avoidable, and square-free infinite words show that aa is 3- avoidable (but not 2-avoidable). Avoidable and unavoidable patterns have been studied by several people (Zimin [48], Schmidt [36], Bean, Ehrenfeucht, McNulty [4], Roth [31], Cassaigne [9], Goralcik, Vanicek [17], Baker, McNulty, Taylor [2], Crochemore, Goralcik [13]).

Problems which have been stated, and partially solved, include the following: given a pattern e, is it avoidable or not ? There is a nice algorithm in [4], and basically the same in [48], to decide whether a pattern is avoidable. The complexity of their algorithm is at least exponential. P. Roth (personal communication) recently has proved that the general problem is NP-complete.

For a pattern e, denote by $\alpha(e)$ the number of distinct letters occurring in e. Every pattern e such that $|e| \geq 2^{\alpha(e)}$ is in fact avoidable, and this is the best possible bound because there exists an unavoidable pattern of length $2^n - 1$ over an n letter pattern alphabet. The next problem is to determine, for some unavoidable pattern e, the size $\mu(e)$ of the smallest k such that e is k-avoidable. The first word that is 4-avoidable but not 3- avoidable has been given by [2]. Upper bounds of μ, as a function of α are also given there. Recently, Roth [31], Cassaigne [9], Goralcik, Vanicek [17] have solved the problem of determining all the 2-avoidable binary patterns.

References

[1] A. ADLER, S. LI, Magic cubes and Prouhet sequences, *American Math. Monthly* (1977), 618–627.

[2] K. A. BAKER, G. F. MCNULTY, W. TAYLOR, Growth problems for avoidable words, *Theoret. Comput. Sci.* **69** (1989), 319–345.

[3] L. BAUM, M. SWEET, Continued fractions of algebraic power series in characteristic 2, *Ann. Math.* **103** (1976), 593–610.

[4] D. R. BEAN, A. EHRENFEUCHT, G. F. MCNULTY, Avoidable patterns in strings of symbols, *Pacific J. of Math.* **85** (1979), 261–294.

[5] F. J. BRANDENBURG, Uniformely growing k-th powerfree homomorphisms, *Theoret. Comput. Sci.* **23** (1983), 69–82.

[6] S. BRLEK, Enumeration of factors in the Thue-Morse word, *Discr. Appl. Math.* **24**, (1989), 83–96.

[7] A. CARPI, On the size of a squarefree morphism on a three letter alphabet, *Inform. Proc. Letters* **16** (1983), 231–236.

[8] A. CARPI, Overlap-free words and finite automata, manuscript, 1990.

[9] J. CASSAIGNE, Unavoidable binary patterns, manuscript, Ecole normale supérieure, Paris, 1991, submitted.

[10] J. CASSAIGNE, Counting overlap-free binary words, manuscript, Ecole normale supérieure, Paris, 1992.

[11] G. CHRISTOL, T. KAMAE, M. MENDÈS FRANCE, G. RAUZY, Suites algébriques, automates et substitutions, *Bull. Soc. Math. France* **108** (1980), 401–419.

[12] A. COBHAM, Uniform tag sequences, *Math. Systems Theory* **6** (1972), 164–192.

[13] M. CROCHEMORE, P. GORALCIK, Mutually avoiding ternary words of small exponents, *Intern. J. Algebra Comput.*, 1992, to appear.

[14] F. DEJEAN, Sur un théorème de Thue, *J. Combin. Th. A* **13** (1972), 90–99.

[15] E. D. FIFE, Binary sequences which contain no *BBb*, *Trans. Amer. Math. Soc.* **261** (1980), 115–136.

[16] W.H. GOTTSCHALK, G.A. HEDLUND, A characterization of the Morse minimal set, *Proc. Amer. Math. Soc.* **15** (1964), 70–74.

[17] P. GORALCIK, T. VANICEK, Binary patterns in binary words, manuscript, 1991, submitted.

[18] M. HALL, Generators and relations in groups – the Burnside problem, in T. L. Saaty (ed) *Lectures on Modern Mathematics* **2**, Wiley, 1964, 42–92.

[19] T. HARJU, On cyclically overlap-free words in binary alphabets, *The Book of L*, Springer-Verlag, 1986, 123–130.

[20] G.A. HEDLUND, Remarks on the work of Axel Thue, *Nordisk Mat. Tidskr.* **15** (1967), 148–150.

[21] Y. KOBAYASHI, Enumeration of irreductible binary words, *Discrete Appl. Math.***20** (1988), 221–232.

[22] M. LOTHAIRE, *Combinatorics on Words*, Addison-Wesley, 1983.

[23] J. H. LOXTON, A. J. VAN DER POORTEN, Arithmetic properties of the solutions of a class of functional equations, *J. reine angew. Math.* **330** (1982), 159–172.

[24] A. A. MARKOV, Impossibility of certain algorithms in the theory of associative systems *Dokl. Akad. Nauk. SSSR* **55** (1941), 587–590.

[25] M. MORSE, Recurrent geodesics on a surface of negative curvature, *Transactions Amer. Math. Soc.* **22** (1921), 84–100.

[26] J. MOULIN-OLLAGNIER, Preuve de la conjecture de Dejean pour des alphabets à 5, 6, 7, 8, 9 lettres, Prépublication du département de mathématique et informatique, Université Paris-Nord, Nr. 89–4, 1989.

[27] J.-J. PANSIOT, A propos d'une conjecture de F. Dejean sur les répétitions dans les mots, *Discrete Appl. Math.*7 (1984), 297–311.

[28] E. L. POST, Recursive unsolvability of a problem of Thue, *J. Symbolic Logic* **11** (1947), 1–11.

[29] M. E. PROUHET, Mémoire sur quelques relations entre les puissances des nombres, *C. R. Acad. Sci. Paris.* **33** (1851), 31.

[30] A. RESTIVO, S. SALEMI, Overlap-free words on two symbols, in: *Automata on infinite words*, Nivat, Perrin (eds), Lect. Notes Comp. Sci.,**192**, Springer-Verlag, 1985, 198–206.

[31] P. ROTH, Every binary pattern of length six is avoidable on the two-letter alphabet, *Acta Informatica* (1992).

[32] G. ROZENBERG, A. SALOMAA, *The Mathematical Theory of L-Systems*, Academic Press, 1980.

[33] W. RUDIN, Some theorems on Fourier coefficients, *Proc. Amer. Math. Soc.* **10** (1959), 855–859.

[34] A. SALOMAA, *Jewels of Formal Language Theory*, Computer Science Press, 1981.

[35] M. SAPIR, Inherently nonfinitely based finite semigroups, *Mat. Sb.* **133** (1987), 154–166.

[36] U. SCHMIDT, Avoidable patterns on two letters, *Theoret. Comput. Sci.* **63** (1989), 1–17.

[37] P. Séébold, Sequences generated by infinitely iterated morphisms, *Discrete Appl. Math.*11, (1985), 255–264.

[38] H. S. SHAPIRO, Extremal problems for polynomials and power series, Thesis, M.I.T., 1951.

[39] R. SHELTON, Aperiodic words on three symbols I, *J. Reine Angew. Math.* **321** (1981), 195–209.

[40] R. SHELTON, R. SONI, Aperiodic words on three symbols II, *J. Reine Angew. Math.* **327** (1981), 1–11.

[41] R. SHELTON, R. SONI, Aperiodic words on three symbols III, *J. Reine Angew. Math.* **330** (1982), 44–52.

[42] R. SHELTON, R. SONI, Chains and fixing blocks in irreducible sequences, *Discrete Math.*54 (1985), 93–99.

[43] A. THUE, Über unendliche Zeichenreihen, *Kra. Vidensk. Selsk. Skrifter. I. Mat.-Nat. Kl.*, Christiana 1906, Nr. 7.

[44] A. THUE, Die Lösung eines Spezialfalles eines generellen logischen Problems, *Kra. Vidensk. Selsk. Skrifter. I. Mat.-Nat. Kl.*, Christiana 1910, Nr. 8.

[45] A. THUE, Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen, *Kra. Vidensk. Selsk. Skrifter. I. Mat.-Nat. Kl.*, Christiana 1912, Nr. 10.

[46] A. THUE, Probleme über Veränderungen von Zeichenreihen nach gegebenen Regeln, *Kra. Vidensk. Selsk. Skrifter. I. Mat.-Nat. Kl.*, Christiana 1914, Nr. 7.

[47] A. THUE, *Selected Mathematical Papers*, edited by T. Nagell, A. Selberg, S. Selberg, K. Thalberg, Universitetsforlaget, Oslo 1977.

[48] A. I. ZIMIN, Blocking sets of terms, *Math. USSR Sb.* **47**, (1984), 353–364.

ARBRES, ARBORESCENCES ET RACINES CARRÉES SYMÉTRIQUES.

PIERRE BOUCHARD, YVES CHIRICOTA, ET GILBERT LABELLE

31 mars 1992

1. Introduction

Soit $\mathbf{C}$ le corps des nombres complexes et $\mathcal{A}$ l'ensemble des espèces atomiques (à isomorphisme près). Étant donnée une espèce $F \in \mathbf{C}[[\mathcal{A}]]$, l'équation $G^2 = F$ n'a pas toujours de solution G dans $\mathbf{C}[[\mathcal{A}]]$. Par exemple, l'équation $G^2 = X$ n'en a pas dans $\mathbf{C}[[\mathcal{A}]]$. Si toutefois il existe une espèce G qui satisfait l'équation, on l'appelle une *racine carrée* de F. Le contexte des espèces nous permet de généraliser la notion de racine carrée en introduisant la *racine carrée symétrique* d'une espèce qui se définit comme suit:

Définition 1.1. *Soit $G \in \mathbf{C}[[\mathcal{A}]]$. On dira que F est une* racine carrée symétrique *de G si l'équation suivante est vérifiée:*

$$E_2(G) = F.$$

Dans le cas d'une racine carrée symétrique, on trouve toujours une solution comme nous le verrons à la section 3. Cette notion trouve une belle application en nous permettant de montrer qu'à une transformation affine près, l'espèce des arborescences est la racine carrée symétrique de l'espèce des arbres.

Nous terminerons en donnant quelques généralisations et directions pour des recherches ultérieures.

2. Substitution générale dans l'espèce E_2

Il est possible de munir l'ensemble des espèces moléculaires (à isomorphisme près) d'un ordre total. Dans un premier temps on peut ordonner ces dernières par degré et ensuite décréter un ordre total sur les espèces de même degré. Par exemple, voici l'ordre total que nous utiliserons pour les espèces moléculaires de degré ≤ 4:

$$1 < X < E_2 < X^2 < E_3 < C_3 < XE_2 < X^3$$
$$< E_4 < E_4^{\pm} < E_2(E_2) < XE_3 < E_2^2$$
$$< P_4^{\mathrm{bic}} < C_4 < XC_3 < X^2E_2 < E_2(X^2) < X^4.$$

Nous noterons $\mathcal{M}$ l'ensemble des espèces moléculaires muni d'un ordre total qui prolonge l'ordre ci-dessus.

Nous utiliserons la proposition suivante qui est une généralisation d'une formule de substitution de Yeh [15],[16] au cas où l'espèce substituée est à terme constant quelconque.

Proposition 2.1. *Soit* $G = \sum_{M \in \mathcal{M}} g_M M$, *alors*

$$(1) \qquad E_2(G) = \sum_{M \in \mathcal{M}} g_M E_2(M) + \sum_{M \in \mathcal{M}} \binom{2}{g_M} M^2 + \sum_{\substack{P,Q \in \mathcal{M} \\ P<Q}} g_P g_Q PQ.$$

Notons que puisque l'espèce $E_2(X)$ est polynomiale, il est possible d'y substituer une espèce à coefficient constant non nul [7]. En particulier, on a que $E_2(1) = 1$.

3. Racine carrée symétrique

Nous débutons avec la proposition centrale qui nous renseigne sur l'existence de solutions G de $E_2(G) = F$ étant donnée $F \in \mathbb{C}[[\mathcal{A}]]$.

Proposition 3.1. *soit* $F \in \mathbb{C}[[\mathcal{A}]]$ *dont la décomposition en espèces moléculaires est* $F = \sum_{M \in \mathcal{M}} f_M M$. *Alors*

(1) *Si le terme constant* f_1 *de* F *est différent de* 0 *ou de* $-1/8$, *alors il existe dans* $\mathbb{C}[[\mathcal{A}]]$ *deux solutions* G *de l'équation* $E_2(G) = F$.
(2) *Si le terme constant* f_1 *de* F *est* $-1/8$, *alors il existe dans* $\mathbb{C}[[\mathcal{A}]]$ *une seule solution* G *de l'équation* $E_2(G) = F$.
(3) *Si le terme constant* f_1 *de* F *est* 0, *alors*
   - *il existe dans* $\mathbb{C}[[\mathcal{A}]]$ *une seule solution* G *de l'équation* $E_2(G) = F$ *dont le terme constant* g_1 *soit non-nul: ce terme constant est alors* $g_1 = -1$;
   - *il existe dans* $\mathbb{C}[[\mathcal{A}]]$ *au plus une solution* G *de l'équation* $E_2(G) = F$ *dont le terme constant* g_1 *soit nul.*

Démonstration: Cherchons les coefficients g_M (pour tout $M \in \mathcal{M}$) tels que si on pose $G = \sum_{M \in \mathcal{M}} g_M M$, alors $E_2(G) = F$.

En utilisant la proposition 2.1, on trouve

$$\begin{aligned}
E_2(G) &= E_2(\sum_{N \in \mathcal{M}} g_N N) \\
&= \sum_{N \in \mathcal{M}} g_N E_2(N) + \sum_{N \in \mathcal{M}} \binom{g_N}{2} N^2 + \sum_{\substack{P,Q \in \mathcal{M} \\ P<Q}} g_P g_Q PQ \\
&= g_1 + g_X E_2(X) + g_{E_2} E_2(E_2(X)) + g_{X^2} E_2(X^2) + \ldots \\
&\quad + \binom{g_1}{2} + \binom{g_X}{2} X^2 + \binom{g_{E_2(X)}}{2} E_2(X)^2 + \binom{g_{X^2}}{2} X^4 + \ldots \\
&\quad + g_1 g_X X + g_1 g_{X^2} X^2 + g_1 g_{E_2} E_2(X) + \ldots \\
&\quad + g_X g_{E_2} X E_2(X) + g_X g_{X^2} X^3 + \ldots \\
&\quad + \ldots \\
(2) \qquad &= \sum_{M \in \mathcal{M}} f_M M
\end{aligned}$$

Montrons par récurrence (en considérant l'ordre sur $\mathcal{M}$ décrit plus haut) que la donnée des f_M détermine les solutions en les g_M. En comparant les coefficients constants de chaque membre de (2) et en isolant g_1, on trouve

$$g_1 = \frac{-1 \pm \sqrt{1+8f_1}}{2},$$

donc à priori g_1 peut prendre deux valeurs distinctes sauf dans le cas où $f_1 = -1/8$ où g_1 ne peut prendre que la valeur $-1/2$. Supposons pour le moment $f_1 \neq 0$. Alors $g_1 \neq 0$. En comparant cette fois les coefficients de X on trouve $g_X g_1 = f_X$. Ce qui permet de trouver $g_X = f_X/g_1$.

Fixons maintenant $M \in \mathcal{M}$ et supposons que si $M_1 < M$ alors la valeur de M_1 est complètement déterminée pour une valeur choisie (ou imposée si aucun autre choix n'est possible) pour g_1. Il faut calculer g_M. Trois cas se présentent.

Premièrement, M est de la forme $M = E_2(N)$. Dans ce cas, il est clair que $N < M$ car le degré de N est plus petit que celui de M et l'ordre total se fait en premier lieu d'après le degré. Donc g_N est complètement déterminé et en comparant les coefficients de M dans l'équation 2, on trouve que $g_N + g_1 g_M = f_M$. Ainsi,

$$g_M = \frac{f_M - g_N}{g_1}.$$

Deuxièmement, M est de la forme $M = N^2$. Dans ce cas on trouve

$$f_M = \binom{g_N}{2} + g_1 g_M + \sum_{\substack{PQ=N^2 \\ 1<P<Q<N^2}} g_P g_Q.$$

Ici aussi on peut utiliser l'hypothèse d'induction pour conclure que g_M est complètement déterminé. Il suffit pour celà d'isoler g_M ci-dessus. On trouve

$$g_M = \frac{\phi_1}{g_1},$$

où ϕ_1 ne dépend que des T tels que $T < M$.

Troisièmement, si M n'est pas de l'une des deux formes précédentes, alors

$$f_M = g_1 g_M + \sum_{\substack{1<P<Q \\ PQ=M}} g_P g_Q,$$

ce qui nous permet d'écrire $g_M = \frac{\phi_2}{g_1}$, où ϕ_2 ne dépend que des T tels que $T < M$.

Dans le cas $f_1 = 0$, la seule possibilité non-nulle pour g_1 est $g_1 = -1$ et le raisonnement ci-dessus nous conduit à une unique solution G. Par contre, si on cherche une solution avec $g_1 = 0$, la comparaison des coefficients dans (2) nous permet de conclure que

$$g_M = f_{E_2(M)}, \tag{3}$$

le choix des g_M est donc forcé, ce qui montre qu'il y a au plus une solution G de $E_2(F) = G$. En plus de la relation 3, d'autres relations possiblement contradictoires avec 3, peuvent s'ajouter, ce qui fait qu'il peut ne pas y avoir de solution. C'est le cas par exemple si $F = X$. ∎

Étant donnée F, il est donc facile de calculer explicitement les valeurs de g_M en fonction des f_M et de g_1, si $g_1 \neq 0$. Remarquons que si M est atomique et n'est pas de la forme $E_2(N)$, alors $g_M = f_M/g_1$. Quelques calculs montrent que

$$g_{E_2} = \frac{g_1 f_{E_2} - f_X}{g_1^2},$$

$$g_{X^2} = \frac{2g_1^2 f_{X^2} - f_X^2 + g_1 f_X}{2g_1^3},$$

$$g_{E_3} = \frac{f_{E_3}}{g_1},$$

$$g_{XE_2} = \frac{g_1^3 f_{XE_2} - g_1 f_X f_{E_2} + f_X^2}{g_1^4},$$

$$g_{C_3} = \frac{f_{C_3}}{g_1},$$

$$g_{X^3} = \frac{2g_1^4 f_{X^3} - 2g_1^2 f_X f_{X^2} - g_1 f_X^2 + f_X^3}{2g_1^5}.$$

Nous noterons $\sqrt{F}^{(S)}$ l'ensemble des racines carrées symétriques de F. Il est intéressant de noter que l'espèce X des singletons admet une unique racine carrée symétrique dont les premiers termes sont donnés par

$$-1 - X - E_2(X) + X^2 + XE_2(X) - X^3 - E_2(E_2(X))$$
$$+E_2(X)^2 - 2X^2E_2(X) + E_2(X^2) + \cdots .$$

4. Arbres et arborescences

Nous allons maintenant voir comment la racine carrée symétrique permet de relier les arborescences et les arbres. Pour ce faire nous utiliserons la formule de dissymétrie pour les arbres due à Leroux [14] et Leroux-Miloudi [3] énoncée plus bas. Notons $\mathbb{A}$ l'espèce des arbres et A celle des arborescences. Les premiers termes de la décomposition moléculaire de l'espèce des arbres sont les suivants

$$\mathbb{A} = X + E_2 + XE_2 + E_2(X^2) + XE_3 + XE_2(X^2) + X^3E_2 + XE_4 + \cdots .$$

L'espèce des arborescences quant à elle se décompose comme suit:

$$A = X + X^2 + X^3 + XE_2 + 2X^4 + X^2E_2 + XE_3 + \cdots .$$

Définition 4.1. *Soit $F \in \mathbb{C}[[\mathcal{A}]]$. On définit une nouvelle espèce que l'on note $F^\bullet$ en posant $F^\bullet = XF'$. On dit que $F^\bullet$ est obtenue de F en* pointant *les F-structures.*

Remarque 4.2. *On a évidemment $A = \mathbb{A}^\bullet$ puisque les arborescences sont précisément les arbres pointés.*

On trouvera la démonstration de la formule suivante dans [14],[3]. Elle a été étendue par G. Labelle dans [10] au cas des arborescences et arbres enrichis (voir aussi [11],[8]).

Proposition 4.3. *On a l'identité suivante (formule de dissymétrie pour les arbres):*

$$\mathbb{A} + A^2 = A + E_2(A). \tag{4}$$

∎

En fait, dans $\mathbf{C}[[\mathcal{A}]]$, on peut écrire cette formule sous une forme équivalente comme le dit la proposition suivante.

Proposition 4.4. *L'identité (4) est équivalente à*

$$E_2(1 - A) = 1 - \mathbb{A}. \tag{5}$$

En d'autres termes, l'espèce (virtuelle) $1 - A$ est une racine carrée symétrique de l'espèce $1 - \mathbb{A}$. Plus explicitement, on peut écrire

$$1 - A \in \sqrt{1 - \mathbb{A}}^{(S)}.$$

Démonstration: En utilisant la proposition 2.1, on a que

$$E_2(1 - A) = 1 - (A - A^2 + E_2(A)),$$

donc, en utilisant (4), on trouve

$$E_2(1 - A) = 1 - \mathbb{A}.$$

∎

Remarque 4.5. *Notons que l'espèce $1-\mathbb{A}$ possède deux racines carrées symétriques, car dans ce cas le coefficient constant est non nul. Toutefois un calcul montre que la seconde racine est en fait une espèce rationnelle:*

$$-2 + \frac{1}{2}X + \frac{3}{4}E_2(X) + \frac{1}{16}X^2 + \cdots.$$

On ne connaît pas encore d'interprétation combinatoire évidente pour cette racine.

Remarque 4.6. *La proposition 4.4 montre qu'on peut pointer algébriquement l'espèce des arbres sans utiliser la formule différentielle usuelle: $F^\bullet = XF'$. Elle permet aussi d'exprimer la décomposition moléculaire des arbres en fonction des arborescences.*

Remarque 4.7. *On a $A = \mathbb{A}^\bullet = X\mathbb{A}'$ et $A = XE(A)$. D'où $\mathbb{A}' = E(A)$. Il s'en suit que $\mathbb{A} \in \int E(A)$, ou encore que $\mathbb{A} = \int_J E(A) + W$ où W est tel que $W' = 0$ (voir [2]) où $\int_J$ désigne l'intégrale virtuelle de Joyal [5]. On a donc d'une part,*

$$\begin{aligned} \mathbb{A} &= \int_J E(A) + W \\ &= XE(A) - E_2(X)E(A)' + E_3(X)E(A)'' - E_4(X)E(A)''' + \cdots + W \\ &= A - E_2(X)E(A)' + E_3(X)E(A)'' - E_4(X)E(A)''' + \cdots + W. \end{aligned}$$

D'autre part en utilisant (4), on trouve que

$$W = E_2(A) - A^2 + E_2(X)E(A)' - E_3(X)E(A)'' + E_4(X)E(A)''' + \cdots.$$

On vérifiera que l'on a bien $W' = 0$. ∎

5. Conclusion

Il est possible d'écrire des formules du type (1) pour les espèces E_n et C_n ($n = 3, 4, 5, \cdots$) (cf [11]). Ceci permet de généraliser la notion de racine carrée symétrique. Dans le premier cas on parlera de *racine n^e symétrique*, dans le deuxième cas de *racine n^e cyclique*. Dénotons par $\sqrt[n]{F}^{(S)}$ l'ensemble des solutions de $E_n(G) = F$ et par $\sqrt[n]{F}^{(C)}$ l'ensemble des solutions de $C_n(G) = F$. Un prolongement naturel du présent travail serait de chercher si certains éléments de $\sqrt[n]{1-\mathbb{A}}^{(S)}$ ou de $\sqrt[n]{1-\mathbb{A}}^{(C)}$ ont une interprétation combinatoire pour $n > 2$. Un autre prolongement naturel serait d'étudier les racines n^e moléculaires d'une espèce F, c'est-à-dire les solutions G de l'équation $M(G) = F$ où M est une espèce moléculaire vivant sur la cardinalité n. L'utilisation de logiciels de calcul symbolique aidera certainement a analyser ces types de problèmes.

Bibliographie

1. Yves Chiricota, *Classification des espèces moléculaires de degré 6 et 7*, Rapport de recherche du département de mathématiques et d'informatique.
2. Yves Chiricota et Gilbert Labelle, *Familles de solutions combinatoires de l'équation différentielle $y' = 1 + y^2$, $y(0) = 0$, et d'équations différentielles autonomes*, À paraître dans Discrete Mathematics, North-Holland 19 pages.
3. Pierre Leroux et Brahim Miloudi, *Généralisations de la formule d'Otter*, À paraître dans les Annales des sciences mathématiques du Québec, 1991.
4. André Joyal, *Une théorie combinatoire des séries formelles*, Advances in Mathematics 42 (1981), 1–82.
5. ———, *Calcul intégral combinatoire et homologie des groupes symétriques*, Comptes rendus mathématiques de l'Academie des sciences du Canada **VII** (1985), no. 6, 337–342.
6. ———, *Règle des signes en algèbre combinatoire*, Comptes rendus mathématiques de l'Academie des sciences du Canada **VII** (1985), 285–290.
7. ———, *Foncteurs analytiques et espèces de structures*, In Labelle et Leroux [12], pp. 126–159.
8. Gilbert Labelle, *Sur la symétrie et l'asymétrie des structures combinatoires*, Actes de Colloque, LaBRI, Université de Bordeaux I, Maylis Delest, Gérard Jacob et Pierre Leroux eds., 3–19.
9. ———, *On combinatorial differential equations*, Journal of Mathematical Analysis and Applications **113** (1986), no. 2, 344–381.
10. ———, *Counting asymmetric enriched trees*, À paraître dans Journal of Symbolic Computation, 36 pages, 1991.
11. ———, *On asymmetric structures*, À paraître dans Discrete Mathematics, North-Holland 30 pages, 1991.
12. Gilbert Labelle et Pierre Leroux, eds., , Lecture Notes in Mathematics, no. 1234, Montréal, Québec, 1986, Université du Québec à Montréal, 1985, Springer-Verlag, Berlin.
13. Jacques Labelle, *Applications diverses de la théorie combinatoire des espèces de structures*, Annales des sciences mathématiques du Québec **7** (1983), no. 1, 59–94.
14. Pierre Leroux, *Methoden der Anzahlbestimmung für einige Klassen von Graphen*, Bayreuther Mathematische Schriften (1988), 1–36, Heft 26.
15. Yeong-Nan Yeh, *On the combinatorial species of Joyal*, Ph.D. thesis, State University of New-York at Buffalo, 1985.
16. ———, *The calculus of virtual species and $\mathbb{K}$-species*, In Labelle et Leroux [12], pp. 351–369.

Département de mathématiques et d'informatique, Université du Québec à Montréal (UQAM), Case postale 8888, Succursale A, Montréal (Québec), Canada H3C 3P8

Determinants of Super-Schur Functions Lattice Paths, and Dotted Plane Partitions

Francesco Brenti
Institut Mittag-Leffler
Auravägen 17
S-182 62 Djursholm
Sweden

Extended Abstract

Let $\mathbf{x} = (x_1, x_2, \ldots)$ and $\mathbf{y} = (y_1, y_2, \ldots)$ be two sequences of independent variables and λ be a partition. We denote by

$$s_\lambda(x_1, x_2, \ldots / y_1, y_2, \ldots)$$

the super-Schur function corresponding to λ in the variables $\mathbf{x}$ and $\mathbf{y}$. These functions arise naturally in the representation theory of Lie superalgebras [6] and were also defined, independently, by Metropolis, Nicoletti, and Rota in [8], under the name of bisymmetric functions. Since then, they have been studied extensively and we refer the reader to [1], [2], or [4] for their definition (they can be defined in several equivalent ways) and further information about them.

The purpose of the present work is to give combinatorial interpretations to the minors of the infinite matrix

$$\mathcal{S}(\mathbf{x}, \mathbf{y}) \stackrel{\text{def}}{=} (s_{(k)}(x_1, \ldots, x_n / y_1, \ldots, y_n))_{n,k \in \mathbf{N}} .$$

Our main results (Theorems 1.1 and 1.3) are proved combinatorially using lattice paths and are stated in terms of dotted and diagonal strict plane partitions, respectively. They also have many applications. As special cases we obtain combinatorial interpretations of determinants of homogeneous, elementary, and Hall-Littlewood symmetric functions, Schur's Q-functions, q-binomial coefficients, and q-Stirling numbers of both kinds. Other applications include the solution of a problem posed by Yahory in [10] and the combinatorial interpretation of a class of symmetric functions first defined, algebraically, by Macdonald in [7]. Many of our results are new even in the case $q = 1$. Others are q-analogues of known results. Our main theorem also has several interesting applications to the theory of total positivity. These are treated in [3].

In order to state the main results we need to define some notation, and terminology. Given an infinite matrix $M = (M_{n,k})_{n,k \in \mathbf{N}}$ (where $M_{n,k}$ is the entry in the n-th *row* and k-th *column* of M) and $\{n_1, \ldots, n_r\}_<$, $\{k_1, \ldots, k_r\}_< \subseteq \mathbf{N}$ we let

$$M \begin{pmatrix} n_1, \ldots, n_r \\ k_1, \ldots, k_r \end{pmatrix} \stackrel{\text{def}}{=} \det \left[(M_{n_i, k_j})_{1 \leq i, j \leq r} \right] .$$

Given an infinite sequence $\{a_i\}_{i\in\mathbf{N}}$ we let

$$\{a_i\}_{i\in\mathbf{N}}\begin{pmatrix} n_1,\ldots,n_r \\ k_1,\ldots,k_r \end{pmatrix} \stackrel{\text{def}}{=} A\begin{pmatrix} n_1,\ldots,n_r \\ k_1,\ldots,k_r \end{pmatrix}$$

where $A \stackrel{\text{def}}{=} (a_{n-k})_{n,k\in\mathbf{N}}$ (and $a_i \stackrel{\text{def}}{=} 0$ if $i<0$).

A *dotted partition* is a partition $\lambda=(\lambda_1,\lambda_2,\ldots,\lambda_n)$ where, for each $i\in\mathbf{P}$ such that $m_i(\lambda)>0$, the rightmost occurrence of i in λ may be dotted. Given two (possibly) dotted integers we will write $a \doteq b$ to indicate that they are equal as *dotted* integers, and $a=b$ if they are only equal as integers (so that, for example, $2=\dot{2}$, $2\doteq 2$, $\dot{2}\doteq\dot{2}$). We will also write $(a+b)^{\cdot}$ instead of the more cumbersome $\overbrace{a+b}^{\cdot}$. Given a partition $\lambda=(\lambda_1,\ldots,\lambda_r)$ a *shifted dotted plane partition* of shape λ is an array of (possibly dotted) positive integers $\pi=(\pi_{i,j})_{1\le i\le r,\, i\le j\le i+\lambda_i-1}$ where each row is a dotted partition and $\pi_{i,j}>\pi_{i+1,j}$ whenever $\pi_{i,j}$ and $\pi_{i+1,j}$ are both defined and $\pi_{i+1,j}$ is *not* dotted. Note that we *do not* require the parts of λ to be distinct. Let π be a shifted dotted plane partition as above. For $k=1,\ldots,\lambda_1$ we let

$$t_k(\pi) \stackrel{\text{def}}{=} \sum_{i=1}^{d_k(\pi)} \pi_{i,i+k-1}\,, \tag{1}$$

where $d_k(\pi) \stackrel{\text{def}}{=} |\{i\in\mathbf{P} : \pi_{i,i+k-1}>0\}|$, and

$$\dot{d}_k(\pi) \stackrel{\text{def}}{=} |\{i\in\mathbf{P} : \pi_{i,i+k-1} \text{ is dotted }\}|\,.$$

Also, given π as above we let $\tilde{\pi} \stackrel{\text{def}}{=} (\pi_{ij})_{1\le i\le r\,;\, i\le j\le i+\lambda_i-2}$. We also let

$$t(\pi) \stackrel{\text{def}}{=} (t_1(\pi),\ldots,t_{\lambda_1}(\pi),0,0,\ldots)\,,$$

$$d(\pi) \stackrel{\text{def}}{=} (d_1(\pi),\ldots,d_{\lambda_1}(\pi),0,0,\ldots)\,,$$

and

$$\dot{d}(\pi) \stackrel{\text{def}}{=} (\dot{d}_1(\pi),\ldots,\dot{d}_{\lambda_1}(\pi),0,0,\ldots)\,.$$

Given a set of variables $\mathbf{x}=(x_1,x_2,x_3,\ldots)$ and a vector $d=(d_1,d_2,d_3,\ldots)$ of integers we let

$$\mathbf{x}^d \stackrel{\text{def}}{=} \prod_{i\ge 1} x_i^{d_i}\,.$$

and $S(d) \stackrel{\text{def}}{=} (d_2,d_3,\ldots)$. Finally, we define the *weight* of π to be

$$w(\pi) \stackrel{\text{def}}{=} \mathbf{y}^{\dot{d}(\tilde{\pi})}\,\mathbf{x}^{t(\tilde{\pi})-\dot{d}(\tilde{\pi})-S(t(\pi))}\,.$$

Our main results are the following.

Theorem 1.1 *Let* $\{n_1, \ldots, n_r\}_>$, $\{k_1, \ldots, k_r\}_> \subseteq \mathbf{N}$. *Then*

$$\mathcal{S}(\mathbf{x},\mathbf{y}) \begin{pmatrix} n_r, \ldots, n_1 \\ k_r, \ldots, k_1 \end{pmatrix} = \sum_{\pi} \mathbf{x}^{t(\bar{\pi}) - \dot{d}(\bar{\pi}) - S(t(\pi))} \, \mathbf{y}^{\dot{d}(\bar{\pi})} \tag{2}$$

where the sum is over all shifted dotted plane partitions π *of shape* $(n_1+1, \ldots, n_r+1)$ *in which the i-th row has smallest part* $\doteq \dot{1}$ *and largest part* $= k_i + 1$ *for* $i = 1, \ldots, r$.

Theorem 1.2 *Let* $\{m_1, \ldots, m_r\}_<$, $\{k_1, \ldots, k_r\}_< \subseteq \mathbf{N}$ *and* $m, n \in \mathbf{P}$, $m > max\{m_r, k_r\}$. *Then*

$$\{s_{(k)}(x_1, \ldots, x_n / y_1, \ldots, y_n)\}_{k \in \mathbf{N}} \begin{pmatrix} m_1, \ldots, m_r \\ k_1, \ldots, k_r \end{pmatrix} = \sum_{\pi} \mathbf{x}^{t(\bar{\pi}) - \dot{d}(\bar{\pi}) - S(t(\pi))} \, \mathbf{y}^{\dot{d}(\bar{\pi})} \tag{3}$$

where the sum is over all shifted dotted plane partitions π *of shape* $((n+1)^r)$ *in which the i-th row has smallest part* $\doteq (m - k_i + 1)^{\cdot}$ *and largest part* $= m - m_i + 1$, *for* $i = 1, \ldots, r$.

Since the minor in the last theorem is just the skew super-Schur function corresponding to the skew shape $(m - m_1 + 1, \ldots, m - m_r + r)/(m - k_1 + 1, \ldots, m - k_r + r)$, this theorem gives a combinatorial interpretation for these skew super-Schur functions. Other combinatorial interpretations have been obtained by Berele and Regev [2], Balantekin and Bars [1], Dondi and Jarvis [4], and Stanley [9].

It is also possible to state the preceding results in terms of diagonal strict plane partitions (i.e., plane partitions in which parts decrease strictly along each diagonal, from upper left to lower right). Let T be a shifted (or skew) plane partition. For $i \in \mathbf{P}$ we let $c_i(T)$ (respectively $r_i(T)$) be the number of columns (respectively rows) of T that contain at least one part equal to i, and $m_i(T)$ be the number of parts of T that are equal to i. We then let

$$c(T) \stackrel{\text{def}}{=} (c_1(T), c_2(T), \ldots),$$

$$r(T) \stackrel{\text{def}}{=} (r_1(T), r_2(T), \ldots),$$

and

$$m(T) \stackrel{\text{def}}{=} (m_1(T), m_2(T), \ldots).$$

Theorem 1.3 *Let* $\{n_1, \ldots, n_r\}_>$, $\{k_1, \ldots, k_r\}_> \subseteq \mathbf{N}$. *Then*

$$\mathcal{S}(\mathbf{x},\mathbf{y}) \begin{pmatrix} n_r, \ldots, n_1 \\ k_r, \ldots, k_1 \end{pmatrix} = \sum_{T} \mathbf{y}^{m(T)-c(T)} \, \mathbf{x}^{m(T)-r(T)} \, (\mathbf{y}+\mathbf{x})^{r(T)+c(T)-m(T)}$$

where the sum is over all diagonal strict shifted plane partitions T *of shape* $(k_1, \ldots, k_r)$ *in which the i-th row has largest part* $\leq n_i$ *and* $\geq n_{i+1} + 1$, *for* $i = 1, \ldots, r$ *(where* $n_{r+1} \stackrel{\text{def}}{=} -1$*).*

Theorem 1.4 *Let* $\{m_1, \ldots, m_r\}_<$, $\{k_1, \ldots, k_r\}_< \subseteq \mathbf{N}$ *and* $m, n \in \mathbf{P}$, $m > max\{m_r, k_r\}$. *Then*

$$\{s_{(k)}(x_1, \ldots, x_n/y_1 \ldots, y_n)\}_{k\in\mathbf{N}} \begin{pmatrix} m_1, \ldots, m_r \\ k_1, \ldots, k_r \end{pmatrix} = \sum_T \mathbf{y}^{m(T)-c(T)} \mathbf{x}^{m(T)-r(T)} (\mathbf{y}+\mathbf{x})^{r(T)+c(T)-m(T)}$$

where the sum is over all diagonal strict plane partitions T *of shape* $(m - m_1 + 1, \ldots, m - m_r + r) \setminus (m - k_1 + 1, \ldots, m - k_r + r)$ *with largest part* $\leq n$.

In the case that $k_i = i$, for $i = 1, \ldots, r$, the preceding theorem first appeared, though without proof, in [9, Theorem 5.2].

In the second part of this work the preceding results are specialized to several interesting cases. In particular, using the fact that

$$e_k(y_1, \ldots, y_n) = s_{(k)}(0/y_1, \ldots y_n),$$

$$h_k(x_1, \ldots, x_n) = s_{(k)}(x_1, \ldots, x_n/0),$$

$$q_k(x_1, \ldots, x_n; \alpha) = s_{(k)}(x_1, \ldots, x_n/ - \alpha x_1, \ldots, -\alpha x_n) \tag{4}$$

and

$$q_k(x_1, \ldots, x_n; -1) = Q_{(k)}(x_1, \ldots, x_n),$$

we can interpret combinatorially several determinants of elementary and complete homogeneous symmetric functions, Hall-Littlewood symmetric functions, and Schur's Q-functions. In some cases we obtain the classical Jacobi-Trudi identity, in others analogs of it. We give two examples of such results here.

Let $T = (T_{i,j})_{1\leq i\leq r,\, i\leq j\leq i+n_i}$ be a shifted plane partition of shape $(n_1+1, \ldots, n_r+1)$. We call a part $T_{i,j}$, of T, *free* if $T_{i-1,j} > T_{i,j} > T_{i,j+1}$ (the inequalities being vacuously satisfied if either one of $T_{i-1,j}$ and $T_{i,j+1}$ are undefined). We let

$$\mathcal{F}(T) \stackrel{\text{def}}{=} \{(i,j) \in sh(T) : \ T_{i,j} \text{ is free}\},$$

and call $\mathcal{F}(T)$ the *free set* of T. Given T as above we define

$$\mathcal{U}(T) \stackrel{\text{def}}{=} \{(i,j) \in sh(T) : \ (i-1,j) \in sh(T), \ T_{i-1,j} = T_{i,j}\},$$

and call $\mathcal{U}(T)$ the *upper set* of T.

Theorem 1.5 *Let* $\{n_1, \ldots, n_r\}_>$, $\{k_1, \ldots, k_r\}_> \subseteq \mathbf{N}$. *Then*

$$(q_k(x_1, \ldots, x_n; \alpha))_{n,k\in\mathbf{N}} \begin{pmatrix} n_r, \ldots, n_1 \\ k_r, \ldots, k_1 \end{pmatrix} = \sum_T \mathbf{x}^{m(T)} (-\alpha)^{|\mathcal{U}(T)|} (1-\alpha)^{|\mathcal{F}(T)|}, \tag{5}$$

where the sum is over all diagonal strict shifted plane partitions T *of shape* $(k_1, \ldots, k_r)$ *in which the i-th row has largest part* $\leq n_i$ *and* $\geq n_{i+1} + 1$, *for* $i = 1, \ldots, r$ *(where* $n_{r+1} \stackrel{\text{def}}{=} -1$*).*

Theorem 1.6 *Let* $\{m_1, \ldots, m_r\}_<$, $\{k_1, \ldots, k_r\}_< \subseteq \mathbf{N}$ *and* $m, n \in \mathbf{P}$, $m > max\{m_r, k_r\}$. *Then*

$$\{q_k(x_1, \ldots, x_n; \alpha)\}_{k \in \mathbf{N}} \begin{pmatrix} m_1, \ldots, m_r \\ k_1, \ldots, k_r \end{pmatrix} = \sum_T \mathbf{x}^{m(T)} (-\alpha)^{|\mathcal{U}(T)|} (1-\alpha)^{|\mathcal{F}(T)|} \tag{6}$$

where the sum is over all diagonal strict skew plane partitions T *of shape* $(m - m_1 + 1, \ldots, m - m_r + r) \setminus (m - k_1 + 1, \ldots, m - k_r + r)$ *with largest part* $\leq n$.

Note that the symmetric function on the LHS of (6) is just the symmetric function $S_{\lambda/\mu}(\mathbf{x}; \alpha)$ defined (in the case that $\mu = \emptyset$) by Macdonald in [7, p.116, eq. (4.5)], where $\lambda \stackrel{\text{def}}{=} (m - m_1 + 1, \ldots, m - m_r + r)$ and $\mu \stackrel{\text{def}}{=} (m - k_1 + 1, \ldots, m - k_r + r)$. Therefore Theorem 1.6 gives a combinatorial interpretation of these symmetric functions.

Finally, by suitably specializing our main results we can give combinatorial interpretations of determinants of q-binomial coefficients, and of q-Stirling numbers of both kinds.

For example, using the fact that

$$\begin{bmatrix} n \\ k \end{bmatrix}_q = h_{n-k}(1, q, q^2, \ldots, q^k), \tag{7}$$

we can obtain the following result (where $B(q)$ denotes the infinite matrix of q-binomial coefficients).

Theorem 1.7 *Let* $\{n_1, \ldots n_r\}_>$, $\{k_1, \ldots, k_r\}_> \subseteq \mathbf{N}$. *Then*

$$B(q) \begin{pmatrix} n_r, \ldots, n_1 \\ k_r, \ldots, k_1 \end{pmatrix} = q^{-n((k_1+1, \ldots, k_r+1)')} \sum_T q^{|\overline{T}|}$$

where the sum is over all row strict shifted plane partitions T *of shape* $(k_1+1, \ldots, k_r+1)$ *in which the i-th row has largest part equal to* $n_i + 1$, *for* $i = 1, \ldots, r$ *(where for a partition* $\lambda = (\lambda_1, \lambda_2, \ldots)$, $n(\lambda) \stackrel{\text{def}}{=} \sum_{i \geq 1} (i-1) \lambda_i$*).*

In the case $q = 1$ the preceding theorem was first proved (though stated in a slightly different way) by Gessel and Viennot [5, Corollary 11].

Given a permutation $\sigma \in S_n$ having k cycles $C_1, \ldots, C_k$ we let $S(\sigma) \stackrel{\text{def}}{=} \{\min(C_1), \ldots, \min(C_k), n+1\}$, and $\{\sigma^{(1)}, \ldots, \sigma^{(k+1)}\}_> \stackrel{\text{def}}{=} S(\sigma)$. We say that σ is written in *normal form* if:

i) each cycle of σ is written with its smallest element first;

ii) the cycles are written in increasing order of their first elements.

The *normal representation* of σ is the word obtained from the normal form of σ by erasing all the parentheses. The number of *inversions* of σ, denoted by $\text{inv}(\sigma)$, is the number of inversions in the normal representation of σ. Given an r-tuple of permutations $(\sigma_1, \ldots, \sigma_r)$ and a partition $\mu = (\mu_1, \ldots, \mu_r)$ we associate to them a shifted skew array, denoted $ST_\mu(\sigma_1, \ldots, \sigma_r)$, by letting $\sigma_i^{(j)}$ be its $(i, i + \mu_i + j - 1)$ entry, for $i = 1, \ldots, r$, $j = 1, \ldots, k_i + 1$ (where k_i is the number of cycles of σ_i, for $i = 1, \ldots, r$). Using the fact that

$$C[n+1, k+1]_q = e_{n-k}([1]_q, [2]_q, \ldots, [n]_q).$$

it is possible to deduce the following result (where $C(q)$ denotes the infinite matrix of q-Stirling numbers of the first kind).

Theorem 1.8 *Let* $\{n_1, \ldots, n_r\}_>$, $\{k_1, \ldots, k_r\}_> \subseteq \mathbf{N}$. *Then*

$$C(q) \begin{pmatrix} n_r, \ldots, n_1 \\ k_r, \ldots, k_1 \end{pmatrix} = \sum_{(\sigma_1, \ldots, \sigma_r)} \prod_{i=1}^{r} q^{inv(\sigma_i)}$$

where the sum is over all r-tuples $(\sigma_1, \ldots, \sigma_r) \in S_{n_1+1} \times \ldots \times S_{n_r+1}$ *such that* $ST(\sigma_1, \ldots, \sigma_r)$ *is a shifted plane partition of shape* $(k_1 + 2, \ldots, k_r + 2)$.

Theorem 1.9 *Let* $\{m_1, \ldots, m_r\}_<$, $\{k_1, \ldots, k_r\}_< \subseteq \mathbf{N}$ *and* $m, n \in \mathbf{P}$, $m > max\{m_r, k_r\}$. *Then*

$$\{c[n+1, k+1]_q\}_{k \in \mathbf{N}} \begin{pmatrix} m_1, \ldots, m_r \\ k_1, \ldots, k_r \end{pmatrix} = \sum_{(\sigma_1, \ldots, \sigma_r)} \prod_{i=1}^{r} q^{inv(\sigma_i)}$$

where the sum is over all r-tuples $(\sigma_1, \ldots, \sigma_r) \in (S_{n+1})^r$ *such that* $ST(\sigma_1, \ldots, \sigma_r)$ *is a skew plane partition of shape* $(m - m_1 + 3, \ldots, m - m_r + r + 2) \setminus (m - k_1 + 1, \ldots, m - k_r + r)$.

Let $m, n \in \mathbf{P}$, $m < n$. Given a partition $\pi = \{B_1, \ldots, B_k\}$ of $[m, n]$ into k blocks we let $S(\pi) \stackrel{\text{def}}{=} \{\max(B_1), \ldots, \max(B_k), m-1\}$ and $\{\pi^{(1)}, \ldots, \pi^{(k+1)}\}_> \stackrel{\text{def}}{=} S(\pi)$. Let now π_i be the (unique) block of π containing $\pi^{(i)}$, for $i = 1, \ldots, k$. We define the *height* of π to be the number

$$ht(\pi) \stackrel{\text{def}}{=} \sum_{i=1}^{k} (i-1)(|\pi_i| - 1).$$

Given an r-tuple of partitions $(\pi_1, \ldots, \pi_r)$ we associate to it a shifted array $ST(\pi_1, \ldots \pi_r)$ by letting the elements of $S(\pi_i)$ (in decreasing order) be the i-th row of it, for $i = 1, \ldots, r$, and then shifting the resulting array. Using the fact that

$$S[n+1, k+1]_q = h_{n-k}([1]_q, [2]_q, \ldots, [k+1]_q).$$

it is possible to deduce the following results (where $S(q)$ denotes the infinite matrix of q-Stirling numbers of the second kind).

Theorem 1.10 *Let* $\{n_1, \ldots, n_r\}_>$, $\{k_1, \ldots, k_r\}_> \subseteq \mathbf{N}$. *Then*

$$S(q)\begin{pmatrix} n_r, \ldots, n_1 \\ k_r, \ldots, k_1 \end{pmatrix} = \sum_{(\pi_1,\ldots,\pi_r)} \prod_{i=1}^{r} q^{ht(\pi_i)},$$

where the sum is over all r-tuples $(\pi_1, \ldots, \pi_r) \in \Pi([n_1+1]) \times \ldots \times \Pi([n_r+1])$ *such that* $ST(\pi_1, \ldots, \pi_r)$ *is a shifted plane partition of shape* $(k_1+1, \ldots, k_r+1)$.

Theorem 1.11 *Let* $\{m_1, \ldots, m_r\}_<$, $\{k_1, \ldots, k_r\}_< \subseteq \mathbf{N}$ *and* $m, k \in \mathbf{P}$, $m > max\{m_r, k_r\}$. *Then*

$$\{S[n+1, k+1]_q\}_{n \in \mathbf{N}} \begin{pmatrix} m_1, \ldots, m_r \\ k_1, \ldots, k_r \end{pmatrix} = \sum_{(\pi_1,\ldots,\pi_r)} \prod_{i=1}^{r} q^{ht(\pi_i)}$$

where the sum is over all r-tuples $(\pi_1, \ldots, \pi_r) \in \Pi([m-k_1+1, m-m_1+1]) \times \ldots \times \Pi([m-k_r+1, m-m_r+1])$ *such that* $ST(\pi_1, \ldots, \pi_r)$ *is a shifted plane partition of shape* $((k+2)^r)$.

The last four results are new even in the case $q = 1$.

References

[1] A. B. Balantekin and I. Bars, *Representations of supergroups*, J. Math. Phys., **22** (1981), 1810-1818.

[2] A. Berele and A. Regev, *Hook Young diagrams with applications to combinatorics and to representations of Lie superalgebras*, Adv. in Math., **64** (1987), 118-175.

[3] F. Brenti, *Combinatorics and Total Positivity*, preprint.

[4] P. H. Dondi and P. D. Jarvis, *Diagram and superfield techniques in the classical superalgebras*, J. Phys., **A 14** (1981), 547-563.

[5] I. Gessel and G. Viennot, *Binomial determinants, paths, and hook length formulae*, Adv. in Math., **58** (1985), 300-321.

[6] V. G. Kač, *Lie superalgebras*, Adv. in Math., **26** (1977), 8-96.

[7] I. G. Macdonald, *Symmetric Functions and Hall Polynomials*, Oxford University Press, New York/London, 1979.

[8] N. Metropolis, G. Nicoletti, and G. C. Rota, *A new class of symmetric functions*, Mathematical Analysis and Applications, Part B, Adv. in Math. Suppl. Stud., **7B** (1981), 563-575.

[9] R. Stanley, *Unimodality and Lie superalgebras*, Studies in Applied Math., **72** (1985), 263-281.

[10] Yahory, Problem Session of the Conference in Combinatorics and Commutative Algebra, Nagoya, 1990.

S-series and Plethysm of Hook-shaped Schur Functions with Power Sum Symmetric Functions

J. B. REMMEL AND J. O. CARBONARA

Department of Mathematics, University of California, San Diego, La Jolla, CA 92093

M. YANG

Department of Mathematics, Western Illinois University, Macomb, IL 61455

ABSTRACT

We present a simple combinatorial rule to expand the plethysm $s_{(1^a,b)}[p_n]$ of a Schur function of hook shape $s_{(1^a,b)}$ and a power symmetric function p_n as a sum of Schur functions. As an application of our rule, we derive explicit formulas for the expansion of the the plethysms $s_2[s_{(1^a,b)}]$ and $s_{1^2}[s_{(1^a,b)}]$ as a sum of Schur functions.

One of the fundamental problems in the theory of symmetric functions is to find a combinatorial rule to find the expansion of the plethysm of two Schur functions $s_\lambda[s_\mu]$ as a sum of Schur functions. Let Λ^n denote the space of homogeneous symmetric polynomials of degree n. Then given symmetric polynomials with integer coefficients, $P \in \Lambda^n$ and $Q \in \Lambda^m$, we can formally define the plethysm $P[Q]$ as follows. First write $Q = \sum_\alpha a_\alpha x^\alpha$ where a_α is an integer and if $\alpha = (\alpha_1, \alpha_2, ...)$, then $x^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \dots$. Then define

$$e_r^Q(x) = \prod_\alpha (1 + t x^\alpha)^{a_\alpha} \mid_{t^r} \tag{1}$$

and for $\lambda = (\lambda_1, \dots, \lambda_k)$,

$$e_\lambda^Q(x) = e_{\lambda_1}^Q(x) \dots e_{\lambda_k}^Q(x). \tag{2}$$

Here given a series $f(t)$, $f(t) \mid_{t^r}$ denotes the coefficient of t^r in $f(t)$. Next since P is symmetric, we can express P in terms of the elementary symmetric functions $e_\lambda(x)$,

$$P(x) = \sum_{\lambda \vdash n} c_\lambda e_\lambda(x). \tag{3}$$

Then by definition,

$$P[Q] = \sum_{\lambda \vdash n} c_\lambda e_\lambda^Q(x) \tag{4}$$

It is easy to see that if $Q = \sum_\alpha a_\alpha x^\alpha$ and a_α is nonnegative for all α, then $P[Q]$ is nothing but the symmetric polynomial which results by substituting the monomials of Q for the variables of P. Thus for example, let $CS(\mu) = \{T_1(\mu), T_2(\mu), ...\}$ denote the set of column strict tableaux of shape μ and use the usual combinatorial definition of s_μ as

$$s_\mu(x) = \sum_{T \in CS(\lambda)} x^T \tag{5}$$

where $x^T = \prod_i x_i^{m_i(T)}$ and $m_i(T)$ equals the number of occurences of i in T. Then

$$s_\lambda[s_\mu](x) = s_\lambda(x^{T_1(\mu)}, x^{T_2(\mu)}, \ldots) \tag{6}$$

The notion of plethysm goes back to Littlewood (ref. [6]). The basic problem of plethysm is to find the coefficients $a_{\lambda,\mu,\nu}$ where

$$s_\lambda[s_\mu] = \sum a_{\lambda,\mu,\nu} s_\nu \tag{7}$$

It is known that $a_{\lambda,\mu,\nu}$ are nonnegative integers. That is, let S_n denote the symmetric group on n letters and given a partition λ of n, let U_λ denote the irreducible S_n-module corresponding to λ. Let $U_\mu^{\otimes n}$ denote the n-fold tensor product of U_μ where μ is a partition of m. Then the wreath product of S_n with S_m, $S_n(S_m)$, acts naturally on $U_\lambda \otimes U_\mu^{\otimes n}$ and $a_{\lambda,\mu,\nu}$ is the multiplicity of U_ν in the $S_{n \cdot m}$-module which results by inducing the action of $S_n(S_m)$ on $U_\lambda \otimes U_\mu^{\otimes n}$ to a representation of $S_{n \cdot m}$. See (ref. [5]) and (ref. [8]) for details.

The problem of computing the coefficients $a_{\lambda,\mu,\nu}$ has proved to be very difficult. Essentially, there are explicit formulas for $a_{\lambda,\mu,\nu}$ only when λ is a partition of 2, 3, or 4 and $\mu = (m)$. Most algorithms for the computation of $a_{\lambda,\mu,\nu}$ reduce the problem to the problem of multiplying Schur functions and finding explicit expansions of $s_\mu[p_k]$ where $p_k = \sum_i x_i^k$ is the power symmetric function. That is, the following properties of plethysm hold.

$$(P_1 \pm P_2)[Q] = P_1[Q] \pm P_2[Q] \tag{8}$$

$$(P_1 \cdot P_2)[Q] = P_1[Q] P_2[Q] \tag{9}$$

$$p_n[Q] = Q[p_n] \tag{10}$$

$$s_n[P \cdot Q] = \sum_\lambda s_\lambda[P] s_\lambda[Q] \tag{11}$$

$$s_\lambda[P+Q] = \sum_{\mu\subseteq\lambda} s_\mu[P] s_{\lambda/\mu}[Q] \tag{12}$$

$$s_\lambda[s_\mu]' = \begin{cases} s_\lambda[s_{\mu'}] & \text{if } |\mu| \text{ is even} \\ s_{\lambda'}[s_{\mu'}] & \text{if } |\mu| \text{ is odd} \end{cases} \tag{13}$$

where for any sum $\sum c_\nu s_\nu$, $(\sum c_\nu s_\nu)'$ denotes the sum $\sum c_\nu s_{\nu'}$ and ν' denotes the conjugate of ν. Given (8)-(10), it follows that since

$$s_\lambda = \sum_{\alpha_1+2\alpha_2+\cdots+n\alpha_n=n} \frac{\chi^\lambda_{(1^{\alpha_1}2^{\alpha_2}\cdots n^{\alpha_n})}}{\alpha_1!\cdots\alpha_n!} \left(\frac{p_1}{1}\right)^{\alpha_1} \cdots \left(\frac{p_n}{n}\right)^{\alpha_n} \tag{14}$$

where $\chi^\lambda_{(1^{\alpha_1}2^{\alpha_2}\cdots n^{\alpha_n})}$ denotes the value of the irreducible character of S_n corresponding to λ at the conjugacy class corresponding to the partition $(1^{\alpha_1}2^{\alpha_2}\cdots n^{\alpha_n})$, then

$$s_\lambda[s_\mu] = \sum_{\alpha_1+2\alpha_2+\cdots+n\alpha_n=n} \frac{\chi^\lambda_{(1^{\alpha_1}2^{\alpha_2}\cdots n^{\alpha_n})}}{\alpha_1!\cdots\alpha_n!} \left(\frac{s_\mu[p_1]}{1}\right)^{\alpha_1} \cdots \left(\frac{s_\mu[p_n]}{n}\right)^{\alpha_n} \tag{15}$$

Hence since we can compute $\chi^\lambda_{(1^{\alpha_1}2^{\alpha_2}\cdots n^{\alpha_n})}$, we can reduce the calculation of $s_\lambda[s_\mu]$ to the problem of multiplying Schur functions if we had a way to compute

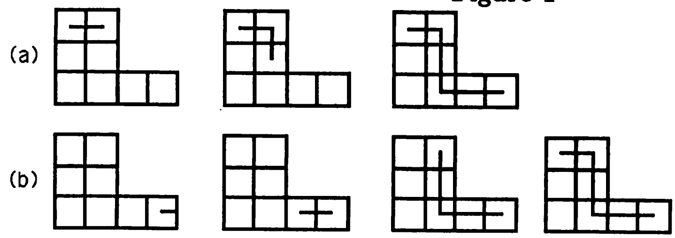
$$s_\mu[p_k] = \sum_\nu d^\nu_{\mu,k} s_\nu \tag{16}$$

or even the special case of (16),

$$s_{(n)}[p_k] = \sum_\nu d^\nu_{(n),k} s_\nu. \tag{17}$$

Now there are algorithms to compute the coefficients $d^\nu_{\mu,k}$ given by Chen, Garsia, and Remmel (ref. [2]) which we shall describe later. Moreover, there is a particularly simple algorithm due to Chen (ref.[1]) to compute the coefficients $d^\nu_{(n),k}$. (A similar algorithm for computing $d^\nu_{(n),k}$ had been given by Duncan (ref. [3])). To describe Chen's algorithm, we need to define the notion of special and transposed special rim hook tabloids. Given a Ferrers diagram λ, a *rim hook* h of λ is a consecutive sequence of cells along the North-East boundary of λ such that any two consecutive cells of h share and edge and the removal of the cells of h from λ results in a Ferrers diagram corresponding to another partition. We let $r(h)$ denote the number of rows of h and $c(h)$ denote the number of columns of h. We say that h is *special* if h has a cell in the first column of λ and h is *transposed special* (t-special) if h has cells in the first row of λ. For example, figure 1(a) pictures all special rim hooks of $\lambda = (2,2,4)$ and figure 1(b) pictures all t-special rim hooks of $\lambda = (2,2,4)$.

Figure 1



This given, a *rim hook tabloid* T of shape λ and type $\mu = (\mu_1, \ldots, \mu_k)$ is a filling of the Ferrers diagram of μ with rim hooks $(h_1, ..., h_k)$ such that $(|h_1|, ..., |h_k|)$ is a rearrangement of $(\mu_1, \ldots, \mu_k)$ where $|h_i|$ denotes the number of cells of h_i. To be more precise, one can think of a rim hook tabloid T as a sequence of shapes $\{\phi = \lambda^{(0)} \subset \lambda^{(1)} \subset \cdots \subset \lambda^{(k)} = \lambda\}$ such that for all $i \geq 1$, $\lambda^{(i)}/\lambda^{(i-1)}$ is a rim hook of $\lambda^{(i)}$ and $(|\lambda^{(1)}/\lambda^{(0)}|, |\lambda^{(2)}/\lambda^{(1)}|, \ldots, |\lambda^{(k)}/\lambda^{(k-1)}|)$ is a rearrangement of μ. T is called a *special (t-special) rim hook tabloid* if for all $i \geq 1$, $\lambda^{(i)}/\lambda^{(i-1)}$ is a special (t-special) rim hook of $\lambda^{(i)}$. The sign of T, $sgn(T)$, is defined by

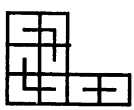
$$sgn(T) = \prod_i sgn(\lambda^{(i)}/\lambda^{(i-1)}) \tag{18}$$

where if h is a rim hook,

$$sgn(h) = (-1)^{r(h)-1}. \tag{19}$$

We emphasize however that the rim hook tabloid T of shape λ is the filling of the Ferrers diagram and is not the sequence of shapes $\{\phi = \lambda^{(0)} \subset \lambda^{(1)} \subset \cdots \subset \lambda^{(k)} = \lambda\}$. That is, figure 2 pictures a rim hook tabloid T of shape $\lambda = (2,2,4)$ and type (2,3,3) whose sign is $(-1)^{2-1}(-1)^{2-1}(-1)^{1-1} = 1$ and gives the two sequence of shapes that can be associated to it. Of course, if T is a special rim hook tabloid or a t-special rim hook tabloid, then there is a unique sequence of shapes $\{\phi = \lambda^{(0)} \subset \lambda^{(1)} \subset \cdots \subset \lambda^{(k)} = \lambda\}$ that can be associated to T.

Figure 2

T =  =

$\{\phi \subset (1,2) \subset (2,2,2) \subset (2,2,4)\} = \{\phi \subset (1,2) \subset (1,4) \subset (2,2,4)\}.$

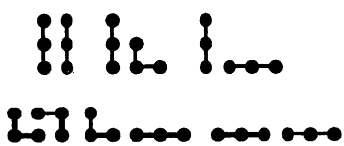
Let $SRHT(\lambda, \mu)$ (t-$SRHT(\lambda, \mu)$) denote the set of special (t-special) rim hook tabloids of shape λ and type μ. Note that if $\mu = (k^r)$ for some r, then there can be at most 1 special (t-special) rim hook of shape λ. This given, Chen's algorithm can be stated as

$$d^{\nu}_{(n),k} = \begin{cases} sgn(T) & \text{if there is a } t\text{-special rim hook tabloid T of shape } \nu \text{ and type } (k^n) \\ 0 & \text{otherwise} \end{cases} \tag{20}$$

Thus to compute $s_n[p_k]$, we need only generate all t-special rim hook tabloids T of type (k^n) and replace each such T by $sgn(T)s_{sh(T)}$ where $sh(T)$ denotes the shape of T.

For example, figure 3 pictures all t-special rim hook tabloids of type (3^2) where instead of drawing a Ferrers diagram, we have indicated the cells of the Ferrers diagram by dots.

Figure 3



Thus $s_2[p_3] = s_{(2^3)} - s_{(1,2,3)} + s_{(1^2,4)} + s_{(3,3)} - s_{(1,5)} + s_{(6)}$.

The main purpose of this paper is to give an extension of Chen's algorithm to compute the plethysm of a power symmetric function and a Schur function of hook shape. That is, we shall give an algorithm to compute $s_{(1^a,b)}[p_k]$ where $a + b = n$. To this end, we define a rim hook tabloid $T = \{\phi \subset \lambda^{(1)} \subset \cdots \subset \lambda^{(k)} = \lambda\}$ to be *bispecial* if for all $i \geq 1$, $\lambda^{(i)}/\lambda^{(i-1)}$ is either a special rim hook of $\lambda^{(i)}$ or $\lambda^{(i)}/\lambda^{(i-1)}$ is a t-special rim hook of $\lambda^{(i)}$. We then say that T is a *$(1^a, b)$-bispecial rim hook tabloid of type (k^n)* if among the rim hooks $\lambda^{(2)}/\lambda^{(1)}, \lambda^{(3)}/\lambda^{(2)} \ldots \lambda^{(k)}/\lambda^{(k-1)}$, there are exactly a special rim hook and *b-1* transposed special rim hooks. That is, in T, the number of special rim hooks is the length of the first column of $(1^a, b)$ and the number of t-special rim hooks is the length of the first row of $(1^a, b)$ where we count the first rim hook $\lambda^{(1)}/\lambda^{(0)}$ as both special and t-special rim hook. Then our main result is the following.

Theorem 1

$$d^{\nu}_{(1^a,b),k} = \sum_{T} sgn(T) \tag{21}$$

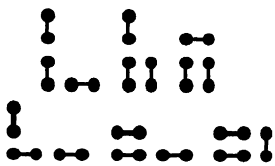
where T runs over all $(1^a, b)$-bispecial rim hook tabloids of shape ν and type (k^{a+b}).

We shall also prove that

Proposition 2 *If T_1 and T_2 are bispecial rim hook tabloids of shape ν and type k^{a+b}, then* $sgn(T_1) = sgn(T_2)$.

Thus there is no cancellation on the right hand side of (21) so that as with Chen's algorithm, to compute $s_{(1^a,b)}[p_k]$ we need only generate all $(1^a, b)$-bispecial rim hook tabloids T of type (k^n) and replace each such T by $sgn(T)s_{sh(T)}$. For example, figure 4 pictures all (1,2)-bispecial rim hook tabloids of type (2^3).

Figure 4



Thus $s_{(1,2)}[p_2] = s_{(1^3,3)} - s_{(1^2,2^2)} + s_{(2^3)} - s_{(1^2,4)} + s_{(2,4)} - s_{(3,3)}$.

The motivation for our result is related to the study of Schur function series, although the proof is through a different channel. Consider the Schur function series of the form:

$$\prod_i \frac{1 + a_1 x_i + a_2 x_i^2 + \cdots + a_p x_i^p}{1 + b_1 x_i + b_2 x_2^2 + \cdots + b_q x_i^q} = 1 + \sum_\alpha d_\alpha s_\alpha(x) \tag{22}$$

where p, q are positive integers, a_i, b_j are real numbers, and $\alpha = (\alpha_1 \le \alpha_2 \le \cdots \le \alpha_\ell)$ is a partition in increasing order. In [10], we have developed a combinatorial method for evaluating the coefficients d_α of Schur functions $s_\alpha(x)$ in the above expansion. Briefly speaking, the coefficients are calculated through the construction of certain bispecial rim hook tabloids of shape α with restricted hook lengths which depend on the generating function on the left hand side of (22). It is known that the series of hook-shaped Schur functions is generated by:

$$\prod_i \frac{1 + x_i}{1 - x_i} = 1 + 2 \sum_{a+b \ge 0} s_{(1^b, a+1)}. \tag{23}$$

Hence the plethysm of the sum of hook-shaped Schur functions with the power sum symmetric function $p_k(x)$ is the series generated by

$$\prod_i \frac{1 + x_i^k}{1 - x_i^k} = 1 + 2 \sum_\alpha k_\alpha s_\alpha(x), \tag{24}$$

where if $|\alpha|$ denotes the size of α, then

$$k_\alpha = \sum_{a+b+1=|\alpha|} s_{(1^b, a+1)}(x^k)|_{s_\alpha(x)}. \tag{25}$$

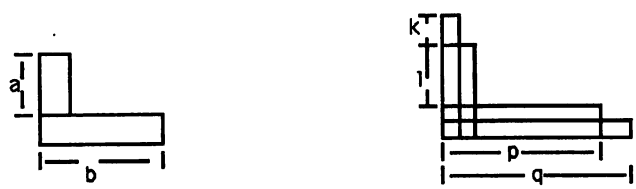
The coefficient k_α can easily be obtained using the general combinatorial construction in [10]. This lead us to conjecture Theorem 1 as a combinatorial algorithm for the Schur function expansion for the plethysm of a single hook-shaped Schur function with the power sum symmetric function.

We do not have the space in these proceedings to give a full proof of Theorem 1. Basically, the proof of Theorem 1 relies on two rules which can be used to compute the general case of plethysm of Schur functions $s_\lambda[s_\mu]$ as outlined in (ref. [2]), namely a version of the Littlewood-Richardson rule for multiplying Schur functions due to Remmel and Whitney (ref. [9]) and the SXP algorithm for computing $s_\lambda[p_k]$ due to Chen, Garsia, and Remmel (ref. [2]). We shall end this paper with a description of the SXP algorithm and a brief indication of how it can be used to prove Theorem 1. However before that we give a nice application of Theorem 1 by explicitly computing the plethysm $s_2[s_{(1^a,b)}]$ and $s_{1^2}[s_{(1^a,b)}]$.

We say λ is of *hook shape* if $\lambda = (1^a, b)$ for some a and b and λ is of *double hook shape* if $\lambda = (1^k, 2^l, p, q)$ where $2 \leq p \leq q$. These shapes are pictured in figure 5.

Figure 5

Hook shapes $(1^a, b)$ — *Double hook shapes* $(1^k, 2^l, p, q)$



Let $\langle , \rangle$ denote the Hall inner product on symmetric functions. Thus $\langle s_\lambda, s_\mu \rangle = \delta_{\lambda,\mu}$.

Theorem 3 *Let $a + b = n$, λ be a partition of $2n$, $u_\lambda = \langle s_2[s_{(1^a,b)}], s_\lambda \rangle$, and $v_\lambda = \langle s_{1^2}[s_{(1^a,b)}], s_\lambda \rangle$.*
Then

(a) (i) $u_\lambda = 0$ *if λ is not a hook shape or double hook shape*

(ii) *If $\lambda = (1^k, 2n-k)$ is a hook shape,*

$$u_\lambda = \begin{cases} 1 & \text{if } k=2a \text{ and } a \text{ is even} \\ 1 & \text{if } k = 2a+1 \text{ and } a \text{ is odd} \\ 0 & \text{otherwise} \end{cases}$$

(iii) *If $\lambda = (1^k, 2^l, p, q)$ where $2 \leq p \leq q$*

$$u_\lambda = \begin{cases} 1 & \text{if } q+p \in \{2b, 2b+2\} \text{ and } \frac{k}{2}+p \text{ is even} \\ 1 & \text{if } q+p=2b+1 \\ 0 & \text{otherwise} \end{cases}$$

(b) (i) $v_\lambda = 0$ *if* λ *is not of hook shape or double hook shape*

(ii) *If* $\lambda = (1^k, 2n-k)$ *is a hook shape,*

$$v_\lambda = \begin{cases} 1 & \text{if } k{=}2a \text{ and } a \text{ is odd} \\ 1 & \text{if } k = 2a{+}1 \text{ and } a \text{ is even} \\ 0 & \text{otherwise} \end{cases}$$

(iii) *If* $\lambda = (1^k, 2^l, p, q)$ *where* $2 \le p \le q$ *is a double hook shape, then*

$$u_\lambda = \begin{cases} 1 & \text{if } q{+}p \in \{2b, 2b{+}2\} \text{ and } \frac{k}{2}{+}p \text{ is odd} \\ 1 & \text{if } q{+}p{=}2b{+}1 \\ 0 & \text{otherwise} \end{cases}$$

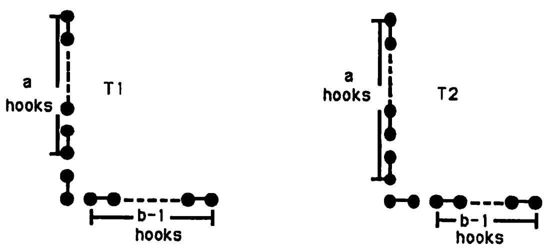
Proof: First we use the fact that $s_2 = \frac{1}{2}(p_1^2 + p_2)$ and $s_{1^2} = (p_1^2 - p_2)$. Moreover, clearly $p_1[s_\lambda] = s_\lambda$ for all λ so that

$$s_2[s_{(1^a,b)}] = \frac{1}{2}(s_{(1^a,b)}s_{(1^a,b)} + s_{(1^a,b)}[p_2]) \tag{26}$$

$$s_{1^2}[s_{(1^a,b)}] = \frac{1}{2}(s_{(1^a,b)}s_{(1^a,b)} - s_{(1^a,b)}[p_2]) \tag{27}$$

Thus we need to compute $s_{(1^a,b)}s_{(1^a,b)}$ and $s_{(1^a,b)}[p_2]$. By Theoreom 1, to compute $s_{(1^a,b)}[p_2]$, we must generate all $(1^a, b)$-bispecial rim hook tabloids T of type (2^{a+b}). Let $B(1^a, b)$ denote the set of all $(1^a, b)$-bispecial rim hook tabloids T of type (2^{a+b}). It is easy to see that if $T \in B(1^a, b)$, then T must be of hook shape or double hook shape. Now consider a $T \in B(1^a, b)$ of hook shape. Other than the initial rim hook of T, we must have a special rim hooks of size 2 all of which lie in the first column of T and $(b-1)$ special rim hooks of size 2 all of which lie in the first row of T. Since the initial rim hook of T is either horizontal or vertical, it is easy to see that there are two posibilities for such T which are pictured in figure 6.

Figure 6



$sh(T_1) = (1^{2a+1}, 2b-1)$ $sh(T_2) = (1^{2a}, 2b)$

$sgn(T_1) = (-1)^{a+1}$ $sgn(T_2) = (-1)^a$

Thus if we let $l_\lambda = \langle s_{(1^a,b)}[p_2], s_\lambda \rangle$, then if $\lambda = (1^k, 2n-k)$ is a hook shape,

$$l_\lambda = \begin{cases} (-1)^{a+1} & \text{if k=2a+1} \\ (-1)^a & \text{if k=2a} \\ 0 & \textit{Otherwise} \end{cases} \tag{28}$$

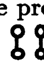
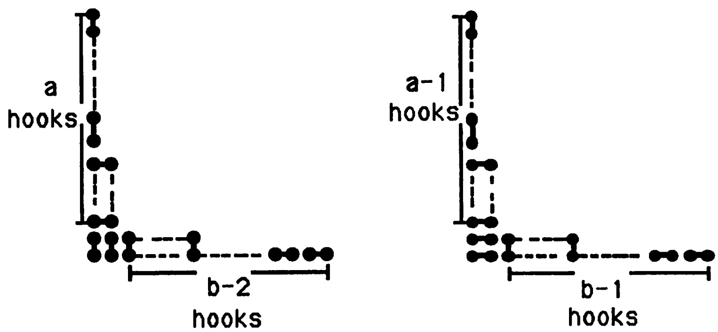
Next consider the $T \in B(1^a, b)$ of double hook shape. There are precisely two ways to fill the 2×2 corner square of T with hooks of size 2, namely  or . Thus any such T have one of the two forms pictured in figure 7.

Figure 7



It follows that the number of squares in the first two rows of T is either $2b$ or $2b+2$. Using this it is easy to show that if $\lambda = (1^k, 2^l, p, q)$ where $2 \le p \le q$ is a double hook shape, then

$$l_\lambda = \begin{cases} (-1)^{(k/2)+p} & \text{if p+q = 2b} \\ (-1)^{(k/2)+p-2} & \text{if p+q = 2b + 2} \\ 0 & \textit{otherwise} \end{cases} \tag{29}$$

Note the fact that λ is a partition of $2n$ and $p+q \in \{2b, 2b+2\}$ automatically forces that k is even.

Next let $t_\lambda = \langle s_{(1^a,b)} s_{(1^a,b)}, s_\lambda \rangle$. One can show by a careful analysis of the Remmel-Whitney rule to expand $s_{(1^a,b)} s_{(1^a,b)}$ as a sum of Schur functions that we have the following. First t_λ=0 unless λ is a hook shape or a double hook shape. If $\lambda = (1^k, 2n-k)$ is a hook shape, then

$$t_\lambda = \begin{cases} 1 & \text{if k=2a or k=2a+1} \\ 0 & \textit{otherwise} \end{cases} \tag{30}$$

Finally if $\lambda = (1^k, 2^k, p, q)$ where $2 \le p \le q$ is a double hook shape, then

$$t_\lambda = \begin{cases} 1 & \text{if p+q} \in \{2b,2b+2\} \\ 2 & \text{if p+q} = 2b+1 \\ 0 & otherwise \end{cases} \tag{31}$$

If we combine the results of (24)-(27) using equations (22) and (23), then one can easily derive the explicit formulas for u_λ and v_λ. □

Next we present the SXP-algorithm of Chen, Garsia and Remmel (ref. [2]) to compute $s_\mu[p_k]$ from which Theorem 1 can be derived. They show that if μ is a partition of n, then

$$s_\mu[p_k] = \sum_{|I_0|+\cdots+|I_{k-1}|=n} c^\mu_{I_0,\ldots,I_{k-1}} SS_{I_0,\ldots,I_{k-1}}(x) \tag{32}$$

where

(a) the sum is to be carried out over all k-tuples of partitions $I_0, \ldots, I_{k-1}$ whose diagrams are contained in μ and whose sum of parts add up to n,

(b) we have

$$c^\mu_{I_0,\cdots,I_{k-1}} = \langle s_{I_0} \cdots s_{I_{k-1}}, s_\mu \rangle \tag{33}$$

and

(c) the expression $SS_{I_0,\ldots,I_{k-1}(x)}$ denotes certain signed Schur function indexed by a partition with empty k-core whose construction is best explained through an example.

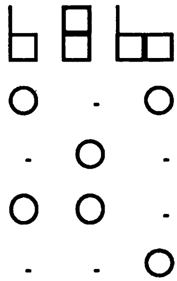
For instance, in the expansion of $s_{113}(x^3)$, since n=5 and k=3, one of the terms in (a) is that which corresponds to the triple of partitions $I_0 = (1), I_1 = (1^2), I_2 = (2)$. By using the Remmel-Whitney rule for multiplying Schur functions (ref. [9]), we obtain

$$c^{(113)}_{(1),(1^2),(2)} = 2.$$

To construct

$$SS_{(1),(1^2),(2)}(x),$$

we proceed as follows. First of all we represent $(1), (1^2)$, and (2) as partitions with a equal number of parts, that is, we write $(0,1)$ instead of (1), (1^2) since it already has two parts, and $(0,2)$ instead of (2). This given, we construct the *circle diagram* given below



(34)

The precise rule for putting together a column of this diagram from a partition $(i_1, i_2, \ldots, i_m)$ is that the distance (in dots) between the s^{th} and $(s+1)^{th}$ circle is given by the difference $i_{s+1} - i_s$, or equivalently the s^{th} circle is at distance $(i_s + s - 1)$ from the top. That is pictorially we have

$(i_1, i_2, ..., i_m)$

.

.

o $\leftarrow i_1$

.

.

o $\leftarrow i_2 + 1$

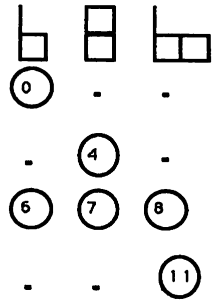
.

.

.

o $\leftarrow i_m + m - 1$

Accordingly, in the column labelled by (1,1) the first circle is at distance 1 from the top and the second circle is at distance one from the first. Proceeding in the same manner for the other two partitions we obtain the circle diagram given in (34). This done, assign to the positions in the diagram (indicated by dots when not by circles) the labels 0,1,2,3,... sucessively from left to right and from top to bottom, and record the label only when it falls in one of the circles. This gives the labelled diagram



(35)

In the case of a general k-tuple $I_0, I_1, \ldots, I_{k-1}$ we obtain a circle diagram with m circles in each column, where m is the maximum number of non-zero parts appearing in any of the partitions $I_0, I_1, \ldots, I_{k-1}$.

Let $b_1 < b_2 < b_3 < \cdots < b_{m \cdot k}$ be the labels placed on the circles and $q_{s,1} < q_{s,2} < \cdots < q_{s,m}$ be the labels appearing in the column corresponding to the partition I_s. Finally, let $inv(I_0, \ldots, I_{k-1})$ denote the number of inversions of the permutation

$$q_{0,1}q_{0,2}\cdots q_{0,m}q_{1,1}q_{1,2}\cdots q_{1,m}\cdots q_{k-1,1}q_{k-1,2}\cdots q_{k-1,m} \tag{36}$$

and let $sh(I_0, \ldots, I_{k-1}) = (b_1, b_2 - 1, b_3 - 2, \ldots, b_{m \cdot k} - mk + 1)$. This given, we set

$$SS_{I_0,\ldots,I_{k-1}} = (-1)^{\binom{m}{2}\binom{k}{2} + inv(I_0,\ldots,I_{k-1})} s_{sh(I_0,\ldots,I_k)} \tag{37}$$

We note that $I_0, \ldots, I_{k-1}$ is also called the k-quotient of $sh(I_0, \ldots, I_{k-1})$, see (ref. [5]). Going back to our particular example, we can easily see that to calculate the number of inversion of the permutation

$$0, 6, 4, 7, 2, 11,$$

we need only count, for each circle in the diagram of (35), the number of circles that are North-East of it, and add all these counts. This gives

$$inv((1), (1^2), (2)) = 0 + 0 + 1 + 2 + 1 + 0 = 4$$

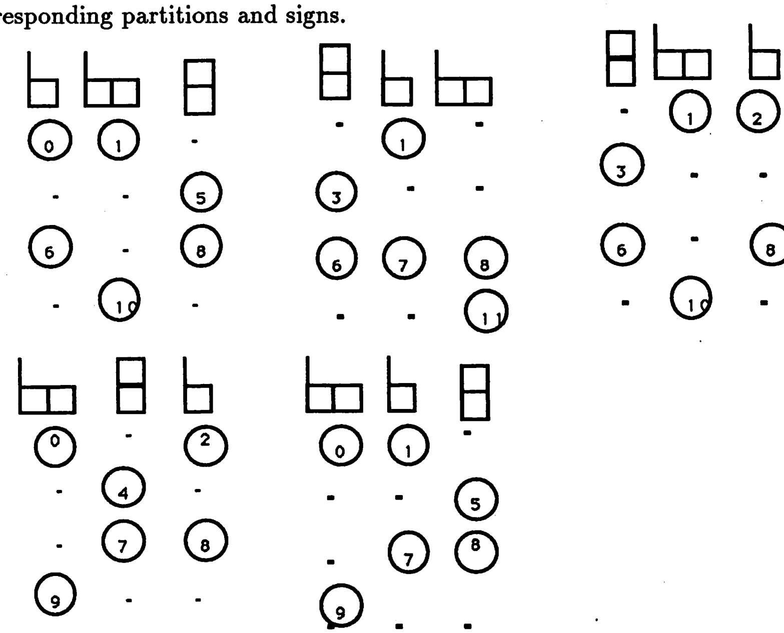
at the same time, we have $\binom{m}{2}\binom{k}{2} = \binom{2}{2}\binom{3}{2} = 3$ and

$$I(I_0, \cdots, I_{k-1}) = (0 - 0, 2 - 1, 4 - 2, 6 - 3, 7 - 4, 11 - 5) = (0, 1, 2, 3, 3, 6)$$

So we finally obtain in this case

$$SS_{(1)(1^2)(2)} = -s_{12336}(x).$$

It is worthwhile noting at this point that, in view of (33), the coefficient $c^{\mu}_{I_0,\ldots,I_{k-1}}$ does not depend on the order in which $I_0, \ldots, I_{k-1}$ are given. This means that in our particular example we can take advantage of the result that $c^{(113)}_{(1),(1^2),(2)} = 2$ and obtain 5 additional terms in the expansion of $s_{113}[p_3]$ by carrying out the above process for each of the remaining permutations of the triplet $((1),(1^2),(2))$. In the table below we give the resulting diagrams and the corresponding partitions and signs.



We thus obtain that the contribution to the expansion of $s_{113}[p_3]$ coming from the triplet $((1),(1^2),(2))$ is the expression

$$-2(s_{123^26}(x) + s_{3^245}(x) - s_{1^33^26}(x) + s_{1^345}(x) - s_{124^3}(x) + s_{34^3}(x)).$$

Taking all of this into account, we can easily see that the SXP algorithm decomposes into successive applications of the following 3 basic steps. Namely, to calculate the expansion of $s_\mu[p_k]$ when μ is a partition of n, we proceed as follows.

Step 1. We pick a k-tuple of partitions $I_0, I_1, \dots, I_{k-1}$ satisfying the 3 conditions

(a) The Ferrers' diagram of each I_s is contained in that of μ.

(b) $| I_0 | \leq | I_1 | \leq \cdots \leq | I_{k-1} |$

(c) $| I_0 | + | I_1 | + \cdots + | I_{k-1} | = n$.

This done we calculate the coefficent $c^{\mu}_{I_0,\dots,I_{k-1}}$ by the Schur function multiplication algorithm. If this coeficient is not zero we proceed to the next step. Otherwise we repeat step 1.

Step 2. Pick a permutation $I_{\sigma_1}, I_{\sigma_2}, \dots, I_{\sigma_k}$ of $I_0, I_1, \dots, I_{k-1}$ and construct the labelled circle diagram whose s^{th} column is indexed by I_{σ_s}.

Step 3. Calculate the partition $sh(I_{\sigma_1}, I_{\sigma_2}, \ldots, I_{\sigma_k})$and the corresponding sign.

Step 2 and 3 are to be repeated over all distinct permutations of $I_0, I_1, \ldots, I_{k-1}$. This done we go back to step 1 and repeat the process over all posible choices of $I_0, I_1, \ldots, I_{k-1}$ satisfying (a) (b) and (c).

We prove Theorem 1 by showing that our algorithm to compute $s_{(1^a,b)}[p_n]$ is equivalent to the SXP algorithm. To do this we need to prove two things. First, it is well known that $c^{\mu}_{I_0,\ldots,I_{k-1}} = 0$ unless $I_j \subseteq \mu$ for all j. Thus when μ is a hook, it must be the case that I_j is hook for $j = 0, \cdots, k-1$. In our case, we must show that if ν is a partition such that there is a $(1^a, b)$-bispecial rim hook tabloid of shape ν and type (k^{a+b}), then ν is a partition with empty k-core and the k-quotient of $\nu = (I_0, \ldots, I_{k-1})$ where I_j is hook for $j = 0, \cdots, k-1$. Second, we must show that if ν is a partition with empty k-core and the k-quotient of $\nu = (I_0, \ldots, I_{k-1})$ where I_j is hook for $j = 0, \cdots, k-1$, then then number of $(1^a, b)$-bispecial rim hook tabloids T of shape ν and type (k^{a+b}) is precisely $c^{(1^a,b)}_{I_0,\ldots,I_{k-1}}$ and the sign of any such T agrees with the sign given in (37). These two facts are proved by studying how the movement of circles in a circle diagram D effects the shape of the partition ν associated with D and the k-quotient of ν.

References

[1] Y.M. Chen *Combinatorial Algorithms for Plethysm*, Doctoral Thesis, UCSD(1982).

[2] Y.M. Chen, A.M. Garsia and J.B. Remmel, *Algorithms for Plethysm*, Contemporary Mathematics, **34**, 1984.

[3] D.G. Duncan, *D.E.Littlewood's Algebra of S-functions*, Can. J. Math. **4**, (1952), pp 504-512.

[4] O. Egecioglu and J.B. Remmel, *A Combinatorial interpretation of the Inverse Kostka Matrix*, Linear and Multilinear Algebra, **26**, (1990), pp 59-84.

[5] G.D. James and A. Kerber, *The Representation Theory of the Symmetric Group*, Encyc. of Math. and its Applications, Eddison-Wesley, (1981).

[6] D.E. Littlewood , *Invariant theory, tensors, and group characters*, Philos. Trans. Roy. Soc. London, ser. A239 (1944), pp 305-355.

[7] D.E. Littlewood, *The Theory of Group Characters*, 2nd edition, Oxford University Press, Oxford (1950).

[8] I.G. Macdonald, *Symmetric Functions and Hall Polynomials*, Clarendon Press, Oxford (1979).

[9] J.B. Remmel and R. Whitney, *Multiplying Schur functions*, J. of Algorithms, **5** , (1984), pp 471-487.

[10] J.B. Remmel and M. Yang, *New Multiplicity-free Schur Function Series*, SIAM J. Disc. Math. 4 253-274 (1991).

From algebraic sets to monomial linear bases by means of combinatorial algorithms.

L. Cerlienco* and M. Mureddu

1.1 Let $\mathbf{N}$ be the monoid of non-negative integers. Denote by $\mathbf{i} := (i_1, \ldots, i_n)$ an arbitrary element in the power $\mathbf{N}^n$. The usual order on $\mathbf{N}$, as well as the partial order it induces on $\mathbf{N}^n$, will be denoted by $\leq$.

Define an *n-dimensional Ferrers diagram* to be any ideal of the poset $\mathbf{N}^n$, i.e. any non-empty subset $\mathcal{F} \subseteq \mathbf{N}^n$ such that $\mathbf{j} < \mathbf{i} \in \mathcal{F} \Longrightarrow \mathbf{j} \in \mathcal{F}$. An element $\mathbf{i} = (i_1, \ldots, i_n) \notin \mathcal{F}$ is said to be a *co-minimal element* for the Ferrers diagram $\mathcal{F}$ if it is a minimal element of the complementary filter $\mathbf{N}^n \setminus \mathcal{F}$, i.e. if $(i_1, \ldots, i_{r-1}, i_r - 1, i_{r+1}, \ldots, i_n) \in \mathcal{F}$ for each r such that $i_r \geq 1$. Of course, $\mathbf{i} \notin \mathcal{F}$ is a co-minimal element iff $\mathcal{F}' := \{\mathbf{i}\} \cup \mathcal{F}$ is a Ferrers diagram.

We will write $\preceq$ for any *term-ordering* on $\mathbf{N}^n$, i.e. a linear ordering which is compatible with the monoid structure on $\mathbf{N}^n$:

$$\mathbf{0} \prec \mathbf{i} \qquad \text{for every } \mathbf{i} \neq \mathbf{0} \text{ in } \mathbf{N}^n$$

$$\mathbf{i} \preceq \mathbf{j} \Longrightarrow \mathbf{i} + \mathbf{r} \preceq \mathbf{j} + \mathbf{r} \qquad \text{for every } \mathbf{i}, \mathbf{j}, \mathbf{r} \in \mathbf{N}^n.$$

It is well known that any term-ordering on $\mathbf{N}^n$ is also a well-ordering.

1.2 Let K be a field and let $X := \{x_1, x_2, \ldots, x_n\}$ be a given set of indeterminates. Let us consider the usual polynomial algebra $K[X] := K[x_1, \ldots, x_n]$. Denote by $M_X \subseteq K[X]$ the free abelian monoid on X. The elements of M_X (i.e. the monic monomials) will be called *terms* of $K[X]$ and denoted by $\mathbf{x}^{\mathbf{i}} := x_1^{i_1} \cdots x_n^{i_n}$ with $\mathbf{i} := (i_1, \ldots, i_n) \in \mathbf{N}^n$. The orders $\leq$ and $\preceq$ on $\mathbf{N}^n$, as well as the notion of Ferrers diagram, extend to M_X in an obvious way.

An ideal J of the algebra $K[X]$ is said to be *cofinite* if

$$\operatorname{codim}(J) := \dim(K[X]/J) < \infty$$

Given a finite set $\mathcal{P} := \{P_1, \ldots, P_N\} \subseteq K^n$, the ideal

$$\Im(\mathcal{P}) := \{p \in K[X] \mid (\forall P \in \mathcal{P})(p(P) = 0)\}$$

*Submitted by the first author.

MB2. [Put the first i points of the list $\mathcal{P}$ in $\mathcal{Q}$.]
Set $\mathcal{Q} \leftarrow \{P_j \mid 1 \leq j \leq i\}$.

MB3. [Find which coordinate of $\mathbf{d}_{i+1}$ has to be changed, say $d_{i+1,s}$.]
Set $s \leftarrow \max\{k \geq 1 \mid \pi_{k-1}(P_j) = \pi_{k-1}(P_{i+1})$, for some $P_j \in \mathcal{Q}\}$.
($s-1$ is the length of the longest initial segment shared by P_{i+1} and some point $P_j \in \mathcal{Q}$. If $s > 1$, then in successive steps this decreases.)

MB4. [Find the points that determine the s-th coordinate of $\mathbf{d}_{i+1}$.]
Set $\mathcal{E} \leftarrow \{j \mid P_j \in \mathcal{Q},\ \ \pi_{s-1}(P_j) = \pi_{s-1}(P_{i+1}),\ \ \pi^{s+1}(\mathbf{d}_j) = \pi^{s+1}(\mathbf{d}_{i+1})\}$.
(Indices of the points of $\mathcal{Q}$ which have the first $s-1$ coordinates equal to those of P_{i+1} and whose corresponding elements in $\underline{\mathcal{F}}$ have the $n-s$ rightmost coordinates equal to those of $\mathbf{d}_{i+1}$. $\mathcal{E}$ is always not-empty.)

MB5. [Assign the value to the s-th coordinate of $\mathbf{d}_{i+1}$.]
Set $d_{i+1,s} \leftarrow (1 + \max\{d_{j,s} \mid j \in \mathcal{E}\})$.

MB6. [Did you determine the first coordinate of $\mathbf{d}_{i+1}$?]
If $s > 1$

MB.6.1. [Find the points that determine another coordinate of $\mathbf{d}_{i+1}$.]
Set $\mathcal{Q} \leftarrow \{P_j \mid 1 \leq j \leq i,\ \pi^s(\mathbf{d}_j) = \pi^s(\mathbf{d}_{i+1}) = (d_{i+1,s}, \ldots, d_{i+1,n})\}$.
(Points of $\underline{\mathcal{P}}$ whose corresponding elements in $\underline{\mathcal{F}}$ have the $n-s+1$ rightmost coordinates equal to those of $\mathbf{d}_{i+1}$.)

MB6.2. [Is $\mathcal{Q}$ empty?]
If $\mathcal{Q} \neq \emptyset$, return to step **MB3.**

MB7. [Increase i]
Set $i \leftarrow i + 1$. If $i < N$, return to step **MB2.**

MB8. [Done] Terminate the algorithm.

We put: $\mathcal{MB}(\mathcal{P}) := \{\mathbf{d}_1, \ldots, \mathbf{d}_N\}$; $\delta_{\underline{\mathcal{P}}}: \mathcal{P} \to \mathcal{MB}(\mathcal{P})$, $P_i \mapsto \mathbf{d}_i$. One could think that $\mathcal{MB}(\mathcal{P})$ is ill-defined, that is it depends on the order which has been used for arranging the points $P_1, \ldots, P_N$ when starting **Algorithm MB** (i.e. on the list $\underline{\mathcal{P}}$) rather than on the set $\mathcal{P} = \{P_1, \ldots, P_N\}$ itself. Well, this is not true. In fact, it is possible to prove the following propositions.

Prop. 1 *Let $\underline{\mathcal{F}'} := (\mathbf{d}'_1, \ldots, \mathbf{d}'_N)$ be the list associated with the list of points $\underline{\mathcal{P}'} = (P_{\sigma(1)}, \ldots, P_{\sigma(N)})$, where $\sigma \in S_N$, by* **Algorithm MB**. *Then, for some $\tau \in S_N$, $(\mathbf{d}'_1, \ldots, \mathbf{d}'_N) = (\mathbf{d}_{\tau(1)}, \ldots, \mathbf{d}_{\tau(N)})$.* □

Corollary 1 *If $\mathcal{P} \subseteq \mathcal{Q}$, then $\mathcal{MB}(\mathcal{P}) \subseteq \mathcal{MB}(\mathcal{Q})$.* □

of the algebraic set $\mathcal{P}$ is cofinite. More generally, when K is algebraically closed, from the *Nullstellensatz* it follows that the ideal $J \subseteq K[X]$ is a cofinite ideal iff the algebraic set of J, i.e.

$$\mathcal{V}(J) := \{P \in K^n \mid (\forall p \in J)(p(P) = 0)\},$$

is finite. In this case we have $\#\mathcal{V}(J) \leq \mathrm{codim}(J)$ (cfr. [3] p.23).

For a given ideal J, any linear basis $\mathcal{L}_J$ of the quotient algebra $K[X]/J$ whose elements are of the form $[\mathbf{x}^{\mathbf{i}}]_J := \mathbf{x}^{\mathbf{i}} + J$ will be called a *monomial basis*. If $\mathcal{L}_J = \{\mathbf{x}^{\mathbf{i}} + J \mid \mathbf{i} \in L \subseteq \mathbf{N}^n\}$ is a monomial basis, we shall say that $\{\mathbf{x}^{\mathbf{i}} \mid \mathbf{i} \in L\}$ is a *system of representatives* for the monomial basis $\mathcal{L}_J$. Obviously, if $\mathcal{L}_J = \{\mathbf{x}^{\mathbf{i}} + J \mid \mathbf{i} \in L\}$ is a monomial basis of $K[X]/J$, then any polynomial $p \in K[X]$ is congruent modulo J to exactly one polynomial of the form $\sum_{\mathbf{i} \in L} a_{\mathbf{i}} \mathbf{x}^{\mathbf{i}}$, $a_{\mathbf{i}} \in K$. Given an arbitrary term-ordering $\preceq$ on M_X, a monomial basis $\mathcal{L}_J = \{\mathbf{x}^{\mathbf{i}_1} + J, \ldots, \mathbf{x}^{\mathbf{i}_N} + J\}$ with $\mathbf{x}^{\mathbf{i}_1} \prec \ldots \prec \mathbf{x}^{\mathbf{i}_N}$ is said to be *minimal* with respect to $\preceq$ if for any other monomial basis $\mathcal{L}'_J = \{\mathbf{x}^{\mathbf{i}'_1} + J, \ldots, \mathbf{x}^{\mathbf{i}'_N} + J\}$ with $\mathbf{x}^{\mathbf{i}'_1} \prec \ldots \prec \mathbf{x}^{\mathbf{i}'_N}$ we have $\mathbf{x}^{\mathbf{i}_j} \preceq \mathbf{x}^{\mathbf{i}'_j}$ for $j = 1, \ldots, N$. Of course, both $\mathcal{L}_J$ and $\mathcal{L}'_J$, $\mathcal{L}_J \neq \mathcal{L}'_J$, could be minimal with respect to different term-orderings. It is not hard to prove that if $\mathcal{L}_J = \{\mathbf{x}^{\mathbf{i}} + J \mid \mathbf{i} \in L\}$ is a minimal monomial basis then $L \subseteq \mathbf{N}^n$ is an n-dimensional Ferrers diagram.

1.3 In the search for a minimal monomial basis $\mathcal{L}_{\mathcal{P}}$, we present a purely combinatorial algorithm to get it from $\mathcal{P}$. In fact, making use of the **Algorithm MB** below, we associate a Ferrers diagram $\mathcal{MB}(\mathcal{P}) = \{\mathbf{d}_1, \ldots, \mathbf{d}_N\} \subseteq \mathbf{N}^n$ to any finite set $\mathcal{P} := \{P_1, \ldots, P_N\} \subseteq K^n$. The Ferrers diagram $\mathcal{MB}(\mathcal{P})$ gives the monomial basis $\mathcal{L}_{\mathcal{P}} = \{\mathbf{x}^{\mathbf{d}} + \Im(\mathcal{P}) \mid \mathbf{d} \in \mathcal{MB}(\mathcal{P})\}$ which is minimal with respect to the inverse lexicographical ordering $\preceq_{i.l.}$ with $x_1 \prec_{i.l.} x_2 \prec_{i.l.} \cdots \prec_{i.l.} x_n$.

Put:

$\mathcal{P} := \{P_1, \ldots, P_N\} \subseteq K^n$
$\underline{\mathcal{P}} := (P_1, \ldots, P_N)$
$\mathbf{d}_j = (d_{j,1}, \ldots, d_{j,n}) \in \mathbf{N}^n$
$\pi_s \colon K^n \longrightarrow K^s, \quad (x_1, \ldots, x_n) \mapsto (x_1, \ldots, x_s)$
($\pi_0(P)$ is assumed to be the empty sequence.)
$\pi^s \colon K^n \longrightarrow K^{n-s+1}, \quad (x_1, \ldots, x_n) \mapsto (x_s, \ldots, x_n)$

ALGORITHM MB. Given a list $\underline{\mathcal{P}} := (P_1, \ldots, P_N)$ of points, we determine an ordered Ferrers diagram $\underline{\mathcal{F}} := (\mathbf{d}_1, \ldots, \mathbf{d}_N)$.

MB1. [Initialize.]
Set $\mathbf{d}_1 \leftarrow \mathbf{d}_2 \leftarrow \ldots \mathbf{d}_N \leftarrow (0, \ldots, 0)$; $i \leftarrow 1$.
(*In the beginning the coordinates of all the elements of* $\underline{\mathcal{F}}$ *are zero.*)

Prop. 2 *It is possible to arrange the points $P_1, \ldots, P_N$ in a suitable list $\underline{\mathcal{P}}' = (P_{\sigma(1)}, \ldots, P_{\sigma(N)})$ such that the elements in the corresponding list $\underline{\mathcal{F}}' = (\mathbf{d}_{\tau(1)}, \ldots, \mathbf{d}_{\tau(N)})$ are arranged according to the inverse lexicographical order $\preceq_{i.l.}$: $s < t \Rightarrow \mathbf{d}_{\tau(s)} \prec_{i.l.} \mathbf{d}_{\tau(t)}$* □

Lemma 1 *Let $\underline{\mathcal{F}} := (\mathbf{d}_1, \ldots, \mathbf{d}_N)$ be the Ferrers diagram associated to $\underline{\mathcal{P}} := (P_1, \ldots, P_N)$ by* **Algorithm MB**; *let $\mathbf{d}_N = (d_{N,1}, \ldots, d_{N,n})$. If $d_{N,i} \neq 0$, then there is some $k < N$ such that $\mathbf{d}_k = (d_{N,1}, \ldots, d_{N,i-1}, d_{N,i}-1, d_{N,i+1}, \ldots, d_{N,n})$.*

Proof. In order to calculate the first $i-1$ coordinates $d_{N,1}, \ldots, d_{N,i-1}$ of $\mathbf{d}_N = \delta_{\underline{\mathcal{P}}}(P_N)$, we have to consider the set $\{P_{j_1}, \ldots, P_{j_s}, P_{j_{s+1}} = P_N\} \subseteq \mathcal{P}$ $j_1 < \ldots < j_s < N$, of all points $P_{j_r} \in \mathcal{P}$ such that $\pi^i(\delta_{\underline{\mathcal{P}}}(P_{j_r}) = (d_{N,i}, \ldots, d_{N,n})$. Putting $\tilde{\mathcal{P}} := \{\pi_{i-1}(P_{j_1}), \ldots, \pi_{i-1}(P_{j_s}), \pi_{i-1}(P_N)\}$, we have $\pi_{i-1}(\delta_{\underline{\mathcal{P}}}(P_N)) = \delta_{\tilde{\underline{\mathcal{P}}}}(\pi_{i-1}(P_N)) = (d_{N,1}, \ldots, d_{N,i-1})$. For every $r \in \{1, \ldots, s+1\}$, there is a point $P_{j'_r} \in \mathcal{P}$, $j'_r < j_r$, such that $d_{j'_r,i} = d_{j_r,i} - 1 = d_{N,i} - 1, d_{j'_r,i+1} = d_{j_r,i+1} = d_{N,i+1}, \ldots, d_{j'_r,n} = d_{j_r,n} = d_{N,n}$ and $\pi_{i-1}(P_{j'_r}) = \pi_{i-1}(P_{j_r})$. Up to a suitable rearrangement of the points in $\underline{\mathcal{P}}$, we may assume that $j'_1 < \ldots < j'_{s+1}$. When calculating $\pi_{i-1}(\delta_{\underline{\mathcal{P}}}(P_{j'_{s+1}}))$ we have to consider the set $\mathcal{Q} := \{P_l \in \mathcal{P} \mid l \leq j'_{s+1}, \pi^i(\mathbf{d}_l) = (d_{j'_{s+1},i}, \ldots, d_{j'_{s+1},n}) = (d_{N,i} - 1, d_{N,i+1} \ldots, d_{N,n})\}$. Of course $P_{j'_r} \in \mathcal{Q}$. Let $\tilde{\mathcal{Q}} := \{\pi_{i-1}(P) \mid P \in \mathcal{Q}\} \subseteq K^{i-1}$; we have $\tilde{\mathcal{P}} \subseteq \tilde{\mathcal{Q}}$. Hence $\mathcal{MB}(\tilde{\mathcal{P}}) \subseteq \mathcal{MB}(\tilde{\mathcal{Q}}) \subseteq \mathbf{N}^{i-1}$; in particular $(d_{N,1}, \ldots, d_{N,i-1}) \in \mathcal{MB}(\tilde{\mathcal{Q}})$. It follows that there exists a point $P_k \in \mathcal{Q}$ such that $\delta_{\underline{\mathcal{P}}}(P_k) = (d_{N,1}, \ldots, d_{N,i-1}, d_{N,i} - 1, d_{N,i+1} \ldots, d_{N,n})$. □

As a straightforward consequence of **Lemma 1** we get:

Prop. 3 *The set $\mathcal{MB}(\mathcal{P})$ is an n-dimensional Ferrers diagram.* □

Prop. 4 *The set $\mathcal{L}_{\mathcal{P}} = \{\mathbf{x}^{\mathbf{d}} + \Im(\mathcal{P}) \mid \mathbf{d} \in \mathcal{MB}(\mathcal{P})\}$ is a monomial linear basis for $K[X]/\Im(\mathcal{P})$.*

Proof. By induction on the dimension n of K^n.

For $n = 1$, we have $\mathcal{P} = \{\rho_1, \ldots, \rho_n\}$, $\rho_i \in K$, and $\Im(\mathcal{P}) = (g)$, with $g = \Pi_{i=1}^N (x - \rho_i)$. **Algorithm MB** gives $\mathcal{MB}(\mathcal{P}) = \{0, 1, \ldots, N-1\}$; hence $\mathcal{L}_{\mathcal{P}} = \{\mathbf{x}^{\mathbf{d}} + \Im(\mathcal{P}) \mid \mathbf{d} \in \mathcal{MB}(\mathcal{P})\} = \{1 + (g), x + (g), \ldots, x^{N-1} + (g)\}$, which is a minimal monomial basis for $K[X]/(g)$.

Suppose now that the statement is true for every finite subset of $K^{n'}$, $n' < n$, and prove it for $\mathcal{P} = \{P_1, \ldots, P_N\} \subset K^n$. As $\dim K[X]/\Im(\mathcal{P}) = N = \#\mathcal{MB}(\mathcal{P})$, it remains to prove that the residue classes mod.$\Im(\mathcal{P})$ of the monomials $\mathbf{x}^{\mathbf{d}}$, $\mathbf{d} \in \mathcal{MB}(\mathcal{P})$, are linearly independent over $K[X]/\Im(\mathcal{P})$; in other words, we have to prove that any polynomial of the form

$$(1) \qquad p(x_1, \ldots, x_n) = \sum_{\mathbf{d} \in \mathcal{MB}(\mathcal{P})} \alpha_{\mathbf{d}} x^{\mathbf{d}} \in \Im(\mathcal{P}), \qquad \alpha_{\mathbf{d}} \in K$$

is the zero polynomial.

Putting $D := \mathcal{MB}(\mathcal{P})$, $D_r := \{\mathbf{d} = (d_1, \ldots, d_n) \in D \mid d_n = r\}$ and $\mathcal{P}_r := \{P \in \mathcal{P} \mid \delta_{\underline{\mathcal{P}}}(P) \in D_r\}$, it is easy to check that

$$(2) \qquad \mathcal{MB}(\pi_{n-1}(\mathcal{P}_r)) = \pi_{n-1}(D_r).$$

Let us write down polynomial (1) in the form

$$(3) \qquad p(x_1, \ldots, x_n) = \sum_{r=0}^{h} p_r(x_1, \ldots, x_{n-1}) x_n^r$$

where $h = \max\{d_n \mid \mathbf{d} = (d_1, \ldots, d_n) \in D\}$ and

$$(4) \qquad p_r(x_1, \ldots, x_{n-1}) = \sum_{(d_1, \ldots, d_{n-1}) \in \pi_{n-1}(D_r)} \alpha_{(d_1, \ldots, d_{n-1}, r)} x_1^{d_1} \cdots x_{n-1}^{d_{n-1}}.$$

The polynomial $p(x_1, \ldots, x_n) \in \Im(\mathcal{P})$ has to vanish at every point in $\mathcal{P}$. Consider a point $P = (a_1, \ldots, a_n) \in \mathcal{P}_h \subseteq \mathcal{P}$; there exist in $\mathcal{P}$ exactly $h + 1$ points that have the same first $n - 1$ coordinates as $P = (a_1, \ldots, a_n)$. It follows that the polynomial

$$p(a_1, \ldots, a_{n-1}, x_n) = \sum_{r=0}^{h} p_r(a_1, \ldots, a_{n-1}) x_n^r$$

vanishes identically. In particular $p_h(a_1, \ldots, a_{n-1}) = 0$ for every $(a_1, \ldots, a_{n-1}) \in$ $\in \mathcal{Q} := \pi_{n-1}(\mathcal{P}_h)$. Hence

$$(5) \qquad p_h(x_1, \ldots, x_{n-1}) \in \Im(\mathcal{Q}) \subseteq K[x_1, \ldots, x_{n-1}].$$

By (2) and (4) we have

$$(6) \qquad p_h(x_1, \ldots, x_{n-1}) = \sum_{(d_1, \ldots, d_{n-1}) \in \mathcal{MB}(\mathcal{Q})} \alpha_{(d_1, \ldots, d_{n-1}, h)} x_1^{d_1} \cdots x_{n-1}^{d_{n-1}}$$

Because of the inductive hypothesis, the set $\{x_1^{d_1} \cdots x_{n-1}^{d_{n-1}} + \Im(\mathcal{Q}) \mid (d_1, \ldots$ $\ldots, d_{n-1}) \in \mathcal{MB}(\mathcal{Q})\}$ is a monomial basis for $K[x_1, \ldots, x_{n-1}]/\Im(\mathcal{Q})$. ¿From this and from (5) we deduce that polynomial (6) vanishes identically. Hence

$$(7) \qquad p(x_1, \ldots, x_n) = \sum_{r=0}^{h-1} p_r(x_1, \ldots, x_{n-1}) x_n^r$$

Arguing for $r = h - 1, h - 2, \ldots, 1, 0$ as for $r = h$, we conclude that $p_r(x_1, \ldots, x_{n-1})$ is the zero polynomial for every $r \in \{0, \ldots, h\}$. □

Prop. 5 *The monomial linear basis $\mathcal{L}_{\mathcal{P}} = \{\mathbf{x}^{\mathbf{d}} + \Im(\mathcal{P}) \mid \mathbf{d} \in \mathcal{MB}(\mathcal{P})\}$ for $K[X]/\Im(\mathcal{P})$ is minimal with respect to the inverse lexicographical order $\preceq_{i.l.}$ on M_X.*

Proof. Let $\underline{\mathcal{P}} = (P_1, \ldots, P_N)$, $\underline{D} := \mathcal{MB}(\mathcal{P}) = (\mathbf{d}_1, \ldots, \mathbf{d}_N)$; $\mathbf{d}_i = (d_{i,1}, \ldots$ $\ldots, d_{i,n})$, $P_i = (a_{i,1}, \ldots, a_{i,n})$ for $i = 1, \ldots, N$. Let $h := \max\{d_{i,n} \mid i = 1, \ldots, N\}$, and $\mathcal{P}_h = \{P_{j_1}, \ldots, P_{j_m}\} \subseteq \mathcal{P}$; for any P_{j_r} there are in $\mathcal{P}$ exactly $h+1$ points which have the same $n-1$ coordinates as P_{j_r}; let us denote them by $Q_{j_r,0}, \ldots, Q_{j_r,h} = P_{j_r}$ Up to a suitable rearrangement we may assume that they are the last $(h+1)m$ elements in the list $\underline{\mathcal{P}}$, i.e.

$$(Q_{j_1,0}, \ldots, Q_{j_1,h}, \ldots, Q_{j_m,0}, \ldots, Q_{j_m,h}) = (P_{N-(h+1)m+1}, P_{N-(h+1)m+2}, \ldots, P_N),$$

so that

$$d_{N-(h+1)(m-1),n} = d_{N-(h+1)(m-2),n} = \cdots = d_{N-(h+1),n} = d_{N,n} = h$$

We have to prove that for any $\mathbf{d}' = (d'_1, \ldots, d'_n)$ such that $\mathbf{d}' \prec_{i.l} \mathbf{d}_N$ there exists in $\Im(\mathcal{P})$ a polynomial of the form

$$\sum_{i=1}^{N-1} \alpha_i \mathbf{x}^{\mathbf{d}_i} + \alpha \mathbf{x}^{\mathbf{d}'} \in \Im(\mathcal{P}). \tag{8}$$

Because of **MB.6.1** of **Algorithm MB**, without loss of generality we may assume that $d'_n < d_{N,n}$. Observe that (8) is equivalent to

$$\sum_{i=1}^{N-1} \alpha_i a_{s,1}^{d_{i,1}} \cdots a_{s,n}^{d_{i,n}} + \alpha a_{s,1}^{d'_1} \cdots a_{s,n}^{d'_n} = 0, \qquad P_s = (a_{s,1}, \ldots, a_{s,n}) \in \underline{\mathcal{P}} \tag{9}$$

Hence, it is enough to prove that the N by N matrix A whose $s-th$ row is

$$a_{s,1}^{d_{1,1}} \cdots a_{s,n}^{d_{1,n}} \quad \ldots\ldots \quad a_{s,1}^{d_{N-1,1}} \cdots a_{s,n}^{d_{N-1,n}} \quad a_{s,1}^{d'_1} \cdots a_{s,n}^{d'_n}$$

(that is, the evaluation at P_s of the list of monomials $\mathbf{x}^{\mathbf{d}_1}, \ldots, \mathbf{x}^{\mathbf{d}_{N-1}}, \mathbf{x}^{\mathbf{d}'}$) is a singular matrix. We shall prove this by showing that the submatrix A' consisting of the last $(h+1)m$ rows of A has rank less than $(h+1)m$.

Let X be any minor of order $(h+1)m$ of A'; notice that X can be given the form

$$X = \sum M_1 \cdot M_2 \cdot \cdots \cdot M_m \tag{10}$$

where M_i is a minor of A' which consists of the $h+1$ rows whose indices are $N-(h+1)(m-i+1)+1, N-(h+1)(m-i+1)+2, \ldots, N-(h+1)(m-i)$. Since all the points $P_{N-(h+1)(m-i+1)+1}, P_{N-(h+1)(m-i+1)+2}, \ldots, P_{N-(h+1)(m-i)}$ have the same $(n-1)$ first coordinates, the minor M_i is different from zero only if its $(h+1)$ columns correspond to monomials $\mathbf{x}^{\mathbf{i}} = x_1^{i_1} \cdots x_{n-1}^{i_{n-1}} x_n^{i_n}$ such that their exponents i_n's have all the possible values $0, 1, \ldots, h$. On the other hand, there are no more than $m-1$ such $(h+1)$-tuples of different columns (since only $m-1$ among the monomials $\mathbf{x}^{\mathbf{d}_1}, \ldots, \mathbf{x}^{\mathbf{d}_{N-1}}, \mathbf{x}^{\mathbf{d}'}$ have h as last exponent);

it follows that at least one of the m minors M_i's in (10) is zero. Hence $X = 0$.
□

1.4 In the case where $n = 2$ **Algorithm MB** can be given the following simplified form.

Assume that

$$\mathcal{P} = \{(a_1, b_{11}), \dots, (a_1, b_{1h_1}), \dots, (a_m, b_{m1}), \dots, (a_m, b_{mh_m})\}$$

with $h_1 + \dots + h_m = N$ and $i \neq j \Rightarrow a_i \neq a_j$. Then,

$$\mathcal{MB}(\mathcal{P}) = \{(p, q) \mid 0 \leq p < m,\ 0 \leq q < h_p\}.$$

1.5 All that we have stated up to now can be generalized to what we might call *algebraic multisets* in the following sense.

Consider the linear map

$$\begin{array}{rcl} \mathbf{D_i}: K[X] & \longrightarrow & K[X] \\ \mathbf{x^h} & \longmapsto & \binom{\mathbf{h}}{\mathbf{i}} \mathbf{x^{h-i}} \end{array}$$

where $\mathbf{h} = (h_1, \dots, h_n)$, $\mathbf{i} = (i_1, \dots, i_n) \in \mathbf{N}$ and $\binom{\mathbf{h}}{\mathbf{i}} := \binom{h_1}{i_1} \cdot \dots \cdot \binom{h_n}{i_n}$. Observe that when the field K has characteristic zero, then

$$\mathbf{D_i} = \frac{1}{\mathbf{i}!}\mathbf{D^i} := \frac{1}{\mathbf{i}!} \frac{\partial^{i_1 + \dots + i_n}}{\partial x_1^{i_1} \cdots \partial x_n^{i_n}}$$

where $\mathbf{i} = (i_1, \dots, i_n)$ and $\mathbf{i}! = i_1! \cdots i_n!$.
Let v_P be the evaluation map at the point P:

$$\begin{array}{rcl} v_P: K[X] & \longrightarrow & K \\ q & \longmapsto & q(P) \end{array}$$

Define the linear map $v_P^{\mathbf{i}}$ as the composition $v_P \circ \mathbf{D_i}$, i.e.

$$\begin{array}{rcl} v_P^{\mathbf{i}}: K[X] & \longrightarrow & K \\ q & \longmapsto & (\mathbf{D_i} q)(P). \end{array}$$

For every ideal J of $K[x_1, \dots, x_n]$ and every $P \in \mathcal{V}(J)$, put

$$\mathcal{F}_J(P) := \left\{\mathbf{i} \in \mathbf{N}^n \mid (\forall p \in J) \left(v_P^{\mathbf{i}}(p) = 0\right)\right\}$$

Prop. 6 *(i) $\mathcal{F}_J(P)$ is a Ferrers diagram; (ii) if $(g_1, \ldots, g_s)$ is a system of generators of J, then $\mathcal{F}_J(P)$ is the largest Ferrers diagram contained in the set*

$$\left\{\mathbf{i} \in \mathbf{N}^n \,\middle|\, v_P^{\mathbf{i}}(g_j) = 0 \text{ for every } 1 \leq j \leq s\right\}.$$

We define a *finite n-dimensional algebraic multiset*, or simply an *algebraic multiset*, to be a set $\wp := \{(P_1, \mathcal{F}_1), \ldots, (P_N, \mathcal{F}_N)\}$; each element $(P_j, \mathcal{F}_j) \in \wp$ consists of a point P_j of K^n together with a Ferrers diagram $\mathcal{F}_j \subset \mathbf{N}^n$, which will be called the *algebraic diagram* of the point P_j. We shall freely make use of the notation $(P, \mathbf{i}) \in \wp$, or also $P \in \wp$, to mean that for some $j \in \{1, \ldots, N\}$, $P = P_j$ and $\mathbf{i} \in \mathcal{F}_j$. With every algebraic multiset $\wp = \{(P_1, \mathcal{F}_1), \ldots, (P_N, \mathcal{F}_N)\}$ we associate the set

$$\Im(\wp) := \left\{p \in K[X] \,\middle|\, (\forall j)(\forall \mathbf{i} \in \mathcal{F}_j)\left(v_{P_j}^{\mathbf{i}}(p) = 0\right)\right\}.$$

It is not difficult to prove that $\Im(\wp)$ is a cofinite ideal of $K[X]$ and that $\mathcal{F}_{\Im(\wp)}(P_j) = \mathcal{F}_j$ for every $j \in \{1, \ldots, N\}$. Moreover, one can prove that

$$\operatorname{codim} \Im(\wp) = \#\wp := \sum_{j=1}^{N} \#\mathcal{F}_j.$$

The question now is: how do we get a monomial linear basis for $K[X]/\Im(\wp)$? Once more the problem can be solved by applying a slightly modified version of **Algorithm MB** (in fact **Algorithm MB** itself with a few obvious changes) to a suitable set $\mathcal{R}(\wp) \subset (K \times \mathbf{N})^n$ associated with the algebraic multiset $\wp$. $\mathcal{R}(\wp)$ will be called the *umbral representation* of $\wp$. To be precise, consider the bijection

$$\begin{array}{rcl} u\colon K^n \times \mathbf{N}^n & \longrightarrow & (K \times \mathbf{N})^n \\ ((a_1, \ldots, a_n), (i_1, \ldots, i_n)) & \longmapsto & ((a_1, i_1), \ldots, (a_n, i_n)). \end{array}$$

Put

$$\mathcal{R} = \mathcal{R}(\wp) := \{u(P, \mathbf{i}) \mid (P, \mathbf{i}) \in \wp\}.$$

and

$$\mathcal{MB}(\wp) := \mathcal{MB}(\mathcal{R}).$$

(the symbol $\mathcal{MB}$ on the right-hand side represents the operator defined by **Algorithm MB**). It is possible to prove that $\mathcal{MB}(\wp)$ satisfies properties analogous to those of $\mathcal{MB}(\mathcal{P})$; in particular, (i) $\mathcal{MB}(\wp)$ is a Ferrers diagram and (ii) the set $B := \{\mathbf{x}^{\mathbf{i}} + \Im(\wp) \mid \mathbf{i} \in \mathcal{MB}(\wp)\}$ is a monomial linear basis of $K[X]/\Im(\wp)$.

1.6 The above algorithms may come in handy for solving various problems. Let us examine a few of them.

I. First of all, let us see how to determine a system of generators $(\gamma_1, \ldots, \gamma_r)$ for the ideal $\Im(\wp)$ of a finite algebraic multiset $\wp := \{(P_1, \mathcal{F}_1), \ldots, (P_N, \mathcal{F}_N)\}$. In fact, the set $\{\gamma_1, \ldots, \gamma_r\}$ we shall obtain is also a *reduced Gröbner basis* of $\Im(\wp)$. It goes without saying that the same procedure works also when $\wp$ is a finite algebraic set.

Let B be the monomial linear basis obtained in **1.5** and let $Y = \{\mathbf{x}^{\mathbf{r}_1}, \ldots, \mathbf{x}^{\mathbf{r}_r}\} \subset M_X$ be the minimal set of terms such that $B = M_X \setminus \sum_{i=1}^r \mathbf{x}^{\mathbf{r}_i} \cdot M_X$. For each $\mathbf{x}^{\mathbf{r}_i} \in Y$ determine a polynomial $\gamma_i \in K[X]$ in the form of a determinant in the following way. The first row of γ_i is the list $(\mathbf{x}^{\mathbf{d}_1}, \ldots, \mathbf{x}^{\mathbf{d}_m}, \mathbf{x}^{\mathbf{r}_i})$ where $\mathbf{x}^{\mathbf{d}_j} \in B$ and $m = \#B = \mathrm{codim}(\Im(\wp))$; the succesive rows are lists of the form $(v_{P_j}^{\mathbf{i}}(\mathbf{x}^{\mathbf{d}_1}), \ldots, v_{P_j}^{\mathbf{i}}(\mathbf{x}^{\mathbf{d}_m}), v_{P_j}^{\mathbf{i}}(\mathbf{x}^{\mathbf{r}_i}))$, one for each $(P_j, \mathbf{i}) \in \wp$. It can be proved that the list $(\gamma_1, \ldots, \gamma_r)$ is a reduced Gröbner basis of $\Im(\wp)$.

II. Consider a linear form $f \in K[X]^* \cong K[[X]]$. If $\mathrm{Ker}(f)$ contains a cofinite ideal J of $K[X]$, then f is said to be an *n-linearly recursive function* and J is called a *characteristic ideal* of f. This notion has been introduced in [4] as a generalization of that of *linearly recursive sequence*, to which it reduces when $n = 1$. n-linearly recursive functions may also be regarded as elements of the dual bialgebra of the usual polynomial bialgebra on $K[X]$. When working on these subjects, it may happen that examples (perhaps, *suitable* examples) of n-linearly recursive functions are needed. How to construct them? It is convenient to divide the answer to this question into two parts.

(A) Let us first suppose that we know a system of generators $(g_1, \ldots, g_s)$ of the characteristic ideal J of the n-linearly recursive functions we are considering. In this case we may calculate a reduced Gröbner basis $G_{\preceq} := RGB(g_1, \ldots, g_s)$ of $(g_1, \ldots, g_s)$ with respect to some term-ordering $\preceq$ on M_X. Let $G_{\preceq} = (\gamma_1, \ldots, \gamma_r)$, $\gamma_j \in K[X]$, and let $\xi_j \in M_X$ be the leading term (with respect to $\preceq$) of the polynomial γ_j. Then the set $B = M_X \setminus \sum_{j=1}^r \xi_j \cdot M_X$ is a monomial linear basis for $K[X]/J$, which is minimal with respect to $\preceq$. It follows that any n-linearly recursive function whose characteristic polynomial is J is uniquely determined by the set of its *initial values* $\{f(\mathbf{x}^{\mathbf{d}}) \mid \mathbf{x}^{\mathbf{d}} \in B\}$, all the other values $f(\mathbf{x}^{\mathbf{i}})$, $\mathbf{x}^{\mathbf{i}} \notin B$, being calculated by making use of the polynomials $\gamma_j \in G_{\preceq}$ as *scales of recurrence*.

(B) If instead the characteristic ideal is not given, we are quite unlikely to obtain one of them (remember it must be cofinite!) just choosing at random a set of generators: most of the times we would get a non-cofinite ideal or, when cofinite, a trivial one. The correct way to do this consists instead in choosing a finite algebraic multiset (possibly, a finite algebraic set) $\wp$ and then determining both the monomial linear basis B of $K[X]/\Im(\wp)$ and a set of generators for $\Im(\wp)$ by means of the machinery described in the previous sections.

III. Lastly, consider the following interpolation problem: given a finite n-dimensional algebraic multiset $\wp := \{(P_1, \mathcal{F}_1), \ldots, (P_N, \mathcal{F}_N)\}$ and a set of values $\{\alpha_{j,\mathbf{i}} \mid j = 1, \ldots, N \text{ and } \mathbf{i} \in \mathcal{F}_j\} \subset K$ determine the *unique* polynomial p of the form $\sum_{\mathbf{x}^{\mathbf{i}} \in B} a_{\mathbf{i}} \mathbf{x}^{\mathbf{i}}$ ($a_{\mathbf{i}} \in K$ and B is a monomial linear basis of $K[X]/\Im(\wp)$) for which we have $v_{P_j}^{\mathbf{i}}(p) = \alpha_{j,\mathbf{i}}$.

This problem appears to be an n-dimensional natural generalization of the unidimensional one which is solved by means of Lagrange interpolation formula (though a thorough analysis of these two shows that in some respects the analogy necessarily fails). Once more, the key point for solving this problem is to determine the monomial basis B. Let $B = \{\mathbf{x}_{\mathbf{d}_1}, \ldots, \mathbf{x}_{\mathbf{d}_m}\}$ and let $\mathcal{A}$ be the $m \times m$ matrix whose rows are of the form $(v_{P_j}^{\mathbf{i}}(\mathbf{x}_{\mathbf{d}_1}), \ldots, (v_{P_j}^{\mathbf{i}}(\mathbf{x}_{\mathbf{d}_m}))$, one for each $(P_j, \mathbf{i}) \in \wp$. Consider the vector $\overline{\alpha}$ whose components are the values $\alpha_{j,\mathbf{i}} = v_{P_j}^{\mathbf{i}}(p)$ (arranged according to the order that has been used for the rows of $\mathcal{A}$). The components of the vector $\overline{\beta} := \mathcal{A}^{-1} \cdot \overline{\alpha}$ are the coefficients of the desired polynomial.

Appendix.

1) Example of **Algorithm MB** for a 4-dimensional set.

$\mathcal{P}$	$\longleftrightarrow$	$\mathcal{F}$
$P_1 = (2,3,9,4)$	$\longleftrightarrow$	$\mathbf{d}_1 = (0,0,0,0)$
$P_2 = (2,5,7,3)$	$\longleftrightarrow$	$\mathbf{d}_2 = (0,1,0,0)$
$P_3 = (2,3,3,2)$	$\longleftrightarrow$	$\mathbf{d}_3 = (0,0,1,0)$
$P_4 = (2,5,5,1)$	$\longleftrightarrow$	$\mathbf{d}_4 = (0,1,1,0)$
$P_5 = (6,1,1,3)$	$\longleftrightarrow$	$\mathbf{d}_5 = (1,0,0,0)$
$P_6 = (2,3,3,6)$	$\longleftrightarrow$	$\mathbf{d}_6 = (0,0,0,1)$
$P_7 = (8,3,4,0)$	$\longleftrightarrow$	$\mathbf{d}_7 = (2,0,0,0)$
$P_8 = (6,5,6,3)$	$\longleftrightarrow$	$\mathbf{d}_8 = (1,1,0,0)$
$P_9 = (4,7,6,6)$	$\longleftrightarrow$	$\mathbf{d}_9 = (3,0,0,0)$
$P_{10} = (4,1,7,7)$	$\longleftrightarrow$	$\mathbf{d}_{10} = (2,1,0,0)$
$P_{11} = (2,5,7,8)$	$\longleftrightarrow$	$\mathbf{d}_{11} = (0,1,0,1)$
$P_{12} = (4,3,0,3)$	$\longleftrightarrow$	$\mathbf{d}_{12} = (0,2,0,0)$
$P_{13} = (1,1,2,5)$	$\longleftrightarrow$	$\mathbf{d}_{13} = (4,0,0,0)$
$P_{14} = (2,1,9,0)$	$\longleftrightarrow$	$\mathbf{d}_{14} = (1,2,0,0)$
$P_{15} = (4,1,7,6)$	$\longleftrightarrow$	$\mathbf{d}_{15} = (1,0,0,1)$
$P_{16} = (8,1,8,0)$	$\longleftrightarrow$	$\mathbf{d}_{16} = (3,1,0,0)$
$P_{17} = (2,1,7,6)$	$\longleftrightarrow$	$\mathbf{d}_{17} = (0,2,1,0)$
$P_{18} = (4,3,3,3)$	$\longleftrightarrow$	$\mathbf{d}_{18} = (1,0,1,0)$
$P_{19} = (4,1,1,0)$	$\longleftrightarrow$	$\mathbf{d}_{19} = (1,1,1,0)$
$P_{20} = (8,1,2,4)$	$\longleftrightarrow$	$\mathbf{d}_{20} = (2,0,1,0)$
$P_{21} = (6,3,4,8)$	$\longleftrightarrow$	$\mathbf{d}_{21} = (2,2,0,0)$
$P_{22} = (2,9,6,7)$	$\longleftrightarrow$	$\mathbf{d}_{22} = (0,3,0,0)$
$P_{23} = (2,3,7,1)$	$\longleftrightarrow$	$\mathbf{d}_{23} = (0,0,2,0)$
$P_{24} = (2,1,7,5)$	$\longleftrightarrow$	$\mathbf{d}_{24} = (0,2,0,1)$
$P_{25} = (4,1,1,2)$	$\longleftrightarrow$	$\mathbf{d}_{25} = (0,0,1,1)$
$P_{26} = (2,7,6,5)$	$\longleftrightarrow$	$\mathbf{d}_{26} = (0,4,0,0)$
$P_{27} = (2,5,5,4)$	$\longleftrightarrow$	$\mathbf{d}_{27} = (1,0,1,1)$
$P_{28} = (2,3,7,7)$	$\longleftrightarrow$	$\mathbf{d}_{28} = (0,1,1,1)$
$P_{29} = (2,3,0,2)$	$\longleftrightarrow$	$\mathbf{d}_{29} = (0,0,3,0)$
$P_{30} = (2,1,1,1)$	$\longleftrightarrow$	$\mathbf{d}_{30} = (0,1,2,0)$
$P_{31} = (4,3,3,7)$	$\longleftrightarrow$	$\mathbf{d}_{31} = (1,1,0,1)$
$P_{32} = (8,5,6,4)$	$\longleftrightarrow$	$\mathbf{d}_{32} = (3,2,0,0)$
$P_{33} = (4,5,5,2)$	$\longleftrightarrow$	$\mathbf{d}_{33} = (1,3,0,0)$
$P_{34} = (2,1,1,8)$	$\longleftrightarrow$	$\mathbf{d}_{34} = (0,2,1,1)$
$P_{35} = (8,1,1,1)$	$\longleftrightarrow$	$\mathbf{d}_{35} = (1,0,2,0)$
$P_{36} = (4,5,5,5)$	$\longleftrightarrow$	$\mathbf{d}_{36} = (1,2,0,1)$
$P_{37} = (6,1,8,8)$	$\longleftrightarrow$	$\mathbf{d}_{37} = (3,0,1,0)$
$P_{38} = (2,7,8,4)$	$\longleftrightarrow$	$\mathbf{d}_{38} = (0,3,1,0)$

The system of representatives for the corresponding monomial basis is given by

1	x_1	x_1^2	x_1^3	x_1^4
x_2	x_1x_2	$x_1^2x_2$	$x_1^3x_2$	
x_2^2	$x_1x_2^2$	$x_1^2x_2^2$	$x_1^3x_2^2$	
x_2^3	$x_1x_2^3$			
x_2^4				

x_3	x_1x_3	$x_1^2x_3$	$x_1^3x_3$
x_2x_3	$x_1x_2x_3$	$x_1^3x_3$	
$x_2^2x_3$			

x_3^2	$x_1x_3^2$
$x_2x_3^2$	

x_3^3

x_4	x_1x_4
x_2x_4	$x_1x_2x_4$
$x_2^2x_4$	$x_1x_2^2x_4$

x_3x_4	$x_1x_3x_4$
$x_2x_3x_4$	
$x_2^2x_3x_4$	

2) Example of **Algorithm MB** for a 3-dimensional algebraic multiset.

Consider the algebraic multiset $\wp$ given by

$$P_1 = (0,0,0) \quad \mathcal{F}_1 = \left[\begin{array}{cccccc} (0,0,0) & (1,0,0) & & (0,0,1) & & \\ (0,1,0) & & & & & \end{array} \right]$$

$$P_2 = (0,0,1) \quad \mathcal{F}_2 = \left[\begin{array}{cccccc} (0,0,0) & (1,0,0) & (2,0,0) & (0,0,1) & (1,0,1) & \end{array} \right]$$

$$P_3 = (1,1,0) \quad \mathcal{F}_3 = \left[\begin{array}{cccccc} (0,0,0) & (1,0,0) & & (0,0,1) & (1,0,1) & (0,0,2) \\ (0,1,0) & (1,1,0) & & (0,1,1) & & (0,1,2) \end{array} \right]$$

$$P_4 = (1,1,1) \quad \mathcal{F}_4 = \left[\begin{array}{c} (0,0,0) \\ (0,1,0) \end{array} \right]$$

The umbral representation $\mathcal{R} = \mathcal{R}(\wp)$ of $\wp$ (whose elements are intentionally arranged at random) as well as the corresponding diagram $\mathcal{MB}(\mathcal{R})$ are

$\wp$		$\mathcal{R}$		$\mathcal{MB}(\mathcal{R})$
$(1,1,0),(0,0,2)$	$\longleftrightarrow$	$R_1 = (10,\ 10,\ 02)$	$\longleftrightarrow$	$\mathbf{d}_1 = (0,\ 0,\ 0)$
$(1,1,0),(0,1,2)$	$\longleftrightarrow$	$R_2 = (10,\ 11,\ 02)$	$\longleftrightarrow$	$\mathbf{d}_2 = (0,\ 1,\ 0)$
$(0,0,1),(1,0,1)$	$\longleftrightarrow$	$R_3 = (01,\ 00,\ 11)$	$\longleftrightarrow$	$\mathbf{d}_3 = (1,\ 0,\ 0)$
$(1,1,0),(1,0,1)$	$\longleftrightarrow$	$R_4 = (11,\ 10,\ 01)$	$\longleftrightarrow$	$\mathbf{d}_4 = (2,\ 0,\ 0)$
$(0,0,1),(0,0,1)$	$\longleftrightarrow$	$R_5 = (00,\ 00,\ 11)$	$\longleftrightarrow$	$\mathbf{d}_5 = (3,\ 0,\ 0)$
$(1,1,0),(0,0,1)$	$\longleftrightarrow$	$R_6 = (10,\ 10,\ 01)$	$\longleftrightarrow$	$\mathbf{d}_6 = (0,\ 0,\ 1)$
$(0,0,1),(2,0,0)$	$\longleftrightarrow$	$R_7 = (02,\ 00,\ 10)$	$\longleftrightarrow$	$\mathbf{d}_7 = (4,\ 0,\ 0)$
$(0,0,0),(1,0,0)$	$\longleftrightarrow$	$R_8 = (01,\ 00,\ 00)$	$\longleftrightarrow$	$\mathbf{d}_8 = (1,\ 0,\ 1)$
$(1,1,0),(0,0,0)$	$\longleftrightarrow$	$R_9 = (10,\ 10,\ 00)$	$\longleftrightarrow$	$\mathbf{d}_9 = (0,\ 0,\ 2)$
$(0,0,0),(0,1,0)$	$\longleftrightarrow$	$R_{10} = (00,\ 01,\ 00)$	$\longleftrightarrow$	$\mathbf{d}_{10} = (1,\ 1,\ 0)$
$(0,0,0),(0,0,0)$	$\longleftrightarrow$	$R_{11} = (00,\ 00,\ 00)$	$\longleftrightarrow$	$\mathbf{d}_{11} = (2,\ 0,\ 1)$
$(0,0,0),(0,0,1)$	$\longleftrightarrow$	$R_{12} = (00,\ 00,\ 01)$	$\longleftrightarrow$	$\mathbf{d}_{12} = (1,\ 0,\ 2)$
$(1,1,0),(0,1,0)$	$\longleftrightarrow$	$R_{13} = (10,\ 11,\ 00)$	$\longleftrightarrow$	$\mathbf{d}_{13} = (0,\ 1,\ 1)$
$(1,1,0),(1,0,0)$	$\longleftrightarrow$	$R_{14} = (11,\ 10,\ 00)$	$\longleftrightarrow$	$\mathbf{d}_{14} = (3,\ 0,\ 1)$
$(0,0,1),(0,0,0)$	$\longleftrightarrow$	$R_{15} = (00,\ 00,\ 10)$	$\longleftrightarrow$	$\mathbf{d}_{15} = (0,\ 0,\ 3)$
$(1,1,0),(0,1,1)$	$\longleftrightarrow$	$R_{16} = (10,\ 11,\ 01)$	$\longleftrightarrow$	$\mathbf{d}_{16} = (0,\ 1,\ 2)$
$(1,1,0),(1,1,0)$	$\longleftrightarrow$	$R_{17} = (11,\ 11,\ 00)$	$\longleftrightarrow$	$\mathbf{d}_{17} = (2,\ 1,\ 0)$
$(0,0,1),(1,0,0)$	$\longleftrightarrow$	$R_{18} = (01,\ 00,\ 10)$	$\longleftrightarrow$	$\mathbf{d}_{18} = (2,\ 0,\ 2)$
$(1,1,1),(0,1,0)$	$\longleftrightarrow$	$R_{19} = (10,\ 11,\ 10)$	$\longleftrightarrow$	$\mathbf{d}_{19} = (1,\ 0,\ 3)$
$(1,1,1),(0,0,0)$	$\longleftrightarrow$	$R_{20} = (10,\ 10,\ 10)$	$\longleftrightarrow$	$\mathbf{d}_{20} = (0,\ 1,\ 3)$

Therefore, a monomial linear basis of $K[X]/\Im(\wp)$ consists of the equivalence classes (modulo $\Im(\wp)$) of the monomials

1	x_1	x_1^2	x_1^3	x_1^4
x_2	x_1x_2	$x_1^2x_2$		

x_3	x_1x_3	$x_1^2x_3$	$x_1^3x_3$
x_2x_3			

x_3^2	$x_1x_3^2$	$x_1^2x_3^2$
$x_2x_3^2$		

x_3^3	$x_1x_3^3$
$x_2x_3^3$	

References

[1] B. Buchberger, *An Algorithmic Method in Polynomial Ideal Theory*, in "N. K. Bose (ed): Recent Trends in Multidimensional System Theory", D. Reidel Publishing Co. , 1985, pp. 184-232

[2] L. Cerlienco, F. Piras *On the Continuous Dual of a Polynomial Bialgebra*, Communications in Algebra, 19(10), 2707–2727 (1991)

[3] W. Fulton *Algebraic curves.* The Benjamin Cummings P.C., 1981

[4] J.P.S. Kung *Jacobi's identity and the König-Egerváry theorem*, Discrete Math. 49(1984) 75-77

[5] L. Robbiano *Introduzione all'algebra computazionale.* Appunti per il corso INDAM. (Roma 1986/87)

[6] H.M. Möller, B. Buchberger *The construction of multivariate polynomials with preassigned zeros*, in "Computer Algebra (Marseille, 1982)", Lecture Notes in Comp. Sci., 144, Springer, Berlin-New York, 1982, pp.24-31

Authors address:

Luigi Cerlienco, Marina Mureddu
Dipartimento di Matematica.
Via Ospedale 72 — I-09124 Cagliari (Italy)
Tel. 070-2000456/7 — Fax 070-2000467
eMail CERLIENCO@VAXCA1.INFN.IT

Some Combinatorial Aspects of Time-Stamp Systems

Robert Cori Eric Sopena

Laboratoire Bordelais de Recherche en Informatique
Unité associée C.N.R.S. 1304
351, cours de la Libération
F-33405 TALENCE

Abstract

Many problems in distributed computing are solved using the paradigm of bounded time-stamps. The basic object of this powerful technique is a finite directed graph called a time-stamp system. The vertices of this graph are used as labels for processes in the system, and the arcs between the labels encode the order of creation of two processes. The requirements generally considered lead to a time-stamp system whose size is an exponential function of the maximal number k of living processes in the system. In the following, we restrict the requirements and we construct time-stamp systems of linear size in k for these restricted problems.

0 Introduction

Recently, many problems in the coordination of concurrent processes were shown to be solvable by using a bounded set of time-stamps [5, 8, 11]. Let us begin by describing the main features of these systems.

In a system, processes are created and die. When a process is created, a time-stamp is assigned to it; it indicates the "logical time" when this creation took place. The aim of this time-stamp is to allow the determination of the most recently created process between any pair of living processes. Generally [10], the set of natural numbers is used as the set of time-stamps. In [8] Israëli and Li proved that, when the number of living processes is assumed to be bounded by an integer k, a time-stamp system with a finite number of elements may be used. They observed the analogy with a combinatorial problem adressed by Erdös [6] and others [7, 12] in the sixties. Although proving the existence of such a system with $(2+\epsilon)^k$ elements, they constructed explicitly a system with 3^{k-1} elements. More recently, Zielonka [14] improved their solution by giving a time-stamp system with $k2^{k-1}$ elements; he also proved that this time-stamp system is optimal if it is required that any stamp also contains in a certain sense the name of the process. A generalization of time-stamp systems was considered in [4] as an application of the construction of a certain family of automata; these may be considered as distributed time-stamp systems.

In the following paper we consider two new problems consisting in building a restricted time-stamp system and we give solutions with a set whose size is a linear function of the maximal number k of living processes. The first restriction concerns the order in which processes die: we restrict the processes dying to the p older ones. We call these systems *p-restricted* time stamp systems. Using lexicographic product on graphs we obtain a p-restricted time-stamp system with $p2^{p-1}(2k-2p+1)$ elements.

The second restriction is obtained by weakening the information asked to the system. It is assumed that there are always exactly k living processes (immediately after the death of any process another one is created), and the determination of the last process created is required when the whole set of labels is given. We call these systems *weak* time-stamp systems. The determination of weak time-stamp systems was already considered [11] and a solution with k^2 time-stamps was given. We improve this result by giving a weak time-stamp system with $2k-1$ elements. Our construction uses a matching from the family of $(k-1)$-subsets of $\{1,\ldots,2k-1\}$ onto the family of its k-subsets. This matching was considered by many authors [1, 2, 9, 13]. Note that weak time-stamp systems may be used to solve the mail-box problem stated in [3].

Only sequential time-stamps are examined here, concurrent ones being left for further investigation.

1 Time-stamps

In this section, we give the definitions and some combinatorial results on time-stamp systems, most of them being due to Israëli & Li. Let us begin giving some notation.

A *directed graph* is defined as a finite set X of *vertices* together with a set of *arcs* which is a subset E of $X \times X$. If (x,y) is an arc, the vertex y is said to dominate x.

The set of all dominators of a vertex x is denoted by $\Gamma_G(x)$.

$$\Gamma_G(x) = \{y \mid (x,y) \in E\}$$

For a subset $Y \subset X$, $\Gamma_G(Y)$ denotes the set of vertices which are dominators of all the elements of Y.

$$\Gamma_G(Y) = \bigcap_{y \in Y} \Gamma_G(y).$$

In the whole paper we only consider loopless and antisymmetric graphs. They satisfy

$$\forall\, x, y \in X,\ (x,x) \notin E \quad \text{and} \quad (x,y) \in E \Rightarrow (y,x) \notin E$$

A sequence $(y_1, y_2, \ldots, y_p)$ of vertices is an *ordered sequence* if for any $1 \leq i < j \leq p$, y_j is a dominator of y_i.

Definition 1.1 *A time-stamp system of order k is a directed graph, in which any ordered sequence having less than k elements has a dominator.*

In such a graph, any vertex belongs to an ordered sequence of cardinality k. A related notion was considered by many authors after Erdös [6, 7, 12] namely that of a tournament (i.e. a directed graph in which for any pair of vertices $\{x, y\}$ one is the dominator of the other) satisfying the so-called property $S(p)$. For such a tournament, any subset of cardinality p has a dominator. Hence, any tournament with property $S(p)$ is a time-stamp system of order $p+1$ but the converse is not true. An example of a tournament which is a time-stamp system of order 4 and which does not satisfy $S(3)$ is given below. The lower bounds found for the number of vertices a tournament must have in order to satisfy $S(p)$, are hence not valid for time-stamps systems; however, similar constructions hold.

For any graph $G = (X, E)$ and any vertex x let us denote by G_x the graph whose vertex set is $\Gamma_G(x)$, and whose edge set is $E \cap (\Gamma_G(x) \times \Gamma_G(x))$. We get :

Proposition 1.2 *If G is a time-stamp system of order k, then for any x in X, G_x is a time-stamp system of order $k-1$.*

Proof. If $(y_1, y_2, \ldots, y_p)$ is an ordered sequence in G_x then $(x, y_1, y_2, \ldots, y_p)$ is an ordered sequence in G. If $p < k-1$, since G is a time-stamp system of order k, the sequence $(x, y_1, \ldots, y_p)$ has a dominator which is in $\Gamma_G(x)$. □

Corollary 1.3 *The number of vertices of a time-stamp system of order k is not less than $2^k - 1$.*

Proof. We use induction on k. For $k = 0, 1$ there is nothing to prove. The first non trivial case is $k = 2$ and the smallest time-stamp system of order 2 is the circuit C_3 with 3 vertices. Let G be a time-stamp system of order $k+1$ having n vertices. By the induction hypothesis and by proposition 1.2 each of the G_x's has not less than $2^k - 1$ vertices. Hence the number of arcs $|E|$ of G satisfies $|E| \geq n(2^k - 1)$. Since G is antisymmetric and loopless $|E| \leq \frac{n(n-1)}{2}$ and the result follows. □

Note that the converse of proposition 1.2 holds:

Proposition 1.4 *Let G be an antisymmetric graph such that for any vertex x, G_x is a time-stamp system of order $k-1$. Then G is a time-stamp system of order k.*

Proof. Let $(x_1, x_2, \ldots, x_l)$, $l < k$ be an ordered sequence in G. Then $(x_2, \ldots, x_l)$ is an ordered sequence in G_{x_1}. By the hypothesis it has a dominator x in G_{x_1}, and x is a dominator of $(x_1, x_2, \ldots, x_l)$. □

The following classical notion in graph theory is useful in order to build time-stamp systems.

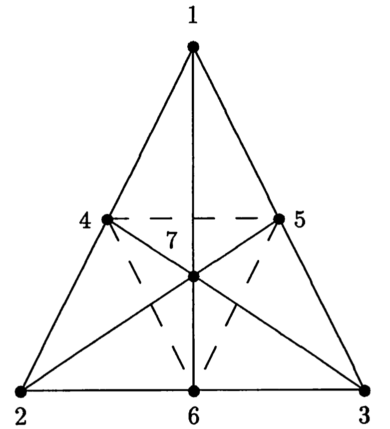
Definition 1.5 *Let $G = (X, E)$ and $H = (Y, F)$ be two directed graphs. The lexicographic product $G \otimes H$ has vertex set $X \times Y$ and its set of arcs is given by*

$$(x', y') \in \Gamma_{G\otimes H}(x, y) \text{ iff } (x, x') \in E \quad \text{or} \quad (x = x' \text{ and } (y, y') \in F)$$

Proposition 1.6 *If G and H are time-stamp systems of respective order k and l, then $G \otimes H$ is a time-stamp system of order $k + l - 1$.*

Proof. Let $(u_1, u_2, \ldots, u_m)$ be an ordered sequence in $G \otimes H$, such that $m < k + l - 1$. Let $u_i = (x_i, y_i)$, then the sequence of x_i's is an ordered sequence in G. Note that the x_i's are not necessarily distinct. If the number of distinct x_i's is less than k, they have a dominator x in X and for any $y \in Y$, (x, y) is a dominator of $(u_1, u_2, \ldots, u_m)$. If the number of distinct x_i's is not less than k, then the number of those equal to x_m is less than l. Let $(y_j, \ldots, y_m)$ be such that $x_j = x_m$ and $x_{j-1} \neq x_m$. This sequence is an ordered sequence in H with less than l elements, thus it has a dominator y and (x_m, y) is a dominator of $(u_1, u_2, \ldots, u_m)$. □

From this proposition follows a method for the construction of time-stamp systems of arbitrary order. Using the graph C_3, it is possible to get a time-stamp system of order k with 3^{k-1} vertices [8]. Other time-stamp systems are known; for little values of k the smallest are given by the tournaments satisfying $S(p)$ and for greater values



$L_1 = \{4,5,6\}$ $L_5 = \{2,3,6\}$

$L_2 = \{1,6,7\}$ $L_6 = \{3,4,7\}$

$L_3 = \{1,2,4\}$ $L_7 = \{1,3,5\}$

$L_4 = \{2,5,7\}$

Figure 1: The Fano plane.

by a construction due to Zielonka [14]. We recall here these results, one of them being that the Fano plane of order 7 gives a time-stamp system of order 3. Consider the dominators of vertex i as a line L_i of this plane. Since a time-stamp system is loopless and antisymmetric, the lines have to be numbered such that

$$i \notin L_i \text{ and } j \in L_i \Rightarrow i \notin L_j$$

This is possible for the Fano plane and this numbering is given in Figure 1.

The corresponding graph F_7 is the smallest time-stamp system of order 3, it has 7 vertices and is given by $\Gamma_{F_7}(i) = L_i$. E. & G. Szekeres [12] gave a tournament satisfying property $S(3)$ with 19 vertices, it is the smallest time-stamp system of order 4, already known. Note that $C_3 \otimes C_3 \otimes C_3$ is a time-stamp system of order 4 which is a tournament but which does not satisfy $S(3)$. The following construction, due to Zielonka [14], gives a time-stamp system of order k with $k2^{k-1}$ vertices; for $k \geq 9$ no time-stamp system with a smaller number of vertices is known.

Consider the subset X_k of $\{1 \dots k\} \times \{0,1\}^k$ consisting of elements $(\alpha, x_1, \dots, x_k)$ such that $x_\alpha = 0$, as a set of vertices of a graph $G = (X_k, E_k)$ and let E_k be such that

$$(\beta, y_1, \dots, y_k) \in \Gamma_G(\alpha, x_1, \dots, x_k) \text{ if } (\alpha > \beta \text{ and } x_\beta \neq y_\alpha) \text{ or } (\alpha < \beta \text{ and } x_\beta = y_\alpha)$$

Proposition 1.7 *G is a time-stamp system of order k having* $k2^{k-1}$ *vertices.*

Proof. Clearly the number of vertices of G is $k2^{k-1}$. Since there are no arcs between two vertices with the same first component, any ordered sequence U of G must have vertices in which all the first components are distinct. Now, if U has less than k elements then at least one $\alpha \in \{1, 2, \ldots, k\}$ is available for the first component of a dominator x of U. To end the proof, it is necessary to define the other components $x_1, \ldots, x_k$ of x. Of course $x_\alpha = 0$; to obtain x_β, if there exists an element $y \in U$ with β as first component take

$$x_\beta = y_\alpha \text{ if } \beta < \alpha \quad \text{and} \quad x_\beta = 1 - y_\alpha \text{ if } \alpha < \beta$$

If no such element exists take $x_\beta = 0$. □

2 Restricted Time-Stamp Systems

Consider the family $\Im$ of ordered sequences having k elements in a time-stamp system of order k; then, for any U and any $u_i \in U$, there exists $v \in X$ such that $U \backslash \{u_i\} \cup \{v\} \in \Im$.

Returning to the labeling of processes, this means that when k processes are living and one dies, then a time-stamp can be given to a new process. Let us restrict the set of processes which may die to the elder ones, introducing the following notion.

Definition 2.1 *A p-restricted time-stamp system of order k is a loopless antisymmetric graph such that there exists a family $\Im$ of ordered sequences with k elements satisfying*

(2.1) *For any vertex* $x \in G$, $\exists\, U \in \Im$ *such that* $x \in U$.

(2.2) *If* $U = (u_1, \ldots, u_k)$ *and if* $i \leq p$, $\exists v$ *such that* $(u_1, \ldots, u_{i-1}, u_{i+1}, \ldots, v) \in \Im$.

We will first build a 1-restricted time-stamp system of order k, then we will show that for a fixed p, there exists a p-restricted time-stamp system of order k with a number of vertices which is a linear function of k.

Definition 2.2 *Let G_k be the graph with vertex set $\{1, \ldots, 2k-1\}$ and for each vertex i, let $\Gamma_{G_k}(i) = \{i+1, i+2, \ldots, i+k-1\}$ where the sums are taken mod$(2k-1)$.*

This graph is a tournament, and moreover each vertex is the dominator of $k-1$ vertices and has $k-1$ dominators. It is not so difficult to verify that :

Proposition 2.3 *G_k is a 1-restricted time-stamp system of order k.*

Proof. Consider the family $\Im$ of all ordered sequences having k elements. Any $U \in \Im$ has the form $(i, i+1, \ldots, i+k-1)$, where the sums are taken $mod(2k-1)$. Since we are only checking the 1-restricted property, it is sufficient to find a dominator for $(i+1, \ldots, i+k-1)$, which is $i+k$. □

The graph G_k allows us to build p-restricted time-stamp systems for any arbitrary integer p since we have :

Proposition 2.4 *Let $H = (X, E)$ be a time-stamp system of order p. Then $H \otimes G_k$ is a p-restricted time-stamp system of order $k + p - 1$.*

Proof. Let us first give some notation. Let Y_k denote the set of vertices of G_k and for any ordered sequence $V = (v_1, \ldots, v_m)$ in $H \otimes G_k$, where $v_i = (x_i, y_i)$, let $\alpha(V)$ be the subset of X consisting of the first components of the v_i's, and let $\beta(V)$ be the subset of Y_k consisting of the second components of the elements whose first one is equal to x_m :

$$\begin{aligned}\alpha(V) &= \{x_i | u_i = (x_i, y_i)\},\\ \beta(V) &= \{y_i | x_i = x_m\}.\end{aligned}$$

Let $\Im$ be the family of ordered sequences V having $k + p - 1$ elements and such that

$$\begin{aligned}&(i) \quad card(\alpha(V)) \leq p,\\ &(ii) \quad \beta(V) = y, y+1, \ldots, y+i \;\; mod(2k-1), \; i < k.\end{aligned}$$

Thus, $\beta(V)$ consists of consecutive elements in Y_k.

We first prove that a vertex (x, y) of $H \otimes G_k$ belongs to at least one element of $\Im$. Consider an ordered sequence U in H of order p and containing x as its last element,

$$U = (x_1, x_2, \ldots, x_{p-1}, x_p = x).$$

Then, the following sequence v is an element of $\Im$:

$$(v_1 = (x_1, y_1), v_2 = (x_2, y_2), \ldots, v_p = (x, y_p), v_{p+1} = (x, y_p{+}1), \ldots, v_{p+k-1} = (x, y_p{+}k{-}1))$$

where the y_i's, $(i = 1, \ldots, p)$ are abitrarily taken in Y_k.

Let now $V = (v_1, v_2, \ldots, v_m)$ be an element of $\Im$ where $m = p + k - 1$, and consider $v_i \in V, i \leq p$. Since $\beta(V)$ is an ordered sequence in G_k we have $card(\beta(V)) \leq k$, hence either $v_i = (x_i, y_i)$ is such that $x_i \neq x_m$ or $i = p$ and $(x_j, y_j) = (x_m, y_i + j - i)$ for $j = i+1, \ldots, m$.

If $x_i = x_m$ or if $\beta(V)$ has less than k elements, let $v = (x_m, y_m)$. Then

$$(v_1, v_2, \ldots, v_{i-1}, v_{i+1}, \ldots, v_m, v)$$

is an element of $\Im$.

If $x_i \neq x_m$ and $\beta(V)$ has k elements, then $\alpha(V)$ has no more than $m-k+1=p$ elements and $\alpha(V)\setminus\{v_i\}$ has less than p elements. Since H is a time-stamp system of order p there exists $x \in X$ such that $\alpha(V)\setminus\{x_i\}\cup\{x\}$ is an ordered sequence in H with x as last element. Let $v=(x,y)$ where y is any element of Y_k. Then $(V=v_1,v_2,\ldots,v_{i-1},v_{i+1},\ldots,v_m,v)$ is an element of $\Im$. □

Corollary 2.5 *There exists a p-restricted time-stamp system of order k with* $p2^{p-1}(2k-2p+1)$ *vertices.*

3 Weak time-stamp systems

A time-stamp system allows us to compare any pair of stamps. In many applications this strong request may be weakened to the determination of the last created process when the whole set of the k living processes (namely their time-stamps) is known. This informal requirement may be made precise by the following definition :

Let X be a finite set and let $\Im$ be a family of k-subsets of X. Then $\Im'$ denotes the family of $(k-1)$-subsets Y' of X such that $\exists Y\in\Im,\ \ Y'\subset Y$.

Definition 3.1 *A weak time-stamp system of order k on the family* $\Im$ *is given by two mappings* α *and* β*:*

$$\begin{aligned}\alpha &: \Im \to X\\ \beta &: \Im' \to X\end{aligned}$$

satisfying:

(1) $\forall x\in X,\ \exists Y\in\Im,\ \ x\in Y$

(2) $\alpha(Y)\in Y$ and $\beta(Y')\notin Y'$

(3) $\alpha(Y'\cup\beta(Y'))=\beta(Y')$ and $\beta(Y\setminus\alpha(Y))=\alpha(Y)$.

Note that the two parts of (3) are equivalent provided that $\forall\ Y\in\Im,\ \ \exists\ Y'\in\Im'$ such that $Y=Y'\cup\beta(Y')$ and $\forall\ Y'\in\Im',\ \exists\ Y\in\Im$ such that $Y'=Y\setminus\alpha(Y)$.

In the context of processes, $\alpha(Y)$ is the last created time-stamp where Y is a set of k living processes, and $\beta(Y')$ is the time-stamp which has to be assigned to a new process when the set of living processes is Y'.

With this definition, it is assumed that there are always k or $k-1$ living processes. If the determination of the last created process is required for any set of less than k processes, then we are lead to a situation more or less similar to that of ordinary time-stamp systems. To verify this fact it suffices to consider the algorithm allowing us to compare any pair of time-stamps contained in a same element Y of $\Im$ by deleting iteratively the last element of Y until one of the two time-stamps to compare is found.

The following proposition allows us to build weak time-stamp systems :

Proposition 3.2 *There exists a weak time-stamp system on* $< X, \Im >$ *if and only if (1) is satisfied and there exists a bijection* λ *of* $\Im$ *onto* $\Im'$ *such that*

$$\forall\, Y \in \Im, \quad \lambda(Y) \subset Y. \tag{4}$$

Proof. Let $\Im$ be a family of k-subsets of X satisfying (1), and let λ be a bijection of $\Im$ onto $\Im'$ satisfying (4). Define α and β by:

$$\begin{aligned} \alpha(Y) &= Y \setminus \lambda(Y), \\ \beta(Y') &= \lambda^{-1}(Y') \setminus Y'. \end{aligned}$$

Clearly, the definitions of α and β imply (2). The verification of (3) is straightforward:

$$\alpha(Y' \cup \beta(Y')) = \alpha(\lambda^{-1}(Y')) = \lambda^{-1}(Y') \setminus \lambda(\lambda^{-1}(Y')) = \beta(Y').$$

Conversely, let $(X, \Im, \alpha, \beta)$ be a weak time-stamp system and consider λ defined by $\lambda(Y) = Y \setminus \alpha(Y))$, it is easy to verify that λ' defined by $\lambda'(Y') = Y' \cup \beta(Y')$ is the inverse of λ. □

Corollary 3.3 *For any weak time-stamp system* $(X, \Im, \alpha, \beta)$, $|X| \geq 2k-1$.

Proof. Consider the bipartite graph whose vertex set is $\Im \cup \Im'$, and whose edge set is given by the pairs $\{Y, Y'\}$ satisfying $Y' \subset Y$. In this graph, every element $Y \in \Im$ has valency k and any element $Y' \in \Im'$ has valency at most $|X| - k + 1$. Thus, if m denotes the cardinality of $\Im$ we get

$$km \leq m(|X| - k + 1)$$

and the result follows. □

Proposition 3.4 *For any* k*, there exists a weak time-stamp system of order* k *with* $2k-1$ *elements. Moreover, the computation of the mappings* α *and* β *can be done with a number of operations which is a linear function of* k*.*

Proof. Let $X = \{1, \ldots, 2k-1\}$ and let $\Im$ be the family of all k-subsets of X. Then $\Im'$ is the family of $(k-1)$-subsets of X. The existence of a matching from $\Im'$ into $\Im$ is a classical result of combinatorial theory. It may be obtained as a consequence of Hall's theorem, also known as the "marriage theorem". The following algorithms allow the computation of $\alpha(Y)$ and $\beta(Y')$; they use a last-in/first-out stack S.

Algorithm 1 : Determination of $\alpha(Y)$

```
. for i := 1 step 1 until 2k − 1 do
.   begin
.     if i ∉ Y then push(S, i)
.     else if notempty(S)
.       then pop(S) else x := i
.   end;
.   α(Y) := x
```

Algorithm 2 : Determination of $\beta(Y')$

```
. for i := 1 step 1 until 2k − 1 do
.   begin
.     if i ∉ Y' then push(S, i)
.     else if notempty(S) then pop(S)
.   end;
.   while notempty(S)
.     do begin x := top(S); pop(S) end;
.   β(Y') := x
```

□

These algorithms can be found in [9] (exercise 1 p 567); they are there attributed to Debruijn et al. [2]. Aigner [1] proposed another algorithm using lexicographic order on the k-subsets of $\{1, 2, \ldots, 2k-1\}$ and Trehel [13] proved that these two algorithms give the same matching.

References

[1] M. Aigner, *Lexicographic matchings in Boolean algebras,* J. Comb. Theory, B, **14** (1973), 187 – 194.

[2] N.G. de Bruijn, van Ebbenhorst Tengbergen, Kuyswijk, *On the set of divisors of a number*, Nieuw Archief voor Wiskunde 2, **23** (1951), 191 – 193.

[3] R.Cori, M. Latteux, M. Roos, E. Sopena, *2-Asynchronous automata,* Theoretical Comp. Sci. **61** (1988), 93 – 102.

[4] R. Cori, Y. Metivier, W. Zielonka, *Asynchronous mappings and asynchronous cellular automata,* (1990), to appear in Information and Computation.

[5] D. Dolev, N. Shavit, *Bounded concurrent time-stamp systems are constructible,* ACM Symposium on Theory of Computing (1989), 454 – 466.

[6] P. Erdős, *On a problem in graph theory,* Math. Gaz. **47** (1963), 220 – 223.

[7] R.L. Graham, J.H. Spencer, *A constructive solution to a tournament problem,* Canad. Math. Bull **14** (1971), 45 – 48.

[8] A. Israëli, M. Li, *Bounded time-stamps,* Proc 28th IEE Symposium on Foundations of Computer Science (1987), 371 – 382.

[9] D.E. Knuth, *The art of computer programming,* Vol 3, *Sorting and searching,* Addison Wesley, Reading (1973).

[10] L. Lamport, *Time, clocks and the ordering of events in a distributed system,* Comm. ACM **21** (1978), 558 – 564.

[11] M. Li, P. Vitanyi, *How to share concurrent asynchronous wait free variables,* Proceedings ICALP 1989 , Lecture Notes in Computer Science **372** (1989), 488 – 507.

[12] E.& G. Szekeres, *On a problem of Schütte and Erdős,* Math. Gaz. **49** (1965), 290 – 293.

[13] M.Trehel, *Deux constructions équivalentes d'un jeu d'arrangements pour les fichiers inverses multi-indices,* R.A.I.R.O Informatique Théorique, **12** (1978), 3 – 14.

[14] W. Zielonka, *Time-stamp systems for a fixed set of agents,* (1990), submitted.

Invisible Permutations and Rook Placements on a Ferrers Board

Kequan Ding
Department of Mathematics, University of Wisconsin
Madison, WI 53705, USA
(preliminary version)

April 7, 1992

1 Introduction

Let λ be a partition of some integer. Write $\lambda = (\lambda_1, \ldots, \lambda_m)$, $\lambda_1 \geq \cdots \geq \lambda_m \geq 1$. Sometimes, it is convenient to write $\lambda = (1^{\nu_1} 2^{\nu_2} \cdots n^{\nu_n})$ where ν_i is the number of λ_j's which are equal to i. In this paper, we view a Ferrers board F_λ of shape λ as a two dimensional subarray of an m by n matrix, where $n = \lambda_1$ and the k-th row has length λ_k, $1 \leq k \leq m$. For example,if $\lambda = (3,1)$, then

$$F_\lambda = \begin{pmatrix} * & * & * \\ & & * \end{pmatrix}.$$

We assume, contrary to the usual convention, that F_λ is right justified; the reason for this is explained below. Let $M(F_\lambda)$ be the set of all m by n matrices (a_{ij}) with a_{ij} in some field K such that $a_{ij} = 0$ for $(i,j) \notin F_\lambda$. Thus, for $\lambda = (3,1)$,

$$\mathbf{M}(F_\lambda) = \left\{ \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ 0 & 0 & a_{23} \end{pmatrix} \right\}.$$

Say that λ is *parabolic* of type $(\mu_1, \mu_2, \ldots, \mu_k)$ if $m = n$ and there exist positive integers $\mu_1, \mu_2, \ldots, \mu_k$ such that

$$\mathbf{M}(F_\lambda) = \left\{ \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1k} \\ 0 & A_{22} & \cdots & A_{2k} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & A_{kk} \end{pmatrix} \right\}. \tag{1}$$

Thus, $\lambda = (4,3,3,1)$ is parabolic with $(\mu_1, \mu_2, \mu_3) = (1,2,1)$. This explains our use of the word parabolic: if λ is parabolic then the invertible elements in $\mathbf{M}(F_\lambda)$ are a parabolic subgroup of $GL_n(K)$. It also explains why our Ferrers boards are right justified. An element of $\mathrm{M}(F_\lambda)$ is a rook placement of shape λ if it is a (0, 1) matrix with at most one 1 in each row and column. The 1's correspond to non-attacking rooks on the board F_λ. If there are r rooks, the matrix has rank r and we say that the rook placement has rank r. If there exists some rook placement σ of rank r on F_λ, then $\lambda_k \geq r-(k-1)$, $1 \leq k \leq r$. Let σ be a permutation of $[n]=\{1,2,\cdots,n\}$, $M(\sigma)$ be the corresponding n by n rook matrix with one 1 in the i-th row and the $\sigma(i)$-th column, $1 \leq i \leq n$. Thus, we have $M(\sigma\tau) = M(\tau)M(\sigma)$. An inversion of σ is a pair $((\sigma(i), \sigma(j))$ where $i<j$ and $\sigma(i)>\sigma(j)$.

Definition 1.1 *Let r, m, n be nonnegative integers such that $r \leq min\{m,n\}$. Let $\Omega^r_{m,n}$ be the set of all integer sequences*

$$(\mathbf{u},\mathbf{v},\mathbf{w}) = (u_1,u_2,\cdots,u_{n-r},v_1,v_2,\cdots,v_r,w_1,w_2,\cdots,w_{m-r})$$

satisfying

$$\begin{cases} 0 \leq u_1 \leq u_2 \leq \cdots \leq u_{n-r} \leq r \\ 0 \leq w_1 \leq w_2 \leq \cdots \leq w_{m-r} \leq r \\ 0 \leq v_i \leq i-1, \quad 1 \leq i \leq r \end{cases} \tag{2}$$

We call the sequences satisfying the inequalities above the inversion number sequences. When $m=n=r$, $\mathbf{u}$ and $\mathbf{w}$ do not exist and the sequence $\mathbf{v}$ defined above is an inversion sequence of an ordinary permutation; see [4]. Sometimes, we write $(\mathbf{u},\mathbf{v},\mathbf{w})$ as $(\mathbf{u}|\mathbf{v}|\mathbf{w})$ to mark the position between two consecutive components, clearly.

Let R^r_λ be the set of all the rook placements of rank r on a Ferrers board of shape λ. Let W_k be the symmetric group on k letters. Let $S(k)=\{(12),(23),\cdots,(k,k-1)\}$ be its distinguished generators, which we sometimes view as k by k matrices. Let $E_{ij} \in \mathrm{M}(F_\lambda)$ be the matrix whith 1 at (i,j) position and 0's elsewhere. Now, we introduce a length function l on R^r_λ.

Definition 1.2 *Let $\nu_r = \sum_{i=1}^r E_{i,n-r+i}$. For $\sigma \in R^r_\lambda$, the length function $l(\sigma)$ is defined by*

$$l(\sigma) = min\{k+h \mid \sigma = s_k \cdots s_1 \nu_r s'_1 \cdots s'_h\}.$$

where $s_i \in S(m)$ and $s'_j \in S(n)$ and

$$s_p \cdots s_1\, \nu_r\, s'_1 \cdots s'_q \in R^r_\lambda$$

for each $p \in [k]$ and $q \in [h]$.

If $\lambda = (n^m)$ and $r=n$, then $R^r_\lambda = W_n$. This function l agrees with the usual length function on W_n in terms of the generators $S(n)$ which counts the number of inversions in a permutation. If $\sigma \in R^r_\lambda$ and $\sigma \in R^r_{\lambda'}$ for two different partitions λ and λ', it is not clear, without proof, that $l(\sigma)$ computed in R^r_λ is the same as $l(\sigma)$ computed in $R^r_{\lambda'}$. We will prove this in Lemma 5.1. Until then we will assume that $l(\sigma)$ is defined with respect to a rectangular board $\lambda=(n^m)$.

In [2], Garsia and Remmel considered another numerical function on R^r_λ.

Definition 1.3 *For each $\sigma \in R_\lambda^r$, place a dot in every cell that is above a rook or to the right of a rook and a circle in each of the remaining cells of F_λ. Let $GR(\sigma)$ denote the number of circles.*

Example 1.4 *Suppose $\lambda = (4,3,2)$ and $\sigma = E_{12} + E_{33}$. Then we get the configuration*

$$\begin{array}{cccc} \circ & 1 & \bullet & \bullet \\ \circ & \bullet & \circ & \\ 1 & \bullet & & \end{array}$$

Thus, $GR(\sigma) = 3$.

Definition 1.5 *Let*

$$R_r(\lambda, q) = \sum_{\sigma\, \in R_\lambda^r} q^{GR(\sigma)},$$

and

$$RL_r(\lambda, q) = \sum_{\sigma\, \in R_\lambda^r} q^{l(\sigma)}.$$

Call them rook polynomial and rook length polynomial, respectively.

The main results of this paper are the following three theorems.

Theorem 1.6 *There is a bijection $\Phi : R_{m,n}^r \longrightarrow \Omega_{m,n}^r$ such that if $\Phi(\sigma) = (\mathrm{u}|\mathrm{v}|\mathrm{w})$, then*

$$l(\sigma) = \sum_{i=1}^{n-r} u_i + \sum_{j=1}^{r} v_j + \sum_{k=1}^{m-r} w_k. \tag{3}$$

The triple $(\mathrm{u}, \mathrm{v}, \mathrm{w})$ may be described as follows: Suppose $\sigma = E_{x_1,y_1} + \cdots + E_{x_r,y_r}$ where $x_1 < x_2 < \cdots < x_r$. Call the rook at (x_i, y_i) the i-th rook or the i-th 1 of σ. Then,

- u_i is the number of rooks to the left of the i-th zero column in σ
- v_i is the number of rooks above and to the right (to the "northeast") of the i-th rook in σ
- w_i is the number of rooks below the i-th zero row in σ

In this paper, we introduce certain permutations, which we call *invisible permutations.* These permutations serve as a bridge between $R_{(n^m)}^r$ and $\Omega_{m,n}^r$. We will postpone the formal description of the invisible permutations to section 2.

Theorem 1.7 *If $\sigma \in R_\lambda^r$, then*

$$GR(\sigma) + l(\sigma) = \sum_{i=1}^{m} \lambda_i - \ r(r+1).$$

Theorem 1.8 *If λ is a partition, then*

$$RL_r(\lambda, q) \;=\; \sum_{1\le i_1<\cdots<i_r\le m} q^{\sum_{j=1}^{r}(i_j-j)} \prod_{j=1}^{r}(\lambda_{i_j}-r+j)_q \tag{4}$$

where $(k)_q = 1 + q + q^2 + \cdots + q^{k-1}$.

These two theorems give an explicit formula for Garsia-Remmel rook polynomial

$$R_r(\lambda, q) \;=\; q^C \sum_{1\le i_1<\cdots<i_r\le m} q^{-\sum_{j=1}^{r}(i_j-j)} \prod_{j=1}^{r}(\lambda_{i_j}-r+j)_{q^{-1}}. \tag{5}$$

where

$$C = \sum_{i=1}^{m} \lambda_i - r(r+1)$$

is the constant in Theorem 1.7.

Corollary 1.9 *For an m by n rectangular board, we have*

$$RL_r(n^m, q) = \begin{bmatrix} m \\ r \end{bmatrix} \begin{bmatrix} n \\ r \end{bmatrix} [r]! \tag{6}$$

where $\left[{k \atop r}\right]$ is the Gaussian binomial coefficient and $[r]! = (1)_q(2)_q\cdots(r)_q$.

If $m = n$, this is proved in [8] by means of the root system of type A_{n-1}.

Corollary 1.10 *If $m = r$, then*

$$RL_r(\lambda, q) \;=\; (\lambda_1 - r + 1)_q(\lambda_2 - r + 2)_q \cdots (\lambda_r)_q \tag{7}$$

Corollary 1.11 *Suppose $m = r$ and λ is parabolic of type $(\mu_1, \cdots, \mu_k)$. Then,*

$$RL_r(\lambda, q) = \prod_{i=1}^{k} [\mu_i]!. \tag{8}$$

In particular, if λ and λ' are parabolic of types $(\mu_1, \cdots, \mu_k)$ and $(\mu'_1, \cdots, \mu'_k)$, respectively, where $(\mu'_1, \cdots, \mu'_k)$ is a permutation of $(\mu_1, \cdots, \mu_k)$, then $RL_r(\lambda, q) = RL_r(\lambda', q)$. Thus, the rook length polynomial is invariant under permutations of the diagonal blocks.

2 A Bijection: Maximum Rank Case

The main idea of this and the next section is to extend a rook placement σ on a m by n rectangular board to a permutation matrix $M(P(\sigma))$ of size $(m+n-r)$ which corresponds to a permutation $P(\sigma)$. We call the rows and the columns in $M(P(\sigma))$ which are not in σ the imaginary part of $M(P(\sigma))$, and call the submatrix of $M(P(\sigma))$ which is identical to σ the real part of $M(P(\sigma))$. Note that there are a lot of different ways to extend a rook placement to a permutation. We choose the permutation so that the column indices of the rooks in the imginary part of $M(P(\sigma))$ increase as the row indices increase. We call the permutation $P(\sigma)$ obtained in this way an *invisible permutation* of σ as this permutation is "behind the curtain". This explains the title of this paper. In the following, we give the definition of invisible permutations and then use an example to illustrate our idea.

Definition 2.1 *Let $\sigma \in R^r_{m,n}$. Let $b_1, b_2, \ldots, b_r$ and $c_1, c_2, \ldots, c_r$ be the column indices and the row indices of σ, respectively, such that the entries of σ at (c_i, b_i) are 1's, $i \in [r]$ and $c_1 < c_2 < \cdots < c_r$. Then, $P(\sigma)$ is a permutation in W_{m+n-r} defined by*

$$P(\sigma) = \left(\begin{array}{ccc|ccc|ccc} 1 & \cdots & n-r & n-r+c_1 & \cdots & n-r+c_r & d_1 & \cdots & d_{m-r} \\ a_1 & \cdots & a_{n-r} & b_1 & \cdots & b_r & n+1 & \cdots & n+m-r \end{array}\right)$$

where $\{a_1, a_2, \ldots, a_{n-r}\}$ is the complement of $\{b_1, b_2, \ldots, b_r\}$ in $[n]$ with $a_1 < a_2 < \cdots < a_{n-r}$, and $\{d_1, d_2, \ldots, d_{m-r}\}$ is the complement of $\{n-r+c_1.\cdots, n-r+c_r\}$ in $[n+m-r]-[n-r]$ with $d_1 < d_2 < \cdots < d_{m-r}$.

Example. Let $r = 2$, $m = n = 4$.

$$\sigma = E_{14} + E_{32} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

According to Definition 2.1, $c_1 = 1$, $c_2 = 3$, $b_1 = 4$ and $b_2 = 2$. Then, $\{a_1, a_2\} = [n] - \{b_1, b_2\} = \{1, 3\}$. In particular, $a_1 = 1$ and $a_2 = 3$. Similarly, $\{d_1, d_2\} = [n+m-r] - [n-r] - \{n-r+c_1, n-r+c_2\} = \{4, 6\}$. Thus, $d_1 = 4$ and $d_2 = 6$. So we have the following invisible permutation.

$$P(\sigma) = \left(\begin{array}{cc|cc|cc} 1 & 2 & 3 & 5 & 4 & 6 \\ 1 & 3 & 4 & 2 & 5 & 6 \end{array}\right).$$

By the correspondence M, as mentioned in the introduction, the rook matrix corresponding to $P(\sigma)$ is

$$M(P(\sigma)) = (E_{11} + E_{23}) + (E_{34} + E_{52}) + (E_{45} + E_{66})$$

$$
= \left(\begin{array}{cccc|cc}
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
\cline{1-4}
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1
\end{array}\right).
$$

In $P(\sigma)$, the columns to the left of the first vertical line are called the column imaginary part of $P(\sigma)$. The columns to the right of the second vertical line are called the row imaginary part of $P(\sigma)$. The columns between the two vertical lines are called the real part of $P(\sigma)$. In the example above, the real part corresponds to the 4 by 4 submatrix of $M(P(\sigma))$ at the southwest corner which is exactly the rook matrix σ itself.
Remark. It might seem "natural" to put the "real" submatrix σ in the northwest corner of $M(P(\sigma))$. But, we find that is not convenient.

In order to make this paper more accessible, we consider first the special case $r = m \leq n$. The general bijection will be given in the next section. For a rectangular board, sometimes we write R^r_{nm} as $R^r_{m,n}$ to match the notation $\Omega^r_{m,n}$. We call the elements of $R^r_{r,n}$ maximum rank rook matrices. Recall the definition of invisible permutation. If $m = r$, then $c_i = i$ for $i = 1, \cdots, r$. So, Definiton 2.1 becomes

$$
P(\sigma) = \left(\begin{array}{ccc|ccc}
1 & \cdots & n-r & n-r+1 & \cdots & n \\
a_1 & \cdots & a_{n-r} & b_1 & \cdots & b_r
\end{array}\right)
$$

Now, we give the main result of this section. If $r = n$, this is Hall's classical theorem and the map in our theorem is Hall's bijection [4]. It is clear that the following theorem is the special case $m = r$ of Theorem 1.6.

Theorem 2.2 *There exists a bijection* $\Phi : R^r_{r,n} \to \Omega^r_{r,n}$ *such that* $\forall \sigma \in R^r_{r,n}$*, if* $(\mathrm{u}|\mathrm{v}) = \Phi(\sigma)$*, then*

$$
l(\sigma) = \sum_{i=1}^{n-r} u_i + \sum_{j=1}^{r} v_j.
$$

Proof. For each $\sigma = \sum_{i=1}^{r} E_{i,b_i} \in R^r_{r,n}$, consider the permutation $P(\sigma) = (\mathrm{a}|\mathrm{b})$. We count the number of inversions in $P(\sigma)$ in two steps. First, note that the imaginary part a is increasing. Any inversion of $P(\sigma)$ with some a_i involved must be of the form (a_i, b_j). Let u_i be the number of inversions of the form (a_i, b_j), $1 \leq i \leq n-r$, $1 \leq j \leq r$. Next, let v_i be the number of inversions of the form (b_j, b_i), $1 \leq i \leq r$, $1 \leq j \leq r$. Thus, we get an integer sequence

$$
(\mathrm{u}|\mathrm{v}) = (u_1, u_2, ..., u_{n-r}|v_1, v_2, ..., v_r).
$$

Write $\Phi(\sigma) = (\mathrm{u}|\mathrm{v})$.

In the rest of this section, we will refer to the inversions of the form (a_i, b_j) and (b_j, b_i) as *column inversions* and *essential inversions*, respectively. Since the imaginary part of $P(\sigma)$ is increasing,

$$0 \leq u_1 \leq u_2 \leq \ldots \leq u_{n-r} \leq r. \tag{9}$$

Notice that for each i, the number of inversions of the form (b_j, b_i) is solely determined by the sign of each difference $(b_j - b_i)$ rather than the actual values of b_j and b_i. The v_i's satisfy the following inequalities, which are exactly the restrictions on the inversion sequences of elements of W_r (see [4], or page 90 in[1]).

$$0 \leq v_k \leq k-1, \quad 1 \leq k \leq r. \tag{10}$$

Hence, the sequence $\Phi(\sigma) = (\mathbf{u}|\mathbf{v})$ satisfies the inequalities in Definition 1.1. Thus, we have $\Phi(\sigma) \in \Omega^r_{r,n}$.

Next, we need to verify that $\Phi(\sigma)$ satisfies the equality in Theorem 2.2. By the definition of the length function l, we can first move all the zero columns of σ to the left of all the nonzero columns, keeping the relative positions of the nonzero columns unchanged. This takes exactly u_{n-r+i} adjacent column swaps for the $(i+1)$-th zero column, $i = 0, 1, \cdots, n-r-1$. Then, the last r columns form a r by r permutation matrix, which has $\sum_{i=1}^r v_i$ inversions. Hence the equality in Theorem 2.2 is true. So far, we have proved that Φ is a mapping from $R^r_{r,n}$ to $\Omega^r_{r,n}$ satisfying the equality in our theorem.

From the definitions of $R^r_{r,n}$ and $\Omega^r_{r,n}$, it is easy to see that they are of the same cardinality $\binom{n}{r} r!$. Therefore, in order to show that Φ is a bijection, we need only show that Φ is surjective. We will do this in two steps.

First, consider an arbitrary sequence $(\mathbf{u}|\mathbf{v}) \in \Omega^r_{r,n}$. We are going to construct an integer sequence $(\mathbf{a}|\mathbf{b})$ such that

1. $(\mathbf{a}|\mathbf{b}) \in W_n$,
2. $\Phi((\mathbf{a}|\mathbf{b})) = (\mathbf{u}|\mathbf{v})$, and
3. $\mathbf{a}$ is a strictly increasing sequence,

The construction can be done by the following algorithm.

Algorithm 2.3

1) *For $i = 1$ to $n-r$, let $a_i = u_i + i$, (Thus, by Definition 1.1, $\{a_1, \cdots, a_{n-r}\}$ is stricly increasing.)*

2) *Define a sequence of integers $b_r, b_{r-1}, \cdots, b_1$ and a sequence of subsets $B_r, B_{r-1}, \cdots, B_1$ of $[n]$ recursively as follows: Let $B_r = [n] - \{a_1, \cdots, a_{n-r}\}$. Let b_r be the (v_r+1)-th largest element in B_r. For $i \geq 1$, let $B_{r-i} = B_{r-i+1} - \{b_{r-i+1}\}$, let b_{r-i} be the $(v_{r-i}+1)$-th largest element in B_{r-i}. (This is possible since $v_i \leq i-1$, by Definition 1.1.)*

Output $\mathbf{a} = (a_1, a_2, ..., a_{n-r})$ *and* $\mathbf{b} = (b_1, b_2, ..., b_r)$.

Clearly, $(\mathbf{a}|\mathbf{b})$ satisfies the conditions 1, 2 and 3, prior to the algorithm above.

The second step is to find an element σ of $R^r_{r,n}$ such that $P(\sigma) = (\mathbf{a}|\mathbf{b})$. To do this, we can simply throw away the imaginary part $\mathbf{a}$, and use the correspondence M, as mentioned in the introduction, on the permutation

$$\sigma = \begin{pmatrix} 1 & 2 & \cdots & r \\ b_1 & b_2 & \cdots & b_r \end{pmatrix}.$$

Then, $P(\sigma) = (\mathbf{a}|\mathbf{b})$, by definition of $P(\sigma)$. Q.E.D.

In particular, the construction of Algorithm 2.3 gives an inverse of the mapping Φ. In order to illustrate the bijection given above, we consider the following example.

Example 2.4 *Let*

$$\sigma = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

Then,

$$P(\sigma) = \left(\begin{array}{cc|cc} 1 & 2 & 3 & 4 \\ 1 & 3 & 4 & 2 \end{array}\right).$$

Thus, $(\mathbf{a}|\mathbf{b}) = (a_1, a_2|b_1, b_2) = (1, 3|4, 2)$.

First, consider the inversions of the form (a_i, b_j). Since there is no b_i which is less than $a_1 = 1$, we have $u_1 = 0$. Again, since there is only one term $b_2 = 2$ in the real part of $P(\sigma)$ which is less than $a_2 = 3$, we have $u_2 = 1$. Next, look at the real part (b_1, b_2). As there is no term in the real part of $P(\sigma)$ which is to the right of $b_1 = 4$ and bigger than 4, $v_1 = 0$. On the other hand, there is exactly one term $b_1 = 4$ in the real part of $P(\sigma)$, which is to the right of $b_2 = 2$ and bigger than 2. So, $v_2 = 1$. Hence, we have $\Phi(\sigma) = (u_1, u_2|v_1, v_2) = (0, 1|0, 1)$.

To recover $P(\sigma)$ from $\Phi(\sigma)$, we need only to add i to u_i for $i = 1, 2$, and then arrange the elements of $\{1, 2, 3, 4\} - \{u_1, u_2\}$ according to the inversion numbers $v_1 = 0$ and $v_2 = 1$. Hence, $u_1 = 0 + 1 = 1$ and $u_2 = 1 + 2 = 3$. Also,

$$B_2 = \{1, 2, 3, 4\} - \{u_1, u_2\} = \{2, 4\}.$$

As $v_2 = 1$, b_2 is the $(1 + 1)$st largest element in B_2, so $b_2 = 2$. Thus, $B_1 = B_2 - \{2\} = \{4\}$. Similarly, $v_1 = 0$ implies that b_1 is the $(0 + 1)$st largest element in B_1 so $b_1 = 4$. To recover σ from $P(\sigma)$, we can simply write $P(\sigma)$ as the permutation matrix $M(P(\sigma))$ and chop off the first two rows. Then, we are done.

3 The Bijection in General Form

In this section, we remove the restriction $r = m$ and prove Theorem 1.6 in general. The idea, here, is similar to that used in the previous section though the techniques are a little more complicated.
Remark. When $m = r$, the mapping Φ is exactly the mapping introduced in the previous section. Because of this, we use the same symbol Φ in this section.
Proof of Theorem 1.6. Consider the sequences $\{a_i\}$, $\{b_i\}$, $\{c_i\}$ and $\{d_i\}$ as defined in Definition 2.1. Now, define

$$\Phi(\sigma) = (\mathbf{u}|\mathbf{v}|\mathbf{w}) = (u_1, u_2, \cdots, u_{n-r}|v_1, v_2, \cdots, v_r|w_1, w_2, \cdots, w_{m-r}) \tag{11}$$

where

- u_i = the number of b_j's smaller than a_i,
- v_i = the number of b_j's larger than b_i with $j < i$,
- w_i = the number of $(n - r + c_j)$'s larger than d_i.

Mimic the terminology used in section 2. We call $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ the *column inversion number sequence*, the *essential inversion number sequence* and the *row imaginary inversion number sequence*, respectively. And, we call $CIN(\sigma) = \Sigma\, u_i$, $EIN(\sigma) = \Sigma\, v_i$ and $RIN(\sigma) = \Sigma\, w_i$ the *column inversion number* of σ , the *essential inversion number* of σ and the *row inversion number* of σ , respectively. By Definition 2.1, it is easy to check that $\Phi(\sigma)$ satisfies the inequalities in Definition 1.1. Further, since $\mathbf{a}$ and $\mathbf{d}$ are both increasing, all the inversions of $P(\sigma)$ are counted in the right hand side of the formula (3) exactly once. Thus,

$$l(P(\sigma)) = \sum_{i=1}^{n-r} u_i + \sum_{j=1}^{r} v_j + \sum_{k=1}^{m-r} w_k$$

where l is the usual length function on W_{m+n-r}. Since $\mathbf{a}$ and $\mathbf{d}$ are both increasing, we have $l(\sigma) = l(P(\sigma))$. Hence, formula (3) is true.

Thus, Φ is a mapping from $R^r_{m,n}$ to $\Omega^r_{m,n}$. Now, we need to show that Φ is a bijection. Clearly, the sets $R^r_{m,n}$ and $\Omega^r_{m,n}$ are of the same cardinality $\binom{m}{r}\binom{n}{r}r!$. Thus, we need only to show that Φ is surjective.

Here, we use the algorithm in section 2. Suppose we have a sequence $(\mathbf{u}|\mathbf{v}|\mathbf{w})$ from $\Omega^r_{m,n}$. First, use 1) of Algorithm 2.3 on $\mathbf{u}$. We get $\mathbf{a} = (a_1, a_2, \cdots, a_{n-r})$. Then, use 2) of the algorithm on $\mathbf{v}$ and $B_r = [n] - \{a_1, a_2, \cdots, a_{n-r}\}$. Then, we get the sequence $\mathbf{b} = (b_1, b_2, \cdots, b_r)$. Next, in a similar way, use 1) of the algorithm on $w_1 + n - r, w_2 + n - r, \cdots, w_{m-r} + n - r$. We get a strictly increasing sequence $\mathbf{d} = (d_1, d_2, \cdots, d_{m-r})$. Let

$$C = [n + m - r] - [n - r] - \{d_1, d_2, \cdots, d_{m-r}\}.$$

Arrange the elements of C, to form a strictly increasing sequence $\mathbf{c} = (n - r + c_1, n - r + c_2, \cdots, n - r + c_r)$. Then,

$$P(\sigma) = \left(\begin{array}{c|c|c} 1\,2\,3 \cdots \; n-r & \mathbf{c} & \mathbf{d} \\ \mathbf{a} & \mathbf{b} & n+1 \; n+2 \; \cdots \; n+m-r \end{array} \right).$$

If we delete the first $(n-r)$ rows and the last $(m-r)$ columns from $M(P(\sigma))$, we get σ. Thus, Φ is surjective and hence is bijective. Q.E.D.

4 The Poset $(R^r_{nm}, \leq)$

In this section, we define a partial order " $\leq$ " on R^r_{nm}.

Definition 4.1 *Define a graph with R^r_{nm} as its vertex set. If $\sigma, \tau \in R^r_{nm}$ and $\sigma \neq \tau$, we say that σ and τ are adjacent if there exists some $s \in S(m)$ such that $s\sigma = \tau$ or there exists some $s' \in S(n)$ such that $\sigma s' = \tau$ where $S(k) = \{(12), (23), \ldots, (k-1,\ k)\}$.*

Note that $S(k)$ generates W_k, and $W_m \times W_n$ acts on R^r_{nm}, transitively, by means of left and right multiplications. The graph defined as above is connected.

Definition 4.2 *Let l be defined as in Definition 1.2. For $\sigma, \tau \in R^r_{nm}$, define $\sigma \leq \tau$ if $\sigma = \tau$ or there exists a sequence of elements $\sigma_1, \sigma_2, \ldots, \sigma_k \in R^r_{nm}$ such that $\sigma = \sigma_1, \tau = \sigma_k$, each σ_i is adjacent to σ_{i+1} in the graph and $l(\sigma_i) < l(\sigma_{i+1})$, for $i \in [k-1]$. If, in particular, $\sigma = \nu_r$, then we call the sequence above a reduced sequence of τ.*

Lemma 4.3 *$(R^r_{nm},\ \leq)$ is a partially ordered set.*

Definition 4.4 *A level set of rank k in $(R^r_{nm}, \leq)$ is defined by*

$$L_k(R^r_{nm}) = \{\sigma \in R^r_{nm} \mid l(\sigma) = k\}.$$

Corollary 4.5 *The sequence $\{|L_k(R^r_{nm})|\}_{k \geq 0}$ of cardinalities of level sets is symmetric and unimodal.*

Proof. The best property of the bijection defined in formula (11) is that it maps a rook matrix to a triple of integer sequences. Each component of the triple satisfies a set of inequalities, independently. Thus, the enumeration of R^r_{nm} is decomposed into the enumeration of the three integer sequences.

By Definition 1.1, a sequence $\mathbf{u}$ satisfying $0 \leq u_1 \leq \cdots \leq u_{n-r} \leq r$ can be identified as a partition with its Ferrers board contained in a $n-r$ by r board, such that the i-th column of the rectangular board contains exactly u_i cells of the Ferrers board, $1 \leq i \leq n-r$. For example, if $n = 7$ and $r = 3$, then the sequence

$$(u_1, u_2, u_3, u_4) = (0, 1, 1, 3)$$

can be identified as a partition with the Ferrers board

$$\begin{pmatrix} 0 & * & * & * \\ 0 & 0 & 0 & * \\ 0 & 0 & 0 & * \end{pmatrix}.$$

Let $Y_{n,r}$ be the set of all partitions with their Ferrers boards contained in a $n-r$ by r board. Clearly, the inclusion relation of Ferrers boards is a partial order on the set $Y_{n,r}$. This poset has the following generating function function

$$\sum_{\mathbf{u}\in Y_{n,r}} q^{|\mathbf{u}|} = \begin{bmatrix} n \\ r \end{bmatrix}$$

where $|\mathbf{u}| = \sum_{i=1}^{n-r} u_i$. (This formula has many proofs. See for example [6, 3]). Similar is the case for the set $\{\mathbf{w}\}$. By Hall's argument, we know that the set $\{\mathbf{v}\}$ satisfying $0 \le v_i \le i-1$, $1 \le i \le r$ is in one to one correspondence with W_r and $\sum v_i$ counts the inversions in the permutation corresponding to $\{\mathbf{v}\}$. Thus, by Rodrigues' formula [7],

$$\sum_{\omega\in W_r} q^{l(\omega)} = \sum_{0\le v_i\le i-1} q^{|\mathbf{v}|} = [r]!$$

where $|\mathbf{v}| = \sum_{i=1}^{r} v_i$. Thus, by Theorem 1.6, we know that

$$\begin{aligned} \sum_{\sigma\in R^r_{nm}} q^{l(\sigma)} &= \sum_{\sigma\in R^r_{nm}} q^{CIN(\sigma)+EIN(\sigma)+RIN(\sigma)} \\ &= (\sum_{\mathbf{u}\in Y_{n,r}} q^{|\mathbf{u}|})(\sum_{0\le v_i\le i-1} q^{|\mathbf{v}|})(\sum_{\mathbf{w}\in Y_{m,r}} q^{|\mathbf{w}|}) \\ &= (\sum_{\mathbf{u}\in Y_{n,r}} q^{|\mathbf{u}|})(\sum_{\omega\in W_r} q^{l(\omega)})(\sum_{\mathbf{w}\in Y_{m,r}} q^{|\mathbf{w}|}) \\ &= \begin{bmatrix} n \\ r \end{bmatrix} [r]! \begin{bmatrix} m \\ r \end{bmatrix} \end{aligned}$$

This is the formula in Corollary 1.9. Since the coefficients of $[{k \atop r}]$ and $[r]!$ are symmetric and unimodal, the coefficients of $\sum_{\sigma\in R^r_{nm}} q^{l(\sigma)}$ is symmetric and unimodal.Q.E.D.

Lemma 4.6 *Let $\lambda = (\lambda_1, \cdots, \lambda_m)$ be a partition with $\lambda_1 = n$. Then, R^r_λ is an order ideal of R^r_{nm}.*

Proof. Suppose for a fixed $\sigma \in R^r_\lambda$, there exists some $\tau \le \sigma$ such that $\tau \notin R^r_\lambda$. Choose such τ so that $l(\tau)$ is maximum. Then, by the definition of our poset $(R^r_{nm}, \le)$, there exists a sequence of elements $\tau_0, \tau_1, \ldots, \tau_t \in R^r_{nm}$ such that $\tau = \tau_0, \sigma = \tau_t$, each τ_i is adjacent to τ_{i+1}, and $l(\tau_i) < l(\tau_{i+1})$, $i \in [t-1]$. Since $l(\tau)$ is maximum, $\tau_i \in R^r_\lambda$, for all $i > 0$. And, there exists some $s \in S(m)$ such that $s\tau = \tau_1$ or some $s' \in S(n)$ such that $\tau s' = \tau_1$. Without loss of generality, we consider the first case, only. Then, there exists some k such that $s = (k, k+1)$. The fact that $\tau \notin R^r_\lambda$ and $\tau_1 \in R^r_\lambda$ implies the following two facts:

- there is a rook in the $(k+1, n-h)$ position of τ such that $n-h > \lambda_{k+1}$, i.e., in τ, this rook is outside the board F_λ.
- if the k-th row of τ is nonzero, the rook in the k-th row is to the right of the $(n-h)$-th column since otherwise, $\tau_1 \notin R^r_\lambda$.

Therefore, $l(\tau) > l(\tau_1)$ by the equality in Theorem 1.6 whether the k-th row is zero or not. Thus, we get a contradiction. Q.E.D.

5 Evaluation of Length Function l

Lemma 5.1 *Suppose λ and μ are partitions and $\sigma \in R^r_\lambda \bigcap R^r_\mu$. Then the length of σ viewed in R^r_λ is the same as the length of σ viewed in R^r_μ.*

Combine the lemma with Theorem 1.6, we have

Corollary 5.2 *If $\sigma \in R^r_\lambda$, then*

$$l(\sigma) = \sum_{i=1}^{n-r} u_i + \sum_{j=1}^{r} v_j + \sum_{k=1}^{m-r} w_k.$$

where $(\mathrm{u}, \mathrm{v}, \mathrm{w})$ *is obtained by the method described in the section before, viewing σ as a m by n rook matrix.*

Proof of Lemma 5.1. Write $l_\lambda(\sigma)$ and $l_\mu(\sigma)$ for the two lengths which we want to prove equal. We may assume that $\mu = (n^m)$ and write $l_\mu(\sigma)$ as $l(\sigma)$ as in Section 2, 3 and 4. For any given $\sigma \in R^r_\lambda$, if

$$\sigma = s_k s_{k-1} \cdots s_2 s_1 \, \nu_r \, s'_1 s'_2 \cdots s'_{h-1} s'_h$$

with $(k+h)$ being minimum, then Lemma 4.6 tells us that each

$$s_{k'} s_{k'-1} \cdots s_2 s_1 \, \nu_r \, s'_1 s'_2 \cdots s'_{h'-1} s'_{h'} \in R^r_\lambda,$$

for $1 \le k' \le k, 1 \le h' \le h$. Thus, each reduced sequence of σ in R^r_λ is also a reduced sequence of σ in $R^r_{n^m}$. Hence, our equality $l_\lambda(\sigma) = l(\sigma)$ is true. Q.E.D.

In fact, the length function l can be evaluated "locally", in the sense of the following corollary, according to the explanation of u_i, v_j, w_k as described in the introduction.

Corollary 5.3 *Let $\sigma \in R^r_\lambda$ such that $\sigma = \sum_{i=1}^r E_{x_i,y_i}$ and $x_1 < x_2 < \cdots < x_r$. Let α_i be the number of zero rows above the x_i-th row in σ, γ_i the number of zero columns to the right of the y_i-th column in σ, and β_i the number of 1's to the "northeast" of the i-th 1. Then,*

$$l(\sigma) = \sum_{i=1}^{r} (\alpha_i + \beta_i + \gamma_i). \tag{12}$$

Proof. Clearly, $v_i = \beta_i$, for each i. Note that

$$\begin{aligned}\sum_{i=1}^{n-1} u_i &= \sum_{i=1}^{n-r} \text{number of rooks to the left of the } i\text{-th zero column in } \sigma \\ &= \sum_{i=1}^{r} \text{number of zero columns to the right of the } i\text{-th rook} \\ &= \sum_{i=1}^{r} \alpha_i.\end{aligned}$$

Similarly, we have $\sum_{i=1}^{m-r} w_i = \sum_{i=1}^{r} \gamma_i$. Q.E.D.

In [8], it was proved that on a square board, let $\{(x_i, y_i)\}_{1 \leq r}$ be as above, then

$$l(\sigma) = \sum_{i=1}^{r} (x_i + n - y_i) + \text{Inv}(y_1, \cdots, y_r) - r(r-1). \tag{13}$$

Here,

$$\text{Inv}(y_1, \cdots, y_r) = \sum_{i=1}^{r} \mu_i$$

where μ_i is the number of y_j such that $y_j > y_i$, and $j < i$, $1 \leq i \leq r$.

Remark 5.4 *The sum in formula (13) can be realized by the following procedure. First, move the i-th 1 of σ to the north until it reaches the top row. Then, move the 1 to the right until it reaches the northeast corner of the Ferrers board. Count the total number of the steps taken by the 1. Now, sum over all i, $1 \leq i \leq r$. Next, delete all the zero rows and all the zero columns from σ. The array obtained in this way is a r by r permutation matrix with inversion number equal to $Inv(y_1, \cdots, y_r)$.*

In the following, we will prove that the formula (13) is true on arbitrary Ferrers boards. Although we could prove this by repeating the argument above, we will show directly that the right hand side of formula (13) equals the right hand side of formula (12).

Lemma 5.5 *Let $\sigma \in R_\lambda^r$. Then,*

$$\sum_{i=1}^{r} (\alpha_i + \beta_i + \gamma_i) = \sum_{i=1}^{r} (x_i + n - y_i) + \text{Inv}(y_1, \cdots, y_r) - r(r-1).$$

Proof. Note that

$$\text{Inv}(b_1, \cdots, b_r) = \sum_{i=1}^{r} \gamma_i.$$

We need only show that

$$r(r-1) + \sum_{i=1}^{r} (\alpha_i + \beta_i) = \sum_{i=1}^{r} (x_i + n - y_i).$$

Note that we can obtain the right hand side as in the first step of the procedure in Remark 5.4. If we look at the procedure more carefully., we see that as the i-th 1 in σ, moves to the top row, it passes every non-zero row above the x_i-th row (say, there are k_i of them), and every non-zero column to the right of the y_i-th column (say, there are h_i of them). Meanwhile, the i-th 1 in σ passes through all the zero rows above the x_i-th row and all the zero columns to the right of the y_i-th column. Hence,

$$\sum_{i=1}^{r}(x_i + n - y_i) = \sum_{i=1}^{r}(\alpha_i + \beta_i + k_i + h_i).$$

Note that $(k_1, \cdots, k_r)$ and $(h_1, \cdots, h_r)$ are two permutations of the set $[r]$, so,

$$\sum_{i=1}^{r} h_i = \sum_{j=1}^{r} k_j = \binom{r}{2}.$$

Hence, our equality is proved, for all $\sigma \in R_\lambda^r$. Q.E.D. Remark. So far, we have four different ways to evaluate our length function l. They are given by Definition 1.2, Theorem 1.6, formula (13) and formula (12). The fourth is the only one which calculates the length function, "locally", i.e., it counts the contribution of each rook in σ, individually. This local property is very useful in the proof of Theorem 1.7 and Theorem 1.8. In fact, Theorem 1.7 gives another way to evaluate our length function.

Proof of Theorem 1.7. By Lemma 4.6, all rook placements of shape λ with r rooks can be obtained by adjacent row interchanges and adjacent column interchanges from the "canonical" rook placement ν_r so that all the intermediate rook placements stay inside F_λ. We need only show that $GR(\sigma) + l(\sigma)$ is an invariant under the action of the transposition $s_i = (i, i+1)$, on the rows or on the columns.

Without loss of generality, we prove that this is true for the action of s_i on the rows, only. The situation concerning the columns is exactly the same. We separate four cases.

a) the i-th and the $(i+1)$-th row are both non-zero,

b) the i-th row is non-zero and the $(i+1)$-th row is zero,

c) the i-th row is zero and the $(i+1)$-th row is non-zero,

d) the i-th and the $(i+1)$-th row are both zero.

In case a), there are two subcases.

a1) The i-th rook is to the right of the $(i+1)$-th rook.

a2) The i-th rook is to the left of the $(i+1)$-th rook.

In case a1), let k and j be the column indices of the i-th and the $(i+1)$-th rook, respectively. Then, $j < k$. Consider the two row array formed by the i-th row and the $(i+1)$-th row.

$$\begin{array}{ccccccc} & & (j) & & (k) & & \\ \cdots & & \bullet & \cdots & 1 & & \cdots \\ \cdots & & 1 & \cdots & \bullet & & \cdots \end{array}$$

By Definition 1.3, if there is a column of the form $\genfrac{}{}{0pt}{}{\bullet}{\bullet}$ in this array, either all columns of the form $\genfrac{}{}{0pt}{}{\bullet}{\bullet}$ are to the right of the k-th column, i.e., they are determined by the i-th rook and the $(i+1)$-th rook or they are both determined by a rook in some row below the $(i+1)$-th row. In both of the cases, the action of s_i does not change this column $\genfrac{}{}{0pt}{}{\bullet}{\bullet}$. If there is a column of the form $\genfrac{}{}{0pt}{}{\circ}{\circ}$ in the array, the column must be to the left of the j-th column. Again, this column will not be affected by the action of s_i. If there is a column of the form $\genfrac{}{}{0pt}{}{\bullet}{\circ}$ then the dot on the top is determined by the rook of the i-th row, which implies that the column is to the right of the k-th column. But, this means the bottom cannot be a "o", by Definition 1.3. So, such a column does not exist. If there is a column of the form $\genfrac{}{}{0pt}{}{\circ}{\bullet}$ the column must be between the j-th and the k-th columns. Thus, the bottom dot is determined by the $(i+1)$-th rook, only. Hence, the action of s_i changes this column into the form $\genfrac{}{}{0pt}{}{\bullet}{\circ}$. At last, the j-th column $\genfrac{}{}{0pt}{}{\bullet}{1}$ is changed into the form $\genfrac{}{}{0pt}{}{1}{\circ}$ and the k-th column $\genfrac{}{}{0pt}{}{1}{\bullet}$ is changed into the form $\genfrac{}{}{0pt}{}{\bullet}{1}$. Therefore, the local property of l tells us that the action of s_i increases GR by 1 and decreases l by 1, so, their sum is not affected by the action of s_i.

By a similar argument, we can prove that the sum of l and GR is not affected by the action of s_i in the cases a2), b) and c). Clearly, in case d), the action of s_i does not change $GR(\sigma)$ or $l(\sigma)$. So, the sum $GR(\sigma) + l(\sigma)$ is an invariant under the action of the symmetric group of the rows. For the columns, the situation is exactly the same. Thus, $GR(\sigma) + l(\sigma)$ is a constant C. The constant C can be figured out from the canonical form ν_r as given in Definition 1.2, in which $l(\nu_r) = 0$ and hence

$$C = GR(\nu_r) = \sum_{i=1}^{l} \lambda_i - r(r+1).$$

Therefore, Theorem 1.7 is proved. Q.E.D.

6 Proof of Theorem 1.8

To illustrate the idea of our argument, we start with the case that $r = m = 2$. Consider a Ferrers board, in which the first row has λ_1 cells and the second row has λ_2 cells. Suppose the rook in the second row is in the k-th cell from the right. Now, consider the position of the first rook. there are two cases: Either the first rook is to the left of the second rook or the first rook is to the right of the second rook.

Case 1. The first rook is to the left of the second rook. In this case, by formula (12), the second rook contributes q^{k-1} to the generating function because it has $(k-1)$ zero columns to the right and there is no rook to the northeast. At the same time, the first rook has a

contribution q^i, $k-1 \le i \le \lambda_1 - 2$.
Case 2. The first rook is to the right of the second rook. In this case, there are $(k-2)$ zero columns to the right of the second rook and there is only one rook to the northeast. Again, by formula (12), the second rook contributes q^{k-1} to the generating function. Note that the first rook contributes q^j, $1 \le j \le k-2$.

Hence, combining the two cases, the generating function of the rook placements of this shape is given by the following expression.

$$\sum_{k=1}^{\lambda_2} q^{k-1}(1 + q + \cdots + q^{\lambda_1 - 2}) = (\lambda_2)_q(\lambda_1 - 1)_q.$$

From the argument above, we can see that the local property of l makes it possible to look at the contribution of each rook to the generating function, individually. With this in our mind, we can go ahead to look at the general case. From the formula obtained above, we conjecture that for $r = m$, we have

$$\begin{aligned} RL_m(\lambda, q) &= (\lambda_1 - m + 1)_q(\lambda_2 - m + 2)_q \cdots (\lambda_{m-1} - 1)_q(\lambda_m)_q \\ &= \prod_{j=1}^{m} (\lambda_j - m + j)_q. \end{aligned} \tag{14}$$

This is the formula in Corollary 1.10. If $m = 1$, this is obvious. When m equals 2, the formula is proved above. Now, we use induction on m. Suppose that the rook in the last row (i.e., the m-th row) is in the k-th cell from the right. Since $m = r$, this rook is the m-th rook. Then, by formula (12), the m-th rook has $k - 1 - t$ zero columns to the right if there are t rooks living to the right of the m-th rook. Hence, the m-th rook contributes q^{k-1} to the generating function. This is independent of the configuration formed by the first $m-1$ rooks. Now, delete the k-th column (from the right side) and the last row from the Ferrers board and use the induction hypothesis on that board, we get the formula 14.

$$\begin{aligned} RL_m(\lambda, q) &= \prod_{j=1}^{m-1}((\lambda_j - 1) - (m-1) + j)_q \sum_{h=1}^{\lambda_m} q^{h-1} \\ &= \prod_{j=1}^{m}(\lambda_j - m + j)_q \end{aligned}$$

Finally, we consider the general case, i.e., $r \le m$. Suppose that the r rooks live in the rows with indices $i_1, i_2, \cdots, i_r$. Thus, by formula (12), the zero rows give an extra factor $q^{\sum_{u=1}^{j}(i_u - i_{u-1} - 1)}$ to the contribution of the rook in the i_j-th row, which does not appear in the case $r = m$. Therefore, using formula (14),

$$RL_r(\lambda, q) = \sum_{1 \le i_1 < \cdots < i_r \le m} q^{\sum_{j=1}^{r}(r-j+1)(i_j - i_{j-1} - 1)} \prod_{j=1}^{r}(\lambda_{i_j} - r + j)_q$$

Here, we used the convention that $i_0 = 0$. Since $\sum_{j=1}^{r}(i_j - j) = \sum_{j=1}^{r}(r - j + 1)(i_j - i_{j-1} - 1)$, the theorem is proved. Q.E.D.

Proof of Corollary 1.11 Let λ be a parabolic board of type $(\mu_1, \cdots, \mu_k)$. Then, $\lambda_s - (m - i) = j$, if $s = \sum_{i=1}^{t} \mu_i + j$ and $1 \le j \le \mu_{t+1}$ for some integer t, $1 \le t \le k$. Therefore, by formula (14), we have

$$RL_m(\lambda, q) = (\lambda_1 - m + 1)_q (\lambda_2 - m + 2)_q \cdots (\lambda_{m-1} - 1)_q (\lambda_m)_q \; = \prod_{i=1}^{k} [\mu_i]!.$$

Since the formula above does not change if we permute the μ_i's, the rook length polynomial is invariant under the permutations of diagonal blocks. Q.E.D.

Comments. The results obtained here have other consequences. For example, Corollary 1.9 and formula (14) tell us that the coefficients of rook (length) polynomials are a unimodal sequence for arbitrary rook placements on a rectangular board and those maximum ranked rook placements on any Ferrers board. These strongly support the general conjecture made by Garsia and Remmel that the sequence of coefficients of $R_r(\lambda, q)$ is unimodal.

The corollary of 1.10 tells us that if we swap the second and the third diagonal blocks in a parabolic Ferrers board of shape $\lambda = (6, 5, 5, 3, 3, 3)$, then the rook length polynomials for the boards with $\lambda = (6, 5, 5, 3, 3, 3)$ and $\lambda' = (6, 5, 5, 5, 2, 2)$ are exactly the same though the two boards are quite different!

```
* * * * * *        * * * * * *
  * * * * *          * * * * *
  * * * * *          * * * * *
      * * *          * * * * *
      * * *                * *
      * * *                * *
```

Masao Ishikawa pointed out that the bijection Φ in section 3 does not preserve the partial order and in general $(R^r_{m,n}, \le)$ is not isomorphic to the direct product $Y_{n,r} \times W_r \times Y_{m,r}$. It is my pleasure to thank him for this and many other suggestions.

References

[1] R. A. Brualdi, Introductory Combinatorics, 2nd edition, North-Holland, 1992, p. 90.

[2] A. M. Garsia and J. B. Remmel, q-counting rook configurations and a formula of Frobenius, J. Combin. Theory Ser. A 41(1986) 246-275.

[3] I. M. Gessel, A q-analog of the exponential formula, Discrete Mathematics 40(1982) 69-80.

[4] M. Hall Jr, Proceedings Symposium in Pure Mathematics, AMS, Providence, 6(1963), 203.

[5] M. Ishikawa, a private letter, December 28, 1991.

[6] A. Nijenhuis, A. E. Solow and H. S. Wilf, Bijective methods in the theory of finite vector spaces, J. of Combin. Theory Ser. A, 37(1984) 80-84.

[7] O. Rodrigues, Note sur les inversions, ou Dérangements produits dans les permutations, J. de Math. 1 Série 4(1839) 236-239.

[8] L. Solomon, The Bruhat decomposition, Tits system and Iwahori ring for the monoid of matrices over a finite field, Geometriae Dedicata 36(1990) 15-49.

[9] M. Wachs and D. White, p,q-Stirling numbers and set partition statistics, J. Combin. Theory Ser. A 56:1(1991) 27-46.

La Correspondance de Robinson-Schensted pour les Tableaux Oscillants Gauches

Serge Dulucq †
Bruce E. Sagan ‡

Abstract. We introduce an analog of the Robinson-Schensted algorithm for Skew Oscillating Tableaux which generalize the well known correspondence for Standard Tableaux. We show that this new algorithm enjoy some of the same properties as the original. In particular, it is still true that replacing a permutation by its inverse exchanges the two output tableaux. These facts permit us to derive a number of identities involving $\tilde{f}_n^{\alpha\to\beta}$.

1. Introduction

Les tableaux de Young, introduits en 1902 par A. Young [You] pour calculer certains idempotents de l'algèbre du groupe symétrique, ont été très abondamment étudiés sous différents points de vue. En particulier, dans le cadre de la théorie des représentations du groupe symétrique, l'algorithme de Robinson-Schensted [Rob, Sche] donne une preuve combinatoire de l'identité

$$n! = \sum_{\lambda \vdash n} \left(f^{\lambda}\right)^2 , \tag{1}$$

identité dans laquelle $n!$ est la dimension de la représentation régulière, et f^{λ} est à la fois le degré et la multiplicité de la représentation associée à la forme λ. Depuis, une riche combinatoire sur les tableaux de Young a vu le jour. Ainsi, plusieurs correspondances analogues à celle de Robinson-Schensted ont été proposées pour divers types de tableaux: tableaux semi-standards [Knu], tableaux oscillants [DDF, Fav, Sun, McL], "shifted" tableaux [Sag, Wor], tableaux gauches [SaSt], ...

† LaBRI Université Bordeaux I, 351 Cours de la Libération, 33405 Talence Cédex.
E-mail : dulucq@labri.greco-prog.fr

‡ Department of Mathematics, Michigan State University, East Lansing, MI 48824-1027.
E-mail : sagan@nsf1.mth.msu.edu

Dans ce travail, nous nous intéressons à la famille des tableaux oscillants gauches, de tels tableaux correspondant à des chemins quelconques allant d'une certaine forme initiale à une certaine forme finale dans le treillis de Young représenté figure 1.

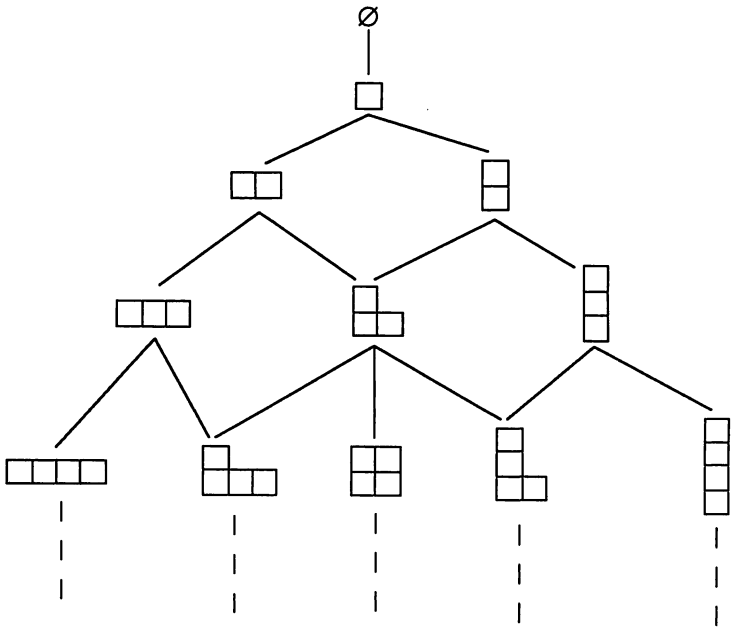


Figure 1

Ces tableaux oscillants gauches généralisent donc les notions de tableaux oscillants (pour lesquels la forme initiale est la forme vide ∅), tableaux standards (la forme initiale est ∅ et le chemin est une chaîne dans le poset représenté figure 1), tableaux gauches (la forme initiale est quelconque et le chemin est une chaîne dans ce même poset).

Dans cet article, nous donnons un analogue de l'algorithme de Robinson-Schensted [Rob, Sche] pour cette famille des tableaux oscillants gauches. Nous montrons également que cette correspondance possède les mêmes propriétés que l'originale, en particulier celle qui fait que le remplacement d'une permutation par son inverse a pour effet d'échanger les deux tableaux obtenus [Schü]. De ces résultats, nous déduisons un certain nombre d'identités faisant intervenir le nombre de ces tableaux.

Le paragraphe suivant présente les définitions et notations que nous utilisons au long de ce travail. Ensuite, nous proposons un algorithme (dit fondamental) associant à un tableau oscillant gauche un triplet constitué d'une involution partielle et de deux tableaux partiels, correspondance donnant une première formule d'énumération pour ces tableaux (en fonction des formes initiale et finale et de la longueur des tableaux). Dans le paragraphe quatre, nous donnons la correspondance de type Robinson-Schensted pour les tableaux oscillants gauches, déduite de l'algorithme fondamental. Nous obtenons ainsi une formule analogue à la formule (1) sur les dimensions. Dans le dernier paragraphe, nous montrons que cette correspondance possède une propriété analogue à l'originale, propriété dont nous déduisons une formule pour le nombre de tableaux oscillants gauches ayant une forme initiale et une longueur fixées.

2. Définitions et notations

Nous utilisons ici la terminologie habituelle sur les tableaux et nous supposons connu l'algorithme de Robinson-Schensted tel que présenté dans [Sche]. Les définitions non précisées ci-dessous pourront être trouvées dans cet article.

Nous notons indifféremment par $\lambda = (\lambda_1, \lambda_2, \ldots\ldots, \lambda_k)$ une *partition* et le *diagramme de Ferrers* correspondant, représenté en notation "Française", c'est à dire avec la plus longue part λ_1 sur la ligne inférieure (ligne *1*) et où les parts nulles sont interdites. La *cellule* du diagramme sur les ligne *i* et colonne *j* est notée *c=(i, j)* et ainsi, $c \in \lambda$ si et seulement si $1 \leq j \leq \lambda_i$. Si $\mu \subseteq \lambda$, la *forme gauche* λ/μ est l'ensemble de cellules $\{c \mid c \in \lambda \text{ et } c \notin \mu\}$. Si $|\lambda/\mu| = n$, nous écrivons $|\lambda/\mu| \vdash n$ et parlons de *partition (gauche)* de n pour λ/μ.

Un *tableau de Young T de forme* λ/μ est un étiquetage des cellules de λ/μ par des entiers positifs de sorte que les lignes et colonnes soient croissantes au sens large. Nous notons T_c ou $T(i, j)$ l'étiquette de la cellule $c=(i, j)$; ainsi, $k \in T$ signifie $k=T(i, j)$ pour un certain couple i, j. Un tableau de Young est *partiel* si ses étiquettes sont distinctes, *standard* s'il est partiel et les étiquettes comprises entre 1 et $n=|\lambda/\mu|$. Par exemple, losque $\lambda=(5, 4, 2)$ et $\mu=(3, 2)$, les deux tableaux suivants sont respectivement partiel et standard.

Figure 2

Les ensembles de tableaux de Young partiels et standards de forme λ/μ sont notés respectivement $PT(\lambda/\mu)$ et $ST(\lambda/\mu)$, l'ensemble $\overline{PT}(\lambda/\mu)$ désignant les tableaux de Young partiels de forme λ/μ avec décroissance en ligne et colonne.

Nous notons $f^{\lambda/\mu}$ le nombre de tableaux standards de forme λ/μ.

Une *involution partielle* π sur $[n]=\{1, 2, ..., n\}$ est un ensemble de paires verticales d'entiers de $[n]$ deux à deux distincts

$$\pi = \begin{pmatrix} i_1 \; i_2 \, ... \, i_k \\ j_1 \; j_2 \, ... \, j_k \end{pmatrix} \quad \text{où} \quad i_1 < i_2 < ...... < i_k < n \;\; \text{et} \;\; i_p > j_p \quad (1 \leq p \leq k),$$

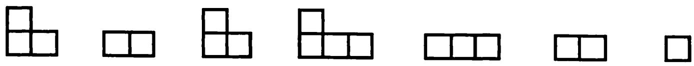
ces paires verticales constituant les cycles de l'involution qui sera toujours sans point fixe. Nous notons $\hat{\pi} = i_1 \; i_2 \, ...\, i_k$ et $\check{\pi} = j_1 \; j_2 \, ...\, j_k$ les lignes supérieure et inférieure de cette involution.
L'ensemble des involutions partielles sur $[n]$ est l'ensemble PINV(n).

Définition 2.1 Un *tableau oscillant gauche* de *formes initiale* α et *finale* β et de *longueur n* est une suite $\tilde{T}$ de diagramme de Ferrers (ou partitions)

$$\tilde{T} = (\alpha = \lambda_0, \lambda_1, ..., \lambda_n = \beta)$$

où λ_k est obtenue à partir de λ_{k-1} par ajout ou suppression d'une cellule. L'ensemble de tels tableaux est noté $TOG_n(\alpha \to \beta)$ et leur nombre $\tilde{f}_n^{\alpha \to \beta}$.

Exemple 2.2

$\tilde{T} =$ 

Ainsi $\tilde{T} \in TOG_6\{(2,1) \to (1)\}$.

Cette notion de tableaux oscillants gauches généralise donc les notions de

- tableaux oscillants [DDF, SUN, DF, MacL] qui correspondent au cas où $\alpha = \varnothing$,
- tableaux de Young standards correspondant au cas où $\alpha = \varnothing$ et où les seules opérations admises dans "l'histoire" du tableau sont des adjonctions de cellules.

Dans le prochain paragraphe, nous présentons l'algorithme fondamental à partir duquel nous obtiendrons dans la section suivante une correspondance de type Robinson-Schensted pour les tableaux oscillants gauches. Pour terminer, nous verrons que cette correspondance possède des propriétés analogues à celle de Robinson-Schensted, en particulier lorsque l'on permute les deux tableaux oscillants gauches obtenus ce qui a pour effet d'inverser (en un certain sens que nous définirons) l'involution partielle.

3. L'algorithme fondamental

Théorème 3.1 Soient α et β deux partitions et n un entier. Il existe une bijection

$$\begin{array}{ccc} \theta : \mathrm{TOG}_n(\alpha \to \beta) & \to & \mathrm{PINV}(n) \times \mathrm{PT}(\beta / \mu) \times \overline{\mathrm{PT}}(\alpha / \mu) \\ \tilde{\mathrm{T}} & \mapsto & (\pi, \mathrm{P}, \mathrm{Q}) \end{array}$$

telle que $\pi \,\dot{\cup}\, \mathrm{P} \,\dot{\cup}\, \mathrm{Q} = [n]$ (où $\dot{\cup}$ désigne l'opérateur union disjointe).

Remarque 3.2 Dans l'énoncé de ce théorème, la partition μ décrit l'ensemble de toutes les partitions λ telles que $\lambda \subseteq \alpha$ et $\lambda \subseteq \beta$.

On déduit immédiatement de ce théorème le résultat suivant.

Corollaire 3.3 Le nombre de tableaux oscillants gauches de formes initiale α et finale β et de longueur n est donné par

$$\tilde{f}_n^{\alpha \to \beta} = \sum_{\substack{k,l \geq 0 \\ n-k-l \text{ pair}}} \binom{n}{k,l,n-k-l} (n-k-l)!! \sum_{\mu : \substack{\alpha/\mu \vdash k \\ \beta/\mu \vdash l}} f^{\alpha/\mu} f^{\beta/\mu} .$$

En considérant des formes particulières pour α et β, nous retrouvons les formules d'énumération suivantes pour les tableaux oscillants [DDF, Fav, Sun].

Corollaire 3.4 (i) $\tilde{f}_n^{\varnothing \to \beta} = \binom{n}{|\beta|} (n - |\beta|)!! \, f^{\beta}$,

(ii) $\tilde{f}_{2n}^{\varnothing \to \varnothing} = (2n)!!$.

Avant de prouver le théorème 3.1, nous allons définir les différents processus de suppression dans un tableau de Young gauche partiel que nous serons amenés à considérer.

Soit $\mathrm{P} \in \mathrm{PT}(\lambda / \mu)$. Trois types de suppression peuvent être réalisées sur ce tableau.

• Suppression d'une cellule vide.

Il s'agit de la suppression d'une cellule appartenant à μ en position *(i, j)* telle que les cellules en position *(i, j+1)* et *(i+1, j)* n'appartiennent ni à μ, ni à λ/μ (figure 3 ci-après, suppression de la cellule vide *(5, 1)*).

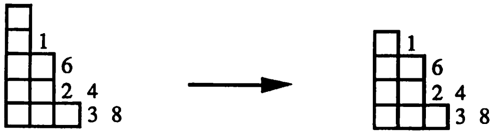


Figure 3

• Suppression d'une cellule contenant un entier.

Il s'agit de la suppression d'une cellule en position *(i, j)* de λ/μ contenant un certain entier *k*, sur le même principe que dans l'algorithme de Robinson-Schensted. Cet entier *k* vient prendre la place du plus grand entier de la ligne *i-1* qui lui soit inférieur (lorsqu'il existe), ce processus étant itéré jusqu'à ce que l'une des deux situations suivantes se produisent.

- Suppression externe :

Le processus se termine par l'expulsion d'un entier de la première ligne de λ/μ (la figure 4 représente la suppression de la cellule contenant *6* ce qui conduit à l'expulsion de l'entier *3*).

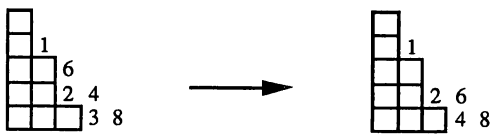


Figure 4

- Suppression interne :

Le processus se termine lorsque, à une certaine étape, l'entier *m* expulsé de la ligne *p+1* est plus petit que tous ceux de la ligne *p* (*p>0*) de λ/μ ou lorsque cette ligne *p* ne comporte aucun entier (cas où $\lambda_p=\mu_p$). Dans ce cas, l'entier *m* est placé dans la dernière cellule de μ sur la ligne *p*. Ainsi, une cellule "coin" de μ est remplie (la figure 5 représente la suppression de la cellule contenant l'entier *1* ce qui a pour effet de remplir la cellule *(3, 2)*).

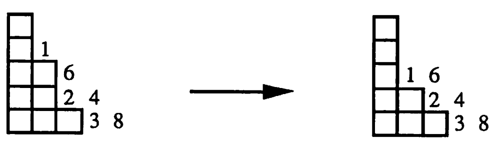


Figure 5

Preuve du théorème 3.1

Soit $\tilde{T} = (\alpha = \lambda_0, \lambda_1, ..., \lambda_n = \beta)$ un tableau appartenant à $TOG_n(\alpha \to \beta)$. On construit une suite ($\varnothing_\alpha$ désignant le tableau vide de forme α/α)

$$(\pi_0, P_0, Q_0)=(\varnothing, \varnothing_\alpha, \varnothing_\alpha), (\pi_1, P_1, Q_1), \;.....,\; (\pi_n, P_n, Q_n)=(\pi, P, Q)$$

à l'aide de l'algorithme suivant (voir exemple 3.5).

Algorithme $\mathcal{A}$**.** Supposons qu'à la kième étape, λ_k soit obtenue à partir de λ_{k-1} par :

(1) ajout d'une cellule en position (i, j) ; dans ce cas

$Q_k= Q_{k-1}$, $\pi_k= \pi_{k-1}$,

$P_k(i, j)=k$ (ajout d'une cellule à P_{k-1} étiquetée k) ;

(2) suppression d'une cellule en position (i, j) ; dans ce cas, on supprime la cellule en position (i, j) de P_{k-1}, le processus conduisant au tableau P_k par

(a) une suppression externe :

$Q_k= Q_{k-1}$,

on ajoute le couple (k,m) à π_{k-1} où m est l'entier expulsé de P_{k-1} ;

(b) une suppression interne :

$Q_k(i, j)=k$ où (i, j) est la cellule de P_{k-1} ayant été remplie,

$\pi_k= \pi_{k-1}$;

(c) une suppression de cellule vide :

$Q_k(i,j)=k$ où (i, j) est la cellule supprimée,

$\pi_k= \pi_{k-1}$.

Un exemple de cette construction est donné ci-dessous (exemple 3.5).

Il est facile de voir que, partant de $\tilde{T} \in TOG_n(\alpha \to \beta)$, cet algorithme fournit

$P \in PT(\beta / \mu)$: à chaque étape de l'algorithme, $P_k \in PT(\lambda_k / \mu_k)$,

$Q \in \overline{PT}(\alpha / \mu)$: en effet, $Q_0=\varnothing_\alpha$ et, à chaque étape, si $P_k \in PT(\lambda_k / \mu_k)$ alors

$Q_k \in \overline{PT}(\alpha / \mu_k)$ (cas 2b et 2c de l'algorithme),

$\pi \in PINV(n)$ par construction.

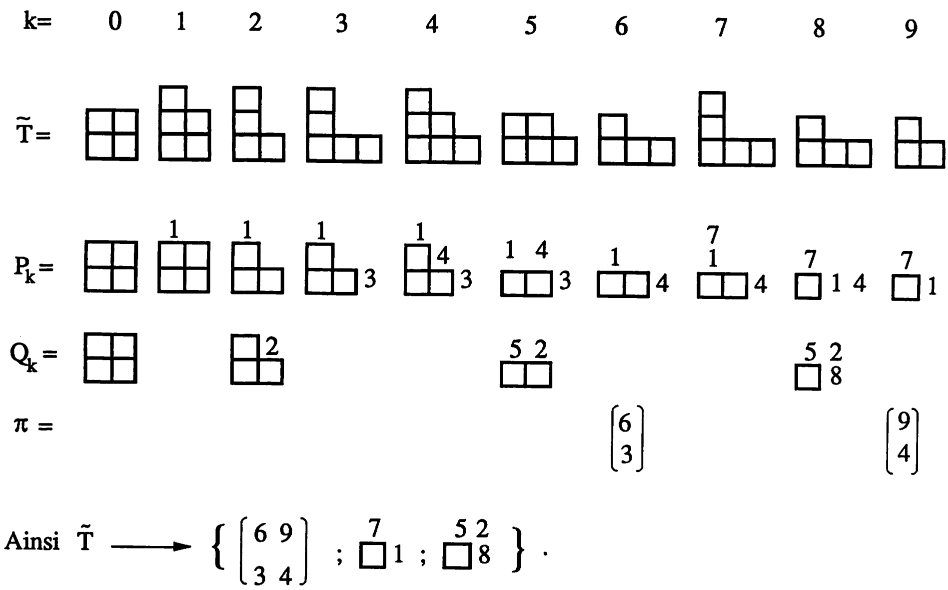
De plus, $\pi \;\dot{\cup}\; P \;\dot{\cup}\; Q = [n]$ car

Q est étiqueté par des entiers correspondant aux étapes de suppression (cas 2b et 2c),

P est étiqueté par des entiers correspondant aux étapes d'insertion (cas 1), entiers non supprimés lors d'étapes ultérieures (cas 2a),

π est constituée de couples dont les entiers correspondent aux étapes de suppression externe (cas 2a) et aux entiers expulsés.

Exemple 3.5



Pour montrer que cette application est bijective, nous allons construire son inverse. Pour cela, nous donnons un algorithme qui, partant d'un triplet (π, P, Q) appartenant à $PINV(n) \times PT(\beta/\mu) \times \overline{PT}(\alpha/\mu)$ tel que $\pi \dot{\cup} P \dot{\cup} Q = [n]$, fournit un tableau $\tilde{T}$ de $TOG_n(\alpha \rightarrow \beta)$.

Soit donc le triplet (π, P, Q).

On construit une suite $(P_n=P, Q_n=Q)$, (P_{n-1}, Q_{n-1}), ..., (P_0, Q_0) à l'aide de l'algorithme suivant.

Algorithme $\mathcal{Q}^{-1}$. Supposons qu'à l'étape k, k variant de n à 1,

(1^{-1}) l'entier k soit l'étiquette d'une cellule de P_k :

on supprime la cellule de P_k contenant cet entier ce qui donne P_{k-1},

$Q_{k-1} = Q_k$;

(2^{-1}) l'entier k n'appartient pas à P_k :

(a) le couple (k,m) appartient à π

on insère (en utilisant l'algorithme de Robinson-Schensted) l'entier m dans P_k pour obtenir P_{k-1},

$Q_{k-1} = Q_k$;

(b) $Q_k(i, j)=k$ et $P_k(i, j)=m$

on vide la cellule (i, j) de Q_k pour obtenir Q_{k-1},

on insère l'entier m dans P_k à partir de la ligne i+1 en utilisant également l'algorithme de Robinson-Shensted et la cellule (i, j) de P_k devient vide pour former P_{k-1},

(c) $Q_k(i, j)=k$ et P_k ne possède pas de cellule en position (i, j)

on vide la cellule (i, j) de Q_k pour obtenir Q_{k-1},

on ajoute la cellule vide (i, j) à P_k pour obtenir P_{k-1}.

On obtient ainsi, à l'issue de cet algorithme, la suite de tableaux ($P_n=P$, P_{n-1}, ..., P_0) où $P_k \in PT(\lambda_k / \mu_k)$. Le tableau $\tilde{T}$ est alors $\tilde{T} = (\alpha = \lambda_0, \lambda_1, ..., \lambda_n = \beta)$.

Il est facile de voir que cet algorithme $\mathcal{A}^{-1}$ correspond exactement à la construction inverse de celle fournie par l'algorithme $\mathcal{A}$,

l'étape (1^{-1}) correspondant à la suppression d'une cellule,

l'étape (2^{-1}) correspondant à l'insertion d'une cellule :

(2^{-1}a) étant une insertion externe,

(2^{-1}b) étant une insertion interne,

(2^{-1}c) étant l'insertion d'une cellule vide.

Un exemple d'application de l'algorithme $\mathcal{A}^{-1}$ est donné dans la deuxième partie de l'exemple 3.8 (3.8.2). Il correspond à la construction inverse de celle de l'exemple 3.5.

Il est maintenant naturel de se demander ce qu'il advient lorsque l'on considère le miroir du tableau oscillant gauche lorsque l'on applique l'algorithme $\mathcal{A}$. Pour cela, nous allons dans un premier temps introduire quelques notations et définitions.

Définition 3.6

(i) A un tableau $\tilde{T} = (\alpha = \lambda_0, \lambda_1, ..., \lambda_n = \beta)$ appartenant à $TOG_n(\alpha \to \beta)$, on associe son *miroir* $\tilde{T}^R = (\beta = \lambda_n, ..., \lambda_1, \lambda_0 = \alpha)$ appartenant à $TOG_n(\beta \to \alpha)$.

(ii) Etant donnés un entier n et un tableau $P \in PT(\lambda / \mu)$ (resp. $P \in \overline{PT}(\lambda / \mu)$) on définit le tableau $P^c \in \overline{PT}(\lambda / \mu)$ (resp. $P \in PT(\lambda / \mu)$) par $P^c(i, j)=n+1-P(i, j)$ pour toute cellule $(i, j) \in \lambda/\mu$.

(iii) A une involution partielle $\pi \in PINV(n)$ ayant pour cycles les couples (i_p, j_p), on associe l'involution $\pi^c \in PINV(n)$ dont les cycles sont les couples$(n+1-j_p, n+1-i_p)$.

Théorème 3.7 Soit $\tilde{T}$ un tableau oscillant gauche tel que $\theta(\tilde{T}) = (\pi, P, Q)$.

Alors $\theta(\tilde{T}^R) = (\pi^c, Q^c, P^c)$.

<u>Preuve</u>. Nous n'en donnerons ici que le principe.

Etant donné un tableau $\tilde{T} = (\alpha = \lambda_0, \lambda_1, \ldots\ldots, \lambda_n = \beta)$ tel que $\theta(\tilde{T}) = (\pi, P, Q)$, posons $(\sigma,\ A,\ B) = (\pi^c, Q^c, P^c)$.

Considérons alors la suite

$$(\pi_0, P_0, Q_0) = (\varnothing, \varnothing_\alpha, \varnothing_\alpha),\ (\pi_1, P_1, Q_1),\ \ldots\ldots,\ (\pi_n, P_n, Q_n) = (\pi, P, Q)$$

obtenue en appliquant l'algorithme $\mathcal{A}$ au tableau $\tilde{T}$,

et la suite

$$(A_n = A,\ B_n = B),\ (A_{n-1},\ B_{n-1}),\ \ldots\ldots,\ (A_0,\ B_0)$$

obtenue en appliquant l'algorithme $\mathcal{A}^{-1}$ au triplet $(\sigma,\ A,\ B)$.

La preuve du théorème consiste à montrer que, à toute étape *k*, $0 \leq k \leq n$, les tableaux P_k et A_{n-k} vérifient

$$P_k \in PT(\lambda_k / \mu_k) \text{ pour une certaine forme } \mu_k,$$
$$A_{n-k} \in PT(\lambda_k / \gamma_k) \text{ pour une certaine forme } \gamma_k,$$

c'est à dire que les formes "globales" de ces deux tableaux sont identiques, fait dont nous déduisons le résultat attendu, soit

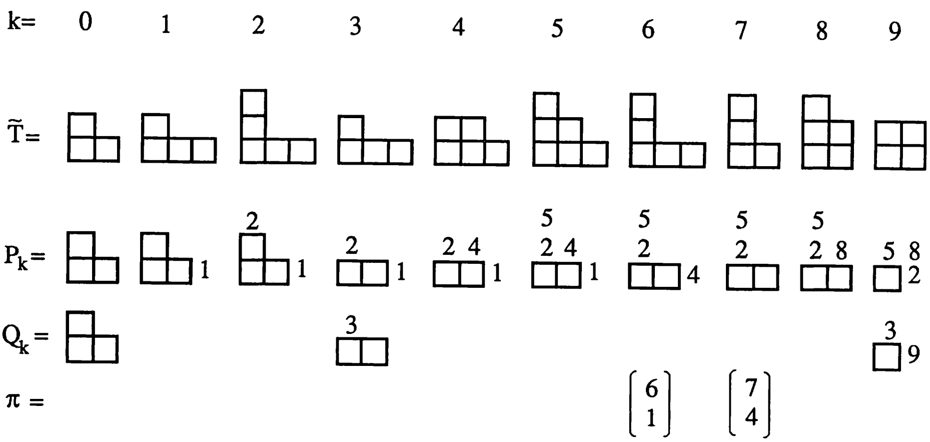
$$\theta^{-1}(\pi^c, Q^c, P^c) = \tilde{T}^R.$$

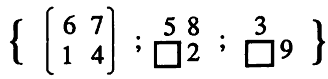
La preuve de cefait consiste à examiner les différents cas intervenant dans les algorithmes $\mathcal{A}$ et $\mathcal{A}^{-1}$ appliqués en parallèle aux suites $(\pi_k,\ P_k,\ Q_k)$ et $(A_{n-k},\ B_{n-k})$, et résulte essentiellement des propriétés de l'algorithme de Robinson-Schensted pour un tableau gauche.

Nous nous contentons de donner ici un exemple illustrant ce résultat, plutôt que d'effectuer cette énumération des différents cas à examiner.

Exemple 3.8

(3.8.1) Le triplet $(\pi,\ P,\ Q)$ obtenu par application de l'algorithme $\mathcal{A}$ au tableau $\tilde{T}$, ce tableau étant le miroir de celui considéré lors de l'exemple 3.5.

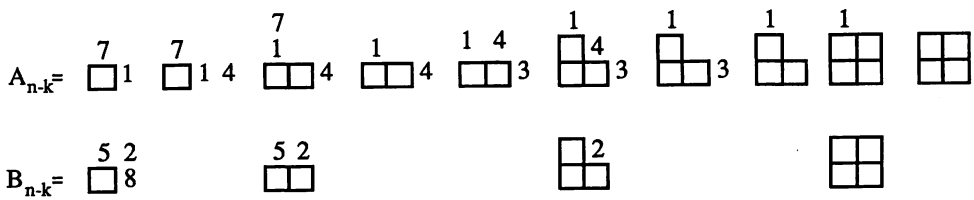


Ainsi $\tilde{T}$ ⟶  et, pour $0 \leq k \leq n=9$,

$P_k \in PT(\lambda_k / \mu_k)$ pour une certaine forme μ_k.

(3.8.2) L'algorithme $\mathcal{A}^{-1}$ appliqué au triplet $(\sigma, A, B) = (\pi^c, Q^c, P^c)$.

k= 0 1 2 3 4 5 6 7 8 9



$$\sigma = \begin{bmatrix} 6 & 9 \\ 3 & 4 \end{bmatrix}$$

On peut constater que $A_{n-k} \in PT(\lambda_k / \gamma_k)$ pour une certaine forme γ_k, et ceci pour tout k compris entre 0 et $n=9$. Le tableau ainsi obtenu est donc exactement le tableau miroir du tableau précédent.

4. La correspondance de Robinson-Schensted pour les tableaux oscillants gauches.

A partir de l'algorithme fondamental présenté précédemment, nous obtenons une correspondance pour les tableaux oscillants gauches analogue à celle de Robinson-Schensted.

Théorème 4.1 Soient α une partition et n un entier. Il existe une bijection

$$\Phi : PINV(2n) \times PT(\alpha/\mu) \times \overline{PT}(\alpha/\mu) \rightarrow TOG_n(\alpha \rightarrow \beta) \times TOG_n(\alpha \rightarrow \beta)$$

$$(\pi, P, Q) \mapsto (\tilde{P}, \tilde{Q})$$

où $\pi \dot{\cup} P \dot{\cup} Q = [2n]$.

Preuve. Elle est immédiate compte-tenu du théorème 3.1. En effet

$$(\pi, P, Q) \stackrel{\theta}{\longleftrightarrow} \tilde{T} = (\alpha = \lambda_0, \ldots\ldots, \lambda_n = \beta, \ldots\ldots, \lambda_{2n} = \alpha)$$

$$\updownarrow$$

$$\left\{\begin{array}{l} \tilde{P} = (\alpha = \lambda_0, \lambda_1, \ldots\ldots, \lambda_n = \beta) \\ \tilde{Q} = (\alpha = \lambda_{2n}, \lambda_{2n-1}, \ldots\ldots, \lambda_n = \beta) \end{array}\right\}.$$

Nous déduisons immédiatement de ce théorème l'identité suivante.

Corollaire 4.2 $$\sum_{\beta} \left(\tilde{f}_n^{\alpha\to\beta}\right)^2 = \sum_k \binom{2n}{k,k,2n-2k} (2n-2k)!! \sum_{\mu:\alpha/\mu \vdash k} \left(f^{\alpha/\mu}\right)^2 .$$

La bijection Φ du théorème 3.1 est l'analogue de la correspondance de Robinson-Schensted pour les tableaux standards. Elle permet, en la particularisant, de retrouver les identités connues pour les tableaux oscillants [DDF, Fav, Sun], les tableaux standards gauches [SaSt] et les tableaux standards [Sche].

Corollaire 4.3 (i) $\sum_{\beta} \left(\tilde{f}_n^{\varnothing\to\beta}\right)^2 = (2n)!!$,

(ii) $\sum_{\beta/\alpha \vdash n} \left(f^{\beta/\alpha}\right)^2 = \sum_k \binom{n}{k}^2 k! \sum_{\alpha/\mu \vdash n-k} \left(f^{\alpha/\mu}\right)^2$,

(iii) $\sum_{\beta \vdash n} \left(f^{\beta}\right)^2 = n!$.

Preuve.

(i) est une conséquence immédiate du corollaire 3.3 en prenant $\alpha=\varnothing$;

(ii) est une conséquence du théorème 3.1 ; il suffit pour cela de se restreindre aux tableaux "oscillants" $\tilde{P}$ et $\tilde{Q}$ pour lesquels aucune suppression de cellule n'a lieu (tableaux standards gauches). La correspondance fournit alors des triplets (π, P, Q) tels que

$\pi \in$ PINV($2n$),

P $\in$ PT(α/μ) avec $\breve{\pi} \mathbin{\dot\cup} \mathrm{P} = \{1, 2, \ldots\ldots, n\}$,

Q $\in \overline{\mathrm{PT}}$($\alpha/\mu$) avec $\hat{\pi} \mathbin{\dot\cup} \mathrm{Q} = \{n, n+1, \ldots\ldots, 2n\}$;

Ceci correspond à ne jamais employer la règle (2c) (réciproquement (2^{-1}c)) de l'algorithme $\mathcal{A}$ (réciproquement $\mathcal{A}^{-1}$).

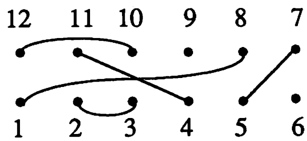
(iii) même raison que précédemment avec $\alpha=\emptyset$.

5. Propriété de la correspondance Φ

Une involution $\pi \in \mathrm{PINV}(2n)$ peut être représentée sous la forme d'un *graphe* $G(\pi)$ ayant pour ensemble de sommets les entiers de $[2n]$ disposés sur deux lignes parallèles, de sorte que les entiers i et $2n+1-i$ soient en correspondance, et pour arêtes les paires verticales constituant π (les cycles de l'involution).

Exemple 5.1

$$\pi = \begin{pmatrix} 3 & 7 & 8 & 11 & 12 \\ 2 & 5 & 1 & 4 & 10 \end{pmatrix} \longleftrightarrow \quad G(\pi) =$$

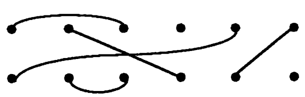
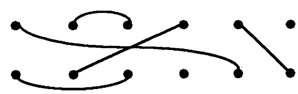


Définition 5.2

Etant donnée $\pi \in \mathrm{PINV}(2n)$, on définit son *inverse* $\pi^{-1} \in \mathrm{PINV}(2n)$ comme étant l'involution dont le graphe $G(\pi^{-1})$ est le graphe symétrique de $G(\pi)$ par rapport à un axe horizontal. Autrement dit, au cycle (i_p, j_p) de π correspond le cycle $(2n+1-j_p, 2n+1-i_p)$ dans π^{-1}, et ainsi $\pi^{-1} = \pi^c$ (définition 3.6).

L'inverse d'une involution partielle ainsi défini correspond exactement à l'inverse d'une permutation du groupe symétrique lorsque celle-ci est représentée de manière analogue sous forme d'un graphe.

Exemple 5.3 A l'involution π de l'exemple précédent correspond l'involution inverse π^{-1} dont le graphe est le suivant.

$G(\pi) =$  $\longrightarrow$ $G(\pi^{-1}) =$ 

Proposition 5.4 Soit $\mathrm{PINV_C}(n)$ l'ensemble des involutions partielles sur $[n]$

- dont les cycles de longueur 2 sont bi-coloriés,
- pouvant admettre des points fixes.

Il existe une bijection $\zeta : \mathrm{PINV_C}(n) \longrightarrow \{\pi \in \mathrm{PINV}(2n) : \pi = \pi^{-1}\}$.

La figure suivante suffit pour se convaincre de la validité de ce résultat.

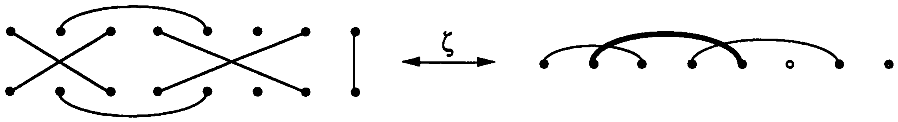


Figure 6

Théorème 5.5 Soient n un entier et α une partition.

Soit $(\pi, \mathrm{P}, \mathrm{Q})$ appartenant à $\mathrm{PINV}(2n) \times \mathrm{PT}(\alpha / \mu) \times \overline{\mathrm{PT}}(\alpha / \mu)$ tel que $\phi(\pi, \mathrm{P}, \mathrm{Q}) = (\tilde{\mathrm{P}}, \tilde{\mathrm{Q}})$.

Alors $\phi(\pi^{-1}, \mathrm{Q}^c, \mathrm{P}^c) = (\tilde{\mathrm{Q}}, \tilde{\mathrm{P}})$.

Ce résultat est une conséquence immédiate des théorèmes 3.7 et 4.1. Nous en déduisons l'identité suivante.

Corollaire 5.6 $$\sum_{\beta} \tilde{f}_n^{\alpha\to\beta} = \sum_k \binom{n}{k} 2^k \, I_{n-k} \sum_{\mu : \alpha/\mu \vdash k} f^{\alpha/\mu}$$

où I_p = coefficient de $\dfrac{x^p}{p!}$ dans e^{x+x^2}.

<u>Preuve</u>.
Soient σ appartenant à $\mathrm{PINV_C}(n)$ telle que $\zeta(\sigma)=\pi$, et P un tableau de $\mathrm{PT}(\alpha/\mu)$ tel que $\pi \,\dot{\cup}\, \mathrm{P} \,\dot{\cup}\, \mathrm{P}^c = [2n]$. Comme $\pi=\pi^{-1}$, nous avons

$$(\sigma, \mathrm{P}) \overset{\zeta}{\longleftrightarrow} (\pi, \mathrm{P}) \longleftrightarrow (\pi, \mathrm{P}, \mathrm{P}^c) = (\pi^{-1}, \mathrm{P}, \mathrm{P}^c) \overset{\phi}{\longleftrightarrow} (\tilde{\mathrm{P}}, \tilde{\mathrm{P}}) \longleftrightarrow \tilde{\mathrm{P}}.$$

On déduit alors immédiatement de ce fait la formule du corollaire.

Comme précédemment, en particularisant la correspondance Φ, nous retrouvons les identités suivantes pour les tableaux oscillants, les tableaux standards gauches et les tableaux standards.

Corollaire 5.7 (i) $\displaystyle\sum_{\beta} \tilde{f}_n^{\varnothing\to\beta} = I_n$ = coefficient de $\dfrac{x^n}{n!}$ dans e^{x+x^2},

$$= \sum_{k=0}^{\lfloor n/2 \rfloor} \frac{n!}{(n-2k)!\, k!}.$$

$$\text{(ii)} \sum_{\beta/\alpha \vdash n} f^{\beta/\alpha} = \sum_k \binom{n}{k} Inv(n-k) \sum_{\mu:\alpha/\mu \vdash k} f^{\alpha/\mu},$$

où $Inv(p)$ = coefficient de $\frac{x^p}{p!}$ dans $e^{x+\frac{x^2}{2}}$.

$$\text{(iii)} \sum_{\beta \vdash n} f^{\beta} = Inv(n).$$

Un raisonnement analogue à celui prouvant le corollaire 4.3 permet d'obtenir ces résultats, compte-tenu du théorème 5.5.

Remarque 5.8 Il est possible de mettre en évidence d'autres propriétés de cette correspondance pour les tableaux oscillants gauches, propriétés classiques dans le cas des tableaux standards. Ceci sera l'objet d'un prochain article.

Références bibliographiques

[DDF] M. Delest, S. Dulucq, L. Favreau, An analogue to Robinson-Schensted correspondence for oscillating tableaux, Actes du 20ième Séminaire Lotharingien de Combinatoire, L. Cerlienco et D. Foata éd., Alghero, Sardaigne, 1988.

[DF] S. Dulucq, L. Favreau, A Combinatorial Model for Bessel Polynomials, Proc. Third International Symposium on Orthogonal Polynomials and their applications, Erice, Sicily, C. Brezinski, L. Gori and A. Ronveaux ed., Lecture Notes in Math., Springer Verlag, 1990, 243-249.

[Fav] L. Favreau, "Combinatoire des tableaux oscillants et des polynômes de Bessel", Thèse, LaBRI Université Bordeaux I, 1991.

[Knu] D.E. Knuth, Permutations, matrices and generalized Young tableaux, *Pacific J. Math.* 34(1970), 709-727.

[McL] T.J. McLarnan, "Tableau recursions and symmetric Schensted Correspondences for ordinary, shifted and oscillating tableaux", Ph.D. Thesis, UCSD, 1986.

[Rob] G. de B. Robinson, On representations of the symmetric group, *Amer. J. Math.* 60(1934), 745-760.

[Sag] B.E. Sagan, Schifted tableaux, Schur Q-functions, and a conjecture of R. Stanley, *J. Comb. Theory (A)*, 45(1987), 62-103.

[SaSt] B.E. Sagan, R.P. Stanley, Robinson-Schensted algorithms for skew tableaux, *J. Comb. Theory (A)* 55(1990), 161-193.

[Sche] C. Schensted, Longest increasing and decreasing subsequences, *Canad. J. Math.* 13(1961), 179-191.

[Schü] M.P. Schützenberger, Quelques remarques sur une construction de Schensted, *Math. Scand.* 12(1963), 117-128.

[Sun] S. Sundaram, "On the combinatorics of representations of $Sp(2n, \mathbb{C})$", Ph.D. Thesis, MIT, 1986.

[Wor] D. Worley, "A theory of shifted Young tableaux", Ph.D. Thesis, MIT, 1984.

[You] A. Young, On quantitative substitutional analysis II, *Proc. London Math. Soc. (1)* 34(1902), 361-397.

Random palindromes: multivariate generating function and Bernoulli density

Ph. Dumas §
L. Thimonier §¶

Abstract

Consider a finite alphabet with a probability distribution $\mathbf{p}$. We study the probability $\delta(\mathbf{p})$ of obtaining a palindrome in a finite time by independant draws. Using a Mahler equation for an associated generating function, we give a closed form expression for $\delta(\mathbf{p})$. Moreover we describe completely the cases where $\delta(\mathbf{p})$ has value less than 1, in connection with the singularities of the generating function. Except for the case of a one or two letters alphabet it is found that $\delta(\mathbf{p})$ is always less than 1.

1 Introduction

Consider a finite alphabet $\mathcal{A}$ and a language $\mathcal{L}$ over $\mathcal{A}$. Fix a probability law $\mathbf{p}$ on $\mathcal{A}$ and choose randomly and indepently letters $w_1, w_2, \ldots, w_t, \ldots$ from $\mathcal{A}$. We look at the first time T when we obtain a word which belongs to $\mathcal{L}$. This time T is a finite integer or infinite. We denote $\delta(\mathcal{L}, \mathbf{p})$ the probability that T is finite. By definition [2, 7] the Bernoulli density of $\mathcal{L}$ is the function $\delta(\mathcal{L}, \mathbf{p})$ of $\mathbf{p}$.

It must be noticed that some classical random allocation problems like the birthday paradox (occurring in collisions in hashing methods) or the coupon collector problem arise in this context [3, p. 297]. There have been only few general attacks for these random words problems until recently, when the use of regular languages equipped with shuffle was proposed [7, 5].

Taking into account that the word obtained at time T has no proper left factor in $\mathcal{L}$, this notion of a density leads us to study the language $\mathcal{F}$ whose elements are the words from $\mathcal{L}$ but without proper prefix in $\mathcal{L}$. We call these words prefix free words and $\mathcal{F}$ the prefix free language associated with $\mathcal{L}$.

§Algorithms Project, INRIA Rocquencourt BP 105, 78153 Le Chesnay Cedex, France
¶Laboratoire de Mathématiques et d'Informatique Fondamentale, Université de Picardie, 80039 Amiens Cedex, France

Our basic tool is generating functions: with a language $\mathcal{G}$ we associate its multivariate generating functions

$$G(\mathbf{z}) := \sum_{\alpha\in\mathbb{N}^m} g_\alpha \mathbf{z}^\alpha.$$

In the expression, m is the number of letters in the alphabet $\mathcal{A} = \{a_1, \dots, a_m\}$, $\mathbf{z}$ is an m-tuple of indeterminates $(z_i)_{1\le i\le m}$, the indeterminate z_i marks the letter a_i and, for each multi-index α, g_α is the number of words in $\mathcal{G}$ for which the letter a_i occurs α_i times.

Frequently we want to take into account the length of words and we bring in an additional indeterminate u. Then the generating series is

$$G(u,\mathbf{z}) := G(u\mathbf{z}) = \sum_{\alpha\in\mathbb{N}^m} g_\alpha u^{|\alpha|}\mathbf{z}^\alpha.$$

We equip the algebras of formal power series $\mathbf{Q}[[\mathbf{z}]]$ and $\mathbf{Q}[[u,\mathbf{z}]]$ with the classical ultrametrics, which make them complete spaces.

When $\mathcal{F}$ is the prefix free language associated with $\mathcal{L}$, we have the equality with $F(u,\mathbf{z})$ the generating function of $\mathcal{F}$,

$$\delta(\mathcal{L},\mathbf{p}) = F(1,\mathbf{p})$$

since the word $a_{i_1}\cdots a_{i_T}$ appears with probability $p_{i_1}\cdots p_{i_T}$.

A palindrome is a word $w = w_1\cdots w_l$ of length l greater than or equal to 2, which is equal to its reversal $\tilde{w} = w_l\cdots w_1$. We now specialize $\mathcal{L}$ to be the language of palindromes and we look for a closed form expression and some properties of its density $\delta(\mathbf{p}) := \delta(\mathcal{L},\mathbf{p})$. To this end we use the language $\mathcal{F}$ of the palindromes without proper prefix in $\mathcal{L}$. So we have $\delta(\mathbf{p}) = F(1,\mathbf{p})$.

The case $m = 1$ is evident; there is only one prefix free palindrome, aa.

The case $m = 2$ is well known: $\mathcal{F}$ is a regular language, the prefix free palindromes over a two letters alphabet being of the form $abb\cdots bba$. The generating series is then rational:

$$F(u,\mathbf{z}) = \frac{u^2z_1^2}{1-uz_2} + \frac{u^2z_2^2}{1-uz_1}.$$

The case $m \ge 3$ has been studied by Beauquier and Thimonier but with univariate generating series. Hence they only deal with the uniform probability distribution case. In Part 2 we use a functional equation to derive a closed form expression of F as a series of rational fractions.

In the very simple case $m = 2$, the density is constant and equal to 1. Hence it is natural to ask if we may have $\delta(\mathbf{p}) = 1$ when $m \ge 3$. Part 3 is devoted to this question. The expression of $\delta(\mathbf{p})$ obtained as a by-product of part 2 is an alternating sum and we cannot estimate it directly. To achieve our goal we relate the density to the radius of convergence of the power series $F(u,\mathbf{p})$. The classification of the singularities of $F(u,\mathbf{p})$ according to the location of $\mathbf{p}$ gives us a very simple answer: we have $\delta(\mathbf{p}) = 1$ only when $m \le 2$.

2 Multivariate generating series

To obtain an expression of the generating series $\mathcal{F}$ of the language of the prefix free palindromes, we use a lemma from Beauquier and Thimonier[2].

Lemma 1 *If a palindrome has a prefix in $\mathcal{F}$ with length l, its length is greater than or equal to $2k-1$.*

The application of the lemma requires a few supplementary notations. First if k is an integer, we put

$$\mathbf{z}^k := (z_1^k, \ldots, z_m^k)$$

and

$$S_k := \sum_{i=1}^{m} z_i^k.$$

Next we introduce the language $\mathcal{S}$ of the symmetrical words: a word is symmetrical if it is equal to its reversal $\tilde{w} = w_l \cdots w_1$. Accordingly a palindrome is a symmetrical word of length at least 2. We introduce also for each $i = 1 \ldots m$, the language $\mathcal{L}_i$ of the palindromes starting with a_i and the language $\mathcal{F}_i$ of the prefix free palindromes starting with a_i. We obtain immediately

$$\begin{aligned} S(\mathbf{z}) &= \frac{1+S_1}{1-S_2}, \\ L_i(\mathbf{z}) &= z_i^2 S(\mathbf{z}), \\ L(\mathbf{z}) &= S_2 S(\mathbf{z}) \end{aligned}$$

and we are looking for

$$F(\mathbf{z}) = \sum_{i=1}^{m} F_i(\mathbf{z}).$$

The preceding lemma translates into a functional equation a la Mahler.

Lemma 2 *The generating series F_i of the language of the prefix free palindromes starting with a_i satisfies a Mahler equation,*

$$L_i(\mathbf{z}) = F_i(\mathbf{z}) + \left(\frac{1}{z_i} + S(\mathbf{z})\right) F_i(\mathbf{z}^2).$$

Proof : If w is a palindrome starting wit a_i: i) either it is prefix free, ii) or it has a unique proper left factor $v = a_i v' a_i$, which is prefix free and has length k. In the latter case: a) either w has length $2k-1$ and is written $v = a_i v' a_i v' a_i$, b) or w has length $2k+l$ ($l \geq 0$) and is written vuv, where u is a symmetrical word of length l. This immediately generates the equation above, by translating union and concatenation of languages (operating unambiguously) into sum and product of generating series. □

Proposition 1 *The generating series F_i has an explicit expression,*

$$F_i(\mathbf{z}) = \sum_{n\geq 0}(-1)^n z_i^{2^n+1}\frac{1+S_{2^n}}{1-S_{2^{n+1}}}\prod_{0\leq k<n}\left(1+z_i^{2^k}\frac{1+S_{2^k}}{1-S_{2^{k+1}}}\right).$$

Proof : Let $G_i(\mathbf{z})$ be the series defined by $F(\mathbf{z}) = z_i^2 G_i(\mathbf{z})$. This is licit because the letter a_i occurs at least twice in every word of $\mathcal{F}_i$. The series G_i satisfies a fixed point equation,

$$G_i(\mathbf{z}) = S(\mathbf{z}) - z_i G(\mathbf{z}^2)\,(1 + z_i S(\mathbf{z}))\,.$$

The mapping $H(\mathbf{z}) \mapsto S(\mathbf{z}) - z_i H(\mathbf{z}^2)(1 + z_i S(\mathbf{z}))$ is a contraction in the complete space $\mathbf{Q}[[\mathbf{z}]]$, whence the result by iteration. □

By rearranging the preceding expression and summing over i, we can prove the next theorem. It shows clearly the symmetrical aspect of the series F with respect to the indeterminates $z_1, \ldots, z_m$. Recall that the indeterminate u labels the length of the words. Moreover $\lg x$ is the logarithm base 2 of x. Last the support of an integer s, $\operatorname{supp} s$, is the set of the indexes for which the corresponding bit of its binary expansion equals 1. For example the support of $13 = 2^0 + 2^2 + 2^3$ is $\{0, 2, 3\}$.

Theorem 1 *The generating series F satisfies*

$$F(u,\mathbf{z}) = \sum_{s\geq 1}(-1)^{\lfloor \lg s\rfloor}u^{s+1}S_{s+1}\prod_{j\in\operatorname{supp} s}\frac{1+u^{2^j}S_{2^j}}{1-u^{2^{j+1}}S_{2^{j+1}}}.$$

This formula allows us to compute the first few terms of F,

$$F(u,\mathbf{z}) = S_2u^2 + (S_1S_2 - S_3)u^3 + (S_2^2 - S_4)u^4 + (S_1S_2^2 - S_1S_4 - S_2S_3 + S_5)u^5 + \cdots.$$

Proof : This result directly follows from the proposition. It suffices to apply the formulae

$$\prod_{0\leq j<n}\left(1 + a_j x^{2^j}\right) = \sum_{0\leq s<2^n} x^s \prod_{k\in\operatorname{supp} s} a_k,$$

$$\sum_{n\geq 0}(-1)^n a_n x^{2^n}\prod_{0\leq j<n}\left(1 + a_j x^{2^j}\right) = 1 + \sum_{s\geq 1}(-1)^{\lfloor \lg s\rfloor}x^s\prod_{k\in\operatorname{supp} s} a_k.$$

□

3 Radius of convergence, Bernoulli density

Until now series were considered as formal power series. Henceforth we think of them as analytic functions. We call $R(\mathbf{p})$ the radius of convergence of $F(u,\mathbf{p})$. Its value is greater than or equal to 1 since $F(u,\mathbf{p})$ is as a generating function in the sense of probability theory.

We use some open disks; we denote $\Delta(0,\rho)$ the open disk with center 0 and radius ρ.

The density of the language of palindromes is the function $\delta : \mathbf{p} \mapsto F(1, \mathbf{p})$. It is defined on the standard simplex of dimension $m-1$,

$$\mathrm{Simp}_{m-1} : p_i \geq 0 \ (i = 1 \dots m), \ \sum_{i=1}^{m} p_i = 1.$$

With a $\mathbf{p}$ from Simp_{m-1}, we associate $p_+ = \max_i p_i$, the Euclidean norm $\|\mathbf{p}\|$ of $\mathbf{p}$ and more generally the norm $\|\mathbf{p}\|_{2^k}$ defined by

$$\|\mathbf{p}\|_{2^k} = \left(\sum_{i=1}^{m} p_i^{2^k}\right)^{1/2^k}.$$

Theorem 2 *Let* $\mathbf{p}$ *be a point of* Simp_{m-1}.

1. *If all the* p_i *but one are zero (*$\mathbf{p}$ *is an extremal point of* Simp_{m-1}*), we have* $\delta(\mathbf{p}) = 1$ *and* $R(\mathbf{p}) = +\infty$.

2. *If only two* p_i *are non zero (*$\mathbf{p}$ *is on an edge of* Simp_{m-1} *but is not a vertex), we have* $\delta(\mathbf{p}) = 1$ *and* $R(\mathbf{p}) = 1/p_+$.

3. *In the other cases (*$\mathbf{p}$ *is not on an edge of* Simp_{m-1}*), we have* $\delta(\mathbf{p}) < 1$ *and* $R(\mathbf{p}) = 1/\|\mathbf{p}\|$. *Moreover* $F(u, \mathbf{p})$ *extends to the disk* $\Delta(0, 1/p_+)$ *as a meromorphic function, whose poles are all simple and equal to* $\omega/\|\mathbf{p}\|_{2^k}$, *with* k *a positive integer and* ω *a* 2^k*-th root of unity.*

The first two cases of the theorem have been seen in the introduction.

The proof follows from four lemmas. To avoid calculations with infinity, we suppose in the sequel that $\mathbf{p}$ is not an extremal point of the simplex.

Lemma 3 *The radius of convergence of* $F(u, \mathbf{p})$ *satisfies*

$$\frac{1}{\|\mathbf{p}\|} \leq R(\mathbf{p}) \leq \frac{1}{p_+}.$$

Proof : The prefix free palindromes are palindromes, therefore $L(u, \mathbf{p})$ is a majorizing series for $F(u, \mathbf{p})$. It follows that $R(\mathbf{p})$ is greater than or equal to the radius of convergence of

$$L(u, \mathbf{p}) = u^2 \|\mathbf{p}\|^2 \frac{1+u}{1-u^2\|\mathbf{p}\|^2};$$

that is to say $R(\mathbf{p}) \geq 1/\|\mathbf{p}\|$.

Likewise, among the prefix free palindromes, there are those in which there occur only two distinct letters. Their generating function is

$$\Phi(u, \mathbf{p}) = \sum_{i \neq j} \frac{u^2 p_i^2}{1 - u p_i} + \frac{u^2 p_j^2}{1 - u p_j}.$$

Hence we have a minorant series for $F(u, \mathbf{p})$.

We call H the set of indices i which satisfy $p_i = p_+$ and h the cardinality of H. Using the classical notation $[u^n]f(u)$ for the coefficient of u^n in the power series $f(u)$, we may write

$$[u^n]\Phi(u,\mathbf{p}) = \sum_{i \neq j} p_i^2 p_j^{n-2} \geq \sum_{i \neq j,\, j \in H} p_i^2 p_j^{n-2} = \sum_{j \in H} \left(\|\mathbf{p}\|^2 - p_+^2\right) p_+^{n-2}$$

and consequently

$$[u^n]F(u,\mathbf{p}) \geq [u^n]\Phi(u,\mathbf{p}) \geq h\left(\|\mathbf{p}\|^2 - p_+^2\right) p_+^{n-2}.$$

Thanks to Hadamard's formula regarding the radius of convergence of power series, we get $1/R(\mathbf{p}) \geq p_+$. □

We have to be more precise on the value of $R(\mathbf{p})$. To this goal we associate with $\mathbf{p}$ another point of the simplex, $\mathbf{p}' = \mathbf{p}^2/\|\mathbf{p}\|^2$.

Lemma 4 *The density at $\mathbf{p}'$ and the radius of convergence at $\mathbf{p}$ are related.*

1. *If $\delta(\mathbf{p}') = 1$ then $R(\mathbf{p}) = 1/p_+$.*
2. *If $\delta(\mathbf{p}') < 1$ then $R(\mathbf{p}) = 1/\|\mathbf{p}\|$ and $F(u,\mathbf{p})$ extends as a meromorphic function over the disk $\Delta(0, 1/p_+)$. The singularities of this function are the simple poles $\omega/\|\mathbf{p}\|_{2^k}$, with $k > 0$ and ω a 2^k-th root of unity.*

Proof : When $|u| \cdot \|\mathbf{p}\| < 1$ we may write Lemma 2 as follows,

$$F_i(u,\mathbf{p}) = -\frac{F_i(u^2,p^2)}{up_i} + \frac{1+u}{1-u^2\|\mathbf{p}\|^2}\left(u^2 p_i^2 - F_i(u^2,p^2)\right).$$

Summing from $i = 1$ to $i = m$, we obtain

$$F(u,\mathbf{p}) = -H(u,p) + \frac{1+u}{1-u^2\|\mathbf{p}\|^2}\left(u^2\|\mathbf{p}\|^2 - F(u^2,p^2)\right).$$

The functions $u \mapsto H(u,\mathbf{p})$ and $u \mapsto F(u^2,p^2)$ are analytic on the disk $\Delta(0, 1/\|\mathbf{p}\|_4)$, because $|u^2|\,\|\mathbf{p}^2\| < 1$ iff $|u|\,\|\mathbf{p}\|_4 < 1$. Accordingly there are only two possibilities.

In the first case the function of u

$$\frac{u^2\|\mathbf{p}\|^2 - F(u^2,p^2)}{1-u^2\|\mathbf{p}\|^2}$$

is analytic on the disk $\Delta(0, 1/\|\mathbf{p}\|_4)$, which means that

$$F\left(\frac{1}{\|\mathbf{p}\|^2}, \mathbf{p}^2\right) = F\left(1, \frac{\mathbf{p}^2}{\|\mathbf{p}\|^2}\right) = \delta(\mathbf{p}') = 1.$$

By recurrence on k, $F(u,\mathbf{p})$ is analytic on the disks $\Delta(0, 1/\|\mathbf{p}\|_{2^k})$ and then on the disk $\Delta(0, 1/p_+)$, since the sequence $(\|\mathbf{p}\|_{2^k})$ is strictly decreasing and converges to the limit p_+ [6,

p. 15, 26]. As a result the radius of convergence satisfies $R(\mathbf{p}) \geq 1/p_+$. From Lemma 3 we conclude $R(\mathbf{p}) = 1/p_+$ in the case under consideration.

In the second case the application is not analytic on $\Delta(0, 1/\|\mathbf{p}\|_4)$. We have $\delta(\mathbf{p}') < 1$ and $R(\mathbf{p}) = 1/\|\mathbf{p}\|$ according to Lemma 3. The function $F(u, \mathbf{p})$ extends to the disk $\Delta(0, 1/\|\mathbf{p}\|_4)$ and its extension has two simple poles, $\pm 1/\|\mathbf{p}\|$. By recurrence it extends to the disk $\Delta(0, 1/p_+)$ as described in the lemma. □

To conclude we need a lemma of independent interest. One may see it as a convexity-like property.

Lemma 5 *Let $\sum_\alpha c_\alpha \mathbf{z}^\alpha$ be a power series in m variables with non negative coefficients. Assume that the power series in one variable u,*

$$\sum_n u^n \sum_{|\alpha|=n} c_\alpha \mathbf{a}^\alpha,$$

has a radius of convergence $R(\mathbf{a}) > 0$ when $\mathbf{a}$ lies in a convex set A from ${\mathbf{R}_+}^m$. Then for $\mathbf{a}$ and $\mathbf{b}$ in A, we have the inequality

$$\frac{1}{R(\lambda\mathbf{a} + (1-\lambda)\mathbf{b})} \geq \begin{cases} \dfrac{1-\lambda}{R(\mathbf{b})} & \text{if } \lambda \in [0, 1/2[, \\ \dfrac{1}{2}\left(\dfrac{1/2}{R(\mathbf{a})} + \dfrac{1/2}{R(\mathbf{b})}\right) & \text{if } \lambda = 1/2, \\ \dfrac{\lambda}{R(\mathbf{a})} & \text{if } \lambda \in]1/2, 1]. \end{cases}$$

Proof : For $\lambda \in [0,1]$, α a multi-index and $\mathbf{a}$, $\mathbf{b} \in A$, we have, according to the binomial formula,

$$\begin{aligned} (\lambda\mathbf{a} + (1-\lambda)\mathbf{b})^\alpha &= \prod_{1 \leq i \leq m} (\lambda a_i + (1-\lambda) b_i)^{\alpha_i} \\ &\geq \prod_{1 \leq i \leq m} (\lambda^{\alpha_i} a_i^{\alpha_i} + (1-\lambda)^{\alpha_i} b_i^{\alpha_i}) \\ &\geq \lambda^{|\alpha|} \prod_{1 \leq i \leq m} a_i^{\alpha_i} + (1-\lambda)^{|\alpha|} \prod_{1 \leq i \leq m} b_i^{\alpha_i}. \end{aligned}$$

We collect these inequalities for all the α such that $|\alpha| = n$ and we use the concavity of the nth root function to obtain

$$\left(\sum_{|\alpha|=n} c_\alpha (\lambda\mathbf{a} + (1-\lambda)\mathbf{b})^\alpha\right)^{1/n} \geq (\lambda^n + (1-\lambda)^n)^{1/n} \times$$

$$\left(\frac{\lambda^n}{\lambda^n + (1-\lambda)^n}\left(\sum_{|\alpha|=n} c_\alpha \mathbf{a}^\alpha\right)^{1/n} + \frac{(1-\lambda)^n}{\lambda^n + (1-\lambda)^n}\left(\sum_{|\alpha|=n} c_\alpha \mathbf{b}^\alpha\right)^{1/n}\right).$$

It suffices to compute the upper limit in order to exhibit the formula. □

We arrive at the last step of our proof. In the simplex Simp_{m-1} we consider a radius from the center of gravity $\mathbf{eq}_m = (1/m, \ldots, 1/m)$, which corresponds to the equally likely case. This radius is parametrically defined by $\mathbf{p} = \mathbf{eq}_m + t\mathbf{q}$, $t \in [0, t_+]$. Here $\mathbf{q}$ is a vector which satisfies $\sum_i q_i = 0$, $\|\mathbf{q}\| = 1$. The number t_+ has value $-1/(mq_-)$, where $q_- = \min_i q_i$ to ensure that the point $\mathbf{p} = \mathbf{eq}_m + t\mathbf{q}$ lies on the simplex.

Lemma 6 *If $m \geq 3$ the radius of convergence satisfies*

$$R(\mathbf{p}) = 1/\|\mathbf{p}\|$$

for arbitrary $\mathbf{p} \in [\mathbf{eq}_m, \mathbf{eq}_m + t_+\mathbf{q}[$.

Proof : By restriction to the segment $[\mathbf{eq}_m, \mathbf{eq}_m + t_+\mathbf{q}]$, the three functions of the point $\mathbf{p}$, namely $1/R(\mathbf{p})$, $\|\mathbf{p}\|$, and p_+, gives us three functions of t. We call them respectively $f(t)$, $g(t)$ and $h(t)$. We know that $f(t)$ can take only two values: $g(t)$ or $h(t)$. Our goal is to prove that $f(t) = g(t)$ for all $t \in [0, t_+[$.

We introduce the set $C = \{t \in [0, t_+[,\ f(t) = g(t)\}$. We will verify that 0 lies in C and C is both open and closed. As $[0, t_+[$ is connex, it will yield $C = [0, t_+[$, as desired.

First $0 \in C$. Indeed we use one more time the fact that prefix free palindromes are palindromes, but some palindromes are not prefix free. Hence $\delta(\mathbf{eq}_m) = F(1, \mathbf{eq}_m) < L(1, \mathbf{eq}_m)$, that is to say

$$\delta(\mathbf{eq}_m) < \frac{2}{m-1} \leq 1.$$

The last inequality follows from the assumption $m \geq 3$. According to Lemma 4, we have $R(\mathbf{eq}_m) = 1/\|\mathbf{eq}_m\|$ since $\mathbf{eq}'_m = \mathbf{eq}_m$. Consequently $f(0) = g(0)$.

Secondly C is open. Let $t_0 \neq 0$ be a point from C. (The treatment of 0 is similar.) According to Lemma 5, we have the inequalities

$$\begin{aligned} f(t) &\geq \frac{t}{t_0} g(t_0) && \text{if } t \in]t_0/2, t_0[, \\ f(t) &\geq \frac{t_+ - t}{t_+ - t_0} g(t_0) && \text{if } t \in]t_0, (t_0 + t_+)/2[. \end{aligned}$$

This permits us to define a continuous piecewise linear function $\varphi(t)$ wich is a lower bound for $f(t)$ on the segment $[t_0/2, (t_+ - t_0)/2]$. The difference $\varphi(t) - h(t)$ is continuous and takes a positive value at t_0. By continuity it takes positive values on an open interval I around t_0. For t in I, we have $f(t) \geq \varphi(t) > h(t)$ therefore $f(t) = g(t)$. Hence I is a subset of C and C is open.

Finally C is closed. Let t_1 be a point adherent to C. We use an integer $N \geq 2$, which will be made precise later. Let ϵ be a positive number that satisfies

$$\epsilon < \min\left(\frac{t_1}{N}, \frac{t_+ - t_1}{N}, \frac{g(t_1)^2}{2N}\right).$$

We choose a point t_0 from C such that $|t_1 - t_0| < \epsilon$. In the first place we prove that

$$f(t_1) > \frac{N}{N+1} g(t_0).$$

i) If $t_1 \leq t_0$, we have $t_1 \in]t_0/2, t_0]$ since $|t_1 - t_0| < \epsilon$ and $\epsilon < t_1/2 \leq t_0/2$. From Lemma 5 we may write $f(t_1) \geq (t_1/t_0)g(t_0)$, but $t_0 - t_1 < \epsilon < t_1/N$ hence $t_1/t_0 > N/(N+1)$. This gives us $f(t_1) > N/(N+1)g(t_0)$. ii) If $t_0 < t_1$, we successively derive by the same method $t_1 \in [t_0, (t_0 + t_+)/2[$, $f(t_1) \geq (t_+ - t_1)/(t_+ - t_0)\, g(t_0)$ and again $f(t_1) > N/(N+1)g(t_0)$. Thereafter we choose N large enough to ensure $f(t_1) > h(t_1)$. Using the fact that $g(t)$ is the Euclidean norm, we have

$$\frac{g(t_0)}{g(t)} = 1 + \frac{t_0 - t}{g(t)^2} + O((t_0 - t)^2).$$

So we can impose

$$\frac{g(t_0)}{g(t_1)} > 1 - \frac{1}{N+1}$$

if N is great enough, since $|t_0 - t_1| < \epsilon < g(t_1)^2/(2N)$. Under the assumption we get

$$g(t_0) > \frac{N}{N+1} g(t_1)$$

and further

$$f(t_1) > \left(\frac{N}{N+1}\right)^2 g(t_1).$$

This relation implies $f(t_1) > h(t_1)$ if N is large enough. So $f(t_1) = g(t_1)$, t_1 is a member of C and C is closed. □

The last lemma gives us the theorem. Indeed it shows that $R(\mathbf{p}) = 1/\|\mathbf{p}\|$ for all the points $\mathbf{p}$ which lie on the interior of a d-dimensional face of the simplex for $d = 2 \ldots m-1$, hence for all the points $\mathbf{p}$ which are not on an edge of the simplex.

4 Conclusion

Except in case m equals 2, the function $F(u, \mathbf{p})$ is transcendental because it has infinitely many singularities. As a result the language $\mathcal{F}$ cannot be an unambiguous context-free language [4]. In fact it is not even context-free [1].

The singularities are poles which aggregate on the circle $C(0, 1/\|\mathbf{p}\|)$. This phenomenon is illustrated by Figure 1 in the case $\mathbf{p} = \mathbf{eq}_3$. It is typical of the solutions of a Mahler equation in the simplest case when the function is not rational. The idea we used may be employed for a univariate function when the equation gives poles with modulus less than 1. It must be noticed that the asymptotic behaviour of the coefficients is immediate in this case. For example the probability that the waiting time T has value n is asymptotically

$$(1 - \delta(\mathbf{p}'))\, \|\mathbf{p}\|^{2\lfloor n/2 \rfloor}$$

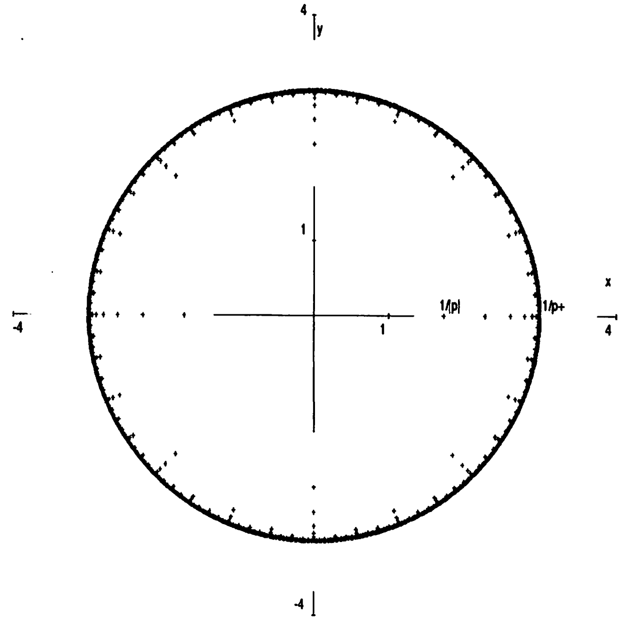


Figure 1: The poles of $F(u, \mathbf{eq}_3)$ agglomerate on the circle of radius 3.

and in particular the number of prefix-free palindromes of length n is equivalent to

$$(1 - \delta(\mathbf{eq}_m))m^{2\lfloor n/2 \rfloor}.$$

These formulae can be used to compute $\delta(\mathbf{eq}_m)$ with the help of the recurrence corresponding to the Mahler equation,

$$u_n = m\,u_{n-2} - u_{\lceil n/2 \rceil} \quad (n \geq 4).$$

For the first few values of m we obtain the following values of $\delta(\mathbf{eq}_m)$ which represents the probability to obtain a palindrome in a finite time with a uniform probability distribution over a m letters alphabet.

m	$\delta(\mathbf{eq}_m)$
2	1.0000000000
3	0.7215100801
4	0.5415013253
5	0.4299516130
6	0.3555383290
7	0.3027139845
8	0.2633895810
9	0.2330230426
10	0.2088879991

We can also get $\delta(\mathbf{p})$ with a great accuracy since the expansion of F that we obtained converges very rapidly. For example we can compute $\delta(\mathbf{p})$ in the case $\mathbf{p} = (1/2, 1/4, 1/4)$

using the approximation,

$$\delta(\mathbf{p}) \simeq \sum_{i=1}^{m}\sum_{n=0}^{N}(-1)^n p_i^{2^n+1}\frac{1+S_{2^n}}{1-S_{2^{n+1}}}\prod_{0\le k<n}\left(1+p_i^{2^k}\frac{1+S_{2^k}}{1-S_{2^{k+1}}}\right).$$

Taking only $N=7$, we get a result which is exact to 38 digits,

$$\delta(1/2,1/4,1/4) \simeq 0.75241112528971363575197933398903183045.$$

The path we took to prove Theorem 2 may seem rather indirect; however δ takes values arbitrary near 1 and the elementary methods, which we have employed, obliged us to some tricks. As an extension of ours results, we can prove that almost surely the process gives palindromes only a finite number of times if $m \geq 2$, thanks to the Borel-Cantelli lemma.

We wish to acknowledge the contribution of Philippe Flajolet in offering valuable suggestions and stimulating comments during the course of this work.

References

[1] J. Beauquier. Prefix and perfect languages. In G. Ausiello and M. Protasi, editors, *CAAP'83 Trees in Algebra and Programming L'aquila, March 1983*, pages 129–140. Springer-Verlag, 1983. Lecture Notes in Computer Sciences 159.

[2] J. Beauquier and L. Thimonier. Prefix-free words of length n over m letters : two-sided well-balanced parentheses and palindromes. In P. Leroux G. Labelle, editor, *Colloque de combinatoire énumérative 1985*, volume 1234 of *Lecture Notes in Mathematics*, pages 27–33 and 385. UQAM, Springer-Verlag, 1986.

[3] L. Comtet. *Advanced Combinatorics.* Reidel, Dordrecht, 1974.

[4] P. Flajolet. Analytic models and ambiguity of context–free languages. *Theoretical Computer Science*, 49:283–309, 1987.

[5] P. Flajolet, D. Gardy, and L. Thimonier. Probabilistic languages and random allocations. In Timo Lepistö and Arto Salomaa, editors, *Automata, Languages and Programming*, volume 317 of *Lecture Notes in Computer Science*, pages 239–253. Springer Verlag, 1988. Proceedings of 15th ICALP Colloquium, Tempere, Finland, July 1988.

[6] G. H. Hardy, J. E. Littlewood, and G. Pólya. *Inequalities.* Cambridge University Press, 2 edition, 1952.

[7] L. Thimonier. *Fonctions génératrices et mots aléatoires.* Doctorat ès sciences, Université de Paris-Sud Orsay, 1988.

The numbers game and Coxeter groups

KIMMO ERIKSSON
Department of Mathematics
Royal Institute of Technology
S-100 44 Stockholm, Sweden
kimmo@nada.kth.se

1 Introduction

In 1988, Professor Anders Björner in Stockholm gave to me, his brand new student, the following problem to work on: What can be said about the numbers game? Three years later new answers to the question are still popping up, as the subject tends to get both deeper and wider the more is extracted from it. I will here mention some basic properties of the game and some connections with Coxeter groups.

The numbers game is a one-player game played on a graph. It was defined by Shahar Mozes [7] as follows. Let G be a simple graph with N nodes. Place a real number on each node. A move now consists of first picking a node i with a negative number p_i, then adding the number p_i to the number at each neighbor j of i, and finally reversing the sign at node i. If no move is possible the game has terminated.

Mozes showed that the process is what I will call *strongly convergent*: Given a starting position either every play sequence diverges (*i.e.* can be continued forever), or every play sequence will converge to the same terminal position in the same number of moves. In other words, the length and result of the game is independent of what choices are made.

Possible questions to ask about the numbers game are for example:

- Are there any natural generalizations that preserve the strong convergence property?
- From which positions does the game converge, and in how many steps?
- Given two positions p and q, is it decidable whether q can be reached by playing from p?

These questions all have sort of a "global" aspect on the game; we need to know every possible way to play from p, and every position reached in this way. Therefore it is natural to study the *game graph* obtained by taking all positions reachable from p as vertices, and taking the moves as directed edges, each labeled with the node fired. The sequences of legal moves from p define a language $\mathcal{L}_p$ in the alphabet of nodes. But now some additional questions arise.

- Is the language $\mathcal{L}_p$ a greedoid? This was known for a related game.
- Is the game graph, viewed as the partially ordered set of positions, a lattice?

In order to get some algebraic tools for dealing with such questions, it is natural to represent a position on a graph with d nodes as a point $p = (p_1, p_2, \ldots, p_d)$ in real d-dimensional space, where the coordinates are given by the numbers on the nodes. A move is then seen to be equivalent to a reflection of the point in some hyperplane [4] [5] [7]. This is the background to why Coxeter groups, which have a standard representation as groups generated by reflections, are interesting in this context.

2 The polygon idea

Björner had observed the following relations in the numbers game. Suppose two nodes x and y are both playable. If x and y are not neighbors, then playing xy or yx leads to equal positions. If x and y are neighbors, then xyx and yxy are legal play sequences leading to equal positions. Thus we get polygons in the game graph.

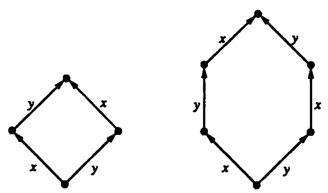


Figure 1: Polygons: $xy = yx$ (left) and $xyx = yxy$ (right)

These polygons imply strong convergence by the following theorem, Eriksson 1991 [4].

Theorem 2.1 *A game has the strong convergence property if and only if whenever two different moves are possible, they define a polygon, that is, two legal play sequences of equal length (possibly infinite) which end (if finite) in the same position.*

Let $(xyx...)_l$ denote the alternating play sequence of length l.

3 Generalizations by weighted graphs

The first generalization we shall make is the introduction of *weights* on the *edges* of the graph. Specifically, to each edge (i, j) we assign two strictly positive weights, k_{ij} and k_{ji}, such that if p_i denotes the number on node i, then the new position after firing node i is computed from the previous position by

$$\begin{cases} p_j := p_j - k_{ij}p_i & \text{for each vertex } j \neq i; \\ p_i := -p_i & \text{(as in the unweighted game).} \end{cases}$$

By geometric considerations one can show that if there are integers $m(i,j) > 2$ for every edge (i,j) such that the weight product $k_{ji}k_{ij} = 4\cos^2(\pi/m(i,j))$ then we get alternating polygons: $(xyx...)_{m(x,y)} = (yxy...)_{m(x,y)}$ which are legal if both x and y are playable. If $k_{ji}k_{ij} \geq 4$, then we get an infinite polygon, *i.e.* both sequences $(xyx...)$ and $(yxy...)$ are playable forever. Other weight products than these do not give polygons, hence it follows that the edge weighted numbers game is strongly convergent if and only if every weight product satisfies one of the conditions above.

Next, it is natural to explore what happens when we define a *node weight*, $w_i > 0$, for each node i, such that firing node i affects the number at the same node by $p_i := -w_i p_i$. In this case things get more complicated, but a thorough analysis along the same lines as in the edge weighted case gives:

Theorem 3.1 *A node weighted game is strongly convergent if and only if for each edge* (i,j) *the corresponding weight product satisfies either* $k_{ji}k_{ij} \geq 2\sqrt{w_i w_j} + w_i + w_j$, *or* $k_{ji}k_{ij} = 2\sqrt{w_i w_j}\cos(2\pi/n) + w_i + w_j$ *for some integer* $n \geq 3$, *and* $w_i = w_j$ *if* n *is odd.*

Note that if all node weights are equal to 1, then we get back to the conditions for the purely edge weighted case. From now on, by an *N-game* I will mean a strongly convergent, edge weighted numbers game.

4 Coxeter groups

Let V be a finite index set. A Coxeter group (W,S) is a group W with a distinguished set of involutory generators, $S = \{s_x : x \in V\}$, and relations $(s_x s_y)^{m(x,y)} = e$ (the identity of the group), with integers $m(x,y) \geq 2$, or $m(x,y) = \infty$, by which we mean that $s_x s_y$ has infinite order in W.

Now think about the moves of an N-game as linear transformations of the position vector. These transformations are clearly involutions, and by the polygon property $(xyx...)_{m(x,y)} = (yxy...)_{m(x,y)}$ we have the Coxeter type of relations. Indeed, the moves generate a group of linear transformations that is isomorphic to the corresponding Coxeter group defined by the $m(x,y)$, as is demonstrated by the following equivalence between the numbers game and the standard representation of a Coxeter group as acting on an $|S|$-dimensional space X with some basis $\{e_x : x \in V\}$ and geometry given by a bilinear form $B(e_x, e_y) = -\cos(\pi/m(x,y))$, and generators $S = \{\sigma_x : x \in V\}$ where every σ_x is a reflection in the hyperplane perpendicular to unit vector e_x, [5]. Hence, for $\lambda \in V$, $\sigma_x(\lambda) = \lambda - 2B(\lambda, e_x)e_x$.

Given a Coxeter group (W,S), construct a graph G with $|S|$ nodes on which to play the numbers game in the following way. For any pair x, y of nodes, put an undirected edge (x,y) in G if the relation exponent $m(x,y) > 2$. Let the edge weights be $k_{xy} = k_{yx} = 2\cos(\pi/m(x,y))$. Then the numbers game on G is an N-game, by Theorem 3.1.

To every vector $\lambda \in V$ we now associate the position in the game where the number written on a node x is $\lambda_x \stackrel{\text{def}}{=} B(\lambda, e_x)$. A reflection σ_x gives a new vector $\sigma_x(\lambda)$, in which the numbers of the associated game position are given by

$$[\sigma_x(\lambda)]_y = \begin{cases} -\lambda_x & \text{if } x = y \ ; \\ \lambda_y + 2\cos(\pi/m(x,y))\lambda_x & \text{if } x \neq y. \end{cases}$$

Hence a hyperplane reflection σ_x from the negative to the positive halfspace is equivalent to the firing of the corresponding node x in the numbers game.

Indeed, if Γ_p is the game graph from a position p where all numbers are negative, then Γ_p is isomorphic to the Cayley graph of the corresponding Coxeter group.

5 Length of game

Now let W be the Coxeter group of linear transformations generated by the set of moves, which we denote S, in the N-game played on $G = (V, E)$. Let $p = (p_1, p_2, \ldots)$ be a position in the game. Let α_i be the functional on positions that returns the number at node i, i.e. $\alpha_i(p) = p_i$. Thus the multiset $\{\alpha_i(w(p)) : w \in W, i \in V\}$ contains all numbers that can ever arise in any node when playing the game from p (backwards and forwards). This can be interpreted as the values obtained by applying the set of functionals $\Phi = \{\alpha_i(w(\cdot)) : w \in W, i \in V\}$ on the start position p. One can see that Φ is (isomorphic to) a *root system*, see [6], of the Coxeter group, where the α_i are the primitive roots of Φ.

It is known that the roots can be partitioned in $\Phi = \Phi^+ \cup \Phi^-$, where every $\phi \in \Phi^+$ is a nonnegative linear combination of the α_i, and every $\phi \in \Phi^-$ is a nonpositive linear combination of the α_i. Further, if $s_i \in S$ is the move of firing node i, then an important property of root systems is that $\Phi^+ \circ s_i = (\Phi^+ - \{\alpha_i\}) \cup \{-\alpha_i\}$. If i is playable, then $p_i = \alpha_i(p)$ is negative, while $-\alpha_i(p)$ is positive. Hence, $\Phi^+ \circ s_i(p)$ has exactly one negative value less than $\Phi^+(p)$. In a terminal position t, $\Phi^+(t)$ has only positive values. Consequently, the length of a game from p is equal to the number of negative values in $\Phi^+(p)$, if finite. If this number is infinite, then the game from p is divergent.

By the above, we can also state a comparison test.

Theorem 5.1 *If $\bar{p} \leq p$ (that is, $\bar{p}_i \leq p_i$ for every node i) then every play sequence that is legal from p is also legal from $\bar{p}$.*

This is immediately clear from the fact that the linear combinations in Φ^+ have nonnegative coefficients.

6 The language of play sequences is a greedoid

Definition. A language $\mathcal{L}$ is a greedoid if it is left-hereditary, which means that

$$\alpha\gamma \in \mathcal{L} \Rightarrow \alpha \in \mathcal{L} \qquad \text{(G1)}$$

and $\mathcal{L}$ satisfies the following exchange condition.

$$\alpha, \beta \in \mathcal{L}, |\beta| > |\alpha| \Rightarrow \exists x \in \beta : \alpha x \in \mathcal{L} \qquad \text{(G2)}$$

It is known that in another vertex-firing game, the chips game of Björner, Lovász and Shor [2], the language of legal play sequences is a greedoid. We now prove this for any N-game.

Lemma 6.1 *Let $G = (V, E)$ be an edge weighted graph that defines an N-game.*

(a) *Given a subset $V' \subseteq V$, let G' be the subgraph of G induced by the nodes in V'. Then The numbers game on G' is strongly convergent, that is, an N-game..*

(b) *If x is a legal move and α is a legal play sequence in a position p on G, such that $x \notin \alpha$, then $x\alpha$ is a legal play sequence from p.*

PROOF. (a) This is obvious from the characterization of N-games in Section 3, since the weight product of any edge in G' is the same as for the corresponding edge in G which defines an N-game.

(b) Let $V' = V - \{x\}$. Let $\bar{p}$ be the position after playing x. Then α is playable from $p|_{G'}$ (the restriction to G'), and $\bar{p}|_{G'} \leq p|_{G'}$, so by Theorem 5.1 α is playable from $\bar{p}|_{G'}$. Accordingly, $x\alpha$ is playable from p. □

In the following, fix $G = (V, E)$, and fix a subset $V' \subseteq V$ with induced subgraph G'. Let $d(\alpha)$ denote the number of moves in play sequence α which are firings of nodes that are not in V'.

Lemma 6.2 *Let p be a position on G where β and α are legal play sequences such that $d(\beta) = 0$ and $d(\alpha) \geq |\alpha| - |\beta|$. Then there exists a play sequence γ such that $d(\gamma) = 0$, $|\gamma| = |\beta| - |\alpha| + d(\alpha)$ and $\alpha\gamma$ is legal from p.*

PROOF. If $d(\alpha) = 0$ we can restrict the game to G' and strong convergence on G' implies the existence of the desired γ. Suppose the theorem has been proved when $d(\alpha) < n$, and suppose this $d(\alpha) = n$. Then α can be written $\alpha_1 y \alpha_2$ where $y \notin V'$, $d(\alpha_1) = n - 1$ and $d(\alpha_2) = 0$. By hypothesis there exists some γ_1 such that $d(\gamma_1) = 0$, $|\gamma_1| = |\beta| - |\alpha_1| + n - 1$ and $\alpha_1\gamma_1$ is legal from p. By Lemma 6.1(b), $\alpha_1 y \gamma_1$ is legal from p. Let q be the position after $\alpha_1 y$. In q we have γ_1 and α_2 legal, $d(\gamma_1) = 0$ and $d(\alpha_2) = 0 \geq |\alpha_2| - |\gamma_1|$ by the above. Thus there exists a γ with $d(\gamma) = 0$, such that $\alpha_1 y \alpha_2 \gamma = \alpha\gamma$ is legal from p, and $|\gamma| = |\gamma_1| - |\alpha_2| = |\beta| - |\alpha| + d(\alpha)$. □

Theorem 6.3 *$\mathcal{L}_p$, the language of legal play sequences from position p in an N-game, is a greedoid.*

PROOF. Obviously, $\mathcal{L}_p$ is left-hereditary, so we only have to verify (G2), the greedoid exchange property. Suppose α and β are legal play sequences with $|\beta| > |\alpha|$. Let V' be the set of vertices fired in β. Then $d(\beta) = 0$ and $d(\alpha) \geq |\alpha| - |\beta|$, so by Lemma 6.2 there is a play sequence γ of length at least 1 and firing only nodes in V', such that $\alpha\gamma$ is legal. Let x be the first move of γ. Then $x \in \beta$ and αx is legal. □

7 Final words

Much more can be said about the numbers game than was possible here. I refer the interested reader to my doctoral thesis, due in spring 1993.

References

[1] A. Björner, On a combinatorial game of S. Mozes, preprint, 1988.

[2] A. Björner, L. Lovasz and P. Shor, Chip-firing games on graphs, *Eur. J. Comb.* **12** (1991), 283–291.

[3] K. Eriksson, Convergence of Mozes's game of numbers, 1990, to appear in *Linear Algebra Appl.*

[4] K. Eriksson, Strong convergence and a game of numbers, 1991, preprint.

[5] K. Eriksson, Games and the weak order of Coxeter groups, 1992, in preparation.

[6] J. Humphreys, *Reflection groups and Coxeter groups*, Cambridge University Press, Cambridge, 1990.

[7] S. Mozes, Reflection processes on graphs and Weyl groups, *J. Combin. Theory*, Ser. A, **53** (1990), 128–142.

FONCTIONS DE BESSEL, EMPILEMENTS ET TRESSES

J.M. Fedou, LaBRI*
Département d'Informatique
351, Cours de la Libération
33405 Talence, Cedex France

Abstract. In this paper, we show that the inverse of the basic Bessel function ${}^qJ_0$ and the quotient of the basic Bessel functions, ${}^qJ_1$ and ${}^qJ_0$ are the generating functions of some heaps of sticks. Moreover, we show that the numerators of the rational functions appearing in the development of these functions are the invariant polynomials of some kind of braids.

Résumé. Nous montrons dans cet article que l'inverse du q-analogue de la fonction de Bessel ${}^qJ_0$ ainsi que le rapport des q-analogues des fonctions de Bessel ${}^qJ_1$ and ${}^qJ_0$ sont les fonctions génératrices de certains empilements de demi-segments. Nous montrons que les numérateurs des fonctions rationnelles apparaissant dans le développement de ces quotients sont des polynômes invariants de certains éléments du groupe des tresses.

1. introduction

Récemment sont intervenus les q-analogues des fonctions de Bessel dans des problèmes d'énumération de polyominos, par exemple les polyominos parallélogrammes ou diagrammes de Ferrers gauches (M.P. Delest et J.M. Fedou [9]) et les polyominos convexes dirigés (M. Bousquet-Melou [5]) selon les paramètres aire et nombre de colonnes.

ENUMERATION DE DIAGRAMMES DE FERRERS GAUCHES ET FONCTIONS DE BESSEL.

Un diagramme de Ferrers est la figure obtenue à partir d'une partition $n_1 \geq n_2 \geq \ldots \geq n_k$ d'un entier $n = n_1 + n_2 + \ldots + n_k$ en juxtaposant des rectangles de hauteur 1 et de largeur n_i. Un diagramme gauche de Ferrers est la différence de deux diagrammes de Ferrers (cf. figure 1).

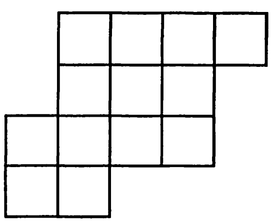


Figure 1. Diagramme de Ferrers gauche d'aire 13 ayant 5 colonnes.

* Unité associée au Centre National de la Recherche Scientifique n° 1304.

This work is partially supported by the PRC Mathématique et Informatique

Ces objets apparaissent également dans la littérature sous le nom de polyominos parallélogrammes . M.P Delest et J.M.Fedou [9] ont montré que la série génératrice F(t;q) des diagrammes de Ferrers gauches selon les paramètres nombre de colonnes et aire s'exprime aisément comme rapport de q-analogue des fonctions de Bessel J_1 et J_0,

$$F(t;q) = \frac{\displaystyle\sum_{n=0}^{+\infty} \frac{(-1)^n q^{\binom{n+1}{2}}}{(q;q)_n (q;q)_{n+1}} q^{n+1} t^{n+1}}{\displaystyle\sum_{n=0}^{+\infty} \frac{(-1)^n q^{\binom{n}{2}}}{(q;q)_n^2} q^n t^n},$$

où $(a;q)_n = (1-a)(1-aq)(1-aq^2)\ldots(1-aq^{n-1})$. On remarque en effet que F(t;q) est à un changement de variables près un q-analogue du quotient $\frac{J_1}{J_0}$.

Rappelons que les fonctions de Bessel J_ν [10] sont définies pour $\nu>-1$, par

$$J_\nu(x) = \sum_{n=0}^{\infty} \frac{(-1)^n (\frac{1}{2}x)^{2n+\nu}}{n!\,\Gamma(\nu+n+1)},$$

et définissons, pour ν entier, le q-analogue de J_ν par

$${}^qJ_\nu(x) = \sum_{n=0}^{+\infty} \frac{(-1)^n q^{\binom{n+\nu}{2}} x^{n+\nu}}{(q;q)_n (q;q)_{n+\nu}}.$$

On a alors $F(t;q) = \frac{{}^qJ_1}{{}^qJ_0}(qx)$.

ANCIENNES FORMULES, NOUVEAUX q-ANALOGUES

Ces problèmes ont fait resurgir d'anciennes formules concernant les fonctions de Bessel. Les développements des fonctions $\frac{1}{J_0}$, $\frac{J_1}{J_0}$ ont été étudiés par Carlitz [6], Lehmer [14], Watson [16] et beaucoup d'autres et il semble que la plupart de ces résultats admettent des q-analogues agréables. La figure 2 donne les premiers termes des développements classiques ainsi que leurs q-analogues. En particulier, les numérateurs intervenant dans tous ces développements sont à coefficients entiers positifs, symétriques et semblent unimodaux.

n	a_n	b_n	$a_n(q)$	$b_n(q)$
1	1	1	1	1
2	3	1	$1+q+q^2$	1
3	19	4	$1+2q+4q^2+5q^3+4q^4+2q^5+q^6$	$1+q+q^2+q^3$
4	211	33	$1+3q+8q^2+16q^3+26q^4+33q^5+37q^6+33q^7+26q^8+16q^9+8q^{10}+3q^{11}+q^{12}$	$1+2q+4q^2+6q^3+7q^4+6q^5+4q^6+2q^7+q^8$

Figure 2. premières valeurs des polynômes $a_n(q)$ et $b_n(q)$.

notations. On note,

$$\frac{1}{{}^{q}J_0(x)}=\sum_{n\geq 0}\frac{a_n(q)}{(q;q)_n^2}\,x^n,$$

$$\frac{{}^{q}J_1(x)}{{}^{q}J_0(x)}=\sum_{n\geq 1}\frac{b_n(q)}{(q;q)_n(q;q)_{n-1}}\,x^n,$$

où $(q;q)_n=(1-q)(1-q^2)\ldots(1-q^n)$.

Nous donnons dans le paragraphe 2 de cet article une interprétation combinatoire de $J_0, \frac{1}{J_0}, \frac{J_1}{J_0}$ en terme d'empilements de demi-segments et montrons au paragraphe 3 que les polynômes $a_n(q)$ et $b_n(q)$ s'expriment très simplement en terme de tresses valuées.

FONCTIONS GENERATRICES, EMPILEMENTS ET PYRAMIDES

La notion d'empilement de pièces, due à Viennot [15], est une version géométrique de la théorie des monoïdes partiellement commutatifs, introduite en 1969 par Cartier et Foata [8]. Soit un ensemble $\mathcal{S}$ muni d'une relation binaire $\mathcal{R}$ réflexive et symétrique, dans lequel deux éléments en relation sont dits *en concurrence*. Un *empilement* est un ensemble fini de couples (A,n), où A est élément de s et n est un entier, appelé *niveau* de A, tel que

- deux pièces distinctes de E en concurrence ne sont jamais au même niveau
- si une pièce P de E est de niveau non nul n, il existe une pièce de E en concurrence avec P et de niveau n-1.

D'un point de vue monoïde de commutation, on peut définir un empilement à l'aide d'une suite d'éléments de $\mathcal{S}$. Deux suites $(A_i)_{i=1..n}$ et $(B_i)_{i=1..n}$ sont dites équivalentes lorsque les deux suites sont formées des mêmes éléments et lorsque les couples d'éléments en concurrence sont dans le même ordre, plus précisément lorsqu'il existe une permutation σ de $\{1..n\}$ telle que

- pour tout i, $B_{\sigma(i)} = A_i$,
- pour tout couple d'entiers $i<j$ tels que $A_i \,\mathcal{R}\, A_j$, $\sigma(i)<\sigma(j)$.

Un *empilement* est une classe d'équivalence pour cette relation. Dans ce qui suit on désignera un empilement à l'aide de l'un de ses représentants. On note $\mathcal{E}(\mathcal{S})$ l'ensemble des empilements d'éléments de $\mathcal{S}$.

Une pièce A_i d'un empilement est dite *maximale* (resp. *minimale*) lorsque aucune des pièces $A_1,\ldots,A_{i-1}$ (resp. $A_{i+1},\ldots,A_n$) n'est en concurrence avec Ai. Les pièces maximales (resp. minimales) correspondent en théorie des monoïdes partiellement commutatifs aux lettres commutant avec toutes les lettres placées à leur gauche (resp. droite). Un empilement dont toutes les pièces sont à la fois maximales et minimales est dit *trivial.* En d'autres termes, un empilement est trivial lorsque aucune de ses pièces n'est en concurrence avec une autre.

L'un des intérêts de la théorie des empilements est de visualiser et d'éclaircir un certain nombre de théorèmes et plus particulièrement les théorèmes d'inversion. Ces théorèmes permettent d'énumérer certaines familles d'empilements relativement à une valuation donnée. On appelle valuation toute application v de l'ensemble des objets dans un anneau de polynômes telle que

- pour tout élément s de $\mathcal{S}$, v(s) est un polynôme sans coefficient constant,
- le nombre d'éléments ayant une composante non nulle relativement à ce monôme est fini, ceci afin de s'assurer de la convergence des séries étudiées.

La valuation d'un empilement est le produit des valuations des pièces qui le composent. On appelle *série génératrice* d'une famille d'empilements la somme des valuations des empilements faisant partie de cette famille. Le théorème 1 ci-dessous est l'analogue du théorème d'inversion de Möbius.

théorème 1 *La série génératrice des empilements de* $\mathcal{E}(\mathcal{A})$ *est*

$$\sum_{E \in \mathcal{E}(\mathcal{A})} v(E) = \frac{1}{\sum_{F \in \mathcal{T}(\mathcal{A})} (-1)^{|F|} v(E)},$$

où $\mathcal{T}(\mathcal{A})$ *désigne l'ensemble des empilements triviaux et* $|F|$ *désigne le nombre de pièces de* F.

théorème 2 *Soit* $\mathcal{M}$ *un sous-ensemble de* $\mathcal{A}$. *La série génératrice des empilements (éventuellement vides) ayant toutes leurs pièces maximales dans* $\mathcal{M}$ *est,*

$$\sum_{E, Max(E) \subset \mathcal{M}} v(E) = \frac{\sum_{F \in \mathcal{T}(\mathcal{M}^c)} (-1)^{|F|} v(E)}{\sum_{F \in \mathcal{T}(\mathcal{A})} (-1)^{|F|} v(E)},$$

où $\mathcal{T}(\mathcal{M}^c)$ *désigne l'ensemble des empilements triviaux dans le complémentaire de* $\mathcal{M}$.

Pour les preuves concernant ces théorèmes ainsi que des compléments sur le sujet, le lecteur pourra se reporter à Viennot [15]. Ceci nous amène à une interprétation combinatoire des numérateurs des développements de $\frac{1}{J_0}$ et $\frac{J_1}{J_0}$ à l'aide de certains éléments du groupe des tresses.

GROUPE DES TRESSES

Le groupe des tresses a été introduit par Artin ([2], [3]). Du point de vue de la théorie des groupes, le groupe des tresses à n+1 brins B_n est le groupe libre engendré par les générateurs $x_1, x_2, \ldots, x_n$, satisfaisant aux relations

(1) pour tout i, $1 \leq i \leq n-1$, $x_i\, x_{i+1}\, x_i = x_{i+1}\, x_i\, x_{i+1}$,

(2) pour tout i,j de $\{1,2\ldots,n\}$ tel que $|\, i-j \,| \geq 2$, $x_i\, x_j = x_j\, x_i$.

On peut visualiser le groupe des tresses de la manière suivante.

Considérons deux ensembles de n+1 points, appelés respectivement origines et extrémités. L'élément x_i du groupe des tresses fait passer un brin joignant i à i+1 au dessus du brin joignant i+1 à i, les autres origines j (j≠i et j≠i+1) étant directement reliées aux extrémités j :

Ces tresses élémentaires se composent aisément ; la tresse f.g s'obtient en reliant les extrémités de f aux origines de g. Ainsi, la relations (1) est représentée figure 4. Les travaux sur le groupe des tresses sont nombreux et variés. Toujours d'un point de vue de la théorie des groupes, le problème du mot ainsi que celui de la conjugaison pour le groupe des tresses ont été résolus par Garside [13]. . Du point de vue topologique, le groupe des tresses est relié à la théorie des noeuds([4]). Récemment, de nombreux développements ont été introduits également en Mécanique Statistique où le groupe des tresses est très proche de la relation de Yang-Baxter ([17]). L'article de Cartier [7] fait le point sur tous les développements récents sur le groupe des tresses.

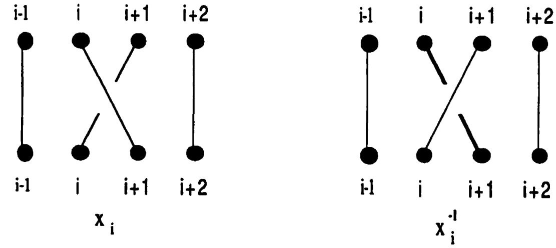


Figure 3. générateurs du groupe des tresses

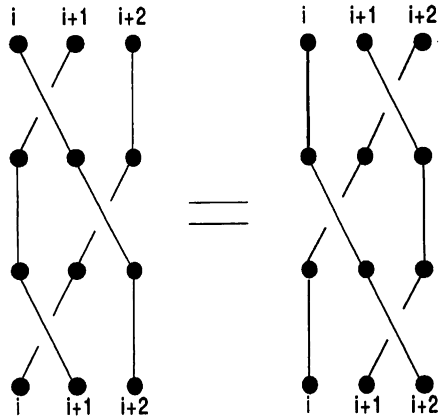


Figure 4. $x_i \, x_{i+1} \, x_i = x_{i+1} \, x_i \, x_{i+1}$

Définition 3. Soit w un mot de $\{x_1, x_2, \ldots, x_{n-1}, x_1^{-1}, x_2^{-1}, \ldots, x_{n-1}^{-1}\}^*$ représentant l'élément t du groupe des tresses B_n. La différence entre le nombre de lettres positives x_i intervenant dans w et le nombre de lettres négatives x_i^{-1} dans w est indépendante du représentant w de la tresse t et est appelée *exposant* de t. On note,

$$\exp(t) = \left(\sum_{i=1}^{n-1} |w|_{x_i} \right) - \left(\sum_{i=1}^{n-1} |w|_{x_i^{-1}} \right).$$

Définition 4. Une fonction Tr de B_n dans $\mathbb{C}$ est une *trace de Markov* lorsqu'elle vérifie les conditions suivantes:

- $Tr(ab) = Tr(a)\, Tr(b)$ pour $a, b \in B_n$,
- $Tr(a\, x_n) = \tau\, Tr(a)$, pour $a \in B_{n-1}$
- $Tr(a\, x_n^{-1}) = \overline{\tau}\, Tr(a)$, pour $a \in B_{n-1}$,

où les constantes τ et $\overline{\tau}$ sont définies par $\tau = Tr(x_n)$ et $\overline{\tau} = Tr(x_n^{-1})$.

En particulier, les traces de Markov sont utilisées pour calculer les polynômes de lien ([4],[17]). Dans ce qui suit, nous utiliserons la trace de Markov triviale Tr définie par $Tr(a) = q^{\exp(a)}$.

Exemple. Si $a = x_1\, x_2\, x_1^{-1}\, x_2$, alors $\exp(a) = 3-1=2$ et $Tr(a) = q^2$.

2. fonctions de Bessel et empilements de demi-segments

2.1. définitions

On considère deux droites parallèles D et D' de $\mathbb{R}^3$, de vecteur directeur $\vec{i}$, et O, O' deux points appartenant respectivement à D et D'. Un point M de D (resp. M' de D') est ainsi entièrement caractérisé par son abscisse m (resp m') dans le repère $(O, \vec{i})$ de D (resp. $(O', \vec{i})$ de D').

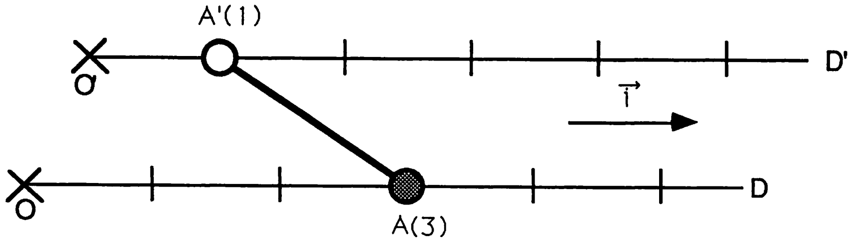


Figure 5. un demi-segment.

définition 5. On appelle *demi-segment* tout intervalle semi-ouvert [A ; A' [où A et A' sont des points à abscisses entières positives a et a' respectivement de D et D'. On appelle *poids* du demi-segment [A,A'[le monôme $\pi([A,A'[)=q^{a+a'}$. Deux demi-segments [A ; A' [et [B ; B' [sont en *concurrence* lorsque ($a\le b$ et $a'>b'$) ou ($a\ge b$ et $a'<b'$) ou ($a=b$ et $a'=b'$).
Exemple. Le demi-segment dessiné figure 5 est [A(3) ; A'(1) [et son poids est q^4.

définition 6. Un *empilement trivial* de demi-segments est la donnée d'une suite de demi-segments deux à deux disjoints. Le *poids* d'un tel empilement est le produit des poids des demi-segments qui le composent.

Exemple. Le poids de l'empilement trivial dessiné figure 6 est q^{30}.

Un empilement trivial ayant n demi-segments est défini par la donnée de n points à abscisses entières $A_1, A_2,\ldots, A_n$ distincts de D et n points à abscisses entières non nécessairement distincts $A'_1, A'_2,\ldots, A'_n$ de D'. L'entier n étant fixé, la somme S_n des poids de tous les empilements triviaux à n pièces est la some des produits $\prod_{i=1}^{n} q^{a_i+a'_i}$, pour $0\le a_1<a_2<\ldots<a_n$ et $0\le a'_1\le a'_2\le\ldots\le a'_n$. En remplaçant a_i par $a_i=\alpha_1+\alpha_2+\ldots+\alpha_i$, et a'_i par $a_i=\alpha'_1+\alpha'_2+\ldots+\alpha'_i$on obtient

$$S_n=\Sigma q^{n\alpha_1+(n-1)\alpha_2+\ldots+2\alpha_{n-1}+\alpha_n} q^{n\alpha'_1+(n-1)\alpha'_2+\ldots+2\alpha'_{n-1}+\alpha'_n},$$

où la somme est prise sur tous les entiers $\alpha_1\ge 0$, $\alpha_2>0,\ldots,\alpha_n>0$, et $\alpha'_1\ge 0,\alpha'_2\ge 0,\ldots,\alpha'_n\ge 0$. Il vient donc

$$S_n=\frac{q^{\binom{n}{2}}}{(q;q)_n^2}.$$

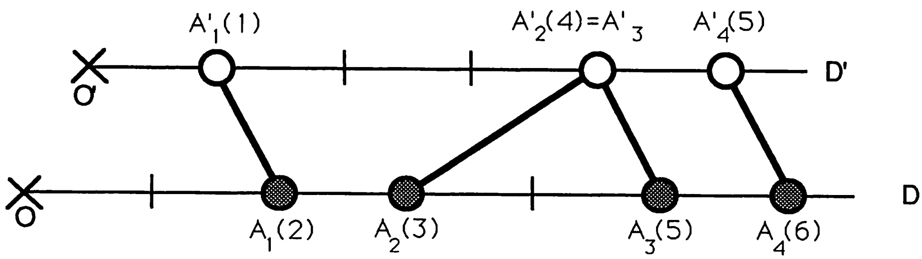


Figure 6. Un empilement trivial.

définition 7. Un *empilement* de demi-segments est la donnée d'une suite de demi-segments. Deux suites de demi-segments définissent le même empilement lorsque elles sont formées des mêmes demi-segments et lorsque les demi-segments en concurrence sont dans le même ordre.

exemple. L'empilement dessiné figure 7 correspond aux suites de points
$([A_1,A'_1[\ , [A_2,A'_2[\ , [A_3,A'_3[\ , [A_4,A'_4[\ )$
$= ([A_2,A'_2[\ , [A_1,A'_1[\ , [A_3,A'_3[\ , [A_4,A'_4[\ )$
$= ([A_2,A'_2[\ , [A_3,A'_3[\ , [A_1,A'_1[\ , [A_4,A'_4[\ )$.

définition 8. On dit qu'un demi-segment $[A_k,A'_k[$ est une *pièce maximale* de l'empilement $([A_i,A'_i[)_{i=1..n}$, lorsque chacun des demi segments $[A_i,A'_i[$, pour i=1 à k-1 n'est pas en concurrence avec $[A_k,A'_k[$.

exemple. Les pièces maximales de l'empilement dessiné figure 7 sont $[A_1,A'_1[$ et $[A_2,A'_2[$.

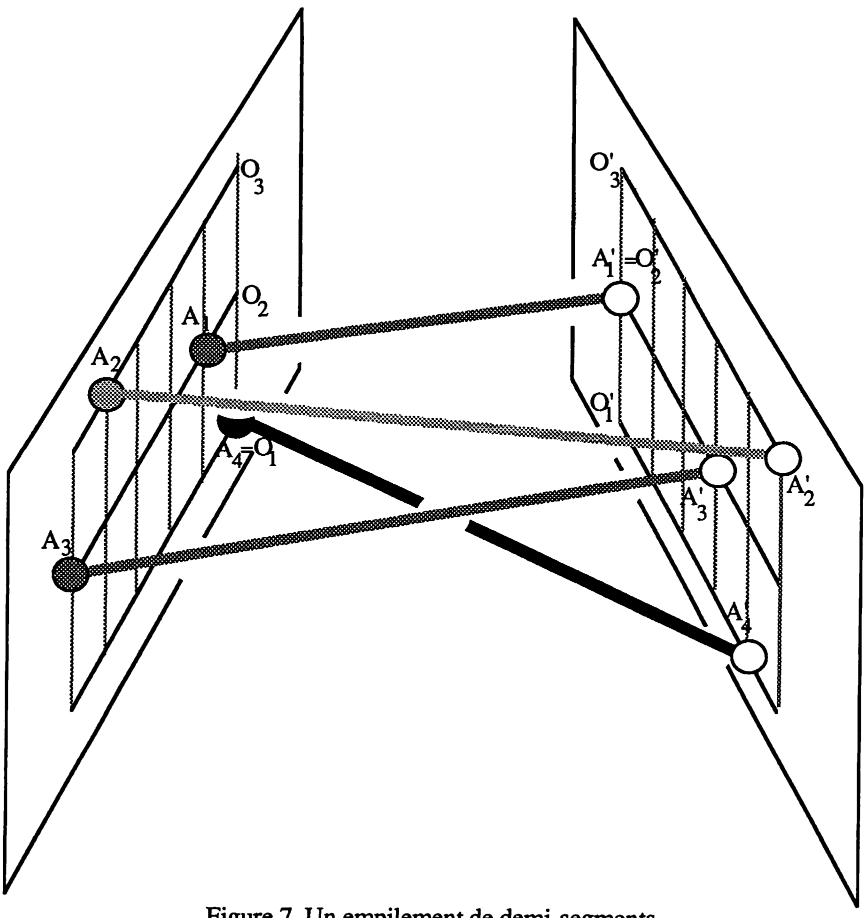


Figure 7. Un empilement de demi-segments

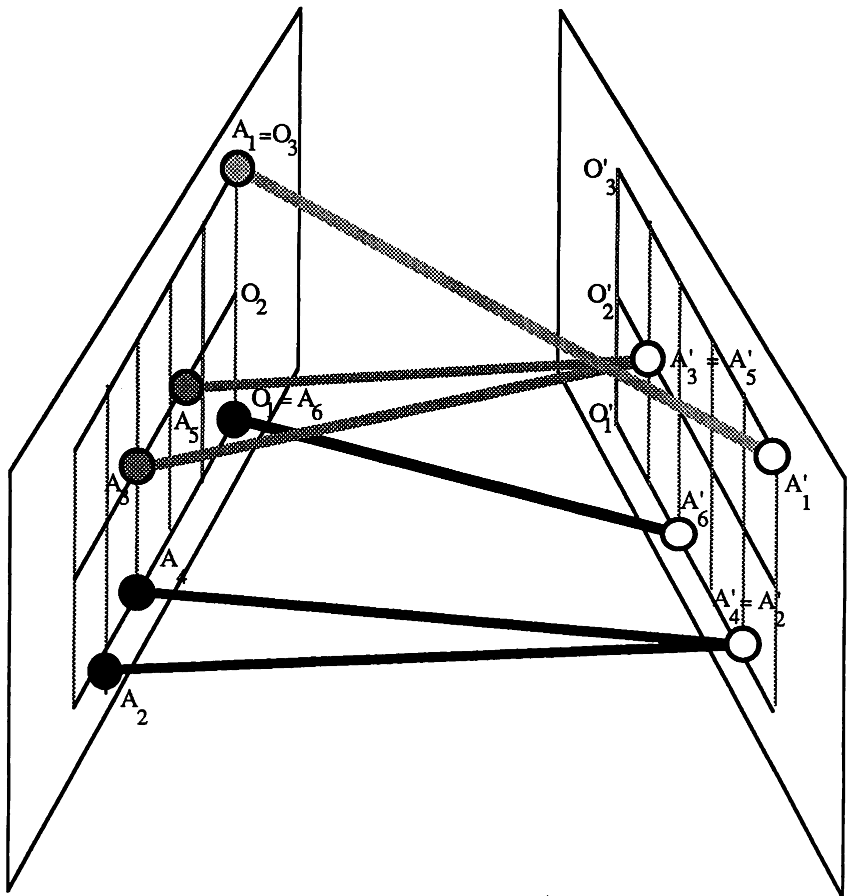


figure 8 Une pyramide de demi-segments

définition 9. On dit que $([A_i,A'_i[)_{i=1..n}$ est l'*écriture standard* d'un empilement de demi-segments lorsque pour tout k, $[A_k,A'_k[$ est la pièce maximale la plus à droite de l'empilement $([A_i,A'_i[)_{i=k..n}$.

exemple. l'écriture standard de l'empilement dessiné figure 5 est
$([A_2,A'_2[\ ,\ [A_1,A'_1[\ ,\ [A_3,A'_3[\ ,\ [A_4,A'_4[\ )$.

2.2. séries génératrices

En appliquant le théorème 1 sur les empilements, on a,

propriété 10. *La série génératrice des empilements de demi-segments est la fonction*
$\frac{1}{qJ_0}$.

De même, la somme des poids des empilements triviaux de n+1 pièces telles que $A_1>0$ est

$$\frac{q^{\binom{n+1}{2}}}{(q;q)_n^2}$$

et l'application du théorème 2 montre immédiatement,

propriété 11. *La série génératrice des empilements non vides de demi-segments tels que la pièce maximale soit de la forme* [O ; M' [*est la fonction* $\frac{qJ_1}{qJ_0}$ *. Ce rapport est également la série génératrice des empilements non vides de demi-segments telles que les pièces maximales soient toutes de la forme* [M ; O' [

exemples. La figure 7 montre, en perspective, un empilement de 4 pièces de poids q^{22}. La figure 8 montre une pyramide de demi-segments dont l'unique pièce maximale est de la forme [O ; M' [.

3. tresses simples

Afin de diviser l'ensemble des empilements en sous classes, nous introduisons la notion de tresse simple. D'une manière intuitive, une tresse simple est une tresse non entrelacée : une tresse simple induit un ordre partiel (dessus, dessous) sur les brins qui la composent. Ainsi, la tresse T de la figure 9 est simple alors que la tresse T' ne l'est pas.

notation. on note $\sigma_{i,j} = x_i\, x_{i+1} \ldots x_{j-1}\, x_j$ pour $i \leq j$, et $\sigma_{i,j} = 1$ lorsque $i > j$.

définition 12. Soit T une tresse de B_n. On dit que le couple (i,j) est un *brin maximal* de la tresse T lorsque la tresse $(\sigma_{i,n})^{-1}\ T\ \sigma_{j,n}$ est une tresse de B_{n-1}.

définition 13. On définit l'ensemble BS_n des *tresses simples* à n+1 brins de B_n par récurrence de la manière suivante,

- *si* n=0, $BS_0 = \{1\}$,
- *si* $n > 0$, *alors* $BS_n = \bigcup_{i=1}^{n+1} \bigcup_{j=1}^{n+1} (\sigma_{i,n})\, BS_{n-1} (\sigma_{j,n})^{-1}$.

exemple. $x_1 = \sigma_{1,1}\ (\sigma_{2,1})^{-1} = \sigma_{1,1}$; $x_1^{-1} = \sigma_{2,1}\ (\sigma_{1,1})^{-1} = (\sigma_{1,1})^{-1}$.

remarque. Les tresses simples de B_n sont donc les éléments T de B_n qui s'écrivent sous la forme,

$$T = \prod_{i=1}^{n} \sigma_{k_i,n+1-i} \prod_{j=1}^{n} \sigma_{h_j,j}^{-1} ,$$

où, pour tout i, j appartenant à $\{1,2,\ldots,n\}$, $1 \leq k_i \leq i+1$ et $1 \leq h_j \leq j+1$.

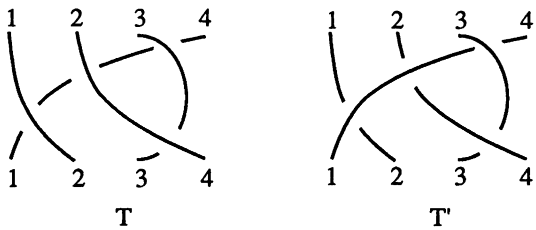
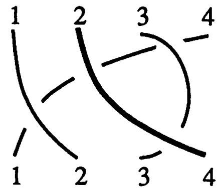


Figure 9. Une tresse simple T, et une tresse non simple T'.

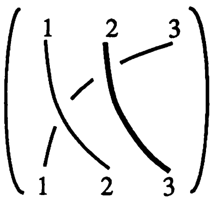
De manière combinatoire, on peut lire l'écriture précédente à partir du dessin d'une tresse simple à l'aide de l'algorithme suivant:

- prendre un brin maximal (k_1,h_1) de T,
- enlever ce brin et considérer la nouvelle tresse T_1 de B_{n-1} obtenue en décrémentant d'une unité les origines de T supérieures à k_1 et les extrémités de T supérieures à h_1.

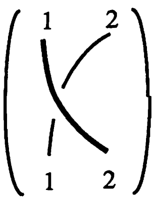
La tresse T_1 ainsi obtenue est une tresse simple sur laquelle on peut itérer ces opérations. On obtient ainsi une suite de tresses simples T_i de B_{n-i} dont un brin maximal est le brin (k_{i+1},h_{i+1}). Cet algorithme permet de lire une tresse simple directement à partir de son dessin. Par exemple, si T désigne la tresse simple de la figure 9, on a,



T

$=\ \sigma_{2,3}$ 

$\sigma^{-1}_{4,3}$

$=\ \sigma_{2,3}\ \sigma_{2,2}$ 

$\sigma^{-1}_{3,2}\ \sigma^{-1}_{4,3}$

$=\ \sigma_{2,3}\ \sigma_{2,2}\,\sigma_{1,1}$

Remarque. Cette écriture n'est pas en général unique. Ainsi, pour la tresse T de la figure 8, les différentes écritures possibles sont,

$$T = \sigma_{2,3}\,\sigma_{2,2}\,\sigma_{1,1} = \sigma_{2,3}\,\sigma_{1,2}\ \sigma_{1,1}\,(\sigma_{2,2})^{-1} = \ \sigma_{1,3}\,\sigma_{1,2}\,(\sigma_{2,3})^{-1}$$

Définition 14. On appelle *écriture standard* d'une tresse simple T l'écriture obtenue dans l'algorithme précédent en prenant à chaque étape le brin maximal le plus à droite.

Exemple. L'écriture standard de la tresse T de la figure 9 est $T = \sigma_{2,3}\,\sigma_{2,2}\,\sigma_{1,1}$.

Remarque. On peut représenter les tresses simples à l'aide de permutations à inversions orientées sans cycles.
En effet, toute tresse T de B_n induit une permutation de S_{n+1}, il suffit de considérer l'image de la tresse T par le morphisme envoyant chaque générateur x_i sur la transposition (i,i+1).Si l'on représente graphiquement une permutation π par la donnée de segments $[i,\pi(i)[$, le nombre de croisements entre ces segments est le nombre d'inversion de la permutation π. Choisir un ordre entre deux segments se croisant revient alors à orienter l'inversion correspondante. On obtient alors une tresse simple en orientant chaque inversion, à condition que le graphe obtenu ne contienne aucun cycle. En effet, si tel est le cas, la tresse obtenue est entrelacée. De plus, la trace de la tresse simple se lit aisément sur la permutation orientée sans cycle en prenant q pour les orientations $(\sigma(i) \longrightarrow \sigma(j))$ et $\frac{1}{q}$ pour les orientations $(\sigma(i) \longleftarrow \sigma(j))$, où $(\sigma(i), \sigma(j))$ est une inversion .

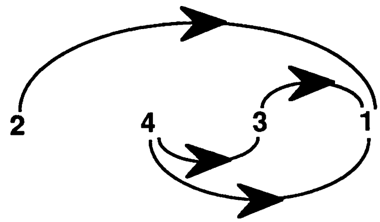


figure 10. Une permutation orientée sans cycle.

Exemple. La tresse T de la figure 9 est représenté par la permutation orientée sans cycle de la figure 10.

4. empilements de demi-segments et tresses simples valuées

4.1. résultats concernant $1/J_0$

Soient $\mathcal{E}(n)$ l'ensemble des empilements de n demi-segments et T(n) l'ensemble des triplets formés d'une tresse simple à n brins, d'une suite croissante de n entiers et d'une suite strictement croissante de n entiers,

$$T(n)=\left\{(t,\{a_k\}_{k=1}^{n},\{a'_k\}_{k=1}^{n}),\left|\begin{array}{l} t\in BS_{n-1} \\ 0\le a_1\le a_2\le\ldots\le a_n \\ 0\le a'_1<a'_2<\ldots<a'_n\end{array}\right.\right\}.$$

théorème 15. *Il existe une bijection* φ_n *de* $\mathcal{E}(n)$ *dans* T(n) *telle que, pour tout* E *de* $\mathcal{E}(n)$,

$$\pi(E)=\pi(t)\,q^{\sum_{i=1}^{n}a_i+a'_i},$$

où $\varphi_n(E)=(t,(a_i)_{i=1}^{n},(a'_i)_{i=1}^{n})$.

preuve. La démonstration se fait par récurrence sur le nombre n de pièces.

Il est clair que $\varphi_1([A(a),A'(a')[\;=(e,(a),(a'))$, où e désigne l'unique tresse à un brin, définit une bijection de $\mathcal{E}_1$ dans T_1 satisfaisant aux conditions du théorème.

Supposons que l'on ait construit la bijection φ_{n-1} de $\mathcal{E}_{n-1}$ dans T_{n-1}.
Soit alors un empilement E de $\mathcal{E}(n)$. Il se décompose de manière unique sous la forme $[A(a),A'(a')[\times F$, où $[A,A'[$ est la pièce maximale la plus à droite de E et F est un empilement à n-1 demi-segments dont toutes les pièces maximales sont soit à gauche de $[A,A'[$, soit en concurrence avec $[A,A'[$. Soit

$$\varphi_{n-1}(F)=(u,(b_i)_{i=1}^{n-1},(b'_i)_{i=1}^{n-1}).$$

On a pour un certain i de $\{1,\ldots,n\}$, et un certain j de $\{1,\ldots,n\}$(on convient que $b_n=b'_n=+\infty$),
$b_1\le b_2\le\ldots\le b_{i-1}\le a<b_i\le\ldots\le b_{n-1}$ et $b'_1<b'_2<\ldots<b'_{j-1}<a'\le b'_j<\ldots<b'_{n-1}$.
Considérons alors les suite $(a_k)_{k=1..n}$ et $(a'_k)_{k=1..n}$ définies par,
$a_k=b_k$ pour $k<i$, $a_i=a$ et $a_k=b_{k-1}-1$ pour $k>i$, et

$a'_k = b'_k$ pour $k<j$, $a'_j = a'$ et $a_k = b_{k-1} + 1$ pour $k>j$.
La suite $(a_k)_{k=1..n}$ est évidemment croissante et la suite $(a'_k)_{k=1..n}$ est strictement croissante. Soit $t = \sigma_{i,n}\, u\, (\sigma_{j,n})^{-1}$, qui est par définition une tresse simple. On montre par récurrence que le mot ainsi récursivement défini est une écriture normale de la tresse associée.
En effet, supposons que E_1 ait pour pièce maximale la plus à droite la pièce $[B(b),B'(b')[$, et que soit par récurrence écrite sous la forme $u= \sigma_{h,n-1}\, v\, (\sigma_{k,n-1})^{-1}$. On a, avec les notations précédentes $b=b_h$ et $b'=b'_k$. Si t n'est pas sous forme normale, alors, $i\leq h$ et $j\leq k$ ce qui entraîne $a<b_h=b$ et $a'\leq b_h=b'$, ce qui contredit le fait que $[A,A']$ est la pièce maximale la plus à droite.
L'application φ_n définie par

$$\varphi_n(E) = (t, (a_i)_{i=1}^n, (a'_i)_{i=1}^n),$$

est la bijection recherchée, la bijection réciproque se définissant de manière analogue.

En sommant le poids des tresses dont l'image par φ_n correspond à une même tresse t_0, on obtient

corollaire 16. *La somme des poids de tous les empilements associés à une même tresse simple* t_0 *est,*

$$\sum_{\varphi_n(E)=(t_0,\ldots,\ldots)} \pi(E) = \frac{tr(t_0)}{(q;q)_n^2}\, q^{\binom{n}{2}}.$$

En sommant sur tous les empilements de demi-segments, il vient,

théorème 17. *Les polynômes* $a_n(q)$ *sont reliés aux tresses simples par la relation,*

$$a_n(q) = q^{\binom{n}{2}} \sum_{t\in BS_{n-1}} tr(t).$$

De plus, les polynômes $a_n(q)$ *sont à coefficients entiers positifs, symétriques de degré* $n(n-1)$.

preuve. La symétrie s'obtient en remarquant que l'inverse d'une tresse simple est une tresse simple et que $\pi(t^{-1}) = \frac{1}{\pi(t)}$. Le degré se déduit en remarquant que la tresse simple de plus fort exposant est la tresse $\sigma_{1,n-1}\, \sigma_{1,n-2} \ldots \sigma_{1,2}\, \sigma_{1,1}$, de poids $q^{n(n-1)/2}$.

4.2. résultats concernant J_1/J_0

Le théorème d'inversion des empilements montre que la fonction $\frac{J_1}{J_0}$ est la série énumératrice dont l'unique pièce maximale est de la forme $[O, M'[$ ou encore la série énumératrice dont l'unique pièce maximale est de la forme $[M, O'[$ Nous montrons ici,

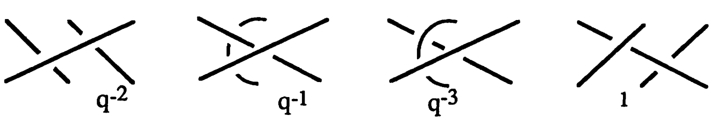


Figure 11. $BS^1{}_2$.

théorème 18. *En notant* $BS1_{n-1}$ *l'ensemble des tresses simples à n brins n'ayant qu'un unique brin maximal de la forme* (i,1), *on a,*

$$b_n(q) = q^{\binom{n}{2}} \sum_{t \in BS^1_{n-1}} tr(t).$$

Ce théorème se déduit en utilisant le corollaire 16 et le lemme suivant,

lemme 19. *L'image par* φ_n *de l'ensemble des empilements de demi-segments dont les pièces maximales sont de la forme* [A,O'[*est l'ensemble des triplets*

$$\left\{ (t, \{a_k\}_{k=1}^n, \{a'_k\}_{k=1}^n), \begin{bmatrix} t \in BS^1_{n-1} \\ 0 \le a_1 \le a_2 \le \ldots \le a_n \\ 0 = a'_1 < a'_2 < \ldots < a'_n \end{bmatrix} \right\}.$$

Reprenons les notations de la construction de la bijection φ_n. Ici, a'=0, et la pièce maximale la plus à droite de l'empilement E_1 est de la forme $[B(b_h),B'(b'_k)[$, où $b_h \le a$ et b'_k quelconque. Par conséquent, $h<i$ et $j=1\le k$, ce qui prouve que la tresse associée appartient bien à $BS1_{n-1}$. De plus, $a'_1=0$, et a est quelconque, d'où la propriété.

exemple. la figure 11 montre les quatre éléments de $BS1_2$ ainsi que leurs poids.

conclusions. Le rapport des fonctions de Bessel ${}^qJ_1$ et ${}^qJ_0$ est apparu lors de l'énumération des diagrammes de Ferrers gauches. La combinatoire liée à ces objets est particulièrement riche. Diverses interprétations combinatoires ont été données, à l'aide de de mots de Dyck, de multichaînes d'un poset, de polytopes rationnels dans la théorie de Ehrhart dans [11], et ici à l'aide des tresses simples. Divers problèmes concernant ces fonctions énumératrices restent ouverts, en particulier la log-concavité ou du moins l'unimodalité des polynômes $a_n(q)$ et $b_n(q)$ ainsi qu'une preuve combinatoire de la symétrie des polynômes $b_n(q)$.

références.

[1] G.E. ANDREWS, q-Series: their development and application in Analysis, Number Theory, Combinatorics, Physics, and Computer Algebra, AMS **66**, (1986).

[2] E. ARTIN, Theori der Zöpfe, Abh. Math. Semin. Univ. **4** (1926) 47-72.

[3] E. ARTIN, Theory of Braids, Annal of Mathematics, **48** (1947) 101-126.

[4] J. BIRMAN and A. LIBGOBER (éditeurs) *Braids*, Proceedings of a research Conference, Contemporary Math. vol. 78, American Math. Soc. (1988).

[5] M. BOUSQUET-MÉLOU, q-Enumération de polyominos convexes, Thèse de Doctorat, Université de Bordeaux I, 1991.

[6] L. CARLITZ, Integers related to the Bessel functions, Proc. of Amer. math. Soc, **14** (1963), 1-9.

[7] P. CARTIER, Développements récents sur les groupes de tresses, Applications à la topologie et à l'algèbre, Séminaire BOURBAKI, dans Astérisque, **716** (1990) 1-42.

[8] P. CARTIER, D. FOATA, *Problèmes combinatoires de commutations et de réarrangements*, Lect. notes. in Math., 85, Springer-Verlag, Berlin (1969).

[9] M.P. DELEST, J.M. FEDOU, Enumeration of skew Ferrers diagrams, to appear in Discrete Math.

[10] A. ERDELYI, et alt., *Higher transcendental functions*, vol.2, McGraw-Hill (1955).

[11] J.M. FEDOU, Enumeration of skew Ferrers diagrams and Basic Bessel functions, à paraitre dans les actes du "2nd conference on lattice path combinatorics and applications" (1990) Hamilton, Canada.

[12] M. ISMAIL, The zeroes of basic Bessel functions and associated orthogonal polynomials, J. of Math. Anal. and Appl. **86** (1982), 1-18.

[13] F.A. GARSIDE, The braid group and other groups, Quart. J. Math. **17** (1969) 245-254.

[14] D.H. LEHMER, Zeros of the Bessel function $J_v(x)$, Math.Tables Aids Comput. **1** (1943-45), 405-407.

[15] X.G. VIENNOT, Heaps of pieces I, Basic definitions and combinatorial lemmas, in Combinatoire énumérative, eds G. Labelle et P. Leroux, Lect. Notes in Math. 1234, Springer-Verlag, Berlin (1986)

[16] G.N. WATSON, *A Treatise on the Theory of Bessel Functions*, Cambridge (1924) 502.

[17] C.N. YANG and M.L. GE (éditeurs) *Braid group, knot theory and statistical mechanics*, Adv. Series Math. Phys. vol 9, World Scientific (1989).

Bijective Principles of Cancellation

(a shortened version of an article
to appear in *Advances in Mathematics*,
as "Producing New Bijections from Old")

March 30, 1992

David Feldman
University of New Hampshire

James Propp
Massachusetts Institute of Technology

ABSTRACT

Combinatorics has many general constructions that produce finite sets from finite sets, such as the Cartesian product and power set constructions. Bijections of the given sets give rise to bijections of the constructed sets in an obvious way. We explore the reverse process, the problem of defining a bijection between the given objects from one between the constructed objects. We formulate a precise interpretation of the question and present a concrete criterion for when such cancellation is or is not possible. We apply the criterion to several standard combinatorial constructions. In particular, we show that a bijection between the mth powers of sets suffices to define a bijection between the sets themselves, but a bijection between power sets generally does not. Our effective cancellation procedures are in some cases computationally infeasible, but we do offer a reasonably explicit procedure, in the spirit of the involution principle of Garsia and Milne, to cancel the construction of forming the Cartesian product of a given finite set with a fixed set that carries a distinguished element. By contrast, our methods allow us to conclude that Cartesian product with a fixed set without extra structure cannot be cancelled.

Key words: Bijection, bijective proof, G-set, Cartesian power, power set

Introduction

Any analytical demonstration of a combinatorial identity compels us to inquire about the existence of a direct demonstration by means of a bijection. While it is conceivable that all true identities will eventually yield to sufficient cleverness, it is also conceivable that this is not the case. To allow for the possibility of settling such questions in the negative, eventually we must formalize the notion of "bijective proof."

Fixing on a good definition seems difficult. Instances of the following sort, however disguised, must surely be excluded: an identity is actually established by analytic means so that the mere existence of a bijection is assured, and by specifying enumerations of the sets on each side of the identity we single one out. Perhaps the bijection itself is to be judged faulty in that it may be inefficient to compute. Perhaps the defect is with the form of the specification of such a bijection, which makes use of the structure of the families of combinatorial objects rather than just the structure of the objects themselves. Or one's complaint may simply be that such a bijection is irrelevant to the proof of the identity.

It is less clear how one should treat instances of the following kind. In the course of showing that two (indexed families of) sets A and B have the same cardinality, a combinatorial construction is performed on each, yielding sets QA and QB and a self-evident bijection between QA and QB is then produced. (We may formalize "combinatorial construction" as a faithful endofunctor on the category of finite sets and bijections. This captures that idea that a construction Q on a set A does not depend on how the elements of A are named.) Assuming it is known that the cardinality of QS determines the cardinality of S, the equicardinality of A and B is proved. This may well be a bijective proof that $|QA| = |QB|$, but is it a bijective proof that $|A| = |B|$?

Sometimes a principle is available that cancels the construction Q, thereby uniformly specifying a bijection between A and B whenever a bijection between QA and QB is given. The involution principle used by Garsia and Milne in their combinatorial proof of the Rogers-Ramanujan identities is a celebrated example, though it must be slightly recast to fall within our rubric. In practice combinatorialists generally accept such proof as "bijective" but the objections raised above still apply. In particular, even when the bijection between QA and QB is computationally efficient the derived bijection between A and B may not be. The specification of the bijection between A and B has a global character, since one takes account of the entire bijection between QA and QB for its definition. Lastly, the truth the identity is evident as soon as one has the bijection from QA to QB, so the bijection between A and B is not strictly relevant to the proof.

One criterion that separates artificial "combinatorifications" of analytic proofs (which we wish to exclude) from proofs that use the involution principle (which we wish to allow) is that the latter may be described as canonical in that they do not involve arbitrary choices. They are, if you will, symmetrical. Hence symmetry considerations will play a pivotal role in our analysis of combinatorial constructions.

We have no hope of settling the debate here as to what truly constitutes a bijective proof. Sidestepping the thorny issue of the absolute existence of bijective proofs of combinatorial identities, we consider more tractable questions of relative existence. A combinatorial cancellation principle for the construction Q takes a bijection that demonstrates the identity $|QA| = |QB|$ and yields a bijection that demonstrates the identity $|A| = |B|$.

We study questions concerning the existence of bijective cancellation principles. The condition we formulate for the cancellation of a construction F is certainly necessary and sufficient. Our main examples show the construction $QA = A^m$, the m^{th} Cartesian power of A, may be cancelled, but $QA = 2^A$, the power set of A, may not.

Even our negative results may have a positive value for the working combinatorialist. Knowing that the power set construction cannot generally be cancelled suggests the possibility of interesting bijections from 2^A to 2^B even in situations where no bijection from A to B is available.

Our methods come from group theory and are purely mathematical, as opposed to metamathematical. Nevertheless it may be fruitful to interpret our results as statements about combinatorics in various toposes, though we do not pursue this.

The organization of our paper is as follows. In the next section, we give a few preliminaries concerning group actions. Then we dive immediately into considerations that lead us to our Fundamental Lemma and Main Theorem, which give us computable criteria for the existence of cancellation principles. The five sections that follow apply this theorem to obtain various positive and negative results, some new and some old: a positive result for cancellation of disjoint union (old), a negative result for Cartesian products (a folk-observation that, while not in the research literature, seems to surprise few seasoned combinatorists), a positive result for Cartesian products with pointed sets (new), a positive result for Cartesian powers (new), and a generally negative result for power sets (new) which is clarified by the solution an interesting group-theoretical problem. The final section situates our work in a more abstract, category-theoretic setting.

Preliminaries

A G-set is a set with an action by a group G (on the left). Two G-sets are isomorphic if and only if there is a bijection from one to the other that commutes with the action by G. A G-orbit is a transitive G-set, and a G-set is a disjoint union of G-orbits. The stabilizer Stab x of an element of a G-set is the set of elements g with $gx = x$. (For background on group actions on finite sets, see Chapter 1 in [Jacobson].)

Note that if S is a G-orbit containing an element x, then the action of G on S is isomorphic to the action of G on the left-cosets of Stab x. Furthermore, if x and y are both in the orbit S, Stab x and Stab y are conjugate subgroups. It follows that a G-orbit is determined up to isomorphism by the conjugacy class of the stabilizer of any of its elements.

If a G-orbit contains an element whose stabilizer is H, we say that it is a G-orbit of H-type. Note that for H, H' conjugate in G, an orbit is of H-type if and only if it is of H'-type.

For each group G, the above remarks allow us to formulate a complete numerical isomorphism invariant for G-sets – indeed, several such invariants, each with its own uses. Let $L(G)$ be the lattice of subgroups of G. For any G-set S, we define maps $\sigma_S, \sigma_S^+, \tau_S$ from $L(G)$ to $\mathbf{Z}$ by letting $\sigma_S(H)$ be the number of elements of S whose stabilizer is H, $\sigma_S^+(H)$ be the number of elements of S that are stabilized by H (i.e., whose stabilizer contains H), and $\tau_S(H)$ be the number of G-orbits of S of H-type. Note that all three functions take a single value on each conjugacy class of subgroups of G. For every non-negative integer-valued function $\tau(H)$ that is constant on conjugacy classes of subgroups there exists an associated G-set.

Equivariance Criterion

Our first goal is to fix a way of making questions about bijective cancellation precise. In particular, we must specify carefully that information which is to be available to a cancellation procedure. Input to the procedure will consist solely of:
(i) rosters of the elements of sets A, B, QA and QB;
(ii) the bijection between QA and QB;
(iii) the declaration of any extra structure that the sets A and B might carry, and whatever additional structure QA and QB acquire through the construction Q.
The procedure must not depend on extraneous information implicit in the presentation of the data. It must make no use of the ordering of the elements on the rosters, no use of the specifics of the coding of their names or anything

else, lest the task of producing a bijection between A and B be trivialized. Now, as we shall indicate, these stipulations may be formalized so as to treat the cancellation procedure as a black box. Having no need to specify a particular model of computation, we do not do so.

We will write $\mathrm{Bij}(X, Y)$ for the set of bijections between X and Y. For the moment, let us fix two sets A and B of cardinality n. Let Q be an endofunctor of the category of finite sets and bijective maps, which is to say, a species in the sense of A. Joyal (see [Joyal]). Thus Q takes a finite set A to a finite set QA, takes a bijection $f : A \to B$ to a bijection $Qf : QA \to QB$, and respects composition of composable bijections. We also ask that Q be faithful, so $Qf = Qg$ if and only if $f = g$. Then if a cancellation procedure for Q exists, it will provide a function

$$\mathcal{F}_{A,B} : \mathrm{Bij}(QA, QB) \to \mathrm{Bij}(A, B) .$$

The procedure cannot depend on the names of the elements of A and B, so the function $\mathcal{F}_{A,B}$ must be invariant under relabelings. More exactly: Write $\mathrm{Sym}\ S$ for the symmetric group on a set S. The group $\mathrm{Sym}\ A \times \mathrm{Sym}\ B$ acts on $\mathrm{Bij}(A, B)$ by

$$(\rho_1, \rho_2)f = \rho_2 f \rho_1^{-1}$$

and on $\mathrm{Bij}(QA, QB)$ by

$$(\rho_1, \rho_2)F = \overline{\rho_2} F \overline{\rho_1^{-1}},$$

where $\overline{\rho}$ is shorthand for $Q\rho$. Invariance under relabelings amounts to

$$\mathcal{F}_{A,B}((\rho_1, \rho_2)f) = (\rho_1, \rho_2)\mathcal{F}_{A,B}(f) ,$$

that is, the equivariance of $\mathcal{F}_{A,B}$ with respect to the action of $\mathrm{Sym}\ A \times \mathrm{Sym}\ B$. The existence of such an equivariant map, for each n, is thus a necessary condition for the existence of an effective procedure of the desired kind. (We do not stipulate that $\mathcal{F}_{A,B}$ should be a one-sided inverse to F, but this is an easy consequence of equivariance, by the fundamental lemma below.)

Is it also a sufficient condition? Write $[n]$ for $\{1, \ldots, n\}$. If an equivariant map exists, for any A and B, then certainly there is a map

$$\mathcal{F}_{[n],[n]} : \mathrm{Bij}(Q[n], Q[n]) \to \mathrm{Bij}([n], [n])$$

equivariant under $\mathrm{Sym}\ [n] \times \mathrm{Sym}\ [n]$. (Of course, producing an equivariant $\mathcal{F}_{[n],[n]}$ from $\mathcal{F}_{A,B}$ requires well-ordering A and B; since A and B are finite sets this can certainly be done.) It is easy to define a lexicographic order on the set of *all* maps from $\mathrm{Bij}(Q[n], Q[n])$ to $\mathrm{Bij}([n], [n])$, since $[n]$ comes with a canonical ordering. So we may effectively single out one equivariant map for each n, namely, the lexicographically earliest one. Let this be done. (Note that in practice this step may be computationally infeasible.)

We are trying to pass from the mere existence of an equivariant $\mathcal{F}_{A,B}$ to an effective implementation by way of $\mathcal{F}_{[n],[n]}$. Suppose we are given a bijection F between QA and QB, where A and B each have cardinality n. Our procedure cannot make use of a particular pair of bijections $a : A \to [n]$ and $b : B \to [n]$, since there is no canonical way choose such. Nevertheless, the procedure may consider the totality of all possible pairs a and b. Each such pair yields bijections $\overline{a} : QA \to Q[n]$ and $\overline{b} : QB \to Q[n]$ from which we obtain the self-bijection of $Q[n]$

$$F' = \overline{b}F\overline{a}^{-1}$$

and the bijection

$$b^{-1}(\mathcal{F}_{[n],[n]}(F'))a$$

between A and B. The equivariance of $\mathcal{F}_{[n],[n]}$ guarantees that the same bijection between A and B will result in every instance. Thus we have a procedure for effectively determining bijection from A to B. We conclude that the existence of an equivariant $\mathcal{F}_{A,B}$ is indeed sufficient.

Some foundational reflection on the last argument may be justified. We are working within the assumptions of classical set theory and logic. Nevertheless, our problem, dealing as it does with sets containing indistinguishable elements, is awkward to formalize in that context. For us, the distinction between a non-empty finite set whose elements are entirely indistinguishable and an empty set is fundamental. The set of well-orderings of a non-empty set of indistinguishable elements may also have no distinguished elements; nevertheless it is non-empty.

Casting the argument somewhat differently may appeal more to some readers. Suppose $\mathcal{F}_{[n],[n]}(F') = f'$. Fix a sufficiently rich first-order language with variables ranging over the elements of $[n]$ (but with no constants for the elements of $[n]$), so that we may characterize the isomorphism type of any F' by a finite predicate. Since F' may have symmetries, some elements of $[n]$ may be indistinguishable by predicates, but equivariance guarantees that we may describe the function f' by predicates nevertheless. Construct such predicates for every pair (F', f'). This family of predicates may now be used to determine when $\mathcal{F}_{A,B}(F) = f$. Thus one never need refer to a well-ordering of A or B. One need only presume a computational model powerful enough to implement the necessary logical procedures.

In view of the above, we adopt the following principle:

Equivariance Criterion: A combinatorial construction Q is bijectively cancellable iff for all A, B there exists a Sym $A \times$ Sym B-equivariant map from Bij(QA, QB) to Bij(A, B).

Main Theorem

We will now concentrate on the question of existence of an equivariant $\mathcal{F}_{A,B}$.

Fundamental Lemma Let G be a (finite) group and S and T be G-sets. (sets with a G action.) Then the following are equivalent:

(i) There is a G-equivariant map h from S to T.

(ii) For every $s \in S$ there is a $t \in T$ such that Stab $s \subseteq$ Stab t.

Proof
(i)⇒(ii): Stab $s \subseteq$ Stab $h(s)$ since $gs = s$ implies $gh(s) = h(gs) = h(s)$.
(ii)⇒(i): Pick a representative $s_{\mathcal{O}}$ of each orbit $\mathcal{O}$ of S. Set $h(s_{\mathcal{O}}) = t_{\mathcal{O}}$ for some element $t_{\mathcal{O}}$ of T satisfying Stab $s_{\mathcal{O}} \subseteq$ Stab $t_{\mathcal{O}}$. Extend h to a map from S to T by sending $h(gs_{\mathcal{O}})$ to $gt_{\mathcal{O}}$. To see that h is well-defined, suppose $g_1 s_{\mathcal{O}} = g_2 s_{\mathcal{O}}$. Then $g_2^{-1}g_1 \in$ Stab $s_{\mathcal{O}}$ and $g_2^{-1}g_1 \in$ Stab $t_{\mathcal{O}}$, that is $g_1 t_{\mathcal{O}} = g_2 t_{\mathcal{O}}$ as desired. To see that h is equivariant, note that for all $g' \in G$, $h(g'(gs_{\mathcal{O}})) = g'gt_{\mathcal{O}} = g'(h(gs_{\mathcal{O}}))$. □

Here is how the Fundamental Lemma is applied.

For $f \in \mathrm{Bij}(A, B)$, Stab f is exactly the group

$$\{ (\rho, f\rho f^{-1}) \mid \rho \in \mathrm{Sym}\ A\} .$$

According to the Fundamental Lemma, there is an equivariant map

$$\mathcal{F}_{A,B} : \mathrm{Bij}(QA, QB) \rightarrow \mathrm{Bij}(A, B)$$

if and only if for every $F \in \mathrm{Bij}(QA, QB)$, Stab F is contained in one of the groups Stab f.

Suppose $(\rho_1, \rho_2) \in$ Stab F. Then ρ_2 is determined by ρ_1. Indeed if Stab F also contains (ρ_1, ρ_2') then it contains $(e, \rho_2(\rho_2')^{-1}) \in$ Stab F as well, where e is the identity map, and

$$\overline{\rho_2(\rho_2')^{-1}} = F\overline{e}F^{-1} = \overline{e} .$$

Since Q is faithful, we have $\rho_2(\rho_2')^{-1} = e$, so $\rho_2 = \rho_2'$.

We may regard A and B as Stab F-sets by taking $(\rho_1, \rho_2)(a) = \rho_1(a)$ for $a \in A$ and $(\rho_1, \rho_2)(b) = \rho_2(b)$ for $b \in B$. In particular, the following two statements are equivalent:
(i) There exists $f : A \rightarrow B$ such that Stab $F \subset$ Stab f.
(ii)There is an $f : A \rightarrow B$ which is an isomorphism of Stab F-sets.

For (i) says that $\rho_2 = f\rho_1 f^{-1}$ and (ii) says that for $(\rho_1, \rho_2) \in \text{Stab } F$ we have $f\rho_1 = \rho_2 f$.

We may also regard QA and QB as (Stab F)-sets by taking $(\rho_1, \rho_2)(\overline{a}) = \overline{\rho_1}(\overline{a})$ for $\overline{a} \in QA$ and $(\rho_1, \rho_2)(\overline{b}) = \overline{\rho_2}(\overline{b})$ for $\overline{b} \in QB$. Certainly $F : QA \to QB$ is an isomorphism of (Stab F)-sets.

In general, if G is a group and A is a G-set, QA is a G-set in a natural way, since Q is a functor. We are now ready for the

Main Theorem The faithful endofunctor Q is bijectively cancellable if and only if for all finite groups G and for all finite sets A and B, QA and QB are isomorphic G-sets if and only if A and B are isomorphic G-sets.

Proof In one direction, the stated condition, along with the recently noted fact that $F : QA \to QB$ is an isomorphism of (Stab F)-sets, implies condition (ii) above. This implies condition (i), which by the Fundamental Lemma implies the existence of an equivariant map.

Conversely, suppose the stated condition fails. Then let A and B be nonisomorphic G-sets with QA and QB isomorphic as G-sets, and let $F : QA \to QB$ be a G-set isomorphism. Then G maps to Stab F, but the equivalence of (i) and (ii) forbids Stab $F \subset$ Stab f for any $f : A \to B$, lest A and B be isomorphic, and the Fundamental Lemma says that no equivariant map exists. □

Cancellation of Disjoint Union

In this and the following sections, we take A, B, and C to be non-empty sets.

First we consider the disjoint union construction $X \mapsto X \mathring{\cup} C$, which we will hereafter write as $X + C$ to suggest the analogy with arithmetic addition. Suppose that one knows a bijection $F : A + C \to B + C$. Given $a \in A$, iterating f sufficiently often on a must eventually produce and element b of B. Thus we define a bijection $\widehat{f} : A \to B$. (See [Stanley].) Notice that the only *a priori* bound on the number of iterations required is the cardinality of C. What is crucial for us is that the construction of $F = \widehat{f}$ used only the information provided by the bijection f, not the properties or names of the elements of A or B. Underlying this situation is the simple group-theoretical fact that, for any finite group G and G-set C, G-sets A and B are isomorphic if and only if G-sets $A + C$ and $B + C$ are isomorphic. (See [Burnside].) More specifically, note that for a fixed group G, the effect of the operation Q on the each of the G-set invariants σ, σ^+, τ is to add a constant vector, which is clearly an invertible operation.

In light of the Main Theorem, the reader may wonder why we allow the possibility that C carries a nontrivial G-action. Q here is the construction "disjoint union with the fixed set C." Since we do not insist on C being disjoint from A and B, the relabeling their elements may induce a permutation on C as well. The possibility of C carrying a non-trivial group actions corresponds to the fact that disjoint union of sets with C can be accomplished in a C-equivariant way (i.e., one that does not depend on the labelling of the elements of C).

Noncancellation of Products

Now we consider the Cartesian product construction $X \mapsto X \times C$. Fix a set X, $|X| > 1$, and set both A and C equal to the set of linear orderings of X. Let B be the set of permutations of X. We may define a bijection between between $A \times C$ and $B \times C$ by using the fact that two linear orderings of X induce a permutation of X. There can be no canonical bijection between A and B, however: The set of permutations of X has a distinguished element, the identity permutation, but the set of linear orderings does not.

Here is the general group-theoretic viewpoint. Let G be a finite group. Let G_t be the G-set obtained by having G act on itself by translation and G_c be the G-set obtained by having G act on itself by conjugation (so that g sends g' to $gg'g^{-1}$). For a nontrivial group, G_t and G_c cannot be isomorphic as G_t is transitive but G_c is not. Nevertheless $h : G_t \times G_t \to G_c \times G_t$ defined by $h(g_1, g_2) = (g_1 g_2^{-1}, g_2)$ is always an isomorphism of G-sets.

Cancellation of Pointed Products

The situation is altogether different when we multiply A and B by a set C with a distinguished point.

From the group-theoretic viewpoint, the analog of C is a G-set with a fixed point.

Theorem A and B and G-sets and and C is a G-set with a distinguished point, and $A \times C$ is isomorphic to $B \times C$, then A is isomorphic to B.

Proof A G-set A is determined by the invariant, σ_A, the integer-function on the set of subgroups of G which records how many times each subgroup of H of G occurs as a stabilizer. Given a point $a \in A$ and a point $c \in C$, the stabilizer of (a, c) in $A \times C$ is $\text{Stab}(a) \cap \text{Stab}(c)$. It follows immediately that $\sigma_{A\times C}$ is linear in σ_A. Represent the linear transformation by the matrix M.

We show that M is nonsingular. Order the subgroups of G so that their cardinalities are non-decreasing. Then M is upper triangular. The fixed point of C guarantees that the diagonal entries of M don't vanish. □

Notice the contrast with the previous situation. The G-set obtained by having G act on itself by translation had no fixed point and this led to an upper triangular matrix with some zeros on the diagonal.

We may also exhibit this cancellation principle more explicitly by the following construction. Now A and B are finite sets, and C is a finite pointed set, with a distinguished element called $*$. Let $f : A \times C \to B \times C$ be a bijection. We define a map $f_* : A \to B$ as follows. $f_*(a)$ is the projection of $f((a,*))$ onto B. A map $f_*^{-1}(= (f^{-1})_*) : B \to A$ is defined similarly. Iteration of the map $f_* \cup f_*^{-1} : A \cup B \to A \cup B$ produces some cycles. We then use f_* to pair any element a of A that occurs in a cycle with an element $f(a)$. This gives us a nontrivial partial bijection between A and B, say from $\tilde{A}$ to $\tilde{B}$. Multiplying by C we get a nontrivial partial from $A \times C$ to $B \times C$ taking $\tilde{A} \times C$ to $\tilde{B} \times C$. This induces a bijection from $(A \setminus \tilde{A}) \times C$ to $(B \setminus \tilde{B}) \times C$ (see the section "Cancellation of Disjoint Union") and now we may iterate the process until we get a bijection from A to B. It should be noted that this construction was only discovered after the group-theoretical approach suggested its feasibility.

Cancellation of Powers

Given a bijection between between the m-fold Cartesian powers of two finite sets, we will use the Main Theorem to define a bijection between the sets themselves.

Theorem Fix a finite group G. If the m-fold powers r^m and r'^m of the permutation representations r and r' of G are isomorphic, then r and r' are isomorphic.

Proof Without loss of generality, assume r and r' both represent G in Sym S for some finite set S. Recall that $L(G)$ is the lattice of subgroups of G, and that the permutation representation r determines a function $\sigma_r : L(G) \to \mathbf{Z}$ which counts the number of times each subgroup of G occurs as the stabilizer of an element in S. The stabilizer of an m-tuple $(s_1, \ldots, s_m)$ is just $\bigcap \text{Stab } s_i$. This observation makes it easy to compute σ_{r^m} from σ_r: for any subgroup G' of G,

$$\sigma_{r^m}(G') = \sum_{\substack{H_1, \ldots, H_m \\ G' = \bigcap H_i}} \sigma_r(H_1) \cdots \sigma_r(H_m).$$

In particular, $\sigma_{r^m}(G) = \sigma_r(G)^m$ since an intersection of subgroups that gives G cannot involve groups other than G itself. We will prove that σ_r can be recovered from σ_{r^m} by induction on the length of the longest chain from a subgroup G' to the top of the lattice. We have already seen that $\sigma_r(G) = (\sigma_{r^m}(G))^{1/m}$. Let G' be a proper subgroup of G, and assume that we have calculated $\sigma_r(H)$ for all groups H strictly containing G'. Then

$$\sigma_{r^m}(G') = \sum_{\substack{H_1,\ldots,H_m \supset G' \\ G' = \bigcap H_i}} \sigma_r(H_1)\cdots\sigma_r(H_m) + \sum_{j=1}^{m} \binom{m}{j} \left(\sum_{H_1,\ldots,H_{m-j} \supset G'} \sigma_r(H_1)\cdots\sigma_r(H_{m-j}) \right) \sigma_r(G')^j$$

where the H_i are subgroups of G that properly contain G'. Since the right side is a polynomial in $\sigma_r(G')$ with positive coefficients, the equation can have at most one non-negative solution. Thus σ_{r^m} determines σ_r. Since $\sigma_{r^m} = \sigma_{r'^m}$, it follows that $\sigma_r = \sigma_{r'}$, which implies that r and r' are isomorphic. □

The Main Theorem now implies that the m-fold Cartesian power functor is cancellable. However, we do not have a satisfying, concrete description of such a cancellation principle.

Noncancellation of Power Sets

The power set of S will be denoted 2^S. We show that a bijection between 2^{S_1} and 2^{S_2} does not generally induce a canonical bijection between S_1 and S_2 by finding a finite group G and a pair of G-sets S_1 and S_2 which are not isomorphic even though 2_1^S and 2_2^S are isomorphic. The counterexample is then the isomorphism between 2_1^S and 2_2^S, considered simply as a bijection between power sets. In this light, the G-action appears as relabelling-symmetries. If there were a canonical induced bijection between S_1 and S_2 it would also possess the same relabelling-symmetries. This is impossible, since G has a different action on the two sets.

Since the stabilizer of an element is H if it is stable under H but under no strictly larger subgroup of G, we have

$$\sigma_S^+(H) = \sum_{K \supseteq H} \sigma_S(K)$$

and

$$\sigma_S(H) = \sigma_S^+(H) - \sum_{K \supset H} \sigma_S(K).$$

The function $\sigma_S(H)$ can be determined from $\sigma_S^+(H)$ working inductively down the lattice of subgroups. Moreover, each orbit of H-type contains $[N_G(H):H]$ elements with stabilizer H, so

$$\sigma_S(H) = [N_G(H):H]\tau_S(H)$$

where $[N_G(H):H]$ is the index of H in its normalizer. The isomorphism type of a G-set S is clearly determined by the function $\tau_S(H)$, and so by either of the functions $\sigma_S^+(H)$, $\sigma_S(H)$.

Let $\mathcal{O}$ be a G-orbit of H-type, with $s \in \mathcal{O}$ having stabilizer H. Pairing elements gs of $\mathcal{O}$ with cosets gH of H gives a G-set isomorphism between $\mathcal{O}$ and G/H. If K is a subgroup of G, the following are clearly all equivalent:
(i) gs is in the same K-orbit as $g's$
(ii) there exist $k \in K$ such that $gs = kg's$
(iii) there exist $k \in K$ such that $gH = kg'H$
(iv) there exist $k \in K$, $h \in H$ such that $g = kg'h$
(v) $KgH = Kg'H$.
Let $o(K,H)$ be the number of K-orbits in G/H. The equivalence of (i) and (v) implies that $o(K,H)$ is also the number of double cosets of the form KgH. For any H' conjugate to H, $\mathcal{O}$ is also of H' type, so $o(K,H') = o(K,H)$. As the number of double cosets KgH equals the number of double cosets HgK (note $(HgK)^{-1} = Kg^{-1}H$), $o(K,H) = o(H,K)$. In particular, $o(K,H)$ also only depends on the conjugacy class of K.

The key formula is

$$\sigma_{2^S}^+(K) = 2^{\sum_H o(H,K)\tau_S(H)},$$

where H ranges over representatives of the conjugacy classes of G. The left side is the number of subsets of S which are stabilized by K. A subset of S is stabilized by K exactly if it is the union of K-orbits of S. The number of K-orbits of S is $\sum_{H \subseteq G} o(H,K)\tau_S(H)$ since each G-orbit of H-type decomposes into $o(H,K)$ K-orbits.

Example 1 Take $G = Z/2Z \times Z/2Z$, the Klein 4-group. Since G is Abelian the issue of conjugacy of subgroups is moot. The subgroups of G are G itself, the three 2-element subgroups R_1, R_2 and R_3, and $\{e\}$. Taking the subgroups in this order, the matrix

$$[o(H,K)] = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 & 2 \\ 1 & 1 & 2 & 1 & 2 \\ 1 & 1 & 1 & 2 & 2 \\ 1 & 2 & 2 & 2 & 4 \end{bmatrix}$$

is singular, with vector $[-2\ 1\ 1\ 1\ -1]$ in the kernel. The linear transformation $[o(H,K)]$ takes the same values on $[2\ 0\ 0\ 0\ 1]$ and $[0\ 1\ 1\ 1\ 0]$. Let S_1 be a G-set with two orbits of type G and one of type $\{e\}$. Let S_2 be a G-set with one orbit of type R_i, $i = 1, 2, 3$. Then 2^{S_1} and 2^{S_2} are isomorphic G-sets.

Example 2 Take $G = S_3$, the symmetric group on three letters. The conjugacy classes of subgroups of G are represented G itself, the 3-element cyclic subgroup C_3, any of the 2-element subgroups R_i generated by a reflection, and $\{e\}$. Taking the subgroups in this order, the matrix

$$[o(H,K)] = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 2 \\ 1 & 1 & 2 & 3 \\ 1 & 2 & 3 & 6 \end{bmatrix}$$

is also singular, with vector $[-2\ 1\ 2\ -1]$ in the kernel, so the linear transformation $[o(H,K)]$ takes the same values at $[2\ 0\ 0\ 1]$ and $[0\ 1\ 2\ 0]$. Let S_1 be a G-set with two orbits of type G and one of type $\{e\}$. Let S_2 be a G-set with one orbit of type C_3, and two of type R_i. Then 2^{S_1} and 2^{S_2} are isomorphic G-sets.

The preceding examples are far from rare, as we will now show.

We thank Goetz Pfeiffer for a demonstration of the following

Theorem The matrix $[o(H,K)]$ has nonzero determinant precisely when G is cyclic.

We have modified Pfeiffer's approach to avoid quoting results from representation theory. First we will need a

Lemma Let G be a finite group with subgroups H and K. Then the number of double cosets KgH is the scalar product of the permutation characters of H and K.

Proof The number of double cosets KgH is the number of K-orbits in G/H. By the orbit counting formula, this is

$$\begin{aligned} \frac{1}{|K|}\sum_{k\in K}|\mathrm{Fix}_{G/H}k| &= \frac{1}{|K|}\sum_{k\in K}\frac{1}{|H|}\sum_{g\in G}\delta(kgH = gH) \\ &= \frac{1}{|K|}\frac{1}{|H|}\sum_{k\in K}\sum_{g\in G}\delta(g^{-1}kg \in H) \end{aligned}$$

Here $\delta(\mathcal{P})$ is 1 if $\mathcal{P}$ is true and 0 if not. Since each pair (k,g) satisfying $k \in K, g^{-1}kg \in H$ gives rise to $|G|$ triples $(x,y,z) = (g'kg'^{-1}, g', g'g)$ satisdetermines the cardinality of A. The categorical view says Q is cancellable if there is a functor from $\mathbf{Fin}_Q$ back to $\mathbf{Fin}$. In general one cannot expect such functors. Indeed there is a paucity of interesting group homomorphisms

$$\mathrm{Hom}_{\mathbf{Fin}_Q}(A,A) = \mathrm{Hom}_{\mathbf{Fin}}(QA,QA) \to \mathrm{Hom}_{\mathbf{Fin}}(A,A)\,.$$

The question arises then, if our equivariant maps

$$\mathcal{F}_{A,B} : \mathrm{Bij}(QA,QB) \to \mathrm{Bij}(A,B)$$

do not define functors, how close do they come? Indeed we have functor-like gadgets $\mathcal{F}$, mapping objects to objects and arrows to arrows from $\mathbf{Fin}_Q$ to $\mathbf{Fin}$. Nevertheless, for composable arrows x and y we only insist that

$$\mathcal{F}(xy) = \mathcal{F}(x)\mathcal{F}(y)$$

when at least one of x or y is in the image of the functor Q. We call such an $\mathcal{F}$ a semifunctor. In general we can consider semifunctors in any category with a distinguished subcategory. Note that a single equivariant bijection $\mathcal{F}_{A,B}$ induces a semifunctor on a component of the category; there are no further compatability constraints. If we take $A = B$, we are just looking at maps from $\mathrm{Bij}(QA,QA)$ to $\mathrm{Bij}(A,A)$ as two sided $\mathrm{Bij}(A,A)$-sets, so we are back to group theory. Since these are both symmetric groups, we frame the following

Problem Classify pairs (G,H), with G and H symmetric groups satisfying $H \subset G$, according to the existence of a map $G \to H$ that respects G and H as two-sided H sets.

REFERENCES

[Burnside] W. Burnside, *Theory of Groups of Finite Order.* Dover Publications, 1955.

[Jacobson] Nathan Jacobson, *Basic Algebra I.* W. H. Freeman, 1974.

[Joyal] André Joyal, *Une théorie combinatoire des séries formelles,* Adv. in Math. **42**, Academic Press (1981), 1-82.

[Stanley] Richard Stanley, *Enumerative Combinatorics I.* Wadsworth, 1986.

That G be cyclic is also sufficient for D to have full rank. Let $\tilde{C} = [\tilde{c}_{ij}]$ be the $s \times s$ matrix, with rows indexed by the subgroups of G, columns indexed by a generator for each subgroup and $\tilde{c}_{ij}$ equal to the number of elements fixed when g_j acts on O_i times the number generators of $\langle g_j \rangle$. Clearly $\tilde{C}\tilde{C}^T = CC^T = D$, since $\tilde{C}$ amounts to "collecting" the identical rows of C. On the other hand $\tilde{C}$ is invertible since its columns are nonzero multiples of the columns in the table of marks. This concludes the proof.

Comment The previous argument gives a formula for the determinant of $D = D(n)$ when $G = \mathbf{Z}/\langle n \rangle$:

$$\mathrm{Det}(D) = \prod_{d|n} d\phi(d) .$$

Alternatively, the determinant of a tensor product is given by the formula

$$\mathrm{Det}(A \otimes B) = \mathrm{Det}(A)^{\dim(B)} \cdot \mathrm{Det}(B)^{\dim(A)} .$$

If $n = pq$ with p and q relatively prime, then essentially

$$D(n) = D(p) \otimes D(q) .$$

If p is prime, then

$$\mathrm{Det}(D(p^n)) = p^{p-1}(p-1) \cdot \mathrm{Det}(D(p^{n-1}))$$

and $\mathrm{Det}(D(p)) = p - 1$.

The theorem guarantees a rich supply of counterexamples to the cancellation of the power set construction, but in a small way it has a positive aspect as well, for it shows that bijection between powers sets which have only cyclic symmetry may in fact be cancelled.

Categorical Viewpoint

Till now, we have regarded constructions Q as endofunctors of the category **Fin** of finite sets and bijections. It is better here to consider the functor $\widehat{Q}$ from **Fin** to $\mathbf{Fin}_Q$, the full image of Q. The objects of $\mathbf{Fin}_Q$ are the objects of **Fin**, but

$$\mathrm{Hom}_{\mathbf{Fin}_Q}(A, B) = \mathrm{Hom}_{\mathbf{Fin}}(QA, QB) .$$

On objects $\widehat{Q}$ is the identity map (this avoids the technical annoyance that $QA = QB$ does not generally imply $A = B$) and on maps $\widehat{Q}$ coincides with Q. Note that $\widehat{Q}$ is surjective on objects where Q is not.

Our notion of a canonical cancellation for a construction Q lies between two extremes. The classical view says Q is cancellable if the cardinality of QA

fying $y^{-1}xy \in K, z^{-1}xz \in H$, we may rewrite what we had as

$$\begin{aligned} & \frac{1}{|G|}\frac{1}{|K|}\frac{1}{|H|}\sum_{x\in G}\sum_{y\in G}\sum_{z\in G}\delta(y^{-1}xy \in K, z^{-1}xz \in H) \\ = & \frac{1}{|G|}\sum_{x\in G}\frac{1}{|K|}\sum_{y\in G}\delta(y^{-1}xy \in K)\frac{1}{|H|}\sum_{z\in G}\delta(z^{-1}xz \in H) \\ = & \frac{1}{|G|}\sum_{x\in G}\frac{1}{|K|}\sum_{y\in G}\delta(xyK = yK)\frac{1}{|H|}\sum_{z\in G}\delta(xzH = zH) \\ = & \frac{1}{|G|}\sum_{x\in G}\mathrm{Fix}_{G/K}x\mathrm{Fix}_{G/H}x \end{aligned}$$

as desired. □

W. Burnside's classic *Theory of Groups of Finite Order* introduces a matrix called there the "table of marks" associated to any finite group G. Let

$$G_1 = \{e\}, G_2, \ldots, G_s = G$$

be a sequence of representatives for conjugacy classes of subgroups of G ordered so that

$$|G_1| \leq |G_2| \leq \cdots \leq |G_s|,$$

and let O_i be a G-orbit of G_i-type, e.g., G/G_i. The "table of marks" is then the $s \times s$ matrix $B = [m_{ij}]$ where m_{ij} is just the number of elements of O_i fixed by G_j. The stabilizers of elements of O_i are conjugates of G_i, so $m_{ij} = 0$ unless G_j is contained in some conjugate of G_i; in particular $m_{ij} = 0$ if $i < j$. On the other hand m_{ii} is always positive since O_i certainly contains at least one G_i-stable element. Thus $[m_{ij}]$ is a lower triangular matrix with non-zero entries on the diagonal, and the determinant of $[m_{ij}]$ does not vanish.

Now let $C = [c_{in}]$ be the $s \times |G|$ matrix, with rows indexed by the conjugacy classes of subgroups of G, columns indexed by the elements of G, and c_{in} equal to the number of elements fixed when g_n acts on O_i. The rows of C are by definition the permutations characters of G. The n^{th} column of C coincides with the column of the table of marks corresponding to the (conjugacy class of) the subgroup $\langle g_n \rangle$ of G generated by g_n, since $\langle g_n \rangle$ and g_n leave the same elements fixed. The rank of C is thus the number of conjugacy classes of cyclic subgroups of G.

By the lemma, CC^T is the $s \times s$ matrix $D = [d_{ij}]$, with rows and columns indexed by conjugacy classes of subgroups of G and $d_{ij} = |G|o(G_i, G_j)$. The rank of D cannot be more the rank of C, so for D to have full rank, every subgroup of G, including G itself, must necessarily be cyclic.

STATISTIQUES PERMUTATIONNELLES ET MULTIPERMUTATIONNELLES

Dominique Foata
Département de mathématiques
Université Louis Pasteur
Strasbourg

RÉSUMÉ

L'introduction des monoïdes partiellement commutatifs en 1969 a d'une part servi de modèle algébrique au parallélisme, d'autre part, permis de construire des transformations sur les classes de réarrangements de mots, envoyant des statistiques sur d'autres statistiques. Ce modèle a été repris récemment par Guo-Niu Han, qui a défini un nouveau type de commutation permettant d'envoyer une bi-statistique sur une autre. On se propose de faire le point sur les constructions de ces transformations.

SERGEY FOMIN*
Theory of Algorithms Laboratory
Institute of Informatics, Russian Academy of Sciences
St.Petersburg 199178, Russia

Dual graphs and Schensted correspondences

Abstract

A graph is said to be graded if its vertices are divided into *levels* numbered by integers, so that the endpoints of any edge lie on consecutive levels.

The following three types of problems are considered:

(1) path counting in graded graphs, and related combinatorial identities;
(2) bijective proofs of these identities;
(3) design and analysis of algorithms establishing corresponding bijections.

The R.P.Stanley's [St88,St90] linear-algebraic approach to (1) is extended to cover a wide range of graded graphs. The main idea is to consider the *pairs* of graded graphs with the common set of vertices and common rank function. Such graphs are said to be *dual* if the associated linear operators satisfy a certain *commutation relation* (the "Heisenberg" one). The algebraic consequences of these relations are then interpreted as combinatorial identities. (This idea is also implicit in [St90].)

Applications include various examples of graded graphs, e.g., the Young, Fibonacci, Young-Fibonacci and Pascal lattices, the graph of shifted shapes, the r-nary trees, the subword order, the lattice of finite binary trees, etc. Many enumerative identities (both known and unknown) are obtained.

These identities can also be derived in a purely combinatorial way by generalizing the Robinson-Schensted correspondence to the class of graphs under consideration(cf. [Fo86]). The same tools can be applied to permutation enumeration, including Ferrers boards and involution counting. The bijective correspondences mentioned above are effected by the RSK-type algorithms a general approach to which is given. As particular cases of the construction we rederive the classical algorithm of Robinson, Schensted, and Knuth [Sc61,Kn70,Sc77], the Sagan-Worley [Sa87,Wo84] and Haiman's [Ha89] algorithms, the algorithm for the Young-Fibonacci graph [Fo86, Ro91]. Among new applications there are RSK-analogues for the infinite binary tree, the Pascal graphs, etc.

*Partially supported by the Mittag-Leffler Institute

An extended version of this contribution appeared in the Mittag-Leffler Institute preprint series [Fo1, Fo2]; the paper [Fo3] presents a related approach to Knuth-type correspondences and respective Schur functions.

1. Graded Graphs

A *graded graph* is a triple $G = (P, \rho, E)$ where

(1) P is a discrete set of *vertices*,

(2) $\rho : P \to \mathbb{Z}$ is a *rank function*,

(3) E is a multiset of *arcs/edges* (x, y) where $\rho(y) = \rho(x) + 1$.

The set $P_n = \{x : \rho(x) = n\}$ is called a *level* of G. We shall always assume that the levels are finite and G has a zero $\hat{0}$, i.e.,

$$P_0 = \{\hat{0}\}, \quad P_{-1} = P_{-2} = P_{-3} = \cdots = \phi.$$

Various examples of graded graphs can be found in [St86,Pr89,etc.].

G can be regarded either as oriented or as non-oriented graph. The accessibility relation in an oriented graph defines a *partial order* on P. If there are no multiple arcs, G turns out to be the Hasse diagram of this poset. Therefore non-oriented paths (i.e. paths in non-oriented graph) are often called *Hasse walks.*

Let $e(x \to y)$ denote the number of *oriented* paths between x and y; also write $e(y) = e(\hat{0} \to y)$. So $e(y)$ is the number of paths going from $\hat{0}$ to y ("a generalized binomial coefficient"). One can similarly define $e(x \to y \to z)$ et al.

The main results we are going to obtain are combinatorial identities involving $e(x)$ and similar enumerative characteristics. The typical example is the Young-Frobenius identity

$$\sum_{x \in P_n} e(x)^2 = n! \tag{1.1}$$

for the Young's lattice. This is not an isolated result. Surprisingly, the same formula proves to be valid for the so-called Fibonacci lattices [St75,Fo86,St88]. Similar identity (with additional coefficients) is known for the graph of shifted shapes (see (4.3)). Another examples are the lattice of binary trees and the binary subword order where

$$\sum_{x \in P_n} e(x) = n! \quad . \tag{1.2}$$

Each of these facts is known to have both "computational" and "bijective" proofs. However these proofs use *individual* properties of the graphs (except for the proof of (1.1) in [St88,Fo86]).

We develop combinatorial techniques providing general results of this type for a wide class of graded graphs. Unified proofs of (1.1), (1.2), and many other both known and unknown enumerative identities (cf., e.g., [St88] and [Sa90]) are given. Then we present a general approach to the Robinson-Schensted-type algorithms *for the same class of graphs.* It allows to provide bijective proofs to the enumerative results of this paper.

2. Linear-Algebraic Approach

A graded graph is completely determined by the *adjacency matrices* of the bipartite graphs formed by consecutive levels and the arcs joining them. Thus graded graphs can be studied as linear-algebraic objects.

Fix a field K of zero characteristic. The finitary K-valued functions on P (or, equivalently, the formal linear combinations of the vertices) form the vector space KP. The space KP can also be equipped with a Euclidean structure by declaring the vertices to form an orthonormal system. However we don't exploit this Euclidean structure; only linear techniques are used.

2.1 Definition. Let $G_1 = (P, \rho, E_1)$ and $G_2 = (P, \rho, E_2)$ be a pair of graded graphs with the common set of vertices and common rank function. Define an *oriented graded graph* $G = (P, \rho, E_1, E_2)$ by directing the G_1-edges "upwards" and the G_2-edges "downwards" (i.e. according to the rank increase or decrease, respectively). Now it is natural to introduce the *up* and *down operators* $U, D \in End(KP)$ associated with the graph G (or with the pair (G_1, G_2)) by

$$Ux = \sum_y a_1(x,y)y \quad ,$$
$$Dy = \sum_x a_2(x,y)x$$

where $a_i(x, y)$ is the multiplicity (weight) of the edge (x, y) in the [non-oriented] graph G_i.

Many path counting characteristics can be easily expressed in terms of operators U and D. For example, let $x \in P_n$. Then

$$U^k x = \sum_{y \in P_{n+k}} e_1(x \to y)y \quad ,$$
$$D^k x = \sum_{y \in P_{n-k}} e_2(y \to x)y$$

where indices in e_1 and e_2 refer to path counting in G_1 and G_2, respectively.

Non-oriented paths (Hasse walks) having arbitrary (but fixed) structure can be dealt with in the same manner.

2.2 Definition. Let G be defined by Definition 2.1. Assume w is a word in the alphabet $\{U, D\}$ ($\{U, D\}$-word, for short). A path p in G is said to *have a structure* w (or to be a w-path) if its consecutive arcs are directed upwards/downwards in accordance with the word w. So U's correspond to up-directed arcs (remind they should be the arcs of G_1) and D's to the down-directed arcs (those of G_2). We emphasize that in this definition a word w is to be read *from the right to the left* since it will be later interpreted as an operator.

Using this terminology, we can say that the number of w-paths in G from x to y is a coefficient of y in the expansion of wx where w is interpreted as an operator in KP (a product of corresponding U's and D's). For example, the coefficient of $\hat{0}$ in $D^n U^n \hat{0}$ is

$$\sum_{x \in P_n} e_1(x) e_2(x) \ .$$

We study the situations when the maps U and D satisfy some algebraic conditions. Algebraic consequences of those conditions can be then restated as combinatorial identities.

2.3 Definition. Let $r \in \mathbb{P}$. Graded graphs G_1 and G_2 with common set of vertices and common rank function (see Definition 2.1) are said to be r-*dual* (or simply *dual* when $r = 1$) if

$$DU = UD + rI \ . \tag{2.1}$$

It means combinatorially that

(1) if x and y are *different* vertices of the same rank then the number of DU-paths from x to y equals the number of UD-paths joining x and y;
(2) for any vertex x the number of DU-paths (loops) from x to x equals the number of UD-loops plus r.

Note that (2.1) is symmetric (invariant) with respect to an interchange of the initial graphs G_1 and G_2. Indeed, a change of D and U into U^* and D^*, respectively, transforms (2.1) into *equivalent* relation

$$U^* D^* = D^* U^* + I$$

(the $*$ stands for a conjugation with respect to a natural pairing in KP).

The case $U^* = D$ or, equivalently, $D^* = U$, $G_1 = G_2$, corresponds to self-dual graphs (differential posets; see Sec.3).

In what follows we assume that the graphs G_1, G_2, G, the operators U and D, etc., are as in Definition 2.1 and satisfy (2.1).

Let $r \in \mathbb{P}$. Define an associative graded algebra $\mathfrak{A}_r$ with identity I generated by elements U and D satisfying $DU = UD + rI$. Given a pair of r-dual graded graphs G_1 and G_2, a natural representation of $\mathfrak{A}_r$ arises. Thus any equality (identity) in $\mathfrak{A}_r$ can be reinterpreted in combinatorial terms. The following are typical examples of such identities.

2.4 Lemma. [St88] *Let k and l be nonnegative integers. Then*

$$D^l U^{k+l} = U^k (UD + (k+1)r) \dots (UD + (k+l)r) \ . \tag{2.2}$$

This lemma can be generalized in the following way.

2.5 Theorem. *For any* $\{U,D\}$*-word* $w = w(U,D)$ *with* m *entries of* D *and* n *entries of* U

$$w(U,D) = \sum_k r^k d_k(w) U^{n-k} D^{m-k}$$

where $d_k(w)$ *is the* k*'th coefficient of the rook polynomial of the Ferrers board (cf.* [St86, Sec.2.4]*) whose boundary is defined by* w*.*

Identities of this type can be converted into enumerative formulae concerning path counting (taking in mind the natural representation of the algebra $\mathfrak{A}_r$ in the space KP). Consider Lemma 2.4 as a typical example.

2.6 Corollary. *For any vertex* $x \in P$ *of rank* k

$$\sum_{y \in P_{k+l}} e_1(y) e_2(x \to y) = e_1(x) r^l (k+l)!/k! \ .$$

Proof. Apply both sides of (2.2) to $\hat{0}$ (note $D\hat{0} = 0$!) and take the coefficient of x.□
Now we state separately the special case $x = \hat{0}$ of Corollary 2.6.

2.7 Corollary.

$$\sum_{y \in P_l} e_1(y) e_2(y) = r^l \ l! \qquad (2.3)$$

We shall demonstrate that (2.3) generalizes (1.1), (1.2), and (4.3). Respective versions of Corollary 2.6 reduce to enumerative formulae involving "skew tableaux".

3. Self-Dual Graphs (Differential Posets)

Let Q be a countable poset. The distributive lattice $J(Q)$ of finite order ideals of Q (cf. [St86, Sec.3.1]) is a graded graph with zero; rank of an ideal is its cardinality.

3.1 Example. *Young graph* [St86, etc.] Let $\mathbb{P}^2 = \{(i,j) : i > 0, \ j > 0; \ i,j \in \mathbb{Z}\}$ be the two-dimensional integral quadrant with the usual (i.e. coordinatewise) partial order. The graph $\mathbb{Y} = J(\mathbb{P}^2)$ is called the *Young graph/lattice*; see Fig.1. The numbers $e(x)$ are the dimensions of irreducible representations of symmetric groups S_n (see, e.g., [St71, Sec.17]).

The Young graph is a *self-dual* graph = differential poset[St88]. Roughly, it is because for each shape there is one more box one can add than one can delete.

3.2 Theorem. [St88, Prop.5.5] *The Young graph is the only self-dual distributive lattice.*

3.3 Example. *Young-Fibonacci graph* [Fo86,St88]. Let $\{1,2\}^*$ denote the set of all the words in the alphabet $\{1,2\}$. Define the *Young-Fibonacci graph* $\mathbb{YF}$ (see Fig.2) as follows:

(1) $\{1,2\}^*$ is the set of its vertices (i.e., vertices are $\{1,2\}$-words);

(2) w' covers w iff either $w' = 1w$ (concatenation) or $w' = 2v$ where w covers v.

So the rank of a $\{1,2\}$-word is the sum of its "digits". We have mentioned yet that (1.1) is known to hold for the Young-Fibonacci graph. In [Fo86,Ro91] the complete analogue of Robinson-Schensted for $\mathbb{YF}$ was also constructed. The *self-duality* of the Young-Fibonacci graph follows from observing that

(1) $\mathbb{YF}$ is a modular lattice [Fo86,St88];

(2) every vertex of $\mathbb{YF}$ has one more successors than predecessors.

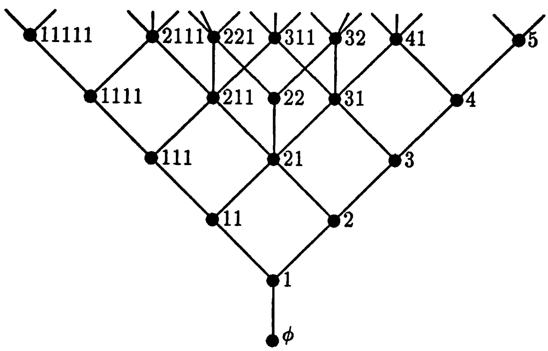


Fig.1 Young graph

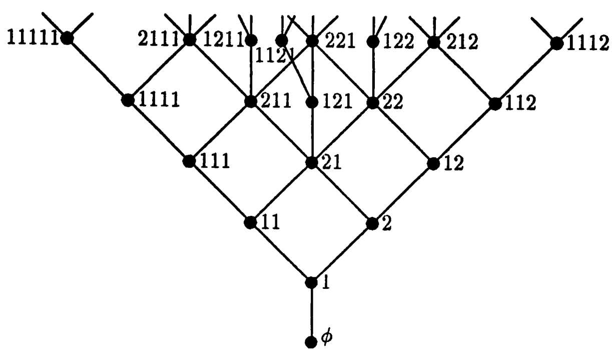


Fig.2 Young-Fibonacci graph

4. Weighted Distributive Lattices

Assume $G_1 = (P, \rho, E_1)$ is a distributive lattice: $P = J(Q)$. Let us try to construct the r-dual graph G_2 as follows. Let $w : Q \to K$ be a weight function (recall K is the basic field). For any edge $(x, y) \in E_1$ consider the only element $q \in y \setminus x$ and define $w(q)$ to be a multiplicity (weight) of an edge (x, y) in G_2. Thus G_2 is a *weighted distributive lattice.* (Generally, G_2 is not even a graph but a graded network.)

We want G_1 and G_2 to satisfy (2.1). Observe that whichever weight function w you choose the matrix $DU - UD$ is *diagonal.* Thus we only need to prove the equality of the diagonal elements of both sides of (2.1).

4.1 Lemma. *Assume $G_1 = J(Q)$. A weight function $w : Q \to K$ defines a weighted distributive lattice G_2 which is dual to G_1 if and only if the following condition holds for any finite order ideal x of Q:*

(4.1) *the total weight of the elements $q \in Q$ able to be added to x (i.e. such that $x \cup \{q\}$ is also an ideal) is r more than the total weight of the elements $q \in Q$ able to be deleted from x.*

The latter condition is very strong; only some special posets Q allow a weight function w to be defined satisfying (4.1).

In the remainder of this section we give several examples.

4.2 Example. *The chain* $\mathbb{N} = \{0, 1, 2, \dots\}$. This graph can be treated as a lattice of ideals of $\mathbb{P} = \{1, 2, 3, \dots\}$ with the usual ordering. The only weight function w on $\mathbb{P}$ satisfying (4.1) with $r = 1$ is $w(q) = q$. Thus the graph dual to $\mathbb{N}$ is the same chain having its qth edge multiplied by q.

4.3 Example. *Pascal graphs.* Let $Q = r\mathbb{P}$, i.e., Q is a union of r disjoint copies of $\mathbb{P}$. Then $\mathbb{N}^r = J(Q)$ is an r-dimensional *Pascal graph.* So $\mathbb{N}^r$ is the lattice of r-dimensional vectors with nonnegative integer coordinates.

To construct an r-dual graph for $\mathbb{N}^r$ (the coincidence of these two r's is *not* accidental) let us make the following general observation.

4.4 Lemma. *Assume the graphs G_1 and G_2 are r-dual, and the graphs H_1 and H_2 are s-dual. Then $G_1 \times H_1$ and $G_2 \times H_2$ are $(r+s)$-dual.*

Thus the r-dual graph for $\mathbb{N}^r$ can be obtained by cartesian multiplication of r copies of a graph dual to $\mathbb{N}$ (see Example 4.2).

By means of Lemma 4.4 one can easily construct new pairs of r-dual graphs from old ones; e.g., the graph $\mathbb{Y}^r$ (the rth cartesian power of the Young graph) is self-r-dual, as is $\mathbb{YF}^r$.

4.5 Example. *Diagrams with $\leq r$ rows.* Let Q be a direct product of an infinite chain $\mathbb{P}$ and a finite chain $[r] = \{1, \dots, r\}$. The distributive lattice $J(Q) = J(\mathbb{P} \times [r])$ is the lattice of Young diagrams containing r rows or less. It is of course a sublattice and an order ideal of the Young lattice. So the constants $e(x)$ are the same as in $\mathbb{Y}$.

4.6 Lemma. *A weight function* $w : \mathbb{P} \times [r] \to K$ *defined by*

$$w((q_1, q_2)) = r + q_1 - q_2 \tag{4.2}$$

satisfies (4.1).

Thus we have constructed a weighted lattice that is r-dual to $J(\mathbb{P} \times [r])$. Note that the r's involved in $\mathbb{P} \times [r]$, (4.2), and (2.1) do coincide.

4.7 Example. *Shifted shapes* [Sa87,Wo84,St90]. Let $Q = J(\mathbb{P} \times [2]) = \mathit{SemiPascal}$ (cf. Example 4.5). The graph $\mathbb{SY} = J(Q)$ is the *graph of shifted shapes* which are the Young diagrams with *strictly* decreasing row lengths; see Fig.3. Since $\mathbb{SY}$ is *not* an order ideal of the Young lattice, the values $e(x)$ in $\mathbb{SY}$ differ from those in $\mathbb{Y}$, x being a shifted shape. The main identity involving $e(x)$'s in $\mathbb{SY}$ is the following [Sc11,Sa79]:

$$\sum_{x \in P_n} e(x)^2 2^{n-h(x)} = n! \tag{4.3}$$

where $h(x)$ is the *height* (=number of rows) of the shape x.

The only weight function on *SemiPascal* satisfying (4.1) with $r = 1$ is

$$w(q) = \begin{cases} 1\,, & \text{if } q \text{ is a diagonal element} \\ 2\,, & \text{otherwise} \end{cases} \tag{4.4}$$

Thus the dual graph for $\mathbb{SY}$ is the same graph having doubled the edges which correspond to adding non-diagonal elements; see Fig.3.

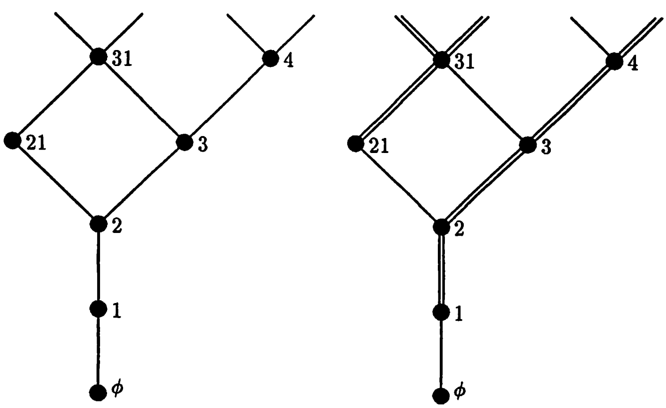


Fig.3 Graph of shifted shapes and its dual

5. Derivations in Graded Algebras

Let A be a graded associative K-algebra with identity; so A (as a vector space) is a direct sum of "homogeneous" subspaces A_n, $n \in \mathbb{N}$; and if $a \in A_n$ and $b \in A_m$ then $ab \in A_{n+m}$. Assume $D : A \to A$ is a derivation, i.e., a linear endomorphism satisfying

$$D(ab) = aD(b) + D(a)b \quad ;$$

assume, in addition, that $rank(D(a)) = rank(a) - 1$. Suppose $t \in A$ is such that $D(t) = r \cdot id$ where id stands for an identity element of A; of course $t \in A_1$. (Informally, D is a "derivation with respect to t/r".) Hence the operator $U \in End(A)$ defined by $U(a) = ta$ and the derivation operator D satisfy the condition (2.1) where $I \in End(A)$ is an identity transformation. (One can also take the *right* multiplication $U(a) = at$ instead.) In case the homogeneous subspaces A_n are finite-dimensional we can fix arbitrary bases in them to obtain a pair of r-dual networks/graphs.

Now we re-construct several examples of Sec.3-4 by means of this tool.

5.1 Example. *The chain* $\mathbb{N} = \{0, 1, 2, \dots\}$ (cf. Example 4.2). Let $A = K[t]$ be the algebra of polynomials in the variable t, and $rank(t^n) = n$. Then U is the operator of multiplication by t, and D is $\frac{d}{dt}$. Now take t^n as the basic vectors to obtain Example 4.2.

5.2 Example. *Young graph* (cf. Example 3.1). Let A be the algebra of symmetric functions, i.e., the symmetric formal power series of bounded degree in commuting variables $x_1, x_2, \dots$ This algebra can be alternatively described as the algebra $K[t_1, t_2, \dots]$ of polynomials in the variables t_n satisfying $rank(t_n) = n$. Now let D be the derivation with respect to t_1 and U the multiplication by t_1 (so $t = t_1$). Since $D(t_1) = 1$, we have a pair of dual graded graphs/networks; to make the construction explicit bases in A_n's should be chosen. In case $t_n = \sum x_i^n$ the basis of the *Schur functions* (see, e.g., [Ma79]) gives rise to the construction of Example 3.1 (the Young graph).

5.3 Example. *Pascal graphs* (cf. Example 4.3). Let A be the algebra $K[t_1, \dots, t_r]$ of polynomials in r commuting variables. Define

$$Uf(t_1, \dots, t_r) = (t_1 + \dots + t_r) f(t_1, \dots, t_r) \ , \tag{5.1}$$

$$D = \frac{\partial}{\partial t_1} + \dots + \frac{\partial}{\partial t_r} \quad . \tag{5.2}$$

The condition (2.1) holds, and we get the pair of Example 4.3.

5.4 Example. *Diagrams with* $\leq r$ *rows* (cf. Example 4.5). Here A is the algebra of symmetric polynomials in r commuting variables $t_1, \dots, t_r$. Like in Example 5.3, define U and D by (5.1) and (5.2). The result coincides with that of Example 4.5.

6. Bracket Tree, Binary Trees, and Subword Order

This section is devoted to constructing dual graphs for two remarkable rooted trees.

6.1 Example. *Lifted binary tree.* This is the infinite binary tree $\mathbb{T}_2$ with an additional edge attached below the old zero. The vertices of this tree can be naturally labelled by the binary notations of the nonnegative integers: 0, 1, 10, 11, 100, 101, 110, ... so that 0 is the root; 1 is the only vertex of rank 1; 10 and 11 are the vertices of rank 2, etc. (see Fig.5). Generally, the rank is the length of the notation, except for the case $\rho(0) = 0$.

Now define the graph *BinWords* (a *lifted binary subword order*; cf. [Bj91]; see Fig.5) with the same set of vertices, the same rank function, and the following covering relation: x covers y iff x can be obtained by deleting a single symbol from y. In addition, 1 covers 0. (For example, 101001 covers 11001, 10001, 10101, and 10100.) We emphasize that *BinWords* is a graph *without multiple edges.*

6.2 Lemma. *BinWords and the lifted binary tree are dual.*

6.3 Example. *Bracket Tree.* This tree (see Fig.4) is defined as follows. The vertices of rank n are the "syntactically correct" formulae defining different versions of calculation of *non-associative* product $x \cdot x \cdot \ldots \cdot x$ containing $n+1$ entries of x. We call such sequences the *bracket schemes.* Two schemes are linked in the tree if one of them results from another by deleting the first entry of x and subsequent removing the pair of "unnecessary" brackets. Indeed, it is a tree.

6.4 Lemma. *The Bracket Tree is dual to the distributive lattice $J(\mathbb{T}_2)$ of the finite order ideals of the infinite binary tree.*

This statement needs explanation: we should demonstrate that, in a sense, the Bracket Tree and $J(\mathbb{T}_2)$ have the same set of vertices (see, e.g., [SW86,Sec.3.1] for another proof). To do that, associate any bracket scheme with an appropriate parsing tree. Remove the leaves of this tree (they correspond to the entries of x) as well as the edges incident to them. As a result we obtain a *marked* binary tree, i.e., an order ideal of the infinite binary tree $\mathbb{T}_2$. See [St75] for additional information concerning $J(\mathbb{T}_2)$.

Now we can apply the enumerative results of Sec.2 to each of the examples listed above (the same is true concerning the generalized Schensted construction of Sec.8).

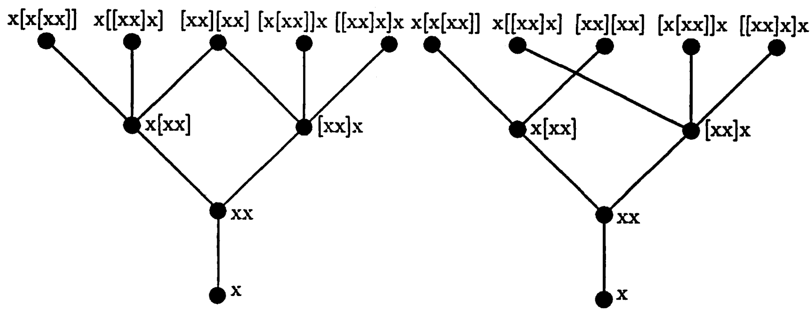


Fig.4 Lattice of binary trees and Bracket Tree

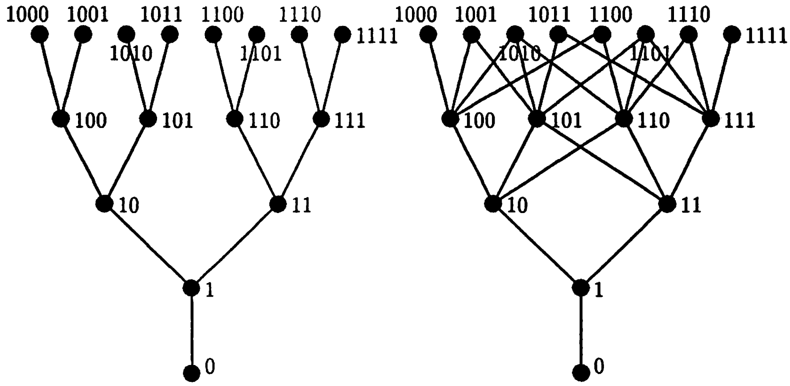


Fig.5 Lifted binary tree and BinWords

7. Some Applications

The identities of Sec.2 can be applied to each of the examples given above. The versions of Corollary 2.6 are stated below.

7.1 Corollary. *Assume G is a self-dual graph. Then*

$$\sum_{x \in P_n} e(x)^2 = n!$$

In particular, this is true for the Young graph $\mathbb{Y}$ and the Young-Fibonacci graph $\mathbb{YF}$.

7.2 Corollary. *Assume G_1 and G_2 are dual graphs, and G_2 is a tree. Then*

$$\sum_{x \in P_n} e_1(x) = n! \quad .$$

In particular, this is the case when G_1 is either BinWords or $J(\mathbb{T}_2)$.

7.3 Corollary. *Assume a weight function w defined on a poset Q satisfies the conditions of Lemma 4.1. Then in the distributive lattice $P = J(Q)$ one has*

$$\sum_{x \in P_n} e(x)^2 \prod_{q \in x} w(q) = r^n n! \ . \tag{7.1}$$

In particular, this identity holds in the following distributive lattices:

(1) *The graph of shifted shapes $\mathbb{SY}$: w is defined by (4.4);*
(2) *The graph of the diagrams with $\leq r$ rows: w is defined by (4.2);*
(3) *Pascal graph $\mathbb{N}^r$:*

$$w(q) = num(q)$$

where $q \in r\mathbb{P}$ is an integer $num(q)$ taken from a certain chain $\mathbb{P}$.

7.4 Comments.

1. In the case of the graph of shifted shapes $\mathbb{SY}$ the identity (7.1) reduces to (4.3).

2. In the case of the graph of r-row diagrams (7.1) turns into

$$\sum_{x=(x_1,\ldots,x_r)\in P_n} e(x)^2 \prod_{j=1}^{r} \frac{(r+x_j-j)!}{(r-j)!} = r^n n!$$

3. In the case of the Pascal graph $\mathbb{N}^r$ a direct simplification transforms (7.1) into the classical $\sum e(x) = r^n$, where $e(x)$ denotes the appropriate multinomail coefficient.

8. Generalized Schensted

All the enumerative identities that can be derived from

$$DU = UD + rI \tag{8.1}$$

can also be proved combinatorially, i.e., by establishing certain explicit bijections. To do this, one can start with fixing a bijection between the objects which are counted in the left-hand and the right-hand sides of (8.1). Then, to get a bijective proof of any identity involving U's and D's (say, (2.2)), one can write a sequence of elementary algebraic transformations which prove this identity (that is, at each step a single substitution $DU \leftarrow UD + rI$ is performed) and then proceed step by step along the lines of this algebraic proof by applying "bijective version" of $DU = UD + rI$ that we fixed in advance. Since some of the involved elementary transformations commute, the resulting algorithm is *parallel.*

This construction, when applied to the identity

$$D^n U^n = (UD + r)(UD + 2r)\ldots(UD + nr)$$

(cf. (2.2)), results in a generalized Schensted algorithm which establishes bijection between pairs of paths and permutations (in case $r > 1$ the latters are colored in r colors). In other words, we obtain bijective proofs of respective versions of Corollary 2.7.

We conclude by describing the basic construction in explicit algorithmic terms.

8.1 Definition. Assume, as before, that G_1 and G_2 are r-dual graded graphs. Let $\Phi = \{\Phi_{xy}\}_{x,y\in P}$ be a family of bijections between the objects counted by "the (x, y)'th matrix elements" of DU and $UD + rI$, respectively, i.e., bijections from

$$B_{xy} = \{(b_1, b_2) : b_1 \in E_1, b_2 \in E_2, end(b_1) = end(b_2), start(b_1) = y, start(b_2) = x\}$$

to

$$A_{xy} = \{(a_1, a_2) : a_1 \in E_1, a_2 \in E_2, start(a_1) = start(a_2), end(a_1) = x, end(a_2) = y\} \cup\{1, \ldots, r\}\ .$$

Such a family is called an *r-correspondence.* Since x and y are determined by either a_i or b_i, we may unambigously regard Φ as a single bijection

$$\Phi: \bigcup B_{xy} \longrightarrow \bigcup A_{xy} .$$

Once an r-correspondence Φ is fixed for a given pair of dual graded graphs, a bijective correspondence between pairs of paths and colored permutations arises.

8.2 Algorithm. *("Generalized Schensted: tableaux to permutations")*

Input:

(i) edges $t_1(1)$, $t_1(2)$, ... , $t_1(n)$ forming a path in G_1 starting at $\hat{0}$;

(ii) edges $t_2(1)$, $t_2(2)$, ... , $t_2(n)$ forming a path in G_2 starting at $\hat{0}$ and having common endpoint with the first path.

Output: colored permutation σ (matrix $n \times n$ with exactly one nonzero element in each row and column; this element should be one of 1, ... ,r).

```
var
  φ1 : array [1..n,0..n] of ( edge of G1  or nil) ;
  φ2 : array [0..n,1..n] of ( edge of G2  or nil) ;
  σ : array [1..n,1..n] of integer;
  a1, a2, b1, b2, k, l : integer;
begin
  for k := 1 to n do φ1(k,n) := t1(k);
  for l := 1 to n do φ2(n,l) := t2(l);
  for (k,l) := (n,n) downto (1,1) do
    begin
      b1 := φ1(k,l) ;   b2 := φ1(k,l);
      case
        b1 = nil or b2 = nil  ⟹  a1 := b1;  a2 := b2;  σ(k,l) := 0;
        Φ(b1,b2) ∈ E1 × E2  ⟹  (a1,a2) := Φ(b1,b2);  σ(k,l) := 0;
        Φ(b1,b2) = i ∈ {1,...,r}  ⟹  a1 := nil;  a2 := nil;  σ(k,l) := i
      endcase ;
      φ1(k,l−1) := a1 ;   φ2(k−1,l) := a2
    end
end
```

Comments: In the third **for**-cycle (n^2 repetitions) a pair (k_1, l_1) should be treated *after* (k_2, l_2) whenever $k_1 \leq k_2$ and $l_1 \leq l_2$. Calculations may be done in parallel regarding this condition.

One can also write, in the same manner, an algorithm realizing the reverse bijection.

8.3 Algorithm. *("Generalized Schensted: permutations to tableaux")*

Input = Output of Algorithm 8.2.

Output = Input of Algorithm 8.2.

```
var
  φ1, φ2 : ... ; {see Algorithm 8.2}
  a1, a2, b1, b2, k, l : integer;
begin
  for k := 1 to n do φ1(k,0) := (0̂,0̂);
  for l := 1 to n do φ2(0,l) := (0̂,0̂);
  for (k,l) := (1,1) to (n,n) do
    begin
      a1 := φ1(k,l-1) ;   a2 := φ1(k-1,l);
      case
        (a1 = nil or a2 = nil) and σ(k,l) = 0  ==>  b1 := a1;  b2 := a2;
        (a1 = nil and a2 = nil) and σ(k,l) ≠ 0  ==>  (b1,b2) := Φ^-1(σ);
        (a1 ≠ nil and a2 ≠ nil)  ==>  (b1,b2) := Φ^-1(a1,a2)
      endcase ;
      φ1(k,l) := b1 ;   φ2(k,l) := b2
    end;
  for k := 1 to n do t1(k) := φ1(k,n);
  for l := 1 to n do t2(l) := φ2(n,l)
end
```

Both algorithms are essentially parallel. To get a sequential version of, e.g., Algorithm 8.2, replace

for $(k, l) := (n, n)$ **downto** $(1, 1)$ **do**

by

for $k := n$ **downto** 1 **do**

  for $l := n$ **downto** 1 **do**

where the last two loops can be interchanged as well. The interior **for**-loop is an analogue of the Schensted insertion and reduces to the latter in the case of the Young lattice.

For the graph of shifted shapes the "row-wise" and "column-wise" sequential versions coincide with the algorithms of Sagan-Worley [Sa87,Wo84] and Haiman [Ha89], respectively (provided the natural r-correspondence is chosen).

The applications to rim hook tableaux are given in [FS92].

For other pairs of dual graphs (including those listed above) respective Schensted analogues can be constructed by specializing the general scheme; see [Fo2]. A unified approach to Knuth-type algorithms (cf. [Kn70]) is suggested in [Fo3]. Other identities in the algebra $\mathfrak{A}_r$ give rise to corresponding modifications of "generalized Schensted", like its "skew version" that extends the constructions of [SS90] (see [Fo2,Fo3] for details).

ACKNOWLEDGEMENTS. I am grateful to Curtis Greene and Dennis Stanton for several helpful comments.

References

[Bj91] A.Björner, The Möbius function of subword order, *Invariant Theory and Tableaux,* D.Stanton, Ed., Springer-Verlag, 1990, 118-124.

[Fo86] S.V.Fomin , Generalized Robinson-Schensted-Knuth correspondence [in Russian], *Zapiski Nauchn. Sem. LOMI* **155** (1986) , 156-175.

[Fo1] S.Fomin, Duality of graded graphs, Report No.15 (1991/92), *Institut Mittag-Leffler,* 1992; submitted to *J.Algebr.Combinatorics.*

[Fo2] S.Fomin, Schensted algorithms for dual graded graphs, Report No.16 (1991/92), *Institut Mittag-Leffler*, 1992; submitted to *J.Algebr.Combinatorics.*

[Fo3] S.Fomin, Schur operators and Knuth correspondences, Report No.17 (1991/92), *Institut Mittag-Leffler*, 1992; submitted to *J.Combin.Theory*, Ser.A.

[FS92] S.Fomin, D.Stanton, Rim hook lattices, preprint, 1992.

[Ha89] M.D.Haiman , On mixed insertion, symmetry, and shifted Young tableaux, *J.Combin. Theory,* Ser.A **50** (1989), 196-225.

[Kn70] D.E.Knuth , Permutations, matrices, and generalized Young tableaux, *Pacific J.Math.* **34** (1970), 709-727.

[Ma79] I.Macdonald, Symmetric functions and Hall polynomials, Oxford Univ. Press , Oxford, 1979.

[Pr89] J.Propp , Some variants of Ferrers diagrams, *J.Combin.Theory, Ser.A* **52** , 1989, 98-128.

[Ro91] T.W.Roby , Applications and extensions of Fomin's generalization of the Robinson-Schensted correspondence to differential posets, Ph.D.thesis, M.I.T., 1991.

[Sa79] B.E.Sagan , An analog of Schensted's algorithm for shifted Young tableaux, *J. Comb. Theory,* Ser.A **27** (1979), 10-18.

[Sa87] B.E.Sagan , Shifted tableaux, Schur Q-functions and a conjecture of R.Stanley, *J. Comb. Theory,* Ser.A **45** (1987), 62-103.

[SS90] B.E.Sagan, R.P.Stanley , Robinson-Schensted algorithms for skew tableaux, *J.Combin. Theory,* Ser.A **55** (1990), 161-193.

[Sc11] I.Schur , Über die Darstellung der symmetrischen und der alternienden Gruppe durch gebrechene leneare Substitutionen, *J.Reine Angew. Math.* **139** (1911), 155-250.

[Sc61] C.Schensted , Longest increasing and decreasing subsequences, *Canad. J. Math.* **13** (1961), 179-191.

[Sc77] M.-P.Schützenberger , La correspondence de Robinson, in *Combinatoire et représentation du groupe symétrique,* D.Foata ed., Lecture Notes in Math. **579** (1977), 59-135.

[St71] R.P.Stanley , Theory and application of plane partitions, Parts 1 and 2, *Studies in Applied Math.* **50** (1971), 167-188, 259-279.

[St75] R.P.Stanley , The Fibonacci lattice, *Fibonacci Quart.* **13** (1975), 215-232.

[St86] R.P.Stanley , Enumerative combinatorics, vol.I, Wadsworth and Brooks/Cole, Pacific Grove, CA, 1986.

[St88] R.P.Stanley , Differential posets, *J.Amer.Math.Soc.* **1** (1988) , 919-961.

[St90] R.P.Stanley , Variations on differential posets , *Invariant Theory and Tableaux,* D.Stanton, Ed., Springer-Verlag, 1990, 145-165.

[SW86] D.Stanton, D.White, Constructive combinatorics, Springer-Verlag, 1986.

[Wo84] D.R.Worley , A theory of shifted Young tableaux, Ph.D.thesis, M.I.T., 1984.

THEORY OF ALGORITHMS LABORATORY, INSTITUTE OF INFORMATICS, RUSSIAN ACADEMY OF SCIENCES, 14TH LINE 39, ST.PETERSBURG 199178, RUSSIA. E-MAIL: SFOMIN@IIAS.SPB.SU

The asymptotic behaviour of coefficients of large powers of functions.

Danièle GARDY *
L.R.I., CNRS-URA 410,
Bât. 490, Université Paris XI,
91405 Orsay Cedex, France

Abstract

We review existing results on the asymptotic approximation of the coefficient of order n of a function $f(z)^d$, when n and d grow large while staying roughly proportional. Then we present extensions of these results to allow more general relationships between n and d and to take into account a multiplicative factor $\psi(z)$.

1 Introduction

Generating functions of the type $\phi(z) = f(z)^d$, where f is a given function with positive coefficients and d is a parameter which tends to infinity, appear in several problems of discrete probability theory, combinatorial enumeration, etc. These problems often require an estimate of the n^{th} coefficient of f^d, which we denote by $[z^n]\{f(z)^d\}$, for large n and d.

For example, let $X_1, \cdots, X_d$ be d random variables, independent and with the same probability distribution defined by the generating function $f(z)$. Their sum $S_d = \sum_{i=1}^{d} X_i$ has for generating function $f^d(z)$, whose coefficient of order n is the probability $\Pr(S_d = n)$. The average value of S_d is $d\, f'(1)$, and its variance is also of order d. The situations where $n = d\, f'(1) + o(\sqrt{d})$, $n = o(d)$ or $d = o(n)$ describe the behaviour of the sum respectively close to the mean (in a range where the central limit theorem applies), before or beyond the mean (in an area of large deviations).

Coefficients of the type $[z^n]\{f^d(z)\}$ appear for example in asymptotic coding theory [10], in the evaluation of some parameters on forests of trees [17, 21], in the evaluation of diagonal coefficients of some bivariate functions $F(z, u)$, and in a class of asymptotic

*This work was partially researched while visiting at the Computer Science Department of Brown University, Providence R.I. 02912 (U.S.A.), and partially supported by the PRC-CNRS Mathématique-Informatique and by ESPRIT-II Basic Research Action No. 3075 (project ALCOM).

distributions related to urn models, which require computing the coefficient $[y^n]\{f(x,y)^d\}$ of a bivariate function: See [14, 15] for a survey of results on urn models and [7, 9] for some applications to relational database theory. Related problems also appear in the evaluation of trie parameters [5] or of the number of lattice points in a ball [16, 20], and in the analysis of a random walk on an hypercube [4].

The present paper is intended as a survey of results on the asymptotic estimation of coefficients of the type $[z^n]\{f^d(z)\}$; it also presents some as yet unpublished results. Its plan is as follows: In order to unify the presentation, we introduce some notations, then recall the basis of the main technic (a saddle-point approximation) in Section 2. Section 3 presents results pertaining to the asymptotic approximation of $[z^n]\{f(z)^d\}$, for large n and d growing at a similar rate. We then extend these results to allow different growth rates for n and d. In Section 4, we allow a multiplicative factor $\psi(z)$ and study the coefficient $[z^n]\{f(z)^d\,\psi(z)\}$. Finally we indicate some applications, mostly related to urn models and Stirling numbers, in Section 5.

2 Notations and methods

2.1 Notations

We consider in this paper functions of one variable which have a power series expansion $f(z) = \sum_{k\geq 0} f_k z^k$. We assume in the sequel that the function f satisfies the following property:

Assumption $\mathcal{A}_1$:
The function f has real positive coefficients with $f_0 \neq 0$ and $f_1 \neq 0$, and a strictly positive, possibly infinite, radius of convergence R. Its coefficients are such that $GCD\{k : f_k \neq 0\} = 1$.

The condition on the GCD can be stated in an equivalent form: There exists no entire function g and no integer $m \geq 2$ such that $f(z) = g(z^m)$. The condition on f_0 simply means that when $f(z)$ has valuation p, we can factor out z^p: If $f(z) = z^p(f_0 + f_1 z + \cdots)$, then $f^d(z) = z^{dp}(f_0 + f_1 z + \cdots)^p$. The restriction on f_1 is a technical one, which might be removed, but this extension implies more restrictive conditions on the relative growths of n and d than those given in some theorems of this paper.

To simplify the notations in the sequel, we define two operators on a function f:

$$\Delta f(z) = z\frac{f'}{f}(z); \qquad \delta f(z) = \frac{f''}{f}(z) - \frac{f'^2}{f^2}(z) + \frac{f'(z)}{zf(z)}.$$

These operators are related by: $z\delta f(z) = (\Delta f)'(z)$. When the function f has real positive coefficients, it is not difficult to show that, for all real positive z smaller than R, the radius of convergence of f, *the value of $\delta f(z)$ is strictly positive and the function Δf is increasing.*

2.2 The saddle-point approximation

Before studying a function $f^d(z)$, we first recall results valid for any analytic function ϕ. Its coefficient of order n is given by Cauchy's formula, where the integration contour is a closed curve around the origin of the complex plane which stays inside the convergence domain:

$$[z^n]\phi(z) = \frac{1}{2i\pi}\oint \phi(z)\frac{dz}{z^{n+1}}.$$

We immediately deduce from it an upper bound $|[z^n]\phi(z)| \leq (1/2\pi)\oint|\phi(z)z^{-n-1}|dz$. Integrating on a circle of radius ρ smaller than the radius of convergence of ϕ gives $|[z^n]\phi(z)| \leq \phi(\rho)\rho^{-n}$, and the best (smallest) upper bound is obtained, when possible, for ρ such that $\rho\phi'(\rho)/\phi(\rho) = n$.

For example, let us assume that ϕ is the generating function of a random variable X, of mean μ, and let $n = (1+\delta)\mu$. Then $\Pr(X = n) = [z^n]\phi(z)$ is bounded from above by $\phi(\rho)\rho^{-n}$. Setting $\rho = e^t$ and using the fact that $\phi(e^t) = E(e^{tX})$, we get:

$$\Pr(X = (1+\delta)\mu) \leq \frac{E(e^{tX})}{e^{t(1+\delta)\mu}}.$$

This is Chernoff's bound, which often gives useful information on the probability that a random variable is at distance at least $1+\delta$ of its mean.

Now assume that X is itself obtained by summing d independent random variables with a common distribution: $\phi(z) = f^d(z)$. Then we have that $\Pr(X = n) \leq f^d(\rho)\rho^{-n}$, and this bound is tightest for ρ such that $\rho f'(\rho)/f(\rho) = n/d$.

The upper bound can be refined to give an approximation of $[z^n]\phi(z)$: Instead of bounding ϕ on the integration circle, we look closely at the points which give the main contribution to the integral. This is the basis of the saddle point method (see for example [3] for a general presentation and Hayman [13][22, Ch.5] for applications to the approximation of generating function coefficients). It turns out that, if we can choose for radius of the integration circle the point ρ defined by the equation $\rho\phi'(\rho)/\phi(\rho) = n$, the main part of the integral often comes from the vicinity of ρ, which is a *saddle point*. Defining $h(z) = \log\phi(z) - (n+1)\log z$, we get:

$$[z^n]\phi(z) = \frac{1}{2i\pi}\oint e^{h(z)}dz \approx \frac{e^{h(\rho)}}{\sqrt{2\pi h''(\rho)}} = \frac{\phi(\rho)}{\rho^{n+1}\sqrt{2\pi h''(\rho)}}.$$

This approximation holds for a large class of functions $\phi(z) = f^d(z)$, when n/d belongs to an interval $[a, b]$ $(0 < a < b)$ and $n, d \to +\infty$ [2, 11, 12].

The application of the saddle-point method to the asymptotic evaluation of coefficients of a function is closely related to Laplace's method for approximating an integral. Although this method is usually applied to integrals depending on one parameter, Fulks [6] and Pederson [19] have studied integrals depending on two parameters, which are in the same vein as the problem of evaluating $[z^n]\{f^d(z)\}$, where we have two parameters n and d.

3 Asymptotic approximations of coefficients

3.1 The case n constant

We include this case for the sake of completeness, although it presents no real difficulty. If n is constant, the saddle-point method does not work; however a direct analysis can give some information. For example, the following result simply means that the first coefficients of $f(z)^d$ behave as those of $(f_0 + f_1 z)^d$.

Theorem 1:
If the function f has real positive coefficients, such that $f_0 \neq 0$ and $f_1 \neq 0$, then, for $d \to +\infty$ and for any fixed n:

$$[z^n]\{f^d(z)\} = \binom{d}{n} f_0^{d-n} f_1^n (1 + O(1/d)).$$

This is proved by expanding the coefficient into a sum of (a fixed number of) multinomial coefficients, which are themselves easily approximated. If we allow n to grow, both the number of terms in the sum and the terms themselves are unbounded, and this proof no longer holds.

3.2 A general formula when $n = \Theta(d)$

The problem of finding the asymptotic value of $[z^n]\{f(z)^d\}$, when $n, d \to +\infty$ and n and d are roughly proportional, was studied for example by Daniels [2] and Greene and Knuth [12], mostly for probability generating functions. As noted by Good [11], this result is actually valid for a larger class of functions, such as entire functions or functions defined on an open disk; moreover it can be improved to give further terms of an asymptotic development. We give below the main result [2][11, p.868].

Theorem 2:
Let f be a function satisfying the assumption $\mathcal{A}_1$ of Section 2.1, and let R be its radius of convergence. Assume that n/d belongs to an interval $[a, b]$, $0 < a < b$, and that $n, d \to +\infty$. Define ρ and σ^2 by $\Delta f(\rho) = n/d$ and $\sigma^2 = \rho^2 \delta f(\rho)$. If $\rho < R$, then:

$$[z^n]\{f(z)^d\} = \frac{f(\rho)^d}{\sigma \rho^n \sqrt{2\pi d}}(1 + o(1)).$$

We can simplify Theorem 2 further if n/d has a finite, non null, limit:

Corollary 1:
Under the assumptions of Theorem 1, if there exist two real strictly positive constants k and m such that $n = kd + m$, then:

$$[z^n]\{f(z)^d\} = \frac{A^d}{B \rho_0^m \sqrt{d}}(1 + o(1)),$$

for suitable constants $A = f(\rho_0)\rho_0^{-k}$ and $B = \sigma\sqrt{2\pi}$, and with ρ_0 the solution (independent of n and d) of $\Delta f(z) = k$. Note that σ too is a constant: $\sigma^2 = \rho_0^2\,\delta f(\rho_0)$. If $n = kd$, i.e. $m = 0$, then we have the simpler formula:

$$[z^n]\{f(z)^d\} = \frac{A^d}{B\sqrt{d}}(1 + o(1)).$$

It is possible to get some information on the variation of the term A when the quotient n/d is finite and bounded away from 0. Let $A = A(k)$ with $k = n/d$. *On a closed interval of $]0, +\infty[$ including $\Delta f(1)$, $A(k)$ is a unimodal function of k, first increasing then decreasing , with a maximum $A(\Delta f(1)) = f(1)$.*

3.3 Function defined by an implicit equation

A recent paper by Meir and Moon [17] deals with the approximation of the coefficient of z^n in $f(z)^d$, when $d, n \to +\infty$ and $d = O(n)$, and with f defined by an implicit equation: $f(z) = z\phi(f(z))$ and $f(0) = 0$. This improves on a former result by Flajolet and Steyaert [21], which was proved for $d = o(n)$, more precisely for $d \le \sqrt{n}/\log^3(n)$. Meir and Moon give the following result:

Theorem 3:
Let ϕ be a function satisfying the assumption $\mathcal{A}_1$ of Section 2.1 and define a function f by $f(z) = z\phi(f(z))$ and $f(0) = 0$. Let $d = \alpha n + \lambda\sqrt{n} + o(\sqrt{n})$, with α a constant such that $0 \le \alpha < 1$ and that $(\Delta\phi)^{-1}(1 - \alpha)$ exists, and with λ a finite (positive or negative, possibly null) constant. Then, for $n, d \to +\infty$

$$[z^n]\{f(z)^d\} = \frac{d}{n\sigma\sqrt{2\pi n}} e^{-\lambda^2/2\sigma^2} \rho^{d-n}\phi(\rho)^n(1 + o(1)),$$

where ρ is defined by $\Delta\phi(\rho) = 1 - \alpha$ and σ^2 by

$$\sigma^2 = \rho^2\frac{\phi''(\rho)}{\phi(\rho)} + \alpha(1 - \alpha) = \rho^2\delta\phi(\rho).$$

Meir and Moon actually prove in passing the following result:

$$[t^n]\{\phi(t)^d\} = \frac{e^{-\lambda^2/2\sigma^2}}{\sigma\sqrt{2\pi d}}\,\frac{\phi(\rho)^d}{\rho^n}(1 + o(1)),$$

and their range of validity is for $n = (1 - \alpha)d + \lambda\sqrt{d} + o(\sqrt{d})$. For $\alpha > 0$, this is basically an extension of Theorem 2 (to allow $\lambda \ne 0$) applied to $n = kd + O(\sqrt{d})$ with a constant $k = 1 - \alpha$ in $]0, 1]$. Theorem 3 is then obtained by an application of the Lagrange inversion formula:

$$[z^n]\{f(z)^d\} = \frac{d}{n}[t^{n-d}]\{\phi(t)^n\}.$$

When $\alpha = 0$ but $\lambda > 0$, Theorem 3 gives an approximation valid for $d = \lambda\sqrt{n}(1 + o(1))$, i.e. $d^2 = \lambda^2 n(1 + o(1))$. If $\alpha = \lambda = 0$, the result holds for $d = o(\sqrt{n})$ i.e. $d^2 = o(n)$. This means that Meir and Moon have results for $d \approx \alpha n$ (when $\alpha \neq 0$) or for $d^2 = O(n)$ (when $\alpha = 0$). If $\lambda = 0$ and $\alpha \neq 0$, and if we have $f(z) = z^q g(z)$ with $g(0) \neq 0$, then either one of Theorem 2 or Theorem 3 can be applied indifferently to evaluate the coefficient $[z^n]\{f^d(z)\} = [z^{n-qd}]\{g^d(z)\}$, for $d = \alpha n + o(\sqrt{n})$.

3.4 The case $n = o(d)$

We study now the case where d and n both grow large, but n stays much smaller than d. Theorem 4 is an extension of Theorem 2 to the case $n = o(d)$.

Theorem 4:
Let f satisfy the assumption $\mathcal{A}_1$ of Section 2.1 and let $n = o(d)$, with $n, d \to +\infty$. Define ρ as the unique real positive solution of $\Delta f(z) = n/d$. Then:

$$[z^n]\{f(z)^d\} = \frac{f(\rho)^d}{\rho^n\sqrt{2\pi n}}(1 + o(1)).$$

Theorem 4 is proved by integrating on a circle going through the saddle point ρ, which becomes $o(1)$ for $n = o(d)$. The singularities are beyond the integration contour as soon as n and d are large enough. The detailed proof can be found in [8].

Theorem 4 is closely related to a result of Odlyzko and Richmond [18], which holds for a class of polynomials f, and for n and d such that, with q denoting the degree of f, $qd - n \to +\infty$. Their result covers the case $n = o(d)$, when f is the generating function of a probability distribution with finite support.

If we have more information on the respective orders of growth of n and d, we can obtain a useful approximation of the saddle point ρ and give a more precise form of Theorem 4 . The following corollary, for example, deals with the cases when $n = o(\sqrt{d})$ or $n = o(d^{2/3})$.

Corollary 2:
If f satisfies the assumption $\mathcal{A}_1$ of Section 3.1 and if $n = o(\sqrt{d})$, with $n, d \to +\infty$, then:

$$[z^n]\{f^d(z)\} = \frac{f_0^d}{\sqrt{2\pi n}}\left(\frac{e f_1 d}{f_0 n}\right)^n (1 + o(1) = \frac{f_0^d}{n!}\left(\frac{f_1 d}{f_0}\right)^n (1 + o(1)).$$

If we only have the weaker condition $n = o(d^{2/3})$, then:

$$[z^n]\{f^d(z)\} = \frac{f_0^d}{\sqrt{2\pi n}}\left(\frac{e f_1 d}{f_0 n}\right)^n \exp\left(\frac{n^2}{d}(\frac{f_2}{f_1} - \frac{1}{2})\right)(1 + o(1)).$$

Intuitively, Corollary 2 means that, when $n = o(\sqrt{d})$, the first two coefficients of f determine the main term in the asymptotic expression of the coefficients of f^d. This result can be compared to the relevant one for an affine function (although an affine function does not satisfy assumption $\mathcal{A}_1$): $[z^n]\{(f_0 + f_1 z)^d\} = \binom{d}{n} f_0^{d-n} f_1^n$; Stirling's formula for the factorials gives an approximation equivalent to the first one of Corollary 2. When n increases with respect to d, the other coefficients are progressively introduced. As long as some relationship $n^l = o(d^q)$ holds, it is possible to get a result similar to Corollary 2. This requires a good approximation of the saddle point ρ, and might become quite involved according to which coefficients of f are null, but it would be possible to work it out for a given function f. However, if for example $n = d/\log d$, we cannot find a relationship $n^l = o(d^q)$ and we have to take all the coefficients of f into account.

3.5 The case $d = o(n)$

When the function f satisfies some functional equation, the result of Meir and Moon presented in Section 3.3 can sometimes be applied. More generally, we can prove analogs of Theorems 2 and 4 for some classes of functions, using similar technics.

Theorem 5:
Let $f(z) = e^{P(z)}$, where $P(z) = \sum_{0 \le i \le q} P_i z^i$ is a polynomial of degree q with positive coefficients. Assume that the coefficients P_0 and P_1 are nonnull. If $n, d \to +\infty$ in such a way that $d = o(n)$, define ρ as the unique real positive solution of $zP'(z) = n/d$. Then:

$$[z^n]\{e^{dP(z)}\} = \frac{e^{dP(\rho)}}{\rho^n \sqrt{2\pi n}}(1 + o(1)).$$

If the function f is not entire, its singularities become important. For example, we can prove the following result for a meromorphic function with one pole on its circle of convergence:

Theorem 6:
Let f be a meromorphic function with positive coefficients, whose singularity of smallest modulus is a pole in 1: $f(z) = g(z)/(1-z)$, where g is a function analytic for $|z| \le 1$. Assume that $f_1 \neq 0$, and define ρ by $\Delta f(\rho) = n/d$. Then, if $d = o(n)$ and $n = o(d^{10/9})$, we have:

$$[z^n]\{f^d(z)\} = \sqrt{\frac{d}{2\pi}} \cdot \frac{f^d(\rho)}{n\rho^n}(1 + o(1)).$$

4 Introducing a factor $\psi(z)$

We now allow a multiplicative factor $\psi(z)$ and study $[z^n]\{f^d(z)\psi(z)\}$. The function ψ may itself depend on d or on other parameters, as long as the following property is satisfied:

Assumption $\mathcal{A}_2$:
The function ψ has positive coefficients, such that $\psi(0) \neq 0$, has a strictly positive radius of convergence, and either is fixed, or is a product of "large" powers of functions. In this case, it has the following form, where p is any fixed integer and the $d_i \to +\infty$:

$$\psi(z) = \prod_{i=1}^{p} g_i(z)^{d_i} \qquad \text{with} \qquad d_i = o(\frac{d}{\sqrt{n}}), 1 \leq i \leq p. \tag{1}$$

We can justify the condition on ψ as follows: An extra factor $\psi(z)$ moves the saddle-point away from the value ρ_0 obtained for f^d; this does not matter as long as the new saddle-point ρ stays close enough, within $o(1/\sqrt{n})$ of ρ_0. The difference $\rho - \rho_0$ is $\Theta(\rho_0(\sum_i d_i/d))$, hence the condition (1).

We now present some theorems which extend the former ones to allow an extra factor ψ. Theorem 7 is an obvious extension of Theorem 1:

Theorem 7:
If f is a function with positive coefficients such that $f_0 \neq 0$ and $f_1 \neq 0$, and if the function ψ satisfies the assumption $\mathcal{A}_2$, then for n constant and $d \to +\infty$:

$$[z^n]\{f^d(z)\psi(z)\} = \binom{d}{n} f_0^{d-n} f_1^n \psi(0)(1 + O(1/\sqrt{d})).$$

When n and d have the same growth rate, we can prove the following result, which is roughly Theorem 2 of [10] (the condition on ψ below is stronger than the assumption $\mathcal{A}_2$):

Theorem 8:
Let f satisfy the assumption $\mathcal{A}_1$ of Section 2.1, and let ψ be a function with positive coefficients and a strictly positive radius of convergence. Assume that the equation $\Delta f(z) = n/d$ has a real positive solution ρ smaller than the radius of convergence of f. Then, for $n, d \to +\infty$ and $n = \Theta(d)$:

$$[z^n]\{f^d(z)\psi(z)\} = \frac{f(\rho)^d \psi(\rho)}{\rho^{n+1}\sqrt{2\pi d\, \delta f(\rho)}}(1 + o(1)).$$

The case $n = o(d)$ is settled by the following theorem, whose proof can be found in [8]:

Theorem 9:
Let f and ψ satisfy repectively the assumptions $\mathcal{A}_1$ and $\mathcal{A}_2$ and let $n = o(d)$, with $n, d \to +\infty$. Define ρ as the unique real positive solution of $\Delta f(\rho) = n/d$. Then:

$$[z^n]\{f(z)^d \psi(z)\} = \frac{f(\rho)^d \cdot \psi(\rho)}{\rho^n \sqrt{2\pi n}}(1 + o(1)).$$

If some relationship $n^l = o(d^q)$ holds, then we have the analog of Corollary 2:

Corollary 3:
If f satisfies the assumptions $\mathcal{A}_1$ of Section 2.1, if ψ satisfies $\mathcal{A}_2$ but with the stronger conditions $d_i = o(d/n)$, and if $n = o(\sqrt{d})$:

$$[z^n]\{f^d(z)\psi(z)\} = \psi(0)\frac{f_0^d}{\sqrt{2\pi n}} \cdot \left(\frac{ef_1 d}{f_0 n}\right)^n (1 + o(1)).$$

If we only have $n = o(d^{2/3})$, then:

$$[z^n]\{f^d(z)\psi(z)\} = \psi(0)\frac{f_0^d}{\sqrt{2\pi n}} \left(\frac{ef_1 d}{f_0 n}\right)^n \exp\left(\frac{n^2}{d}(\frac{f_2}{f_1} - \frac{1}{2})\right) (1 + o(1)).$$

5 Some applications

An easy check of our formulæ is provided by the function $f(z) = e^z$. The saddle point is $\rho = n/d$ and we get, for $d \to +\infty$ and for n either fixed or going to infinity

$$[z^n]\{e^{dz}\} = d^n/n! = \frac{e^n d^n}{n^n\sqrt{2\pi n}}(1 + o(1)),$$

which is simply Stirling's formula for $n!$.

One of the basic constructions for obtaining combinatorial structures is to take a sequence of simpler objects. Let $f(z)$ be the generating function enumerating these objects according to their size; the generating function enumerating the sequences of d basic objects, according to their global size, is $f(z)^d$, and the coefficient $[z^n]\{f(z)^d\}$ enumerates the number of sequences of d basic objects of size n. The same approach can also be used to analyze the abelian partitional complex, whose bivariate generating function has the form $\exp(xf(y))$.

However, we do not count the structures of size 0 and we have $f(0) = 0$ and $n \geq d$. Let us define $f(y) = yg(y)$ with $g(0) \neq 0$; we have that $[z^n]\{f^d(z)\} = [z^{n-d}]\{g^d(z)\}$. The results presented above can now be applied to evaluate the number of composed objects of size $n \geq d$ which are a sequence of d simpler objects.

Classical examples are the Stirling numbers of the first and the second type. Stirling numbers of the first type enumerate, among other things, the number of permutations of n objects with k cycles; their exponential generating function is $\sum_{n,k} s_{n,k}x^k y^n/n! = \exp(x\log(1/1-y))$; hence

$$s_{n,k} = \frac{n!}{k!}[y^{n-k}]\{f(y)^k\} \qquad \text{with} \qquad f(y) = \frac{1}{y}\log\frac{1}{1-y} = \sum_{n\geq 0}\frac{y^n}{n+1}.$$

For example, we can get an asymptotic equivalent for $n = k + o(k)$, or equivalently $k = n - o(n)$, but still $n - k \to +\infty$. The saddle point ρ is approximately $2(n-k)/k$ and

Corollary 2 gives, for $n = k + o(\sqrt{k})$:

$$s_{n,k} = \frac{n!}{k!\sqrt{2\pi(n-k)}}\left(\frac{ek}{2(n-k)}\right)^{n-k}(1+o(1)),$$

i.e., using Stirling's approximation for $(n-k)!$ backwards:

$$s_{n,k} = \binom{n}{k}(k/2)^{n-k}(1+o(1)).$$

Let $S_{n,k}$ be a Stirling number of second type, enumerating for example the number of partitions of n objects into k blocks. These numbers have for exponential generating function $\sum_{n,k} S_{n,k}x^k y^n/n! = \exp(x(e^y-1))$, hence

$$S_{n,k} = \frac{n!}{k!}[y^{n-k}]\{f(y)^k\} \qquad \text{with} \qquad f(y) = \frac{e^y-1}{y} = \sum_{n\geq 0}\frac{y^n}{(n+1)!}.$$

For $n = k + o(\sqrt{k})$ and $n - k \to +\infty$, Corollary 2 applied to $f(y)$ gives

$$S_{n,k} = \frac{n!}{k!\sqrt{2\pi(n-k)}}\left(\frac{ek}{2(n-k)}\right)^{n-k}(1+o(1)) = \binom{n}{k}(k/2)^{n-k}(1+o(1)).$$

This asymptotic expression, which is also given for example in [1, p. 825], is the same as the one for the Stirling numbers of the first type: From Corollary 2, only f_0 and f_1 are important if $n-k = o(\sqrt{k})$. However, if the difference $n-k$ is of order at least $\sqrt{k}$, the next coefficients become important. For example, if $n-k \to +\infty$ with only $n-k = o(k^{2/3})$, then the second part of Corollary 2 shows that the Stirling numbers of the first and second type have a different behaviour:

$$\begin{aligned} s_{n,k} &= \binom{n}{k}(k/2)^{n-k}e^{(n-k)^2/6k}(1+o(1)); \\ S_{n,k} &= \binom{n}{k}(k/2)^{n-k}e^{-(n-k)^2/6k}(1+o(1)). \end{aligned}$$

Stirling numbers of the second type also appear in a classical occupancy problem of discrete probability theory: *We throw n balls into k urns randomly and independently; what is the number of urns with at least one ball?* Let $N_{n,d}$ be the number of ways of assigning the n balls to exactly d urns. If the balls are undistinguishable and if the urns have unbounded capacity, the associated generating function is [14]:

$$\Phi(x,y) = \sum_{n,d} N_{n,d}\, x^d\, \frac{y^n}{n!} = (1 + x(e^y-1))^k.$$

Let $f(y) = (e^y-1)/y$; we have $N_{n,d} = n!\,\binom{k}{d}\,[y^{n-d}]\{f(y)^d\}$, which can be expressed using Stirling numbers of the second type: $N_{n,d} = d!\binom{k}{d}S_{n,d}$.

6 Conclusion

We have presented results on the asymptotic approximation of coefficients of the type $[z^n]\{f^d(z)\}$, with applications, and on the asymptotic approximation of the coefficient $[z^n]\{f^d(z)\psi(z)\}$; examples of applications using such coefficients can be found in [10]. Possible extensions include:

- Allowing the second coefficient of f to be null: $f_1 = 0$. This corresponds to a function $f(z) = 1 + f_2z^2 + \cdots$. Preliminary studies indicate that such an extension considerably restricts the respective ranges of n and d.

- Allowing $d = o(n)$ for more general functions than those considered in Section 3.5. Here again we may have to introduce further growth restrictions on n and d, depending on the singularities of the funtion $f(z)$; we may also have to use a technic more adapted to the nature of the singularities than the saddle point method.

- Removing the restriction that ψ has positive coefficients. This does not seem to pose any real difficulty, as opposed to the fact that the similar condition on f is essential.

- Obtaining further terms of an asymptotic expansion. This is similar to the extension of the results of Daniels [2] by Good [11], and should not introduce major difficulties.

References

[1] M. ABRAMOWITZ and I. STEGUN. *Handbook of Mathematical Functions.* Dover, New York, 1964.

[2] H.E. DANIELS. Saddlepoint approximations in statistics. *Annals of Mathematical Statistics*, 25:631–650, 1954.

[3] N. G. deBRUIJN. *Asymptotic Methods in Analysis.* Dover, New York, 1981 (reprint).

[4] P. DIACONIS, R.L.GRAHAM, and J.A. MORRISON. Asymptotic analysis of a random walk on a hypercube with many dimensions. *Random Structures and Algorithms*, 1(1):51–72, 1990.

[5] P. FLAJOLET. On the performance evaluation of extendible hashing and trie searching. *Acta Informatica*, 20:345–369, 1983.

[6] W. FULKS. A generalization of Laplace's method. *American Mathematical Society Proceedings*, 2(4):613–622, August 1951.

[7] D. GARDY. Join sizes, urn models and normal limiting distributions. Technical Report 600, Laboratoire de Recherche en Informatique, Université de Paris-Sud, France, October 1990.

[8] D. GARDY. On coefficients of powers of functions. Technical Report CS-91-53, Dept. of Computer Science, Brown University, Providence, R.I. (USA), August 1991.

[9] D. GARDY. Normal limiting distributions for projection and semijoin sizes. *SIAM Journal on Discrete Mathematics*, 5(2), 1992.

[10] D. GARDY and P. SOLE. Saddle-point techniques in asymptotic coding theory. In *Algebraic Coding, First french-Soviet Workshop*, pages 75–81, Paris, July 1991. Lecture Notes in Computer Science, Springer Verlag, No. 573.

[11] I.J. GOOD. Saddle-point methods for the multinomial distribution. *Annals of Mathematical Statistics*, 28:861–881, 1957.

[12] D.H. GREENE and D.E. KNUTH. *Mathematics for the Analysis of Algorithms.* Birkhäuser, 1982.

[13] W.K. HAYMAN. A generalisation of Stirling's formula. *Journal für die reine und angewandte Mathematik*, 196:67–95, 1956.

[14] N.L. JOHNSON and S. KOTZ. *Urn models and their application.* Wiley & Sons, 1977.

[15] V. KOLCHIN, B. SEVAST'YANOV, and V. CHISTYAKOV. *Random Allocations.* Wiley & Sons, 1978.

[16] J.E. MAZO and A.M. ODLYZKO. Lattice points in high-dimensional spheres. Technical report, AT&T Bell Laboratories, Murray Hill, USA, 1990.

[17] A. MEIR and J.W. MOON. The asymptotic behaviour of coefficients of powers of certain generating functions. *Journal Européen de Combinatoire*, 11:581–587, 1990.

[18] A.M. ODLYZKO and L.B. RICHMOND. On the unimodality of high convolutions of discrete distributions. *Annals of Probability*, 13(1):299–306, 1985.

[19] R.N. PEDERSON. Laplace's method for two parameters. *Pacific Journal of Mathematics*, 15(2):585–596, 1965.

[20] J.A. RUSH and N.J.A. SLOANE. Bound for packing superballs. *Mathematika*, 34:8–18, 1987.

[21] J.-M. STEYAERT and P. FLAJOLET. Patterns and pattern-matching in trees: an analysis. *Information and Control*, 58(1-3):19–57, July/August/September 1983.

[22] H.S. WILF. *Generatingfunctionology.* Academic Press, 1990.

RECENT PROGRESS
on
THE MACDONALD q,t-KOSTKA CONJECTURE

A. M. Garsia (*)

ABSTRACT. In this lecture I shall present recent progress, in a joint effort with Mark Haiman, toward proving the Macdonald q,t-Kostka conjecture. The main thrust of our work has been towards the construction of a representation theoretical setting for the Macdonald basis $\{P_\mu(X,q,t)\}_\mu$. The original goal was to obtain new methods for attacking some of the problems and conjectures arising from Macdonald work [14]. This effort has been met with success beyond our best expectations. In particular, it has already brought to light some truly remarkable properties and facts concerning these polynomials. It has also opened up a new area of investigation with a wide variety of exciting algebraic and combinatorial problems and conjectures. The feeling prevails that this is only the the tip of a mathematical iceberg that could keep many investigators occupied for a few years to come. I can give here only a sample of the results and problems that stem from this development. We refer to [3], [4] and [5] for a more complete treatment.

—o——o——o——o——o——o—

We recall that Macdonald in [14] shows the existence of a family of polynomials $\{P_\lambda(x;q,t)\}$ which are uniquely characterized by the following conditions

$$\begin{aligned} a) \quad & P_\lambda = S_\lambda + \sum_{\mu<\lambda} S_\mu\, \xi_{\mu\lambda}(q,t) \\ b) \quad & \langle P_\lambda\, ,\, P_\mu \rangle_{q,t} \;=\; 0 \quad for \quad \lambda \neq \mu \end{aligned} \tag{1}$$

Where S_λ denotes the Schur function indexed by λ and $\langle\ ,\ \rangle_{q,t}$ denotes the scalar product of symmetric polynomials defined by setting for the power basis $\{p_\rho\}$

$$\langle p_{\rho_1}\, ,\, p_{\rho_2} \rangle_{q,t} \;=\; \begin{cases} z_\rho\, p_\rho\left[\frac{1-q}{1-t}\right] & \text{if } \rho_1=\rho_2=\rho \text{ and} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

Here we use λ-ring notation and z_ρ is the integer that makes $n!/z_\rho$ the number of permutations with cycle structure ρ. There are a number of outstanding conjectures concerning these polynomials (see [14]). Here we shall be dealing with those involving the so called ***integral forms*** $J_\mu(x;q,t)$ and their associated Macdonald-Kostka coefficients $K_{\lambda\mu}(q,t)$. We shall use the same notation as in [14]. In particular $\{Q_\lambda(x;q,t)\}$ denotes the basis dual to $\{P_\lambda(x;q,t)\}$ with respect to the scalar product $\langle\ ,\ \rangle_{q,t}$. Clearly, (1) b) gives

$$Q_\lambda(x;q,t) \;=\; d_\lambda(q,t)\, P_\lambda(x;q,t)\ , \tag{3}$$

(*) Work carried out under NSF grant support.

for a suitable rational function $d_\lambda(q,t)$. However in [14] it is shown that

$$d_\lambda(q,t) = \frac{h_\lambda(q,t)}{h'_\lambda(q,t)}$$

with

$$h_\lambda(q,t) = \prod_{s\in\lambda}(1-q^{a_\lambda(s)}t^{l_\lambda(s)+1}) \quad , \quad h'_\lambda(q,t) = \prod_{s\in\lambda}(1-q^{a_\lambda(s)+1}t^{l_\lambda(s)})$$

where s denotes a generic lattice square and $a_\lambda(s)$, $l_\lambda(s)$ respectively denote the *arm* and the *leg* of s in the Ferrers' diagram of λ.

We recall from [14] that

$$J_\mu(x;q,t) = h_\mu(q,t)P_\mu(x;q,t) = h'_\mu(q,t)Q_\mu(x;q,t) \ , \tag{4}$$

and the coefficients $K_{\lambda\mu}(q,t)$ are defined through an expansion which in λ-ring notation may be written as

$$J_\mu(x;q,t) = \sum_\lambda S_\lambda[X(1-t)]\, K_{\lambda\mu}(q,t) \tag{5}$$

Macdonald conjectures that these coefficients are polynomials in q and t with non-negative integer coefficients. We shall refer to this here and after as the MPK conjecture. Macdonald derives a number of properties of the $K_{\lambda\mu}(q,t)$; in particular he shows that for any partition μ

$$K_{\lambda\mu}(1,1) = f_\lambda \tag{6}$$

where f_λ denotes the number of standard tableaux of shape λ. This given, the MPK conjecture is equivalent to the statement that for each μ there exists an S_n-module M_μ yielding a bigraded version of the left regular representation whose character has the expansion

$$char\ M_\mu = \sum_\lambda \chi^\lambda\, K_{\lambda\mu}(q,t) \ . \tag{7}$$

More precisely, if $\mathcal{H}_{h,k}(M_\mu)$ denotes the submodule of M_μ consisting of its bihomogeneous elements of bidegree (h,k) and we set

$$p^\mu(q,t) = \sum_{h,k\geq 0} q^h t^k\, char\ \mathcal{H}_{h,k}(M_\mu) \tag{8}$$

then (7) should hold true with $char\ M_\mu = p^\mu(q,t)$. In this vein, the symmetric polynomial

$$H_\mu(x;q,t) = \sum_\lambda S_\lambda\, K_{\lambda\mu}(q,t) = J_\mu[X/(1-t);q,t] \tag{9}$$

may be viewed as the Frobenius characteristic of M_μ, while the expression

$$F_\mu(q,t) = \sum_\lambda f_\lambda\, K_{\lambda\mu}(q,t) \tag{10}$$

should give its Hilbert series, that is the polynomial

$$F_\mu(q,t) \;=\; \sum_{h,k\geq 0} q^h t^k \, dim\, \mathcal{H}_{h,k}(M_\mu) \;. \tag{11}$$

For technical reasons it is preferable to work with the modified versions of $H_\mu(x;q,t)$ and $F_\mu(q,t)$ obtained by setting

$$\tilde{H}_\mu(x;q,t) \;=\; H_\mu(x;q,1/t)t^{n(\mu)} \quad , \quad \tilde{F}_\mu(q,t) \;=\; F_\mu(q,1/t)t^{n(\mu)} \;. \tag{12}$$

It will also be convenient to set

$$\tilde{K}_{\lambda\mu}(q,t) \;=\; K_{\lambda\mu}(q,1/t)t^{n(\mu)} \;,$$

where, for $\mu = (\mu_1 \geq \mu_2 \geq \cdots \geq \mu_k)$

$$n(\mu) \;=\; \sum_{i=1}^{k} (i-1)\mu_i \;. \tag{13}$$

In the fall of 1989 I set myself the task of finding a module M_μ whose bigraded character, as defined by (8), has the expansion

$$char\, M_\mu \;=\; \sum_\lambda \chi^\lambda \, \tilde{K}_{\lambda\mu}(q,t) \;. \tag{14}$$

Since setting $q=0$ in $K_{\lambda\mu}(q,t)$ yields the Kostka-Foulkes coefficients $K_{\lambda\mu}(t)$ a good starting point appeared to be some early work of the algebraic geometers (see [6] and the references quoted there) which yielded the first proof of the analogous positivity result for $K_{\lambda\mu}(t)$. The basic ingredient that may be extracted from this literature is a certain graded S_n-module $\mathbf{R}_\mu$ in whose character the coefficients $K_{\lambda\mu}(t)$ appear as t-multiplicities of irreducibles. To be precise, let p^μ_m denote the character of the action of S_n on the m^{th} graded component of $\mathbf{R}_\mu$, and set

$$p^\mu(t) \;=\; \sum_{m\geq 0} t^m \, p^\mu_m \;. \tag{15}$$

Expanding in terms of the the irreducible characters χ^λ we may also write

$$p^\mu(t) \;=\; \sum_\lambda \chi^\lambda C_{\lambda\mu}(t) \;, \tag{16}$$

where $C_{\lambda\mu}(t)$ is the polynomial whose coefficient of t^m gives the multiplicity of χ^λ in p^μ_m. Now a sequence of deep developments (see [13] II §3 ex. 1 p. 92 and III §7 ex. 9 p.136) yields that

$$K_{\lambda\mu}(t) \;=\; C_{\lambda,\mu}(t^{-1})t^{n(\mu)} \;. \tag{17}$$

However, although $\mathbf{R}_\mu$ was later shown by Kraft [7] and DeConcini-Procesi [2] to have an elementary direct definition as a quotient of the polynomial ring $\mathbf{Q}[x_1,\ldots,x_n]$, the relation in (17) had only been established in a setting that not only required t to be the power of a prime, but also relied on some of the deepest results and tools of Algebraic Geometry. This given, I set myself the task of providing a purely representation theoretical setting for the study of $\mathbf{R}_\mu$ and its graded character $p^\mu(t)$. This program was successfully carried out in joint work with Procesi in [6]. In particular a new proof of (17) was obtained which only used elementary representation theoretical tools. In an earlier work [8]-[12] Lascoux and Schützenberger stated without proof that that a suitable submodule $\mathbf{R}_\mu[X]$ of the S_n-harmonic polynomials had also the same graded character. Subsequently, Bergeron and Garsia [1] were able to provide a proof of this fact that used only elementary tools of commutative algebra. This done, the next task was to try and see to what extent these methods could be applied to the q,t-case. This program is in the process of being carried out in joint work with Mark Haiman. The starting point is the brilliant idea of Haiman to try and construct the desired bigraded module by working in an analogous manner with polynomials in two sets of variables $X = \{x_1, \ldots, x_n\}$ and $Y = \{y_1, \ldots, y_n\}$. This viewpoint led to the construction of an S_n-module $M_D[X, Y]$ which yields a bigraded version of the left regular representation of S_n for each lattice square diagram D. During more than a year we have been involved in an intensive study of the module M_D and the various problems that have arisen from our efforts to compute its character. During this period we have gathered overwhelming evidence that when D is the Ferrers' diagram of a partition μ the resulting module $M_\mu[X, Y]$ has a bigraded character given by (14). To describe this evidence we need some notation. For a given diagram D we let $p^D(q,t)$ denote the bigraded character of M_D, and let $G[D](x; q, t)$ be its Frobenius characteristic. We also set

$$p^D(q,t) \;=\; \sum_\lambda \chi^\lambda\, C_{\lambda D}(q,t) \ . \tag{18}$$

Clearly we must also have then

$$G[D](x;q,t) \;=\; \sum_\lambda S_\lambda(x)\, C_{\lambda D}(q,t) \ . \tag{19}$$

When D is the Ferrers' diagram of a partition μ, $p^D(q,t)$, $G[D](x;q,t)$ and $C_{\lambda D}(q,t)$ will be simply represented by $p^\mu(q,t)$, $G_\mu(x;q,t)$ and $C_{\lambda\mu}(q,t)$ respectively. Our efforts have been directed towards proving that

$$C_{\lambda\mu}(q,t) \;=\; \tilde{K}_{\lambda\mu}(q,t) \tag{20}$$

So far this has now been proved by Garsia and Haiman ([4],[5]) in the following cases

(1) For all μ when λ is a hook,
(2) For all λ when μ is a hook,
(3) For all λ when μ any two row or two column partition,
(4) For all λ when μ is a partition of $n \le 6$.
(5) Verified by computer for all λ when μ a partition of $n \le 7$

The completion of this work, and a proof of the conjecture in full generality, hinges on the establishment of a number of properties of the modules $\mathbf{R}_\mu[X, Y]$ which have emerged from theoretical considerations combined with computer data. I can give a brief view of some the work that has been done. In the one parameter case, the modules $\mathbf{R}_\mu$ and $\mathbf{R}_\mu[X]$ studied in [1] and [6] respectively have different definitions but they are shown in [1] to be equivalent as graded S_n-modules. The module $\mathbf{R}_\mu$ studied in [6] is a quotient of the ring of polynomials in $x_1, x_2, \ldots, x_n$, while

$\mathbf{R}_\mu[X]$ is defined in [1] as the linear span of the derivatives of the Garnir polynomials corresponding to standard tableaux of shape μ. The module $\mathbf{R}_\mu[X,Y]$ referred to above is a natural bigraded extension of $\mathbf{R}_\mu[X]$. It may be defined as the linear span of the derivatives of a single bihomogeneous polynomial $\Delta_\mu(x,y)$ in the variables X and Y. Extending the definition of the ring $\mathbf{R}_\mu$ given in [6] to the two variable case, leads to two separate constructions and two additional spaces ${}^{bg}\mathbf{R}_\mu$ and ${}^{sg}\mathbf{R}_\mu$. Both these spaces are obtained by working on polynomials in X and Y. While ${}^{bg}\mathbf{R}_\mu$ is bigraded, it is not defined as a quotient ring, in contrast ${}^{sg}\mathbf{R}_\mu$ is a quotient ring but only known to be singly graded. All three spaces $\mathbf{R}_\mu[X,Y]$, ${}^{bg}\mathbf{R}_\mu$ and ${}^{sg}\mathbf{R}_\mu$ are S_n modules under the diagonal action of S_n. This is the action defined by setting, for $\sigma = \sigma_1 \cdots \sigma_n \in S_n$:

$$\sigma P(x_1,x_2,..,x_n;y_1,y_2,..,y_n) = P(x_{\sigma_1},x_{\sigma_2},..,x_{\sigma_n};y_{\sigma_1},y_{\sigma_2},..,y_{\sigma_n}) \,. \tag{21}$$

The first task is to prove (as in the one parameter case), that these three spaces are equivalent as graded S_n-modules. All evidence gathered so far supports this conclusion. It is interesting to have a look at some of the properties that have been established concerning these three spaces. We shall also agree to use the same notation for the bigraded characters of $\mathbf{R}_\mu[X,Y]$ and ${}^{bg}\mathbf{R}_\mu$. Let us keep in mind that (due to (6)) the validity of (20) implies that $\mathbf{R}_\mu[X,Y]$ and ${}^{bg}\mathbf{R}_\mu$ should be bi-graded versions of the left regular representation of S_n. In particular, all the three spaces should have dimension $n!$. Under this conjecture, the polynomial

$$F_\mu(q,t) = \sum_\lambda f_\lambda \, \tilde{K}_{\lambda\mu}(q,t) \tag{22}$$

should give a bigraded version of the Hilbert series of these two modules, while

$$F_\mu(q) = F_\mu(q,q) = \sum_\lambda f_\lambda \, \tilde{K}_{\lambda\mu}(q,q)$$

should give the Hilbert series of ${}^{sg}\mathbf{R}_\mu$ and the singly graded one for $\mathbf{R}_\mu[X,Y]$ and ${}^{bg}\mathbf{R}_\mu$.

Macdonald also shows a number of identities relating the coefficients $K_{\lambda\mu}(q,t)$ for various values of λ, μ, q, t. For instance, from the results in [14] it can be derived that

1) $F_\mu(q,t)$ is symmetric, that is $F_\mu(q,t) = F_\mu(q^{-1},t^{-1})q^{n_\mu}t^{n_{\mu'}}$.
2) $K_{\lambda\mu}(0,t) = K_{\lambda\mu}(t)$,
3) $\tilde{K}_{\lambda\mu}(q,t) = \tilde{K}_{\lambda\mu'}(t,q)$,
4) $\tilde{K}_{\lambda\mu}(q,t) = q^{n_{\mu'}}t^{n_\mu} \, \tilde{K}_{\lambda'\mu}(q^{-1},t^{-1})$,

where, priming a partition here represents conjugation. This given, the following table summarizes some of the results presented in [4] and [5], which yield further evidence in support of (20).

PROPERTY	$\mathbf{R}_\mu[X,Y]$	${}^{bg}\mathbf{R}_\mu$	${}^{sg}\mathbf{R}_\mu$
Symmetric Hilbert series	yes	yes	?
Quotient ring	$\asymp$	?	yes
Regular representation	?	yes	yes

Dimension $n!$	?	yes	yes
Bigraded	yes	yes	?
$C_{\lambda\mu}(q,1) = \tilde{K}_{\lambda\mu}(q,1)$	?	yes	$\asymp$
$C_{\lambda\mu}(q,0) = \tilde{K}_{\lambda\mu}(q,0)$	yes	yes	$\asymp$
$C_{\lambda\mu}(q,t) = \tilde{C}_{\lambda\mu'}(t,q)$	yes	yes	$\asymp$
$C_{\lambda\mu}(q,t) = q^{n_{\mu'}}t^{n_\mu}C_{\lambda'\mu}(q^{-1},t^{-1})$	yes	yes	$\asymp$

The symbol $\asymp$ is to signify here that the property in question is not applicable to the given module. The question mark "?" represents the conjecture that we should have a *yes*. Note that each property holds true for at least one of the modules. Remarkably, it can be shown that the removal of a single question mark "?" (that is replacing it by a *yes*) in this table removes them all and forces all three constructions to yield the same bigraded S_n-module.

However the strongest evidence supporting the validity of (20), is that the modules $M_D[X,Y]$ have suggested us identities involving the polynomials $G[D](x;q,t)$ which we were in fact able to prove within the theory of Macdonald polynomials. We shall only give a brief view of this development and refer the reader to [3] for a more detailed presentation.

We shall say that two lattice square diagrams D_1 and D_2 are equivalent and write $D_1 \approx D_2$ if and only if D_2 can be obtained from D_1 by a sequence of row and column rearrangements. If D is a lattice square diagram, the diagram obtained by reflecting D across the diagonal line $x = y$ will be called the *conjugate* of D and denoted by D'. Similarly, the reflection of a lattice square s across $x = y$, will be denoted by s'. Finally, if D may be decomposed into the union of two diagrams D_1 and D_2 in such a manner that no square of D_2 is in the rook domain of a square of D_1, then we shall say that D is *decomposable* and we write $D = D_1 \times D_2$. This given, the construction of the module M_D suggests that the family of polynomials $\{G[D](x;q,t)\}_D$ has the following basic properties

$$\begin{cases} (1) \quad G[D_1](x;q,t) = G[D_2](x;q,t) & \text{if } \; D_1 \approx D_2 \\ (2) \quad G[D_1](x;q,t) = G[D_2](x;t,q) & \text{if } \; D_2 \approx D_1' \\ (3) \quad G[D](x;q,t) = G[D_1](x;q,q)G[D_2](x;q,t) & \text{if } \; D \approx D_1 \times D_2 \end{cases} \tag{23}$$

The validity of (20) yields the further relation

$$G[D](x;q,t) = \tilde{H}_\mu(x;q,t) \quad \textit{if} \;\; D \;\; \textit{is} \;\; \textit{the} \;\; \textit{diagram} \;\; \textit{of} \;\; \mu \tag{24}$$

which may be interpreted as an initial condition. Moreover, a study of the behavior of the S_n-module M_D under restriction to S_{n-1} suggests recursions for some of the polynomials $G[D](x;q,t)$ which, together with equation (24) above completely determine them as linear combinations of the polynomials $\tilde{H}_\mu(x;q,t)$. This permits the study of the the polynomials $G[D](x;q,t)$ independently of the validity of (20). In particular by recursing through this extended family, in [3], we were able to rederive some of the Stanley-Macdonald [15]-[14] Pieri rules in a manner which unravels their original intricacy as a combination of successive, simple elementary steps. The remarkable agreement of the resulting identities with those that can be derived from the theory of Macdonald polynomials, offers what is so far the best evidence of the validity of (20). In our lecture we shall present some of the latter developments in a *Viennotique* of lattice square diagrams manipulations.

REFERENCES

[1] N. Bergeron and A. Garsia, *On certain spaces of Harmonic Polynomials*, Contemporary Mathematics, (to appear).

[2] C. de Concini and C. Procesi, *Symmetric functions, conjugacy classes and the flag variety*, Invent.Math., 64 (1981), 203-230.

[3] A. M. Garsia and M. Haiman, *Factorizations of Pieri rules for Macdonald polynomials*, UCSD Preprint (1992).

[4] A. M. Garsia and M. Haiman, *Some bigraded S_n-modules and the Macdonald q,t-Kotska coefficients* (to appear)

[5] A. M. Garsia and M. Haiman, *Orbit Harmonics and Graded Representations* (with M. Haiman), (Research Monograph to appear as part of the Collection Published by the Lab. de Comb. et Informatique Mathématique, edited by S. Brlek, U. du Quèbec á Montréal)

[6] A. M. Garsia and C. Procesi, *On certain graded S_n modules and the q-Kostka polynomials* To appear in Advances of Mathematics.

[7] H. Kraft , *Conjugacy classes and Weyl group representations*, Proc. 1980 Torun Conf. Poland Astérisque, 87-88 (1981), 195-205.

[8] A. Lascoux, *Cyclic permutations on words, Tableaux and Harmonic polynomials*, (to appear)

[9] A. Lascoux and M.P. Schützenberger, *Sur une conjecture de H. O. Foulkes*, C.R.Acad.Sc.Paris, 286 (1978), 323-324,

[10] A. Lascoux and M.P. Schützenberger, *Croissance des polynômes de Foulkes/Green* , C.R. Acad. Sc. Paris, 288 (1979), 95-98.

[11] A. Lascoux and M.P. Schützenberger, *Croissance des Tableaux de Young et foncteurs de Schur*, Astérisque, 87-88 (1981), 191-205,

[12] A. Lascoux and M.P. Schützenberger, *Cyclic permutations, Tableaux and Harmonic Polynomials*, (preliminary version of item [8] above).

[13] I. G. Macdonald, *Symmetric functions and Hall polynomials*, Clarendon Press, Oxford (1979).

[14] I. G. Macdonald, *A new class of symmetric functions*, Actes du 20^e Séminaire Lotharingien, Pub. IRMA Strasbourg (1988) 131-171

[15] R.Stanley, *Some combinatorial properties of Jack Symmetric functions*, Adv. Math. 77 (1989) 76-117.

A DECOMPOSITIONS FOR GRAPHS RELATED TO THE TUTTE POLYNOMIAL

IRA M. GESSEL*
DEPARTMENT OF MATHEMATICS
BRANDEIS UNIVERSITY
WALTHAM, MA 02254-9110

1. Introduction. In this paper, we study the enumerative consequences of a very simple way of decomposing a graph: choose a vertex and remove it and its incident edges, keeping track of the number of connected components and edges. By applying this decomposition to connected graphs, we recover some known formulas for counting connected graphs by edges and for counting trees by inversions. Applying the decomposition to arbitrary graphs, we add another parameter to these formulas, counting graphs by edges and connected components, and counting trees by inversions and another statistic described below. The corresponding two-variable generalization of the inversion enumerator for trees turns out to be a well-known graph polynomial: the Tutte polynomial of the complete graph. There are many equivalent definitions of the Tutte polynomial $t_G(\alpha, \beta)$ of a graph G, but for our purposes the most useful one is as a polynomial obtained by a simple change of variables from the generating function for subgraphs of a given graph by connected components and edges.

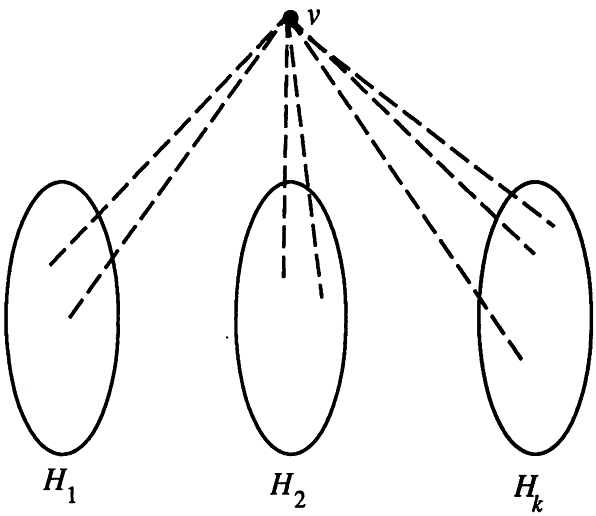
It is natural then to try to generalize our formulas to the Tutte polynomial of an arbitrary graph G by restricting the decomposition to subgraphs of G. The formula for the inversion enumerator for trees generalizes nicely to an arbitrary graph G, giving a new interpretation to $t_G(1, \beta)$ as counting spanning trees of G, by inversions, but only some inversions are counted. In particular, we find a combinatorial interpretation for any graph G (without loops) of $t_G(1, -1)$. (In the case of the complete graph, this number is a tangent or secant number.) We also generalize our interpretation of the complete Tutte polynomial $t_G(\alpha, \beta)$ in the case in which G has a vertex adjacent to every other vertex.

2. The depth-first decomposition. Let H be a connected graph rooted at the vertex v. Let $H_1, H_2, \ldots, H_k$ be the connected components of the graph obtained by deleting v and its incident edges. We call $H_1, \ldots, H_k$ the *depth-first components* of H rooted at v. The reason for this terminology is that if for each i we choose an edge joining v to a vertex v_i in H_i, and then apply this procedure recursively to each H_i rooted at v_i, we obtain a depth-first spanning tree of H. We refer the reader to [8] and [6] for the enumerative consequences of the complete depth-first search. In this paper we study the formulas that arise from a single application of the depth-first decomposition, without actually constructing the depth-first search spanning trees.

Given a set of connected graphs $H_1, \ldots, H_k$ on disjoint vertices and a new vertex v, we can construct a graph rooted at v whose depth-first components are $H_1, \ldots, H_k$ by

*partially supported by NSF grant DMS-9101516

adding edges from v to a subset of the vertices of $H_1, \ldots, H_k$; this subset must include at least one vertex from each H_i.



Now let $C_n(\beta) = \sum_i C_{n,i}\beta^i$, where $C_{n,i}$ is the number of connected graphs with vertex set $[n] = \{1, 2, \ldots, n\}$ and with i edges. It is well known that

$$\sum_{n=0}^{\infty} C_n(\beta)\frac{u^n}{n!} = \log\left(\sum_{n=0}^{\infty}(\beta+1)^{\binom{n}{2}}\frac{u^n}{n!}\right). \tag{1}$$

If we consider connected graphs on $[n]$ rooted at 1, then the depth-first decomposition, together with elementary properties of exponential generating functions, gives the following formula (which can also be derived algebraically from (1):

Theorem 1.

$$\sum_{n=0}^{\infty} C_{n+1}(\beta)\frac{u^n}{n!} = \exp\left(\sum_{m=1}^{\infty}\left((\beta+1)^m - 1\right)C_m(\beta)\frac{u^m}{m!}\right). \quad \square \tag{2}$$

A recurrence equivalent to Theorem 1 was given by Leroux [11, p. 15], who also generalized it to species. The special case $\beta = 1$ was stated by Harary and Palmer [9, p. 8], who attributed it to John Riordan, though it does not appear in the paper of his that they cite [12].

Since every connected graph with n vertices has at least $n-1$ edges, $C_n(\beta)$ is divisible by β^{n-1}. Thus we may define a polynomial $I_n(\beta)$ by

$$C_n(\beta) = \beta^{n-1} I_n(\beta+1). \tag{3}$$

The polynomial $I_n(\beta)$ is called the *inversion enumerator for trees* because of its combinatorial interpretation, described below. If we replace u with u/β in (2) and then replace β with $\beta - 1$ we obtain:

Theorem 2.

$$\sum_{n=0}^{\infty} I_{n+1}(\beta)\frac{u^n}{n!} = \exp\left(\sum_{m=1}^{\infty} \frac{\beta^m - 1}{\beta - 1} I_m(\beta)\frac{u^m}{m!}\right)$$
$$= \exp\left(\sum_{m=1}^{\infty} (1+\beta+\cdots\beta^{m-1}) I_m(\beta)\frac{u^m}{m!}\right). \quad \square \tag{4}$$

It is clear from (4) that the coefficients of $I_n(\beta)$ are nonnegative and that $I_n(-1)$ is also nonnegative. In the next section we give combinatorial interpretations to these quantities.

3. Inversions in trees. We first recall some standard notation for rooted trees. If T is a tree rooted at a vertex v, and x and y are distinct vertices of T, then we say that x is an *ancestor* of y, and y is a *descendant* of x, if x lies on the unique path from v to y. (This includes the case that $x = v$.) If x is an ancestor of y and x and y are adjacent, we call x the *parent* of y and we call y a *child* of x.

Now let T be a rooted tree on a totally ordered vertex set. An *inversion* in T is a pair (i, j) of vertices of T such that i is an ancestor of j and $i > j$. We define inversions in an unrooted tree (with a totally ordered vertex set) by rooting the tree at its least vertex.

The next result is due to Mallows and Riordan [12]. (See also Foata [4].)

Theorem 3. *The coefficient of β^i in $I_n(\beta)$ is the number of trees on $[n]$ with i inversions.*

Proof. For the moment let $J_m(\beta)$, for $m \geq 1$, be the inversion enumerator for unrooted trees on $[m]$, which we root at vertex 1. Then the enumerator for trees on $[m]$ rooted at i is easily seen to be $\beta^{i-1}J_m(\beta)$, and thus the enumerator for all rooted trees on $[m]$ is $(1+\beta+\cdots+\beta^{m-1})J_m(\beta)$. Now the inversions of a tree rooted at 1 are the same as the inversions of the subtrees rooted at the children of 1. We deduce (4) with $J_n(t)$ replacing $I_n(t)$. Since $I_n(t)$ is uniquely determined by (4), we must have $I_n(t) = J_n(t)$. $\square$

In view of the combinatorial interpretations we have for $C_n(t)$ and $I_n(t)$, it is natural to ask for a combinatorial interpretation of (3). Such a combinatorial interpretation has been given in [8], and the approach taken there, which is further studied in [6], can be used to give combinatorial proofs of the generalizations of (3) that follow.

If we set $\beta = -1$ in (4) we obtain

$$\sum_{n=0}^{\infty} I_{n+1}(-1)\frac{u^n}{n!} = \exp\left(\sum_{m \text{ odd}} I_m(-1)\frac{u^m}{m!}\right). \tag{5}$$

From (5) we can easily derive a combinatorial interpretation of $I_m(-1)$. Let us say that a rooted tree with a totally ordered vertex set is *increasing* if each vertex is less than all its children.

Theorem 4. *$I_n(-1)$ is the number of increasing trees on $[n]$ in which every vertex other than the root has an even number of children.*

Proof. It follows from (5) that $I_n(-1)$ is the number of trees on $[n]$ that are increasing and that have the following property: any subtree consisting of a nonroot vertex and all its descendants contains an odd number of vertices. This is easily seen to be equivalent to the condition stated in the theorem. □

A bijective proof of Theorem 4 has been given by Pansiot [13]. As noted by Kreweras [10] and Gessel [5], (4) implies that $\sum_{n=0}^{\infty} I_{n+1}(-1)u^n/n! = \sec u + \tan u$. Some analogous formulas for counting other types of trees by inversions can be found in Gessel, Sagan, and Yeh [7].

We can also apply the depth-first decomposition to arbitrary (not necessarily connected) graphs. In the general case, if H is a graph rooted at v then the number of connected components of H is one more than the number of depth-first components of H which are not connected to v. Let $S_n(\alpha,\beta) = \sum_{i,j} S_{n,i,j}\alpha^i\beta^j$, where $S_{n,i,j}$ is the number of graphs on $[n]$ with i connected components and j edges. Thus $C_n(\beta)$ is the coefficient of α in $S_n(\alpha,\beta)$. The depth-first decomposition yields:

Theorem 5.

$$\sum_{n=0}^{\infty} S_{n+1}(\alpha,\beta)\frac{u^n}{n!} = \alpha\exp\left(\sum_{m=1}^{\infty}\left((\beta+1)^m - 1 + \alpha\right)C_m(\beta)\frac{u^m}{m!}\right). \quad \square \tag{6}$$

Substituting $C_m(\beta) = \beta^{m-1}I_m(\beta+1)$, replacing u with u/β, and then replacing β with $\beta - 1$ in (6), we get

$$\sum_{n=0}^{\infty}(\beta-1)^{-n}S_{n+1}(\alpha,\beta-1)\frac{u^n}{n!} = \alpha\exp\left(\sum_{m=1}^{\infty}\frac{\alpha+\beta^m-1}{\beta-1}I_m(\beta)\frac{u^m}{m!}\right). \tag{7}$$

Now let us define polynomials $t_n(\alpha,\beta)$ for $n > 0$ by

$$t_n(\alpha,\beta) = (\alpha-1)^{-1}(\beta-1)^{-n}S_n\big((\alpha-1)(\beta-1),\beta-1\big),$$

so that $S_n(\alpha,\beta) = \alpha\beta^{n-1}t_n(\alpha/\beta+1,\beta+1)$. Replacing α with $(\alpha-1)(\beta-1)$ in (7), we obtain:

Theorem 6.

$$\sum_{n=0}^{\infty} t_{n+1}(\alpha,\beta)\frac{u^n}{n!} = \exp\left(\sum_{m=1}^{\infty}(\alpha+\beta+\beta^2+\cdots+\beta^{m-1})I_m(\beta)\frac{u^m}{m!}\right). \quad \square \tag{8}$$

It follows from (8) that $t_n(\alpha,\beta)$ is a polynomial with nonnegative integer coefficients and that $t_n(1,\beta) = I_n(\beta)$. It is not difficult to derive a combinatorial interpretation from (8) that refines our interpretation for $I_n(\beta)$:

Theorem 7. *The coefficient of $\alpha^i\beta^j$ in $t_n(\alpha,\beta)$ is the number of trees T on $[n]$ with j inversions such that vertex 1 is adjacent to exactly i vertices which are less than all their descendants.* □

From (8) it is also easy to derive a combinatorial interpretation for the coefficients of $t_n(\alpha+1,-1)$. We leave this to the reader.

In the next section we shall find similar formulas to those given here, when we restrict ourselves to the subgraphs of a fixed connected graph. Thus what we have done so far is the case of complete graphs. We shall see that $t_n(\alpha,\beta)$ is the instance for the complete graph on n vertices of a well-known polynomial called the *Tutte polynomial*, which is defined for any graph (and more generally for any matroid). Most of the formulas we obtained for $t_n(\alpha,\beta)$ and its specializations can be generalized to the Tutte polynomial of an arbitrary graph.

4. The Tutte polynomial. Let G be a graph with vertex set V. We shall assume that G has no loops or multiple edges, though most of our results will hold in a slightly modified form if loops and multiple edges are allowed.

We consider the polynomial

$$S_G(\alpha,\beta) = \sum_H \alpha^{c(H)}\beta^{e(H)},$$

where the sum is over all spanning subgraphs H of G; here $c(H)$ is the number of connected components of H and $e(H)$ is the number of edges of H. Now every spanning subgraph of G has at least as many connected components as G, so $c(H) \geq c(G)$. Moreover, a subgraph with j components must have at least $|V| - j$ edges. Thus $e(H) \geq |V| - c(H)$. The difference $e(H) - |V| + c(H)$ is sometimes called the *cycle rank* or *cyclomatic number* of H; it is the maximum number of edges that can be removed from H without increasing the number of components.

Thus we may consider the polynomial

$$R_G(\alpha,\beta) = \sum_H \alpha^{c(H)-c(G)}\beta^{e(H)-|V|+c(H)},$$

which is related to $S_G(\alpha,\beta)$ by

$$R_G(\alpha,\beta) = \alpha^{-c(G)}\beta^{-|V|}S_G(\alpha\beta,\beta)$$

and

$$S_G(\alpha,\beta) = \alpha^{c(G)}\beta^{|V|-c(G)}R_G(\alpha/\beta,\beta).$$

We now define the *Tutte polynomial* of G by

$$t_G(\alpha,\beta) = R_G(\alpha-1,\beta-1). \tag{9}$$

Accounts of the basic properties of Tutte polynomials can be found in Biggs [2] and Björner [3]. We note here only that it is well known that the coefficients of $t_G(\alpha,\beta)$ are nonnegative for any graph, and as Tutte showed, they can be interpreted as counting spanning trees of G by statistics called *internal* and *external activity*. A generalization of these statistics, which includes the interpretations discussed in this paper, can be found in [6].

The Tutte polynomial $t_G(\alpha,\beta)$ is related to $S_G(\alpha,\beta)$ by

$$\begin{aligned} t_G(\alpha,\beta) &= (\alpha-1)^{-c(G)}(\beta-1)^{-|V|}S_G\big((\alpha-1)(\beta-1),\beta-1\big) \\ S_G(\alpha,\beta) &= \alpha^{c(G)}\beta^{|V|-c(G)}t_G(\alpha/\beta+1,\beta+1). \end{aligned} \tag{10}$$

Note that each of the three graph polynomials S_G, R_G, and t_G is multiplicative in the sense that its value for any graph is the product of its values for the connected components of the graph. From now on we assume that G is connected.

We now derive analogs for an arbitrary connected graph of the formulas of Section 1. Instead of exponential generating functions, we get formulas involving sums over partitions. First we fix a vertex v of G and consider the depth-first decomposition applied to connected subgraphs of G rooted at v. We see that every connected subgraph of G can be obtained uniquely by first choosing a partition $\{V_1,\dots,V_k\}$ of $V-\{v\}$, and then choosing, for each i from 1 to k, a connected subgraph H_i of G with vertex set V_i and a nonempty subset of the set of edges in G joining V_i to v. Now let $C_G(\beta)$ count connected subgraphs of G by edges, so that $C_G(\beta)$ is the coefficient of α in $S_G(\alpha,\beta)$, and let $I_G(\beta)=t_G(1,\beta)$. Then by (10),

$$C_G(\beta)=\beta^{|V|-1}I_G(\beta+1). \tag{11}$$

For each subset U of $V-\{v\}$, let $G[U]$ be the induced subgraph of G with vertex set U, and let $\epsilon(U)$ be the number of vertices of U adjacent to v. We can now give the generalizations of Theorems 1 and 2:

Theorem 8.

$$C_G(\beta)=\sum_{V_1,V_2,\dots,V_k}\prod_{i=1}^{k}\big((\beta+1)^{\epsilon(V_i)}-1\big)C_{G[V_i]}(\beta), \tag{12}$$

$$I_G(\beta)=\sum_{V_1,V_2,\dots,V_k}\prod_{i=1}^{k}(1+\beta+\cdots+\beta^{\epsilon(V_i)-1})I_{G[V_i]}(\beta), \tag{13}$$

where the sums are over all partitions $\{V_1,\dots,V_k\}$, for all $k>0$, of $V-\{v\}$ with the property that each $G[V_i]$ is connected. (We interpret $1+\beta+\cdots+\beta^{m-1}$ as 0 for $m=0$.)

Proof. (12) follows immediately from the depth-first decomposition. From (11) and (12) we have

$$I_G(\beta+1)=\sum_{V_1,V_2,\dots,V_k}\prod_{i=1}^{k}\left(\frac{(\beta+1)^{\epsilon(V_i)}-1}{\beta}\right)I_{G[V_i]}(\beta+1), \tag{14}$$

and replacing β with $\beta-1$ in (14) we obtain (13). □

We can conclude from (13) that $I_G(\beta)$ has nonnegative coefficients and deduce from it a combinatorial interpretation for $I_G(\beta)$. It is clear from (13) that $I_G(1)$ is the number of spanning trees of G. To give a combinatorial interpretation to $I_G(\beta)$ via (13) in terms of a statistic on spanning trees of G, we need inductively a combinatorial interpretation to each $I_{G[V_i]}(\beta)$ (which will depend on the choice of a root for $G[V_i]$), and then we need a bijection between $\{0, 1, \ldots, \epsilon(V_i) - 1\}$ and the set of $\epsilon(V_i)$ edges joining V_i to v. One way to do this is to start by totally ordering V and always choosing the least possible vertex as the root. When we do this, we arrive at the following statistic on spanning trees of G. First we root G at its least vertex, say v. Now to any edge $f = \{x, y\}$ of T, we assign an integer $\kappa_T(f)$: without loss of generality suppose that x is the parent of y. Then y is greater than exactly $\kappa_T(f)$ of the vertices that are descendants of y in T and that are adjacent to x in G. We define $\kappa(T)$ to be $\sum_f \kappa_T(f)$, where the sum is over all edges f of T.

It is easily seen that if G is a complete graph then $\kappa(T)$ is the number of inversions of T. In the general case, $\kappa(T)$ is the number of inversions (y, z) of T such that the parent of y is adjacent in G to z. Then we have the following generalization of Theorem 3:

Theorem 9. *The coefficient of β^i in $I_G(\beta)$ is the number of spanning trees T of G with $\kappa(T) = i$.* □

We can also generalize (5) to

$$I_G(-1) = \sum_{\substack{V_1, \ldots, V_k \\ \epsilon(V_i) \text{ is odd}}} \prod_{i=1}^{k} I_{G[V_i]}(-1), \tag{15}$$

and (15) yields a combinatorial interpretation to $I_G(-1) = t_G(1, -1)$ generalizing Theorem 4:

Theorem 10. *$I_G(-1)$ is the number of spanning trees T of G with $\kappa(T) = 0$ and such that for every pair of vertices $\{x, y\}$ with x the parent of y, x is adjacent in G to an even number of descendants of y.* □

The generalization of Theorem 5 is completely straightforward:

Theorem 11.

$$S_G(\alpha, \beta) = \alpha \sum_{V_1, V_2, \ldots, V_k} \prod_{i=1}^{k} \left((\beta + 1)^{\epsilon(V_i)} - 1 + \alpha\right) S_{G[V_i]}(\alpha, \beta). \quad \square \tag{16}$$

We would now like to generalize Theorem 4 to an arbitrary connected graph. Unfortunately, a completely satisfactory generalization seems to exist only in the case in which v is adjacent to every other vertex of G. From (16) we deduce that

$$t_G(\alpha + 1, \beta) = \sum_{V_1, V_2, \ldots, V_k} \prod_{i=1}^{k} (\alpha + 1 + \beta + \cdots + \beta^{\epsilon(V_i) - 1}) I_{G[V_i]}(\beta), \tag{17}$$

recalling that $1 + \beta + \cdots + \beta^{m-1}$ is interpreted as 0 for $m = 0$. Note that if we set $\beta = -1$ in (17), we find that the coefficients of $t_G(\alpha + 1, -1)$ are nonnegative, and it is easy to give a combinatorial interpretation to them.

We may replace α with $\alpha - 1$ in (17) but if $\epsilon(V_i) = 0$ for some i then we will have an undesirable factor of $\alpha - 1$. However, if v is adjacent to every other vertex (as happens in particular for complete graphs) then there is no problem, and we have a nice generalization of (8):

Theorem 12. *Suppose that v is adjacent to every other vertex of G. Then*

$$t_G(\alpha,\beta) = \sum_{V_1,V_2,\ldots,V_k} \prod_{i=1}^{k} (\alpha+\beta+\cdots+\beta^{\epsilon(V_i)-1}) I_{G[V_i]}(\beta), \tag{18}$$

and the coefficient of $\alpha^i\beta^j$ in $t_G(\alpha,\beta)$ is the number of spanning trees T of G with $\kappa(T)=j$ and such that $\kappa(f)=0$ for exactly i edges f incident with v. □

References

1. J. S. Beissinger, *On external activity and inversions in trees*, J. Combin. Theory Ser. B **33** (1982), 87–92.
2. N. Biggs, *Algebraic graph theory*, Cambridge University Press, Cambridge, 1974.
3. A. Björner, *The homology and shellability of matroids and geometric lattices*, Matroid Applications, ed. N. White, Cambridge University Press, 1991, pp. 226–283.
4. D. Foata, *La série génératrice exponentielle dans les problèmes d'énumeration*, Presses de l'Université de Montréal, 1974.
5. I. Gessel, *A noncommutative generalization and q-analog of the Lagrange inversion formula*, Trans. Amer. Math. Soc. **257** (1980), 455-482.
6. I. M. Gessel and B. E. Sagan, *The Tutte polynomial of a graph, depth-first search, and simplicial complex partitions*, in preparation.
7. I. M. Gessel, B. E. Sagan, and Y.-N. Yeh, *Enumeration of trees by inversions*, preprint.
8. I. M. Gessel and D.-L. Wang, *Depth-first search as a combinatorial correspondence*, J. Combin. Theory Ser. A **26** (1979), 308–313.
9. F. Harary and E. M. Palmer, *Graphical Enumeration*, Academic Press, New York, 1973.
10. G. Kreweras, *Une famille du pôlynomes ayant plusieurs propriétés énumératives*, Periodica Math. Hungarica **11** (1980), 309–320.
11. P. Leroux, *Methoden der Anzahlbestimmung für einege Klassen von Graphen*, Bayreuther mathematische Schriften, no. 26, 1988, pp. 1–36.
12. C. L. Mallows and J. Riordan, *The inversion enumerator for labeled trees*, Bull Amer. Math. Soc. **74** (1968), 92–94.
13. J.-J. Pansiot, *Nombres d'Euler et inversions dans les arbres*, Europ. J. Combin. **3** (1982), 259–262.

Free Hyperplane Arrangements Interpolating Between Root System Arrangements

by
Tadeusz Józefiak
and
Bruce Sagan[1]

Abstract

Let R and S be root systems with $R \subset S$. By adding the roots of $S \setminus R$ to R one at a time, one obtains a sequence of subsets each of which determines a hyperplane arrangement. It turns out that these arrangements are often free and so the associated characteristic polynomials have non-negative integer roots. Zaslavsky [Zas 81] was the first to consider the family of hyperplane arrangements interpolating between D_n and B_n. These investigations were continued by Cartier [Car 82], Orlik and Solomon [O-S 83], Orlik-Solomon-Terao [J-T 80, Example 2.6], Ziegler [Zie 90], and Hanlon [Han pr]. Surprisingly, other interpolating families seem not to have been studied previously. In the present work we will show that some of these families are free by explicitly calculating bases for the corresponding modules of derivations. As immediate corollaries, we can read off the roots of their characteristic polynomials.

Let

$$\mathcal{A} = \{H_1, \cdots, H_k\} \tag{1}$$

be an arrangement (set) of hyperplane subspaces in the Euclidean space $\mathbf{R}^n$. Let $L = L(\mathcal{A})$ be the poset of intersections of these hyperplanes ordered by reverse inclusion. Thus L has a unique minimal element $\hat{0}$ corresponding to $\mathbf{R}^n$, an atom corresponding to each H_i, and a unique maximal element $\hat{1}$ corresponding to $\cap_{1 \leq i \leq k} H_i$. It is well-known that L is a geometric lattice with rank function

$$\operatorname{rk} X = n - \dim X$$

for any $X \in L$. Let $\mu(X) = \mu(\hat{0}, X)$ denote the Möbius function of the lattice. Then the *characteristic polynomial of* L is

$$\chi(L, t) = \sum_{X \in L} \mu(X) t^{\dim X}.$$

Now consider the polynomial algebra $A = \mathbf{R}[x_1, \ldots, x_n] = \mathbf{R}[x]$ with the usual grading by total degree $A = \oplus_{i \geq 0} A_i$. A *derivation* is an $\mathbf{R}$-linear map

$$\theta : A \to A$$

[1]Both authors are currently at: Department of Mathematics, Michigan State University, East Lansing, MI 48824-1027

satisfying

$$\theta(fg) = f\theta(g) + g\theta(f)$$

for any $f, g \in A$. The set of all derivations is a A-module. It is graded by saying that θ has degree d if $\theta(A_i) \subseteq A_{i+d}$. This module is also free with a basis given by the operators $\partial/\partial x_1, \ldots, \partial/\partial x_n$. It will often be convenient to display a derivation as a column vector whose entries are its components with respect to this basis. Thus if

$$\theta = p_1(x)\partial/\partial x_1 + \cdots + p_n(x)\partial/\partial x_n$$

where $p_i(x) \in \mathbf{R}[x]$ for all i, then we write

$$\theta = \begin{bmatrix} p_1(x) \\ \vdots \\ p_n(x) \end{bmatrix} = \begin{bmatrix} \theta(x_1) \\ \vdots \\ \theta(x_n) \end{bmatrix}.$$

Let $e_1, \ldots, e_n$ denote the coordinate vectors in $\mathbf{R}^n$ with the variables $x_1, \ldots, x_n$ being considered as elements of the corresponding dual basis. So any hyperplane $H \subseteq \mathbf{R}^n$ is defined by an equation

$$p_H(x_1, \ldots, x_n) = 0$$

where p_H is a linear polynomial. Thus the arrangement $\mathcal{A}$ in (1) is defined by the form

$$Q = Q(\mathcal{A}) = \prod_i p_{H_i}.$$

Consider the associated *module of $\mathcal{A}$-derivations* defined by

$$D(\mathcal{A}) = \{\theta \mid \theta \text{ a derivation and } \theta(Q) \in Q \cdot \mathbf{R}[x]\}.$$

We say that $\mathcal{A}$ is a *free arrangement* if $D(\mathcal{A})$ is a free module. Terao first introduced free arrangements and proved the following fundamental theorem [Ter 81, Ter 83]. A simpler proof was obtained with Solomon [S-T 87].

Theorem 1 *If $\mathcal{A}$ is free then*

1. *$D(\mathcal{A})$ has a homogeneous basis $\theta_1, \ldots, \theta_n$,*

2. *the set*

$$\{d_1, \ldots, d_n\} = \{\deg \theta_1, \ldots, \deg \theta_n\}$$

depends only on $\mathcal{A}$,

3. *the characteristic polynomial of $\mathcal{A}$ factors as*

$$\chi(L(\mathcal{A}), t) = \prod_i (t - d_i - 1). \blacksquare$$

In order to find such homogeneous bases, we use a result whose holomorphic version is due to Saito [Sai 80], and whose algebraic analogue comes from Terao [Ter 83] and Solomon-Terao [S-T 87]. Given any set of derivations $\theta_1, \ldots, \theta_n$, consider the rectangular matrix whose columns are the corresponding column vectors

$$\Theta = [\theta_1, \ldots, \theta_n] = [\theta_j(x_i)].$$

Theorem 2 *Suppose $\theta_1, \ldots, \theta_n \in D(\mathcal{A})$ where $\mathcal{A}$ has defining form Q. Then the following conditions are equivalent:*

1. $\det \Theta = cQ$ *where $c \in \mathbf{R}$ is non-zero,*

2. *$\mathcal{A}$ is free with basis $\theta_1, \ldots, \theta_n$.* ■

Thus we can prove that an arrangement $\mathcal{A}$ is free by constructing homogeneous derivations that

1. are in the submodule of $\mathcal{A}$-derivations and

2. have the proper determinant.

Often, the hardest part of the proof is showing that the scalar c in part 1 of Theorem 2 is non-zero. In some cases this step involves interesting new determinants related to those of Jacobi-Trudi [Sag ta].

To state our results, we will need a bit more notation. Any finite set $P \subseteq \mathbf{R}^n$ of vectors gives rise to the arrangement whose hyperplane subspaces are $H = p^\perp$ for $p \in P$. Let $\chi(P, t)$ and $\Theta(P)$ stand, respectively, for the corresponding characteristic polynomial and matrix for a basis of derivations. Also define column vectors

$$x^k = \begin{bmatrix} x_1^k \\ \vdots \\ x_n^k \end{bmatrix} \qquad \text{and} \qquad \hat{x} = \begin{bmatrix} \hat{x}_1 x_2 \cdots x_n \\ x_1 \hat{x}_2 \cdots x_n \\ \vdots \\ x_1 x_2 \cdots \hat{x}_n \end{bmatrix}$$

where $\hat{x}_i$ means that x_i is omitted.

We first interpolate between the root systems D_n and B_n.

Theorem 3 *Let*

$$DB_{n,k} = D_n \cup \{e_1, \ldots, e_k\}.$$

Then $DB_{n,k}$ is free with basis matrix

$$\Theta(DB_{n,k}) = [x^1, x^3, \ldots, x^{2n-3}, \theta_n]$$

where

$$\theta_n = x_1 x_2 \cdots x_k \hat{x}.$$

Thus

$$\chi(DB_{n,k}, t) \text{ has roots } 1, 3, \ldots, 2n-3, n+k-1. \blacksquare$$

By symmetry, it is clear that adding the e_i in any order would produce a free arrangement with the same characteristic polynomial.

When interpolating between A_{n-1} and B_n, the order in which the roots are added matters. First we add $e_1, \ldots, e_n$. The remaining roots can be listed in a triangular array

$$\begin{array}{cccc} e_1+e_2 & e_1+e_3 & \cdots & e_1+e_n \\ & e_2+e_3 & \cdots & e_2+e_n \\ & & & \vdots \\ & & & e_{n-1}+e_n \end{array}$$

We can add the $e_i + e_j$ *by columns* where we read each column from top to bottom, or *by rows* where we read each row from left to right. To describe the basis matrices, let

$$E_k(t) = t(t-x_1)(t-x_2)\cdots(t-x_k)$$

so that the coefficients of powers of t in $E_k(t)$ are elementary symmetric functions in the first k variables. The corresponding column vectors are

$$E_k = \begin{bmatrix} E_k(x_1) \\ \vdots \\ E_k(x_n) \end{bmatrix}.$$

Note that the first k entries of E_k are zero.

Theorem 4 *Interpolate from A_{n-1} to B_n by columns by letting*

$$AB^c_{n,k,l} = A_{n-1} \cup \{e_1, \ldots, e_n\} \cup \{e_1+e_2, e_1+e_3, e_2+e_3, \ldots, e_k+e_l\}.$$

Then $AB^c_{n,k,l}$ is free with basis matrix

$$\Theta(AB^c_{n,k,l}) = [x^1, x^3, \ldots, x^{2l-3}, \theta_l, E_l, E_{l+1}, \ldots, E_{n-1})$$

where

$$\theta_l = (x_1+x_l)(x_2+x_l)\cdots(x_k+x_l)E_{l-1}$$

Thus

$$\chi(AB^c_{n,k,l}, t) \textit{ has roots } 1, 3, \ldots, 2l-3, k+l, l+1, l+2, \ldots, n. \ \blacksquare$$

To interpolate by rows, define

$$E_{k,l}(t) = t(t+x_1)(t+x_2)\cdots(t+x_k)(t-x_1)(t-x_2)\cdots(t-x_l).$$

with associated column vector

$$E_{k,l} = \begin{bmatrix} E_{k,l}(x_1) \\ \vdots \\ E_{k,l}(x_n) \end{bmatrix}.$$

Theorem 5 *Interpolate from A_{n-1} to B_n by rows by letting*

$$AB^r_{n,k,l} = A_{n-1} \cup \{e_1, \ldots, e_n\} \cup \{e_1 + e_2, \ldots, e_1 + e_n, e_2 + e_3, \ldots, e_2 + e_n, \ldots, e_k + e_l\}.$$

Then $AB^r_{n,k,l}$ is free with basis matrix

$$\Theta(AB^r_{n,k,l}) = [x^1, x^3, \ldots, x^{2k-1}, E_{k,k}, E_{k,k+1}, \ldots, E_{k,l-1}, E_{k-1,l}, E_{k-1,l+1}, \ldots, E_{k-1,n-1}].$$

Thus

$\chi(AB^r_{n,k,l}, t)$ *has roots* $1, 3, \ldots, 2k-1, 2k+1, 2k+2, \ldots, k+l, k+l, k+l+1, \ldots, n.$ ∎

Similar theorems hold for interpolation between A_{n-1} and D_n. One can also get results for arrangement interpolating between a root system and itself, e.g., from D_n to D_{n+1}. It is interesting to note that many, although not all, of the results we have obtained can be generalized to the Dowling lattices (using hyperplanes of the form $x_i + \zeta x_j$ as ζ runs through all rth roots of unity).

Other methods for proving these results are also being investigated. One can compute individual Möbius values in the various families introduced above and prove that their characteristic polynomials factor directly as Hanlon did for $DB_{n,k}$. Finally, Curtis Bennett and Sagan have developed a generalization of the notion of supersolvability which can be used to combinatorially prove factorization of $\chi(DB_{n,k}, t)$ though the lattices are not supersolvable for $k < n$. This method should extend to the other cases under consideration as well.

Acknowledgement. This research was begun while Sagan was visiting UCSD. He would like to thank Adriano Garsia for posing the problem of explaining the factorization of characteristic polynomials for certain non-supersolvable posets, in particular for the lattice associated with D_n. We would also like to thank Günter Ziegler for helpful comments.

References

[Car 82] P. Cartier, Les arrangements d'hyperplans: un chapitre de géométrie combinatoire, in "Séminaire Bourbaki, Volume 1980/81, Exposés 561–578," Lecture Notes in Math., Vol. 901, Springer-Verlag, New York, NY, 1982, 1–22.

[Han pr] P. Hanlon, A combinatorial construction of posets that intertwine the independence matroids of B_n and D_n, preprint.

[J-T 80] M. Jambu and H. Terao, Free arrangements of hyperplanes and supersolvable lattices, *Adv. in Math.* **52** (1984), 248–258.

[Orl 89] P. Orlik, "Introduction to Arrangements," Regional Conference Series in Math., No. 72, American. Math. Soc., Providence, RI.

[O-S 83] P. Orlik and L. Solomon, Coxeter arrangements, in Proc. Symp. Pure Math., Vol. 40, part 2, Amer. Math. Soc., Providence, RI, 1983, 269–291.

[Sag ta] B. E. Sagan, Log concave sequences of symmetric functions and analogs of the Jacobi-Trudi determinants, *Trans. Amer. Math. Soc.*, to appear.

[Sai 80] K. Saito, Theory of logarithmic differential forms and logarithmic vector fields, *J. Fac. Sci. Univ. Tokyo Sec. 1A Math.* **27** (1980), 265–291.

[S-T 87] L. Solomon and H. Terao, A formula for the characteristic polynomial of an arrangement, *Adv. in Math.* **64** (1987), 305–325.

[Ter 80] H. Terao, Arrangements of hyperplanes and their freeness, *J. Fac. Sci. Univ. Tokyo Sec. 1A Math.* **27** (1980), 293–312.

[Ter 81] H. Terao, Generalized exponents of a free arrangement of hyperplanes and the Shepherd-Todd-Brieskorn formula, *Invent. Math.* **63** (1981), 159–179.

[Ter 83] H. Terao, Free arrangements of hyperplanes over an arbitrary field, *Proc. Japan Acad. Ser. A Math* **59** (1983), 301–303.

[Zas 81] T. Zaslavsky, The geometry of root systems and signed graphs, *Amer. Math. Monthly* **88** (1981), 88–105.

[Zie 90] G. Ziegler, Matroid representations and free arrangements, *Trans. Amer. Math. Soc.* **320** (1990), 525–541.

Counting tableaux with row and column bounds

C. KRATTENTHALER†
S. G. MOHANTY‡

ABSTRACT. It is well-known that the generating function for tableaux of a given skew shape with r rows where the parts in the i'th row are bounded by some nondecreasing upper and lower bounds which depend on i can be written in form of a determinant of size r. We show that the generating function for tableaux of a given skew shape with r rows and c columns where the parts in the i'th row are bounded by nondecreasing upper and lower bounds which depend on i *and* the parts in the j'th column are bounded by nondecreasing upper and lower bounds which depend on j can also be given in determinantal form. The size of the determinant now is $r + 2c$. We also show that determinants can be obtained when the nondecreasingness is dropped. Subsequently, analogous results are derived for (α, β)-plane partitions.

1. Introduction and Definitions. Let $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_r)$ and $\boldsymbol{\mu} = (\mu_1, \dots, \mu_r)$ be r-tupels of integers satisfying $\lambda_1 \geq \cdots \geq \lambda_r$, $\mu_1 \geq \cdots \geq \mu_r$, and $\boldsymbol{\lambda} \geq \boldsymbol{\mu}$, meaning $\lambda_i \geq \mu_i$ for all i. A *tableau of shape* $\boldsymbol{\lambda}/\boldsymbol{\mu}$ is an array π

$$
\begin{array}{ccccccc}
 & & & \pi_{1,\mu_1+1} & \cdots\cdots\cdots\cdots & & \pi_{1,\lambda_1} \\
 & \pi_{2,\mu_2+1} & \cdots & \pi_{2,\mu_1+1} & \cdots\cdots & \pi_{2,\lambda_2} & \\
 & \cdot^{\cdot^{\cdot}} & & \vdots & & \cdot^{\cdot^{\cdot}} & \\
\pi_{1,\mu_r+1} & \cdots\cdots\cdots\cdots & & & \pi_{1,\lambda_r} & &
\end{array}
\tag{1.1}
$$

of integers π_{ij}, $1 \leq i \leq r$, $\mu_i + 1 \leq j \leq \lambda_i$, such that the rows are weakly and the columns are strictly increasing. The number of entries in the tableau in (1.1) is $(\lambda_1 - \mu_1) + \cdots + (\lambda_r - \mu_r)$, for which we write $|\boldsymbol{\lambda} - \boldsymbol{\mu}|$. The entries will be called *parts* of the tableau. The sum of all parts of a tableau π is called the *norm*, in symbols $n(\pi)$, of the tableau.

In order to make $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ unique, we always assume that $\mu_r = 0$. Sometimes we will call λ_1 the *width*, and r the *depth* of a shape $\boldsymbol{\lambda}/\boldsymbol{\mu}$. If $\boldsymbol{\mu} = 0$ we shortly write $\boldsymbol{\lambda}$ for the shape $\boldsymbol{\lambda}/\boldsymbol{\mu}$.

The weight $w(\pi)$ of a tableau π under consideration will be $\prod x_{\pi_{ij}}$ where the product is over all parts π_{ij} of π.

†Institut für Mathematik der Universität Wien, Strudlhofgasse 4, A-1090 Wien, Austria.
E-mail: KRATT@AWIRAP.BITNET
‡McMaster University, Hamilton, Ontario, Canada L8S 4K1.
E-mail: MOHANTY@MCMASTER.BITNET

It is well-known (cf. [2,4,5,9]) that the generating function $\sum w(\pi)$ summed over all tableaux π of shape $\boldsymbol{\lambda}/\boldsymbol{\mu}$ where the parts in row i are at most a_i and at least b_i for some r-tupels $\boldsymbol{a} = (a_1, \dots, a_r)$ and $\boldsymbol{b} = (b_1, \dots, b_r)$ satisfying $a_1 \le a_2 \le \cdots \le a_r$, $b_1 \le b_2 \le \cdots \le b_r$, and $\boldsymbol{a} \ge \boldsymbol{b}$, can be written in form of an $r \times r$-determinant,

$$\det_{1\le s,t\le r} \left(h_{\lambda_s - s - \mu_t + t}(\boldsymbol{x}; a_t, b_s)\right) ,$$

where $h_n(\boldsymbol{x}; A, B)$ is the complete homogenous symmetric functions of order n in the variables $x_B, x_{B+1}, \dots, x_A$,

$$h_n(\boldsymbol{x}; A, B) := \sum_{B\le i_1 \le i_2 \le \cdots \le i_n \le A} x_{i_1} x_{i_2} \cdots x_{i_n} .$$

A natural generalization of this problem is to ask for the generating function $\sum w(\pi)$ for tableaux with row bounds *and* column bounds, to be precise, for the generating function for tableaux of shape $\boldsymbol{\lambda}/\boldsymbol{\mu}$ where the parts in row i are at most a_i and at least b_i, and the parts in column j are at most c_j and at least d_j, for some tupels $\boldsymbol{a} = (a_1, \dots, a_r)$, $\boldsymbol{b} = (b_1, \dots, b_r)$, $\boldsymbol{c} = (c_1, \dots, c_{\lambda_1})$, and $\boldsymbol{d} = (d_1, \dots, d_{\lambda_1})$ satisfying $a_1 \le a_2 \le \cdots \le a_r$, $b_1 \le b_2 \le \cdots \le b_r$, $c_1 \le c_2 \le \cdots \le c_{\lambda_1}$, $d_1 \le d_2 \le \cdots \le d_{\lambda_1}$, $\boldsymbol{a} \ge \boldsymbol{b}$, and $\boldsymbol{c} \ge \boldsymbol{d}$. In Theorem 1 we show that this generating function can also be written in form of a determinant whose entries are complete homogenous symmetric functions. The size of the determinant is $r + 2\lambda_1$, i.e. it is the depth plus twice the width of the shape. Also in section 2, from this theorem we deduce determinant formulas for the norm generating function of (α, β)-reverse plane partitions (which are generalizations of tableaux, cf. section 2 for the definition) with row and column bounds, thus obtaining Corollary 2. Finally, in section 3 we show how to get determinants if the monotonicity of the row and column bounds is dropped. Now in adverse choices of the shape and the row and column bounds, the size of the determinant might explode.

2. Monotone row and column bounds. Recall (Gessel and Viennot [3,4]) that a tableau π of shape $\boldsymbol{\lambda}/\boldsymbol{\mu}$ with the parts in row i being at most a_i and at least b_i can be bijectively mapped onto a family $(P_1, \dots, P_r)$ of nonintersecting lattice paths. "Nonintersecting" in this context means that each two paths of this family have no point in common. This correspondence maps the i-th row of π to the i-th path P_i in the family such that P_i starts at $(\mu_i + r + 1 - i, b_i)$ and terminates at $(\lambda_i + r + 1 - i, a_i)$, and such that the parts of the i-th row can be read off the heights of the horizontal steps in P_i. Figure 1 gives a simple example for $r = 3$, $\boldsymbol{\lambda} = (5,4,4)$, $\boldsymbol{\mu} = (2,1,0)$, $\boldsymbol{a} = (8,9,12)$, and $\boldsymbol{b} = (1,3,6)$.

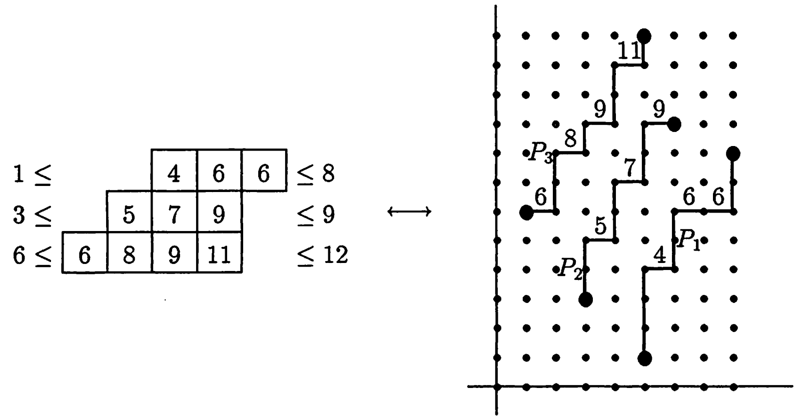


Figure 1

The condition that π has strictly increasing columns corresponds to the condition that the paths are nonintersecting. In addition, this bijection is weight-preserving if we define the weight of a family $\mathcal{P} = (P_1, \dots, P_r)$ of paths to be

$$w(\mathcal{P}) = \prod x_h \ ,$$

where the product is over the heights h of all the horizontal steps of the paths.

We want to compute the generating function $\sum w(\pi)$ for tableaux π of shape $\boldsymbol{\lambda}/\boldsymbol{\mu}$ where the parts in row i are at most a_i and at least b_i, $i = 1, 2, \dots, r$, and the parts in column j are at most c_j and at least d_j, $j = 1, 2, \dots, \lambda_1$. Note that we always assume $\mu_r = 0$ so that there are lower and upper bounds for each column. To abbreviate the notation, for these tableaux we shall often use the terminology *tableaux which obey the row bounds* $\boldsymbol{a}, \boldsymbol{b}$ *and the column bounds* $\boldsymbol{c}, \boldsymbol{d}$, or even shorter *tableaux obeying* $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \boldsymbol{d}$, always assuming that $\boldsymbol{a}$ and $\boldsymbol{b}$ are the upper and lower row bounds while $\boldsymbol{c}$ and $\boldsymbol{d}$ are the upper and lower column bounds, respectively.

A simple trick enables us to use nonintersecting paths for this generalized problem either. This is accomplished by adding $2\lambda_1$ "dummy paths" of length 0. In fact, tableaux π of shape $\boldsymbol{\lambda}/\boldsymbol{\mu}$ obeying $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \boldsymbol{d}$ bijectively correspond to families $(P_1, \dots, P_r, P_{r+1}, \dots, P_{r+2\lambda_1})$ of nonintersecting lattice paths where for $i = 1, \dots, r$ by using the

Gessel/Viennot bijection the path P_i is obtained from the i'th row of π respecting the row bounds $\boldsymbol{a}$ and $\boldsymbol{b}$, $P_i : (\mu_i + r + 1 - i, b_i) \to (\lambda_i + r + 1 - i, a_i)$, while the paths P_l, $l = r+1, \ldots, r + 2\lambda_1$, are dummy paths of length 0, the starting and final points of which are given below.

$$P_l: \ (l' + M(l'), d_{l'} - 1) \to (l' + M(l'), d_{l'} - 1) \qquad l = r+1, \ldots, r+\lambda_1,$$

with $l' = l - r$, and

$$P_l: \ (l'' + \Lambda(l''), c_{l''} + 1) \to (l'' + \Lambda(l''), c_{l''} + 1) \qquad l = r + \lambda_1 + 1, \ldots, r + 2\lambda_1,$$

with $l'' = l - r - \lambda_1$. The auxiliary functions M and Λ are defined by

$$M(I) = \sum_{j=1}^{r} \chi(I > \mu_j) \quad \text{and} \quad \Lambda(I) = \sum_{j=1}^{r} \chi(I > \lambda_j) \ ,$$

where $\chi(\mathcal{A})$ is the usual truth function ($\chi(\mathcal{A}) = 1$ if $\mathcal{A}$ is true, and $\chi(\mathcal{A})$=0 otherwise). Figure 2 gives an example for this correspondence with $r = 4$, $\lambda = (8,6,4,3)$, $\mu = (3,2,0,0)$, $\boldsymbol{a} = (10,12,13,13)$, $\boldsymbol{b} = (1,1,2,3)$, $\boldsymbol{c} = (6,6,9,10,11,11,12,12)$, $\boldsymbol{d} = (2,2,3,3,3,4,6,7)$.

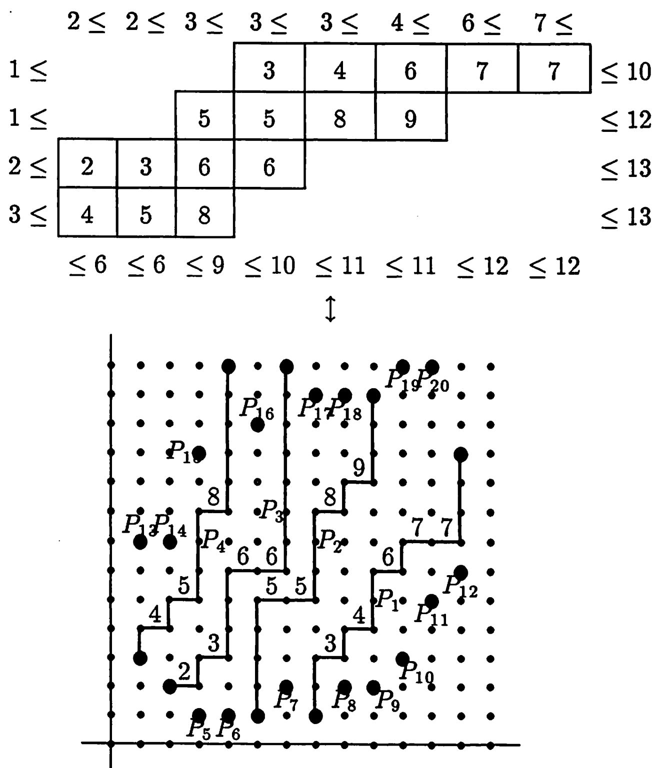


Figure 2

One easily gets convinced that $P_{r+1},\dots,P_{r+\lambda_1}$ force $P_1,\dots,P_r$ to stay above them, and, analogously, that $P_{r+\lambda_1+1},\dots,P_{r+2\lambda_1}$ force $P_1,\dots,P_r$ to stay below them; otherwise $(P_1,\dots,P_r,P_{r+1},\dots,P_{r+2\lambda_1})$ would not be nonintersecting. But that means that the corresponding tableau obeys the column bounds $\boldsymbol{d}$ and $\boldsymbol{c}$.

Now, since we have reduced our tableaux counting problem to the problem of counting families of nonintersecting paths, we may apply the main theorem of Gessel and Viennot.

Proposition 1. *(Gessel, Viennot [4, sec. 2]) The generating function $\sum w(\mathcal{P})$ for families $\mathcal{P}=(P_1,\dots,P_n)$ of nonintersecting paths, where P_i goes from (C_i,D_i) to (A_i,B_i) is given by*

$$\det_{1\le s,t\le n}\left(h_{A_s-C_t}(\boldsymbol{x};B_s,D_t)\right). \qquad \square \tag{1.2}$$

This yields the following theorem:

Theorem 1. *Let $\boldsymbol{a},\boldsymbol{b},\boldsymbol{c},\boldsymbol{d}$ as above. The generating function $\sum w(\pi)$ for tableaux π of shape $\boldsymbol{\lambda}/\boldsymbol{\mu}$ where the parts in row i are at most a_i and at least b_i, and the parts in column j are at most c_j and at least d_j is given by*

$$\det_{1\le s,t\le r+2\lambda_1}\left(h_{A_s^{(1)}-C_t^{(1)}}(\boldsymbol{x};B_s^{(1)},D_t^{(1)})\right), \tag{1.3}$$

where $A_i^{(1)},B_i^{(1)},C_i^{(1)},D_i^{(1)}$ are displayed in the table below.

	$1\le i\le r$	$r+1\le i\le r+\lambda_1$ $i'=i-r$	$r+\lambda_1+1\le i\le r+2\lambda_1$ $i''=i-r-\lambda_1$
$A_i^{(1)}$	$\lambda_i+r+1-i$	$i'+M(i')$	$i''+\Lambda(i'')$
$B_i^{(1)}$	a_i	$d_{i'}-1$	$c_{i''}+1$
$C_i^{(1)}$	$\mu_i+r+1-i$	$i'+M(i')$	$i''+\Lambda(i'')$
$D_i^{(1)}$	b_i	$d_{i'}-1$	$c_{i''}+1$

Table 1

$\square$

Setting $x_i=q^i$ for all i, we obtain as a corollary the norm generating function for tableaux with row and column bounds.

Corollary 1. *Let $\boldsymbol{a},\boldsymbol{b},\boldsymbol{c},\boldsymbol{d}$ as in Theorem 1. The generating function $\sum q^{n(\pi)}$ for tableaux π of shape $\boldsymbol{\lambda}/\boldsymbol{\mu}$ where the parts in row i are at most a_i and at least b_i, and the parts in column j are at most c_j and at least d_j is given by*

$$\det_{1\le s,t\le r+2\lambda_1}\left(q^{D_t^{(1)}(A_s^{(1)}-C_t^{(1)})}\begin{bmatrix}A_s^{(1)}+B_s^{(1)}-C_t^{(1)}-D_t^{(1)}\\ B_s^{(1)}-D_t^{(1)}\end{bmatrix}\right), \tag{1.4}$$

where $A_i^{(1)},B_i^{(1)},C_i^{(1)},D_i^{(1)}$ are displayed in Table 1. The q-binomial coefficient is defined by

$$\begin{bmatrix}n\\ k\end{bmatrix}=\frac{(1-q^n)(1-q^{n-1})\cdots(1-q^{n-k+1})}{(1-q^k)(1-q^{k-1})\cdots(1-q)},$$

if $n \geq k \geq 0$, *and* 0 *otherwise.* □

Applying this result to tableaux of shape $(8,6,4,3)/(3,2,0,0)$ obeying $\boldsymbol{a} = (10,12,13,13)$, $\boldsymbol{b} = (1,1,2,3)$, $\boldsymbol{c} = (6,6,9,10,11,11,12,12)$, and $\boldsymbol{d} = (2,2,3,3,3,4,6,7)$, a typical example of which is displayed in Figure 2, we obtain that the norm generating function for the 196.650.160 tableaux of this type is given by

$$
\begin{aligned}
&q^{63}+7q^{64}+31q^{65}+108q^{66}+318q^{67}+830q^{68}+1967q^{69}+4309q^{70}+8825q^{71}+17054q^{72}+31300q^{73}+54857q^{74}+92202q^{75}\\
&\quad+149150q^{76}+232896q^{77}+351926q^{78}+515736q^{79}+734334q^{80}+1017550q^{81}+1374118q^{82}+1810680q^{83}+2330671q^{84}\\
&\quad+2933374q^{85}+3613028q^{86}+4358418q^{87}+5152677q^{88}+5973797q^{89}+6795371q^{90}+7588103q^{91}+8321344q^{92}\\
&\quad+8965234q^{93}+9492509q^{94}+9880603q^{95}+10112996q^{96}+10180583q^{97}+10081990q^{98}+9823778q^{99}+9419561q^{100}\\
&\quad+8889120q^{101}+8256606q^{102}+7549054q^{103}+6794310q^{104}+6019584q^{105}+5249728q^{106}+4506378q^{107}+3806938q^{108}\\
&\quad+3164499q^{109}+2587580q^{110}+2080681q^{111}+1644535q^{112}+1277020q^{113}+973600q^{114}+728280q^{115}+534013q^{116}\\
&\quad+383492q^{117}+269382q^{118}+184885q^{119}+123772q^{120}+80708q^{121}+51140q^{122}+31435q^{123}+18679q^{124}+10708q^{125}\\
&\quad+5890q^{126}+3102q^{127}+1549q^{128}+733q^{129}+322q^{130}+132q^{131}+48q^{132}+16q^{133}+4q^{134}+q^{135}.
\end{aligned}
$$

Corollary 1 can be generalized in the following way. Call an array $\overline{\pi}$ of the form (1.1) an (α,β)-*reverse plane partition of shape* $\boldsymbol{\lambda}/\boldsymbol{\mu}$ if

$$
\begin{aligned}
&\overline{\pi}_{ij} + \alpha \leq \overline{\pi}_{i,j+1} \qquad 1 \leq i \leq r,\ \mu_i + 1 \leq j < \lambda_i \\
(1.5) \qquad &\text{and} \\
&\overline{\pi}_{ij} + \beta \leq \overline{\pi}_{i+1,j} \qquad 1 \leq i < r,\ \mu_i + 1 \leq j \leq \lambda_{i+1}\ .
\end{aligned}
$$

This definition comprises several classes of reverse plane partitions. In particular, tableaux are $(0,1)$-reverse plane partitions.

Given a tableau π of shape $\boldsymbol{\lambda}/\boldsymbol{\mu}$, the transformation

$$\pi_{ij} \to \pi_{ij} + i(\beta - 1) + j\alpha\ ,$$

applied to every part π_{ij} of π, maps π to an (α,β)-reverse plane partition. Clearly, this mapping is a bijection between tableaux and (α,β)-reverse plane partitions. This bijection does not preserve $w(\pi)$ nor the norm $n(\pi)$. But for the norm we have the simple assertion that the norm of the (α,β)-reverse plane partition which was obtained from a certain tableau under this transformation differs from the norm of this tableau by $\sum \big(i(\beta-1)+j\alpha\big)$, where the sum is over all i,j with $1 \leq i \leq r$ and $\mu_i + 1 \leq j \leq \lambda_i$. This quantity only depends on the shape $\boldsymbol{\lambda}/\boldsymbol{\mu}$ and not on the tableau involved. Therefore, using this bijection it is an easy task to transfer Corollary 1 to the more general case of (α,β)-reverse plane partitions. One only has to find out how the row and column bounds change under this transformation.

Corollary 2. *Let* $\boldsymbol{a},\boldsymbol{b}$ *be* r*-tupels and* $\boldsymbol{c},\boldsymbol{d}$ *be* λ_1*-tupels of integers satisfying*

$$
(1.6) \qquad
\begin{aligned}
a_i + (\beta - 1) \leq a_{i+1}, \quad b_i + (\beta - 1) \leq b_{i+1} \\
c_i + \alpha \leq c_{i+1}, \quad d_i + \alpha \leq d_{i+1}\ .
\end{aligned}
$$

The generating function $\sum q^{n(\overline{\pi})}$ *for* (α,β)*-reverse plane partitions* $\overline{\pi}$ *of shape* $\boldsymbol{\lambda}/\boldsymbol{\mu}$ *where the last part in row* i *is at most* a_i *and the first part in row* i *is at least* b_i*, and where the down-most part in column* j *is at most* c_j *and the upper-most part in column* j *is at least* d_j*, is given by*

$$(1.7)\quad q^{(\beta-1)\sum_{i=1}^{r} i(\lambda_i-\mu_i)+\alpha\sum_{i=1}^{r}\left[\binom{\lambda_i+1}{2}-\binom{\mu_i+1}{2}\right]} \times \det_{1\le s,t\le r+2\lambda_1}\left(q^{D_t^{(2)}(A_s^{(2)}-C_t^{(2)})}\begin{bmatrix} A_s^{(2)}+B_s^{(2)}-C_t^{(2)}-D_t^{(2)} \\ B_s^{(2)}-D_t^{(2)}\end{bmatrix}\right),$$

where $A_i^{(2)}, B_i^{(2)}, C_i^{(2)}, D_i^{(2)}$ *are displayed in Table 2.*

	$1\le i\le r$	$r+1\le i\le r+\lambda_1$, $i'=i-r$	$r+\lambda_1+1\le i\le r+2\lambda_1$, $i''=i-r-\lambda_1$
$A_i^{(2)}$	$\lambda_i+r+1-i$	$i'+M(i')$	$i''+\Lambda(i'')$
$B_i^{(2)}$	$a_i-(\beta-1)i-\alpha\lambda_i$	$d_{i'}-(\beta-1)(\mu'_{i''}+1)-\alpha i''-1$	$c_{i''}-(\beta-1)\lambda'_{i''}-\alpha i''+1$
$C_i^{(2)}$	$\mu_i+r+1-i$	$i'+M(i')$	$i''+\Lambda(i'')$
$D_i^{(2)}$	$b_i-(\beta-1)i-\alpha(\mu_i+1)$	$d_{i'}-(\beta-1)(\mu'_{i''}+1)-\alpha i''-1$	$c_{i''}-(\beta-1)\lambda'_{i''}-\alpha i''+1$

Table 2

$\boldsymbol{\lambda}'/\boldsymbol{\mu}'$ *is the conjugate shape of* $\boldsymbol{\lambda}/\boldsymbol{\mu}$. □

REMARKS. 1) Theorem 1 and Corollary 2 immediately can be used for column-strict plane partitions and (α,β)-plane partitions (cf. [5]), respectively, since rotation by 180° turns them into tableaux and (α,β)-reverse plane partitions, respectively.

2) Very often the size of the determinant can be reduced by ignoring superflous paths. (In Figure 2 $P_5, P_6, P_{17}, P_{19}, P_{20}$ are superflous.) Also, (as long as the monotonicity of the row and column bounds is preserved) the entries in the determinant can be reduced by replacing a_i by $\min\{a_i, c_{\lambda_i}\}$, b_i by $\max\{b_i, d_{\mu_i+1}\}$, etc. For a rectangular shape (c^r) it is easily seen that after these manipulations P_{r+1} and P_{r+2c} are always superflous so that in this case the size of the determinant in fact is at most $r+2c-2$.

3) Reflection in the main diagonal turns an (α,β)-reverse plane partition of shape $\boldsymbol{\lambda}/\boldsymbol{\mu}$ into an (β,α)-reverse plane partition of shape $\boldsymbol{\lambda}'/\boldsymbol{\mu}'$. Therefore the norm generating function for (α,β)-reverse plane partitions of a given rectangular shape (c^r) and with given row and column bounds can be computed in two different ways using Corollary 2. The first way is to use Corollary 2 directly, thus (by Remark 2) obtaining a determinant of size $(r+2c-2)$. Or, secondly, one could use the reflection which exchanges α and β, r and c, $\boldsymbol{a}$ and $\boldsymbol{c}$, and $\boldsymbol{b}$ and $\boldsymbol{d}$, and then apply Corollary 2 with these new parameters. This time a determinant of size $(c+2r-2)$ is obtained. Therefore, in order to minimize the size of the determinant, one should first check whether $r\ge c$ or not, in the first case use Corollary 2 directly, in the latter Corollary 2 should be used only after first having performed the reflection. With a skew shape, in general Corollary 2 will not be applicable in two ways as described above because after the reflection the new bounds might not satisfy (1.6). But as will be seen later, it is possible to give a determinant for the "reflected" problem, either.

4) Of course, our formula can also be used if there are bounds only on three sides. In fact, in this case the determinants in Theorem 1 or Corollaries 1,2 trivially reduce to determinants of size $r + \lambda_1$. It should be noted that three bound counting for rectangular shapes has been considered earlier by Chorneyko and Zing [1, 8, Theorems 1.2.1 and 1.2.2]. Using the Narayana/Mohanty [cf. 7] method of successively building up determinants, they derived determinants for the number of rectangular tableaux with bounds $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ or $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{d}$, respectively. However, Chorneyko and Zing's determinants slightly differ from our $q = 1$-results. But they can also be explained by nonintersecting lattice paths. These determinants correspond to a slightly different choice of the dummy paths. For example, to obtain Chorneyko and Zing's determinant in the $(\boldsymbol{a},\boldsymbol{b},\boldsymbol{d})$-case for a shape (c^r) [8, Theorem 1.2.2], one had to take the dummy paths $P_l : (l, b_1 - 1) \to (l, \max\{b_1, d_{l-r}\} - 1)$, $l = r+1, \ldots, r+c$, instead of $P_l : (l, d_{l-r} - 1) \to (l, d_{l-r} - 1)$. The corresponding picture of a typical example is given in the figure below.

	$2 \le$	$4 \le$	$4 \le$	$5 \le$	
$1 \le$	3	4	6	7	≤ 8
$3 \le$	5	5	7	9	≤ 9
$6 \le$	6	8	9	11	≤ 12

$\longleftrightarrow$

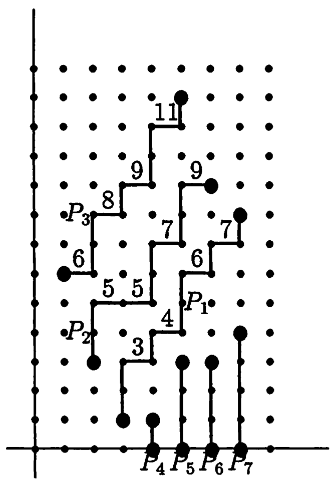


Figure 3

Clearly, the dummy paths in Figure 3 have just the same effect as those in Figure 2. We decided to take paths of length 0 because this choice causes the entries in the determinant to become smaller. $\square$

3. Unrestricted row and column bounds. It turns out that even if we drop the condition of the bounds being nondecreasing, the generating functions for tableaux can be given in determinantal form. Let $\boldsymbol{a},\boldsymbol{b}$ be arbitrary r-tupels of integers and $\boldsymbol{c},\boldsymbol{d}$ be arbitrary λ_1-tupels of integers only satisfying $\boldsymbol{a} \geq \boldsymbol{b}$ and $\boldsymbol{c} \geq \boldsymbol{d}$. We want to compute the generating function for tableaux of shape $\boldsymbol{\lambda}/\boldsymbol{\mu}$ where the parts in row i are at most a_i and at least b_i, and the parts in column j are at most c_j and at least d_j. Since now a formula for the entries of the resulting determinant would involve very clumsy expressions, it is more convenient to only give a rough description of the procedure which finally leads to the determinantal formula. Moreover, as will be seen later, it is better not to rigorously stick to a formula because very often in the last step of this procedure the size of the determinant can be significantly reduced, which would not be observed when a general formula is directly used.

Let us give a sketch of this procedure, which is performed in four steps. First, consider the case $\boldsymbol{\lambda} = (5,4,3)$, $\boldsymbol{\mu} = (2,0,0)$, $\boldsymbol{a} = (14,10,13)$, $\boldsymbol{b} = (4,1,4)$, $\boldsymbol{c} = (6,8,16,15,12)$, $\boldsymbol{d} = (2,7,1,7,5)$ which is illustrated on the left-hand side of Figure 4.

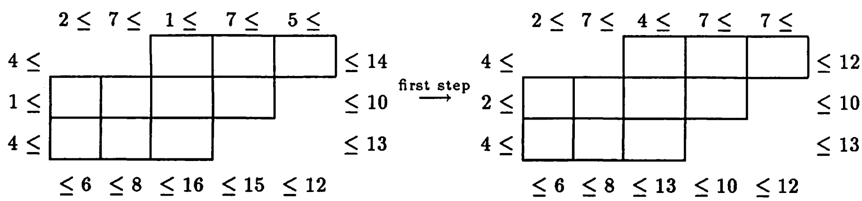


Figure 4

Obviously, some entries in $\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}$, or $\boldsymbol{d}$, respectively, could be replaced by greater respectively smaller ones without changing the set of tableaux obeying these bounds. For example, $d_5 = 5$ could be replaced by 7, $a_1 = 14$ by 12, etc. Formally, we may replace $\boldsymbol{a}$ by $\bar{\boldsymbol{a}}$, $\boldsymbol{b}$ by $\bar{\boldsymbol{b}}$, ... , where

$$\bar{a}_i = \begin{cases} \min\{a_i, a_{i+1}, \dots, a_{\lambda'_{\lambda_i}}\} & \lambda_{i+1} = \lambda_i \\ \min\{a_i, c_{\lambda_i}\} & \lambda_{i+1} < \lambda_i \end{cases}, \quad \bar{b}_i = \begin{cases} \max\{b_{\mu'_{\mu_i}+1}, \dots, b_{i-1}, b_i\} & \mu_{i-1} = \mu_i \\ \max\{b_i, d_{\mu_i+1}\} & \mu_{i-1} > \mu_i \end{cases},$$

$$\bar{c}_i = \begin{cases} \min\{c_i, c_{i+1}, \dots, c_{\lambda_{\lambda'_i}}\} & \lambda'_{i+1} = \lambda'_i \\ \min\{c_i, a_{\lambda'_i}\} & \lambda'_{i+1} < \lambda'_i \end{cases}, \quad \bar{d}_i = \begin{cases} \max\{d_{\mu_{\mu'_i}+1}, \dots, d_{i-1}, d_i\} & \mu'_{i-1} = \mu'_i \\ \max\{d_i, b_{\mu'_i+1}\} & \mu'_{i-1} > \mu'_i \end{cases}.$$

In our example, this normalization of the bounds is displayed on the right-hand side of Figure 4.

In the second step, tableaux of shape $\boldsymbol{\lambda}/\boldsymbol{\mu}$ obeying $\bar{\boldsymbol{a}},\bar{\boldsymbol{b}},\bar{\boldsymbol{c}},\bar{\boldsymbol{d}}$ are interpreted as families of lattice paths as was done before in order to prove Theorem 1.

	$2\le$	$7\le$	$4\le$	$7\le$	$7\le$	
$4\le$			4	9	9	≤ 12
$2\le$	3	7	7	10		≤ 10
$4\le$	4	8	8			≤ 13
	≤ 6	≤ 8	≤ 13	≤ 10	≤ 12	

$\downarrow$ **second step**

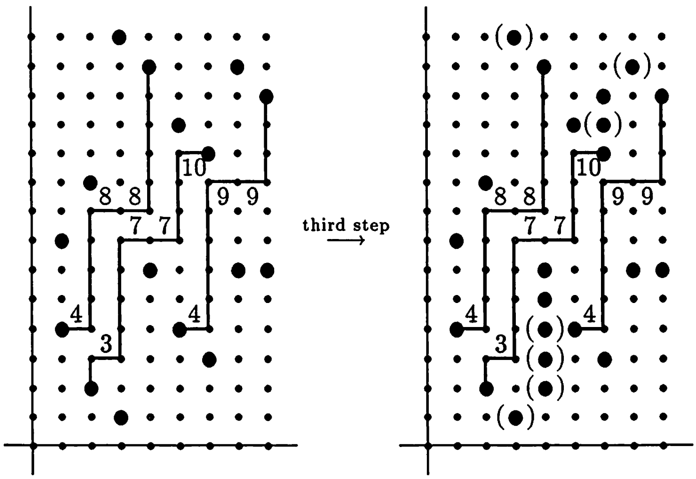
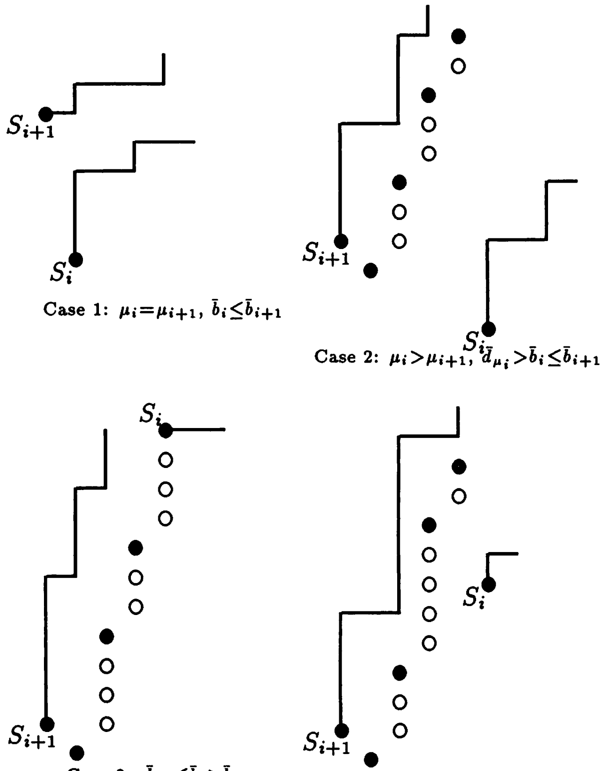


Figure 5

But now there are not enough dummy paths to guarantee that every family of nonintersecting paths corresponds to a tableaux of the desired type. Therefore, in the third step we have to insert additional dummy paths to build up "barriers" at some places. Let us consider the starting points S_i, S_{i+1} of the paths P_i, P_{i+1} which correspond to the i'th and $(i+1)$'th row of the tableau, respectively. Depending on which of the relations $\mu_i = \mu_{i+1}$, $\bar{b}_i \le \bar{b}_{i+1}$, $\bar{d}_{\mu_i} \le \bar{b}_i$ hold or not, we obtain four cases which schematically are described below.



Case 1: $\mu_i=\mu_{i+1}$, $\bar{b}_i\le\bar{b}_{i+1}$

Case 2: $\mu_i>\mu_{i+1}$, $\bar{d}_{\mu_i}>\bar{b}_i\le\bar{b}_{i+1}$

Case 3: $\bar{d}_{\mu_i}\le\bar{b}_i>\bar{b}_{i+1}$

Case 4: $\bar{d}_{\mu_i}>\bar{b}_i>\bar{b}_{i+1}$

The boldface dots different from S_i and S_{i+1} indicate the dummy paths which were inserted during Step 2. The circles indicate the dummy paths which have to be added to build up barriers which prevent P_{i+1} from going below P_i or not obeying $\boldsymbol{d}$, respectively. Only in Case 1 nothing has to be done. Considering the end points of P_i and P_{i+1}, dummy paths are added in an analogous manner in order to guarantee that the corresponding tableau is of the desired type. The result of this third step applied to our example is the family of paths on the right-hand side of Figure 5.

Finally, in the fourth step we look after paths which are superflous. These subsequently are dropped. In the right-hand side family of lattice paths in Figure 5 the superflous paths are put into parentheses. (It should be observed that paths should only be dropped sequentially because the same path could be superflous with respect to some set of paths which already has been dropped but not superflous with respect to another set. There can be several different sets of paths which can be legitimately dropped. One will choose that one which causes the corresponding determinant to become as "small" as possible.)

Now Proposition 1 can be applied thus again obtaining a determinant for the generating function. In our running example (Figures 4,5) this procedure yields that there are 17.163 tableaux of the desired type, the norm generating function of which is given by

$$q^{62}+6\,q^{63}+21\,q^{64}+55\,q^{65}+118\,q^{66}+221\,q^{67}+372\,q^{68}+572\,q^{69}+812\,q^{70}+1072\,q^{71}+1322\,q^{72}+1530\,q^{73}$$
$$+1665\,q^{74}+1707\,q^{75}+1650\,q^{76}+1503\,q^{77}+1290\,q^{78}+1041\,q^{79}+789\,q^{80}+561\,q^{81}+373\,q^{82}+231\,q^{83}$$
$$+132\,q^{84}+69\,q^{85}+32\,q^{86}+13\,q^{87}+4\,q^{88}+q^{89}.$$

Of course, with the help of this procedure the norm generating function for (α, β)-plane partitions or (α, β)-reverse plane partitions of shape $\boldsymbol{\lambda}/\boldsymbol{\mu}$ which obey arbitrary $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \boldsymbol{d}$, can also be computed by transforming the problem to the corresponding tableaux problem.

References

1. I. Z. Chorneyko, L. K. K. Zing, *K-sample analogues of two-sample Kolmogorov–Smirnov stastistics*, Selecta Statistica Canadiana **6** (1982), 37–62.
2. W. Chu, *Dominance method for plane partitions. IV: Enumeration of flagged skew tableaux*, Linear and Multilinear Algebra **28** (1990), 185–191.
3. Gessel, I., G. Viennot, *Binomial determinants, paths, and hook length formulae*, Adv. in Math. **58** (1985), 300–321.
4. I. Gessel, G. Viennot, *Determinants, paths, and plane partitions*, preprint, 1989.
5. C. Krattenthaler, *Generating functions for plane partitions of a given shape*, Manuscripta Math. **69** (1990), 173–202.
6. Macdonald, I. G., *Symmetric Functions and Hall Polynomials*, Oxford University Press, New York/London, 1979.
7. S. G. Mohanty, *Lattice path counting and applications*, Academic Press, New York, 1979.
8. L. K. K. Zing, *K-sample analogues of Kolmogorov–Smirnov statistics and group tests*, Ph. D. Thesis, McMaster University, 1978.
9. M. Wachs, *Flagged Schur functions, Schubert polynomials, and symmetrizing operators*, J. Combin. Theory A **40** (1985), 276–289.

1991 *Mathematics subject classifications*: Primary 05A15; Secondary 05A17, 05E10, 11P81.

Keywords and phrases: tableaux, generating functions, plane partitions, nonintersecting lattice paths.

Polynômes de Schubert

Une approche historique

Alain Lascoux

(Le nous *parsemant ce texte n'a pour fonction que de té moigner que tous mes travaux concernant les polynômes de Schubert sont le fruit d'une collaboration avec M.P.Schützenberger)*

Du temps de Newton, la meilleure manière de contrôler une fonction d'une variable, que ce soit le taux de change Livre/écu ou la course d'une planète, était la technique remontant aux babyloniens de noter suffisamment de ses valeurs à intervalles de temps réguliers. Par différences successives, on simplifie usuellement la fonction considérée. Dans le cas particulier d'un polynôme par exemple, on aboutit à la fonction nulle.

Le respect de la semaine anglaise introduisant des perturbations dans la régularité journalière des observations, Newton proposa dans les Principia le remède de normaliser les différences par l'intervalle de temps auxquelles elles correspondent :

$$\begin{array}{ccccccc} f(a) & & f(b) & & f(c) & & f(d) \\ & \frac{f(a)-f(b)}{(a-b)} & & \frac{f(b)-f(c)}{(b-c)} & & \frac{f(c)-f(d)}{(c-d)} & \\ & & \frac{\frac{f(a)-f(b}{a-b}-\frac{f(b)-f(c)}{b-c}}{a-c} & & \frac{\frac{f(b)-f(c)}{b-c}-\frac{f(c)-f(d)}{c-d}}{b-d} & & \end{array}$$

En termes plus algébriques, étant donnée une fonction de plusieurs variables $g(a,b,\ldots)$, la différence divisée de g par rapport à la paire de variables b,c est la fonction

$$\partial_{bc}(g) := \big(g(a,b,c,d,\ldots) - g(a,c,b,d,\ldots)\big)/(b-c) \ ,$$

les fonctions d une variable a étant, comme l'avait déjà remarqué Newton, des fonctions de plusieurs variables de degré nul en $b,c\ldots$

La théorie des différences divisées était, jusqu'à très récemment, réfugiée dans les livres d'analyse numérique.

Si l'on me permet une incidente, c'est d'ailleurs dans [MT] que je l'y rencontrai pour la première fois et vis qu' elle était un outil commode pour l'étude des fonctions de Schur drapeaux. En effet, les images des fonctions complètes en $\{a,b,c\}$ par rapport à la différence divisée ∂_{cd} sont très exactement les fonctions complètes en $\{a,b,c,d\}$. Cela permet la manipulation des déterminants en les fonctions complètes par induction sur la taille des alphabets.

Revenons au cours historique principal. Aucun évènement notable en Europe jusqu'en 1800, où Rothe [Ro] définit le diagramme d'une permutation, qui exhibe la matrice représentant une permutation en même temps que ses inversions. Chaque inversion : $i<j, \mu_i > \mu_j$ (i.e. paires de 1 en position relative $\begin{smallmatrix}1 & \\ 0 & 1\end{smallmatrix}$, crée une grappe en colonne j et ligne i :

$$\begin{array}{cc} 1 & \\ 0 & 1 \end{array} \longrightarrow \begin{array}{cc} 1 & \\ \text{❀} & 1 \end{array}$$

Par exemple, la permutation $\mu = 764198532$ est représentée par la matrice

$$\begin{array}{ccccccccc}
0&1&0&0&0&0&0&0&0\\
0&0&1&0&0&0&0&0&0\\
0&0&0&0&1&0&0&0&0\\
0&0&0&0&0&0&0&1&0\\
0&0&0&0&0&0&0&0&1\\
1&0&0&0&0&0&0&0&0\\
0&0&0&1&0&0&0&0&0\\
0&0&0&0&0&1&0&0&0\\
0&0&0&0&0&0&1&0&0
\end{array}$$

et son diagramme de Rothe est (en gardant les 1 pour plus de clarté) :

$$\begin{array}{ccccccccc}
-&1&-&-&-&-&-&-&-\\
-&\clubsuit&1&-&-&-&-&-&-\\
-&\clubsuit&\clubsuit&-&1&-&-&-&-\\
-&\clubsuit&\clubsuit&-&\clubsuit&-&-&1&-\\
-&\clubsuit&\clubsuit&-&\clubsuit&-&-&\clubsuit&1\\
1&-&-&-&-&-&-&-&-\\
\clubsuit&\clubsuit&\clubsuit&1&-&-&-&-&-\\
\clubsuit&\clubsuit&\clubsuit&\clubsuit&\clubsuit&1&-&-&-\\
\clubsuit&\clubsuit&\clubsuit&\clubsuit&\clubsuit&\clubsuit&1&-&-
\end{array}$$

On peut aussi lire sur ce diagramme une décomposition de la permutation en un produit de transpositions simples, en numérotant consécutivement par lignes les grappes, et en commençant par le numéro de la ligne:

$$\begin{array}{ccccccccc}
-&-&-&-&-&-&-&-&-\\
-&8&-&-&-&-&-&-&-\\
-&7&8&-&-&-&-&-&-\\
-&6&7&-&8&-&-&-&-\\
-&5&6&-&7&-&-&8&-\\
-&-&-&-&-&-&-&-&-\\
3&4&5&-&-&-&-&-&-\\
2&3&4&5&6&-&-&-&-\\
1&2&3&4&5&6&-&-&-
\end{array}$$

Lisant la mode occidentale de gauche droite, puis de haut en bas, on obtient le mot 8 78 678 5678 345 23456 123456 qui est une d composition r duite de μ^{-1}.

Nouvelle interruption de l'histoire jusqu'en 1973 où Bernstein, Gelfand & Gelfand [B-G-G] et Demazure [D1] publient simultanément deux articles importants ; ils montrent entre autres que les différences divisées satisfont aux relations de Coxeter. En d'autres termes, pour chaque permutation μ, il existe une différence divisée ∂_μ, le cas de Newton correspondant aux transpositions simples. [B-G-G] et [D1] utilisent ces opérateurs pour définir une base linéaire de l'anneau de cohomologie de la variété de drapeaux, c'est-à-dire du quotient $\mathcal{H}(A)$ de l'anneau des polynômes $Z[a_1, \dots, a_n]$ par l'idéal engendré par les fonctions symétriques sans terme constant. En degré maximum $\binom{n}{2}$ $\mathcal{H}(A)$ est de dimension

1. [B-G-G] et [D1] prouvent que les images par différences divisées de tout élément non nul de degré maximal forment une base linéaire de $\mathcal{H}(A)$. La classe canonique de départ choisie par [B-G-G] et [D1] est le Vandermonde $\Delta(A) := \prod_{i<j}(a_i - a_j)$ (*cf.* plus bas pour un exemple).

M.P. Schützenberger et moi même étions à la recherche d'un outil permettant l'étude des deux ordres sur le groupe symétrique, en particulier pour calculer les polynômes de Kahzdan & Lusztig (*cf.*[L&S1]). [B-G-G] et [D1] nous apportèrent la révélation que les arêtes de l'hexagone doivent être interprétées comme les différences divisées ∂_{ab} et ∂_{bc}. Mais plutôt que de faire agir cet hexagone sur le Vandermonde $(a-b)(a-c)(b-c)$, nous avons pris le polynôme le plus simple sans symétries, qui se trouve être le monôme $a^2b^1c^0$. Plus généralement, les *Polynômes de Schubert* sont définis comme les images par les différences divisées du monôme

$$a_1^{n-1}a_2^{n-2}\cdots a_n^0 .$$

Il sont indicés par les permutations, puisque les différences divisées le sont, et contiennent comme sous-famille les fonctions de Schur. Par exemple, la fonction de Schur

$$a^3b + a^2b^2 + ab^3$$

est égale au polynôme de Schubert X_{25134} qu'un lecteur averti distinguera du polynôme correspondant

$12\ cde^2 + 12\ cd^2e + 12\ c^2de - 6\ bde^2 - 6\ bd^2e - 6\ bce^2 - 36\ bcde - 6\ bcd^2 - 6\ bc^2e - 6\ bc^2d + 6\ b^2e^2 + 12\ b^2de + 6\ b^2d^2 + 12\ b^2ce + 12\ b^2cd + 6\ b^2c^2 - 6\ b^3e - 6\ b^3d - 6\ b^3c - 6\ ade^2 - 6\ ad^2e - 6\ ace^2 - 36\ acde - 6\ acd^2 - 6\ ac^2e - 6\ ac^2d + 36\ abde + 36\ abce + 36\ abcd - 18\ ab^2e - 18\ ab^2d - 18\ ab^2c + 18\ ab^3 + 6\ a^2e^2 + 12\ a^2de + 6\ a^2d^2 + 12\ a^2ce + 12\ a^2cd + 6\ a^2c^2 - 18\ a^2be - 18\ a^2bd - 18\ a^2bc - 6\ a^3e - 6\ a^3d - 6\ a^3c + 18\ a^3b$

de [B-G-G] et [D1], les deux polynômes étant, un facteur prè

Le calcul algébrique sur les polynômes de Schubert est pris en compte par la manipulation d'objets combinatoires (permutations, codes, diagrammes de Rothe & Riguet, tableaux, Weintrauben, etc.) ou par l'utilisation d'un produit scalaire et d'opérateurs auto-adjoints.

Cette combinatoire est très semblable à, et même englobe, la combinatoire des fonctions de Schur ; les partitions doivent ainsi être considérées comme des permutations spéciales.

Pour illustrer ce propos, je me permets sans pudeur de détailler la première note sur le sujet [L&S2]. On y trouve

- une formule de Pieri (i.e. un produit sans multiplicité de polynômes de Schubert)
- un produit scalaire pour lequel les polynômes de Schubert sont auto-adjoints, la base adjointe des monômes étant les produits de fonctions élémentaires sur des drapeaux d'alphabets
- une caractérisation des permutations spéciales (dites*vexillaires*) pour lesquelles les polynômes de Schubert sont des fonctions de Schur drapeaux

• un préordre sur les permutations qui permet de décomposer un polynôme de Schubert X_μ et en donne la partie symétrique stable F_μ (qui peut être considérée comme la limite de X_μ pour l'injection itérée $S(n) \to S(1) \times S(n)$. Ce préordre fournit des arbres dont les sommets sont des permutations ; les cycles (classes ultimes) correspondent très exactement aux permutations vexillaires

• la remarque que l'on peut déformer les différences divisées de manière à obtenir une réalisation concrète, mais avec plus de paramètres, de l'algèbre de Hecke comme algèbre d'opérateurs sur l'anneau des polynômes, *cf.* [Ch]. Les cas dégénérés dont les générateurs satisfont à $D^2 = 0$ ou $D^2 = D$ sont les plus importants pour la géométrie et correspondent à l'anneau de cohomologie ou de Grothendieck de la variété de drapeaux.

Le *Comité de Lecture* de l'Académie des Sciences nous pria de sacrifier l'adjectif *vexillaire*, ainsi qu'un exemple de drapeaux associés, sur l'autel de la clarté et de la qualité internationale des *Comptes Rendus*. On trouvera les drapeaux manquants dans [Wa], le mot vexillaire ayant quant à lui été réintroduit, après la trève estivale, dans la note suivante.

La formule la plus fondamentale que satisfasse les fonctions de Schur est la *formule de Cauchy* (dont une forme est la décomposition de la *Résultante* $\prod(a_i - \beta_j)$ de deux alphabets. Les polynômes de Schubert vérifient eux aussi une formule de Cauchy [L1]. Plus généralement, on définit les polynômes de Schubert doubles en prenant comme point de départ la polarisation du Vandermonde $X_\omega := \prod_{i+j \leq n}(a_i - \beta_j)$, les différences divisées n'agissant que sur l'alphabet A des lettres romaines. Alors tout polynôme de Schubert double se décompose en une somme de produits de polynômes de Schubert simples relatifs à chacun des alphabets A et B de même cardinal n.

De nombreuses formules peuvent en fait être déduites de la simple observation que X_ω s'annule pour toute spécialisation de A en une permutation de B, sauf pour la spécialisation $a_1 \to \beta_n, a_2 \to \beta_{n-1}, \dots a_n \to \beta_1$, auquel cas on obtient $\Delta(A)$ (*cf.* [L& S12]).

L'anneau des polynômes symétriques est canoniquement muni d'une structure d'algèbre de Hopf ([Ze]) qui permet de couper l'alphabet en intervalles. La propriété correspondante des polynômes de Schubert est donnée dans [L& S4], où l'on trouve aussi la définition des *polynômes de Grothendieck* qui correspondent aux différences divisées isobares :

$$\pi_{ab}(f) := \left(af - bf^{\sigma_{ab}}\right)/(a - b)$$

au lieu des différences divisées.

L'algèbre des diffèrences divisèes apparait naturellement en géométrie des variétés de drapeaux ([Ar], [B-G-G], [D3], [Ko-K], [A-L-P]); on retrouve les polynômes de Schubert doubles comme coefficients de la matrice de changement de base : différences divisées $\to$ permutations ou son inverse ([L&S7], [L&S12]).

A la même époque, le professeur R. Stanley se penchait sur les chemins dans le permutoedre (*décompositions réduites*). On lui doit l'observation fondamentale que le nombre de décompositions réduites d'une permutation est une somme de dimensions de représentations du groupe symétrique. Plus précisément, il conjectura l'existence de certainess fonctions symétriques F'_μ indicées par les permutations. Décomposant ces fonctions symétriques dans la base des fonctions de Schur, on obtient les dimensions cherchées. Cette conjecture me fut communiquée par Björner en septembre 82; les quelques exemples joints étaient suffisants pour montrer (la preuve est directe) que la fonction F'_μ n'est autre que la partie symétrique stable du polynôme de Schubert X_μ mentionnée plus haut. Restait à relier cette propriété (démontrée par Stanley [St]) à la combinatoire des tableaux associés aux polynômes de Schubert.

En particulier, en remplaçant les relations de Coxeter par les relations

$$i(i+1)i \cong (i+1)i(i+1)$$

$$i < j < k \Rightarrow \; jik \cong jki \quad \& ikj \cong kij$$

on obtient un monoide que nous avons appelé *monoide nilplaxique* en [L& S4]; ce monoide est très semblable au *monoide plaxique* défini par les relations élémentaires de Knuth, quoique il relève l'algèbre du groupe symétrique alors que le monoide plaxique permet d'étudier l'anneau des fonctions symétriques. Dans les deux cas chaque classe d'équivalence contient un tableau et un seul qui est donné par l'*algorithme de Schensted* ou le *jeu de taquin*, à une variante près . Cette adaptation de l'algorithme de Schensted a aussi été obtenue par Edelman et Greene [E-G]. Ces auteurs nénamoins préfèrent mettre l'accent sur d'autres objets qu'ils avaient découverts, à savoir les *tableaux balancés* qui décrivent l'ordre dans lequel une décomposition réduite d'une permutation crée ses inversions. Ainsi, la décomposition $\sigma_2\sigma_1\sigma_3\sigma_2\sigma_1\sigma_3$ correspond au chemin

$$1234 \to 1324 \to 3124 \to 3142 \to 3412 \to 4312 \to 4321$$

et au tableau balancé $\begin{smallmatrix}6\\21\\435\end{smallmatrix}$. Edelman et Greene ont depuis étendu leur théorie des tableaux balancés à des diagrammes plus généraux que les diagrammes de Ferrers.

La combinatoire des décompositions réduites a été généralisée aux groupes de Weyl par Kraskiewicz [Kr2], le cas des groupes hyperoctahédraux se trouvant dans [Kr1] et [Ha].

Les différences divisées diminuent le degré, ce qui complique fortement leur relèvement à l'algèbre libre, si l'on veut garder la compatibilité avec les tableaux et la construction de Schensted. La solution est beaucoup plus simple dans le cas des différences isobares: en effet l'opérateur π_{ab} défini plus haut envoie a^4b sur $a^4b + a^3b^2 + a^2b^3 + ab^4$, c'est-à-dire interpole entre le monôme $a^kb^i, k \geq i$ et son image par la transposition σ_{ab}, qui est a^ib^k. Comme on a une action du groupe symétrique sur l'algèbre libre, on peut aisément étendre les différences divisées π_{ab} par interpolation, à l'algèbre libre. Ces nouveaux opérateurs ne vérifient toutefois pas les relations de Coxeter [L& S10].

Conservant les relations de compatibilité entre polynômes de Schubert et opérateurs π_{ab}, on obtient ainsi des sommes de tableaux qui relèvent les polynômes de Schubert, en prenant comme points de départ tous les tableaux de *Yamanouchi* $\begin{smallmatrix}\cdots\\c\cdots c\\b\cdots\cdots b\\a\cdots\cdots\cdots a\end{smallmatrix}$, dont les images commutatives sont les monômes *dominants* $a^kb^jc^i\cdots$, $k \geq j \geq i \geq \cdots$.

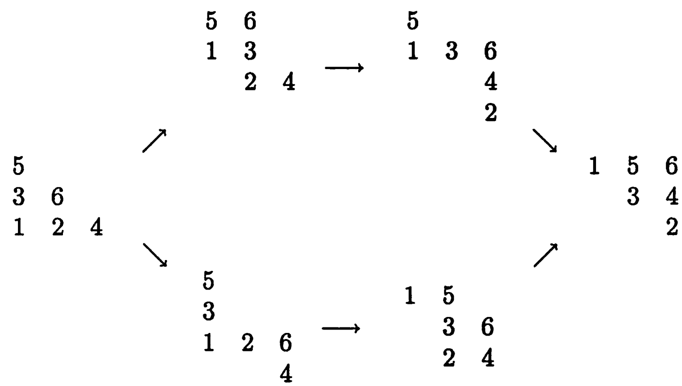
Mais pour décrire précisément les tableaux dans le développement d'un polynôme de Schubert X_μ non commutatif (en particulier, montrer que c'est une somme positive), on a besoin de faire intervenir les décompositions réduites de μ. Ainsi $\mu = 25143$ a deux décompositions réduites qui sont des tableaux:

$$\begin{array}{|c|cc}\hline 4 & & \\ \cline{1-1} 2 & & \\ \hline 1 & \multicolumn{1}{|c|}{3} & \multicolumn{1}{c|}{4} \\ \hline\end{array} \qquad \begin{array}{|c|c|c}\hline 2 & 4 & \\ \hline 1 & 3 & \multicolumn{1}{c|}{4} \\ \hline\end{array}\,.$$

Le lien entre les deux types de tableaux peut être donné par une construction qui s'applique aussi bien au monoide plaxique que nilplaxique : elle consiste à faire agir un groupe symétrique soit sur les lignes, soit sur les colonnes des tableaux. L'action des transpositions simples se définit comme la transformation d'un tableau à deux lignes (resp. deux colonnes) en un contretableau. Ainsi l'on a, grâce au jeu de taquin,

$$\begin{array}{cc}7&\\5&\\3&6\\1&2\end{array} \equiv \begin{array}{cc}7&\\5&6\\3&\\1&2\end{array} \equiv \begin{array}{cc}7&\\5&6\\&3\\1&2\end{array} \equiv \begin{array}{cc}7&\\5&6\\1&3\\&2\end{array} \equiv \begin{array}{cc}&7\\5&6\\1&3\\&2\end{array} \equiv \begin{array}{cc}5&7\\&6\\1&3\\&2\end{array} \equiv \begin{array}{cc}5&7\\1&6\\&3\\&2\end{array}$$

Répétant ces opérations, on obtient une action du groupe symétrique :



Si l'on revient au taquin à deux colonnes, on remarque que, de constructio, les deux colonnes de droite sont emboîtées, et qu'il en est de même des colonnes de gauche par symétrie. Ainsi, la colonne $\begin{smallmatrix}6\\4\end{smallmatrix}$ est contenue dans la colonne $\begin{smallmatrix}6\\4\\2\end{smallmatrix}$, et $\begin{smallmatrix}5\\1\end{smallmatrix}$ est contenue dans la colonne $\begin{smallmatrix}5\\3\\1\end{smallmatrix}$.

Les colonnes de droite figurant dans l'hexagone ci-dessus sont donc emboîtées; en d'autres termes, ce sont les colonnes d'une *clef* de même forme que le tableau qui est dite *clef droite* du tableau de départ.

L'emboîtement des colonnes gauches permet de définir la *clef gauche*, du tableau. Pour l'exemple en cours, le tableau est $\begin{smallmatrix}5\\36\\124\end{smallmatrix}$, la clef gauche est $\begin{smallmatrix}5\\35\\111\end{smallmatrix}$ et la clef droite $\begin{smallmatrix}6\\46\\244\end{smallmatrix}$.

On peut noter, et c'est aussi une conséquence triviale tout autant que remarquable du taquin, que la clef gauche d'un tableau est inférieure à sa clef droite[L& S10-11].

Les clefs correspondent bijectivement aux permutations, ce qui permet de lier la combinatoire des tableaux aux propriétés de l'ordre d'Ehresmann (dit aussi de Bruhat) sur le groupe symétrique. En particulier, on a besoin, en géométrie, des chaînes de permutations pour décrire la postulation des variétés de Schubert ([D1], [D2], [LMS]).

Quoique nous ayons obtenu les polynômes de Schubert non commutatifs comme suite directe de [L& S4], ceux-ci n'ont été soumis à publication que bien postérieurement [L& S9], car il subsiste un problème toujours non tranché : dans quelle catégorie peut-on simultanément utiliser les opérateurs ∂_μ et π_μ ?

Ceci est grave , car pour un géomètre, l'anneau de cohomologie et l'anneau de Grothendieck de la variété de drapeaux sont deux interprétations différentes de l'anneau des polynômes.

Le préordre sur les permutations mentionné plus haut fournit facilement la partie symétrique stable d'un polynôme de Schubert. En particulier, si l'on part d'une permutation qui est un produit direct de deux permutations grassmanniennes, le polynôme de Schubert correspondant est le produit de deux fonctions de Schur et le préordre donne un algorithme de décomposition de ce produit différent de la fameuse *Règle de Littlewood-Richardson* [L& S6].Plutôt qu'un produit direct de permutations grassmanniennes, A.Kohnert [Ko1] prend le cas de polynômes de Schubert images par différences divisées de fonctions de Schur, ce qui lui permet de donner la version duale de la règle de Littlewood-Richardson exprimant le développement d'une fonction de Schur gauche (c'est l'algorithme très performant utilisé dans le système SYMMETRICA dont Kohnert est le maître d'oeuvre).

Remarquant que le produit scalaire avec un polynôme de Schubert peut s'écrire comme l'action d'une différence divisée, puisque celles-ci sont auto-adjointes, on traduit le fait que les polynômes de Schubert sont une base auto-adjointe en une formule d'interpolation pour les polynômes de plusieurs variables:

$$P(B) = \sum\nolimits_\mu \pm \partial_\mu\big(P(A)\big) \cdot X_\mu(A, B) \ .$$

Dans le cas des fonctions d'une variable, cette formule coincide très exactement avec l'interpolation de Newton. Toutefois, dans ce dernier cas, seules apparaissent les différences divisées correspondant aux cycles $\mu = 2 \cdots n\, 1\, n{+}1 \cdots$; les polynômes obtenus par Newton : $(a_1 - \beta)(a_2 - \beta) \cdots (a_n - \beta)$ sont bien des polynômes de Schubert.

Ce nom illustre et leur concision par rapport aux polynômes de Demazure et Bernstein-Gelfand-Gelfand devrait, à notre sens, donner motif à la reconnaissance de leur pedigree (si ce franglais est permis au Quebec) par les savants anglo-saxons.

C'est l'interprétation des fonctions de Schur en termes de représentations qui justifie le fait que ces fonctions portent le nom de Schur et non de leurs "découvreurs" Cauchy et Jacobi. Pour les polynômes de Schubert, le lien avec la théorie des représentations est établi par Kraskiewicz et Pragacz [Kr-Pr] et sera certainement l'objet de développements importants.

Une autre approche [L& S8] consiste à partir de l'expression des polynômes de Schubert dans la base des produits de fonctions élémentaires (pour des drapeaux d'alphabets); on obtient des modules tout autant que des sommes de tableaux, en remplaçant les fonctions élémentaires par des puissances extèrieures de modules libres, ou par des sommes de tableaux-colonnes. Par exemple, le polynôme de Schubert X_{31452} est égal à $(abc + abd + acd + bcd) \cdot (a) - (abcd)$; on obtient le module correspondant comme conoyau du morphisme

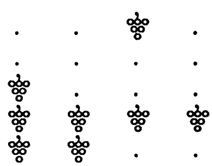
$$\Lambda^3(V_4) \otimes V_1 \longleftarrow \Lambda^4(V_4) \longleftarrow 0$$

et les tableaux à l'aide de l'identité dans l'algèbre plaxique :

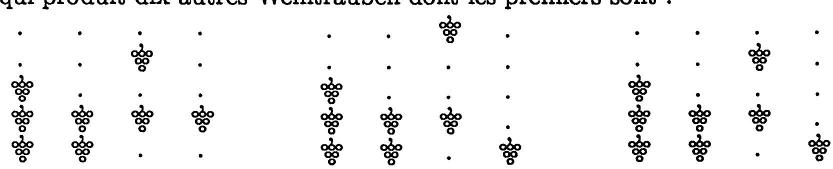
$$\begin{smallmatrix}3\\2\\11\end{smallmatrix} + \begin{smallmatrix}4\\2\\11\end{smallmatrix} + \begin{smallmatrix}4\\3\\11\end{smallmatrix} \equiv \left(\begin{smallmatrix}3\\2\\1\end{smallmatrix} + + \begin{smallmatrix}4\\2\\1\end{smallmatrix} + \begin{smallmatrix}4\\3\\1\end{smallmatrix} + \begin{smallmatrix}4\\3\\2\end{smallmatrix} \right) \cdot 1 - \begin{smallmatrix}4\\3\\2\\1\end{smallmatrix}$$

En fait, l'objet principal de l'article [L& S8] est de donner une formule de dualité entre monoïdes plaxique et nilplaxique et de montrer la possibilité d'un "calcul de Schubert" en les variables non commutatives que sont les différences divisées élémentaires.

Cent quatre vingt huit ans après Rothe, la combinatoire du groupe symétrique s'enrichit d'un nouvel apport lotharingien avec les "Weintrauben" d'A. Kohnert [Ko2]. Au lieu d'inscrire des informations dans les boîtes ou grappes du diagramme de Rothe & Riguet, on déforme celui-ci avec la règle on ne peut plus élémentaire suivante : toute grappe la plus à droite de son niveau doit un jour ou l'autre choir d'un étage, poussant éventuellement les grappes dans la même colonne qui gênent son mouvement. De chaque Weintraub ainsi obtenu, on peut lire plusieurs informations, en particulier un mot : lire par colonnes, de droite à gauche, la suite des niveaux où se trouve une grappe. Ce mot est transformé ensuite en tableau à l'aide de l'algorithme de Schensted ou par le jeu de taquin. Ainsi la permutation $\mu = 362154$ a pour diagramme de Rothe & Riguet



qui produit dix autres Weintrauben dont les premiers sont :



et la lecture de ces quatre Weintrauben donne les mots et tableaux suivants:

$$25221321 \equiv \begin{array}{llll} 5 & & & \\ 3 & & & \\ 2 & 2 & & \\ 1 & 1 & 2 & 2 \end{array} \; ; \; 24221321 \equiv \begin{array}{llll} 4 & & & \\ 3 & & & \\ 2 & 2 & & \\ 1 & 1 & 2 & 2 \end{array} \; ; \; 15221321 \equiv \begin{array}{llll} 5 & & & \\ 3 & & & \\ 2 & 2 & & \\ 1 & 1 & 1 & 2 \end{array}$$

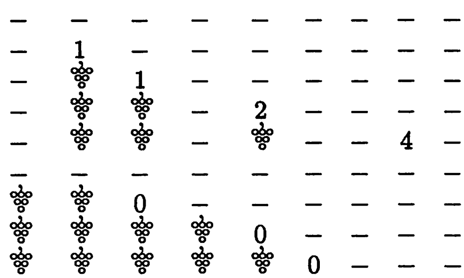
; 14221321 ≡

4			
3			
2	2		
1	1	1	2

.

Kohnert prouve dans sa thèse que dans le cas d'une permutation vexillaire, on obtient bien le polynôme de Schubert non commutatif obtenu en [L& S9]. Les ingrédients pour conclure le cas général sont maintenant réunis, mais il est plus intéressant d'expliciter les liens étroits entre les Weintrauben et les nouvelles constructions sur les polynômes de Schubert obtenues récemment, plutôt que de donner une preuve indépendante.

Un point de vue entièrement différent est dû à Fulton qui retourne aux sources du "calcul de Schubert". Etant donnée une matrice, disons dont les éléments sont des polynômes, Schubert, puis de manière beaucoup plus approfondie, Giambelli [Gi], ont examiné la variété définie par l'annulation des mineurs d'un certain ordre de cette matrice. Si l'on considère la matrice comme représentant un morphisme ϕ de fibrés vectoriels, c'est la stratification des points de la variété de base en fonction du corang de ϕ que l'on cherche à décrire. Si l'on filtre le morphisme par un drapeau source et un drapeau but, on a plus généralement à décrire le rang de familles de mineurs de la matrice de départ. Le problème est tout d'abord de décrire ce que l'on entend par "conditions de rang". La réponse est apportée par Fulton [Fu] : évaluer le rang des sous-matrices prises sur des lignes et des colonnes consécutives à partir de l'origine. Dans le cas, en fait générique, d'une matrice de permutation, les coins des composantes connexes (que Fulton appelle *points essentiels*, contiennent l'information voulue. En d'autres termes, pour la permutation $\mu = 764198532$ qui nous sert d'exemple générique, il suffit de connaître la valeur du rang en les grappes qui ont été remplacées par la valeur de ce rang :



Une fois éclaircie la notion de rang, la géométrie peut développer ses puissantes méthodes et aboutir à l'interprétation des polynômes de Schubert doubles comme classes dans l'anneau de cohomologie des sous-variétés correspondant à des conditions de rang appropriés pour un morphisme entre drapeaux de fibrés vectoriels.

D'autres applications des polynômes de Schubert sont données dans [A-L-P] et [L-P].

Le statut social de Schubert s'améliora considérablement, du jour où commencèrent à circuler les notes de I.G.Macdonald [M] concernant les polynômes du même nom. Rédigées avec la même élégance, clarté et précision que le livre "Symmetric Functions and Hall polynomials" dont elles constituent une suite naturelle, elles suscitèrent immédiatement l'intérêt sur de nombreux continents – on a signalé des photocopies pirates à l'université de Gifu, préfecture de Nagoya.

Un des apports esentiels de ces notes concernent les chemins du permutoedre. Macdonald découvre que la fonction génératrice du nombre de ces chemins, les arêtes étant convenablement pondérées, est égale à la spécialisation du polynôme de Schubert en l'alphabet $\{1,\ldots,1\}$. Changeant ensuite le poids des arêtes, il énonce ensuite la conjecture profonde que l'on trouve ainsi la spécialisation des polynômes de Schubert en $\{1,q,q^2,\ldots\}$. Elle vient d'être prouvée dans des articles non encore publiés que je ne peux qu' évoquer rapidement (les chemins pour l'ordre d'Ehressmann sont aussi liés aux polynômes de Schubert, *cf.* [L& S3]).

Durant l'été 91, R.Stanley proposa une énumération élégante des monômes apparaissant dans le développement d'un polynôme de Schubert. A chaque décomposition réduite $s = s_1s_2\cdots s_r$, Stanley associe l'ensemble $W(s)$ des mots décroissants w dominés (composante à composante) par s et ayant des décroissances contraintes aux décroissances de s, c'est-à-dire des mots tels que

$$s_i \geq w_i \qquad ; \qquad s_i > s_{i+1} \Rightarrow w_i > w_{i+1}$$

Par exemple, $s = \sigma_5\ \sigma_3\ \sigma_2\sigma_3\sigma_4\sigma_5\ \sigma_1\sigma_2$ majore les mots dé croissants 5 3 2222 11 et 4 3 2222 11 et $s = \sigma_5\ \sigma_3\ \sigma_2\sigma_3\sigma_4\ \sigma_1\sigma_2\sigma_5$ majore les mots dé croissants 5 3 222 111 et 4 3 222 111.

En utilisant le tableau d'insertion du mot s, on obtient des tableaux plutôt que de simples mots. La conjecture de Stanley, prouvée dans [B-J-S], est que la somme $\sum_{s,w\in W(s)} w$, où s parcourt l'ensemble des décompositions réduites de μ^{-1}, est égale au polynôme de Schubert X_μ.

On d couvre alors que les quatre mots que nous venons de donner produisent très exactement les quatre tableaux obtenus à l'aide des Weintrauben quelques lignes plus haut.

Plutôt que d'invoquer le hasard, on peut tout d'abord remarquer que la majoration par les décompositions s est redondante. Comme l'ont montré [R-S], il suffit de remplacer les tableaux qui sont des décompositions réduites par leurs clefs gauches, puis de considérer les classes plaxiques de ces dernières. Ceci revient à remplacer les deux décompositions réduites $s = \sigma_5\ \sigma_3\ \sigma_2\sigma_3\sigma_4\sigma_5\ \sigma_1\sigma_2$ et $s = \sigma_5\ \sigma_3\ \sigma_2\sigma_3\sigma_4\ \sigma_1\sigma_2\sigma_5$ par les mots 5 3 2222 11 et 5 3 222 111 qui majorent les mêmes mots décroissants.

Donnons un autre exemple, complet : pour $\mu = 412365$, les décompositions réduites sont $\sigma_1\sigma_2\sigma_3\sigma_5$, qui est le tableau ligne $\begin{array}{|c|c|c|c|}\hline 1&2&3&5\\\hline\end{array}$ seul dans sa classe, et $\sigma_5\sigma_1\sigma_2\sigma_3$, $\sigma_1\sigma_5\sigma_2\sigma_3$, $\sigma_1\sigma_2\sigma_5\sigma_3$ dans la classe du tableau $\begin{array}{|c|cc}\hline 5 \\\hline 1&\multicolumn{1}{|c|}{2}&\multicolumn{1}{c|}{3}\\\hline\end{array}$.

Les clefs gauches sont $\begin{array}{|c|c|c|c|}\hline 1&1&1&1\\\hline\end{array}$ et $\begin{array}{|c|cc}\hline 5 \\\hline 1&\multicolumn{1}{|c|}{1}&\multicolumn{1}{c|}{1}\\\hline\end{array}$, dont les classes plaxiques sont respectivement $\{1111\}$ et $\{5111, 1511, 1151\}$. Ceci donne instantanément $X_{412365} = a_1a_1a_1a_1 + a_5a_1a_1a_1 + a_4a_1a_1a_1 + a_3a_1a_1a_1 + a_2a_1a_1a_1$, et illustre le lien entre la construction de Stanley et [L& S9-10].

Reiner et Shimozono [R-S] étudient en outre les propriétés des polynômes-clefs ("Key polynomials"), pour donner en particulier le développement d'une fonction de Schur drapeau gauche. L'apport essentiel de leur article me semble toutefois l'interprétation directe des polynômes clefs comme caractères, sans recourir comme le faisait Demazure [D1] aux sections des fibrés inversibles au dessus des variétés de Schubert, mais par généralisation de la construction par Young de ses fameux idempotents.

La preuve par Fomin et Stanley [Fo-S] de la conjecture de Macdonald repose sur le choix de $X^{FS} = (1+a\partial_2)(1+a\partial_1)(1+b\partial_2)$ comme élément de départ, les différences divisées étant relatives à un deuxième alphabet. Malgré la non-commutativité des ∂_i, il est facile de faire agir sur X^{FS} les différences divisées relatives à l'alphabet $\{a,b,c\}$. L'ingrédient essentiel du calcul est la relation de Yang-Baxter

$$(1+x\partial_2)(1+(x+y)\partial_1)(1+y\partial_2) = (1+y\partial_1)(1+(x+y)\partial_2)(1+x\partial_1) ,$$

pour tous paramètres x, y indépendants de l'alphabet sur lequel agissent les ∂_i (pour des relations plus générales, voir [B-K]).

Nous voici donc arivés aux frontières de la physique moderne. La France vient de décider la suspension de ses essais nucléaires et je dois donc obtempérer, me contentant d'évoquer l'importance des algèbes de Hecke en théorie des cordes, des noeuds et de leurs énergies inépuisables et non polluantes.

Bibliographie

[A-L-P] E.AKYILDIZ, A.LASCOUX , P.PRAGACZ , Cohomology of Schubert subvarieties of GL_n/P (1989) ,to appear in *J. of Topology* .
[Ar] A. ARABIA, Cycles de Schubert et Cohomologie K-équivariante de K/T, *Inv. Math.* **85** , 39-52(1986).
[B-K] A.BERENSTEIN, A.KIRILLOV, Groups generated by involutions, Gelfand-Tsetlin patterns and combinatorics of Young tableaux, *preprint RIMS-866,* Kyoto (1992).
[B-G-G] I.N.BERNSTEIN, I.M.GELFAND, S.I.GELFAND, Schubert cells and cohomology of the spaces G/P, *Uspekhi.Mat.Nauk*, **28** (1973) 3-26.
[B-J-S] S.BILLEY, W.JOCKUSCH, R.STANLEY, Some combinatorial properties of Schubert polynomials, *prépublication de l'Institut Mittag-Leffler* (1992).
[Ch] I.CHEREDNIK, A unification of Knizhnik-Zamolodchikov and Dunkle operators via affine Hecke algebras, *Inv. in M.* , (1991).
[D 1] M.DEMAZURE, Désingularisation des variétés de Schubert, *Annales E.N.S.* **6** 163-172(1974) .
[D 2] M.DEMAZURE, Une formule des caractères, *Bull.Sc.Math.* **98** 163-172 (1974).
[D 3] M.DEMAZURE Invariants symétriques entiers des groupes de Weyl et torsion, *Inventiones* **21** 287-301(1973).
[E-G] P.EDELMAN, C.GREENE, Balanced Tableaux, Adv.in Math.**63** 42-99 (1987).
[E] C.EHRESMANN, Sur la topologie de certains espaces homogènes, Annals of Math.(2)**35**, 396-443 (1934).
[Fu] W.FULTON, Flags, Schubert polynomials, degeneraci loci and determinantal formulas, *Duke M.J.,* **65** , 381-420 (1991).

[Fo-S] S.FOMIN, R.STANLEY, Schubert polynomials and the nilCoxeter algebra, *prépublication de l'Institut Mittag-Leffler* (1992).
[Gi] G.Z.GIAMBELLI, Rizoluzione del problema degli spazi secanti, Memorie Acc.Torino, **52**, 171-211 (1902).
[Ha] M.D.HAIMAN, Dual equivalence with applications including a conjecture of Proctor, *preprint M.I.T.* (1989).
[Ke] A.KERBER, Algebraic combinatorics via finite group actions, Wissenschaftsverlag, Mannheim, Wien, Zürich (1991).
[Ko1] A.KOHNERT, Schubert Polynomials and skew Schur Functions, à paraître.
[Ko2] A.KOHNERT, Weintrauben, Polynome, Tableaux Bayreuther Mathematische Schriften Vol. 38, p. 1-97 (1991).
[Ko-K] B.KOSTANT, S.KUMAR , The Nil Hecke ring and Cohomology of G/P for a Kac-Moody group G, *Advances in Math.* **62** 187-237(1986) .
[Kr1] , W.KRASKIEWICZ, Reduced decompositions in hyperoctahedral groups, *C.R. Acad. Sc Paris,* **309** , 903-907 (1989).
[Kr2] , W.KRASKIEWICZ, Reduced decompositions in Weyl groups, *preprint , Torun* , (1991).
[Kr-Pr] W.KRASKIEWICZ, P.PRAGACZ : Foncteurs de Schubert, *C.R. Ac.Sc.Paris* **304** , 209-211 (1987).
[L-M-S] V.LAKSHMIBAI, C.MUSILI, C.S.SESHADRI, Geometry of G/P IV, Proc.Indian Acad.Sc. **88A** , 280-362 (1979).
[L1] A.LASCOUX, Classes de Chern des variétés de drapeaux, *Comptes Rendus* **295** (1982) 393 **[L2]** A.LASCOUX, Anneau de Grothendieck de la variété de drapeaux, in The Grothendieck Festtchrift, vol III, Birkhäuser(1990) 1-34.
[L-P] A.LASCOUX, P.PRAGACZ, Divided differences and ideals generated by symmetric polynomials, *prePublications of Bergen University, n° 59-25-04-91* (1991), to appear in Discrete Math.
[L&S1] A.LASCOUX &M.P.SCHÜTZENBERGER, Polynômes de Kazdan-Lusztig pour les grassmanniennes, *in Tableaux de Young et Foncteurs de Schur, Torun1980, Astérique* **87-88**(1981).
[L&S2] , Polynômes de Schubert, *Comptes Rendus* **294** 51982) 447.
[L&S3] , Symmetry and Flag manifolds, *in Invariant Theory, Springer L.N.* **996** (1983) 118-144.
[L&S4] , Structure de Hopf..., *Comptes Rendus* **295** (1982) 629.
[L&S5] , Interpolation de Newton à plusieurs variables, *Séminaire d'Algèbre M.P.Malliavin 1983-84, Springer L.N.* **1146** (1985) 161-175.
[L&S6] , Schubert polynomials and the Littlewood-Richardson rule,*Letters in Math. Phys.***10**, 111-124(1985).
[L&S7] , Symmetrization operators on polynomial rings, Funct.Anal.**21** , 77-78(1987).
[L&S8] , Fonctorialité des polynômes de Schubert, *Contempory Maths* **88** , 585-598(1989).
[L&S9] , Tableaux and non commutative Schubert Polynomials, Funk.Anal **23** (1989) 63-64.
[L&S10] , Keys and Standard Bases, *in* "Invariant Theory and Tableaux", IMA vol. in Maths and Appl. **19** , Springer,125-144 (1990) .

[L&S11] , Arêtes et Tableaux, 20^{eme} Séminaire Lotharingien de Combinatoire, Alghero (sept.1988) 109-120.
[L&S12] , Algèbre des différences divisées, *à paraître dans Discrete Math.* (1991).
[M] I.G.MACDONALD, Notes on Schubert polynomials, *Publications du LACIM, Montr al,* (1991).
[MT] , L.M.MILNE-THOMSON, The calculus of finite differences, London (1933).
[Re-S] , V.REINER, M.SHIMOZONO, Key polynomials and a flagged Littlewood-Richardson rule, *preprint* (1992).
[Ro] H.A.ROTHE, Ueber permutationen, in Beziehung auf die Stellen ihrer Elemente, *Sammlung combinatorisch -analytischer Abhandlungen, Leipzig,* (1800).
[St] R.STANLEY, On the number of reduced decompositions of elements of Coxeter groups, Journal Europ. de Comb.**5**, 359-372 (1984).
[Wa] M.WACHS, Flagged Schur functions, Schubert polynomials, J.of Comb. theory **A** **40**, 276-289 (1985).
[Ze] A.ZELEVINSKY, Representations of finite classical groups, *Springer L.N.* **869** (1981).

CONSEQUENCES OF THE A_ℓ AND C_ℓ BAILEY TRANSFORM AND BAILEY LEMMA

STEPHEN C. MILNE AND GLENN M. LILLY

1. INTRODUCTION

The purpose of this talk is to discuss some applications of the higher-dimensional generalization of the Bailey Transform and Bailey Lemma in the setting of basic hypergeometric series very well-poised on unitary A_ℓ or symplectic C_ℓ groups in [Lil91, LM91, Mil91a-f, ML91]. The derivation of the C_ℓ case in [LM91, ML91] is closely related to the previous analysis of the unitary A_ℓ, or equivalently $U(\ell+1)$ case from [Mil91a,c,d]. This program is based upon the A_ℓ and C_ℓ terminating very well-poised ${}_6\phi_5$ summation theorems which are extracted from [Mil85, Mil87, Mil91a] and [Gus89], respectively. Both types of very well-poised series are directly related [Gus89, Mil85] to the corresponding Macdonald identities. The classical case of all this work, corresponding to A_1 or equivalently $U(2)$, contains an immense amount of the theory and application of one-variable basic hypergeometric series [And76, And86a, Bai35, GR90, Sla66], including elegant proofs of the Rogers-Ramanujan-Schur identities. The ordinary ($q = 1$) case of some of the multiple series in [Mil87] first appeared in certain applications of mathematical physics and the unitary groups $U(n+1)$, or equivalently A_n. This earlier work on the theory of Wigner coefficients for $SU(n)$ was due to Biedenharn, Holman, and Louck [BL68–BL81b, Hol80, HBL76]. They showed in [Hol80, HBL76] how the classical work on ordinary hypergeometric series is intimately related to the irreducible representations of the compact group $SU(2)$. Their work was done in the context of the quantum theory of angular momentum [BL81a-b] and the special unitary groups $SU(n)$.

The classical A_1 Bailey Transform [And86a] and Bailey Lemma [And86a] were ultimately inspired by Rogers' [Rog17] second proof of the Rogers-Ramanujan-Schur identities [And76, And86a, GR90, Rog94, Rog95]. The Bailey Transform was first formulated by Bailey [Bai47, Bai49], utilized by Dyson in [Dys43], applied by Slater in [Sla51–Sla66], and then recast by Andrews [And79] as a fundamental matrix inversion result. This last version of the Bailey Transform has immediate applications to connection coefficient theory and "dual" pairs of identities [And79, And84, And86a, GS83, GS86], and q-Lagrange inversion and quadratic transformations [GS83, GS86]. The most important application of the Bailey Transform is the Bailey Lemma. This result was mentioned by Bailey [Bai49;

1991 *Mathematics Subject Classification*. Primary: 33D70, 05A19; Secondary: 33D20.

S. C. Milne was partially supported by NSF grants DMS 86-04232, DMS 89-04455, DMS 90-96254(transfer), NSA supplements to these NSF grants, and by NSA grants MDA 904-88-H-2010 and MDA 904-91-H-0055.

G. M. Lilly was fully supported by NSA supplements to the above NSF grants and by NSA grant MDA 904-88-H-2010.

§4], and he described how the proof would work. However, he never wrote the result down explicitly and thus missed the full power of *iterating* it. Andrews first established the Bailey Lemma explicitly in [And84] and realized its numerous possible applications in terms of the iterative "Bailey chain" concept. This iteration mechanism enabled him to derive many q-series identities by "reducing" them to more elementary ones. For example, two iterations of the Bailey Lemma reduce the Rogers-Ramanujan-Schur identities to the q-binomial theorem [And84, And86a]. The process of iterating Bailey's Lemma has led to a wide range of applications in additive number theory, combinatorics, special functions, and mathematical physics. For example, see [And84–And86b, ABF84, ADH88, Bax82, Pau82, Pau85, Sla51–Sla66]. The Bailey Transform is a consequence of the terminating very well-poised ${}_4\phi_3$ summation theorem. The Bailey Lemma is derived in [AAB87] directly from Rogers' [Rog95] terminating very well-poised ${}_6\phi_5$ summation theorem and the matrix inversion formulation [And79, GS83, GS86] of the Bailey Transform. The terminating very well-poised ${}_6\phi_5$ summation theorem is crucial to this entire program.

At this point, it is useful to survey the classical Bailey Transform and Bailey Lemma.

Let q be a complex number such that $|q| < 1$. Define

$$(\alpha)_\infty \equiv (\alpha;\, q)_\infty := \prod_{k\geq 0}(1-\alpha q^k) \tag{1.1a}$$

and, thus,

$$(\alpha)_n \equiv (\alpha;\, q)_n := (\alpha)_\infty \,/\, (\alpha q^n)_\infty\,. \tag{1.1b}$$

We then have Andrews' [And79] matrix inversion in

Theorem 1.2 (Classical Bailey Transform for A_1). *Let a be indeterminate and $i, j \geq 0$ be integers. Let the matrices M and M^* be defined as in*

$$M(i;\, j;\, A_1) := (q)_{i-j}^{-1}\,(aq)_{i+j}^{-1}\,; \tag{1.3a}$$

and

$$M^*(i;\, j;\, A_1) := (1-aq^{2i})\,(aq)_{i+j-1}\,(q)_{i-j}^{-1}\,(-1)^{i-j}q^{\binom{i-j}{2}}. \tag{1.3b}$$

Then M and M^* are inverse, infinite, lower-triangular matrices. That is,

$$\delta(i,\, j) = \sum_{j\leq y\leq i} M(i;\, y;\, A_1)\, M^*(y;\, j;\, A_1), \tag{1.4}$$

where $\delta(r,\, s) = 1$ if $r = s$, and 0 otherwise.

Theorem 1.2 follows from the terminating very well-poised ${}_4\phi_3$ summation theorem and a termwise rewriting of the (i, j) entry in the matrix product MM^*. Earlier, Carlitz [Car73; Theorem 5], and then later Al-Salam and Verma [AV84] had obtained bibasic matrix inversion results whose $p = q$ case is equivalent to Theorem 1.2. More recently, Gessel and Stanton [GS83; Theorem 1.2] proved several q-series identities using Theorem 1.2. Gasper

[Gas89] recently derived bibasic extensions and analogs of Theorem 1.2, and the earlier work of Carlitz, Al-Salam, and Verma. Bressoud [Bre83] has deduced an elegant extension of Theorem 1.2 for matrices $M_{a,b}$, with two free parameters, from the terminating very well-poised ${}_6\phi_5$ summation theorem. He proved that $M_{a,b}$ and $M_{b,a}$ are inverse, infinite, lower-triangular matrices. All of this work, as well as [AAB87, And79], provides a natural setting for Theorem 1.2.

Equation (1.3) motivates the definition of the A_1 Bailey Pair.

Definition 1.5 (A_1 Bailey Pair). Let $n \geq 0$ and $y \geq 0$ be integers and $\alpha = \{\alpha_y\}$ and $\beta = \{\beta_y\}$ be sequences. Let M and M^* be as in (1.3). Then we say that α and β form an A_1 Bailey Pair if

$$\beta_n = \sum_{0 \leq y \leq n} M(n;\, y;\, A_1)\, \alpha_y, \tag{1.6}$$

for all $n \geq 0$.

The study of A_1 Bailey Pairs $\{\alpha_n, \beta_n\}$ satisfying (1.6) goes back to L. J. Rogers' [Rog94, Rog17] proofs of the Rogers-Ramanujan-Schur identities, and more recently to L. J. Slater [Sla51–Sla66], D. M. Bressoud [Bre81], and G. E. Andrews [And84].

Equation (1.4) and Definition 1.5 immediately give

Corollary 1.7 (A_1 Bailey Pair Inversion). *α and β satisfy equation* (1.6) *if and only if*

$$\alpha_n = \sum_{0 \leq y \leq n} M^*(n;\, y;\, A_1)\, \beta_y. \tag{1.8}$$

Corollary 1.7 is responsible for the dual pairs of identities in [And79, And86a, GS86]. For example, with suitable α_n and β_n, it follows that (1.6) and (1.8) correspond to Rogers' [Rog95] terminating very well-poised ${}_6\phi_5$ summation [Bai35, GR90], and Jackson's [Jack10] terminating balanced ${}_3\phi_2$ summation [Bai35, GR90], respectively.

Andrews' explicit formulation of the Bailey Lemma is provided by

Theorem 1.9 (Classical Bailey Lemma for A_1). *Let the sequences $\alpha = \{\alpha_n\}$ and $\beta = \{\beta_n\}$ form an A_1 Bailey Pair. If $\alpha' = \{\alpha'_n\}$ and $\beta' = \{\beta'_n\}$ are defined by,*

$$\alpha'_n := \frac{(\rho)_n\,(\sigma)_n}{(aq/\rho)_n\,(aq/\sigma)_n}\,(aq/\rho\sigma)^n\,\alpha_n \tag{1.10a}$$

and

$$\beta'_n := \sum_{0 \leq y \leq n} \frac{(\rho)_y\,(\sigma)_y\,(aq/\rho\sigma)_{n-y}}{(q)_{n-y}\,(aq/\rho)_n\,(aq/\sigma)_n}\,(aq/\rho\sigma)^y\,\beta_y, \tag{1.10b}$$

then α' and β' also form an A_1 Bailey Pair.

Andrews notes in [And84] that Watson's [Wat29] q-analog of Whipple's transformation is an immediate consequence of the second iteration of Theorem 1.9, starting from one of the simplest A_1 Bailey Pairs. In fact, Andrews' infinite family of extensions of Watson's q-Whipple's transformation in [And75] is just a consequence of continued iteration of this same case of Theorem 1.9. Even Whipple's original work in [Whi24, Whi26] fits into the $q = 1$ case of this analysis. Paule [Pau82, Pau85] independently discovered important special cases of Theorem 1.9 and observed how these results could be iterated. Essentially all the depth of the classical Rogers-Ramanujan-Schur identitites and their iterations is embedded in the A_1 Bailey Lemma.

We organize the rest of this talk as follows. Let G denote A_ℓ or C_ℓ. In §2 we state the G terminating very well-poised ${}_6\phi_5$ summations from [LM91, Mil87, Mil91a] which we need in our subsequent work. We indicate in §3 how the G Bailey Transform of [LM91, Mil91a] is obtained from a suitably modified G terminating very well-poised ${}_4\phi_3$ summation theorem and termwise transformations. It is then interpreted as a matrix inversion result for two infinite, lower-triangular matrices. This provides a higher-dimensional generalization of Theorem 1.2. As in Definition 1.5 and Corollary 1.7, the concept of a G Bailey Pair is introduced, and then inverted. This G inversion applied to the G terminating very well-poised ${}_6\phi_5$ summations from §2 yields the G terminating balanced ${}_3\phi_2$ summations in §4. This is just a sample of the new A_ℓ terminating balanced ${}_3\phi_2$ summations from [Mil91a]. We describe in §5 how the G Bailey Lemma from [LM91, Mil91c] is obtained directly from a G terminating very well-poised ${}_6\phi_5$ summation theorem and the matrix inversion formulation of the G Bailey Transform. It shows how to construct another G Bailey Pair from an arbitrary G Bailey Pair, and thus extends Theorem 1.9. The concepts of an ordinary G Bailey Chain and a bilateral G Bailey Chain are introduced. Finally, appealing to the second iterate of the G Bailey Lemma, if time permitts, we will state , as an example, one A_ℓ and one C_ℓ q-Whipple transformation. These examples will appear in §6 of our longer paper based on this talk. Several A_ℓ q-Whipple transformations, including this one, are derived in [Mil91b-c]. Many other consequences of the G Bailey Transform and Lemma appear in [Lil91, LM91, Mil91a-f, ML91].

2. Background Information

The main results in this talk depend upon an A_ℓ and a C_ℓ terminating very well-poised ${}_6\phi_5$ summation theorem from [Mil85, Mil87, Mil91a] and [Gus89, LM91], respectively. Here, we state these two ${}_6\phi_5$ summations in a form convenient for our applications. The $\ell = 1$ case of each is the classical terminating ${}_6\phi_5$ summation in equation (II.21) of [GR90; pp. 238].

We start with

Theorem 2.1 (An A_ℓ terminating ${}_6\phi_5$ summation theorem). *Let a, b, c and $x_1, \ldots, x_\ell$ be indeterminate, let N_i be non-negative integers for $i = 1, 2, \ldots, \ell$ with $\ell \geq 1$, and suppose that none of the denominators in* (2.2) *vanishes. Then*

$$\left\{ \frac{(aq/bc)_{N_1+\cdots+N_\ell}}{(aq/b)_{N_1+\cdots+N_\ell}} \prod_{k=1}^{\ell} \frac{\left(\frac{x_k}{x_\ell} aq\right)_{N_k}}{\left(\frac{x_k}{x_\ell} aq/c\right)_{N_k}} \right\} \tag{2.2a}$$

$$= \sum_{\substack{0\le y_k\le N_k\\ k=1,2,\dots,\ell}} \left\{ \prod_{1\le r<s\le \ell} \left[\frac{1-\frac{x_r}{x_s}q^{y_r-y_s}}{1-\frac{x_r}{x_s}} \right] \prod_{k=1}^{\ell} \left[\frac{1-\frac{x_k}{x_\ell}aq^{y_k+(y_1+\cdots+y_\ell)}}{1-\frac{x_k}{x_\ell}a} \right] \right.$$

$$\times \prod_{r,s=1}^{\ell} \left[\frac{\left(\frac{x_r}{x_s}q^{-N_s}\right)_{y_r}}{\left(q\frac{x_r}{x_s}\right)_{y_r}} \right] \prod_{k=1}^{\ell} \left[\frac{\left(\frac{x_k}{x_\ell}a\right)_{y_1+\cdots+y_\ell}}{\left(\frac{x_k}{x_\ell}aq^{1+N_k}\right)_{y_1+\cdots+y_\ell}} \right]$$

$$\times \frac{(c)_{y_1+\cdots+y_\ell}}{(aq/b)_{y_1+\cdots+y_\ell}} \prod_{k=1}^{\ell} \left[\frac{\left(\frac{x_k}{x_\ell}b\right)_{y_k}}{\left(\frac{x_k}{x_\ell}aq/c\right)_{y_k}} \right]$$

$$\left. \times \left[\left(\frac{aq^{1+(N_1+\cdots+N_\ell)}}{bc} \right)^{y_1+\cdots+y_\ell} q^{y_2+2y_3+\cdots+(\ell-1)y_\ell} \right] \right\}. \tag{2.2b}$$

Proof. First, rewrite Theorem 1.38 of [Mil87] by replacing n by $\ell+1$, making the substitutions

$$a_{\ell+1,\ell+1} = b/a, \quad z_\ell/z_{\ell+1} = a, \tag{2.3a}$$

and then taking $m = N$, and

$$a_{ii} = c_i \quad \text{and} \quad z_i = x_i, \qquad \text{for } i = 1,2,\dots,\ell. \tag{2.3b}$$

By the

$$c_i = q^{-N_i}, \qquad \text{for } i = 1,2,\dots,\ell, \tag{2.4}$$

case of this result and an elementary calculation involving its product side, it follows that the identity (2.2) holds for $c = q^{-N}$, with N any non-negative integer. However, (2.2) is a polynomial identity in c^{-1}, whose degree is a finite function of $\{N_1,\dots,N_\ell\}$. Hence, Theorem 2.1 is true in general. □

Remark. This is the proof of Theorems 2.1 and 2.4, respectively, in [Mil91a], with n replaced by ℓ. The paper [Mil91a] contains three additional A_ℓ terminating very well-poised ${}_6\phi_5$ summation theorems.

Remark. The $\ell = 1$ and $N_1 = n$ case of (2.2) is equation (II.21) of [GR90; pp. 238].

Next, Gustafson's C_ℓ ${}_6\psi_6$ summation theorem from [Gus89] leads in [LM91] to

Theorem 2.5. (The C_ℓ terminating ${}_6\phi_5$ summation theorem). *Let a, b and $x_1,\dots,x_\ell$ be indeterminate, let N_i be non-negative integers for $i = 1,2,\dots,\ell$ with $\ell \ge 1$, and suppose that none of the denominators in* (2.6) *vanishes. Then*

$$\sum_{\substack{0\le y_k\le N_k\\ k=1,2,\dots,\ell}} \left\{ \prod_{k=1}^{\ell} \left[\frac{1-x_k^2q^{2y_k}}{1-x_k^2} \right] \prod_{1\le r<s\le\ell} \left[\frac{1-\frac{x_r}{x_s}q^{y_r-y_s}}{1-\frac{x_r}{x_s}} \, \frac{1-x_rx_sq^{y_r+y_s}}{1-x_rx_s} \right] \right.$$

$$\times \prod_{r,s=1}^{\ell}\left[\frac{\left(\frac{x_r}{x_s}q^{-N_s}\right)_{y_r}(x_rx_s)_{y_r}}{\left(q\frac{x_r}{x_s}\right)_{y_r}(qx_rx_sq^{N_s})_{y_r}}\right]\prod_{k=1}^{\ell}\left[\frac{(ax_k)_{y_k}(qx_kb^{-1})_{y_k}}{(bx_k)_{y_k}(qx_ka^{-1})_{y_k}}\right]$$

$$\times q^{(N_1+\cdots+N_\ell)(y_1+\cdots+y_\ell)}q^{y_2+2y_3+\cdots+(\ell-1)y_\ell}\left(\frac{b}{a}\right)^{y_1+\cdots+y_\ell}\Bigg\} \tag{2.6a}$$

$$=\prod_{k=1}^{\ell}\left[\frac{(qx_k^2)_{N_k}}{(bx_k)_{N_k}(qa^{-1}x_k)_{N_k}}\right]\prod_{1\le r<s\le\ell}\left[\frac{(qx_rx_s)_{N_r}}{(qx_rx_sq^{N_s})_{N_r}}\right]$$

$$\times\left(\frac{b}{a}\right)_{N_1+\cdots+N_\ell}. \tag{2.6b}$$

Proof. We begin with Gustafson's C_ℓ ${}_6\psi_6$ summation theorem from [Gus89 ;Theorem 5.1]. Specializations serve to terminate this summation theorem from below and then from above. This yields the C_ℓ ${}_6\phi_5$ summation theorem , and then the C_ℓ terminating ${}_6\phi_5$ summation in Theorem 2.5, respectively.

Before carrying out the above specializations, we first make the following substitutions in Gustafson's C_ℓ ${}_6\psi_6$ summation theorem:

$$\begin{aligned} &a_i \mapsto a_iq^{-z_i}, &&\text{for } i=1,2,\dots,\ell;\\ &a_{\ell+1}\mapsto a;\\ &b_i\mapsto b_iq^{-z_i}, &&\text{for } i=1,2,\dots,\ell;\\ &b_{\ell+1}\mapsto b. \end{aligned} \tag{2.7}$$

Now set $b_1=b_2=\cdots=b_\ell=q$ in the resulting multiple Laurent series identity to terminate the sum side from below. Next, take $a_i=q^{-N_i}$ for $i=1,2,\dots\ell$, where each N_i is a non-negative integer. This terminates the sum side from above, and gives a summation theorem for a terminating multiple power series.

We then obtain Theorem 2.5 by first making the substitution $x_k=q^{z_k}$, for $k=1,2,\dots,\ell$, and then using $(a)_n=(a)_\infty/(aq^n)_\infty$ and $(a)_{-n}=(-q/a)^nq^{\binom{n}{2}}(q/a)_n^{-1}$ to simplify the product and sum side, respectively. □

Remark. A summary of the above substitutions that transform Gustafson's C_ℓ ${}_6\psi_6$ into Theorem 2.5 is given by:

$$\begin{aligned} &a_i\mapsto a_iq^{-z_i}\mapsto q^{-N_i}q^{-z_i}\mapsto q^{-N_i}x_i^{-1}, &&\text{for } i=1,2,\dots,\ell;\\ &a_{\ell+1}\mapsto a;\\ &b_i\mapsto b_iq^{-z_i}\mapsto q^{1-z_i}\mapsto qx_i^{-1}, &&\text{for } i=1,2,\dots,\ell;\\ &b_{\ell+1}\mapsto b;\\ &q^{z_i}\mapsto x_i. \end{aligned} \tag{2.8}$$

Remark. The $\ell=1$ case of (2.6) is the classical terminating ${}_6\phi_5$ summation in equation (II.21) of [GR90; pp. 238] in which $a\mapsto x_1^2$, $n\mapsto N_1$, $b\mapsto ax_1$, $c\mapsto qx_1b^{-1}$. That is, they are equivalent.

See §2 of [LM91] for the detailed proof of Theorem 2.5.

3. The G Bailey Transform

In this section we discuss the A_ℓ and C_ℓ multivariable extension of the classical A_1 Bailey Transform in Theorem 1.2. Motivated by Andrews [And79], Gessel and Stanton [GS83, GS86], and Agarwal, Andrews and Bressoud [AAB87] we generalize the matrix inversion formulation. This requires matrices M and M^* whose rows and columns are indexed by vectors of length ℓ of non-negative integers.

Throughout this talk, let $\boldsymbol{i} := (i_1, \dots, i_\ell)$, $\boldsymbol{j} := (j_1, \dots, j_\ell)$, $\boldsymbol{N} := (N_1, \dots, N_\ell)$, and $\boldsymbol{y} := (y_1, \dots, y_\ell)$ be vectors of length ℓ with non-negative integer components.

Define the Bailey transform matrices, M and M^*, as follows.

Definition 3.1 (M and M^* for A_ℓ). Let $a, x_1, \dots, x_\ell$ be indeterminate. Suppose that none of the denominators in (3.2) vanishes. Then let

$$M(\boldsymbol{i};\, \boldsymbol{j};\, A_\ell) := \prod_{r,s=1}^{\ell} \left(q\frac{x_r}{x_s} q^{j_r - j_s}\right)^{-1}_{i_r - j_r} \prod_{k=1}^{\ell} \left(aq\frac{x_k}{x_\ell}\right)^{-1}_{i_k + (j_1 + \cdots + j_\ell)}; \tag{3.2a}$$

and

$$\begin{aligned} M^*(\boldsymbol{i};\, \boldsymbol{j};\, A_\ell) := {} & \prod_{k=1}^{\ell} \left[1 - a\frac{x_k}{x_\ell} q^{i_k + (i_1 + \cdots + i_\ell)}\right] \prod_{k=1}^{\ell} \left(aq\frac{x_k}{x_\ell}\right)_{j_k + (i_1 + \cdots + i_\ell) - 1} \\ & \times \prod_{r,s=1}^{\ell} \left(q\frac{x_r}{x_s} q^{j_r - j_s}\right)^{-1}_{i_r - j_r} (-1)^{(i_1 + \cdots + i_\ell) - (j_1 + \cdots + j_\ell)}\, q^{\binom{(i_1 + \cdots + i_\ell) - (j_1 + \cdots + j_\ell)}{2}}. \end{aligned} \tag{3.2b}$$

Definition 3.3 (M and M^* for C_ℓ). Let $x_1, \dots, x_\ell$ be indeterminate. Suppose that none of the denominators in (3.4) vanishes. Then let

$$M(\boldsymbol{i};\, \boldsymbol{j};\, C_\ell) := \prod_{r,s=1}^{\ell} \left[\left(q\frac{x_r}{x_s} q^{j_r - j_s}\right)^{-1}_{i_r - j_r} \left(q x_r x_s q^{j_r + j_s}\right)^{-1}_{i_r - j_r}\right]; \tag{3.4a}$$

and

$$\begin{aligned} M^*(\boldsymbol{i};\, \boldsymbol{j};\, C_\ell) := {} & \prod_{r,s=1}^{\ell} \left[\left(q\frac{x_r}{x_s} q^{j_r - j_s}\right)^{-1}_{i_r - j_r} \left(x_r x_s q^{j_r + i_s}\right)^{-1}_{i_r - j_r}\right] \\ & \times \prod_{1 \le r < s \le \ell} \left[\frac{1 - x_r x_s q^{j_r + j_s}}{1 - x_r x_s q^{i_r + i_s}}\right] (-1)^{(i_1 + \cdots + i_\ell) - (j_1 + \cdots + j_\ell)}\, q^{\binom{(i_1 + \cdots + i_\ell) - (j_1 + \cdots + j_\ell)}{2}}. \end{aligned} \tag{3.4b}$$

Remark. The $\ell = 1$ case of (3.2) is the matrices in (1.3), and the $\ell = 1$ case of (3.4) is entrywise different than (1.3), but equivalent to it.

As in the classical case [AAB87], termwise transformations of a suitably modified A_ℓ or C_ℓ terminating very well-poised ${}_4\phi_3$ summation theorem lead to

Theorem 3.5 (Bailey Transform for A_ℓ and C_ℓ). *Let $G = A_\ell$ or C_ℓ. Let M and M^* be defined as in (3.2) and (3.4), with rows and columns ordered lexicographically. Then M and M^* are inverse, infinite, lower-triangular matrices. That is,*

$$\prod_{k=1}^{\ell} \delta(i_k, j_k) = \sum_{\substack{j_k \le y_k \le i_k \\ k=1,2,\ldots,\ell}} M(i;\, y;\, G)\, M^*(y;\, j;\, G), \tag{3.6}$$

where $\delta(r, s) = 1$ if $r = s$, and 0 otherwise.

Proof. In each case, A_ℓ and C_ℓ, we begin with a terminating very well-poised ${}_4\phi_3$ summation theorem. The A_ℓ ${}_4\phi_3$ summation follows immediately from the $b = aq/c$ case of Theorem 2.1 and the C_ℓ ${}_4\phi_3$ summation is similarly the $a = b$ case of Theorem 2.5.

We then multiply both the sum and product sides of the suitably specialized A_ℓ and C_ℓ terminating ${}_4\phi_3$ summations by some additional factors.

For A_ℓ, we multiply each side of the $N_k \mapsto i_k - j_k$, $\quad x_k \mapsto x_k q^{j_k}$, $\quad a \mapsto aq^{j_n + (j_1 + \cdots + j_\ell)}$ case of the A_ℓ terminating ${}_4\phi_3$ summation by the product

$$\prod_{r,s=1}^{\ell} \left(q \frac{x_r}{x_s} q^{j_r - j_s}\right)^{-1}_{i_r - j_r} \prod_{k=1}^{\ell} \left[\frac{\left(\frac{x_k}{x_\ell} aq\right)_{j_k + (j_1 + \cdots + j_\ell)}}{\left(\frac{x_k}{x_\ell} aq\right)_{i_k + (j_1 + \cdots + j_\ell)}} \right]. \tag{3.7}$$

For C_ℓ, we multiply each side of the $N_k \mapsto i_k - j_k$, $\quad x_k \mapsto x_k q^{j_k}$ case of the C_ℓ terminating ${}_4\phi_3$ summation by the product

$$\prod_{r,s=1}^{\ell} \left[\left(q \frac{x_r}{x_s} q^{j_r - j_s}\right)^{-1}_{i_r - j_r} \left(q x_r x_s q^{j_r + j_s}\right)^{-1}_{i_r - j_r} \right]. \tag{3.8}$$

In either case, the modified product side is seen to be the product of delta functions in the left-hand side of (3.6). The modified sum side is transformed term-by-term to yield the sum side of (3.6). The analysis here for the sum side consists of a lengthly series of elementary calculations. □

Remark. The detailed proof of the A_ℓ case of Theorem 3.5 is in §3 of [Mil91a], with ℓ replaced by n, and the above steps reversed into a verification proof. See §3 of [LM91] for the detailed analysis in the proof of the C_ℓ case.

Equations (3.2) and (3.4) motivate the definition of the A_ℓ and C_ℓ Bailey pair.

Definition 3.9 (G Bailey Pair). Let $G = A_\ell$ or C_ℓ. Let $N_k \ge 0$ be integers for $k = 1, 2, \ldots, \ell$. Let $A = \{A_{(y;\, G)}\}$ and $B = \{B_{(y;\, G)}\}$ be sequences. Let M and M^* be as above. Then we say that A and B form a G Bailey Pair if

$$B_{(N;\, G)} = \sum_{\substack{0 \le y_k \le N_k \\ k=1,2,\ldots,\ell}} M(N;\, y;\, G)\, A_{(y;\, G)}. \tag{3.10}$$

As a consequence of Theorem 3.5 and Definition 3.9 we have the following result.

Corollary 3.11 (G Bailey Pair Inversion). *A and B satisfy equation* (3.10) *if and only if*

$$A_{(N;\,G)} = \sum_{\substack{0\le y_k\le N_k\\ k=1,2,\dots,\ell}} M^*(N;\,y;\,G)\,B_{(y;\,G)}. \tag{3.12}$$

We study an important application of Corollary 3.11 in the next section.

4. G Balanced $_3\phi_2$ Summation Theorems

Corollary 3.11 applied to the G Bailey Pairs $(A_{(y;\,G)},\,B_{(y;\,G)})$ determined by Theorems 2.1 and 2.5 from §2 yields the corresponding G terminating balanced $_3\phi_2$ summations, and vice-versa. These calculations provide a G generalization of Andrews' application in [And79] of Corollary 1.7. The A_ℓ results here are contained in [Mil91a]. The $\ell = 1$ case of the summation theorems in this section are the corresponding classical results in [GR90].

In §4 of [Mil91a] we apply Corollary 3.11 to Theorem 2.1 to obtain

Theorem 4.1 (An A_ℓ generalization of the terminating balanced $_3\phi_2$ summation theorem). *Let a, b, c and $x_1, \dots, x_\ell$ be indeterminate, let N_i be non-negative integers for $i = 1, 2, \dots, \ell$ with $\ell \ge 1$, and suppose that none of the denominators in* (4.2) *vanishes. Then*

$$\left\{\frac{(c/a)_{N_1+\cdots+N_\ell}}{(c/ab)_{N_1+\cdots+N_\ell}}\prod_{k=1}^{\ell}\frac{\left(\frac{x_k}{x_\ell}c/b\right)_{N_k}}{\left(\frac{x_k}{x_\ell}c\right)_{N_k}}\right\} \tag{4.2a}$$

$$= \sum_{\substack{0\le y_k\le N_k\\ k=1,2,\dots,\ell}}\left\{\prod_{1\le r<s\le\ell}\left[\frac{1-\frac{x_r}{x_s}q^{y_r-y_s}}{1-\frac{x_r}{x_s}}\right]\right.$$

$$\times\prod_{r,s=1}^{\ell}\left[\frac{\left(\frac{x_r}{x_s}q^{-N_s}\right)_{y_r}}{\left(q\frac{x_r}{x_s}\right)_{y_r}}\right]\prod_{k=1}^{\ell}\left[\frac{\left(\frac{x_k}{x_\ell}a\right)_{y_k}}{\left(\frac{x_k}{x_\ell}c\right)_{y_k}}\right]$$

$$\left.\times\left[\frac{(b)_{y_1+\cdots+y_\ell}}{\left((ab/c)q^{1-(N_1+\cdots+N_\ell)}\right)_{y_1+\cdots+y_\ell}}\,q^{y_1+2y_2+\cdots+\ell y_\ell}\right]\right\}. \tag{4.2b}$$

Proof. We begin by multiplying both sides of (2.2) by

$$\left\{\prod_{r,s=1}^{\ell}\left(q\frac{x_r}{x_s}\right)_{N_r}^{-1}\prod_{k=1}^{\ell}\left(\frac{x_k}{x_\ell}aq\right)_{N_k}^{-1}\right\}, \tag{4.3}$$

and simplifying. By Definition 3.9, the product and sum sides of the resulting identity determine $B_{(N;\,G)}$ and $A_{(y;\,G)}$, respectively. Substitute this A_ℓ Bailey Pair into (3.12), simplify the resulting sum side termwise, and apply the relation $(a)_n = (-a)^n q^{\binom{n}{2}}\left(a^{-1}q^{1-n}\right)_n$

to suitable factors on the product side. Theorem 4.1 then follows once we make the substitutions $a \mapsto aq^{-(N_1+\cdots+N_\ell)}$, $c \mapsto (aq/c)q^{-(N_1+\cdots+N_\ell)}$, $b \mapsto c/b$, with x_i, N_i, q unchanged. □

Remark. The $\ell = 1$ and $N_1 = n$ case of (4.2) is equation (II.12) of [GR90; pp. 237].

We then went on in Theorem 4.15 of [Mil91a] to show that Theorem 4.1 and a polynomial argument lead to a summation theorem equivalent to (4.2) in which $b = q^{-N}$, $a = b$, and q^{-N_i} is replaced by a_i, for $i = 1, 2, \ldots, \ell$. The multiple sum in the second identity is taken over $y_1, \ldots, y_\ell \geq 0$ and $0 \leq y_1 + \cdots + y_\ell \leq N$, where N is a non-negative integer. The two identities are equivalent since the second one is a polynomial identity in each of a_i^{-1}, whose degree is a finite function of N, and (4.2) implies that the second holds for $a_i = q^{-N_i}$. Letting $N \to \infty$ in this second A_ℓ terminating balanced ${}_3\phi_2$ summation theorem then led in Theorem 5.1 of [Mil91a] to the A_ℓ Gauss summation theorem. This, in turn, yielded an A_ℓ q-Chu-Vandermonde summation and the non-terminating A_ℓ refinement of the q-binomial theorem. Many more analogous special limiting cases of additional A_ℓ terminating balanced ${}_3\phi_2$ summations can be found in §5 of [Mil91a].

We now consider the C_ℓ case. Applying Corollary 3.11 to Theorem 2.5 yields

Theorem 4.4. (A C_ℓ generalization of the terminating balanced ${}_3\phi_2$ summation theorem). *Let a, b and $x_1, \ldots, x_\ell$ be indeterminate, let N_i be non-negative integers for $i = 1, 2, \ldots, \ell$ with $\ell \geq 1$, and suppose that none of the denominators in (4.5) vanishes. Then*

$$\left\{\prod_{k=1}^{\ell}\left[\frac{(ax_k)_{N_k}\,(qx_kb^{-1})_{N_k}}{(bx_k)_{N_k}\,(qx_ka^{-1})_{N_k}}\right]\left(\frac{b}{a}\right)^{N_1+\cdots+N_\ell}\right\} \tag{4.5a}$$

$$= \sum_{\substack{0\leq y_k\leq N_k\\k=1,2,\ldots,\ell}}\left\{\prod_{1\leq r<s\leq\ell}\left[\frac{1-\frac{x_r}{x_s}q^{y_r-y_s}}{1-\frac{x_r}{x_s}}\;\frac{1-x_rx_sq^{y_r+y_s}}{1-x_rx_s}\right]\right.$$

$$\times\prod_{r,s=1}^{\ell}\left[\frac{\left(\frac{x_r}{x_s}q^{-N_s}\right)_{y_r}}{\left(q\frac{x_r}{x_s}\right)_{y_r}}\;\frac{\left(x_rx_sq^{N_s}\right)_{y_r}}{(qx_rx_s)_{y_r}}\right]$$

$$\times\prod_{1\leq r<s\leq\ell}\left[\frac{(qx_rx_s)_{y_r}}{(qx_rx_sq^{y_s})_{y_r}}\right]\prod_{k=1}^{\ell}\left[\frac{(qx_k^2)_{y_k}}{(bx_k)_{y_k}\,(qa^{-1}x_k)_{y_k}}\right]$$

$$\left.\times\left[\left(\frac{b}{a}\right)_{y_1+\cdots+y_\ell}q^{y_1+2y_2+\cdots+\ell y_\ell}\right]\right\}. \tag{4.5b}$$

Proof. We begin by multiplying both sides of (2.6) by

$$\left\{\prod_{r,s=1}^{\ell}\left(q\frac{x_r}{x_s}\right)_{N_r}^{-1}\,(qx_rx_s)_{N_r}^{-1}\right\}, \tag{4.6}$$

and simplifying. By Definition 3.9, the product and sum sides of the resulting identity determine $B_{(N;\,G)}$ and $A_{(y;\,G)}$, respectively. Substitute this C_ℓ Bailey Pair into (3.12), simplify the resulting sum side termwise, rewrite the product side, and then Theorem 4.4 follows. □

Remark. Note that there is some cancellation of factors in (4.5b). This allows us to write (4.5b) more compactly. In particular, the diagonal ($r = s$) factors in

$$\prod_{r,s=1}^{\ell} (qx_rx_s)_{y_r}^{-1} \qquad \text{cancell the factors} \qquad \prod_{k=1}^{\ell} (qx_k^2)_{y_k}.$$

Remark. The $\ell = 1$ case of (4.5) is the classical terminating balanced ${}_3\phi_2$ summation in equation (II.12) of [GR90; pp. 237] in which $n \mapsto N_1$, $a \mapsto x_1^2q^{N_1}$, $b \mapsto b/a$, $c \mapsto bx_1$. That is, they are equivalent.

Just as in the A_ℓ case, Theorem 4.4 and a polynomial argument lead to a summation theorem equivalent to (4.5) in which $b = aq^{-N}$, $a = b$, and q^{-N_i} is replaced by a_i, for $i = 1, 2, \ldots, \ell$. The two identities are equivalent since the second one is a polynomial identity in each of a_i^{-1}, whose degree is a finite function of N, and (4.5) implies that the second holds for $a_i = q^{-N_i}$. That is, we have

Theorem 4.7.(Second C_ℓ generalization of the terminating balanced ${}_3\phi_2$ summation theorem). *Let $a_1, \ldots, a_\ell, b$ and $x_1, \ldots, x_\ell$ be indeterminate, let N be a nonnegative integer, let $\ell \geq 1$, and suppose that none of the denominators in* (4.8) *vanishes. Then*

$$\left\{\prod_{k=1}^{\ell}\left[\frac{\left(qa_kx_k^{-1}b^{-1}\right)_N\left(qx_ka_k^{-1}b^{-1}\right)_N}{\left(qx_k^{-1}b^{-1}\right)_N\left(qx_kb^{-1}\right)_N}\right]\right\} \tag{4.8a}$$

$$= \sum_{\substack{y_1,\ldots,y_\ell\geq 0\\ 0\leq y_1+\cdots+y_\ell\leq N}} \left\{\prod_{1\leq r<s\leq \ell}\left[\frac{1-\frac{x_r}{x_s}q^{y_r-y_s}}{1-\frac{x_r}{x_s}}\,\frac{1-x_rx_sq^{y_r+y_s}}{1-x_rx_s}\right]\right.$$
$$\times\prod_{r,s=1}^{\ell}\left[\frac{\left(\frac{x_r}{x_s}a_s\right)_{y_r}}{\left(q\frac{x_r}{x_s}\right)_{y_r}}\,\frac{\left(x_rx_sa_s^{-1}\right)_{y_r}}{\left(qx_rx_s\right)_{y_r}}\right]$$
$$\times\prod_{1\leq r<s\leq\ell}\left[\frac{(qx_rx_s)_{y_r}}{\left(qx_rx_sq^{y_s}\right)_{y_r}}\right]\prod_{k=1}^{\ell}\left[\frac{\left(qx_k^2\right)_{y_k}}{\left(bx_kq^{-N}\right)_{y_k}\left(qb^{-1}x_k\right)_{y_k}}\right]$$
$$\left.\times\left[\left(q^{-N}\right)_{y_1+\cdots+y_\ell}q^{y_1+2y_2+\cdots+\ell y_\ell}\right]\right\}. \tag{4.8b}$$

Remark. Note that there is the same cancellation of factors in (4.8b) as there was in (4.5b).

Remark. The $\ell = 1$ case of (4.8) is the classical terminating balanced ${}_3\phi_2$ summation in equation (II.12) of [GR90; pp. 237] in which $n \mapsto N$, $a \mapsto a_1$, $b \mapsto x_1^2a_1^{-1}$, $c \mapsto qx_1b^{-1}$. That is, they are equivalent.

Letting $N \to \infty$ in Theorem 4.7 leads to the C_ℓ Gauss summation theorem. This, in turn, yields a C_ℓ q-Chu-Vandermonde summation and the non-terminating C_ℓ refinement of the q-binomial theorem. We include these results in our paper based on the longer version of this talk.

5. The G Bailey Lemma

In this section we motivate and then state the A_ℓ and C_ℓ generalization of the classical A_1 Bailey Lemma in Theorem 1.9. It shows how to construct another G Bailey Pair from an arbitrary G Bailey Pair.

Consider the sequence $A' = \{A'_{(N;\,G)}\}$ defined by

$$A'_{(N;\,G)} := C_N \; A_{(N;\,G)}, \tag{5.1}$$

where the sequence $C = \{C_y\}$ is as of yet unchosen, and A and B form a G Bailey Pair. We want to find a sequence $B' = \{B'_{(y;\,G)}\}$ so that A' and B' also form a G Bailey Pair. That is, we need

$$B'_{(N;\,G)} = \sum_{\substack{0 \le y_k \le N_k \\ k=1,2,\ldots,\ell}} M(N;\, y;\, G)\, A'_{(y;\,G)}. \tag{5.2}$$

Assume that (3.10), (3.12), (5.1), and (5.2) hold, and that $M(i;\, j;\, G) \equiv M(i;\, j)$ and $M^*(i;\, j;\, G) \equiv M^*(i;\, j)$. Then

$$B'_{(N;\,G)} = \sum_{\substack{0 \le y_k \le N_k \\ k=1,2,\ldots,\ell}} \left\{ M(N;\, y)\, C_y\, A_{(y;\,G)} \right\} \tag{5.3a}$$

$$= \sum_{\substack{0 \le y_k \le N_k \\ k=1,2,\ldots,\ell}} \left\{ M(N;\, y)\, C_y \sum_{\substack{0 \le m_i \le y_i \\ i=1,2,\ldots,\ell}} \left[M^*(y;\, m)\, B_{(m;\,G)} \right] \right\} \tag{5.3b}$$

$$= \sum_{\substack{0 \le m_k \le N_k \\ k=1,2,\ldots,\ell}} \left\{ B_{(m;\,G)} \sum_{\substack{m_i \le y_i \le N_i \\ i=1,2,\ldots,\ell}} \left[M(N;\, y)\, M^*(y;\, m)\, C_y \right] \right\} \tag{5.3c}$$

$$= \sum_{\substack{0 \le m_k \le N_k \\ k=1,2,\ldots,\ell}} \left\{ B_{(m;\,G)} \sum_{\substack{0 \le y_i \le N_i - m_i \\ i=1,2,\ldots,\ell}} \left[M(N;\, y+m)\, M^*(y+m;\, m)\, C_{y+m} \right] \right\}. \tag{5.3d}$$

We want to choose $C = \{C_y\}$ so that each C_{y+m} can be factored into a function that is independent of y times a function of m and y. The expression that is independent of y will then be pulled outside the sum. We also desire that the remaining terms combine with those in the inner sum of (5.3d) to form an easily summable expression. In effect, C allows us to pass from a G ${}_4\phi_3$ to a G ${}_6\phi_5$ which is summable by either Theorem 2.1 or 2.5. Such a choice of C allows us to sum the inner sum in (5.3d) and derive a more compact expression for $B'_{(N;\,G)}$.

Keeping in mind the above discussion, we first define the sequences $A' = \{A'_{(y;\,A_\ell)}\}$ and $B' = \{B'_{(y;\,A_\ell)}\}$ by

$$A'_{(N;\,A_\ell)} := \prod_{k=1}^{\ell} \left(\frac{aq}{\rho}\frac{x_k}{x_\ell}\right)_{N_k}^{-1} \prod_{k=1}^{\ell}\left(\sigma\frac{x_k}{x_\ell}\right)_{N_k} \times \frac{(\rho)_{N_1+\cdots+N_\ell}}{(aq/\sigma)_{N_1+\cdots+N_\ell}} (aq/\rho\sigma)^{N_1+\cdots+N_\ell} A_{(N;\,A_\ell)} \tag{5.4a}$$

and

$$B'_{(N;\,A_\ell)} := \sum_{\substack{0\le y_k\le N_k\\ k=1,2,\ldots,\ell}} \Bigg\{ \prod_{k=1}^{\ell}\left[\left(\sigma\frac{x_k}{x_\ell}\right)_{y_k}\left(\frac{aq}{\rho}\frac{x_k}{x_\ell}\right)_{N_k}^{-1}\right] \prod_{r,s=1}^{\ell}\left(q\frac{x_r}{x_s}q^{y_r-y_s}\right)_{N_r-y_r}^{-1} \times \frac{(aq/\rho\sigma)_{(N_1+\cdots+N_\ell)-(y_1+\cdots+y_\ell)}\,(\rho)_{y_1+\cdots+y_\ell}}{(aq/\sigma)_{N_1+\cdots+N_\ell}} (aq/\rho\sigma)^{y_1+\cdots+y_\ell} B_{(y;\,A_\ell)}\Bigg\} \tag{5.4b}$$

We next define the sequences $A' = \{A'_{(y;\,C_\ell)}\}$ and $B' = \{B'_{(y;\,C_\ell)}\}$ by

$$A'_{(N;\,C_\ell)} := \prod_{k=1}^{\ell}\left[\frac{(\alpha x_k)_{N_k}\,(qx_k\beta^{-1})_{N_k}}{(\beta x_k)_{N_k}\,(qx_k\alpha^{-1})_{N_k}}\right]\left(\frac{\beta}{\alpha}\right)^{N_1+\cdots+N_\ell} A_{(N;\,C_\ell)} \tag{5.5a}$$

and

$$B'_{(N;\,C_\ell)} := \sum_{\substack{0\le y_k\le N_k\\ k=1,2,\ldots,\ell}} \Bigg\{ \prod_{k=1}^{\ell}\left[\frac{(\alpha x_k)_{y_k}\,(qx_k\beta^{-1})_{y_k}}{(\beta x_k)_{N_k}\,(qx_k\alpha^{-1})_{N_k}}\right] \prod_{r,s=1}^{\ell}\left(q\frac{x_r}{x_s}q^{y_r-y_s}\right)_{N_r-y_r}^{-1} \times \prod_{1\le r<s\le\ell}\left[\left(qx_rx_sq^{y_r+y_s}\right)_{N_s-y_s}^{-1}\left(qx_rx_sq^{N_s-y_s}\right)_{N_r-y_r}^{-1}\right] \times \left(\frac{\beta}{\alpha}\right)_{(N_1+\cdots+N_\ell)-(y_1+\cdots+y_\ell)}\left(\frac{\beta}{\alpha}\right)^{y_1+\cdots+y_\ell} B_{(y;\,C_\ell)}\Bigg\} \tag{5.5b}$$

These definitions lead to

Theorem 5.6 (The G generalization of the classical A_1 Bailey Lemma). *Let $G = A_\ell$ or C_ℓ. Suppose $A = \{A_{(N;\,G)}\}$ and $B = \{B_{(N;\,G)}\}$ form a G Bailey Pair. If $A' = \{A'_{(N;\,G)}\}$ and $B' = \{B'_{(N;\,G)}\}$ are as above, then A' and B' also form a G Bailey Pair.*

Proof. The definitions in (5.4) and (5.5) are substituted into (3.10). After an interchange of summation, the inner sum is seen to be a special case of the appropriate ${}_6\phi_5$. The ${}_6\phi_5$ is then summed, and the desired result follows. □

Corollary 5.7. *With $A' = \{A'_{(y;\,G)}\}$ and $B' = \{B'_{(y;\,G)}\}$ defined as in Theorem* 5.6, *A' and B' satisfy equation* (3.12).

Notice that we may apply the G Bailey Lemma to the new G Bailey Pair A' and B'. Call the resulting G Bailey Pair (A'', B''). We may continue applying the G Bailey Lemma and create a sequence of G Bailey Pairs

$$(A,B) \to (A',B') \to (A'',B'') \to \cdots .$$

We call this sequence the "G Bailey Chain." This definition is motivated by Andrews [And86a].

We may also move from (A',B') back to (A,B). Given a G Bailey Pair (A',B'), we may determine A from equation (5.4a) or (5.5a) and then B from equation (3.10). Thus, we can move from right to left in the G Bailey Chain. This gives us the "bilateral G Bailey Chain"

$$\cdots \to (A^{(-2)},B^{(-2)}) \to (A^{(-1)},B^{(-1)}) \to (A,B) \to (A',B') \to (A'',B'') \to \cdots .$$

Many of the classical applications mentioned just after Theorem 1.9 extend to the setting of the above G Bailey Chains.

References

[AAB87] A. K. Agarwal, G. Andrews, and D. Bressoud, *The Bailey lattice*, J. Indian Math. Soc. **51** (1987), 57–73.

[AV84] W. A. Al-Salam and A. Verma, *On quadratic transformations of basic series*, SIAM J. Math. Anal. **15** (1984), 414–420.

[And75] G. E. Andrews, *Problems and prospects for basic hypergeometric functions*, Theory and Applications of Special Functions (R. Askey, ed.), Academic Press, New York, 1975, pp. 191–224.

[And76] ———, *The Theory of Partitions*, Encyclopedia of Mathematics and Its Applications (G.-C. Rota, ed.), vol. 2, Addison-Wesley, Reading, Mass., 1976; reissued by Cambridge University Press, Cambridge, 1985.

[And79] ———, *Connection coefficient problems and partitions*, Proc. Sympos. Pure Math. (D. Ray-Chaudhuri, ed.), vol. 34, 1979, pp. 1–24.

[And84] ———, *Multiple series Rogers–Ramanujan type identities*, Pacific J. Math. **114** (1984), 267–283.

[And86a] ———, *q-Series: Their development and application in analysis, number theory, combinatorics, physics and computer algebra*, NSF CBMS Regional Conference Series, vol. 66, 1986.

[And86b] ———, *The fifth and seventh order mock theta functions*, Trans. Amer. Math. Soc. **293** (1986), 113–134.

[ABF84] G. E. Andrews, R. J. Baxter, and P. J. Forrester, *Eight-vertex SOS model and generalized Rogers–Ramanujan-type identities*, J. Statist. Phys. **35** (1984), 193–266.

[ADH88] G. E. Andrews, F. J. Dyson, and D. Hickerson, *Partitions and indefinite quadratic forms*, Invent. Math **91** (1988), 391–407.

[Bai35] W. N. Bailey, *Generalized Hypergeometric Series*, Cambridge University Press, Cambridge, 1935; reprinted by Stechert-Hafner, New York, 1964.

[Bai47] ———, *Some identities in combinatory analysis*, Proc. London Math. Soc. (2) **49** (1947), 421–435.

[Bai49] ———, *Identities of the Rogers–Ramanujan type*, Proc. London Math. Soc. (2) **50** (1949), 1–10.

[Bax82] R. J. Baxter, *Exactly Solved Models in Statistical Mechanics*, Academic Press, London and New York, 1982.

[BL68] L. C. Biedenharn and J. D. Louck, *A pattern calculus for tensor operators in the unitary groups*, Comm. Math. Phys. **8** (1968), 89–131.

[BL81a] ———, *Angular Momentum in Quantum Physics: Theory and Applications*, Encyclopedia of Mathematics and Its Applications (G.-C. Rota, ed.), vol. 8, Addison-Wesley, Reading, Mass., 1981.

[BL81b] ———, *The Racah-Winger Algebra in Quantum Theory*, Encyclopedia of Mathematics and Its Applications (G.-C. Rota, ed.), vol. 9, Addison-Wesley, Reading, Mass., 1981.

[Bre81] D. M. Bressoud, *Some identities for terminating q-series*, Math. Proc. Cambridge Philos. Soc. **89** (1981), 211–223.

[Bre83] ———, *A matrix inverse*, Proc. Amer. Math. Soc. **88** (1983), 446–448.

[Car73] L. Carlitz, *Some inverse relations*, Duke Math. J. **40** (1973), 803–901.

[Dys43] F. J. Dyson, *Three identities in combinatory analysis*, J. London Math. Soc. **18** (1943), 35–39.

[Gas89] G. Gasper, *Summation, transformation, and expansion formulas for bibasic series*, Trans. Amer. Math. Soc. **312** (1989), 257–277.

[GR90] G. Gasper and M. Rahman, *Basic Hypergeometric Series*, Encyclopedia of Mathematics and Its Applications (G.-C. Rota, ed.), vol. 35, Cambridge University Press, Cambridge, 1990.

[GS83] I. Gessel and D. Stanton, *Applications of q-Lagrange inversion to basic hypergeometric series*, Trans. Amer. Math. Soc. **277** (1983), 173–201.

[GS86] ———, *Another family of q-Lagrange inversion formulas*, Rocky Mountain J. Math. **16** (1986), 373–384.

[Gus89] R. A. Gustafson, *The Macdonald identities for affine root systems of classical type and hypergeometric series very well-poised on semi-simple Lie algebras*, Ramanujan International Symposium on Analysis (Dec. 26th to 28th, 1987, Pune, India) (N. K. Thakare, ed.), 1989, pp. 187–224.

[Hol80] W. J. Holman, III, *Summation Theorems for hypergeometric series in $U(n)$*, SIAM J. Math. Anal. **11** (1980), 523–532.

[HBL76] W. J. Holman III, L. C. Biedenharn, and J. D. Louck, *On hypergeometric series well-poised in $SU(n)$*, SIAM J. Math. Anal. **7** (1976), 529–541.

[Jack10] F. H. Jackson, *Transformations of q-series*, Messenger of Math. **39** (1910), 145–153.

[Lil91] G. M. Lilly, *The C_ℓ generalization of Bailey's transform and Bailey's lemma*, Ph.D. Thesis (1991), University of Kentucky.

[LM91] G. M. Lilly and S. C. Milne, *The C_ℓ Bailey Transform and Bailey Lemma*, preprint.

[Mil85] S. C. Milne, *An elementary proof of the Macdonald identities for $A_\ell^{(1)}$*, Adv. in Math. **57** (1985), 34–70.

[Mil87] ———, *Basic hypergeometric series very well-poised in $U(n)$*, J. Math. Anal. Appl. **122** (1987), 223–256.

[Mil91a] ———, *Balanced ${}_3\phi_2$ summation theorems for $U(n)$ basic hypergeometric series*, in preparation.

[Mil91b] ———, *New Whipple's transformations for basic hypergeometric series in $U(n)$*, in preparation.

[Mil91c] ———, *A $U(n)$ generalization of Bailey's lemma*, in preparation.

[Mil91d] ———, *An extension of little q-Jacobi polynomials for basic hypergeometric series in $U(n)$*, in preparation.

[Mil91e] ———, *A $U(n)$ generalization of the Bailey lattice*, in preparation.

[Mil91f] ———, *Iterated multiple series expansions of basic hypergeometric series very well-poised in $U(n)$*, in preparation.

[ML91] S. C. Milne and G. M. Lilly, *The A_ℓ and C_ℓ Bailey transform and lemma*, Bull. Amer. Math. Soc. (N.S.) , in press.

[Pau82] P. Paule, *Zwei neue Transformationen als elementare Anwendungen der q-Vandermonde Formel*, Ph.D. Thesis (1982), University of Vienna.

[Pau85] ———, *On identities of the Rogers–Ramanujan type*, J. Math. Anal. Appl. **107** (1985), 255–284.

[Rog94] L. J. Rogers, *Second memoir on the expansion of certain infinite products*, Proc. London Math. Soc. **25** (1894), 318–343.

[Rog95] ———, *Third memoir on the expansion of certain infinite products*, Proc. London Math. Soc. **26** (1895), 15–32.

[Rog17] ———, *On two theorems of combinatory analysis and some allied identities*, Proc. London Math. Soc (2) **16** (1917), 315–336.

[Sla51] L. J. Slater, *A new proof of Roger's transformation of infinite series*, Proc. London Math. Soc (2) **53** (1951), 460–475.

[Sla52] ———, *Further identities of the Rogers–Ramanujan type*, Proc. London Math. Soc (2) **54** (1952), 147–167.

[Sla66] ———, *Generalized Hypergeometric Functions*, Cambridge University Press, London and New York, 1966.

[Wat29] G. N. Watson, *A new proof of the Rogers–Ramanujan identities*, J. London Math. Soc. **4** (1929), 4–9.

[Whi24] F. J. W. Whipple, *On well-poised series, generalized hypergeometric series having parameters in pairs, each pair with the same sum*, Proc London Math. Soc. (2) **24** (1924), 247–263.

[Whi26] ———, *Well-poised series and other generalized hypergeometric series*, Proc. London Math. Soc. (2) **25** (1926), 525–544.

Department of Mathematics, The Ohio State University, Columbus, Ohio, 43210

E-mail: milne@function.mps.ohio-state.edu

Une preuve bijective d'une formule de Touchard-Riordan.+

Jean-Guy Penaud

LaBRI*

Résumé. Une preuve bijective d'une formule tirée par Riordan d'un papier de Touchard est présentée. Cette preuve procède par un enchaînement de bijections allant des involutions aux polyominos, en passant par les histoires de fichiers, les mots, les arbres à deux types de sommets, les couples de suites d'entiers jusqu'à l'objet final où la propriété devient évidente.

Abstract. A pure combinatorial proof of a formula which was found by Riordan from a paper of Touchard is shown.

1. La formule de Touchard-Riordan

Il est classique de représenter une involution sans point fixe sur les entiers de 1à 2n (ensemble que l'on note [1..2n]) par un diagramme composé de n cordes joignant deux à deux 2n points distincts répartis sur un cercle. Le lecteur trouvera dans un autre papier de Dulucq et l'auteur un aperçu des relations entre cordes, arbres et permutations [8]. La figure 1 ci-contre montre la configuration de cordes associée à l'involution

$$\alpha = (1,4)\ (2,6)\ (3,5)\ (7,8).$$

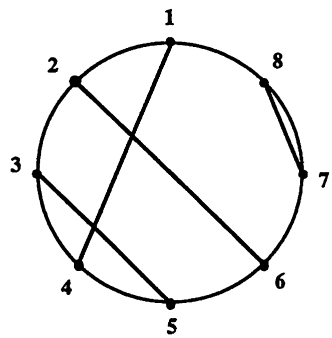


Figure 1.

Les questions qui viennent alors naturellement sont celles relatives aux *points de croisement* des cordes, c'est-à-dire, en termes d'involution sur [1,2n], aux quadruplets (i,j,k,l) tels que $1 \leq i<j<k<l \leq 2n$ et que (i,k) et (j,l) soient des cycles de l'involution.

— Quel est le nombre de configurations de cordes qui ne se coupent pas?

— Quel est le nombre de configurations ayant un nombre de croisements donné?

C'est, d'après Riordan [20], à Errera [9] que l'on doit d'avoir le premier posé et résolu le problème de compter le nombre de configurations où les cordes ne se coupent pas. Ces configurations sont, comme beaucoup d'autres objets combinatoires, comptées par les nombres de Catalan,

+ Travail effectué avec le soutien du PRC Mathématiques et Informatique.
* UA CNRS n° 726, Université de Bordeaux I, 33405 Talence, France.

$$C_n = \frac{1}{n+1}\binom{2n}{n}. \tag{1}$$

Dans une série de trois articles consacrés à l'étude de configurations associées au *problème du timbre poste*, Touchard ([22], [23], [24]) étudie la seconde question. Mais plutôt que d'utiliser un cercle de rayon fini pour représenter une involution sans point fixe, Touchard utilise un vocabulaire architectural et considère les cordes comme un *système d'arches* dont les extrémités sont des points d'un axe horizontal qui s'échangent dans l'involution (notons que d'autres comme Read, [20], les représentent par un *diagramme en dents de scie)*. La figure 2 montre ces deux présentations d'une même involution sans point fixe α.

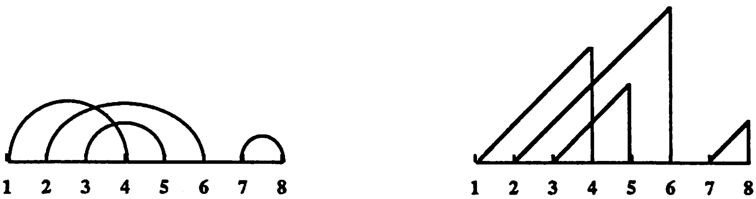


Figure 2. Deux représentation d'une involution sans point fixe.

Touchard trouve une solution dont l'expression, reformulée par Riordan [19] de manière plus condensée et plus explicite, s'écrit à l'aide des *nombres de Delannoy* (appelés *ballot numbers* par l'école anglo-saxonne) . En reprenant les notations de Touchard [23], c'est-à-dire en appelant $T_n(x)$ le polynôme énumérateur des configurations de n cordes selon leur nombre de croisements, et $t_n(x)$ le polynôme dont les coefficients $t_{i\,j}$ sont les nombres de Delannoy, on peut écrire,

$$t_n(x) = (1-x)^n\ T_n(x) = \sum_{j=0,n} (-1)^j t_{n,j} x^{\frac{j(j+1)}{2}}, \tag{2}$$

avec

$$t_{n,j} = D_{n+j}^{n-j} = \binom{2n}{n-j} - \binom{2n}{n-j-1}. \tag{3}$$

REMARQUE 1.1

Le polynôme $T_n(x)$ est aussi le moment μ_n des q-analogues des polynômes d'Hermite H_n(x;q) introduits par Rogers pour démontrer les célèbres identités de Rogers-Ramanujan.

Dans cet article nous nous proposons de donner une preuve bijective des formules (2) et (3) que nous appellerons formules de Touchard-Riordan, en interprétant les coefficients $t_{n\,j}$ à l'aide d'une classe particulière d'involutions sans point fixe. Nous utilisons, outre les configurations de cordes, d'autres objets combinatoires en bijection avec les involutions, que sont les mots, les chemins et les arbres. Mais c'est grâce à un sixième objet, la famille des *polyominos horizontalement convexes*, que la preuve apparaît clairement.

2. La chaîne d'objets utilisés

a) Les involutions sans points fixes et leurs histoires

La figure suivante montre une involution sans point fixe et l'*histoire de fichier* associée est constituée d'un chemin de Dyck, avec des valuations sur les pas descendants indiquant le nombre de croisements (cf [12], [13]). C'est ici une *valuation de Hermite*, c'est-à-dire nulle sur les pas montants, et comprise entre 0 et la hauteur de l'extrémité du pas pour les pas descendants. Elle correspond a une histoire de file de priorité.

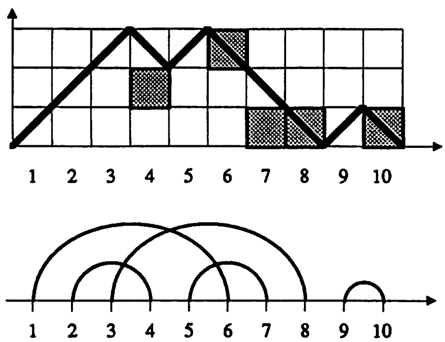


Figure 3.
Une involution et son histoire.

(La somme des hauteurs des bases des carrés donne le nombre de points de croisements de l'involution).

REMARQUE 2.1 Les q-analogues des polynômes d'Hermite, encore appelés polynômes de q-Hermite, $H_n(x;q)$, évoqués à la remarque 1 satisfont la récurrence,

$$H_{n+1}(x;q) = xH_n(x;q) - (1+q+q^2+...+q^{n-1})H_{n-1}(x;q) .$$

D'après la théorie combinatoire des polynômes orthogonaux (voir Flajolet [11] ou Viennot [25]), la valuation des chemins de Dyck interprétant les moments est celle des histoires fdp. Ceci démontre la remarque 1.1, à savoir que le moment des polynômes de q-Hermite est le polynôme énumérateur des involutions selon le nombre de croisements. L'identité (2) de Riordan-Touchard peut aussi être déduite du calcul des moments des polynômes de q-Hermite (voir [16]). Rappelons enfin que la combinatoire des croisements de ces involutions joue un rôle fondamental dans la partie bijective de l'évaluation de l'intégrale de Askey-Wilson ([2], [16]).

Soit T(z,x) la série commutative énumérant les involutions selon le nombre de paires et celui de croisements. En utilisant le polynôme $T_n(x)$ introduit dans la formule de Touchard-Riordan en (2), on peut écrire

$$T(z,x)=\sum_{n\geq 0} T_n(x)z^n.$$

b) Involutions tout-ou-rien connexes (TORC)

Particularisons les involutions. Nous dirons qu'une involution sur [1..2n] est *connexe* si et seulement s'il n'existe pas d'entier k tel que $1 \leq k < 2n$ et que $\alpha(i) \leq k$ pour tout i tel que $1 \leq i \leq k$.

D'autre part, considérons les involutions vérifiant l'alternative suivante, pour toute fermante, c'est-à-dire tout entier k tel que $\alpha(k) = i$ et $i < k$,

ι) k est une *fermante minimale*, i.e.tout ouvrante inférieure à k se ferme avant, soit

$$i < j < k \Rightarrow \alpha(j) < k,$$

ιι) k est une *fermante maximale*, i.e. toute ouvrante inférieure à α(k) se ferme avant k, soit $j < i \quad \Rightarrow \alpha(j) < k.$

La figure 4 montre une involution TORC. Ces involutions TORC ayant n+1 arches sont codés par des histoires fdp associées à des mots de Dyck premiers à 2n+2 lettres, donc des mots de Dyck à 2n lettres.

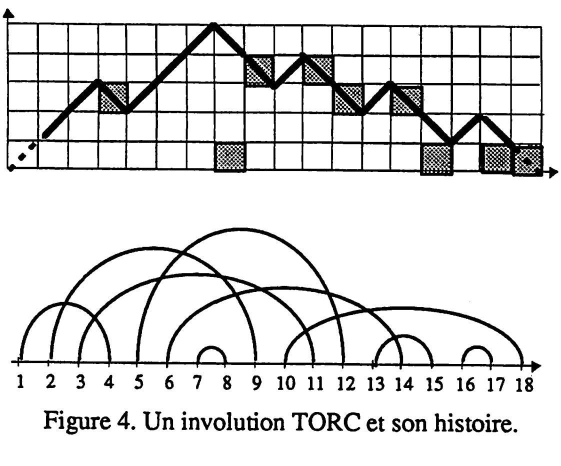


Figure 4. Un involution TORC et son histoire.

A toute involution TORC codée par un mot f, on peut associer un *signe* égal à $(-1)^{r_f}$ où r_f est le nombre de fermantes maximales.

Appelons $\Theta_{n,p,r}$ le nombre d'involutions ayant n+1 arches, p croisements et r fermantes maximales, et posons

$$\Theta_{n,p} = \sum_{r=1}^{p} (-1)^r \Theta_{n,p,r} \ , \tag{4}$$

Θn(x) le polynôme en x ainsi défini,

$$\Theta_n(x) = \sum_{p=0}^{n} \Theta_{n,p} x^p \ ,$$

et

$$\Theta(z,x) = \sum_{n\geq 1} \Theta_n(x) z^n.$$

Pour établir bijectivement la formule de Touchard-Riordan, il suffit de montrer les deux propositions suivantes.

PROPRIETE 2.2 *Pour tout* $n \geq 1$ *on a l'égalité,*

$$\Theta_n(x) = (1-x)^n \, T_n(x). \qquad (5)$$

PROPRIETE 2.3 *Pour toutes valeurs de* n *et de* p, $\Theta_{n,p}=0$, *sauf si* $p = 1+2+...+k = \binom{k+1}{2}$ *alors* $\Theta_{n,p} = (-1)^k t_{n,k}$ *avec* $t_{n,k} = D^{n-k}_{n+k} = \binom{2n}{n-k} - \binom{2n}{n-k-1}$.

La première est une conséquence directe du codage, et la suite del'article est consacrée à la preuve bijective de la seconde.

c) Mots de Dyck bicolorés et arbres à cerises.

Puisqu'il n'y a que deux cas possibles pour une fermante d'une involution TORC, ces involutions peuvent être codées par un mot sur un alphabet à trois lettres, la seule lettre z codant une ouvrante, et les deux lettres $\bar{z}$ et $\bar{y}$ codant une fermante suivant que celle-ci est minimale ou maximale. On a ainsi le langage L d'équation,

$$L = \varepsilon + z L \bar{z} L + z L \bar{y} L. \qquad (6)$$

Pour tout mot g de L codant une involution TORC, $|g|_{\bar{y}}$ est égal au nombre de fermantes maximales, et le *poids* de g, égal à la somme des hauteurs des $\bar{y}$, s'exprime en posant pour tout mot u, $\delta(u) = |u|_z - \left(|u|_{\bar{z}} + |u|_{\bar{y}}\right)$, par,

$$\varpi(g) = \sum_{g=u\bar{y}v} \delta(u).$$

Ce poids est égal au nombre de croisements.

Ainsi le nombre $\Theta_{n,p,r}$, défini précédement, est égal au nombre de mots g de L tels que $|g| = n$, $|g|_{\bar{z}} = r$, et $\varpi(g) = p$, c'est-à-dire aux involutions à n+1 arches, p croisements et r fermantes maximales.

C'est aussi le nombre d'une certaine famille d'arbres, les arbres à cerises. En effet les arbres codés par les mots du langage L ont des sommets de deux types: les cerises (carrées), et les simples. Un sommet sera dit *carré* si et seulement si on le quitte par une arête codée $\bar{y}$. Ainsi $\Theta_{n,p,r}$ sera le nombre d'arbres planaires ayant n arêtes, r sommets carrés, et dont la longueur de *cheminement* des sommets carrés est égale à p.

Exemple. Le mot z z $\bar{y}$ z z z $\bar{z}$ $\bar{y}$ z $\bar{y}$ $\bar{y}$ z $\bar{y}$ $\bar{z}$ z $\bar{z}$, correspondant à l'involution de la figure 4 code aussi l'arbre représenté dans la figure 5 qui a 5 carrés et une longueur de cheminement de 12.

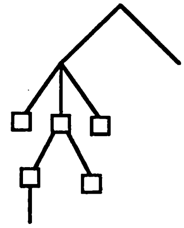


Figure 5. Un arbre à carrés.

Considérons les involutions TORC possédant la propriété supplémentaire que les fermantes qui suivent immédiatement une ouvrante sont maximales. Le langage associé P est un sous-ensemble de L vérifiant l'équation,

$$P = z\dot{z} + z\dot{z}P + zP\dot{z} + zP\dot{z}P + zP\dot{y} + zP\dot{y}P. \qquad (7)$$

Pour les arbres codés par ce langage, toutes les feuilles sont des sommets carrés.

On a alors une propriété de factorisation des mots de L analogue à la propriété de la *factorisation de Catalan* de Chottin et Cori, [4], qu'énonce le lemme suivant, en notant D le langage de Dyck sur z et $\dot{z}$.

LEMME 2.4. *Tout mot* f *de* L *a une unique factorisation* , $f = f_1u_1f_2u_2...f_{2p}u_{2p}f_{2p+1}$, *vérifiant les conditions suivantes,*

les f_i *sont des mots de* D, $(1 \le i \le 2p+1)$,

u_i, $1 \le i \le 2p$, *est une lettre de* $\{z, \dot{z}, \dot{y}\}$,

$u_1u_2...u_{2p}$ *est un mot de* P.

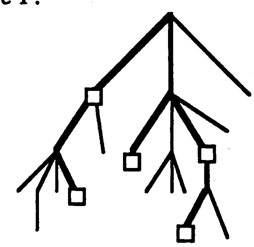


Figure 6. Un arbre à sommets de deux types et son squelette (en gras).

On appelle *squelette* de l'arbre le sous arbre maximal dont toutes les feuilles sont des carrés. La figure 6 ci-dessus montre le squelette en trait épais. Le mot code est le mot suivant où les lettres en gras forment le mot de P, qui code clairement le sous-arbre.

z z z ż z z ż ż z ż ż z ẏ ż z ż ẏ z z ẏ z z ż z ż ż z ż ż z z z z ẏ z ż ż ẏ z ż ż z ż

Exemples. a) $\Theta_{n,1} = -\Theta_{n,1,1}$ où $\Theta_{n,1,1}$ compte le nombre d'involutions sans point fixe ayant n+1 arches, un croisement et une fermante maximale de hauteur nécessairement 1. C'est donc aussi le nombre de mots de L qui les code; or ces mots possèdent un seul $\dot{y}$ et cet $\dot{y}$ est de hauteur 1 dans le mot. D'après le lemme précédent ce sous-ensemble de L est égal à D z D $\dot{y}$ D, or ce langage est clairement en bijection avec le langage D z D z D c'est-à-dire le langage des facteurs gauches de Dyck de hauteur finale 2, et l'on a

$$\Theta_{n,1,1} = D^{n-1}_{n+1} .$$

b) De même pour p=2, on a, d'après la formule (4),

$$\Theta_{n,2} = \sum_{r=1}^{2} (-1)^r \Theta_{n,2,r} .$$

Or d'après le lemme précédent, $\Theta_{n,2,1}$ compte le nombre de mots de longueur 2n du langage D z D z D $\bar{y}$ D $\bar{z}$ D et $\Theta_{n,2,2}$ compte le nombre de mots de longueur 2n du langage D z D $\bar{y}$ D z D $\bar{y}$ D. Ces deux langages sont clairement en bijection et l'on a $\Theta_{n,2,1}=\Theta_{n,2,2}$, donc $\Theta_{n,2}= 0$.

Cette propriété est due d'une part au fait que les mots z z $\bar{y}$ $\bar{z}$ et z $\bar{y}$ z $\bar{y}$ ont même poids et des signes opposés, et d'autre part que tout mot de L a une unique factorisation. Ce résultat se généralise et l'on a le théorème,

THEOREME 2.5 *Il existe une involution qui opère sur les mots de* P. *Cette involution conserve la longueur et le poids du mot, c'est-à-dire la somme des hauteurs des lettres* $\bar{y}$ *et change le signe, c'est-à-dire la parité du nombre de lettres* $\bar{y}$. *Ses seuls points fixes sont les mots* $z^p\bar{y}^p$, $(p\geq 1)$.

Avant d'aborder la preuve du théorème, qui sera faite à l'aide d'un nouvel objet, remarquons que la propriété 2.3 est un corollaire immédiat de ce théorème.

PREUVE DE LA PROPRIETE 2.3

Soit f un mot de L. D'après le lemme 2.4, il s'écrit,

$$f = f_1u_1f_2u_2...f_{2p}u_{2p}f_{2p+1}.$$

Si son squelette $u_1u_2...u_{2p}$ n'est pas de la forme $z^r\bar{y}^r$, il lui correspond d'après le théorème 2.5 un mot différent de même longueur et de même poids mais de signe opposé, $v_1v_2...v_{2p}$, donc le mot g= $f_1v_1f_2v_2...f_{2p}v_{2p}f_{2p+1}$ a même longueur, même poids que f et signe opposé, et donc leurs contributions s'annulleront mutuellement dans le calcul de $\Theta_{n,p}$.

Par contre tous les mots ayant même longueur et comme squelette $z^r\bar{y}^r$ vont avoir même poids $p = 1+2+...+r = \frac{r(r+1)}{2}$ et même signe, et donc leur contribution va s'ajouter. Le coefficient de x^p dans $\Theta_{n,p}$ sera donc au signe près le nombre de mots de longueur 2n dans $(D\ z)^r\ (D\ \bar{y})^r\ D$, clairement en bijection avec le langage $(D\ z)^r\ (D\ z)^r\ D$. C'est donc le nombre de facteurs gauches de mots de Dyck de longueur 2n et de hauteur finale 2r, c'est-à-dire le nombre de Delannoy D_{n+r}^{n-r}.

Dans le polynôme

$$\Theta_n(x) = \sum_{p\geq 0}\Theta_{n,p}x^p,$$

les seuls coefficients non nuls sont ceux pour lesquels p est un entier vérifiant,

— $0 \leq p \leq \frac{n(n+1)}{2}$,

— il existe un entier r tel que $p = \frac{r(r+1)}{2}$,

et alors ce coefficient vaut

$$(-1)^r\left(\binom{2n}{n-r}-\binom{2n}{n-r-1}\right).$$

•

d) polyominos convexes

Remarquons que le théorème 2.5 est évident pour les mots de P de longueur 2 et 4, comme le montre la figure 7 où l'on a représenté les arbres à n arêtes (n =1, 2 ou 3) dont les sommets sont carrés ou non et dont toutes les feuilles sont carrées, les mots de P de longueur 2n qui les codent, et les chemins associés.

Mais il n'est plus évident pour n≥6, pour établir le théorème dans le cas général, nous allons utiliser un autre objet en bijection avec les mots de P, les *polyominos convexes*. Cette bijection est inspirée par une bijection établie par Fédou [10] dans un cadre plus général et dont elle peut être vue comme un cas particulier. Puis nous allons définir par une simple manipulation sur les cellules, une involution sur ces polyominos qui change la parité du nombre de bandes en laissant invariant la différence entre l'aire et le nombre de ces bandes.

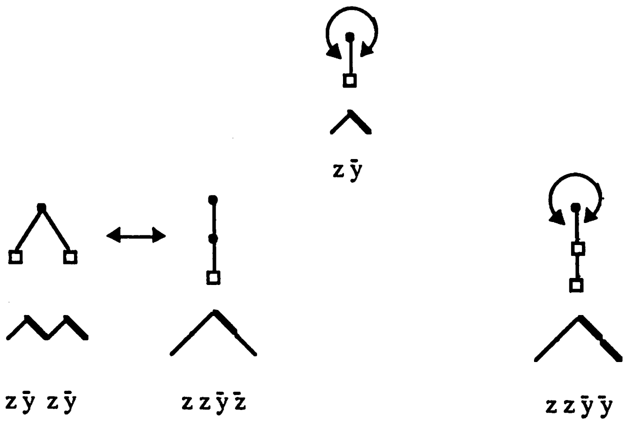


Figure 7. Les 4 premiers mots de P et les arbres associés.

Nous allons coder les mots de P de longueur 2n par deux suites d'entiers de longueur k, (a_i) et (b_i), $1 \le i \le k$, telles que k est le nombre d'occurrences de ȳ dans le mot, a_i est le nombre de z qui précèdent la $i^{ème}$ occurrence de ȳ et b_i le nombre de z̄ ou ȳ qui précèdent cette même occurrence. Plus précisément, on a la proposition,

PROPOSITION 2.6 *Les mots de* P *de longueur* 2n *sont en bijection avec deux suites d'entiers* (a_i) *et* (b_i), $1 \le i \le 2n$, *vérifiant les conditions suivantes,*

— $1 \le a_1 \le a_2 \le \ldots \le a_k \le n$,

— $0 = b_1 < b_2 < \ldots < b_k$

— $b_i < a_i$, $1 \le i \le k$.

De plus il est clair que si le mot w de P est représenté par le couple de listes (a_i) et et (b_i) de longueur k, le poids de w est $\sum_{i=1}^{n}(a_i - b_i)$, et le signe sgn(w) est $(-1)^k$.

D'autre part considérons dans le plan combinatoire $\Pi = \mathbb{Z} \times \mathbb{Z}$ les figures composées de *bandes horizontales* superposées, chaque bande étant formée de carrés unitaires consécutifs ayant en commun un côté vertical. Une bande est repérée par son *abscisse*, position du bord gauche, sa *longueur*, égale au nombre de carrés unitaires qui la composent, et son numéro ou *ordonnée*. L'ensemble forme un *polyomino horizontalement convexe*.

On peut alors définir la construction suivante, notée ψ, qui associe à tout mot de P un ensemble de bandes, ainsi déterminé,

si w est représenté par le couple $(a_i,b_i)_{1\leq i\leq k}$, $\psi(w)$ sera constitué de k +1 bandes; pour $1\leq i\leq k$, β_i a pour abscisse b_i, pour ordonnée i et pour longueur a_i-b_i+1, comme indiqué sur la figure 8 ci-contre. β_{k+1} est constituée d'une seule cellule d'abscisse a_i.

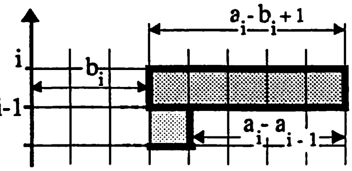


Figure 8.
Repérage de la i^ème bande.

Exemple. Le mot w=z z z ȳ ż ȳ z z ȳ z z z ȳ ż ȳ ȳ , apparu dans l'exemple de la figure 6 est représenté par les 2 suites d'entiers, $(a_i)_{1\leq i\leq 6} = (3, 3, 5, 8, 8, 8)$ d'une part, et $(b_i)_{1\leq i\leq 6} = (0, 2, 3, 4, 6, 7)$. La figure 9 montre le chemin, l'arbre et l'ensemble de bandes associé à w.

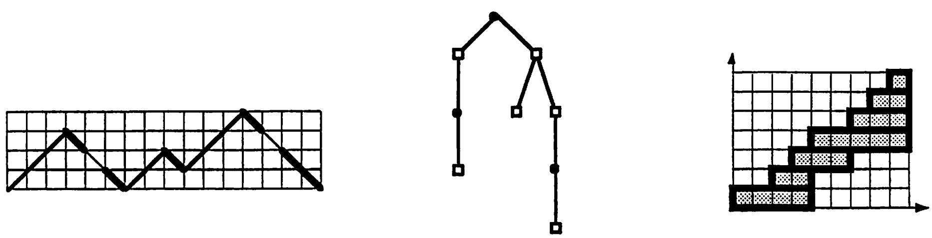


Figure 9. Un mot de P, l'arbre et le polyomino associés.

On a alors le résultat suivant,

PROPOSITION 2.7

La construction ψ *est une bijection entre les mots* w *de P de longueur* 2n *de poids* p *et ayant q lettres* ȳ *et les polyominos parallélogrammes constitués de* q+1 *bandes horizontales, et n+1 colonnes, d'aire* p+q+1 *, n'ayant pas deux bandes horizontales adjacentes de même abscisse, et dont la bande supérieure est composée d'une seule cellule.*

Preuve. La vérification, conséquence directe des propriétés générales des mots de Dyck, et de celles particulières au langage P, est laissée au lecteur. Remarquons que grâce à l'ajout d'une cellule dans chaque bande horizontale par rapport à la valeur $a_i - b_i$ qui donne la hauteur de la $i^{ème}$ occurrence de la lettre $\bar{y}$ dans le mot, on est assuré de l'abscence de points d'articulation dans le polyomino.

Exemple. le mot $z^2 \bar{y}\, \dot{z}\, \bar{y}\, z^2\, \bar{y}\, z^3\, \bar{y}\, \dot{z}\, \bar{y}\, \bar{y}$ de l'exemple précédent a 16 lettres, 6 $\bar{y}$ et un poids de 13. Le polyomino associé par ψ possède 7 bandes, 9 colonnes et 20 cellules carrées .

Les mots $z^q \bar{y}^q$, $(q\geq 1)$, points fixes de l'involution à construire du théorème 5.2, ont pour polyomino associé par ψ un polyomino parallélogramme tel que les abscisses initiales des bandes superposées consécutives diffèrent exactement d'une unité, et que les abscisses finales sont toutes égales à q+1. La figure ci-contre montre $\psi(z^6 \bar{y}^6)$ qui constitue un diagramme de Ferrers en "escalier" régulier.

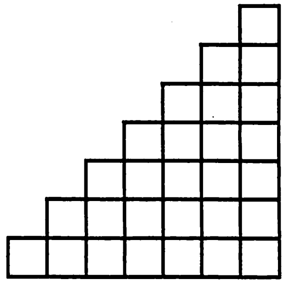


Figure 10. Un diagramme de Ferrers en escalier régulier.

Soit $Q = \psi(P)$, et $Q_0 = \psi(\{z^q \bar{y}^q,\ 1\leq q\})$. Nous allons alors définir, pour achever ce travail, une application de Q dans Q , que l'on notera κ, qui laissera invariants les polyominos de Q_0 et échangera 2 à 2 ceux de $Q \backslash Q_0$ en changeant la parité du nombre de bandes.

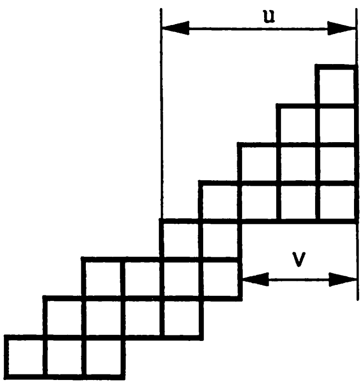


Figure 11. Un polyomino parallélogramme et ses deux paramètres u et v.

Pour cela introduisons deux paramètres u et v sur tout élément π de Q. Si π est un polyomino composé de q+1 bandes d'abscisse b_i $(1\leq i\leq q+1)$ et n+1 colonnes, on pose

— ι) $u = n+1- b_k$ s'il existe un entier k $(1<k\leq q)$ tel que $b_k > b_{k-1}+1$, sinon $u = n+1$;

— $\iota\iota$) $v = a_{\ell+1} - a_l$ où ℓ est le plus grand entier $(1\leq \ell<q)$ s'il existe tel que $a_{\ell+1} = n$ et $a_\ell < n$, sinon $v = n+1$.

Ces paramètres sont illustrés sur la figure 11 ci-contre qui montre un élément de l'ensemble Q_0. Clairement, Q_0 est l'ensemble des polyominos π de Q tels que $u(\pi) = v(\pi) = n+1$ (où n est le nombre de colonne). Posons

$$Q^+ = \{\pi \,/\, \pi \in Q \backslash Q_0 \,,\, u(\pi) \leq v(\pi)\ \},$$

$$Q^- = \{\pi \,/\, \pi \in Q \backslash Q_0 \,,\, u(\pi) > v(\pi)\ \}.$$

Nous sommes alors en mesure de définir κ ,

a) $\underline{\pi \in Q^+}$. Soit k tel que $b_k = n+1-u(\pi)$, et ℓ tel que $a_\ell - a_{\ell-1} = v(\pi)$. Le fait que π soit connexe et sans point d'articulation implique que $k>\ell$. Informellement, pour construire π' à partir de π, on enlève les u cellules les plus à droite de la bande ℓ et on ajoute une cellule à droite de chacune des bandes k,k+1, ...,q+1 puis on couvre le tout par une $(q+2)^{\text{ième}}$ bande d'une seule cellule.

b) $\underline{\pi \in Q^-}$. On comble la bande ℓ par v cellules carrées et l'on enlève sur les bandes supérieures v+1 cellules, une par bande.
Ces opérations sont décrites sur la figure 12 où le signe - marque les cellules à enlever et le signe + les cellules à ajouter.

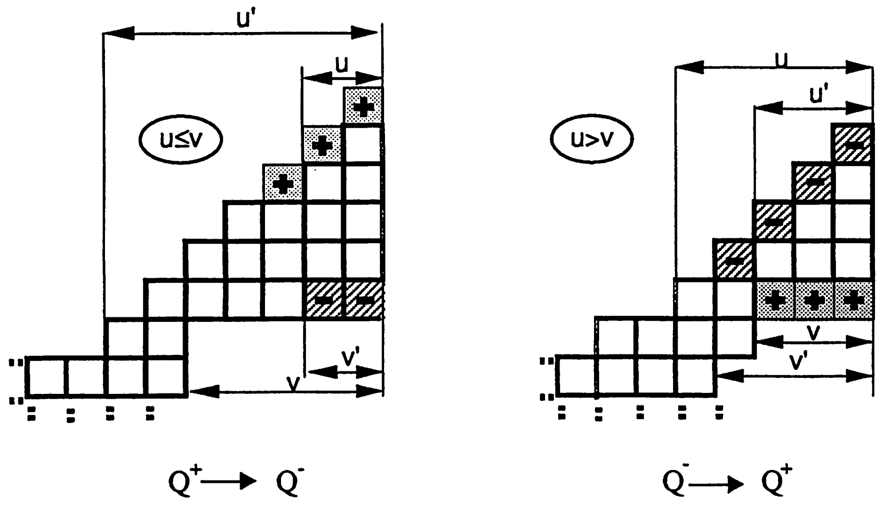


Figure 12. Deux exemples de l'involution κ.

Définissons le poids d'un polyomino π comme la différence entre le nombre de cellules et le nombre de rangées; on établit alors le résultat final, qui achève la preuve du théorème 2.5,

PROPOSITION 2.8 *L'application* κ *est une involution qui laisse* Q_0 *invariant et telle que l'image d'un polyomino* π *de* Q^+ *est un polyomino de* Q^- *et inversement. De plus* κ *conserve le nombre de colonnes, le poids et change la parité du nombre de rangées.*

La figure 12 ci-dessus illustre les deux cas de la définition de κ, et la figure 13 montre le polyomino, le chemin, le mot et l'arbre en correspondance avec le polyomino, le chemin et le mot de l'exemple de la figure 9.

REMARQUE 2.9 Il y a équivalence entre les polyominos parallélogramme et les "Ferrers gauches" définis par un couple de partition d'entiers λ et μ telles que $\mu \subset \lambda$. Cette bijection agissant sur les polyominos parrallélogrammes a des analogies avec celle (plus simple) de Franklin définie sur les diagrammes de Ferrers et démontrant le théorème pentagonal d'Euler (voir le livre d'Andrews [1] p 10).

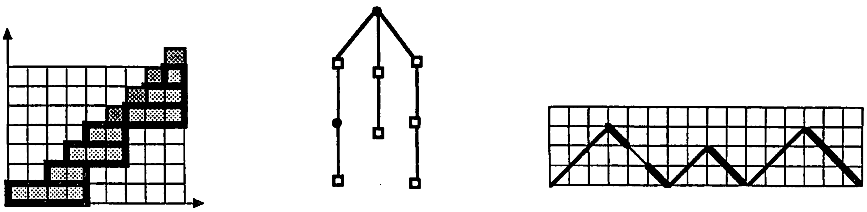


Figure 13. Le polyomino, l'arbre et le mot de P en bijection avec ceux de la figure 9.

BIBLIOGRAPHIE

1. G.E. Andrews, *The Theory of partitions,* Encyclopedia of Mathematics and its Applications, vol n°2, ed Gian-Carlo Rota, Addison-Wesley, Reading, Massachussets, 1976.
2. R. Askey et J. Wilson, Some basic hypergeometric orthogonal polynomial that generalize Jacobi polynomials, Memoirs Amer. Math. Soc. 319 (1985).
3. J. Berstel et C. Reutenauer, *Les séries rationnelles et leurs langages,* Masson, Paris, 1984.
4. L. Chottin et R. Cori, Une preuve combinatoire de la rationnalité d'une série génératrice associée aux arbres, RAIRO Inf. Th.,16 (1982), 113-128.
5. M.P. Delest, *Utilisation des langages algébriques et du calcul formel pour le codage et l'énumération des polyominos*, Thèse d'Etat, Université de Bordeaux I, 1987.
6. M.P. Delest, S. Dulucq, Enumeration of directed column-convex animals with given perimeter and area, rapport LaBRI n° 86-15, Université de Bordeaux I.
7. M.P.Delest, G.X.Viennot, Algebraic langages and polyominoes enumeration, Theor. Comp.Sci. 34 (1984), 169-206 .
8. S. Dulucq, J.G.Penaud, Cordes, Arbres et Permutations, à paraître dans Discrete Math.
9. A. Errera, Un problème d'énumération, Mém. Acad. Royale Belgique ,8°,2, v.11,1931.
10. J.M.Fédou, *Grammaires et q-générations de polyominos,* Thése, Université de Bordeaux I, 1989.
11. P. Flajolet, Combinatorial aspects of continued fractions, Discrete Math., 32 (1980) 125-161.

12. P. Flajolet, J. Françon et J. Vuillemin, Sequence of operations analysis for Dynamic data structures, J. of Algorithm 1 (1980), 111-141.

13. J. Françon, Histoire de Fichiers, RAIRO Inf. Th., 12 (1978), 49-62.

14. F. Franklin, Sur le développement du produit infini $(1-x)(1-x^2)(1-x^3)$..., C.R.A.S 82, (1881), 448-450.

15. J. Françon et X.G. Viennot, Permutations selon les pics, creux, doubles montées, doubles descentes, nombres d'Euler et nombres de Genocchi, Discrete Math. 28 (1979), 21-35.

16. M. E. H. Ismail, D. Stanton et X. G. Viennot, The Combinatorics of q-Hermite polynomials and the Askey-Wilson Integral, Europ. J. Combinatorics (1987) 8, 379-392.

17. D. Knuth, *The art of computer programming*, vol 3, Addison-Wesley, Reading, 1973.

18. M. Lothaire, *Combinatorics on Words,* Encyclopedia of Mathematics and its Applications, vol 17, ed Gian-Carlo Rota , Addison-Wesley, Reading, Massachussets, 1983.

19. R. C. Read, The Chord Intersection Problem,Second International Conf. on Combinatorial Mathematics, Ann. New-York Acad. Sc., vol 319 (1979), 444-454.

20. J. Riordan, The distribution of Crossings of Chords Joining Pairs of 2n Points on a Circle, Mathematics of Computation, v.29, n° 129, (1975), 215-222.

21. M.P.Schützenberger, Context-free langages and pushdown automata, Information and Control, 6 (1963), 246-264.

22. J.Touchard, Sur un problème de configurations, C. R. Acad. Sc. Paris, Vol. 230, (1950), 1997-1998.

23. J.Touchard, Contributions à l'étude du problème des timbres poste, Canad. J. Math.,V. 2, (1950), 385-398.

24. J. Touchard, Sur un problème de configurations et sur les fractions continues, Canad. J. Math, V. 4, (1952), 2-25.

25. X.G. Viennot, *Une théorie combinatoire des polynômes orthogonaux généraux*, notes de cours , Université du Québec à Montréal, Montréal 1983.

26. X.G. Viennot, Enumerative combinatorics and algebraic languages, in *Fundamantal of Computation Theory* , Proc. FCT'85, Lectures Notes in Comp. Sci. n° 199, ed. L. Budach, Springer-Verlag, 1985.

Weyl Group Symmetric Functions and the Representation Theory of Lie Algebras

Arun Ram
Department of Mathematics
Massachusetts Institute of Technology
Cambridge, MA 02139
April 1, 1992

0. Introduction

In view of the many applications of the theory of symmetric functions to representation theory it seems desirable to have a theory of symmetric functions in the spirit of Macdonald's book [Mac] for Weyl groups other than the symmetric group. In spite of the feeling that finding suitable generalizations may seem to be an intractable problem I remain optimistic and want to study symmetric functions from the point of view of representation theory of Lie algebras whenever possible. We must keep an open mind and we must not expect our generalizations to have *all* the magical properties that abound in the classical case. The idea is to find an algebraic structure which motivates each statement in the classical symmetric function theory. If this algebraic notion occurs across the board then this should indicate what the proper generalization is for other types. Note that from this point of view there may be several useful generalizations of a given concept depending on what symmetric function are desirable.

The goal of this paper is to offer a suggestion for the analogue of the basis of homogeneous symmetric functions for Weyl group symmetric functions. In this case the definition is motivated by the theory of centralizer algebras. The idea motivating the generalization is that it is really the Frobenius image of the homogeneous symmetric function that is the useful object. It is clear from the double centralizer theory that an analogue of the Frobenius characteristic map is a feature of the double centralizer mechanism, see [R]. With this point of view one finds an analogue of the "Jacobi-Trudi" formula in the work of Verma [V], Zelevinsky [Z] and Goodman-Wallach [GW]. In this paper I simply offer a mechanism by which to transfer their results to Weyl group symmetric functions.

I would like to thank D.-N. Verma for so patiently explaining the many many things about representations of Lie algebras and their representations which he considered necessary for me to know. In particular, he showed me the representation theoretic result, Theorem (4.6) in this paper, which motivates the definition of the homogeneous symmetric functions. I would like to thank N. Wallach for further useful discussions with me on this topic. I would like to thank the Tata Institute of Fundamental Research for their hospitality during my visit.

I have tried to organize this paper to motivate the concepts of symmetric functions by facts from representation theory. My hope is that this may serve as an introduction to representation theory for algebraic combinatorists who do not already know the subject.

I begin with a brief resume of the classical symmetric function theory. Then in section 2 I try to copy this theory except in the context of a general Weyl group. The remaining sections are an attempt to explain the representation theoretic motivations behind these generalizations.

1. Classical symmetric functions

This section gives a brief summary of the classical symmetric function theory. See [Mac] Chapter 1 for a complete treatment.

Fix a positive integer n. A *partition* $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n)$ is a sequence $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$ of nonnegative integers. Let $\mathcal{P}$ denote the set of partitions. We have the following sequence of inclusions

$$\mathcal{P} \subset \mathbb{Z}^n \subset \mathbb{C}^n \,. \tag{1.1}$$

There is a partial ordering, the *dominance ordering*, on elements of $\mathbb{Z}^n$ given by

$$\gamma \geq \kappa \quad \text{if} \quad \gamma_1 + \gamma_2 + \cdots + \gamma_i \geq \kappa_1 + \kappa_2 + \cdots \kappa_i, \text{ for all } i. \tag{1.2}$$

Let S_n denote the *symmetric group.* The *sign* $\varepsilon(w)$ of a permutation $w \in S_n$ is the determinant of the corresponding permutation matrix. S_n acts on elements of $\mathbb{Z}^n$ by permuting the positions.

For each $1 \leq i < j \leq n$ the *raising operator* R_{ij} is the operator which acts on elements of $\mathbb{Z}^n$ by

$$R_{ij}(\gamma_1, \gamma_2, \ldots, \gamma_n) = (\gamma_1, \gamma_2, \ldots, \gamma_i + 1, \ldots, \gamma_j - 1, \ldots, \gamma_n). \tag{1.4}$$

Let $x_1, x_2, \ldots, x_n$ be commuting variables and for each $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_n) \in \mathbb{Z}^n$ define

$$x^\gamma = x_1^{\gamma_1} x_2^{\gamma_2} \cdots x_n^{\gamma_n}. \tag{1.5}$$

Define an action of S_n on monomials by

$$w x^\gamma = x^{w\gamma}. \tag{1.6}$$

The ring

$$\Lambda_n = \mathbb{Q}\,[x_1, x_2, \ldots, x_n]^{S_n}$$

is the ring of symmetric functions.

Bases of symmetric functions

For each partition λ define the *monomial symmetric function* by

$$m_\lambda = \sum_{\gamma \in S_n \lambda} x^\gamma, \tag{1.7}$$

where the sum runs over all $\gamma \in \mathbb{Z}^n$ in the S_n orbit of λ, i.e., over all distinct permutations of λ.

The *Schur functions* are given by

$$s_\lambda = \frac{\sum_{w\in S_n} \varepsilon(w)x^{w(\lambda+\rho)}}{\sum_{w\in S_n} \varepsilon(w)x^{w\rho}}, \tag{1.8}$$

where $\rho = (n-1, n-2, \ldots, 1, 0)$.

The *elementary symmetric functions* are given by defining

$$\begin{aligned} e_0 &= 1, \\ e_r &= \sum_{1\le i_1<\cdots<i_r\le n} x_{i_1}x_{i_2}\cdots x_{i_r}, \end{aligned}$$

for each positive integer r, and

$$e_\lambda = e_{\lambda_1}e_{\lambda_2}\cdots e_{\lambda_n}, \tag{1.9}$$

for all partitions λ.

The *homogeneous symmetric functions* are given by defining

$$\begin{aligned} h_0 &= 1, \\ h_r &= \sum_{1\le i_1\le\cdots\le i_r\le n} x_{i_1}x_{i_2}\cdots x_{i_r}, \end{aligned}$$

for each positive integer r, and

$$h_\gamma = h_{\gamma_1}h_{\gamma_2}\cdots h_{\gamma_n},$$

for all sequences $\gamma \in \mathbb{Z}^n$.

Define integers $K_{\lambda\mu}$ by

$$s_\lambda = \sum_{\mu\in\mathcal{P}} K_{\lambda\mu}m_\mu. \tag{1.10}$$

One has the following (nontrivial) facts:

(a) The $K_{\lambda\mu}$ are nonnegative integers.

(b) $K_{\lambda\lambda} = 1$ for all $\lambda \in \mathcal{P}$.

(c) $K_{\lambda\mu} = 0$ if $\mu \not\leq \lambda$.

Each of the sets $\{m_\lambda\}_{\lambda\in\mathcal{P}}$, $\{s_\lambda\}_{\lambda\in\mathcal{P}}$, $\{e_\lambda\}_{\lambda\in\mathcal{P}}$, $\{h_\lambda\}_{\lambda\in\mathcal{P}}$, forms a $\mathbb{Z}$-basis of Λ_n.

Inner product

There is an inner product on the ring of symmetric functions given by making the Schur functions orthonormal,

$$\langle s_\lambda, s_\mu\rangle = \delta_{\lambda\mu}. \tag{1.11}$$

Further facts

The homogeneous symmetric functions are the dual basis to the basis of monomial symmetric functions,

$$\langle h_\lambda, m_\mu \rangle = \delta_{\lambda\mu}. \tag{1.12}$$

A consequence of this is that

$$h_\mu = \sum_\lambda s_\lambda K_{\lambda\mu}. \tag{1.13}$$

One has the following "Jacobi-Trudi" formula for the Schur functions in terms of the homogeneous symmetric functions,

$$s_\lambda = \sum_{w \in S_n} \varepsilon(w) h_{\lambda+\rho-w\rho}. \tag{1.14}$$

There is also a formula for the Schur function in terms of raising operators and the homogeneous symmetric function.

$$s_\lambda = \prod_{i<j}(1 - R_{ij}) h_\lambda. \tag{1.15}$$

2. Weyl group symmetric functions

Each of the simple root systems is determined by a *Cartan matrix* C. A list of the Cartan matrices for simple root systems can be found in [Bou] p. 250-258. We shall denote the (i,j) entry of the Cartan matrix by $\langle \alpha_i, \alpha_j \rangle$ so that

$$C = (\langle \alpha_i, \alpha_j \rangle).$$

Let n be the dimension of the Cartan matrix.

Let $\omega_1, \omega_2, \ldots, \omega_n$ be basis vectors in a vector space. Define

$$\mathfrak{H}^* = \sum_i \mathbb{C}\, \omega_i, \quad P = \sum_i \mathbb{Z}\, \omega_i, \quad P^+ = \sum_i \mathbb{N}\, \omega_i,$$

where $\mathbb{N}$ denotes the nonnegative integers. The elements of $\mathfrak{H}^*, P$, and P^+ are called the *weights*, the *integral weights*, and the *dominant integral weights*, respectively. The ω_i are called the *fundamental weights*. We have the following sequence of inclusions

$$P^+ \subseteq P \subseteq \mathfrak{H}^*. \tag{2.1}$$

Let $\gamma = \sum_i \gamma_i \omega_i$ be an element of P. We shall use the notation $\langle \gamma, \alpha_i \rangle$ for the integer γ_i. The *simple roots* α_i are given in terms of the entries of the Cartan matrix,

$$\alpha_i = \sum_j \langle \alpha_i, \alpha_j \rangle \alpha_j.$$

There is a partial ordering on the weight lattice given by

$$\gamma \geq \kappa \quad \text{if} \quad \kappa = \gamma - \sum k_i \alpha_i, \tag{2.2}$$

for nonnegative integers k_i. We say that $\gamma \geq \kappa$ in *dominance*.

Define linear operators $s_i: P \to P$ by

$$s_i \gamma = \gamma - \langle \gamma, \alpha_i \rangle \alpha_i.$$

The *Weyl group* is the group generated by the s_i: $W = < s_1, s_2, \ldots, s_n >$. The *sign* of an element $w \in W$ is $\varepsilon(w) = (-1)^p$, where p is the smallest nonnegative integer such that there exists an expression $s_{i_1} s_{i_2} \cdots s_{i_p} = w$. We will need the following proposition, see [Bou] Ch. 6, §1 Thm. 2.

(2.3) Proposition.

(a) Every Weyl group orbit $W\gamma$, $\gamma \in P^+$ contains a unique element in P^+.

(b) If $\lambda, \mu \in P^+$ and $\rho = \sum_i \omega_i$ then, for $v, w \in W$,

$$w(\lambda + \rho) = v(\mu + \rho) \iff v = w.$$

$\alpha \in P$ is a root if $\alpha = w\alpha_i$ for some $w \in W$ and simple root α_i. Let Φ be the set of roots and let $\Phi^+ = \{\alpha \in \Phi | \alpha > 0\}$ and $\Phi^- = \{\alpha \in \Phi | \alpha < 0\}$ where the ordering is as in (2.2). It is true that $\Phi = \Phi^+ \cup \Phi^-$. The elements of Φ^+ and Φ^- are called positive and negative roots respectively. The *raising operator* R_α associated to a positive root α is the operator which acts on elements of P by

$$R_\alpha \gamma = \gamma + \alpha. \tag{2.4}$$

Corresponding to each $\lambda \in P$ we write, formally, e^λ so that

$$e^\lambda e^\mu = e^{\lambda + \mu}.$$

In particular if $\lambda = \sum_i \lambda_i \omega_i$ then

$$\begin{aligned} e^\lambda &= e^{\lambda_1 \omega_1} e^{\lambda_2 \omega_2} \cdots e^{\lambda_n \omega_n} \\ &= (e^{\omega_1})^{\lambda_1} (e^{\omega_2})^{\lambda_2} \cdots (e^{\omega_n})^{\lambda_n}. \end{aligned} \tag{2.5}$$

(If one finds this "exponential" notation unsettling one can substitute z_i for e^{ω_i} and write $z^\lambda = z_1^{\lambda_1} z_2^{\lambda_2} \cdots z_n^{\lambda_n}$ instead of e^λ.) Define an action of the Weyl group by

$$we^\lambda = e^{w\lambda}, \tag{2.6}$$

for each $w \in W$ and $\lambda \in P$. Define

$$A^W = \mathbb{Z}\, [e^{\omega_1}, e^{-\omega_1}, \cdots, e^{\omega_n}, e^{-\omega_n}]^W.$$

Bases of A^W

For each $\lambda \in P^+$ define the *orbit sum*, or *monomial symmetric function*, by

$$m_\lambda = \sum_{\nu \in W\lambda} e^\nu. \tag{2.7}$$

For each $\lambda \in P^+$ define the *Weyl character* by

$$\chi^\lambda = \frac{\sum_{w\in W} \varepsilon(w) x^{w(\lambda+\rho)}}{\sum_{w\in W} \varepsilon(w) x^{w\rho}}, \tag{2.8}$$

where $\rho = \sum_i \omega_i$.

The *elementary, or fundamental, symmetric functions* are given by defining

$$e_0 = 1,$$
$$e_r = \chi^{\omega_r},$$

for each positive integer r, and

$$e_\lambda = e_1^{\lambda_1} e_2^{\lambda_2} \cdots e_n^{\lambda_n}, \tag{2.9}$$

for all elements $\lambda = \sum_i \lambda_i \omega_i$ in P^+.

Define integers $K_{\lambda\mu}$ by the identity

$$\chi^\lambda = \sum_{\mu \in P^+} K_{\lambda\mu} m_\mu. \tag{2.10}$$

It is true that
(a) The $K_{\lambda\mu}$ are nonnegative integers.
(b) $K_{\lambda\lambda} = 1$ for all $\lambda \in P^+$.
(c) $K_{\lambda\mu} = 0$ if $\mu \not\leq \lambda$.
All of these facts follow from representation theory see §3 (3.5). I know of no easy way to prove these results without using representation theory.

Each of the sets

$$\{m_\lambda\}_{\lambda \in P^+},$$
$$\{\chi_\lambda\}_{\lambda \in P^+},$$
$$\{e_\lambda\}_{\lambda \in P^+},$$

forms a $\mathbb{Z}$-basis of A^W. The fact that the m_λ form a $\mathbb{Z}$ basis of A^W follows immediately from (2.3). One can show by elementary techniques and without using representation theory, see [Bou] Ch. 6, §3, that the χ^λ, $\lambda \in P^+$ form a $\mathbb{Z}$ basis of A^W. This fact also follows from the facts about the numbers $K_{\lambda\mu}$ above. Assuming that the χ^λ form a $\mathbb{Z}$-basis of A^W it follows that the e_λ, $\lambda \in P^+$ form a $\mathbb{Z}$-basis of A^W simply by expanding e_λ in terms of e^μ, $\mu \in P^+$. One can also obtain this result in a different fashion by using representation theory.

Inner product

Let

$$d=\sum_{w\in W}\varepsilon(w)e^{w\rho},$$

where $\rho=\sum_i\omega_i$. If $f=\sum_{\nu\in P}f_\nu e^\nu$ then define $\bar{f}=\sum_\nu f_\nu e^{-\nu}$. Let $[f]_1$ denote taking the coefficient of the identity, e^0, in f. Then define

$$\langle f,g\rangle=\frac{1}{|W|}[fd\bar{g}\bar{d}]_1,$$

where $|W|$ is the order of the Weyl group.

(2.11) Proposition. ([Mac2]) *The inner product defined above satisfies*

$$\langle\chi^\lambda,\chi^\mu\rangle=\delta_{\lambda\mu}.$$

Proof. Since $\chi^\lambda=d^{-1}\sum_{w\in W}\varepsilon(w)e^{w(\lambda+\rho)}$,

$$\langle\chi^\lambda,\chi^\mu\rangle=\frac{1}{|W|}\sum_{v,w\in W}\varepsilon(v)\varepsilon(w)[e^{v(\lambda+\rho)}e^{-w(\mu+\rho)}]_1.$$

This is zero if $\lambda\neq\mu$, because, (2.3), the orbits $W(\lambda+\rho)$ and $W(\mu+\rho)$ do not intersect. If $\lambda=\mu$, then $v(\lambda+\rho)=w(\lambda+\rho)\Leftrightarrow v=w$. Thus $\langle\chi^\lambda,\chi^\lambda\rangle=\frac{1}{|W|}\sum_{w\in W}1=1$. □

If one prefers one may simply *define* the inner product by making the Weyl characters orthonormal.

Homogeneous symmetric functions

Let $\kappa\in P^+$ and define

$$\Gamma_\kappa=\{\mu\in P^+|\mu\leq\kappa\},\text{ and }\Lambda_\kappa=\text{ span}\{\chi^\mu|\mu\in\Gamma_\kappa\}.$$

Since Γ_κ is always finite Λ_κ is always finite dimensional.

Define an inner product on Λ_κ by defining

$$\langle\chi^\lambda,\chi^\mu\rangle=\delta_{\lambda\mu}$$

for all $\lambda,\mu\in\Gamma_\kappa$. Then define the homogeneous symmetric functions h_λ, $\lambda\in\Gamma_\kappa$ to be the dual basis to the monomial symmetric functions,

$$\langle h_\lambda,m_\mu\rangle=\delta_{\lambda\mu}. \tag{2.12}$$

Using the integers $K_{\lambda\mu}$ defined in (2.10), the homogeneous symmetric functions are given in terms of the Weyl characters by

$$h_\mu=\sum_{\lambda\in\Gamma_\kappa}\chi^\lambda K_{\lambda\mu}, \tag{2.13}$$

for all $\mu \in \Gamma_\kappa$. The h_μ, $\mu \in \Gamma_\kappa$ form a basis of Λ_κ.

Each of the sets

$$\{m_\lambda\}_{\lambda \in P^+},$$
$$\{\chi_\lambda\}_{\lambda \in P^+},$$
$$\{e_\lambda\}_{\lambda \in P^+},$$
$$\{h_\lambda\}_{\lambda \in P^+},$$

forms a $\mathbb{Z}$-basis of Λ_κ. To see this choose some total ordering of the elements of Γ_κ which is a refinement of the dominance partial order. Then, by (2.10a-c), the matrix, with rows and columns indexed by elements of Γ_κ, having $K_{\lambda\nu}$ as the λ, ν entry is upper unitriangular with nonnegative integer entries. This implies that it is invertible as a matrix with integer entries. The fact that the χ^μ are a basis of Λ_κ is by definition. The other two statements now follow from (2.10) and (2.13).

"Jacobi-Trudi" formulas

Fix $\kappa \in P^+$. Define

$$h_{w\lambda} = h_\lambda$$

for all $\lambda \in \Gamma_\kappa$ and all $w \in W$ so that h_λ is defined for all $\lambda \in W\Gamma_\kappa$. One has the following "Jacobi-Trudi" type identity for the Weyl characters in terms of the h_λ.

(2.14) Theorem. *Let $\rho = \sum_i \omega_i$. Then for each $\lambda \in \Gamma_\kappa$*

$$\chi^\lambda = \sum_{w \in W} \varepsilon(w) h_{\lambda+\rho-w\rho}.$$

Proof. We show that elements $\sum_{w \in W} \varepsilon(w) h_{\lambda+\rho-w\rho}$ are the dual basis to the basis χ^μ, $\mu \in \Gamma_\kappa$.

$$\chi^\lambda = \sum_{\mu \in W\Gamma_\kappa} \langle \chi^\lambda, h_\mu \rangle e^\mu.$$

Expanding χ^λ by (2.8) and clearing denominators we have that

$$\begin{aligned} \sum_{w \in W} \varepsilon(w) e^{w(\lambda+\rho)-\rho} &= (\sum_{w \in W} \varepsilon(w) e^{w\rho-\rho})(\sum_{\mu \in W\Gamma_\kappa} \langle \chi^\lambda, h_\mu \rangle e^\mu) \\ &= \sum_{\mu \in W\Gamma_\kappa} \sum_{w \in W} \varepsilon(w) \langle \chi^\lambda, h_\mu \rangle e^{\mu+w\rho-\rho} \end{aligned}$$

Substitute $\gamma = \mu + w\rho - \rho$ to get

$$\sum_{w \in W} \varepsilon(w) e^{w(\lambda+\rho)-\rho} = \sum_{\mu \in W\Gamma_\kappa} \langle \chi^\lambda, \sum_{w \in W} \varepsilon(w) h_{\gamma+\rho-w\rho} \rangle e^\gamma.$$

Compare coefficients of e^γ for $\gamma \in P^+$ on each side of this equation. Since $\lambda \in P^+$ we know by (2.3) that $w(\lambda+\rho)-\rho$ is not an element of P^+ for any $w \in W$ except the identity. Thus we know that if $\mu \in P^+$ then

$$\langle \chi^\lambda, \sum_{w \in W} \varepsilon(w) h_{\mu+\rho-w\rho} \rangle = \delta_{\lambda\mu}. \qquad \square$$

Recall that raising operators act on elements of P. We allow the raising operators to act upon the h_λ by defining

$$R(h_\lambda) = h_{R(\lambda)},$$

for each sequence $R = R_{\beta_1} R_{\beta_2} \cdots R_{\beta_k}$. (Note: It is important to keep in mind that raising operators act on elements of P and not on symmetric functions.)

(2.15) Corollary. *For all* $\lambda \in \Gamma_\kappa$,

$$\chi^\lambda = \prod_{\alpha>0} (1 - R_\alpha) h_\lambda.$$

Proof. A sketch of the proof is as follows. Evaluating the right hand side of the above we get

$$\prod_{\alpha>0} (1 - R_\alpha) h_{\lambda+\rho-\rho} = \sum_{E \subseteq \Phi^+} (-1)^{|E|} h_{\lambda+\rho+(-\rho+\sigma_E)},$$

where $\sigma_E = \sum_{\alpha \in E} \alpha$. An element $\gamma = \sum_i \gamma_i \omega_i$ in P^+ is called regular if $\gamma_i > 0$ for all i. The sets $\{-\rho + \sigma_E | E \subseteq \Phi^+, \rho + \sigma_E \text{ regular}\}$ and $\{-w\rho | w \in W\}$ are equal. This is proved by expressing ρ in the form $\rho = \sum_{\alpha \in \Phi^+} \alpha$ and using [Bou] Ch. 6, §1 Cor. 2. Under this bijection $(-1)^{|E|} = \varepsilon(w)$. The terms arising from the subsets E for which $-\rho + \sigma_E$ is not regular cancel with each other. This can be shown by showing that $\prod_{\alpha>0}(1 - R_\alpha)(-\rho)$ is skew-symmetric with respect to W and that if $\gamma \in P^+$ is not regular then $\sum_{w \in W} \varepsilon(w) w\gamma = 0$. These arguments show that

$$\prod_{\alpha>0} (1 - R_\alpha) h_\lambda = \sum_{w \in W} \varepsilon(w) h_{\lambda+\rho-w\rho}. \qquad \square$$

The proof of Cor. (2.15) was motivated by the proof of the Weyl denominator formula given in [Mac2].

Direct limits

The above definition defines an analogue of homogeneous symmetric functions for the spaces Λ_κ. One would like to say that in some sense the h_λ are well defined on all of A^W. With this in mind we introduce the following.

For each pair $\beta, \kappa \in P^+$ such that $\beta \leq \kappa$ define a linear map $f_{\beta\kappa}: \Lambda_\kappa \to \Lambda_\beta$ by

$$s_\lambda \mapsto \begin{cases} s_\lambda, & \text{if } \lambda \leq \beta; \\ 0, & \text{if } \lambda \not\leq \beta. \end{cases}$$

It is clear that

1) If $\beta \leq \gamma \leq \kappa$ then $f_{\beta\kappa} = f_{\beta\gamma} \circ f_{\gamma\kappa}$,

2) For each $\beta \in P^+$, $f_{\beta\beta}$ is the identity on Λ_β.

Thus $(\Lambda_\beta, f_{\beta\gamma})$ form an inverse system of vector spaces, see Bourbaki Theory of Sets I §7, and Bourbaki Algebra I §10. Define

$$\Lambda = \varprojlim(\Lambda_\beta, f_{\beta\gamma}).$$

Then the homogeneous symmetric function h_λ is a well defined element of Λ for all $\lambda \in P^+$ and is equal to

$$h_\mu = \sum_{\lambda \in P^+} s_\lambda K_{\lambda\mu}.$$

An alternate option is to view the homogeneous symmetric function as an element in the direct product of vector spaces

$$\prod_{\lambda \in P^+} \mathbb{Z}\, \chi^\lambda.$$

Depending on what one would like to compute this can create problems with infinite sums. The direct limit approach allows one to control these problems by fixing an ordering on infinite sums.

3. Representation theory

Fix a Cartan matrix $C = (\langle \alpha_i, \alpha_j \rangle)$ and define $\mathfrak{U}$ to be the associative algebra (over $\mathbb{C}$) with 1 generated by $x_i, y_i, h_i, 1 \le i \le n$ with relations (Serre relations)

$$[h_i, h_j] = 0 \qquad (1 \le i, j \le n), \tag{S1}$$

$$[x_i, y_i] = h_i, \qquad [x_i, y_j] = 0 \text{ if } i \neq j, \tag{S2}$$

$$[h_i, x_j] = \langle \alpha_j, \alpha_i \rangle x_j, \qquad [h_i, y_j] = -\langle \alpha_j, \alpha_i \rangle y_j, \tag{S3}$$

$$(\text{ad } x_i)^{-\langle \alpha_j, \alpha_i \rangle + 1}(x_j) = 0 \qquad (i \neq j), \tag{S_{ij}^+}$$

$$(\text{ad } y_i)^{-\langle \alpha_j, \alpha_i \rangle + 1}(y_j) = 0 \qquad (i \neq j). \tag{S_{ij}^-}$$

Here $[a, b] = ab - ba$ and $(\text{ad } a)^k(b) = [a, [a, [a, \cdots, [a, b]] \cdots]$.

Let $\mathfrak{h}^* = \sum_i \mathbb{C}\, \omega_i$. Let V be a $\mathfrak{U}$ module. A vector $v \in V$ is called a weight vector if, for each i,

$$h_i v = \lambda_i v,$$

for some constant $\lambda_i \in \mathbb{C}$. We associate to v the weight $\lambda = \sum_i \lambda_i \omega_i$.

If v is a weight vector of weight γ then $x_i v$ is a weight vector of weight $\gamma + \alpha_i$. This is the motivation for the definition of the raising operators R_α. This also gives the motivation for the definition of the dominance partial order as $\lambda \le \mu$ if and only if $\mu = R\lambda$ for some sequence of raising operators $R = R_{\beta_1} R_{\beta_2} \cdots R_{\beta_k}$.

The dominant weights appear as a result of the following fact.

(3.1) Theorem. (see [Hu]) *There is a unique finite dimensional irreducible representation V^λ of $\mathfrak{U}$ corresponding to each dominant weight $\lambda \in P^+$. This irreducible representation is characterized by the fact that it contains a unique vector, up to scalar multiples, which is a weight vector of weight λ.*

(3.2) Theorem. (Weyl) *Every finite dimensional representation V of $\mathfrak{U}$ (corresponding to a simple Cartan matrix) is completely decomposable as a direct sum of irreducible representations V^λ, $\lambda \in P^+$.*

The motivation for Weyl group symmetric functions is that they are the generating functions of the weights in these representations. Let V be a finite dimensional $\mathfrak{U}$-module. Let B be a basis of $\mathfrak{U}$ such that each element of B is a weight vector. Then the character of V is the generating function of the weights of B,

$$\chi_V = \sum_{b \in B} e^{wt(b)}.$$

The irreducibles V^λ, combinatorially.

In this section we shall define the vector spaces V^λ affording the irreducible $\mathfrak{U}$ representations in a combinatorial fashion to motivate the Weyl group, the Weyl characters and the monomial symmetric functions.

Recall the notations $[a,b] = ab - ba$ and $(\text{ad } a)^k(b) = [a,[a,[a,\cdots,[a,b]]\cdots]$. Let $y_1, y_2, \cdots, y_n$ be letters and suppose that they satisfy the relations

$$(\text{ad } y_i)^{-\langle \alpha_j, \alpha_i \rangle + 1}(y_j) = 0, \qquad (i \neq j).$$

(We should not be surprised by the appearance of the values $\langle \alpha_i, \alpha_j \rangle$ as it is clear that our basic object, the Cartan matrix, must come into the picture in a crucial way.) We shall study words in the letters y_i. Let v^+ be a dummy letter marking the end of a word. Fix a dominant integral weight $\lambda \in P^+$. Define V^λ to be the vector space which is a linear span of words in the y_i ending with v^+,

$$V^\lambda = \text{span}\{y_{i_1} y_{i_2} \cdots y_{i_k} v^+\},$$

with the *additional relations*

$$y_j^{\lambda_j + 1} v^+ = 0.$$

Given that the vectors of the form $y_{i_1} \cdots y_{i_k} v^+$ span V^λ it is possible to choose a basis B of V^λ of vectors of this form. Define the weight of a word $v = y_{i_1} \cdots y_{i_k} v^+$ to be

$$wt(v) = \lambda - \sum_{j=1}^{k} \alpha_{i_j},$$

where the α_{i_j} are simple roots. Define the Weyl character χ^λ to be the generating function of the weights of the basis B,

$$\chi^\lambda = \sum_{b \in B} e^{wt(b)}. \tag{3.3}$$

(This is analogous to the combinatorial definition of the Schur functions in terms of column strict tableaux.) The definition of the Weyl characters is motivated by the following theorem.

(3.4) Theorem. (Weyl character formula) *For each* $\lambda \in P^+$,

$$\chi^\lambda = \frac{\sum_{w\in W} \varepsilon(w) e^{w(\lambda+\rho)}}{\sum_{w\in W} \varepsilon(w) e^{w\rho}},$$

where $\rho = \sum_i \omega_i$.

If we define

$$h_i v^+ = \lambda_i v^+,$$
$$x_i v^+ = 0,$$

for the generators $x_i, h_i, 1 \le i \le n$ of $\mathfrak{U}$ then we can use the defining relations for $\mathfrak{U}$ to rewrite the expressions of the form $x_i y_{i_1} y_{i_2} \cdots y_{i_k} v^+$ and $h_i y_{i_1} \cdots y_{i_k} v^+$ as elements of V^+. In this way we get an action of $\mathfrak{U}$ on V^λ. V^λ is an irreducible $\mathfrak{U}$-module.

Weight multiplicities

Let $\mu \in P$ and define

$$(V^\lambda)_\mu = \ \mathrm{span}\{v = y_{i_1} \cdots y_{i_k} v^+ | wt(v) = \mu\}.$$

Define

$$K_{\lambda\mu} = \dim((V^\lambda)_\mu). \tag{3.5}$$

It is clear from the definitions that

(a) The $K_{\lambda\mu}$ are nonnegative integers.
(b) $K_{\lambda\lambda} = 1$ for all $\lambda \in P^+$.
(c) $K_{\lambda\mu} = 0$ if $\mu \not\le \lambda$.

The key fact motivating the Weyl group is the following.

(3.6) Proposition.

$$K_{\lambda\mu} = K_{\lambda,w\mu},$$

for all $w \in W$.

To show this one uses the $\mathfrak{U}$ action on V^λ to define a map $\tilde{s}_i$ on V^λ by

$$\tilde{s}_i v = exp(x_i) exp(-y_i) exp(x_i) v.$$

One can show that $\tilde{s}_i$ is a bijection from $(V^\lambda)_\mu$ to $(V^\lambda)_{s_i\mu}$ for each $\mu \in P$. Then if $w = s_{i_1} \cdots s_{i_p}$ is an element of the Weyl group, $\tilde{s}_{i_1} \cdots \tilde{s}_{i_p}$ is a bijection between $(V^\lambda)_\mu$ and $(V^\lambda)_{w\mu}$. This gives that $K_{\lambda\mu} = K_{\lambda,w\mu}$.

From (3.3) and (2.3) we have that

$$\begin{aligned}\chi^\lambda &= \sum_{\mu\in P} K_{\lambda\mu} e^\mu \\ &= \sum_{\nu\in P^+} K_{\lambda\nu} \sum_{\mu\in W\nu} e^\mu \\ &= \sum_{\nu\in P^+} K_{\lambda\nu} m_\nu.\end{aligned} \tag{3.7}$$

Note that

$$\dim V^\lambda = \sum_{\mu\in P^+} \sum_{W\mu} K_{\lambda\mu}.$$

Since the orbits $W\mu$ are finite and the integers $K_{\lambda\mu}$ are finite we see that V^λ is finite dimensional.

Tensor products

Let $\lambda, \mu \in P^+$ and let V^λ and V^μ be as given above so that $V^\lambda = \text{span}\{y_{i_1}\cdots y_{i_k}v^+\}$, and $V^\mu = \text{span}\{y_{i_1}\cdots y_{i_k}\bar{v}^+\}$. Then the vector space $V^\lambda \otimes V^\mu$ is

$$V^\lambda \otimes V^\mu = \text{span}\{y_{i_1}\cdots y_{i_r}v^+ \otimes y_{j_1}\cdots y_{j_s}\bar{v}^+\}.$$

The weight of a composite word is given by

$$wt(y_{i_1}\cdots y_{i_r}v^+ \otimes y_{j_1}\cdots y_{j_s}\bar{v}^+) = wt(y_{i_1}\cdots y_{i_r}v^+) + wt(y_{j_1}\cdots y_{j_s}\bar{v}^+).$$

Now let B be a basis of V^λ of vectors of the form $y_{i_1}\cdots y_{i_r}v^+$ and let $\bar{B}$ be a basis of V^μ of vectors of the form $y_{j_1}\cdots y_{j_s}\bar{v}^+$. Then the words $b\otimes\bar{b}$, $b\in B$, $\bar{b}\in\bar{B}$ form a basis of $V^\lambda\otimes V^\mu$. Then the character of $V^\lambda\otimes V^\mu$, the generating function of the basis $B\otimes\bar{B}$, is

$$\sum_{b\otimes\bar{b}} e^{wt(b\otimes\bar{b})} = \sum_{b\in B}\sum_{\bar{b}\in\bar{B}} e^{wt(b)}e^{wt(\bar{b})} = \chi^\lambda\chi^\mu.$$

So even the multiplication of elements of A^W is "coming from" representation theory!

$V^\lambda\otimes V^\mu$ is also a $\mathfrak{U}$-module. If g is one of the generators x_i, y_i or h_i then we define

$$g(y_{i_1}\cdots y_{i_r}v^+\otimes y_{j_1}\cdots y_{j_s}\bar{v}^+) = gy_{i_1}\cdots y_{i_r}v^+\otimes y_{j_1}\cdots y_{j_s}\bar{v}^+ + y_{i_1}\cdots y_{i_r}v^+\otimes gy_{j_1}\cdots y_{j_s}\bar{v}^+.$$

This defines a $\mathfrak{U}$ action on $V^\lambda\otimes V^\mu$.

The elementary symmetric functions e_λ are the characters of the tensor product $(V^{\omega_1})^{\otimes\lambda_1}\otimes\cdots\otimes(V^{\omega_n})^{\otimes\lambda_n}$. By Weyl's theorem (3.2) we know that this tensor product can be decomposed as a direct sum of irreducible representations. In fact, for each λ,

$$e_\lambda = \chi^\lambda + \sum_{\mu<\lambda}\tilde{K}_{\lambda\mu}\chi^\mu, \tag{3.8}$$

for nonnegative integers $\tilde{K}_{\lambda\mu}$. This is the motivation for the definition of the elementary symmetric functions. Since the ω_r are the basic generators of the weight lattice P, the $e_r = \chi^{\omega_r}$ are the most fundamental Weyl characters. (3.8) shows that these do indeed generate A^W.

4. Centralizer algebras

Let M be a finite dimensional module for a semisimple Lie algebra $\mathfrak{U}$. By Weyl's complete reduciblity theorem we know that M decomposes as a direct sum of irreducible $\mathfrak{U}$ modules with certain multiplicities c_λ,

$$M \cong \bigoplus_{\lambda \in \hat{M}} c_\lambda V^\lambda.$$

$\hat{M}$ denotes the set of $\lambda \in P^+$ such that V^λ appears in the decomposition of M. Let C be the centralizer of the action of $\mathfrak{U}$ on M, i.e. $C = \mathrm{End}_{\mathfrak{U}}(M)$. The following theorem is the basic result of the double centralizer theory.

(4.1) Theorem. *M is a bimodule for $C \times \mathfrak{U}$ and*

$$M \cong \bigoplus_{\lambda \in \hat{M}} C_\lambda \otimes V^\lambda,$$

where C_λ is an irreducible module for C.

This decomposition induces a pairing between the irreducible modules C_λ of C and irreducible $\mathfrak{U}$ modules V^λ, $\lambda \in \hat{M}$. It is true that the set of C_λ, $\lambda \in \hat{M}$, is a complete set of irreducible modules of C. The pairing between irreducible C modules and the irreducible $\mathfrak{U}$ modules appearing as factors in M can be given in terms of characters. By definition, a character of C is a linear functional $\xi: C \to \mathbb{C}$ such that

$$\xi(c_1c_2) = \xi(c_2c_1),$$

for all $c_1, c_2 \in \mathbb{C}$. Let $\hat{C}$ be the vector space of characters of C. Given a C-module M the character ξ of C corresponding to M is given by defining $\xi(c)$, $c \in C$, to be the trace of the action of c on M. Let ξ_λ denote the irreducible character of C corresponding to the irreducible C-module C_λ. For each $\lambda \in P^+$ let χ^λ denote the corresponding Weyl character for $\mathfrak{U}$. Let $\Lambda_{\hat{M}}$ be the vector space which is the span of the χ^λ, $\lambda \in \hat{M}$. Define the characteristic map ch to be the map given by

$$\begin{array}{cccc} ch: & \hat{C} & \to & \Lambda_{\hat{M}} \\ & \xi_\lambda & \mapsto & \chi^\lambda \end{array}$$

For each $\mu = \sum_i \mu_i \omega_i \in P$ define the μ weight space of M to be the vector space

$$M_\mu = \{m \in M \mid \text{for each } 1 \leq i \leq n,\ h_i m = \mu_i m\}.$$

Let $\mathfrak{U}(\mathfrak{H})$ be the subalgebra of $\mathfrak{U}$ generated by the h_i. M decomposes as a direct sum of weight spaces under the action of $\mathfrak{U}(\mathfrak{H})$,

$$M \cong \bigoplus_{\mu \in P} M_\mu. \tag{4.2}$$

Since $\mathfrak{U}(\mathfrak{H})$ is a subalgebra of $\mathfrak{U}$ each M_μ is a C module. The bicharacter of M is defined to be

$$\text{bichar}M = \sum_\mu \eta_\mu e^\mu, \tag{4.3}$$

where η_μ denotes the character of M_μ as a C-module. In view of the decomposition in (4.1) we also have

$$\text{bichar}M = \sum_{\lambda \in \hat{M}} \xi_\lambda \chi^\lambda. \tag{4.4}$$

Construction of irreducible modules for centralizer algebras

Consider the action of the generators $x_i \in \mathfrak{U}$ on the weight spaces M_μ of M. For each $\mu \in P$

$$x_i : M_\mu \to M_{\mu+\alpha_i}.$$

Let $\ker_\mu(x_i)$ denote the kernel of the map x_i acting on M_μ. Let

$$\bar{C}_\mu = \bigcap_i \ker_\mu(x_i).$$

(4.5) Theorem.

1) *$\bar{C}_\mu$ is either 0 or an irreducible C module. Furthermore all irreducible C modules can be obtained in this fashion.*
2) *$\bar{C}_\mu \neq 0 \iff \mu \in \hat{M}$.*

Proof. Using (4.1) we have that $M_\mu = \oplus_{\lambda \in \hat{M}} C_\lambda \otimes (V^\lambda)_\mu$. But the only weight in V^λ killed by *all* x_i is λ. So there are no elements of $(V^\lambda)_\mu$ in the $\cap_i \ker_\mu(x_i)$ unless $\lambda = \mu$. When $\lambda = \mu$ we have $(V^\mu)_\mu \subseteq \ker_\mu(x_i)$. Thus

$$\begin{aligned} \bar{C}_\mu &= \bigcap_i \ker_\mu(x_i) \\ &\cong \bigoplus_{\lambda \in \hat{M}} C_\lambda \otimes (V^\lambda)_\mu \delta_{\lambda\mu} \\ &\cong C_\mu, \end{aligned}$$

as C-modules. □

A character formula for C_λ

(4.6) Theorem. *The character of the irreducible C-module C_λ can be given by*

$$\xi_\lambda = \sum_{w \in W} \varepsilon(w) \eta_{\lambda+\rho-w\rho},$$

where η_μ is the character of M_μ as a C-module.

Proof. Equating the expressions (4.3) and (4.4) and rewriting χ^λ by using (3.4) we have

$$\sum_\lambda \xi_\lambda \sum_{w\in W} \varepsilon(w)e^{w(\lambda+\rho)-\rho} = \sum_{\mu\in P}\sum_{w\in W} \eta_\mu \varepsilon(w)e^{\mu+w\rho-\rho}.$$

Substitute $\gamma = \mu + w\rho - \rho$ to get

$$\sum_\lambda \xi_\lambda \sum_{w\in W} \varepsilon(w)e^{w(\lambda+\rho)-\rho} = \sum_{\mu\in P}(\sum_{w\in W} \varepsilon(w)\eta_{\gamma+\rho-w\rho})e^{\gamma}.$$

Compare coefficients of e^γ for $\gamma \in P^+$ on each side of this equation. Since $\lambda \in P^+$ we know by (2.3) that $w(\lambda+\rho)-\rho$ is not an element of P^+ for any $w \in W$ except the identity. Thus

$$\xi_\lambda = \sum_{w\in W} \varepsilon(w)\eta_{\lambda+\rho-w\rho}. \qquad \square$$

(4.7) Corollary. *The dimension c_λ of the irreducible C-module C_λ is given by*

$$c_\lambda = \sum_{w\in W} \varepsilon(w)d_{\lambda+\rho-w\rho},$$

where d_μ is the dimension of the weight space M_μ.

Proof. Evaluate the identity in Theorem (4.6) at the identity of C. $\square$

5. References

[Bou] N. Bourbaki, Groupes et algebres de Lie, Ch. 4-6, Masson, Paris (1981).

[GW] R. Goodman and N. Wallach, Vector and Tensor Invariants: An Introduction to Complex Reductive Groups and their Representations, book in preparation, Chapter 5. Cohomology and Kostant's Theorem, preprint (1990).

[Hu] J. Humphreys, Introduction to Lie Algebras and Representation Theory, "Graduate Texts in Mathematics," Vol. 40, Springer-Verlag, New York/Berlin, 1972.

[Mac] I.G. Macdonald, Symmetric Functions and Hall Polynomials, Clarendon Press, Oxford (1979).

[Mac2] I.G. Macdonald, Lecture notes on root systems, UCSD (1991).

[R] A. Ram, *Representation theory and character theory of centralizer algebras*, Dissertation, Univ. of Calif., San Diego, 1991.

[R2] A. Ram, Representations of semisimple Lie algebras, Lecture notes, MIT, 1992.

[V] D.-N. Verma, *A New BGG style resolution conjectured*, preprint (1984).

[Ze] A.V. Zelevinskii, *Resolvents, dual pairs, and character formulas*, Funct. Anal. Appl. **21**, 152-154 (1987).

Counting Lattice Points in Pyramids

Patrick Solé *

Abstract

The old problem of counting lattice points in euclidean spheres leads to use Jacobi Theta functions and its relatives as generating functions. Important lattices (root systems, the Leech lattice) can be constructed from algebraic codes and analogies between codes and lattices have been extensively studied by coding theorists and number theorists alike . In this dictionnary, the MacWilliams formula is the finite analog of the Poisson formula.
The new problem of counting lattice points in spheres for the L^1 distance leads to hyperbolic trigonometric functions. The same analogy exists but the L^1 counterpart of the Poisson formula is missing. The MacWilliams formula lead to such a duality formula for those lattices which are constructed from codes via Construction A. A connection with Ehrart's enumerative theory of polytopes is pointed out. Both problems have important applications in multidimensional vector quantization.

1 Motivation

An n-dimensional lattice Λ_n is defined as a set of vectors

$$\Lambda_n = \{X \in \mathbb{R}^m | X = u_1 a_1 + \ldots + u_n a_n\}$$

where $a_1, ..., a_n$ are linearly independent vectors in m-dimensional real Euclidean space $\mathbb{R}^m$ with $m \geq n$, and $u_1, ..., u_n$ are in $\mathbb{Z}$. A well known result of Lloyd-Max is that, for quantizing an uniform distribution on the real line, the decision intervals have to be taken of the form $[n - \frac{1}{2}, n + \frac{1}{2}]$. The centers of the intervals constitute a one dimensional lattice ($n = 1$). In the general n-dimensional case, the role of the integers is played by the lattice points, and the role of the decision intervals is played by the so-called Voronoi region of the Lattice [2]. It was shown by Zador [9], extending pionner work of Schützenberger [6], that the quadratic error depended crucially on the geometry of the lattice by a term called the second moment of the lattice [6]. If G_n denotes the average mean squared error per dimension for the best quantizing lattice in n dimensions, then Zador showed that: $\lim(G_n) = \frac{1}{2\pi e} < \frac{1}{12}$ for large n This demonstrates the interest of using multidimensionnal lattices. As Conway and Sloane put it :"it pays to procrastinate". In general, finding the best

*CNRS, I3S , 250, rue A. Einstein, 06 560 Valbonne, FRANCE. Work supported in part by PRC Math-Info.

quantizing lattices in n dimensions is a difficult task. They are only known for some integers $n \leq 24$. More applicative information can be found in the companion paper [1].

When using an infinite lattice for source coding one needs to truncate it according to a shape suited to the probability law of the source to be encoded : euclidean spheres for a Gaussian distribution, pyramids (or hyperoctahedra) for an exponential distribution. It is of practical importance to know how many lattice points remain in the truncated lattice. This task is performed with the help of generating functions

- the old theta functions for spheres
- the new nu function for pyramids

This work emphasizes two analogies:

- codes vs lattices
- spheres vs pyramids

The paper is organized as follows. After reviewing the first analogy in section 1, we develop the second in section 2. In section 3 we derive the generating functions associated with the problem of counting lattice points for pyramids. In particular we treat root systems D_n, E_8. A few numerical results are presented in section 4. We highlight differences, in particular the non-invariance of pyramids by rotations of $\mathbb{R}^n$, and the absence of a Poisson summation formula (subsection 3.3). The connection with Ehrart's enumerative theory of Polytopes is developed in section 5.

2 Theta functions of Lattices

2.1 Lattices

The theta function of a lattice Λ is defined as the formal power series:

$$\theta_\Lambda(q) = \sum_{x \in \Lambda} q^{||x||^2},$$

where $||x||^2 = x.x$, the squared standard euclidean norm. In words the coefficient N_m of q^{m^2} in $\theta_\Lambda(q)$ counts the number of lattice points at distance m from the origin in $\mathbb{R}^n$. The classical Jacobi theta function is

$$\Theta(\xi|z) = \sum_{m=-\infty}^{+\infty} e^{2mi\xi + \pi i z m^2}.$$

The sum converges for $\Im(z) > 0$(the so-called Poincaré upper half-plane). It is classical to set $q = e^{\pi i z}$.For us, q will be, however an indeterminate. Special instances of this formula which will be of use here are

$$\theta_3(q) = \Theta(0|z) = \sum_{m=-\infty}^{+\infty} q^{m^2},$$

the theta function of $\mathbb{Z}$, and

$$\theta_2(q) = e^{\frac{2\pi z}{4}}\Theta(\frac{\pi z}{2}|z) = \sum_{m=-\infty}^{+\infty} q^{(m+\frac{1}{2})^2}.$$

If we extend the definition of theta functions from lattices to cosets of lattices we see that $\theta_{2\mathbb{Z}+1}(q) = \theta_2(q^4)$.. It will also be noticed that $\theta_{2\mathbb{Z}}(q) = \theta_3(q^4)$. Finally, we will also need the relation

$$\theta_4(q) = \theta_3(q^4) - \theta_2(q^4), \tag{1}$$

where

$$\theta_4(q) = \Theta(\frac{\pi}{2}|z) = \sum_{m=-\infty}^{+\infty} (-q)^{m^2}.$$

as well as the simpler

$$\theta_3(q) = \theta_3(q^4) + \theta_2(q^4), \tag{2}$$

which follows from the coset decomposition $\mathbb{Z} = 2\mathbb{Z} + (2\mathbb{Z} + 1)$.

2.2 Algebraic Codes

For us a binary linear *code* of length n will be a linear subspace of $\mathbb{F}_2^n$ coordinatized w.r.t. a special basis of $\mathbb{F}_2^n$. The *weight* of a binary vector u, henceforth denoted by $|u|$ is the number of its nonzero coordinates on this basis. The weight enumerator of a code C is defined as

$$W_C(x,y) = \sum_{u \in C} x^{n-|u|}y^{|u|}.$$

We denote by $|C|$ the cardinality of C. For more details on block codes we refer to [2].

2.3 Construction A

The most simple way to associate a lattice with a code is construction A.

$$A(C) := \{(x_1, \ldots, x_n) \in \mathbb{Z}^n | \exists c \in C,\ x \equiv c((2))\}.$$

Then the following result is classical (Theorem 3 of Chapter 7 of [2]).

$$\theta_{A(C)} = W_C(\theta_3(q^4), \theta_2(q^4)).$$

Denoting by U_n the universe code $\mathbb{F}_2^n$, we see that $W_{U_n}(x,y) = (x+y)^n$, and that $A(U_n) = \mathbb{Z}^n$, so that

$$\theta_{\mathbb{Z}^n}(q) = (\theta_3(q^4) + \theta_2(q^4))^n = \theta_3(q)^n,$$

(by equation (2)) which was to be expected by the product rule for ordinary generating functions since $\mathbb{Z}^n$ is a cartesian product.

The root lattice D_n (dubbed checkerboard lattice in [2] is defined as follows

$$D_n = \{(x_1, \dots, x_n) \in \mathbb{Z}^n | \sum_{i=1}^{n} x_i \equiv 0(2)\}.$$

Clearly, $D_n = A(EW_n)$ where EW_n is the even weight code of dimension $n-1$. Its weight enumerator collects the even powers of y in the w.e. of U_n.

$$W_{EW_n}(x,y) = \frac{1}{2}((x+y)^n + (x-y)^n).$$

By equations (1) and (2) we see that

$$\theta_{D_n}(q) = \frac{1}{2}(\theta_3(q)^n + \theta_4(q)^n).$$

A slightly more advanced example is obtained for the root lattice E_8 which is produced by construction A applied to the Hamming code H_8 of length 8 of weight enumerator

$$W_{H_8} = x^8 + 14x^4y^4 + y^8.$$

This yields equation (101) of Chapter 4 of [9]

$$\theta_{E_8} = \theta_2(q)^8 + 14\theta_2(q)^4\theta_3(q)^4 + \theta_3(q)^8.$$

2.4 Construction B

Construction B of [2] is defined as follows . Let C denote a binary linear code with weights multiple of 4. (Such a code is usually called "doubly even"). With this code construction B associates a lattice $B(C)$ by the formula

$$B(C) := \{(x_1, \dots, x_n) \in \mathbb{Z}^n | \exists c \in C,\ x \equiv c((2)) \sum_{i=0}^{n} x_i \equiv 0((4))\}.$$

The following result is classical (Theorem 15 of Chapter 7 of [2]).

$$\theta_{B(C)}(q) = \frac{1}{2}W_C(\theta_3(q^4), \theta_2(q^4)) + \frac{1}{2}\theta_4(q^4)^n.$$

2.5 Poisson and MacWilliams

The dual of a code C is its dual w.r.t to the standard inner product $x.y = \sum_{i=1}^{n} x_i y_i$, and is usually denoted by $C^{\perp}$. The celebrated MacWilliams formula connects the w.e. of a code with the w.e. of its dual.

$$W_{C^{\perp}}(x,y) = \frac{1}{|C|}W_C(x+y, x-y).$$

The dual of a lattice Λ is defined w.r.t. to the same inner product as

$$\Lambda^* = \{x \in \mathbb{R}^n | \forall y \in \Lambda\ x.y \in \mathbb{Z}\}$$

The Poisson formula connects the theta function of a lattice with the theta function of its dual.We denote by $|L|$ the elementary volume of the lattice L. Letting $q = e^{\pi i z}$, we have

$$\theta_{L^*}(e^{\pi i z}) = |L|(\frac{i}{z})^{n/2}\theta_L(e^{-(i\pi)/z}).$$

Construction A and B behave nicely w.r.t. duality

$$A(C)^* = \frac{1}{2}A(C^\perp) \quad (3)$$

$$B(C)^* = \frac{1}{2}B(C^\perp). \quad (4)$$

Indeed one could derive the MacWilliams formula from the Poisson formula by using relation 3 and construction A.

3 The Nu function of Lattices

For $x \in \mathbf{R}^n$, let $||x||_1 = \sum_{i=1}^n |x_i|$, and $||x|| = \sqrt{\sum_{i=1}^n |x_i|^2}$. For a lattice Λ, we define its function Nu by the relation

$$\nu_\Lambda(z) = \sum_{y\in\Lambda} z^{||y||_1} = \sum_{n=0}^{+\infty} z^n |\{y \in \Lambda | ||y||_1 = n\}|.$$

Note the analogies and differences with the definition of the theta function [2]

$$\theta_\Lambda(q) = \sum_{y\in\Lambda} q^{||y||^2} = \sum_{n=0}^{+\infty} q^{n^2} |\{y \in \Lambda | ||y|| = n\}|.$$

Caution Unlike the euclidean norm , the norm $||.||_1$ depends on the chosen orthogonal basis. Different bases may yield different functions ν_Λ for the same lattice Λ. An example of this situation will be given for $\Lambda = E_8$, where construction A and construction B yield different Nu functions.

3.1 Easy Examples: $\mathbb{Z}$, $2\mathbb{Z}$ and $\mathbb{Z}^n$.

First we start with dimension 1, where the well-known uniform quantifier (lattice $\mathbb{Z}$) has Nu function

$$\nu_{\mathbb{Z}}(z) = 1 + 2\sum_{n=0}^{+\infty} z^n = 1 + \frac{2z}{1-z} = \frac{1+z}{1-z},$$

a geometric series. The lattice of even integers has Nu function (by scaling)

$$\nu_{2\mathbb{Z}}(z) = \sum_{y\in\mathbb{Z}} z^{2||y||_1} = \nu_{\mathbb{Z}}(z^2) = \frac{1+z^2}{1-z^2},$$

and the set of odd integers, which is a coset of the preceding into $\mathbb{Z}$ (by difference)

$$\nu_{1+2\mathbb{Z}}(z) = \nu_{\mathbb{Z}}(z) - \nu_{2\mathbb{Z}}(z) = \frac{2z}{1-z^2}.$$

Since the cubic lattice is a cartesian product of n linear lattices, we get

$$\nu_{\mathbb{Z}^n}(z) = (\nu_{\mathbb{Z}}(z))^n = (\frac{1+z}{1-z})^n.$$

Since D_n consists of those vectors of $\mathbb{Z}^n$ whose L_1 norm is even, we see that its Nu function is the even part of $\nu_{\mathbb{Z}^n}(z)$ that is

$$\nu_{D_n}(z) = \frac{1}{2}(\nu_{\mathbb{Z}^n}(z) + \nu_{\mathbb{Z}^n}(-z)).$$

3.2 Connection with block codes

The next result generalizes to the L^1 metric the Theorem 3 of Chapter 7 of [2].

Theorem 1 *Let C be a binary linear block code with weight enumerator $W_C(x,y)$. Then*

$$\nu_{A(C)} = W(\nu_{2\mathbb{Z}}(z), \nu_{1+2\mathbb{Z}}(z)) = W_C(1+z^2, 2z)(1-z^2)^{-n}.$$

Alternatively, for an indeterminate α, we have

$$\nu_{A(C)}(\tanh(\frac{\alpha}{2})) = W(\cosh(\alpha), \sinh(\alpha)).$$

Proof: Let $c \in C$ and $\bar{c}$ denote the following lattice

$$\{y \in \mathbb{Z}^n | y \equiv c[2]\}$$

Then $A(C)$ is a disjoint union of such $\bar{c}$ for $c \in C$. This entails

$$\nu_{A(C)} = \sum_{c \in C} \nu_{\bar{c}}(q).$$

To compute each summand, observe that $\bar{c}$ is a cartesian product $(2\mathbb{Z})^{n-|c|}(1+2\mathbb{Z})^{|c|}$.Therefore

$$\nu_{\bar{c}}(q) = \nu_{(2\mathbb{Z})}^{n-|c|}(q)\nu_{(1+2\mathbb{Z})}^{|c|}(q).$$

The result follows. □

Let $R_n := \{0, 1\}$ denote the repetition code. Clearly $W_{R_n}(x,y) = x^n + y^n$. From the dual of the repetition code, we get $\nu_{D_n} = \frac{1}{2}(\nu_{\mathbb{Z}^n}(z) + \nu_{\mathbb{Z}^n}(-z))$, as was expected from section 3.1. The dual code yields the ν function of the dual lattice.

$$\nu_{D_n^*} = 2((1+z^2)^n + 2^n z^n)/(2-2z^2)^n.$$

From the Hamming code of length 8, we get

$$\nu_{E_8}(z) = [(1+z^2)^8 + 224z^4(1+z^2)^4 + 256z^8]/(1-z^2)^8 =$$

$$1 + 16z^2 + 352z^4 + 3376z^6 + 19648z^8 + 82256z^{10} + O(z^{12}).$$

Applying the complex variable techniques of [4] yields the following asymptotic estimate (a more elementary and more tedious proof can be obtained by partial fraction expansion)

Corollary 1

$$[z^m]\frac{\nu_{A(C)}(z)}{1-z} \sim |C|\frac{m^n}{n!}.$$

Proof: From Theorem 1 we see that $\nu_{A(C)}(z)$ has a singularity in $z=1$ where it is equivalent to $\frac{|C|}{(1-z)^n}$, since W being homogeneous $W(2,2)=2^n$. Now the term in z^m in $\frac{1}{(1-z)^n}$ is $\binom{n+m-1}{m-1}$. The result follows by the tranfer Theorems of Flajolet-Odlyzko. □

Since the volume of the elementary parallelotope of $A(C)$ is $\frac{2^n}{|C|}$, and the volume of the unit L^1 sphere is $\frac{2^n}{n!}$, Corollary 1 says that the number of lattice points at L^1 distance m from the origin is asymptotically equivalent to the volume of the L^1 sphere of radius m divided by the volume of the lattice. This intuitive result is known as "Gauss counting principle"[5].

Theorem 2 *Let C be a doubly even binary code of length n, and weight enumerator W. Its Nu function is*

$$\nu_{B(C)}(z) = \frac{1}{2}W(\frac{1+z^2}{1-z^2}, \frac{2z}{1-z^2}) + \frac{1}{2}(\frac{1-z^2}{1+z^2})^n.$$

Proof: Let $f(q,a)$ denote the bivariate generating function

$$(\nu_{4\mathbb{Z}}(q) + a\nu_{4\mathbb{Z}+2}(q))^{n-|c|}(\nu_{4\mathbb{Z}+1}(q) + a\nu_{4\mathbb{Z}-1}(q))^{|c|}$$

We claim that, with the notations of the proof of Theorem 1, we have

$$\nu_{\bar{c}}(q) = \frac{1}{2}(f(q,1) + f(q,-1))$$

Since $\nu_{4\mathbb{Z}+1}(q) = \nu_{4\mathbb{Z}-1}(q)$, the only contribution of $f(q,-1)$ comes from the term in $|c|=0$. Then the values of $\nu_{4\mathbb{Z}}(q)$ and $\nu_{4\mathbb{Z}+2}(q)$ are easily computed by scaling from the values of $\nu_{(2\mathbb{Z})}$ and $\nu_{(2\mathbb{Z}+1)}$ calculated in section (4.1). The result follows.
To prove the claim, let $c \in B(C)$ and $y \in \bar{c}$ and let N_i, $i = 0, \pm 1, 2$, denote the number of coordinates of y congruent to i mod 4. Since $|c|$ is a multiple of 4 we see that $N_1 + N_{-1} \equiv 0[4]$. From there it follows that the sum $\sum_i y_i$, which is $N_1 + 2N_2 - N_{-1}$, is congruent to 0 modulo 4. as it should in construction B, iff $N_2 + N_{-1}$ is even. But this latter term is obtained by summing up the even terms in a in $f(q,a)$. □

We are indebted to Dr. Sloane for the L^2 version of this proof.

Applied to the code R_8 this theorem yields a different result for the nu function of E_8.

$$\nu_{E_8}(z) = \frac{1}{2}\frac{(1+z^2)^8 + 256z^8}{(1-z^2)^8} + \frac{1}{2}\frac{(1-z^2)^8}{(1+z^2)^8} = 1 + 128z^4 + 2944z^8 + 1024z^{10} + O(z^{12})$$

This could be expected from asymptotic considerations since we have by [4]

Corollary 2

$$[z^m]\frac{\nu_{B(C)}(z)}{1-z} \sim \frac{|C|}{2}\frac{m^n}{n!}.$$

Proof: From Theorem 2 we see that $\frac{\nu_{B(C)}(q)}{1-q}$ has a singularity at $z = 1$ where it is equivalent to $\frac{|C|}{2(1-q)^{n+1}}$, since $W(2,2) = 2^n$. The rest is analogous to the proof of Corollary 1. □

This shows that if a lattice can be constructed both by construction A and construction B, they will yield different orientations, and, in addition, orientation B will have half as many points in pyramids, which is interesting in quantizing applications.

3.3 Nu Functions of Dual lattices

There is no known analog of the Poisson-Jacobi formula. The combination of MacWilliams formula and of the relation 3 for construction A leads to the following conjecture.

Conjecture 1 *Let the parameters α and β be connected by the relation $e^{-2\beta} = \tanh(\alpha)$, and let L be a lattice. The functions ν_L and ν_{L^*} satisfy the identity*

$$2^{n/2}\nu_{L^*}(\tanh^2(\frac{\beta}{2})) = |L|(\sinh(2\beta))^{n/2}\nu_L(\tanh(\frac{\alpha}{2})).$$

The relation between α and β is indeed symmetric: $e^{-2\alpha} = \tanh(\beta)$.

4 Numerical Examples

In the following tables the second line indicates the number of lattice points *within* a pyramid of height m. The fourth line indicates the number of lattice points *within* a sphere of radius m. In all examples the pyramids of sufficiently large radius present much fewer lattice points than the euclidean spheres of the same radius. In the case of cubic lattices this can be paralleled with the fact that the volume of the unit n-sphere is $2^{n/2}/\Gamma(n/2+1)$, whereas the volume of the unit n-pyramid is $2^n/n!$, which is much smaller for large n. All calculations were performed in MAPLE on a DEC 500.

4.1 Plane Cubic Lattice $\mathbb{Z}^2$

m	1	2	3	4
$[z^m]\nu(z)/(1-z)$	5	13	25	41
m^2	1	4	9	16
$[q^{m^2}]\theta(q)/(1-q)$	5	13	25	51

4.2 Cubic Lattice in dimension 3: $\mathbb{Z}^3$

m	1	2	3	4	5
$[z^m]\nu(z)/(1-z)$	7	25	63	129	231
m^2	1	4	9	16	25
$[q^{m^2}]\theta(q)/(1-q)$	7	27	90	224	482

Note that the surpopulation of spheres as compared to pyramids starts earlier ($m =$ 2) than in the preceding example.

4.3 Lattice D_4

m	1	2	3	4	5	6	7	8
$[z^m]\nu(z)/(1-z)$	1	33	33	225	225	833	833	2241
m^2	1	4	9	16	25	36	49	64
$[q^{m^2}]\theta(q)/(1-q)$	1	49	149	605	1435	3307	5659	9979

4.4 Lattice E_8

Note the factor 10 between the two entries for $m = 4$.

m	1	2	3	4
$[z^m]\nu(z)/(1-z)$	1	129	129	3073
m^2	1	4	9	16
$[q^{m^2}]\theta(q)/(1-q)$	1	2401	26641	340321

5 Nu functions of sub-lattices of $\mathbb{Z}^n$.

In this section, we shall show that the function $\nu(z)$ of a full rank sublattice of $\mathbb{Z}^n$ is always a rational function. We shall rely on a result of Ehrart [8, Theorem 4.6.25], and will assume some familiarity with polytopes.

Lemma 1 *Let $\mathcal{P}$ denote a convex rational polytope of $\mathbb{R}^n$, with vertex set V and let $p_m = |m\mathcal{P} \cap \mathbb{Z}^n|$. Then*

$$\sum_{m \geq 0} p_m z^m = \frac{P(z)}{\prod_{a \in V}(1 - z^{d(a)})},$$

where $d(a)$ is the smallest integer q such that $qa \in \mathbb{Z}^n$. Further, $deg(P) < \sum_{a \in V} d(a)$.

Let H_n denote the n-dimensional octahedron, namely

$$H_n = \{x \in \mathbb{R}^n \mid ||x||_1 = 1\}.$$

Alternatively, H_n can be thought of as the convex hull of the $2n$ vectors $\pm e_i$ where $e_i, i = 1, 2, \ldots n$ is the canonical basis. By the multiplicative property of the L^1 norm, for an x of $\mathbb{R}^n$, we have $||x||_1 = m$ iff $x \in m\mathcal{P}$. We are now in a position to state the main result of this section.

Theorem 3 *Let $\Lambda \subseteq \mathbb{Z}^n$ be a lattice with generating matrix A of full rank. Let a_i be the rows of A^{-1}. Then*

$$\nu_\Lambda(z) = \frac{P(z)}{\prod_{i=1}^n (1 - z^{d(a_i)})},$$

where $P \in \mathbb{Z}[X]$ and $deg(P) < \sum_{i=1}^n d(a_i)$. In particular, the roots of the denominator are complex N^{th} roots of unity with $N = \det(A)$.

Proof: By definition of the Nu function, we have

$$[z^m]\nu_\Lambda(z) = |mH_n \cap \Lambda| = |mH_n A^{-1} \cap \mathbb{Z}^n|.$$

Now, let $\mathcal{P} = H_n A^{-1}$, with vectors of H_n considered as row vectors. Like H_n, the body $\mathcal{P}$ is a convex rational polytope with vertices $\pm e_i A^{-1}$. By Cramer's rule $d(a_i)$ divides $\det(A)$. The result follows by the preceding Lemma. □

As an immediate application of the Theorem, we can recover part of Theorem 1. In construction A it is known [2] that, if C has dimension k, and binary generating matrix $(I_k|B)$, then a generating matrix of $A(C)$ is

$$A = \left(\begin{array}{c|c} I_k & B \\ 0 & 2I_{n-k} \end{array} \right).$$

It is then easy to see that for this matrix $d(a_i) = 2$ for every i. This is consistent with Theorem 1 where the expression $(1 - z^2)^n$ was found for the denominator of Nu. It would be nice to have a combinatorial interpretation of the numerator in the general case. In view of Gauss counting principle and of the results of the preceding section it is natural to make the following conjecture.

Conjecture 2 *Let $\Lambda \subseteq \mathbb{Z}^n$ be a lattice with generating matrix of full rank A. Then, when z is near 1*

$$\frac{\nu_\Lambda(z)}{1 - z} \sim \frac{1}{\det(A)(1 - z)^{n+1}}.$$

In particular, if $P(1) \neq 0$, then $P(1) = 1$.

6 Acknowledgement

The author wishes to thank N.J.A. Sloane for [7] and for helpful discussions and R.A. Calderbank for inviting him to give a talk in Bell Labs on this topic in October 1991. He also thanks Loic Pottier for helpful discussions on polytopes.

References

[1] M. Barlaud, P. Solé, M. Antonini, P. Mathieu, T. Gaidon, "A Pyramidal Scheme for Lattice Vector Quantization of Wavelet Transform Coefficients Applied to Image Coding," submitted to IEEE Trans. on Image Processing.

[2] J.H. Conway, N.J.A. Sloane, *Sphere Packings, Lattices and Groups*, Springer (1988).

[3] E. Ehrart, *Polynômes Arithmétiques et Méthodes des polyèdres en Combinatoire*Birkhäuser (1977).

[4] P. Flajolet, A. Odlyżko,"Singularity Analysis of generating Functions", Rapport de Recherche INRIA 826, April 88.

[5] J. E. Mazo, A.M. Odlyzko,"Lattice Points in High-Dimensional Spheres" Mh. Math. 110, 47-61 (1990).

[6] M-P. Schützenberger,"On the quantization of finite dimensional messages", Info and control, vol.1, (1958)353-380.

[7] N.J.A. Sloane, letter to the author, August 1^{st}, 1991.

[8] R.P. Stanley,*Enumerative Combinatorics*, Wadworsth Brooks/ Cole (1986).

[9] P. Zador, "Asymptotic Quantization Error of Continuous Signals and their Quantization Dimension", IEEE Trans. on Information Theory , vol.IT-28 , 1982.

ON PERMUTATION REPRESENTATIONS OF WEYL GROUPS, DESCENT NUMBERS, AND THE FACE RING OF THE COXETER COMPLEX

JOHN R. STEMBRIDGE

Department of Mathematics
University of Michigan
Ann Arbor, Michigan 48109–1003 USA

Extended Abstract

Let R be a (reduced, crystallographic) root system with Weyl group $W = W(R)$ and Coxeter complex Δ_R. This talk will be concerned with a certain representation ρ^R of W that has been largely ignored until recently. It can be succinctly described as the representation carried by the cohomology of the toric variety of Δ_R, although it can also be given a purely algebraic definition as the representation carried by a certain quotient of the face ring of Δ_R. It should be emphasized that this is not the representation one obtains from the homology of the Coxeter complex itself; this latter representation has received considerably more attention, thanks to the work of Björner [B], Garsia-Stanton [GS], and Stanley [St1].

Our main result is the fact that ρ^R is, for no easily explainable reason, a *permutation* representation; i.e., there exists a basis for ρ^R such that the Weyl group acts by permuting this basis. Unfortunately however, there are two senses in which we regard our proof of this result as unsatisfying. First, it must be applied to each root system on a case-by-case basis. Second, it is non-constructive—we lack an explicit basis for ρ^R that is permuted by W, even for the root systems of type A. Even more vexing is the fact that we can exhibit a number of beautiful properties that come close to characterizing ρ^R as a permutation module, but we are unable to explicitly construct (except in particular cases) a simple set of combinatorial objects permuted naturally by W in a manner isomorphic to ρ^R. In fact, since $\dim \rho^R = |W|$, this means that W itself is an obvious choice for the set of objects. It is hard to imagine that there could exist a natural permutation action of W upon itself

Partially supported by grants from the NSF and a Sloan Fellowship. .

that does not admit a simple description, such as left-multiplication or conjugation, but that is the state we are in.

On the positive side, in the course of proving this theorem we did make a number of interesting combinatorial discoveries, mostly involving descent numbers (i.e., the Weyl group generalization of the Eulerian numbers). Another byproduct of this work is a collection of *Maple* procedures we created for manipulating Weyl groups and root systems. Aside from the easy case $R = G_2$, our proofs for the cases involving the exceptional root systems rely on these procedures, and thus are computer-based.

Acknowledgements

I am indebted to Roger Carter, Igor Dolgachev, and Richard Stanley for some valuable discussions, and also to Alex Ryba, for assistance in extracting some information about $W(E_8)$ from the group theory package *Cayley*.

Origins

The original motivation for this work can be found in Stanley's discussion of some interesting combinatorial properties of the special case $R = A_n$ in [St2, pp. 524–529]. This provoked me into studying this particular case in more detail [Ste]. More recently, Dolgachev and Lunts proved a nice character formula for ρ^R in the general case [DL]. This reawakened my interest in the subject, and led to the discoveries I will report on here.

Some Details about the Representation

Let V be an n-dimensional real Euclidean space and let R be a root system in V. Associated with R there is a hyperplane arrangement $\mathcal{H}_R = \{\alpha^\perp : \alpha \in R\}$. The Coxeter complex Δ_R can be defined geometrically as the simplicial decomposition of the unit sphere S^{n-1} in V induced by $\mathcal{H}_R$.

The Coxeter complex can also be regarded as a (complete, simplicial) fan in V with respect to the weight lattice P. That is, it is a decomposition of V into strongly convex (simplicial) cones, each generated by certain integral weights in P. As is the case with any fan, there is a toric variety X_R naturally associated with Δ_R [O]. The Weyl group acts naturally on Δ_R, and therefore also on X_R and the cohomology ring $H^*(X_R, \mathbf{C})$.

Let f_i denote the number of i-dimensional faces of Δ_R (with $f_{-1} = 1$), and define

$$P_R(q) = \sum_{i=0}^{n} f_{i-1}(q-1)^{n-i} = \sum_{i=0}^{n} h_i q^{n-i}.$$

The first equality serves to define the polynomial $P_R(q)$, and the second equality serves to define the h-vector $h(R) = (h_0, \ldots, h_n)$ of Δ_R. By a theorem of Danilov and Jurkiewicz

[D], one knows that

$$\dim H^{2i}(X_R, \mathrm{C}) = h_i(R), \qquad \dim H^{2i+1}(X_R, \mathrm{C}) = 0,$$

so $P_R(q)$ is essentially the Poincaré polynomial of X_R.

In particular, it follows that ρ^R, the W-representation carried by $H^*(X_R)$, has $n+1$ (nonzero) graded components, and the total dimension of ρ^R is $P_R(1) = f_{n-1}(R) = |W|$. Let χ_q^R denote the graded character of ρ^R; i.e., for $w \in W$,

$$\chi_q^R(w) = \sum_{i=0}^{n} \mathrm{tr}(H^{2i}(X_R), w) q^i,$$

where $\mathrm{tr}(U, w)$ denotes the trace of w on U.

Theorem (Dolgachev-Lunts [DL]). *For $w \in W$, let $\delta(w) = \dim\{v \in V : wv = v\}$ and $\Delta_R^w = \{F \in \Delta_R : w(F) = F\}$ (a subcomplex of Δ_R). We have*

$$\chi_q^R(w) = P_{R,w}(q) \frac{\det(1 - qw)}{(1-q)^{\delta(w)}},$$

where (1) $P_{R,w}(q)$ denotes the Poincaré polynomial of Δ_R^w, and (2) the determinant is evaluated with respect to the reflection representation.

It is also possible to give a purely algebraic definition of the cohomology ring $H^*(X_R, \mathrm{C})$ and the representation ρ^R it carries. To describe this, let $v_i \in V$ denote the set of vertices of Δ_R, with i ranging over some suitable index set I. Recall that the face ring (or Stanley-Reisner ring) $\mathcal{F}_R$ of Δ_R is the quotient $\mathrm{C}[x_i : i \in I]/\Psi$, where Ψ is the ideal generated by monomials whose supports (i.e., subset of vertices with nonzero exponent) are not faces of Δ_R. If we define

$$\theta_j = \sum_{i \in I} \langle v_i, \varepsilon_j \rangle x_i,$$

where $\varepsilon_1, \ldots, \varepsilon_n$ is some basis for V, then $\Theta = (\theta_1, \ldots, \theta_n)$ forms a system of parameters for $\mathcal{F}_R$. By a theorem of Danilov [D], it is known that

$$H^*(X_R, \mathrm{C}) \cong \mathcal{F}_R/\Theta$$

is an isomorphism of graded rings (as well as W-modules). This isomorphism, together with the fact that $\mathcal{F}_R$ is Cohen-Macaulay, can be used to give an alternative proof of the Dolgachev-Lunts formula.

The Main Result

First consider some generalities about permutation representations.

Suppose G is some finite group acting by permutations on a finite set X. Let $X = X_1 \cup \cdots \cup X_l$ be the partition of X into G-orbits. The action of G on X_i is isomorphic to left multiplication on the cosets of H_i in G, where H_i denotes the stabilizer of some $x \in X_i$. Thus the permutation module for G on X is determined up to isomorphism by the list of point-stabilizers $H_1, \ldots, H_l$. In general, there exist non-isomorphic permutation representations of finite groups G that become isomorphic once they are linearized (i.e., once one permits linear changes of basis). Thus in the following, the assertion that the character of ρ^R agrees with the character of some permutation representation does not necessarily imply that there is only one such isomorphism class of permutation representations (even after allowing for conjugacy).

To state the main result, let $S = \{s_1, \ldots, s_n\}$ denote the set of simple reflections for $W(R)$. Define χ^R to be the specialization of χ_q^R at $q = 1$; i.e., the W-character of $H^*(X_R, \mathbf{C})$, or equivalently of $\mathcal{F}_R/\Theta$, *with the grading ignored.*

Theorem.

(a) *χ^R is the character of some permutation representation π^R of W.*

(b) *The degree of π^R is $|W|$, and the number of orbits is 2^n.*

(c) *The point-stabilizers of π^R are generated by reflections, but not necessarily by simple reflections.*

(d) *If R is reducible; say, $R = R_1 \oplus R_2$, then $\pi^R \cong \pi^{R_1} \otimes \pi^{R_2}$ (outer tensor product).*

From now on, assume R is irreducible. In that case, R has a unique highest root α_0. Let s_0 denote the corresponding reflection across $\alpha_0^{\perp}$, and set $S' = S \cup \{s_0\}$. For any nonempty subset J of S', let $W(J)$ denote the subgroup of W generated by $S' - J$. (If J includes s_0, then W will be a parabolic subgroup, but not otherwise.) Note that if $|J| = r + 1$, then $W(J)$ is a reflection group of rank $n - r$.

(e) *The point-stabilizers are all of the form $W(J)$ for various J ($\varnothing \neq J \subset S'$).*

(f) *The number of point-stabilizers that are reflection groups of rank $n-r$ is $\binom{n+1}{2r+1}$.*

(g) *For each $(2r+1)$-subset J of S', it is possible to choose an $(r+1)$-subset J' of J so that the point-stabilizers of π^R are precisely $\{W(J') : |J| \text{ odd}\}$; i.e.,*

$$\chi^R = \sum_{|J| \text{ odd}} 1_{W(J')}^W. \qquad (*)$$

REMARKS.

(1) The reflections S' are the W-images of the simple reflections for the affine Weyl group $\tilde{W}$ attached to W.

(2) By (f), the smallest possible rank of any point-stabilizer is $n/2$.

(3) Unfortunately, the only rules we have for choosing J' from J are *ad hoc*. For example, for the root systems of types A or C, it is possible to linearly order the reflections S' so that if $J = \{\beta_1 < \cdots < \beta_{2r+1}\}$, then $J' = \{\beta_1, \beta_3, \ldots, \beta_{2r+1}\}$. This rule does not seem to generalize.

(4) The permutation representation π^R depends on R itself, not just $W(R)$. Indeed, even though $W(B_n) = W(C_n)$ and the corresponding representations ρ^R are isomorphic, the highest roots of B_n and C_n are different, and the corresponding decompositions in (*) are not equivalent.

(5) For a given root system R, there may be several possible ways to choose J' from J so that (*) is satisfied. However, in the special case $r = 0$, the constraints of (g) are unambiguous—if J is a singleton, then $J' = J$. In other words, the rank n point-stabilizers that occur in π^R are (with multiplicity) the $n+1$ subgroups of W generated by the n-subsets of S'.

(6) For some root systems (notably types A and C), one can show that the decomposition of χ^R implied by this result is consistent with the grading of $H^*(X_R, \mathbf{C})$ (or equivalently, $\mathcal{F}_R/\Theta$). That is, it is possible to assign a grading to the orbits of π^R so that χ_q^R is the graded character of π^R. For other root systems, such as type D, one can show that χ_q^R is the character of a graded permutation representation, but not one whose point-stabilizers are all generated by reflections in $W(D_n)$ (in violation of (c)). For still other root systems, such as G_2, the grading of $H^*(X_R, \mathbf{C})$ is not consistent with any grading of any permutation representation of $W(G_2)$.

(7) If an explicit construction of a permutation representation π^R satisfying all of (a)–(g) can be found, it is reasonable to expect that this should be accompanied by a combinatorial bijection explaining the evaluation of (*) at $w = 1$; i.e.,

$$|W| = \sum_{|J| \text{ odd}} |W|/|W(J')|.$$

This is non-trivial even for type A (see [Ste]).

(8) The complex Δ_R is known to be shellable [B]. Therefore by a result of Garsia [B, Thm. 1.7], one can construct from the shelling a canonical basis for $\mathcal{F}_R/\Theta$, and hence for ρ^R. However, this basis is not permuted by $W(R)$.

Descent Numbers

As in the previous section, let $S = \{s_1, \ldots, s_n\}$ denote the set of simple reflections. The descent set of any $w \in W$ is defined to be $D(w) = \{i : \ell(ws_i) < \ell(w)\}$, where $\ell(\cdot)$ denotes the length function with respect to S. The shellability of Δ_R leads to a nice combinatorial interpretation of the Poincaré polynomial $P_R(q)$, or equivalently, of the h-vector for Δ_R.

Theorem (essentially [B, Thm. 2.1]). $P_R(q) = \sum_{w \in W} q^{|D(w)|}$.

In the case $R = A_{n-1}$, $h_i(R)$ is thus the number of permutations in S_n with i descents. These are the classical Eulerian numbers and $qP_{A_{n-1}}(q)$ is the classical Eulerian polynomial. Although these numbers do not have simple explicit formulas, there is a well-known closed formula for the exponential generating function:

$$\sum_{n \geq 0} P_{A_{n-1}}(q) \frac{t^n}{n!} = \frac{(1-q)e^{(1-q)t}}{1 - qe^{(1-q)t}}.$$

Likewise for the B/C-series, there is a similar formula that is widely known (but seemingly unpublished):

$$\sum_{n \geq 0} P_{B_n}(q) \frac{t^n}{n!} = \frac{(1-q)e^{(1-q)t}}{1 - qe^{2(1-q)t}}.$$

For the D-series, it turns out that there is also a closed formula for the exponential generating function. It is an easy consequence of the following surprising relationship.

Proposition 1. $P_{B_n}(q) = P_{D_n}(q) + 2^{n-1} nq P_{A_{n-2}}(q)$.

We have two proofs, including a bijective one.

In another direction, the following result applies to all root systems except types D and E. In particular, it applies to the non-crystallographic cases $I_2(m)$, H_3, and H_4. We write $W_{[i,j]}$ for the parabolic subgroup of W generated by $\{s_i, s_{i+1}, \dots, s_j\}$.

Proposition 2. *If S can be linearly ordered so that s_i and s_j commute for $|i-j| > 1$ (i.e., the Dynkin diagram of R has no forks), then we have*

$$P_R(q) = |W| \cdot \det[a_{ij}]_{1 \leq i,j \leq n+1},$$

where $a_{ij} = 0$ for $i - j > 1$, $a_{ij} = 1 - q$ for $i - j = 1$, and $a_{ij} = 1/|W_{[i,j-1]}|$ for $i \leq j$.

In analyzing the character of $H^*(X_R, \mathbf{C})$, one needs to know the Poincaré polynomials of not just the Coxeter complexes Δ_R, but also the various fixed-point subcomplexes Δ_R^w (cf. the Dolgachev-Lunts formula). Usually these restrictions turn out to be Coxeter complexes of smaller rank, but not always. Among the classical cases, the only new simplicial complexes that arise in this way are (the complexes associated with) the following simplicial hyperplane arrangements:

$$\mathcal{H}_{k,n} = \{\varepsilon_i^\perp : 1 \le i \le k\} \cup \{(\varepsilon_i \pm \varepsilon_j)^\perp : 1 \le i < j \le n\} \qquad (0 \le k \le n).$$

Note that we recover the Coxeter complexes for D_n and B_n at $k = 0$ and n, respectively. The Poincaré polynomials $P_{n,k}(q)$ for these complexes satisfy a simple relationship generalizing Proposition 1.

Proposition 3. $P_{n,k}(q) = P_{D_n}(q) + 2^{n-1}kqP_{A_{n-2}}(q)$.

In the general case, it is possible to give explicit (but more complicated) formulas for the Poincaré polynomials of the fixed-point subcomplexes. Up to isomorphism, these complexes are indexed by $J \subset S$. More precisely, any Δ_R^w is conjugate under the Weyl group to some $\Delta_R^{w_J}$, where w_J denotes a Coxeter element for the parabolic subgroup W_J generated by some $J \subset S$. Using $P_{R,J}(q)$ to denote the corresponding Poincaré polynomial, the following result gives a formula for $P_{R,J}(1)$, the number of maximal faces in $\Delta_R^{w_J}$.

Proposition 4. $P_{R,J}(1) = \dfrac{|N(W_J)|}{|W_J|} \cdot |\{K \subset S : W_K \sim W_J\}|$, *where* $N(W_J)$ *denotes the normalizer of* W_J *in* W, *and* $\sim$ *denotes conjugacy of subgroups.*

References

[B] A. Björner, Some combinatorial and algebraic properties of Coxeter complexes and Tits buildings, *Adv. in Math.* **52** (1984), 173–212.

[D] V. I. Danilov, The geometry of toric varieties, *Russian Math. Surveys* **33** (1978), 97–154. Translated from *Uspekhi Mat. Nauk* **33** (1978), 85–134.

[DL] I. Dolgachev and V. Lunts, A character formula for the representation of a Weyl group in the cohomology of the associated toric variety, preprint.

[GS] A. M. Garsia and D. Stanton, Group actions of Stanley-Reisner rings and invariants of permutation groups, *Adv. in Math.* **51** (1984), 107–201.

[O] T. Oda, "Convex bodies and algebraic geometry," Springer-Verlag, Berlin-New York, 1985.

[St1] R. P. Stanley, Some aspects of groups acting on finite posets, *J. Combin. Theory Ser. A* **32** (1982), 132–161.

[St2] R. P. Stanley, Log-concave and unimodal sequences in algebra, combinatorics and geometry, *Ann. New York Acad. Sci.* **576** (1989), 500–535.

[Ste] J. R. Stembridge, Eulerian numbers, tableaux, and the Betti numbers of a toric variety, *Discrete Math.*, to appear.

Combinatorics of special functions : facets of Brock's identity

Volker Strehl
Computer Science Department - Informatik I
University of Erlangen-Nürnberg, Germany

Abstract

These notes are intended to give an overview of the relation between an old binomial identity, originally proposed as a problem by P. Brock, its various extensions and, in particular, its relation to combinatorial models for special functions that have been studied in more recent years. A comprehensive treatment of the results mentioned here is given in [33]. The list of references is selective and contains only articles mentioned in this overview.

1 Introduction

In these notes and in my talk I will try to explain the intimate link between two combinatorial results which, at first sight, appear quite unrelated.

- *Brock's identity* : in 1960 the following was presented by P. Brock [2] to the readers of the SIAM Review as problem 60-2:
  for integers $A, B \geq 0$ let

$$H(A,B) := \sum_{i=0}^{A}\sum_{j=0}^{B} \binom{i+j}{j}\binom{A-i+j}{j}\binom{B+i-j}{B-j}\binom{A-i+B-j}{B-j}. \tag{1}$$

  Then show that

$$H(A,B) - H(A-1,B) - H(A,B-1) = \binom{A+B}{A}^2 , \tag{2}$$

where $H(-1,B) = H(A,-1) = 0$.

This problem arose in the study of a sorting problem, or, to be a bit more precise, it reflects a particular way of counting permutations having exactly one increasing and one decreasing subsequence of maximal length. A rather involved inductive proof of the above identity is given in an article by R.M. Baer and P. Brock [1] on 'natural sorting'.

More on solutions and extensions of Brock's problem will be said in sec. 3 below.

- *PLI-endofunctions* : for any integer $p \geq 1$ and for any p-tuple $\mathbf{U} = (U_1, \ldots, U_p)$ of finite sets let ${}^{[p]}PLI[\mathbf{U}]$ denote the set of all functions $f : |\mathbf{U}| \to |\mathbf{U}|$, where $|\mathbf{U}| = U_1 \cup \ldots \cup U_p$, such that $f(U_i) \subseteq U_i \cup U_{i+1}$ and such that all the restrictions $f|_{U_i} : U_i \to U_i \cup U_{i+1}$ are injective maps ($1 \leq i \leq p$, reading indices modulo p).

Let a_i ($1 \leq i \leq p$) and β be parameters and put a weight

$$\prod_{1 \leq i \leq p} (1 + \alpha_i)^{cyc_i(f)} \cdot (1 + \beta)^{cyc_m(f)} \tag{3}$$

on each $f \in {}^{[p]}PLI[\mathbf{U}]$, where (for $1 \leq i \leq p$) $cyc_i(f)$ denotes the number of ('pure') f-cycles contained in U_i, and $cyc_m(f)$ denotes the number of the remaining ('mixed') f-cycles.

The exponential generating function associated to this p-sorted species of *periodic, locally injective endofunctions* ${}^{[p]}PLI(X_1, \ldots, X_p)$ under this weight can then be written in two different ways:
First we have

$$\frac{\prod_{1 \leq i \leq p} (1 + \xi_i(\mathbf{x}))^{1+\alpha_i}}{\left(1 - \prod_{1 \leq i \leq p} \xi_i(\mathbf{x})\right)^{1+\beta}} \quad , \tag{4}$$

where $\mathbf{x} = (x_1 \ldots x_p)$ and where the functions $\xi_j = \xi_j(\mathbf{x})$ are determined by the 'cyclic' implicit system

$$\xi_j = x_j \cdot (1 + \xi_{j-1}) \cdot (1 + \xi_j) \qquad (1 \leq j \leq p) \ , \tag{5}$$

again reading indices modulo p.
On the other hand we get

$$\sum_{\mathbf{n}} \frac{\mathbf{x^n}}{\mathbf{n}!} \prod_{1 < i < p} (1 + \alpha_i + n_{i+1})_{n_i} \cdot {}_{p+1}F_p \left[\begin{matrix} \beta \, , \, -n_1 \ldots - n_p \\ \ldots 1 + \alpha_j + n_{j+1} \ldots \end{matrix} ; (-1)^p \right] =$$

$$= \sum_{\mathbf{n}} \mathbf{x}^{\mathbf{n}} \prod_{1 \le i \le p} \binom{\alpha_i + n_i + n_{i+1}}{n_i} \cdot {}_{p+1}F_p[\cdots] \ , \tag{6}$$

where the summation runs over all $\mathbf{n} = (n_1, \ldots, n_p) \in \mathbf{N}^p$ and $\mathbf{x}^{\mathbf{n}} = x_1^{n_1} \ldots x_p^{n_p}$, $\mathbf{n}! = n_1! \ldots n_p!$.

This result, which reflects on the generating function level two different combinatorial views of the underlying structures is, in fact, the specialization of something quite more general. Comments on the general situation will be made in sec. 4. Some consequences of this result will be presented in the next section. The relation between this kind of result and Brock's identity and its various extensions will be described in sec. 5.

2 Some consequences

In this section I will briefly discuss several simple, yet interesting particular cases of the result on generating functions for PLI-endofunctions just mentioned.

- in the case $p = 1$ and $\alpha = \beta$ we find

$$\sum_{n \ge 0} (1 + \beta + n)_n \, {}_2F_1 \left[\begin{matrix} \beta \, , \, -n \\ 1 + \beta + n \end{matrix} ; -1 \right] \frac{x^n}{n!} = \left[\frac{1 + \xi}{1 - \xi} \right]^{1+\beta} \ , \tag{7}$$

where $1 + \xi = (1 - \sqrt{1 - 4x})/(2x)$, and hence

$$\frac{1 + \xi}{1 - \xi} = (1 - 4x)^{-1/2}$$

from solving the implicit equation (5). Comparison of coefficients on both sides of (7) then shows

$${}_2F_1 \left[\begin{matrix} \beta \, , \, -n \\ 1 + \beta + n \end{matrix} ; -1 \right] = \frac{\Gamma(1 + \beta + n)\Gamma(1 + \frac{\beta}{2})}{\Gamma(1 + \frac{\beta}{2} + n)\Gamma(1 + \beta)} \ ,$$

which is Kummer's formula, cf. [24].

- in the case $p=2, \beta=0$ we get the classical generating function for the Jacobi polynomials (writing now (α,β) in place of (α_1,α_2)) :

$$\begin{aligned}
&\sum_{k,m\geq 0} \frac{x^k}{k!}\frac{y^m}{m!}(1+\alpha+m)_k(1+\beta+k)_m = \\
&= \frac{(1+\xi_1(x,y))^{1+\alpha}\,(1+\xi_2(x,y))^{1+\beta}}{1-\xi_1(x,y)\,\xi_2(x,y)} \\
&= \left(\frac{1-x+y-\mathcal{R}}{2y}\right)^{\alpha}\left(\frac{1+x-y-\mathcal{R}}{2x}\right)^{\beta}\mathcal{R}^{-1} \quad ,
\end{aligned} \tag{8}$$

where

$$\xi_1(x,y)=\frac{1-x-y-\mathcal{R}(x,y)}{2y} \quad , \quad \xi_2(x,y)=\frac{1-x-y-\mathcal{R}(x,y)}{2x} \quad ,$$

and

$$\mathcal{R}=\sqrt{[1-x-y]^2-4xy} \quad .$$

Note that

$$P_n^{(\alpha,\beta)}(x)=\sum_{k+m=n}\frac{(1+\alpha+m)_k}{k!}\frac{(1+\beta+k)_m}{m!}\left(\frac{x+1}{2}\right)^k\left(\frac{x-1}{2}\right)^m$$

is one of the various ways of writing the Jacobi polynomials, see [24].

Note furthermore that the combinatorial model of PLI-endofunctions reduces in this case to the model introduced by D. Foata and P. Leroux in [15], a model which has subsequently turned out to be very fruitful for the study of combinatorial properties of the Jacobi polynomials, see e.g. [23].

- in the case $p=2, \alpha_1=\alpha_2=\beta$, we can evaluate the ${}_3F_2$-term in (6) via Dixons formula [12], [24]:

$$\begin{aligned}
&{}_3F_2\left[\begin{matrix}\beta\ ,\ -n_1\ ,\ -n_2 \\ 1+\beta+n_1\ ,\ 1+\beta+n_2\end{matrix};1\right] = \\
&= \frac{\Gamma(1+\frac{\beta}{2})\Gamma(1+\beta+n_1)\Gamma(1+\beta+n_2)\Gamma(1+\frac{\beta}{2}+n_1+n_2)}{\Gamma(1+\beta)\Gamma(1+\frac{\beta}{2}+n_1)\Gamma(1+\frac{\beta}{2}+n_2)\Gamma(1+\beta+n_1+n_2)}
\end{aligned}$$

which after simplification leads to the classical generating function for the Gegenbauer (ultrasherical) polynomials

$$P_n^{\nu}(x) = \frac{(2\nu)_n}{(\nu+\frac{1}{2})_n} P_n^{(\nu-\frac{1}{2},\nu-\frac{1}{2})} \quad ,$$

namely:

$$\sum_{n\geq 0} P_n^{\nu}(x)\, z^n = (1-2xz+z^2)^{-\nu} \quad .$$

This generating function cannot be obtained from the Foata-Leroux model mentioned above. Instead, a completely different approach for that purpose, using the model of 'complete oriented matchings', was presented by myself at the Montreal colloquium in 1985, see [29]. Putting this latter approach together with the combinatorial (!) proof of the result on PLI-endofunctions mentioned above, one obtains - as a by-product - another combinatorial proof of Dixon's formula.

3 On proofs and extensions of Brock's identity

It has been mentioned above that P. Brock himself (together with R.M. Baer) in [1] gave a rather involved inductive proof of (2). A better way of attacking the problem combinatorially has been pointed out independently by C.A. Church, jr. in [8] and myself in [32], where an appropriate class of configurations (pairs of crossing lattice paths or marked permutations) is examined. Both approaches turn out to be 'bijectively equivalent' via the Robinson-Schensted correspondence. This 'geometric' way of proving Brock's identity can also be used to obtain almost 'visually' various extensions that appear in the literature together with laborious analytic proofs (e.g. [9], [10]).

More interesting for us, however, is the kind of proof that has been initiated by D. Slepian's original analytic solution [27] of Brock's problem, using generating functions and complex integration. This kind of approach was later simplified and largely extended by L. Carlitz in a series of articles, beginning with [3].

L. Carlitz first proves

$$\sum_{i,j,k,l\geq 0} \binom{i+j}{j}\binom{j+k}{k}\binom{k+l}{l}\binom{l+i}{i} u^i v^j w^k x^l =$$
$$\left([(1-v)(1-x)-w+u(1-w)]^2 - 4u(1-v-w)(1-w-x)\right)^{-1/2} ,$$

from which he gets

$$\sum_{m,n\geq 0} H(m,n)\; u^m v^n = \frac{1}{1-u-v} \sum_{m,n\geq 0} \binom{m+n}{n}^2 u^m v^n \tag{9}$$

by identification of variables $u = w, v = x$. Note that (9) is obviously the generating function equivalent of Brock's original assertion (2).

Carlitz then proceeds to consider generating functions for 'cyclic products of binomial coefficients'

$$\binom{n_1+n_2}{n_2}\binom{n_2+n_3}{n_3}\cdots\binom{n_r+n_1}{n_r}$$

in general, and he obtains various identities of Brock-type for the numbers

$$H(n_1, n_2, \ldots, n_r) := \sum_{\substack{i_s+i_{r+s}=n_s \\ 1\leq s\leq r}} \binom{i_1+i_2}{i_2}\binom{i_2+i_3}{i_3}\cdots\binom{i_{2r}+i_1}{i_1} . \tag{10}$$

To give an idea of his results, I just state the two simplest ones:
case $r = 3$:

$$H(m,n,p) - H(m-1,n,p) - H(m,n-1,p) - H(m,n,p-1) = \binom{m+n}{n}\binom{n+p}{p}\binom{p+m}{m} \tag{11}$$

case $r = 4$:

$$\begin{aligned} &H(n_1,n_2,n_3,n_4) - \\ &H(n_1-1,n_2,n_3,n_4) - H(n_1,n_2-1,n_3,n_4) - \\ &H(n_1,n_2,n_3-1,n_4) - H(n_1,n_2,n_3,n_4-1) + \\ &H(n_1-1,n_2,n_3-1,n_4) + H(n_1,n_2-1,n_3,n_4-1) \\ &= \binom{n_1+n_2}{n_2}\binom{n_2+n_3}{n_3}\binom{n_3+n_4}{n_4}\binom{n_4+n_1}{n_1} \end{aligned} \tag{12}$$

Another line of generalization is opened by Carlitz in [4] by introducing extra parameters, i.e. by considering

$$H(n_1,\ldots,n_r|\alpha_1,\ldots,\alpha_{2r}) = \sum_{\substack{i_s+i_{r+s}=n_s \\ 1\le s\le r}} \binom{\alpha_1+i_1+i_2}{i_1}\binom{\alpha_2+i_2+i_3}{i_2}\cdots\binom{\alpha_{2r}+i_{2r}+i_1}{i_{2r}}\ . \tag{13}$$

Here (in the case $r = 2$) Brock's identity generalizes to

$$\begin{aligned} H(m,n|\alpha,\beta,\gamma,\delta) - H(m-1,n|\alpha,\beta,\gamma,\delta) - H(m,n-1|\alpha,\beta,\gamma,\delta) \\ = \binom{\alpha+\gamma+m+n}{m}\binom{\beta+\delta+m+n}{n} \end{aligned} \tag{14}$$

(As an aside: this is one of the extensions of Brock's identity which can be readily obtained from the 'geometric' approach mentioned above.) It is interesting to note that that for this generalization Carlitz starts from the classical generating function for the Jacobi polynomials (8).

A final word in this direction has been said by J.P. Singhal, who in [26] gives generating functions for cyclic product of parametrized binomial coefficients, which in turn can be used to obtain Brock-Carlitz type identites for the numbers $H(n_1,\ldots,n_r|\alpha_1,\ldots,\alpha_{2r})$.

One should note that all these results just mentioned, which are also (at least partially) treated in Riordan's [25] and Egorychev's [13] books, were obtained by rather involved calculations on formal series, without reference to any kind of combinatorial structures. This makes it difficult to digest the results in their general form and to notice the common pattern underlying these complicated looking generating functions and coefficient identities. A combinatorial view explains much of the mystery and opens the way for further extensions.

4 PLI-endofunctios and locally structured endofunctions

The result on the generating functions for periodic, locally injective endofunctions mentioned at the beginning of this note is indeed a specialization of a much more general result about the generating functions of combinatorial structures which I call *locally structured endofunctions.* Loosely speaking, given any p-tuple $\mathbf{A} = (A^{(1)}, \ldots, A^{(p)})$ of p-sorted species we may define the p-sorted species of $\mathbf{A}$-endofunctions by associating with each p-tuple $\mathbf{U} = (U_1, \ldots, U_p)$ of finite sets the set of all $\langle f, (a_u)_{u \in |\mathbf{U}|} \rangle$, where $f : |\mathbf{U}| \to |\mathbf{U}|$ is a mapping, and for each vertex $u \in U_i$ we require that

$$a_u \in A^{(i)}[f^{-1}(u) \cap U_1, \ldots, f^{-1}(u) \cap U_p] \quad (1 \leq i \leq p) \quad ,$$

i.e. there is an $A^{(i)}$-structure associated with the f-preimage of each element $u \in |\mathbf{U}|$, depending on the sort i of vertex u.

A cycle weight may be associated with such objects, as has been done in (3) for the particular case of PLI-endofunctions. In that particular situation

$$A^{(i)} = (1 + X_i) \cdot (1 + X_{i-1}) \quad (1 \leq i \leq p) \quad ,$$

where X_i is simply the species 'vertices of sort i' (again reading indices modulo p). The concept of locally structured endofunctions allows for the specification of a wide variety of combinatorial structures which can be defined characterized by 'local' conditions. In particular, most of the models that were introduced for the study of classical orthogonal polynomials from a combinatorial point of view belong to this class, see [14], [21], [22], [23], for example.

For any species of $\mathbf{A}$-endofunctions the generating function (taking cycle weights into account) may be presented in two different ways

- by viewing $\mathbf{A}$-endofunctions as permutations of $\mathbf{A}$-contractions - which yields an expression in terms of the generating function of the implicitly defined species of $\mathbf{A}$-trees. See (4,5) for this type of result in the particular case of PLI-endofunctions.

- by constructing $\mathbf{A}$-endofunctions (or more generally: partial $\mathbf{A}$-endofunctions) by using a combinatorial differential operator. This leads to an expression which contains no implicitly defined functions - at the expense of using a certain diagonalization operator. In the particular case of PLI-endofunctions this version of the generating function can be made explicit and leads to (6) above.

It should be mentioned that the first way of presenting the generating function owes much to and indeed generalizes the combinatorial approach to multivariable Lagrange inversion as proposed by I. Gessel [16] and further extended and simplified by J. Zeng [34]. The second way comes from a combinatorial interpretation of Hurwitz' parametrized Lagrange inversion formula [19], appropriately generalized to a multivariable and cycle-weighted situation.

The comparison of both ways of writing the generating function for A-endofunctions gives a result (not reproduced here) which contains many interesting special cases of varying degree of generality, e.g

- I. Gessel's multivariate Lagrange inversion formula and J. Zeng's β-generalization of it, as already mentioned;
- Joni's 'general formula' [20];
- the series expansions given by L. Carlitz e.g. in [5],[6] and [7];
- various generalizations and variants of the Pfaff-Saalschütz formula, due (with different proofs, analytical and combinatorial) to I. Gessel, D. Stanton, D. Sturtevant, I. Constantineau and myself ([17], [18], [11]);
- generating functions for the Jacobi polynomials with a linear shift in the parameters, due to H.M. Srivastava and J.P. Singhal in [28], and already treated from a combinatorial perspective by myself in [31].

All these results can be further generalized by making use of the weight function put on f-cycles in full generality.

5 PLI-endofunctions and Brock-Carlitz type identities

The most interesting application (from my point of view) of the general theory just sketched deals with the situation of PLI-endofunctions. In this case both versions of the generating function can be made completely explicit. Indeed, the implicit system (5) for the p tree generating functions $\xi(\mathbf{x}) = \xi_i(x_1, \ldots, x_p)$ $(1 \leq i \leq p)$ can be solved in terms of matching polynomials (which are, to be a bit more specific, multivariable analogues of the Chebychev polynomials). The appearance of matching polynomials in this context appears quite natural from the kind of combinatorial objects considered.

In order to state the central result, let us put for $\mathbf{n} = (n_1, \ldots, n_p) \in \mathbf{N}^p$, parameters $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)$ and β

$$
{}^{[p]}H(\mathrm{n}|\alpha;\beta) = \qquad (15)
$$
$$
\prod_{1\le i\le p} \binom{\alpha_i + \beta + n_i + n_{i-1}}{n_i} \; {}_{p+1}F_p \left[\begin{matrix} \beta \; - n_1 \ldots - n_p \\ \ldots 1 + \alpha_j + \beta + n_{j+1} \ldots \end{matrix} ; (-1)^p \right] .
$$

Then the generating function for the species of p-sorted PLI-endofunctions ${}^{[p]}PLI(X_1, \ldots, X_p)$ (with a slight variation of the weight function) can be written as

$$
{}^{[p]}PLI^{(\alpha;1+\beta)}(\mathrm{x}) = \sum_{\mathbf{n}} {}^{[p]}H(\mathrm{n}|\alpha;\beta)\; \mathrm{x}^{\mathbf{n}} \quad .
$$

On the other hand

$$
\begin{aligned}
{}^{[p]}PLI^{(\alpha;\beta)}(\mathrm{x}) &= \frac{\prod_{1\le i\le p}(1+\xi_i)^{\alpha_i+\beta}}{\left[1-\prod_{1\le i\le p}\xi_i(\mathrm{x})\right]^{\beta}} \qquad (16) \\
&= \frac{\prod_{1\le i\le p}(1+\xi_i)^{\alpha_i}}{\left[c_p(\mathrm{x}) - 4\prod_{1\le i\le p} x_i\right]^{\beta/2}} \quad ,
\end{aligned}
$$

where the $\xi_i = \xi_i(\mathrm{x})$ are as in (5) above and $c_p(\mathrm{x}) = c_p(x_1, \ldots, x_p)$ is a matching polynomial for cycles of length p with weights $x_1, \ldots x_p$ put on the edges.

Here I will not give the explicit form of the functions ξ_i (which can be expressed in terms of $c_p(\mathrm{x})$ and a similar matching polynomial $l_{p-1}(\mathrm{x})$ for a line of length $p-2$). The precise knowledge is not even necessary because in order to obtain Brock-Carlitz type identities knowledge of the denominator in the above presentation is sufficient.

Now let $r = p \cdot q$, take parameters $\boldsymbol{\alpha} = (\alpha_1, \ldots \alpha_r)$ and β. Denote by ${}^{[p,q]}PLI(X_1, \ldots, X_p)$ the p-sorted species obtained from the r-sorted species ${}^{[r]}PLI(X_1, \ldots, X_r)$ by identifying sorts of vertices modulo p, i.e.

$$
X_i \equiv X_j \;\text{ iff }\; i \equiv j \bmod p \quad .
$$

Accordingly, let $\tilde{\boldsymbol{\alpha}} = (\tilde{\alpha}_1, \ldots, \tilde{\alpha}_p)$ with

$$
\tilde{\alpha}_j = \sum_{0\le i<q} \alpha_{j+i\cdot p} \quad (1 \le j \le p) \quad ,
$$

and similarly, for each $\mathbf{m} = (m_1, \ldots, m_r) \in \mathbf{N}^r$ let $\widetilde{\mathbf{m}} = (\widetilde{m}_1, \ldots, \widetilde{m}_p) \in \mathbf{N}^p$ with

$$\widetilde{m}_j = \sum_{0 \le i < q} m_{j+i\cdot q} \quad (1 \le j \le p) \ .$$

We know that the generating function for ${}^{[p,q]}PLI(X_1, \ldots, X_p)$ can be written as

$$^{[p,q]}PLI^{(\alpha,1+\beta)}(\mathbf{x}) = \sum_{\mathbf{n} \in \mathbf{N}^p} {}^{[p,q]}H(\mathbf{n}|\alpha;\beta)\, \mathbf{x}^{\mathbf{n}} \ , \tag{17}$$

where

$$^{[p,q]}H(\mathbf{n}|\alpha;\beta) = \sum_{\widetilde{\mathbf{m}}=\mathbf{n}} {}^{[r]}H(\mathbf{m}|\alpha;\beta) \ ,$$

the sum on the right running over all $\mathbf{m} \in \mathbf{N}^r$ such that $\widetilde{\mathbf{m}} = \mathbf{n}$.

Note that for $q = 2, \beta = 0$ we obtain the Brock numbers $H(m,n)$ in (1) as well as Carlitz' generalizations (10),(13) mentioned in sec. 3.

With all the combinatorial machinery behind our structures properly put into action (in particular, duplication and composition properties of the matching polynomials play a rôle), one obtains from (15) the following identity for generating functions:

$$^{[p,q]}PLI^{(\alpha,\beta)}(\mathbf{x}) = \frac{1}{\lambda_{q-1}(\prod_i x_i, c_p(\mathbf{x}))^\beta}\ {}^{[p]}PLI^{(\widetilde{\alpha},\beta)}(\mathbf{x}) \ , \tag{18}$$

where $\lambda_{q-1}(u,v)$ is another matching polynomial; indeed,

$$\lambda_n(u^2, v) = u^n \cdot U_n(v/2u) \ ,$$

where U_n is the classical Chebychev polynomial.

In order to see how this extends the Brock-Carlitz identities, consider now the case $q = 2$ and put $\beta = 1$ in (18) (which means taking $\beta = 0$ in (15) and (17)). We have $\lambda_1(u,v) = v$, so that the denominator in (18) (for $\beta = 1$) is nothing but $c_p(x_1, \ldots, x_p)$. ¿From the combinatorial signification it is obvious that

$$\begin{aligned} c_2(x_1,x_2) &= 1 - x_1 - x_2 \ , \\ c_3(x_1,x_2,x_3) &= 1 - x_1 - x_2 - x_3 \ , \\ c_4(x_1,x_2,x_3,x_4) &= 1 - x_1 - x_2 - x_3 - x_4 + x_1x_3 + x_2x_4 \ , \end{aligned}$$

which brings us back to (2) (or (9) resp.), (11) and (12).

It can be shown that the general 'rule' for obtaining Brock type identities in the case $q = 2$, established by Carlitz and extended by Singhal in the parametrized situation, is just a cryptic way of stating that the matching polynomial $c_p(x_1, \ldots, x_p)$ appears as denominator polynomial in an identity like (18) relating the appropriate generating functions.

To mention just the simplest case of a binomial identity not covered by previous results, let us put $q = 3$ and $p = 2$. We have $\lambda_2(u, v) = v^2 - u$ and $c_2(x_1, x_2)$ as above. This leads to

$$\lambda_2(x_1x_2, c_2(x_1, x_2)) = 1 - 2x_1 - 2x_2 + x_1^2 + x_1x_2 + x_2^2 \ ,$$

which shows that the numbers

$$h(m,n) = \sum_{\substack{i_1+i_3+i_5=m \\ i_2+i_4+i_6=n}} \binom{i_1+i_2}{i_1}\binom{i_2+i_3}{i_2}\cdots\binom{i_6+i_1}{i_6}$$

satisfy

$$\begin{aligned} & h(m,n) - 2h(m-1,n) - 2h(m,n-1) \\ & +h(m-2,n) + h(m-1,n-1) + h(m,n-2) = \binom{m+n}{n}^2 \ . \end{aligned}$$

Obviously, many more results can be pulled out of this general approach. The 'opposite' situation to the Brock-Carlitz case $p = 2$, namely $q = 2$ and p arbitrary is of particular interest due to its intimate relation to the Jacobi polynomials. In particular, the classical generating function for the Jacobi polynomials can be presented as the limit of a rapidly converging (in the sense of formal series) sequence of rational generating functions which are, in fact, quotients of matching polynomials, see [30] for a direct approach.

References

[1] R.M. Baer and P. Brock. Natural sorting. *SIAM Journal*, 10:284–304, 1962.

[2] P. Brock. On a binomial identity arising from a sorting problem. *SIAM Review*, 2:40, 1960. problem 60-2.

[3] L. Carlitz. A binomial identity arising from a sorting problem. *SIAM Review*, 6:20–30, 1964.

[4] L. Carlitz. Some generating functions and a binomial identity. *Duke Math. J.*, 35:541–548, 1968.

[5] L. Carlitz. An application of MacMahon's master theorem. *SIAM J. Appl. Math.*, 26:431–436, 1974.

[6] L. Carlitz. Some expansion and convolution formulas related to MacMahon's master theorem. *SIAM J. Math. Anal.*, 8:320–336, 1977.

[7] L. Carlitz. Multiple binomial and power sums. *Houston J. Math.*, 6:331–354, 1980.

[8] C.A. Church, jr. Combinatorial proof for a sorting problem identity. *Fibonacci Quart.*, 33:366–368, 1985.

[9] M.E. Cohen. Some extensions of Carlitz on cyclic sums and generating functions. *J. Math. Anal. Appl.*, 60:493–501, 1977.

[10] M.E. Cohen and H.S. Sun. Some extensions of the Brock-Carlitz identity. *Proc. Amer. Math. Soc.*, 76:178–185, 1979.

[11] I. Constantineau. A generalization of the Pfaff-Saalschütz theorem. *Studies in Appl. Math.*, 85:243–248, 1991.

[12] A.C. Dixon. Summation of a certain series. *Proc. London Math. Soc.*, 35:284–289, 1903.

[13] G.P. Egorychev. *Integral Representations and the Computation of Combinatorial Sums*, volume 59 of *Translations of Mathematical Monographs.* Amer. Math. Soc., Providence, R.I., 1984.

[14] D. Foata and J. Labelle. Modèles combinatoires pour les polynômes de Meixner. *Europ. J. Combinatorics*, 4:305–311, 1983.

[15] D. Foata and P. Leroux. Polynômes de Jacobi, interprétation combinatoire et fonction génératrice. *Proc. Amer. Math. Soc.*, 87:47–53, 1983.

[16] I. Gessel. A combinatorial proof of the multivariable Lagrange inversion formula. *J. Combin. Theory, series A*, 45:178–195, 1987.

[17] I. Gessel and D. Stanton. Short proofs of Saalschütz and Dixon's theorem. *J. Combin. Theory, series A*, 38:87–90, 1985.

[18] I. Gessel and D. Sturtevant. A combinatorial proof of Saalschütz' theorem. preprint, 1989.

[19] A. Hurwitz. Über einige Verallgemeinerungen der Leibniz'schen Differentiationsformel und des polynomischen Lehrsatzes. *Z. Math. Phys.*, 35:56–58, 1890.

[20] S.A. Joni. Expansion formulas I. A general method. *J. Math. Anal. Appl.*, 81:364–377, 1981.

[21] J. Labelle and Y.-N. Yeh. A combinatorial model for Hahn polynomials. In D. Foata, editor, *Actes Séminaire Lotharingien de Combinatoire , 16 session*, pages 97–107. Publications de l' IRMA, Strasbourg, 341/S–16, 1987.

[22] J. Labelle and Y.-N. Yeh. The combinatorics of Laguerre, Charlier and Hermite polynomials. *Studies in Appl. Math.*, 80:25–36, 1989.

[23] P. Leroux and V. Strehl. Jacobi polynomials: combinatorics of the basic identities. *Discrete Mathematics*, 57:167–187, 1985.

[24] E. Rainville. *Special Functions.* Chelsea, Bronx, N.Y., 1971.

[25] J. Riordan. *Combinatorial Identities.* J. Wiley, New York, 1968.

[26] J.P. Singhal. A generating function and a binomial identity. *Duke Math. J.*, 39:25–30, 1972.

[27] D. Slepian. Solution to problem 60-2. *SIAM Review*, 4:396–398, 1962.

[28] H.M. Srivastava and J.P. Singhal. New generating functions for Jacobi and related polynomials. *J. Math. Anal. Appl.*, 41:725–748, 1972.

[29] V. Strehl. Combinatorics of Jacobi-configurations I : Complete oriented matchings. In G. Labelle et al., editor, *Combinatoire Énumérative, Montréal 1985*, pages 294–307, Berlin, 1986. Springer-Verlag. (*Lecture Notes in Math., vol. 1234*).

[30] V. Strehl. Combinatorics of Jacobi-configurations II : A rational approximation via matching polynomials. In G. Nicoletti, editor, *Actes Séminaire Lotharingien de Combinatoire , 13e session*, pages 112–133. Publ. IRMA Strasbourg, 316/S–13, 1986.

[31] V. Strehl. Combinatorics of Jacobi-configurations III : The Srivastava-Singhal generating relation revisited. *Discrete Math.*, 73:221–232, 1988.

[32] V. Strehl. Endliche Funktionen und spezielle Funktionen : eine kombinatorische Theorie zu einer Identität von P. Brock und deren Erweiterungen. In D. Foata, editor, *Actes Séminaire Lotharingien de Combinatoire , 17e session*, pages 23–67. Publ. IRMA Strasbourg, 348/S–17, 1988.

[33] V. Strehl. *Zykel-Enumeration bei lokal-strukturierten Funktionen.* Universität Erlangen-Nürnberg, 1990.

[34] J. Zeng. La β-extension de la formule d'inversion de Lagrange à plusieurs variables. In L. Cerlienco and D. Foata, editors, *Actes Séminaire Lotharingien de Combinatoire, 20e session*, pages 23–38, Strasbourg, 1988. Publications de l' I.R.M.A., 372/S-20.

The homology representation of the symmetric group on Cohen-Macaulay subposets of the partition lattice

Extended Abstract

Sheila Sundaram

Université du Québec à Montréal
Montréal H3C 3P8, Québec, Canada

0. Introduction

The action of the symmetric group S_n on the top homology $\tilde{H}_{n-3}(\Pi_n)$ of the partition lattice Π_n has been studied from various viewpoints in recent years, beginning with a character computation by Hanlon ([H1]). Following questions raised by Stanley ([St]), in this paper we consider the S_n-representation on the homology of certain Cohen-Macaulay subposets of Π_n. We present a general technique for manipulating these homology modules. The unique properties of the partition lattice allow further simplification of these formulas, culminating in plethystic generating functions which, by recursive computation, yield the Frobenius characteristic of the representation. We illustrate our technique by giving simple derivations of three known formulas:

1. a formula for the plethystic inverse of the sum of the cycle indicators of the symmetric groups; this is essentially equivalent to Cadogan's formula ([C]);
2. a plethystic formula which determines the characteristic of the homology representation on the lattice Π_n^d of partitions of a set of size nd all of whose blocks are multiples of d; this formula was originally derived in a paper of Calderbank, Hanlon and Robinson ([CHR]);
3. a plethystic generating function identity, also due to these authors, which determines the homology representation on the subposet $\Pi_n^{(i,d)}$ of partitions of n elements into blocks of size congruent to i mod d.

The idea behind our basic observation, Lemma 2.1, is implicitly used to calculate the Möbius functions of fixed-point posets by Calderbank, Hanlon and Robinson ([CHR], Theorem 2.2). Our approach has the advantage of being more conceptual from the representation-theoretic point of view: we avoid intricate Möbius function computations by manipulating virtual homology modules, and by exploiting the considerable machinery of symmetric function theory as presented by Macdonald.

Our main result (Theorem 1.12) is a recursive formula for the characteristic of the homology representation on an arbitrary rank-selected subposet of Π_n. This formula is easily adapted to handle rank-selection in the posets Π_n^d and $\Pi_n^{(1,d)}$. In particular, our methods give elegant plethystic formulas for the characteristic of the homology representation on the following rank-selected subposets of Π_n :

1. when the ranks are selected in arithmetic progression: in particular for initial or final segments of consecutive ranks;
2. when the subposet is obtained by deleting a single rank from Π_n;
3. when the subposet consists of arbitrarily placed segments of consecutive ranks in Π_n;
4. when the subposet is as in 3. but with one rank deleted.

Using these formulas we can extract a considerable amount of concrete information about the representation.

The same idea allows us to give recursive formulas for the permutation representation on the rank-selected chains of Π_n. In particular we determine completely the orbit decomposition of the representation on the maximal chains of Π_n. The multiplicities we obtain turn out to enumerate a known refinement of the Euler numbers. In fact we show that the action of S_n on the maximal chains is closely related to an action of S_2^{n-1} on the set of $n!$ permutations, defined by Foata and Strehl ([FS1], [FS2]) in their enumerative work on the Euler numbers.

A result of Stanley ([St]) states that when all the rank-selected homologies of the partition lattice are combined into one S_n-module, the multiplicity of the trivial representation in this direct sum is the Euler number E_{n-1} of alternating permutations in S_{n-1}. Our computations enable us to describe combinatorially the multiplicity of the trivial representation in the homology module in some specific cases of rank-selection, although it is as yet unclear how these multiplicities enumerate a refinement of the alternating permutations. By examining the restrictions of these S_n-homology modules to S_{n-1}, we obtain a second (different) refinement of the Euler number E_n into nonnegative integers also indexed by subsets.

Our methods apply to the posets Π_n^d and $\Pi_n^{(i,d)}$ as well. As a by-product of our study we show that one can associate to these posets a vector-space structure which strikingly resembles that of the Orlik-Solomon algebra for Π_n.

Since we are interested primarily in group actions, all homology will be taken over the complex field.

1. Results

Our results are described most conveniently by using the theory of symmetric functions (see [M]). Recall that the complete homogeneous symmetric function h_n is the Frobenius characteristic of the trivial representation of the symmetric group S_n, while the elementary symmetric function e_n is the characteristic of the sign representation. We shall write π_n for the characteristic of the top homology module $\tilde{H}_{n-3}(\Pi_n)$ of the partition lattice. Following [M], denote by ω the involution on the ring of symmetric functions which takes h_n to e_n. A well known result on the homology of Π_n, which follows from Hanlon's computation of the character values of π_n and earlier results of Witt and Brandt, states that $\pi_n = \omega(\ell_n)$, where ℓ_n is the characteristic of the S_n-representation on the nth graded component of the free Lie algebra. Joyal gives a direct functorial proof of this result in [J].

The plethysm operation occurs naturally in connection with the partition lattice; for symmetric functions f and g we write $f[g]$ for the plethysm of f with g.

We begin by looking at particular cases of rank-selection. The partition lattice is known to be Cohen-Macaulay, and this property is preserved by rank-selection. The problem is therefore to determine the S_n-module structure of the unique non-vanishing reduced homology (in the highest degree) of the rank-selected subposet.

Write $\Pi_n(\underline{r})$ for the subposet consisting of the first r ranks of Π_n, excluding the 0-element at rank 0. Also, for any Cohen-Macaulay poset P, we generally suppress the homology degree and write $\tilde{H}(P)$ for the (reduced) top homology of P. Denote by $V_n(\underline{r})$ the S_n-representation on the top homology of $\Pi_n(\underline{r})$. (Define $V_n(\underline{0}) = h_n$; if $r > n-2$ or $r < 0$, set $V_n(\underline{r}) = 0$.)

Theorem 1.1

1. The characteristic of the representation $V_n(\underline{r})$ is given by the degree n term in the plethysm

$$(-1)^r(h_{n-r} + h_{n-r+1} + \ldots + h_n)[\pi_1 - \pi_2 + \ldots + (-1)^{n-1}\pi_n].$$

2. Using down and up arrows to indicate respectively restriction and induction of modules, we have the following recursive property relating the S_n-module $V_n(\underline{r})$ to its structure as an S_{n-1}-module:

$$V_n(\underline{r})\downarrow_{S_{n-1}} \simeq V_{n-1}(\underline{r}) \oplus V_{n-1}(\underline{r-1})\downarrow_{S_{n-2}}\uparrow^{S_{n-1}},$$

for all $r = 1, \ldots, n-2$.

3. Let σ be a permutation in S_n of type $\prod_i i^{m_i}$, that is, with m_i cycles of length i. The generating function for the character values $tr(\sigma)$ of σ on $V_n(\underline{r})$, $0 \le r \le n-2$, is

$$\sum_{r=0}^{n-2} tr(\sigma)\big|_{V_n(\underline{r})} u^r = \frac{1}{1+u} \prod_i \prod_{k=0}^{m_i-1} \left(\sum_{d|i} \mu(d) u^{i-i/d} - ki \right).$$

The expression $(1+u) \sum_{r=0}^{n-2} tr(\sigma)\big|_{V_n(\underline{r})} u^r$ is computed in [OS] (p. 183, Example 4.10) for small values of n. Note that this is essentially a kind of "characteristic polynomial "of the subposet of Π_n fixed by σ. Also note that in general the factors in the numerator of the right hand side are not linear.

Next let $V_n(\bar{r})$ denote the subposet obtained by selecting the top r ranks of Π_n, excluding the top element of rank $(n-1)$. The ranks chosen are thus $n-1-r, n-r, \ldots, n-3, n-2$. We have

Theorem 1.2

1. The characteristic of the homology representation of $V_n(\bar{r})$ is given by the degree n term in the plethysm

$$(\pi_{r+1} - \pi_r + \ldots + (-1)^r \pi_1)[h_1 + h_2 + \ldots + h_n].$$

2. The restriction of $V_n(\bar{r})$ to S_{n-1} is a permutation module, whose orbit decomposition is given by the characteristic

$$\sum_{\substack{\lambda \vdash (n-1), \ell(\lambda) = r+1 \\ \lambda = 1^{m_1} 2^{m_2} \ldots}} \binom{r+1}{m_1, m_2, \ldots} h_\lambda,$$

where the sum runs over all partitions λ of $n-1$ of length $r+1$, and m_i denotes the multiplicity of the part i in λ.

Now let $d \ge 2$, and let R_k denote the characteristic of the S_k-representation on the top homology of the subposet of Π_k obtained by selecting all ranks $i \ge 1$ such that $k - i$ is a multiple of d. (Recall that the rank of a partition x in Π_n is n minus the number of blocks in x; the subposet therefore consists of all partitions with *number* of blocks divisible by d). For $1 \le i \le d-1$, define $R_i = h_i$. Note that $R_d = h_d$, the poset in this case being empty.

The following plethystic identity completely determines the characteristics R_k :

Theorem 1.3 Assume $d \ge 2$. Then

$$\sum_{n \ge 0} (-1)^n \sum_{i=1}^{d-1} R_{nd+i} = (h_1 - R_d + R_{2d} - \ldots)[h_1 + h_2 + \ldots].$$

Let $Q_k(i)$ be the characteristic of the S_k action on the top homology of the subposet of Π_k obtained by selecting those ranks corresponding to partitions such that the number of blocks is congruent to i mod d. Define $Q_i(i) = h_i$ and for $1 \le j \le d-1$, set $Q_j(i) = h_j$.

One has the following plethystic generating functions:

Theorem 1.4 Let $d \geq 2$.
1. For $i = 2, \ldots, d-1$:

$$(h_1 - Q_i(i) + Q_{i+d}(i) - Q_{i+2d}(i) + \ldots + (-1)^n Q_{i+d(n-1)}(i) + \ldots)[\sum_{i\geq 1} h_i]$$
$$= \sum_{n\geq 0} (-1)^n \sum_{0\leq j\leq d-1, j\neq i} Q_{nd+j}(i)$$

2. For $i = 1$:

$$(h_1 - Q_{1+d}(1) + Q_{1+2d}(1) - \ldots + (-1)^{n-1} Q_{1+(n-1)d} + \ldots)[\sum_{i\geq 1} h_i]$$
$$= \sum_{n\geq 1} (-1)^{n-1} \sum_{0\leq j\leq d-1, j\neq 1} Q_{nd+j}(1).$$

Theorem 1.5 Let Q_k denote the poset obtained from Π_n by deleting all elements at rank k. The characteristic of S_n acting on the top homology of Q_k is given by the degree n term in the symmetric function

$$(-1)^k \pi_{n-k}[\pi_1 - \pi_2 + \ldots + (-1)^{n-1}\pi_n] - \pi_n.$$

For instance, one easily concludes from this formula that the homology of the subposet obtained by deleting all the atoms is given by the symmetric function $h_2 h_1^{n-2} - \pi_n$.

It is not hard to describe how to compute the representation when two ranks are selected. Let $a < b$ be two ranks chosen among the $n-2$ nontrivial ranks $\{1, 2, \ldots, n-2\}$ of Π_n. Denote the corresponding rank-selected subposet by $\Pi_n\{a, b\}$, and the characteristic of the homology representation by $\beta_n(\{a, b\})$. We will give two equivalent formulas for $\beta_n(\{a, b\})$.

For each integer partition λ of n with $n - b$ parts, we define the symmetric function

$$G_a^\lambda(n) = \sum_{(\alpha^{(1)}, \ldots, \alpha^{(s)}, \ldots)} \prod_s \prod_{k=1}^{s} h_{m_k(\alpha^{(s)})}[\, h_k[\sum_{i\geq 1} h_i]\big|_{deg\ s}\,],$$

where the sum runs over all sequences of integer partitions $(\alpha^{(1)}, \ldots, \alpha^{(s)}, \ldots)$ such that
1. $\alpha^{(i)}$ is a nonempty partition iff the part i appears in λ with positive multiplicity $m_i(\lambda) > 0$;
2. every part of $\alpha^{(i)}$ is of size at most i;
3. $\alpha^{(i)}$ has exactly $m_i(\lambda)$ parts;
4. writing $|\alpha^{(i)}|$ for the sum of the parts of $\alpha^{(i)}$, we must have $\sum_i |\alpha^{(i)}| = n - a$.

Note that the degree of $G_a^\lambda(n)$ is $\sum_{\{s: m_s(\lambda)>0\}} s \sum_{k=1}^{s} m_k(\alpha^{(s)}) = \sum_{\{s: m_s(\lambda)>0\}} s\ m_s(\lambda) = n$. Now define $G_a^b(n) = \sum_{\{\lambda\vdash n,\ \ell(\lambda)=n-b\}} G_a^\lambda(n)$. When $a = b$, this expression is still meaningful, and simplifies considerably to give $G_a^a(n) = h_{n-a}[\sum_{i\geq 1} h_i]\big|_{deg\ n}$.

Theorem 1.6
0. The characteristic $\beta_n(\{a\})$ of the homology of $\Pi_n\{a\} = \{x \in \Pi_n : x \text{ is of rank } a\}$ is

$$G_a^a(n) - h_n.$$

1. The characteristic $\beta_n(\{a,b\})$ of the homology representation of $\Pi_n\{a,b\}$ is

$$G_a^b(n) - G_a^a(n) - G_b^b(n) + h_n.$$

2. We also have the simpler recurrence

$$\beta_n(\{a,b\}) + \beta_n(\{b\}) = \beta_{n-a}(\{b-a\})[\sum_{i\geq 1} h_i]|_{deg\, n}.$$

We turn next to the more general case of segments of consecutive ranks. For $1 \leq k \leq r \leq n-2$ define symmetric functions g_{n-k}^r by the formula

$$g_{n-k}^r = (-1)^{k-r}(h_{n-r} + h_{n-r+1} + \ldots + h_{n-k})[\pi_1 - \pi_2 + \ldots + (-1)^{n-k-1}\pi_{n-k}]|_{deg\,(n-k)}.$$

Comparison with Theorem 1.1 shows that g_{n-k}^r is in fact the characteristic of the S_{n-k}-module $V_{n-k}(\underline{r-k})$. Thus $g_n^r = ch(V_n(\underline{r})) = ch(V_n([1,r]))$, while $g_{n-r}^r = h_{n-r}$. For $0 \leq r < s \leq n-2$, denote by $V_n([r+1,s])$ the reduced top homology of the subposet of Π_n obtained by selecting all elements in ranks $r+1, r+2, \ldots, s$. By convention $V_n([0,s]) = 0$, while $V_n([s+1,s])$ is the trivial module. We are now ready to give a recursive formula for computing $V_n([r+1,s])$ as an S_n-module.

Theorem 1.7 With the preceding definitions, writing $ch(V)$ for the Frobenius characteristic of the S_n-module V, one has, for $0 \leq a \leq r \leq n-2$:

$$ch(V_n[a+1,r]) + ch(V_n[a,r]) = g_{n-a}^r[\sum_{i\geq 1} h_i]|_{deg\, n}.$$

Consequently one immediately has the formulas

$$\begin{aligned} &ch(V_n([a,r])) \\ &= (g_{n-a}^r - g_{n-a-1}^r + \ldots + (-1)^{r-a} g_{n-r}^r)[\sum_{i\geq 1} h_i]|_{deg\, n} - (-1)^{r-a} h_n \\ &= (g_{n-a+1}^r - g_{n-a+2}^r + \ldots + (-1)^{k-1} g_{n-a+k}^r + \ldots + (-1)^{a-1} g_{n-r}^r)[\sum_{i\geq 1} h_i]|_{deg\, n}. \end{aligned}$$

Theorem 1.8 The homology representation of the subposet of Π_n consisting of the interval $[1,r]$ with the rank k removed, is given by

$$(-1)^k g_{n-k}^r[\pi_1 - \pi_2 + \ldots]|_{deg\, n} - V_n([1,r]).$$

At the end of the next section we give an alternative computation for $V_n([2,n-3])$.

Using these results and the underlying methods, one can also obtain information about the multiplicity of the trivial representation (and other irreducibles in specific instances) in the homology. As an example, either directly from Part (1) or by recursively applying Part (2) of Theorem 1.1, one concludes that the trivial representation does not appear in $V_n(\underline{r})$, for any $r \geq 2$. This was first proved by Hanlon ([H2]) by more intricate calculations. We now present two theorems which strengthen this result:

Theorem 1.9 Consider the rank-selected subposet $\Pi_n(S)$, where the set S of ranks chosen is of the form $[1,r] \cup \{a\}$, for $a \geq r \geq 2$. Then the trivial representation does not occur in $\tilde{H}(\Pi_n(S))$, provided one of the following conditions is satisfied:
(1) $n \leq 3r+2$;
(2) $a < 2(r+1)$, or $a \geq n-r$; or $r \geq 3$ and $a \geq n-r-1$;
(3) $r < n-a < \binom{r+1}{2}$.

One is tempted to conjecture that the trivial representation does not appear for large subsets S of ranks which contain an interval of the form $[1,r]$, of length at least 2. That this is false is seen by using Theorems 1.5 and 1.8:

Theorem 1.10

1. Let Q_k denote the subposet $\Pi_n - \{\text{elements at rank } k\}$. The trivial representation does not occur in the homology of Q_k if $n \leq 2k$. Otherwise it appears with nonzero multiplicity equal to

$$\frac{1}{n-k} \sum_{d|(k,n-k)} \mu(d)(-1)^{n-\frac{n}{d}} \binom{\frac{n-k}{d}}{\frac{k}{d}}.$$

This is also the number of standard Young tableaux of hook shape $(k, 1^{n-2k})$ or $(k+1, 1^{n-2k-1})$ whose major index is congruent to 1 modulo $n-k$. (See [KW]). For some special values of n and k, this expression has a more elegant combinatorial interpretation which follows from work of Gessel and Reutenauer ([GR], Theorem 9.4). Let $C_{m,i}$ denote the number of m cycles in S_m with a unique descent, in position i. Then (for $n > 2k$) the above multiplicity equals $C_{n-k,k} + C_{\frac{n-k}{2},\frac{k}{2}}$ if n and k are both twice an odd number, and it equals $C_{n-k,k}$ otherwise. (It follows for instance that if $k=1$ or $n=2k+1$, the multiplicity is always 1).

2. The trivial representation does not occur in the homology of the subposets obtained by taking a segment of consecutive ranks $[1,r]$, and then deleting a rank k, provided $n \leq 2k (\leq 2r)$, or if $n = 2k+1$ and $r \leq n-3$.

Similarly, using Theorem 1.7, we obtain

Theorem 1.11

1. The multiplicity of the trivial representation in the homology of the subposet consisting of a segment of consecutive ranks, beginning with rank 2, is always 1;
2. The multiplicity of the trivial representation in the homology of the subposet consisting of a segment of consecutive ranks, beginning with rank 3, that is, in $V_n([3,r])$, is given below. First let $r = n-2$. Then the multiplicity is

$$\begin{cases} \lfloor \frac{n}{2} \rfloor - 1 & \text{if } n \text{ is odd or if } n \equiv 0 \bmod 4; \\ \frac{n}{2} - 2 & \text{if } n \equiv 2 \bmod 4. \end{cases}$$

Now assume $r \leq n-3$. Then the multiplicity is

$$\begin{cases} r-1 & \text{if } r \equiv 2 \bmod 4 \text{ or } r \equiv 3 \bmod 4; \\ r-2 & \text{if } n \equiv 0 \bmod 4 \text{ or } r \equiv 1 \bmod 4. \end{cases}$$

The preceding formulas were obtained by technical calculations with symmetric functions. The proof of Theorem 1.11 also required knowledge of the behaviour of the S_n-modules $V_n(\underline{r})$

upon restriction to $S_2 \times S_{n-2}$. The ultimate simplicity of these formulas is surprising, given the relative intricacy of our plethystic expressions in Theorems 1.7 and 1.8.

The homology computations discussed above are particular cases of the following recurrence for arbitrary rank-selection:

Theorem 1.12 Let $S = \{s_1 < s_2 < \ldots < s_r\}$ be a subset of the ranks $\{1, \ldots, n-2\}$ of Π_n, and denote by $\beta_n(S)$ the Frobenius characteristic of the homology of the rank-selected subposet of Π_n corresponding to the set S. We have the recurrence

$$\beta_n(S) + \beta_n(S \backslash \{s_1\}) = \beta_{n-s_1}(S - s_1)\,[\sum_{i\geq 1} h_i]|_{deg\ n},$$

where $S - s_1$ denotes the subset of ranks $\{s_2 - s_1 < s_3 - s_1 < \ldots < s_r - s_1\}$ in Π_{n-s_1}.

Denote by $\alpha_n(S)$ the permutation representation of S_n acting on the maximal chains of the rank-selected subposet $(\Pi_n)_S$ of Π_n corresponding to the subset S. A standard Hopf trace formula computation shows that

$$\beta_n(T) = \sum_{S \subseteq T} \alpha_n(S)(-1)^{|T|-|S|},$$

thereby expressing the homology module as an alternating sum of the permutation modules of chains. This is valid for any Cohen-Macaulay poset. Theorem 1.12 in effect gives an expression for the homology modules β_n as an alternating sum of *homology* modules of lower degree. An analogue of Theorem 1.12 holds for arbitrary Cohen-Macaulay posets; see Section 2, Theorem 2.8.

The permutation representation $\alpha_n(S)$ is consequently the sum of the homology modules $\beta_n(T)$, as T ranges over subsets of S. That is, the homology modules $\beta_n(T)$, $T \subseteq S$, *refine* the permutation module $\alpha_n(S)$ of maximal chains. This permutation representation is frequently interesting in its own right. In the case of the partition lattice, Theorem 1.12 gives

Corollary 1.13 $\alpha_n(S) = \alpha_{n-s_1}(S - s_1)[\sum_{i\geq 1} h_i]|_{deg\ n}$.

In particular, we obtain an elegant orbit decomposition for the action α_n of S_n on the maximal chains of Π_n :

Theorem 1.14

$$\alpha_n = \sum_{i=1}^{\lfloor n/2 \rfloor} a_i(n)\ \ 1\uparrow_{S_2^i \times S_1^{n-2i}}^{S_n},$$

where the $a_i(n)$ are determined by the recurrence

$$a_i(n+1) = i a_i(n) + (n - 2i + 2) a_{i-1}(n),$$

with initial conditions $a_0(1) = 1 = a_1(2)$, $a_0(n) = 0$, $n > 1$, and $a_i(n) = 0$ if $2i > n$. (Observe that the character values are supported on the set of involutions).

The proof consists of translating Corollary 1.13 into the following recursive description of the α_n :

$$\alpha_n = \left(1_{S_2} \otimes \alpha_{n-1}\downarrow_{S_{n-2}}\right)\uparrow^{S_n}.$$

The statement of the theorem is then obtained by an inductive computation. Note that the point stabilisers of this action are all Young subgroups of the form $S_2^i \times S_1^{n-2i}$.

The recurrence for the $a_i(n)$ is remarkably similar to the one satisfied by the Eulerian numbers $A(n,i)$ (which count the number of permutations in S_n with i descents). Based on this observation, R. Simion found the following combinatorial description for the multiplicity $a_i(n)$: It is the number of permutations in S_{n-2} with exactly $i-1$ descents, none consecutive, and with the "hereditary"property that when the letters $n-2$, $n-3, \ldots 3, 2, 1$ are erased in succession, the property of non-consecutive descents is preserved after each erasure. A slightly different but equinumerous set of permutations appears in work of Foata and Schützenberger ([FSch]), who call them André permutations. In fact in two subsequent papers ([FS1], [FS2]), Foata and Strehl construct an action of S_2^{n-1} on the set of $n!$ permutations whose orbit decomposition is remarkably reminiscent of ours. The precise connection between the Foata-Strehl representation $W_{FS}(n)$ of S_2^{n-1} and the action α_n of S_n on the maximal chains of Π_n is as follows:

Proposition 1.15

$$\alpha_{n+1}\uparrow_{S_{n+1}}^{S_{2n}} = (W_{FS}(n)\otimes 1_{S_2})\uparrow_{S_2{}^{n-1}\times S_2}^{S_{2n}}.$$

In [St], Stanley shows that the multiplicity of the trivial representation in α_n is the Euler number E_{n-1}. It follows from Theorem 1.14 that the André permutations in S_{n-1} counted by $a_i(n)$ refine the number of alternating permutations E_{n-1} in S_{n-1}. Such refinements are the subject of the papers [FSch], [FS1], [FS2]. In particular, in [FSch] the authors give explicit bijections to explain these refinements. The recurrences of [FSch] for the numbers $a_i(n)$ can all be derived by counting relative orbits of maximal chains in Π_n.

Since the rank-selected homologies $\beta_{n+1}(S)$ (in Π_{n+1}) refine the S_{n+1}-module α_{n+1}, the multiplicities $b_S(n+1)$ of the trivial representation in the homology $\beta_{n+1}(S)$ provide another refinement of the Euler number E_n into nonnegative integers indexed by subsets $S \subseteq \{1, 2, \ldots, n-1\}$. More generally, if we write $b_S(\lambda)$ for the multiplicity of the irreducible indexed by the partition λ of n in the homology of the rank-selected subposet $(\Pi_n)_S$, and $A_n(\lambda)$ for the multiplicity of the irreducible λ in α_n, then it follows that $\sum_S b_S(\lambda) = A_n(\lambda)$, where the sum runs over all subsets of $\{1, 2, \ldots, n-2\}$.

We use Theorem 1.14 to compute the multiplicities of certain irreducibles $A_n(\lambda)$ in the permutation representation α_n.

Proposition 1.16

0. ([St]) $A_n((n)) = E_{n-1}$.
1. The multiplicity of the trivial representation in the *restriction* of α_n to S_{n-1} is the Euler number E_n. Equivalently, $A_n((n-1,1)) = E_n - E_{n-1}$.
2. $A_n((3, 1^{n-3})) = 2^{n-2} - 1$;
3. $A_n((2^2, 1^{n-4})) = 2^{n-2} - 2$;
4. $A_n((2, 1^{n-2})) = 1$ and $A_n((1^n)) = 0$ for all n.

The preceding proposition suggests that the irreducible $(3, 1^{n-3})$ appears exactly once in each rank-selected homology corresponding to a nonempty subset S. This is, however, false, as can be seen by examining the top homology π_n itself: in π_6, this irreducible appears twice. There is consequently a different refinement of 2^{n-2}, also indexed by subsets. We also conclude that the irreducible $(2, 1^{n-2})$ appears exactly once in π_n, and never in any $\beta_n(S)$ for $S \neq \{1, 2, \ldots, n-2\}$.

In view of Part 1. of Propostion 1.16, it is also of enumerative interest to determine the restrictions of the rank-selected homology modules of Π_n to S_{n-1}. For, denoting by $b'_S(n)$ the multiplicity of the trivial representation in this restricted rank-selected homology module $\beta_n(S)\downarrow_{S_{n-1}}^{S_n}$, one has a *second* refinement of the Euler number E_n into nonnegative integers, now indexed by subsets $S \subseteq \{1, 2, \ldots, n-2\}$. Note that $b'_S(n)$ is always greater than or equal to $b_S(n)$.

Finally we remark that from Theorems 1.1 and 1.2 we can compute $b'_n(S)$ when S is an initial or final segment of consecutive ranks. We record these and other assorted results in

Proposition 1.17

0. Let $1 \le r \le n-2$. Then $b_{\{r\}}(n) = p(n, n-r) - 1$, and $b'_{\{r\}}(n) = \sum_{i=n-r}^{n-1} p(i, n-r-1)$, (where $p(m,k)$ denotes the number of integer partitions of m into k parts).
1. Let $S = \{1, r\}$, $1 < r \le n-2$. Then $b_S(n) = b'_{\{r-1\}}(n-1) - b_{\{r\}}(n)$.
2. Let S be the interval $[1, r]$ of ranks in Π_n. Then $b_S(n) = 0$, and $b'_S(n) = 1$ for all $r = 1, \ldots, n-2$.
3. Let S be the interval $[n-1-r, n-2]$ of ranks in Π_n. Then for all $r = 1, \ldots, n-2$,
$$b'_S(n) = \binom{n-2}{r}.$$
4. Let $S = [1, n-2] \backslash \{k\}$. Then
$$b'_S(n) = \sum_{r=0}^{min(n-k-1,k)} \binom{n-k-1}{r} - 1.$$
In particular, if $2k > n-1$, then $b'_S(n) = 2^{n-k-1} - 1$ (while $b_S(n) = 0$ by Theorem 1.10).
5. Let $n \ge 6$, $S = [2, n-3]$. Then $b'_S(n) = 2n - 7$.

2. Methods and further results

Let P be a poset with least element 0. Following Baclawski, let $D_i(P)$ be the free abelian group generated by all $i-1$-chains $a_1 < \ldots < a_i$ of elements of $P - \{0\}$. Set $D_{-1}(P) = 0$. For $i > 0$ and $D_i(P) \ne 0$, define a differential $d_i^W : D_i(P) \longrightarrow D_{i-1}(P)$ by
$$d_i^W(a_1, \ldots, a_i) = \sum_{j=1}^{i-1} (-1)^i (a_1, \ldots, \hat{a}_j, \ldots, a_i);$$
define $d_i^W = 0$ otherwise. (As usual the hat over an element in the chain denotes suppression of that element).

The homology of the algebraic complex $(D(P), d^W)$ is the *Whitney* homology of P. It was defined and studied first by Baclawski ([Bal]) and then by Björner ([Bj]), who related it to the usual order homology of the poset P. We write $WH_i(P)$ for the ith Whitney homology group, that is, $WH_i(P) = ker\ d_i^W / im\ d_{i+1}^W$. Note that if P has a least element 0 and a greatest element 1, the Whitney homology in the highest degree coincides with the top homology of the order complex of $P - \{0, 1\}$.

A fundamental theorem is that, when P is Cohen-Macaulay, the homology groups of the complex $(D(P), d^W)$ are free. This was proved by Baclawski for geometric lattices ([Bal]), and follows from work of Björner ([Bj]) in the general case.

Our first basic result is:

Lemma 2.1 Let P be a poset with least element 0 and greatest element 1; let G be the automorphism group of P. Assume that the Whitney homology groups are free in all degrees. Then each Whitney homology is a G-module. If r is the length of the longest chain in $P - \{0\}$, then as a virtual sum of G-modules, one has
$$WH_r(P) - WH_{r-1}(P) + \ldots + (-1)^{r-1} WH_1(P) + (-1)^r WH_0(P) = 0.$$

Since $WH_r(P) = \tilde{H}_{r-2}(P)$, ($P$ has a greatest element 1), one has, equivalently, the equality of G-modules

$$\tilde{H}_{r-2}(P) = WH_{r-1}(P) - WH_{r-2}(P) + \ldots + (-1)^{r-1}WH_0(P).$$

The proof follows from a routine application of the Hopf trace formula to the Whitney complex $(D(P), d^W)$. By combining this lemma with work of Björner, we obtain a powerful tool for computing the top homology of the order complex of P in the case when the poset P is Cohen-Macaulay.

To apply this lemma to obtain our results, we need to compute the Whitney homology of the partition lattice, as an S_n-module. For a finite poset P with 0, write $WH(P)$ for the direct sum of all the Whitney homology groups of P. A theorem of Orlik and Solomon states that $WH(\Pi_n)$ is isomorphic to the cohomology ring of the complement of the thick diagonal in $\mathbf{C}_n$, and consequently has the structure of a graded anti-commutative algebra, in this case the Orlik-Solomon algebra for the root system A_{n-1} (See [OS], Corollary 5.6). In fact Orlik and Solomon show that for any geometric lattice, the associated graded anti-commutative algebra satisfies the acyclicity property of Lemma 2.1. (See [OS], Lemma 2.18). Our Lemma 2.1 points out that this is true in the more general context of the Whitney homology of any poset. Orlik and Solomon also show that for any geometric lattice L, the Whitney homology $W(L)$ does indeed have the structure of a graded anti-commutative algebra.

The S_n-module structure of $WH(\Pi_n)$ was determined by Lehrer and Solomon, (and in somewhat more general terms in [OS]), although this information can also be extracted from previous calculations of Hanlon. We state Lehrer and Solomon's result below, in a form more suited to our requirements:

Theorem 2.2 ([LS], Theorem 4.5) For $i = 0, 1, \ldots n-1$, the characteristic of the ith Whitney homology of Π_n, is given by

$$\sum_{\substack{\lambda \vdash n,\ \ell(\lambda)=n-i \\ \lambda=(1^{m_1}, 2^{m_2}, \ldots)}} h_{m_1}[\pi_1] e_{m_2}[\pi_2] \ldots e_{m_{2i}}[\pi_{2i}] h_{m_{2i+1}}[\pi_{2i+1}] \ldots,$$

the sum ranging over all (integer) partitions λ of n with exactly $n-i$ parts; m_i denotes the multiplicity of the part i.

We can give a direct proof of a more general result on the module structure of any interval in Π_n.

The results presented in Theorems 1.1 and 1.2 depend on the following preliminary observation:

Proposition 2.3 Assume P is a Cohen-Macaulay poset of rank n; denote by $P(\underline{r})$ the subposet of P obtained by selecting ranks $\{1, \ldots r\}$ in P. Let G be the automorphism group of P. Then the Whitney homology of $P(\underline{r})$ coincides with that of P in degrees $0, 1, \ldots r$. In particular, the order homology of $P(\underline{r})$ is completely determined by the Whitney homology of P, by means of the G-equivariant recurrence

$$\tilde{H}(P(\underline{r})) \oplus \tilde{H}(P(\underline{r-1})) = WH_r(P).$$

Here $1 \leq r \leq n-1$, and by definition $\tilde{H}(P(\underline{0})) = WH_0(P)$, while $\tilde{H}(P(\underline{r})) = 0$ if $r < 0$ or $r > n-1$. Using Björner's characterisation of $WH_r(P)$ in terms of the order homology of P ([Bj]), we can construct a surjective group-equivariant map from $WH_r(P)$ onto $\tilde{H}(P(\underline{r-1}))$, whose kernel is $\tilde{H}(P(\underline{r}))$.

Part (1) of Theorem 1.1 is proved using Lehrer and Solomon's result in conjunction with the basic Lemma 2.1. The second part follows by suitable manipulation of the symmetric functions in Part (1). Alternatively, noting that the subposet $\Pi_n(\underline{r})$ is also a geometric lattice, we use Björner's theory of NBC bases to construct a basis for the homology which makes the statement transparent for all values of r except $r = n-2$. In the latter case we are dealing with the top homology of the partition lattice, and the statement, which is equivalent to the fact that S_{n-1} simply permutes the $(n-1)!$ homology spheres in Π_n, is then due to Stanley ([St]). Curiously enough, this fact is not obvious from Björner's NBC basis for homology. In [Wa], Wachs constructs a new basis which does reflect this property.

To prove Theorem 1.2, we compute the Whitney homology of the *dual* of the partition lattice. (Although the dual is no longer a geometric lattice, it is still Cohen-Macaulay). Write $WH_r^*(\Pi_n)$ for the rth Whitney homology of the dual poset of Π_n. We have

Theorem 2.4 The characteristic of the S_n-module $WH_r^*(\Pi_n)$ is given by the degree n term in the plethysm

$$\pi_{r+1}[h_1 + h_2 + \ldots],\ r = 0, 1, \ldots, n-1.$$

The general technique therefore consists of first determining the Whitney homology of the rank-selected subposet, and then applying Lemma 2.1.

Examining Lemma 2.1 more carefully, we find that, when combined with Theorems 2.2 and 2.4, we obtain the following plethystic inverse identity:

Corollary 2.5 The plethystic inverse of the series $T = \sum_{i\geq 1} h_i$ is the alternating sum $I = \sum_{i\geq 1}(-1)^{i-1}\pi_i$. More specifically, Theorem 2.2 corresponds to the identity $T[I] = h_1$, while Theorem 2.4 corresponds to $I[T] = h_1$.

On the other hand, it follows from a computation of Hanlon ([H1]) on the fixed-point partition lattices that $\pi_n = \frac{1}{n}\sum_{d|n}\mu(d)p_d^{\frac{n}{d}}(-1)^{n-\frac{n}{d}}$, where p_d denotes the dth power-sum. Consequently we recover Cadogan's formula ([C]), which says that the plethystic inverse of the series $T = \sum_{i\geq 1} h_i$, is the series $I = \sum_{d\geq 1}\mu(d)/d\ log(1+p_d)$.

This is the basis for our derivation of the generating functions of Calderbank, Hanlon and Robinson for the other subposets of partitions described in the introduction. We mention briefly a particular result for the lattice Π_n^d of partitions of a set of size nd whose block sizes are all multiples of d. A nice basis for the top homology of this lattice was found recently by Wachs ([Wa]). Her basis suggests that the S_{nd}-representation on $\tilde{H}(\Pi_n^d)$ is in fact a submodule of the wreath product of $\tilde{H}(\Pi_n)$ with the trivial representation of S_d. Using the techniques described above we can easily prove this conjecture, and identify the complementary module as follows:

Proposition 2.6 Write π_n^d for the characteristic of $\tilde{H}(\Pi_n^d)$, and let $q_n^d(1)$ denote the characteristic of the top homology of the subposet of Π_n^d obtained by deleting the atoms. Then one has

$$\pi_n[h_d] = \pi_n^d + q_n^d(1).$$

Wachs has observed that a constructive proof of this observation follows from the results in [Wa], by writing down bases for all the modules involved.

Our next theorem shows that the S_{nd}-module structure of the Whitney homology of the lattice Π_n^d is remarkably similar to that of the partition lattice.

Theorem 2.7 The rth Whitney homology of Π_n^d, as an S_{nd}-module, for $0 \le r \le n$, is given by

$$WH_r(\Pi_n^d) = \sum_{\substack{\lambda \vdash n,\ \ell(\lambda)=n+1-r \\ \lambda=(1^{m_1},2^{m_2},\ldots)}} h_{m_1}[\pi_1^d]e_{m_2}[\pi_2^d]\ldots e_{m_{2i}}[\pi_{2i}^d]h_{m_{2i+1}}[\pi_{2i+1}^d]\cdots,$$

with all notation as in Theorem 2.2.

Our second basic technique, which is most useful for describing the homology of the rank-selected subposet when the rank set is large, is the following:

Theorem 2.8 Let P be a Cohen-Macaulay poset of rank r with automorphism group G, and S a subset of the ranks $\{1,\ldots,r-1\}$. Let P_S be the rank-selected subposet of P, consisting of the ranks in S. Then as G-modules:

$$\begin{aligned}&(-1)^{r-|S|}\tilde{H}(P_S) - \tilde{H}(P) \\ &= \bigoplus_{\substack{0<x_1<\ldots<x_k<1 \\ rank(x_i)\notin S}} (-1)^k \left(\tilde{H}(0,x_1)_P \otimes \tilde{H}(x_1,x_2)_P \otimes \ldots \otimes \tilde{H}(x_k,1)_P\right)\big\uparrow_{stab(x_1,\ldots,x_k)}^{G};\end{aligned}$$

where

1. the sum runs over all chains of elements *not* in S;
2. the homologies involved are all top homologies of open intervals $(x_i,x_{i+1})_P$ in the poset P;
3. $stab(x_1,\ldots,x_k)$ is the intersection of the stabilisers of all $(k+1)$ intervals $(0,x_1)_P$, $(x_i,x_{i+1})_P$, $(x_k,1)_P$, in P;
4. the tensor product is an *internal* tensor product of representations of $stab(x_1,\ldots,x_k)$; this representation is then induced up to G.

The starting point for the proof is a lemma of Baclawski ([Ba2]), which is precisely the Möbius function version of the formula above. Baclawski's lemma applies to arbitrary posets; our proof of the group-equivariant result requires that the poset be Cohen-Macaulay.

In the particular case when exactly one rank is removed, this theorem when applied to the partition lattice, yields the formula of Theorem 1.5. One obtains less extensive simplification when two ranks are deleted, but ultimately one can write down an explicit closed form for the homology representation. We mention one particular example:

Let $V_n([2,n-3])$ be as in Theorem 1.7. Since the subposet in question is obtained by deleting ranks 1 and $n-2$ from Π_n, we may use the preceding result to calculate the homology representation. We have

Proposition 2.9

$$\begin{aligned}&ch(V_n([2,n-3]) \\ &= \pi_n + h_2(\pi_{n-2} + h_1\pi_{n-3} + h_1^2\pi_{n-4} + \ldots + h_1^{n-5}\pi_3 + h_1^{n-4}\pi_2) \\ &- \sum_{1\le k<n-k} \pi_k\pi_{n-k} - f_n;\end{aligned}$$

where the last term f_n is nonzero only if n is even, in which case it is $h_2[\pi_{\frac{n}{2}}]$ if $\frac{n}{2}$ is odd, and equals $e_2[\pi_{\frac{n}{2}}]$ if $\frac{n}{2}$ is even.

It is clear from this expression that the trivial representation appears exactly once (since it appears in π_n iff $n \le 2$).

REFERENCES

[Ba1] K. Baclawski, Whitney numbers of geometric lattices, Advances in Math. 16 (1975), 125-138.

[Ba2] K. Baclawski, Cohen-Macaulay connectivity and geometric lattices, Europ. J. Comb. 3 (1982), 293-305.

[Bj] A. Björner, On the homology of geometric lattices, Algebra Universalis 14 (1982), 107-128.

[Br] A. J. Brandt, The free Lie ring and Lie representations of the full linear group, Amer. Math. Soc. Trans. 56 (1944), 528-536.

[C] C. C. Cadogan, The Möbius function and connected graphs, J. Comb. Theory (B), 11 (1971), 193-200.

[CHR] A. R. Calderbank, P. Hanlon and R. W. Robinson, Partitions into even and odd block size and some unusual characters of the symmetric groups, Proc. London Math. Soc. (3), 53 (1986), 288-320.

[FSch] D. Foata and M-P. Schützenberger, Nombres d'Euler et permutations alternantes, manuscript, University of Florida, Gainesville (1971). First part published in "A Survey of Combinatorial Theory," J. N. Srivastava et al., eds., North-Holland (1973), 173-187.

[FS1] D. Foata and V. Strehl, Rearrangements of the Symmetric Group and enumerative properties of the tangent and secant numbers, Math. Zeitschrift, 137 (1974), 257-264.

[FS2] D. Foata and V. Strehl, Euler numbers and Variations of Permutations, *in* Atti Dei Convegni Lincei 17, Colloquio Internazionale sulle Teorie Combinatorie (1973), Tomo I, Roma, Accademia Nazionale dei Lincei, 1976, 119-131.

[GR] I. Gessel and C. Reutenauer, Counting permutations with given cycle structure and descent set, J. Comb. Theory (A), to appear.

[H1] P. Hanlon, The fixed-point partition lattices, Pacific J. Math., 96 (1981), 319-341.

[H2] P. Hanlon, A proof of a conjecture of Stanley concerning partitions of a set, European J. Comb. 4 (1983), 137-141.

[J] A. Joyal, Foncteurs Analytiques et Espèces de Structures, *in* Lecture Notes in Mathematics 1234 (1986), 126-160, Springer-Verlag, Berlin/Heidelberg/New York.

[KW] W. Kraskiewicz and J. Weyman, Algebra of invariants and the action of a Coxeter element, Math. Inst. Copernicus Univ., Chopina, Poland (preprint) (1987).

[LS] G. I. Lehrer and L. Solomon, On the Action of the Symmetric Group on the Cohomology of the Complement of its Reflecting Hyperplanes, J. of Algebra 104 (1986), pp 410-424.

[M] I. G. Macdonald, "Symmetric functions and Hall polynomials", Oxford University Press (1979).

[OS] P. Orlik and L. Solomon, Combinatorics and topology of complements of hyperplanes, Invent. Math., 56 (1980), 167-189.

[St] R. P. Stanley, Some aspects of groups acting on finite posets, J. Comb. Theory (A) 32, no. 2 (1982), 132-161.

[Wa] M. Wachs, A basis for the homology of d-divisible partition lattices, 1991, submitted.

[W] E. Witt, Treue Darstellung Liescher Ringe, Journal für die reine und angewandte Mathematik, Band 177 (1937), pp. 152-160.

391

Hadamard invertibility of linearly recursive sequences in several variables

Earl J. Taft
Department of Mathematics
Rutgers University
New Brunswick, NJ 08903
U.S.A.

1. Introduction:

Several authors have considered the algebra of linearly recursive sequences over a field k under the Hadamard (pointwise) product. Recently, R. Larson and the author, [L–T], gave a Hopf algebraic proof of the result determining the invertible elements of this algebra as the set of sequences with non–zero initial data which, starting from some coordinate, become interlacings of a finite–number of non–zero geometric sequences. This result had been obtained earlier in [B] by analytic methods for characteristic k zero, and in [R] using algebraic methods. Our Hopf algebra methods are also effective, in the sense of finite algorithms (see Section 2). In this communication, we extend the discussion to the case of $n>1$ variables, i.e., to multisequences $(f_{i_1 i_2 \cdots i_n})$, $i_j \geq 0$ for $1 \leq j \leq n$, which for each such j, and each choice of i_t for $t \neq j$, each row $(f_{i_1 \cdots i_{j-1} s i_{j+1} \cdots i_n})_{s \geq 0}$ satisfies a linearly recursive relation (independent of the choice of i_t for $t \neq j$). The criterion for Hadamard invertibility will still be that a finite set of initial entries should be non–zero, and that each row should eventually be the interlacing of geometric series. Moreover, the procedure is effective.

See also [C–P] for a discussion of the coalgebraic aspects of linearly recursive sequences in several variables.

2. The case of $n=1$ variable

We recall here the ideas and methods of [L–T]. Let $A = k[x]$ be the polynomial algebra in one variable x, with bialgebra structure given by $\Delta x = x \otimes x$ and $\epsilon(x) = 1$, i.e., Δ and ϵ are algebra homomorphisms from A to $A \otimes A$ and k respectively such that $\Delta(x^n) = x^n \otimes x^n$ and $\epsilon(x^n) = 1$ for all $n \geq 0$. We consider the dual space A^* of linear functions on A as all sequences $(f_n)_{n \geq 0}$ for f_n in k, where such a linear function f is identified with (f_n) for $f_n = f(x^n)$ for $n \geq 0$. A^* has a subspace A^o, called the continuous linear dual of A, consisting of those f which vanish on some ideal of finite codimension in A (see [S]). Since each such ideal in A is generated by a monic polynomial, one sees that A^o consists precisely of the linearly recursive sequences $f = (f_n)$, i.e., there is a polynomial $h(x) = x^r - h_1x^{r-1}\cdots -h_r$ such that $f_n = h_1f_{n-1} + h_2f_{n-2}+\cdots+h_rf_{n-r}$ for all $n\geq r$ (see [P–T]). We say that (f_n) satisfies the relation $h(x)$. If r is minimal, we call $h(x)$ the minimal recursive polynomial of f. A^o is a bialgebra, whose algebra structure depends on the coalgebra structure of A. In [P–T], we chose $\Delta x = 1\otimes x + x\otimes 1$ and $\epsilon(x) = 0$, yielding the Hurwitz (or divided power) product (see also [T]). In this paper, $\Delta x = x\otimes x$ and $\epsilon(x) = 1$, which yields the Hadamard (or pointwise) product $(f_n)(f'_n) = (f''_n)$, where $f''_n = f_nf'_n$ for all $n\geq 0$.

A geometric sequence (ar^n) for $a,r\neq 0$ in k satisfies the relation $x-r$, and its inverse $(a^{-1}r^{-n})$ satisfies $x-r^{-1}$. We note that we do not assume f in A^o satisfies $h(x)$ from the beginning, i.e., $h(0)\neq 0$. For example, we allow $(\pi,\sqrt{6},1,2,4,8,16,\cdots)$ which satisfies $x^3 - 2x^2$. It's inverse $(\frac{1}{\pi}, \frac{1}{\sqrt{6}}, 1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16},\cdots)$ satisfies $x^3 - \frac{1}{2}x^2$. More generally, if $(e_n = ar^n)$, $(f_n = bs^n),\cdots,(g_n = ct^n)$ are a finite number m of non-zero geometric sequences, then their interlacing $(e_0,f_0,\cdots,g_0,e_1,f_1,\cdots,g_1,e_2,f_2,\cdots,g_2,e_3,\cdots)$ satisfies $(x^m - a)(x^m - b)\cdots(x^m - c)$, so is linearly recursive. It's inverse is of the same form, so is also linearly recursive. In [L–T], we prove that f in A^o is Hadamard invertible if and only if all $f_n\neq 0$ and, except for a finite number of terms, f is an interlacing of geometric sequences.

In order to use Hopf algebra ideas, one needs to extend $A = k[x]$ to $H = k[x,x^{-1}]$, Laurent polynomials in x. H is a Hopf algebra, and so is its continuous dual H^o. H^o consists of doubly–infinite sequences (f_n) for n in the integers Z, satisfying a recursive relation $f_n = h_1f_{n-1}+ h_2f_{n-2}+\cdots+ h_rf_{n-r}$ for all $n\in Z$, i.e., the polynomial $h(x)$ satisfies $h(0)\neq 0$. The restriction map H^o to A^o is a bialgebra homomorphism. There is an important bialgebra homomorphism from A^o to H^o called backsolving. One writes the minimal recursive relation for f as $h(x) = p(x)x^k$ with $p(0)\neq 0$. Then we delete $f_o,\cdots,f_{k-1}$, and starting with f_k, we use $p(x)$ to backsolve to the left. For example, $f = (2,5,1,1,2,3,5,8,\cdots)$ satisfies $x^4-x^3-x^2 = x^2(x^6-x-1)$. We delete $f_o=2$ and $f_1=5$, and obtain a doubly–infinite sequence $g = (\cdots,g_2=2,g_{-1}=-1,g_0=1,g_1=0,g_2=1,g_3=1,g_4=2,g_5=3,g_6=5,g_7=8,\cdots)$. This enables us to transfer the determination of the invertible elements of A^o to those of H^o. The group of group–like elements of H^o, i.e., those $f = (f_n)_{n\in Z}$ with $\Delta f = f\otimes f$ and $f\neq 0$ is precisely the group of geometric sequences $e(r) = (r^n)$, with $e(r)e(s) = e(rs)$. We then use the Hopf algebra structure of H^o to show that each invertible element is an interlacing of non–zero geometric sequences.

The procedure is effective, provided one knows some information about the non–zero roots of a recursive polynomial $h(x)$ for a sequence $(f_n)_{n\geq 0}$. We write $h(x) = x^k p(x)$ where $p(0)\neq 0$. Then we back–solve to consider a doubly–infinite sequence (F_n) for n in Z. We must check that $f_0,\cdots,f_{k-1}$ are non–zero, and then test (F_n) for being an interlacing of geometric sequences. We consider the subgroup of the multiplicative group k^* of non–zero elements of k generated by the roots of $p(x)$. It is a finitely generated abelian group, with a torsion subgroup of finte order m. If k has characteristic $p>0$, let p^r be the smallest power of p not less than the largest multiplicity of the roots of $p(x)$. Then F is invertible in H^o if and only if F is the interlacing of m geometric sequences (or mp^r such at characteristic $p>0$). So the finite algorithm for f is to determine if $f_0,\cdots,f_{k-1}$ are non–zero, and then to examine the first $2m$ (or $2mp^r$) terms if F to predict the m ratios $r_1,\cdots,r_m$ and the polynomial $(x^m-r_1),(x^m-r_2)\cdots(x^m-r_m)$ that F would satisfy if it were the interlacing of m (or mp^r) geometric sequences. Then we see if $p(x)$ divides this polynomial. Clearly this is a finite procedure, once m has been determined.

3. The case of $n>1$ variables

We consider the bialgebra $A = k[x_1,\cdots,x_n]$ with each x_i (and hence each mononomial $x_1^{i_1}\cdots x_n^{i_n}$) group–like. $A = k[x_1]\otimes k[x_2]\otimes\cdots\otimes k[x_n]$ as an algebra (also as a coalgebra), and thus $A^o \cong k[x_1]^o\otimes\cdots\otimes k[x_n]^o$ as a bialgebra.

We identify each f in A^* as a multisequence $(f_{i_1 i_2\cdots i_n})$ for all $i_1,\cdots,i_n \geq 0$, where $f_{i_1 i_2\cdots i_n} = f(x_1^{i_1}x_2^{i_2}\cdots x_n^{i_n})$. A "row" of such a multisequence is a sequence $\{f_{i_1\cdots i_{\ell-1} j\, i_{\ell+1}\cdots i_n} | j\geq 0\}$ for a fixed $i_1,\cdots,i_{\ell-1},i_{\ell+1},\cdots,i_n\geq 0$, which we say is parallel to the x_ℓ – axis. The product in A^* (and in A^o) is the Hadamard product.

Let f be in A^o, $f(J) = 0$ for a cofinite ideal J of A. For each $1 \le i \le n$, the powers of x_i span a finite–dimensional space in A/J, so there is a minimal monic $h_i(x)$ in $k[x]$ such that each row of f parallel to the x_i–axis satisfies $h_i(x)$. Thus J contains the cofinite elementary ideal Γ generated by $h_1(x_1), \cdots, h_n(x_n)$.

We introduce $H = k[x_1, x_1^{-1}, x_2, x_2^{-1}, \cdots, x_n, x_n^{-1}]$, a Hopf algebra isomorphic to $h[x_1, x_1^{-1}] \otimes \cdots \otimes h[x_n, x_n^{-1}]$, so $H^o \cong k[x_1, x_1^{-1}]^0 \otimes \cdots \otimes k[x_n, x_n^{-1}]^0$. We think of each f in H^o as a scalar field attached to the integral lattice points in n – space. So $f = (f_{i_1, \cdots, i_n})$ for $i_1, i_2, \cdots, i_n$ in Z, $f_{i_1 \cdots i_n} = f(x_1^{i_1} \cdots x_n^{i_n})$. The rows of f are doubly–infinite sequences with all rows parallel to the x_i – axis satisfying a $p_i(x_i)$ with $p_i(0) \neq 0$. The restriction map $H^o \longrightarrow A^o$ is a bialgebra homomorphism. We use the elementary ideal Γ to backsolve from A^o to H^o. We write $h_i(x_i) = p_i(x_i) x_i^{k_i}$ with $p_i(0) \neq 0$. In each row parallel to the x_i–axis, we backsolve starting from the k_i–th coordinate. Let α: $A^o \longrightarrow H^o$ be this backsolving map. If $\alpha(f)$ is not invertible, i.e., if some row of $\alpha(f)$ is not an interlacing of geometric sequences, then f is not invertible. f is invertible if and only if all $f_{i_1 \cdots i_n} \neq 0$ and each row of f is eventually an interlacing of geometric sequences. This is because all rows parallel to a fixed axis must satisfy the same recursive relation, and be themselves invertible. This transfers the problem to showing that in H^o, each row of $\alpha(f)$ is an interlacing of geometric sequences.

We claim the procedure is finite, i.e., we only have to check a finite number of rows in $\alpha(f)$ for interlacing of geometric sequences, as well as checking only a finite number of coordinates of f for being non–zero. Starting at the position $k_1 k_2 \cdots k_n$, we consider the finite hypercube having $f_{k_i \cdots k_n}$ in the upper left corner, and which includes all the initial data beyond this point, i.e., in each row parallel to the x_i–th axis, include positions k_i to degree $h_i(x) - 1$. Then, we backsolve along each of the (finite number of) rows through and in front of the hypercube, which involves deleting only a finite number of coordinates of f. If some backsolved row is not an interlacing of geometric sequences, then f is not invertible. If each of these backsolved rows is an interlacing of geometric sequences, then $\alpha(f)$ is invertible, because outside of these rows, the recursive relation will guarantee the interlacing of geometric sequences in the other rows. Then we have only to check that the entries deleted in the back–solving were not zero, and then f is invertible, since the reciprocal sequence f^{-1} will satisfy $x_i^{k_i}$ times an appropriate interlacing relation for each $1 \le i \le n$. Thus we have an effective procedure to accompany the following theorem.

Theorem: f in $k[x_1, \cdots, x_n]^o$ is Hadamard invertible if and only if there is a finite hypercube in the multisequence representation of f, such that the rows through and in front of the hypercube have non-zero coordinates before the hypercube, and are eventually interlacing of non-zero geometric sequences.

We illustrate with $n=2$, and $f \in k[x,y]^o$ in tableau form

	1	x	x^2	x^3	x^4	x^5	x^6
1	*	*	*	······	······	······	······
y	*	*	*	······	······	······	······
y^2	*	1	5	2	10	4	20···
y^3	*	7	9	14	18	28	36···
y^4	⋮	3	15	6	30	12	60···
y^5	⋮	21	27	42	54	84	108···
	⋮	⋮	⋮	⋮	⋮	⋮	⋮

satisfying $x^3 - 2x = x(x^2-2)$ and $y^4 - 3y^2 = y^2(y^2 - 3)$. We must first check that the * entries are non-zero. Then we test separately the first two rows and the first column (here we use row and column in the usual 2 by 2 sense) by the $n=1$ method. Then, starting with the hypercube $\boxed{\begin{matrix}1 & 5\\ 7 & 9\end{matrix}}$, we test the 2 rows to the right, and the two columns going down. This will certainly be invertible if all the * entries are non-zero.

References

[B] B. Benzaghou, Algèbres de Hadamard, Bull. Soc. Math. France 98 (1970), 209–252.

[C–P] L. Cerlienco and F. Piras, On the continuous dual of a polynomial bialgebra, Comm. Alg. 19 (1991), 2707–2727.

[L–T] R.G. Larson and E.J. Taft, The algebraic structure of linearly recursive sequences under Hadamard product, Israel J. Math. 72 (1990), 118–132.

[P–T] B. Peterson and E.J. Taft, The Hopf algebra of linearly recursive sequences, Aequationes Math. 20 (1980), 1–17.

[R] C. Reutenauer, Sur les éléments inversible de l'algèbre de Hadamard des séries rationelles, Bull. Soc. Math. France 110 (1982), 225–232.

[S] M. Sweedler, Hopf Algebras, Benjamin, New York, 1969.

[T] E.J. Taft, Hurwitz invertibility of linearly recursive sequences, Cong. Numer. 73 (1990), 37–40.

A SURVEY OF POLYOMINO ENUMERATION

Xavier Gérard Viennot [1]
LaBRI & CNRS [2]
Université Bordeaux I

Abstract. In this talk I will give a survey of the state of the art in the enumeration of polyominoes and lattice animals. A visual walk is given through the zoo of families of polyominoes which can be enumerated, among the garden of related bijections, with some glimpses on various related topics.

Keywords: Classical analytic methods (recurrence relations, continued fractions, q-series, ...), transition matrix, Temperley methodology, DSV and q-DSV methodology, theoretical computer science (automata, algebraic languages, Dyck words, attribute grammars), bijections, heaps of pieces, commutations rules, Cartier-Foata trace monoids, trees (binary, ternary, colored, "guingois", ...), basic hypergeometric functions, Ehrhart' theory for convex polytopes, random generation, fractal dimension, braids, computer algebra, critical exponents, phase transition , statistical mechanics.

1. Introduction.

An *elementary cell* is a square $[i,i+1]\times[j,j+1] \subseteq \mathbb{R}\times\mathbb{R}$ with i and j integers. A *polyomino* is a connected union of elementary cells such that the interior is also connected. Polyominoes are defined up to a translation. Two main parameters are defined for polyominoes. The *area* is the number of elementary cells. The *perimeter* is the number of edges (of the lattice $Z\times Z$) on the border of the polyomino. A major open problem in combinatorics (and also in statistical physics) is to give a formula for the enumeration of polyominoes according to the area or to the perimeter (or to both parameters). To the knowledge of the author, not a single formula of any kind (recurrence for the number of polyominoes, explicit or implicit equation for the generating function, ...) is known.

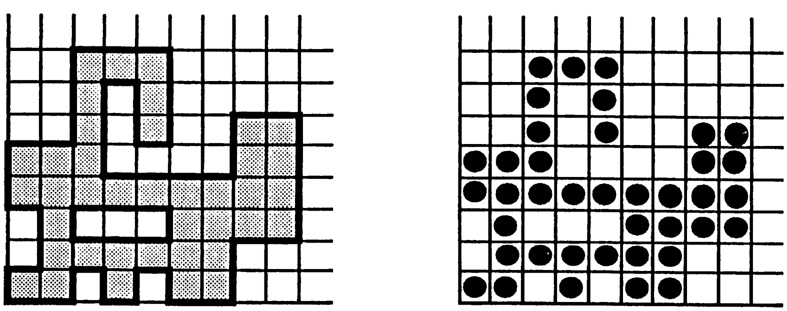


Fig. 1. A polyomino and its associated animal.

1. Partially supported by the Mittag-Leffler Institute and by the PRC "Mathématiques et Informatique"
2. LABRI, Unité associée au Centre National de la Recherche Scientifique 1304, 33405, Talence, France.

In the past few years, there have been an intense activity for the search of enumerative formulae for subclasses of polyominoes. About a dozen families are known with explicit enumeration formula. Some formulae are surprisingly simple, some are very complicated. Most families are enumerated according to both parameters area and perimeters. Many q-series and algebraic generating functions appear. Slowly, some patterns begin to emerge. In this survey I try to put some order in this jungle of formulae disseminated into about 90 papers listed below. I have classified the methods used into six classes: transition matrix, "Temperley methodology", "DSV methodology", heaps methodology, classical analytic (recurrences, continued fractions,...), pure bijective. Of course this is not a rigid classification, there are many overlaps. Many concepts, tools, models, and bijections of other parts of enumerative and algebraic combinatorics appear here.

The problem takes its roots in physics where polyominoes are incarnate under the equivalent notion of *lattice animals*. An animal is obtained from a polyomino by taking the center of each cell (see Fig. 1). In other words, an animal (on the square lattice) is a set α of points of $Z\times Z$ such that any two points can be connected by a path having "elementary steps" North, South, East or West. The physics is the study of models for phase transition and critical phenomena, with computation of the so-called *critical exponents* and *partition function*. For each family of polyominoes (or animals), the asymptotic behavior of the number a_n of polyominoes in the class usually define a critical exponent θ in the form $a_n \sim \mu^n\ n^{-\theta}$. This exponent is analog to the one defined in thermodynamic models (for some models, this analogy can be the identity). Moreover, it appears that some partition functions are exactly the generating function for some classes of animals. Other parameters are also introduced: width, height, ... of the animal; various other lattices are also considered (hexagonal, triangular, "checkerboard", ..), leading to an avalanche of enumeration problems. Thermodynamic models and families of polyominoes are classified in physics by *"universality classes"*, according to their critical exponents. Physical methods can give some approximations to these exponents, or even explicit values (but with no rigorous mathematical proof).

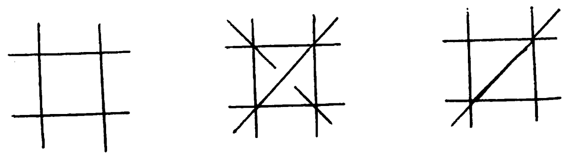
A famous related problem is the enumeration of the so-called *self-avoiding paths* (also completely open). In the case of *polygons* (i.e. closed loops on $Z\times Z$), the problem is equivalent to enumerate simply connected (i.e. having no *"holes"*) polyominoes according to the perimeter. A huge literature exists in physics and theoretical chemistry for animals and related paths (walks) problems. Here we will not be concerned with asymptotic considerations and universality classes, although de Gennes says that identification of universality classes for various connected cluster models is crucial in the scaling theory of branched polymer configurations (de Gennes 1979). We limit our interest to the classification and understanding of formulae giving the exact number of polyominoes or the associated generating function. Also, we will not consider here another active area of research around polyominoes: tilings problems with polyominoes where many people bring contributions (J.H. Conway, M.Gardner, S.Golomb, Gordon, D. Klarner, J.C. Lagarias, ... , and the active french school around D.Beauquier and M.Nivat). Note that other connections exist with computer science as for example the appearance of directed animals in binary search networks (Barcucci, Pinzani, Rodella 1990).

2. The polyominoes zoological garden.

Most of the families of the zoo can be defined by combining two main and simple concepts: "convexity" and "directed".

Let denote by $\mathbb{Sqr}$ the square lattice whose vertices are $Z\times Z$ and whose edges are the edges of elementary cells. Denote by $\mathbb{Full}$ the (highly non planar) lattice obtained by adding the two diagonal edges of each elementary cells of $\mathbb{Sqr}$. The triangular lattice $\mathbb{Tri}$ and hexagonal lattice $\mathbb{Hex}$ can be identified as sub-lattices of $\mathbb{Full}$. A *path* (or *walk*) on a lattice $\mathbb{La}$ is a sequence of vertices $\omega = (s_0,.., s_i, s_{i+1}, ..., s_n)$ such that each pair (s_i, s_{i+1}) of consecutive vertices (i.e. *elementary steps*) are connected by an edge in the lattice. An *animal* on the lattice $\mathbb{La}$ will be a finite *'connected'* set of vertices, that is a set such that each pair of vertices of the animal can be joined by a path of $\mathbb{La}$ contained in the animal. Usually, $\mathbb{Sqr}$ is referred with 4-connexity and $\mathbb{Full}$ with 8-connexity. Altough some papers deal with animals on other lattices, here we will only consider animals on the square lattice $\mathbb{Sqr}$.

Fig. 2. Lattices



Sqr   Full   Tri

We will denote by V, H, D, Δ respectively the vertical axis, horizontal axis, main diagonal and anti-diagonal (diagonal perpendicular to the main diagonal) of R×R. Let L be any one of these four lines. An animal $\alpha \in Z\times Z$ of the square lattice Sqr is said to be *L-convex* iff the intersection of α with any line parallel to L is a connected set of the lattice Full. Let $\vec{L}$ be one of the 8 oriented possible cardinal directions (East, .., North-East, ..) denoted respectively E, N, W, S, NE, NW, SW, SE. A path of the lattice Sqr is said to be *$\vec{L}$-directed* iff its projection on a line parallel to the direction L never goes backward in the reverse direction to L. An animal α (subset of Z×Z) is said to be *$\vec{L}$-directed* iff any point of the animal can be reached from a single point (called *source point*) by a directed path contained in the animal. For example, a NE-directed animal (here called for short *directed animal*) is an animal such that each point can be reached from the origin (0,0) by a path contained in the animal and having elementary steps only North or East. These qualities will be also defined for polyominoes via their underlying animal.

We can also extend the definition of directed animals with several sources points. We will say that the animal is *$\vec{L}$-cs-directed* (*compact source directed animal*) iff there exist a set of points (called *source points*) on the lattice Sqr which are on a line M perpendicular to the line L associated to $\vec{L}$, which are connected for the connexity in Full (8-connexity), and such that every point of the animal can be reached from one of these source points by a directed path in the square lattice Sqr.

The square lattice zoo.

We can combine the different conditions and define plenty of polyominoes families. All of them can be defined by various combination of H, V, D, Δ -convex, NE, NW, SE, SW -directed, and cs-directed properties. For 'name' = 'convex', 'directed 'or 'cs-directed', we will use the notations "X-Y-..-Z-name" to say that the family of polyominoes has the property 'name' relatively to X, Y, ..., and Z. The *zoo of polyominoes* is the set of all the possible families of polyominoes obtained by various combination of these properties (at least one !) and classified by non-isomorphic classes (i.e. up to the group of symmetries acting on the square and up to the relations existing between these different properties). For example the properties Δ-convex and NE-SE-directed imply V-H-convex (parallelogram polyominoes). Also for L= E, N, W, S, 'L-directed' is equivalent to 'L-cs-directed'. The total theoretical number of possible families is $2^8\, 3^4 - 1$.

In the case of directed animal, the property "convex" is sometimes called in the literature *"compact"* or *"fully compact"*. The E-directed animals are also called *partially directed* animals, and no formula is known. I will called the *friendly zoo* the family of polyominoes of the zoo not using in their definition one of the X-directed (or X-cs-directed) for X=N, E, S, or W. The total theoretical number of families of the friendly zoo becomes $6^4 - 1$. I have not listed the number of non trivial, non isomorphic classes. What I know, is that all the families of polyominoes (square lattice) appearing in the ninety papers listed below giving some explicit enumeration formulae belong to the friendly zoo. Only 14 of these families appear. In some papers appear for technical reasons some secondary sub-families which are some slight modification of the main family. Here are the 14 families.

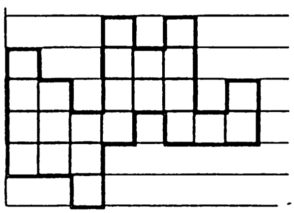


Fig. 3. Column-convex polyomino. (V-convex)

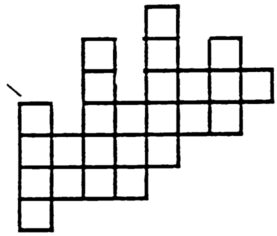


Fig. 4. Directed column-convex animal. (V-convex + NE-directed)

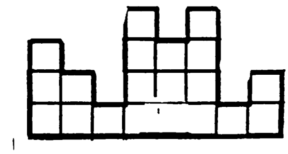


Fig. 5. Wall polyomino (= V-convex + NE-NW directed)

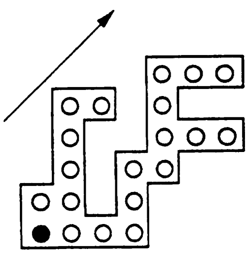


Fig. 6. Directed animal (NE-directed)

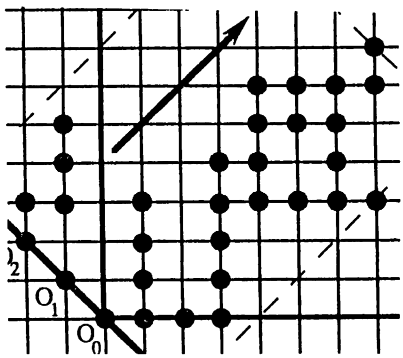


Fig. 7 Compact source directed animal. (NE-cs-directed)

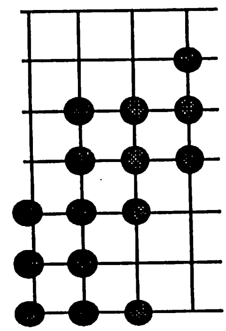


Fig. 8. Directed diagonally convex animal (Δ-convex + NE-directed)

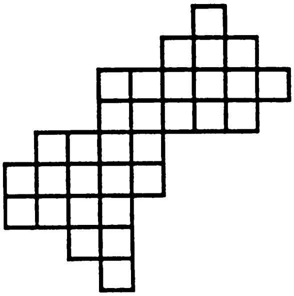


Fig. 9. Convex polyomino. (V-H-convex)

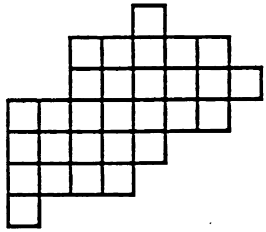


Fig. 10. Directed convex polyomino. (V-H-convex + NE-directed)

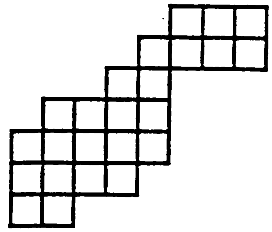


Fig. 11. Parallelogram polyomin (= V-H-Δ-convex + NE-SW-directed)

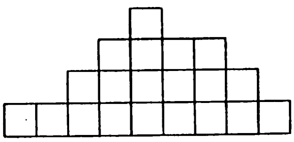


Fig. 12 Stack polyomino.
(= V-H-convex + NE-NW directed)

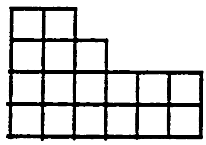


Fig. 13. Ferrers polyomino.
(= V-H-D-convex + NE-directed)

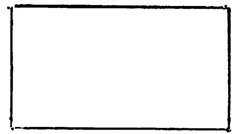


Fig. 14. Rectangle
(=V-H-D-Δ-convex
+ NE-SE-NW-SW-directed (!))

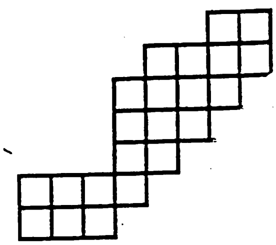


Fig. 15. Circles-wall polyomino
(= V-H-Δ-convex
+ NE-SE-directed + NW-cs-directed)

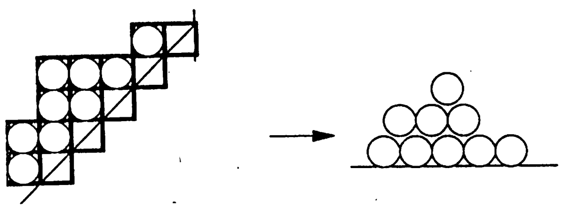


Fig. 16. Circles-stack polyomino
(= V-H-D-Δ-convex
+ NE-SE-directed + NW-cs-directed)

A *convex polyomino* is a polyomino which is both row- and column-convex. A characteristic property is that the perimeter is equal to the perimeter of the minimum rectangle containing the polyomino. There are five non-isomorphic classes of convex polyominoes: Ferrers diagram, stack , parallelogram (also called stair-case, and corresponding to skew Ferrers diagrams) , directed convex and convex. Parallelogram polyominoes are usually defined as polyominoes contained between two paths, each path having only North and East elementary steps, and such that the paths are disjoint, except at their common ending points. For the particular case where the lower path is an "enlarged staircase", we have the polyominoes displayed on Fig. 15. A subclass is in bijection with Andrews's quasi-partitions (Andrews 1981) and with "wall of circles" as considered in Privman, Švrakić 1989.

Each of these 14 families have at least one explicit enumeration formula. In many cases various formulae exist giving the triple generating function according to both area, perimeter and width. For the convex polyominoes, this distribution is equivalent to the distribution according to area, width and height. The only main problem which remains unsolved is the enumeration of the directed animals according to the perimeter. In fact 'perimeter' should be here replaced by 'directed perimeter'. Two perimeters exist for polyominoes: the one defined above is the *bond perimeter,* the *site perimeter* is the number of points outside of the animal which are neighbour (in the square lattice) to a point of the animal. More generally if F is a family of animals, the *F-perimeter* of an animal α of the family F is the number of points x outside the animal α, such that $\alpha \cup \{x\}$ still belongs to the family F. For F = {directed animals}, the F-perimeter is called the *directed perimeter*.

Here are two examples of non trivial enumeration formulae.

$$Y = y\frac{R-\hat{N}}{N},$$

$$N = \sum_{n\geq 0} \frac{(-1)^n x^n q^{\binom{n+1}{2}}}{(q)_n (yq)_n}, \quad \hat{N} = \sum_{n\geq 1} \frac{(-1)^n x^n q^{\binom{n+1}{2}}}{(q)_{n-1} (yq)_n},$$

$$R = y \sum_{n\geq 2} \left[\frac{x^n q^n}{(yq)_n} \left(\sum_{m=0}^{n-2} \frac{(-1)^m q^{\binom{m+2}{2}}}{(q)_m (yq^{m+1})_{n-m-1}} \right) \right].$$

Fig. 17. Generating fucntions for directed convex polyominoes
(according to width x , height y and area q)

3^n

Fig. 18. The number of NE-cs-directed animals.

yes it is really 3^n (!)

The first of the two formulae is not the most complicated. Just have a look at the paper Tzeng, Lin 1991, giving a four variables enumeration of column-convex animals (refinement of the enumeration according to the perimeter) to know what I mean. The second above formula is certainly the most simple non-trivial formula of the garden. Fig. 37 shows a bijection in action (between word of length n on the alphabet {A,B,C} and such cs-directed animals. Try to guess the construction !

If you take a guided tour among the known formulae, methods, technics, bijections related to the friendly zoo, may be you will make the suggestion, as I do , that all polyominoes of the friendly zoo should be enumerated (at least for one parameter) with the same kind of tools.

Curiously, some a priori unrelated regular patterns may appear. For example, look at the appearance of the radical Δ in the following four formulae enumerating, according to width and length respectively, the parallelogram, directed convex, convex polyominoes and a special type of convex polyominoes.

$$X = \frac{1-x-y-\sqrt{\Delta}}{2}, \qquad Y = \frac{xy}{\sqrt{\Delta}}, \qquad \Delta = 1-2x-2y-2xy+x^2+y^2.$$

$$Z = \frac{xy}{\Delta^2}\left(1-3x-3y+3x^2+3y^2+5xy-x^3-y^3-x^2y-xy^2-xy(x-y)^2\right)-\frac{4x^2y^2}{\Delta^{3/2}},$$

$$B = xy\frac{(1-x)(1-x-2y+y^2-xy)}{(1-x-y)\Delta},$$

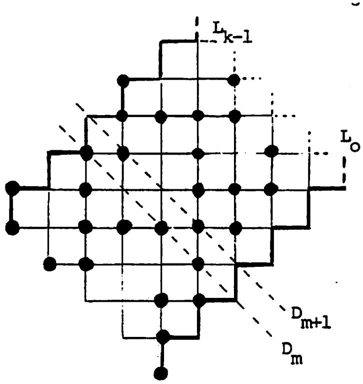
Generating functions for parallelograms, directed convex, convex and convex type B (width and height).

3. Technics and methodologies for polyominoes enumeration.

I have classified the different methods and tools used in the ninety papers listed below into six main classes. Of course, this is not a rigid classification and many papers uses several methods at the same times.

a) Transfer matrix methodology.

This method is very classical in statistical mechanics (see for example Baxter 1982). The combinatorial objects are put in bijection with some paths on a finite graph $\{1,\ldots, p\}$, such that the enumerative generating function becomes the generating function of weighted paths on the graph. Denote by A is the p p matrix with term (i,j) equal to the weight of the edge joining i to j. Then apply the inversion matrix formula for (I - At) and get a rational generating function. This method fits for animals on a bounded strip, as for example directed animals on a bounded strip (Fig. 19). Here is a formula for them (conjectured in Nadal, Derrida,Vannimenus, (1982) and proved in Hakim, Nadal, (1983)



$$\alpha_p=(2p+1)\frac{\pi}{2k}$$

$$\overline{a}_n^k(C) = \frac{1}{k}\sum_p (-1)^p \sin\alpha_p \prod_{i=1}^{k-1}\left(\frac{\sin[(i+\frac{1}{2})\alpha_p]}{\sin\frac{\alpha_p}{2}}\right)^{N_i} (1+2\cos\alpha_p)^{n-1}$$

Fig 19. Directed animal on a strip with enumerating formula (fixed sources points).

b) Temperley methodology.

This methodology has been introduced in Temperley (1956). It fits well with the different families of convex polyominoes. It has been used intensively used in the papers of Brak, Enting, Guttmann, Lin, Chang, Tzeng, Wu, Privman, Forgacs and Švrakič.

If t,x,y, .. are the different "edges" parameters as perimeter, width, height,.. and q is the parameter referring to the area, then the generating function f is decomposed into a sum of partial generating functions f_k satisfying a recurrence relation of the following type:

$$a_k f_k + (a_{k+1} f_{k+1}) + \ldots\ldots (a_{k+r} f_{k+r}) = 0,$$

where $a_k, \ldots\ldots, a_{k+r}$ are polynomials in the variables t,x,y, ..., q. Usually, f_k is the generating function for the polyominoes of the family under consideration having the first "machin" of size k, where "machin" means something like first row, first column or first diagonal. In general, for q=1, the polynomials a_{k+i} in the variables t,x,y, .. depends only upon i . The degree in the variable q of the monomials depends upon k. Of course, the resolution of the above recurrence may be not be easy and each case may take a lot of work. Continued fractions expansions may help.

c) DSV-methodology.

This methodology was introduced (and named "DSV") by M.P. Schützenberger (1962, 1963) thirty years ago, in relation with his work with N. Chomsky on the theory of algebraic languages (i.e. context-free). The background comes from linguistic and theoretical computer science (automata and languages theory). It fits very well in the case of an algebraic generating function. The principle is the following.

Denote by A^* the *free monoid* generated by A (or set of words on the alphabet A). A subset of A^* is called a *language*. Let Z<<A>> be the algebra of *non-commutative power series* in variables A and coefficients in Z. To each language $L \subseteq A^*$, we define the non-commutative generating function $\underline{\underline{L}}$ of L as the formal sum of all the words of L. For the reader not familiar with the (classical in computer science) notions of *algebraic* and *rational language, algebraic grammar* and *non-ambiguous grammar*, we give a simple (and fundamental) example.

The Dyck language D is the set of words w of $\{x,\bar{x}\}^*$ satisfying the two following conditions:

(i) the number $|w|_x$ of occurrences of the letter x in w is equal to the number $|w|_{\bar{x}}$ of occurrences of the letter $\bar{x}$,

(ii) for any left factor u of w = uv, $|w|_x \geq |w|_{\bar{x}}$.

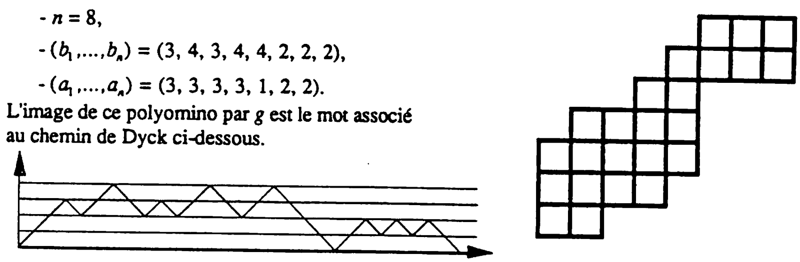


Fig. 20. Bijection between parallelogram polyominoes and Dyck words.

Such words are visualized by the so-called *Dyck paths* (see Fig. 20). Then D can be generated by applying several times the following substitutions rules:

$$D \to x D \bar{x} D \quad \text{or} \quad D \to e,$$

where e denote the *empty word*. This two rules define a "non-ambiguous grammar". The corresponding non-commutative generating function is

$$\underline{\underline{D}} = 1 + x\bar{x} + x\bar{x}x\bar{x} + xx\bar{x}\bar{x} + x\bar{x}x\bar{x}x\bar{x} + x\bar{x}xx\bar{x}\bar{x} + xx\bar{x}\bar{x}x\bar{x} + xx\bar{x}x\bar{x}\bar{x} + xxx\bar{x}\bar{x}\bar{x} + \dots$$

and satisfies the following algebraic equation in $Z\langle\langle\{x,\bar{x}\}\rangle\rangle$:

$$\underline{\underline{D}} = 1 + x\,\underline{\underline{D}}\,\bar{x}\,\underline{\underline{D}}.$$

This equation appears as a "linearization" of the non-ambiguous grammar. If we send all variables x and $\bar{x}$ onto t, then we go back to the classical commutative case giving the generating function for Catalan numbers: $\underline{\underline{D}}$ becomes the generating function y for the number a_n of words of D length n, the algebraic equation becomes $y = 1 + t^2y^2$.

Another classical algebraic language is the *bilateral Dyck* language, i.e. words of $\{x,\bar{x}\}^*$ satisfying only condition (i) of Dyck words. A visualization is given with path in Fig. 21. They are enumerated by the binomial coefficient $\binom{2n}{n}$.

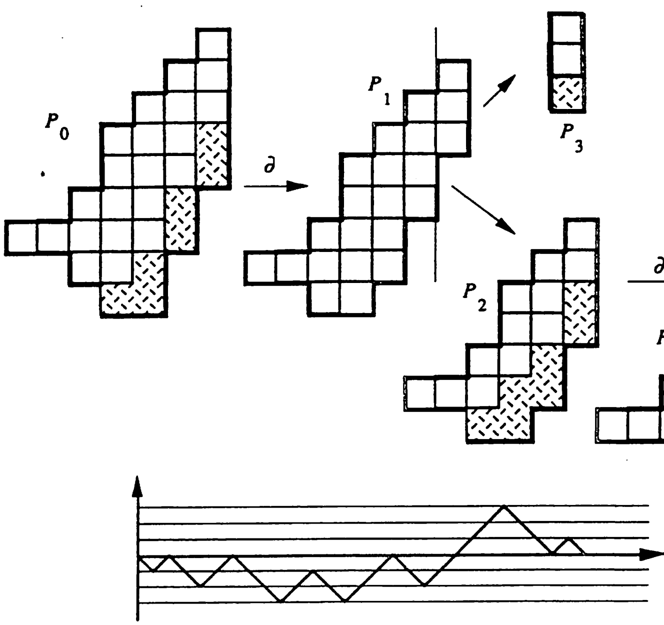


Fig. 21. Bijection between directed convex polyominoes and bilateral Dyck words.

The DSV methodology is in three steps. Let F be a class of combinatorial objects. The most simple form of the problem is to enumerate the objects having *"size"* n (size means any enumerative parameter). Let $f(t) = \Sigma_{n>0}\, a_n\, t^n$ be the corresponding generating function.

- (DSV1) Construct a bijection between the class F and some words of an algebraic language $L \subseteq A^*$ defined by a non-ambiguous grammar such that the the size of the combinatorial objects is the length of the corresponding words (or a linear function of it).
- (DSV2) Write a non-ambiguous algebraic grammar and "translate" it into an algebraic system of equations in the algebra $Z\langle\langle A\rangle\rangle$.
- (DSV3) Applying the morphism sending the letters onto t, we get an algebraic equation for f(t), which can be possibly simplified or computed.

Example: the appearance of the common radical Δ in the four formulae shown at the end of section 2 can be better understood with the DSV methodology. The coding and the corresponding formulae for the perimeter enumeration (obtained by identifying the width and height parameter into a single variable) are the following (respectively parallelogream, directed convex, convex, convex type "B" polyominoes).

Proposition. *(i) Le nombre de polyominos parallélogrammes de périmètre 2n+2 est le $n^{ième}$ nombre de Catalan, soit* $\frac{1}{n+1}\binom{2n}{n}$.
(ii) Le nombre de polyominos convexes dirigés de périmètre 2n+4 est $\binom{2n}{n}$.
(iii) Le nombre de polyominos convexes de périmètre 2n+8 est $(2n+11)4^n - 4(2n+1)\binom{2n}{n}$.
(iv) Le nombre de polyominos de $\mathfrak{B}$ *de périmètre 2n+8 est* $6.4^n + 2^n$.

If the problem contains several enumeration parameters, the method is exactly the same if the coding can be done such that each letter corresponds to one of the parameters, with a linear relation between the length of the word and the "size" of the combinatorial objects. In the last step, we just commute the variables and get the multivariate generating function.

DSV methodology was first illustrated by R. Cori and his students with planar maps (Cori 1975, Cori, Vauquelin 1981, for example), explaining the reason of the algebricity of various Tutte's formulae. After, DSV methodology was used to solve open problems in combinatorics, as for example a problem posed by M. Waterman: enumeration of *secondary structures of single stranded nucleic acids* having a given *complexity* (Vauchaussade de Chaumont, Viennot, 1985). The first open problem in polyominoes enumeration solved by this method is the number of convex polyominoes according to the perimeter (Delest, Viennot, 1984). The reader will find a survey in Viennot, 1985.

In the case of a language accepted by a *finite automaton* (*recognizable* or *rational* langage), then the ordinary generating function will be rational and DSV methodology is similar to transition matrix method.

For all the families of the friendly polyominoes zoo, DSV methodology can be applied (at least for one of the parameter area or perimeter) as shown by Gouyou-Beauchamps and the bordelais group (see papers of Betrema, Bousquet-Mélou, Delest, Dulucq, Fedou, Lalanne, Penaud, Viennot, ..) where DSV is very popular. For the double generating function, a q-analog of DSV can be introduced (see papers of Delest, Fedou and Bousquet-Melou). The area is coded by some powers of the variable q which appears in the algebraic grammar of the language. These *"q-grammars"* have some analogy with the so called *attribute grammars* introduced by Knuth. A recent survey of DSV and q-DSV methodology applied to polyominoes enumeration is given in Delest 1991.

d) Heaps of pieces and Cartier-Foata commutation monoid.

Let P be a set (called set of *basic pieces*). Let C be a binary symmetric and reflexive relation on P called *concurrency relation*. Thus the pair (P,C) defined a graph, called the *concurrency graph.*

A *heap* is finite set E of pairs $\alpha = \{p, j\}$ with $p \in P$, $j \in N$ satisfying the two following relations:
(i) if $\{p, j\}$ and $\{q, k\}$ are two elements of E with $p\,Cq$, then $j \neq k$.
(ii) if $\{p, j\}$ is in E and $j = 0$, then there exist $\{q, k\}$ in E such that pCq and $k = j-1$.

The elements $\{p, j\}$ of E are called *pieces*. The basic piece p is called the *projection* of $\{p, j\}$, while i is called the *level* of the piece $\{p, j\}$. The set of all heaps on (P,C) is denoted by Heap(P,C),

This concept was introduced in Viennot (1985), and has been useful in various part of combinatorics including combinatorial proof in classical linear algebra, combinatorial theory of general orthogonal polynomials, algebraic graph theory, and combinatorial problems related to statistical physics, in particular animals enumeration. The intuitive idea behind the heap concept is better understood in the case of heaps of subsets. Here the set P is a certain collection of subsets of a set B. The concurrency relation C is defined by: aCb iff $a \cap b \neq \emptyset$.

Suppose B is $R \times R$ and represented by an horizontal board, and that each piece α is represented by a solid piece of wood with small constant width, and projection a on B. Then the concept of heap corresponds to the picture obtained by putting one by one solid pieces on B. See Fig. 23, 24, where the basic pieces are respectively *dominos* (polyominos with two cells) of the chessboard or *hexagons* of the triangular lattice. Another example is displayed on Fig. 25. Here B = Z, the basic pieces are segments [i,j] of Z, the concurrency relation is the same as defined above. A partially order relation, called *"to be below"* can be defined for any heap. Intuitively α is below β iff one has to remove first β in order to remove α. The Hasse diagram of the order relation is displayed on Fig. 28. Conversely, any poset can be represented as a heap of pieces. A *pyramid* is a heap having only one maximal piece.

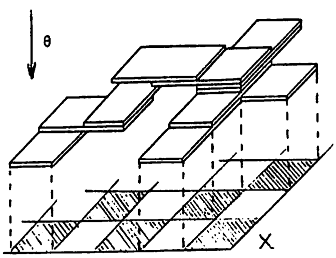


Fig. 23. Heap of "solid dimers" over R R.

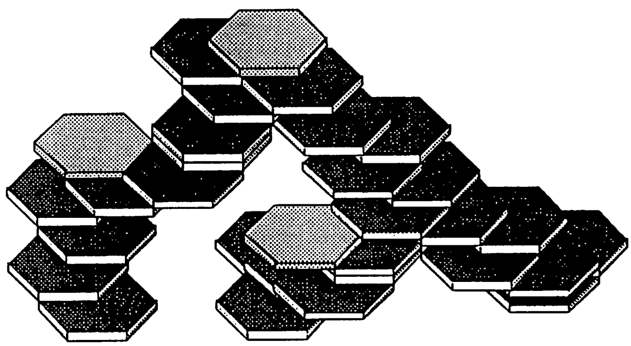
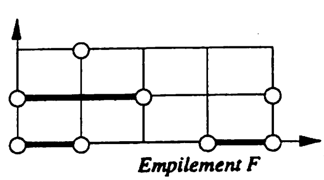
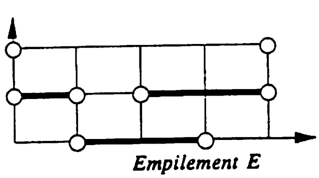


Fig. 24 Heaps of "solid" hexagons.

In fact one can defined a product of two heaps E and F: intuitively put F far above E and let it fall down on E. The set Heap(P,C) becomes a monoid, and this monoid can be defined by some partial commutation rules. Let $\equiv_C$ be the congruence on the free monoid P^* generated by the commutation ab = ba for each pair of basic pieces which are not in concurrence. Then the heap monoid Heap(P,C) is isomorphic to the quotient monoid $P^*/\equiv_C$. Such monoids have been introduced in Cartier, Foata 1969. They have been intensively used in theoretical computer science as a model for concurrency and parallelism problems. They are called *trace monoids* (also *commutation monoid*). Conversely every trace monoid is a heap monoid and the two concepts are equivalent. The advantage of heaps is to provide a geometric interpretation of the equivalence classes of words (called *traces*) with a powerful spatial intuition.





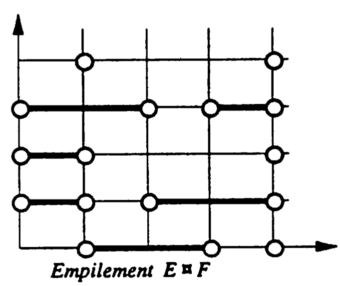


Fig. 25. Product of two heaps.

ad = da
bc = cb
cd = dc

Fig 26. Commutation relation

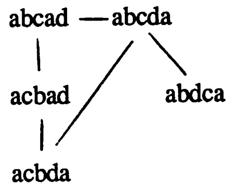


Fig. 27. A trace (equivalence class)

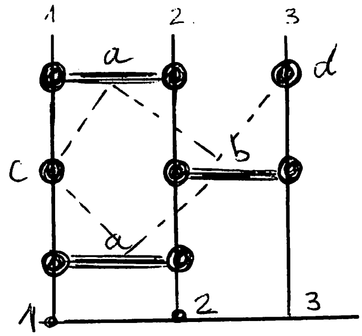


Fig. 28. Realization as a heap and associated Hasse diagram

Heaps have been very useful in polyominoes enumeration. Some families of polyominoes can be put in bijection with some heaps. In particular the parallelogram polyominoes are in bijection with pyramids of segments of N such that the projection of the maximal piece is a segment [0,j]. The directed animals are in bijection with certains pyramids of dimers on Z , and through another bijection with pyramids of momers and dimers on Z (*dimers* are segments [i,i+1], while *monomers* are segments [i]).

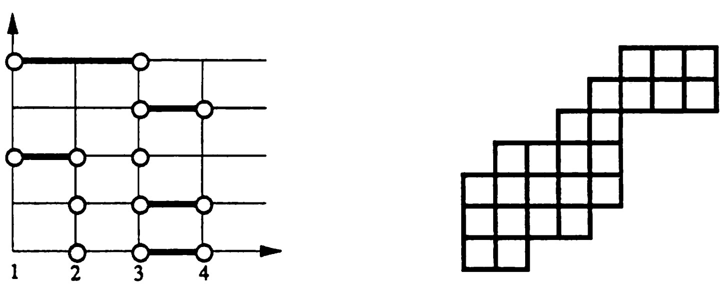


Fig. 29 Bijection between parallelogram polyominoes and pyramid of segments.

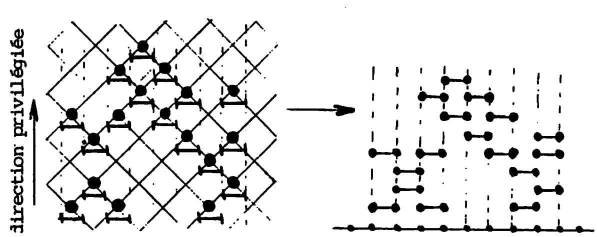


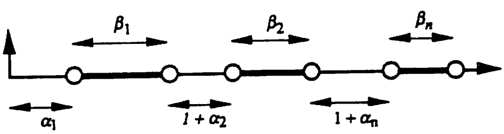
Fig. 30. Bijection between directed animal and heap of dimers.

Three basic facts on heaps are used. These lemma can be summarized schematically by the following:

Paths=Heaps; Pyramids= d/dt(Heaps); Heaps= 1/D and Heaps$^{(*)}$= N/D.

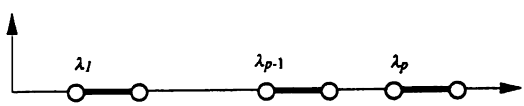
The first lemma described a bijection between paths (on any lattice) and heaps. The three others identities are generating functionsof weighted heaps: each basic piece a is given a weight v(a) and the weight of a heap is the product of the weight of the projection of its pieces. The last equation represents the generating function for weighted heaps such that the projection of the maximal pieces are in a given fixed set $M \subseteq P$. The numerator N and the denominator D are the generating function for trivial heaps (i.e. all the pieces are at level 0). For more details see Viennot, 1985.

In the case of convex polyominoes (parallelograms, directed convex and convex), heaps of segments are considered with weight $v([i,j]) = t\, u^{(j-i)} q^j$. Trivial heaps of segments generate the q-Bessel fuunctions (see Fig. 31) appearing in Fig. 18. In the case of trivial heap of dimers, left hand side of Rogers-Ramanujan identies appear (see Fig. 32. in relation with Fig. 15 and 16). (see Bousquet-Mélou 1991 and Bousquet-Mélou, Viennot 1992)



$$N = \sum_{n\ge 0} \frac{(-1)^n x^n q^{\binom{n+1}{2}}}{(q)_n (yq)_n},$$

Fig. 31. Trivial heap of segments over Z (q-Bessel fonctions).



$$\sum_{n\ge 0} \frac{(-x)^n q^{n+n^2}}{(q)_n} \Big/ \sum_{n\ge 0} \frac{(-x)^n q^{n^2}}{(q)_n},$$

Fig. 32. Trivial heap of dimers over Z (left hand sides of Rogers-Ramanujan identities).

$$\cfrac{1}{1-\cfrac{xq}{1-\cfrac{xq^2}{1-\cfrac{xq^3}{1-\dots}}}}.$$

Another example of the power of heap methodology is a very comprehensive proof for the number of directed animals on a bounded circular strip. The formula given in Fig 19 corresponds to a heaps generating function N/D, where N and D are respectively the generating function for trivial heaps of dimers (here q = 1) on a disjoint union of segments and on a circle (i.e. , up to a change of variable, Tchebycheff polynomials of first and second first kind).

e) Classical analytic methods.

In this category I would list the classical analytic tools used in combinatorial problems: recurrence relations, q=Lagrange inversion, continued fractions expansions, ...

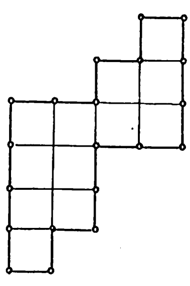
f) Bijective methods

Here I would list enumeration formulae obtained from the construction of explicit bijections between animals and other combinatorial objects which are easier to enumerate. We have already mentioned such bijections with words of algebraic languages and with heaps. Another families of bijections which appear in many papers involve bijections between polyominoes and various families of trees. Fig. 35, 36, 37 show binary trees in the case of parallelogram polyominos, colored binary trees for convex polyominoes, *"guingois"* trees for directed animal. In fact algebraic equation on trees can be defined, analogous to algebraic grammars of words. A surprise is the existence of a bijection between *ternary trees* and directed diagonally convex animals (enumerated according to the directed perimeter).

g) Experimental combinatorics.

A few words about experimental hunting for finding or proving enumeration formulae. Here the tools are a computer with a symbolic algebraic package (as for example your favorite MATHEMACSYMAPLE) and the Sloane'book (thank's to Simon Plouffe, a new version is coming with more than 4000 sequences). Some formulae have been discovered or guessed with these tools. Some technics have been developed in order to guess from the first terms of the sequence (20 to 100) a possible explicit expression for the generating function. In case of a rational power series, excellent algorithms are used based on Pade approximants theory, you cannot miss it . For P-recursive generating functions, the so-called differential Pade method have been introduced in Joyce, Guttmann (1972) and is popular in statistical mechanics for experimental asymptotic results. In the case of algebraic power series, see Brak, Guttmann, (1990a). Contrarily to the case of rational fractions, the method is rather brute force. Nevertheless, some algebraic functions have been guessed this way. The generating function for convex polyominoes was guessed in Guttmann, Enting (1988a), independently of the paper Delest, Viennot (1984). Of course such experimental methods give only mathematical conjectures and have not to be compared with the six methods exposed above. Usually, in the physics literature, the distinction between establishing a formula from computer experiments and giving a mathematical proof is not clearly stated. The experiments can be fundamental for guessing the formula. When the number of known elements of the sequence is much bigger that the number of elements needed to guess the exact formula, the probability that the formula is wrong is infinitesimal, and thus the formula can be considered as an experimental true statement. But mathematically it is just a conjecture waiting for a proof, and may be more: a crystal-clear understanding. Remark that some rigorous proof for polyominoes enumeration, need some huge computation on a computer. The computer is used at an other level, as part of the proof.

4) A nice example: parallelogram polyominoes



All the methodologies mentioned above can be applied for parallelograms polyominoes. The enumeration according to the perimeter gives the classical Catalan numbers. There is plenty of related bijections involving Dyck words, binary trees, heaps of segments, .. Some magic coincidences appears on Fig, 33. Penaud has used similar bijections with parallelogramm polyominoes in order to give a bijective proof of a formula of Riordan-Touchard giving the moments of some q-Hermite polynomials. (see Penaud, 1992 at this collque)

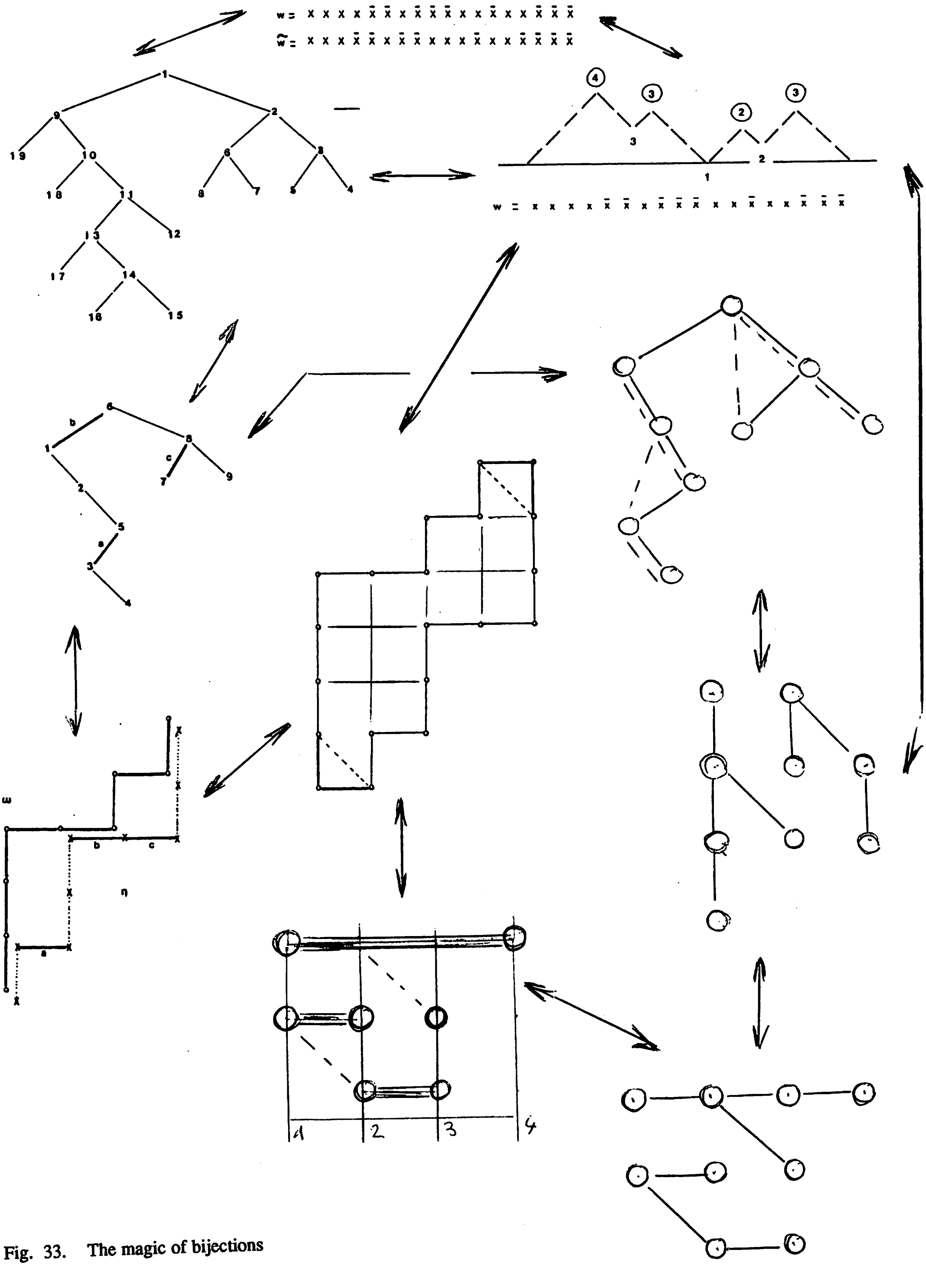


Fig. 33. The magic of bijections

5) The garden of polyominoes bijections.

After the guided tour in the zoo, and the magic network of bijections for parallelograms polyominoes, here are some example of polyominoes bijections. Some are sophisticated.

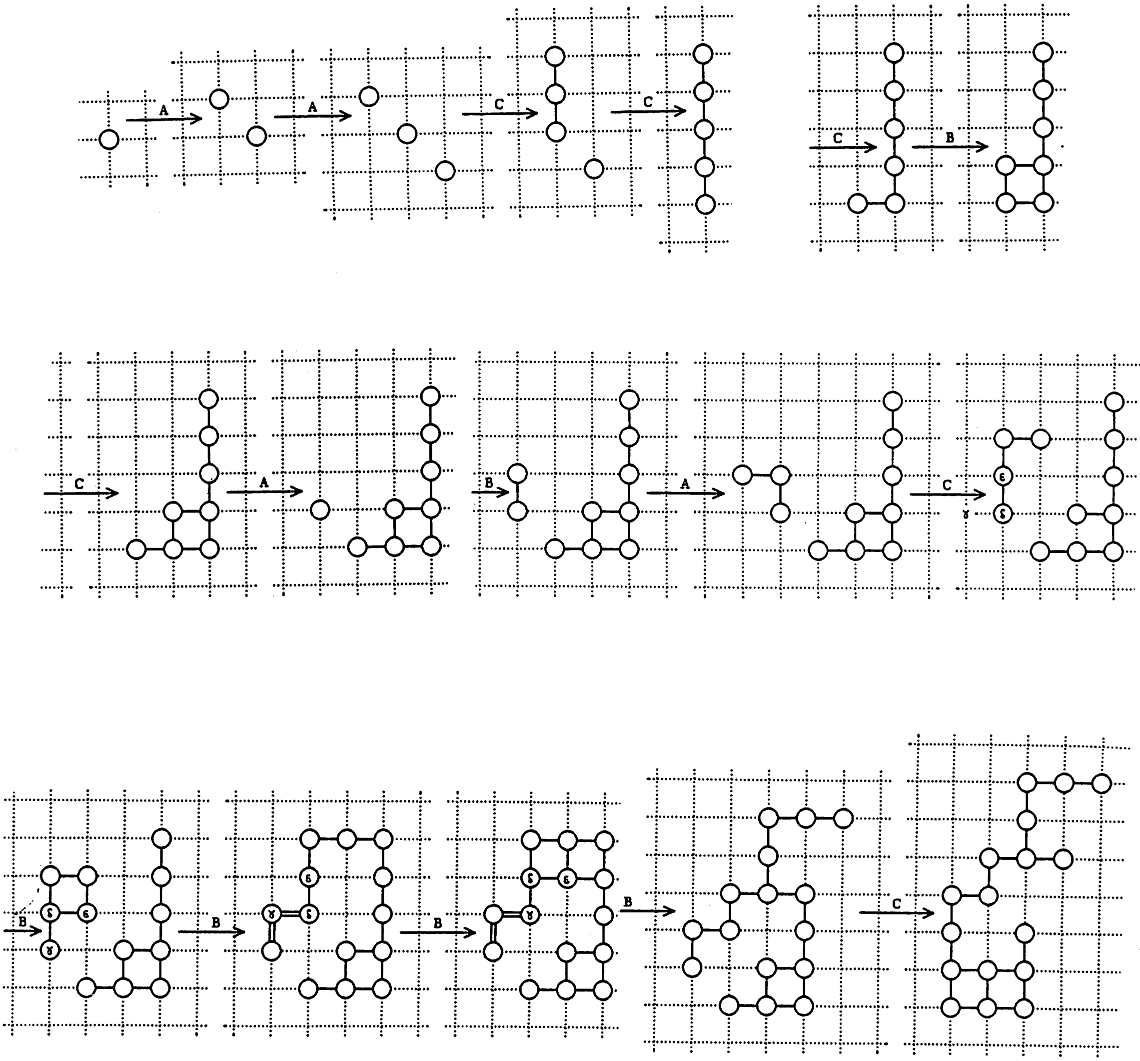


Fig. 34. A bijection for 3^n (compact source directed animals).
try to guess the construction !

(Gouyou-Beauchamps - Viennot, 1988)

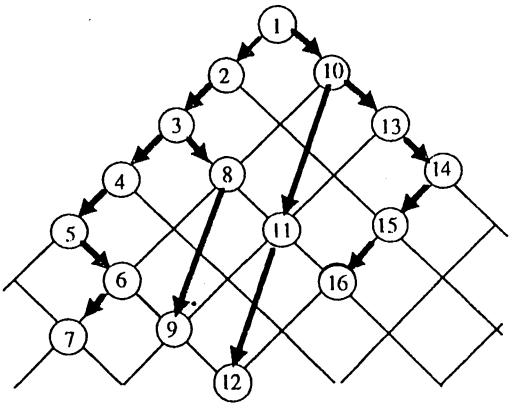


Fig. 35. Bijection between directed animals and "guingois" trees.

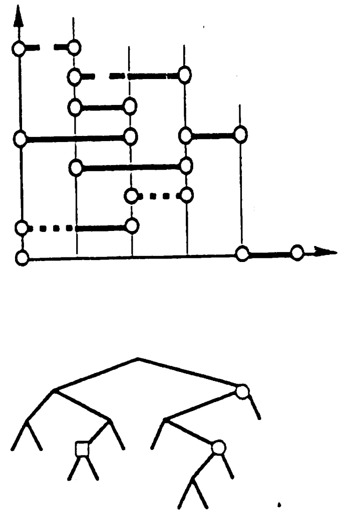


Fig. 36. Bijection between colored heaps of segments and convex polyominoes with projection on acolored binary tree.

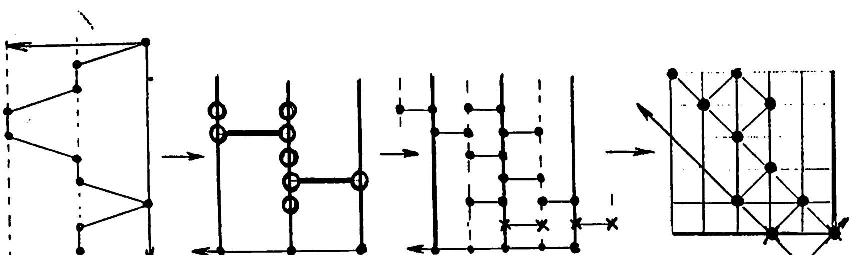


Fig. 37. Path, dimers-monomers heap and directed animal.

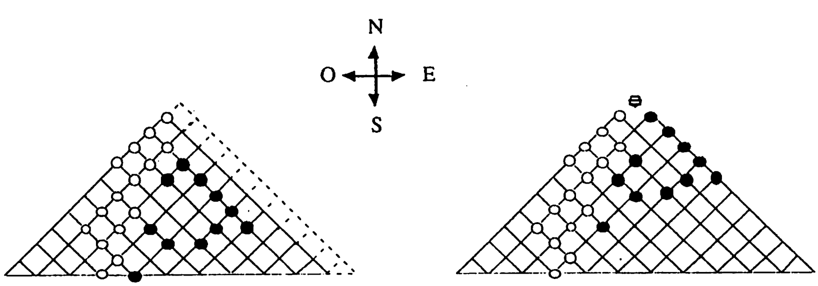


Fig. 38. Penaud's operator for directed animal.

6) Related topics

One of the advantage of using bijective methods in polyominoes enumeration is to show connections with some other apparently completely unrelated topics. For example, we quote here: Ehrhart' theory for counting points with integers coordinate in a convex polytope (see Stanley 1986, Fedou 1989, Delest-Fedou 1991b); basic hypergeometric functions (in particular some q-Bessel functions); the computation of the directed percolation probability (see Fig. 39 and Bousquet-Melou, 1990b); Roger-Ramanujan identidies and continued fractions apppearing in the special class of parallelograms polyominoes displayed on Fig. 15, 16 (in bijection with Andrew's quasi-partitions (Andrews, 1981) and "circles wall" animals (Privman, Švrakič 1989)).

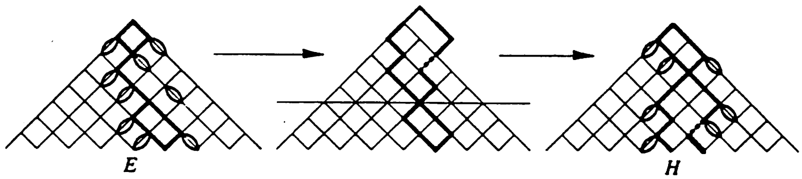


Fig. 39 Modified parallelogram polyominoes appearing in the directed percolation problem.

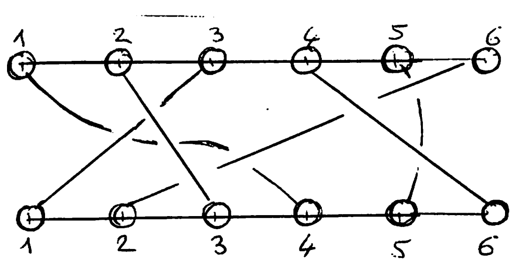
Another advantage of bijective methods is in the random generation of polyominoes. For example, using the bijection with "guingois" trees (Betrema, Penaud, 1991) and an algorithm of Barcucci, Pinzani, 1991, random animals can be generated in linear time. This means that each directed animal having n points appears with same probability (as soon as one can generate equiprobably a number between 1 and k). This operation is supposed to be a primitive with a cost O(1). Fig. 40 shows some random animals. Their fractal dimensions seems to be related to the critical exponents for width and length.

Fig. 40 Random directed animals (size 1000 and 2000).

I will finish with a last surprise. The generating function for parallelogram polyominoes according to the area and having a fixed number of columns is a rationnal fraction of the following form: $(b)^{n}q \,/\, (q;q)_n(q;q)_{n-1}$ where $(q;q)_n= (1-q)(1-q^2)\ldots(1-q^n)$.

The numerator b_n (q) is a polynomial with positive integers coefficients (Fedou 1989). Fedou has just proved (see Fedou 1992 at this colloque) that these coefficients enumerate certain braids of the braid group B_n according to a certain parmeter analog to a Markov trace.

Fig. 41. Heaps and Braids



References

G.E. ANDREWS (1986), *q-Series: their development and applications in analysis, number theory, combinatorics, physics, and computer algebra*, Amer.Math. Soc.

G.E. ANDREWS (1981), *The Rogers-Ramanujan reciprocal and Minc's partition function*, Pacific J. of Math. 95, 251-256.

F.C. AULUCK (1951), Proc. Cambridge. Phil. Soc. 47, 679

E. BARCUCCI, R. PINZANI and E. RODELLA (1990), *Some properties of binary search networks*, Research report 1/90, Dipartimento di Sistemi e Informatica, University of Firenze.

E. BARCUCCI, R. PINZANI and E. RODELLA (1991), *Enumeration of directed vertically convex animals, Research report*, Dipartimento di Sistemi e Informatica, University of Firenze.

E. BARCUCCI, R. PINZANI, *Generation of random directed animals*, in Proc. Atelier franco-québecois de combinatoire, Bordeaux, May 1991, Publication du LACIM, n°9, Université du Québec a Montréal, Montréal,, to appear.

R. J. BAXTER (1982), *Exactly solved models in statistical mechanics*, Academic Press, New-York, .

R.J. BAXTER and A.J. GUTTMANN (1988), *Series expansions of the percolation probability for the directed square lattice*, J. Phys. A: Math. Gene. 21, 3193-3204.

E. BENDER (1974), *Convex n-ominoes*, Discrete Math 8 , 219-226.

J. BETREMA, J.G. PENAUD (1991), *Animaux et arbres guingois*, in Proc. 3ème Colloque "Séries formelles et combinatoire algébrique", LaBRI, Université Bordeaux I.

V.K.BHAT, H.L. BHAN and Y. SINGH (1986), J. Phys.A: Math. Gene 19, 3261-3265.

M. BOUSQUET-MELOU (1990a), *Une bijection entre les polyominos convexes dirigés et les mots de Dyck bilatères*, rapport LaBRI 90-77, to be published in Info. Th. et applications.

M. BOUSQUET-MELOU (1990b), *Nombres de Catalan et percolation dirigée*, rapport LaBRI 90-79.

M. BOUSQUET-MELOU (1990c), *Codage des polyominos convexes et équations pour l'énumération selon l'aire*, Discrete Applied Maths, to be published

M. BOUSQUET-MELOU (1991a), *q-énumération de polyominos convexes*, Publication du LACIM, n° 9, Université du Québec à Montréal, Montréal, 168 p. (Thèse de Doctorat, Université Bordeaux I).

M. BOUSQUET-MELOU (1991b), *q-Enumération de polyominos convexes*, J. Combinatorial Th. A, to be published.

M. BOUSQUET-MELOU (1991c), *Convex polyominoes and heaps of segments* J. Phys. A: Math. Gene., to be published.

M. BOUSQUET-MELOU (1991d), *Convex polyominoes and algebraic languages*, J. Phys. A: Math. Gene., to be published.

M. BOUSQUET-MELOU and X.G.VIENNOT (1991), *Empilements de segments et q-énumération de polyominos convexes dirigés*, J. Combinatorial Th. A, to be published.

R.BRAK and A.J. GUTTMAN (1990a), *Algebraic approximants: a new method of series analysis*, J. Phys. A: Math.Gene. 23, L1331-L1377.

R.BRAK and A.J. GUTTMAN (1990b), *Exact solution of the staircase and row convex polygon perimeter and area generating function*, J.Phys A, 23, (1990) 4581-4588.

R.BRAK, A.J. GUTTMAN and I.G.ENTING (1990), *Exact solution of the row convex polygon perimeter generating function*, J. Phys. A Math. Gene. 23, 2319-2326.

L. CARLITZ (1967), *Integers related with Bessel functions*, Proc. Amer. Math. Soc, 14, 1-9.

P. CARTIER and D.FOATA (1969), *Problèmes combinatoires de commutations et réarrangements*, Lecture Notes in Maths, n° 85, Springer-Verlag, Berlin.

R. CORI (1975)*Un code pour les graphes planaires et ses applications*, Astérisque, Soc.Math.France n° 27.

R. CORI et B. VAUQUELIN (1981), *Planars maps are well labeled trees*, Can J. Math. 33 , 1023-1042.

M.P. DELEST (1987), *Utilisation des langages algébrique et du calcul formel pour le codage et l'enumeration de polyominos*, Thèse d'Etat, Université Bordeaux I.

M.P. DELEST (1988), *Generating functions for column-convex polyminoes*, J. Combinatorial Th. A, 48, 12-31.

M.P.DELEST (1991a), *Enumeration of Polyominoes using MACSYMA*, Theor. Comp. Sci, , to be published.

M.P.DELEST (1991b), *Polyominoes and animals: some recent results*, Proc. MATH/CHEM/COMP 90,Dubrovnik, Croatie, J. of Computer and Chemistry, 8, 3-18.

M.P. DELEST and S. DULUCQ (1986), *Enumeration of directed column-convex animals with given perimeter and area,* rapport LaBRI n° 86-15, Université de Bordeaux I.

M.P.DELEST and J.M. FEDOU (1989), *Exact formulas for fully compact animals*, rapport LaBRI 89-06.

M.P.DELEST, J.M. FEDOU (1990), *Counting polyominoes using attribute grammars*, Proceedings Workshop on attribute grammars 1990, P. Deransart and M. Jourdan eds, Lecture Notes in Computer Science, 461 , 46-60..

M.P.DELEST and J.M. FEDOU (1991a), *Enumeration of Skew Ferrers diagrams*, to appear in Discrete Math. (prepublication in Actes du seminaire lotharingien 1990, publi. IRMA, Université Strasbourg I).

M.P.DELEST and J.M. FEDOU (1991b), *Introduction aux polyèdres en combinatoire d'après E. Ehrhart et R. Stanley*, Info. Théor. et Appl., to be published.

M.P.DELEST and J.M. FEDOU (1991c), *Attribute grammars are useful for combinatorics*, Theor. Comp. Sci., to be published.

M. DELEST, D. GOUYOU BEAUCHAMPS et B. VAUQUELIN (1987), *Enumeration of parallelogram polyominoes with given bond and site perimeter*, Graphs and Combinatorics, 3, 325-339.

M.P.DELEST and G.VIENNOT (1984), *Algebraic langages and polyominoes enumeration*, Theor. Comp. Sci. 34 , 169-206.

B. DERRIDA and J.P. NADAL (1984) J. Physique , 45, L701-L706.

S.DULUCQ (1987), *Etude combinatoire de problèmes d'énumeration, d'algorithmique sur les arbres et de codage par des mots*, Thèse d'Etat, Université Bordeaux I.

G. de GENNES (1979), "Scaling concepts in polymer Physics" (Cornell. Univ. Press, Ithaca, NY).

T. de NEEF and I.G.ENTING (1977), J. Phys. A : Math. Gene., 10, 801-805.

D. DHAR (1982), *Equivalence of the two-dimensional directed animal problem to Baxter hard-square lattice-gas model*, Phys. Rev Lett. 49, 959-962.

D. DHAR (1983), *Exact solution of a directed-site animals enumeration in 3 dimensions*, Phys. Rev Lett. 59 , 853-856.

D. DHAR, M.K. PHANI, M. BARMA (1982), *Enumeration of directed site animals on two-dimensional lattices,* J.Phys.A : Math Gen.15, L 279 -L 284.

I.G. ENTING (1980), J. Phys. A : Math. Gen. 13, 3713-3722.

I.G. ENTING and A.J. GUTTMANN (1985), J. Phys. A : Math. Gen. 18, 1007-1017.

I.G. ENTING and A.J. GUTTMANN (1989a), J. Phys. A : Math. Gen. 22, 1371-1384.

I.G. ENTING and A.J. GUTTMANN (1989b), *Area-weighted moments of convex polygons on the square lattice*, J. Phys. A : Math. Gen. 22, 2639-2642.

I.G. ENTING, A.J. GUTTMANN, L.B. RICHMOND, N.C. WORMALD (1991), in "Random structures and algorithms".

J.M. FEDOU, (1989) *Grammaires et q-énumeration de polyominos,* Thèse de Doctorat, Université Bordeaux I.

J.M. FEDOU (1990), *Enumeration of skew Ferrers diagrams and basic Bessel functions*, in Proc. 2nd conference on lattice path and combinatorics and applications, Hamilton 1990.

J.M. FEDOU and X.G.VIENNOT (1990), *Chains in poset and enumeration of skew Ferrers diagrams*, prepint LaBRI.

P.FLAJOLET (1982), *Combinatorial aspect of continued fractions*, Discrete Maths., 41, 145-153.

P. FLAJOLET (1991), *Polya' festoons*, preprint , INRIA, Rocquencourt.

W. GAUTSCHI (1967), SIAM Review, 9, 24.

I.GESSEL (1980), *A non-commutative generalization and q-analog of the Lagrange inversion formula*, Trans. Amer. Math. Soc., 257, 455-482.

I. GESSEL (1990), *On the number of convex polygons,* preprint, Brandeis University.

M.L. GLASSER, V. PRIVMAN and N.M. SVRAKIC (1987), J. Phys. A: Math. Gene. L1275.

D. GOUYOU-BEAUCHAMPS (1985), *Codage par des mots et des chemins: problèmes combinatoires et algorithmiques*, Thèse d'Etat, Université Bordeaux I

D. GOUYOU-BEAUCHAMPS and X.G. VIENNOT (1988), *Equivalence of the two-dimensional directed animal problem to a one-dimensional path problem*, Advances in Applied Mathematics 9, 334-357.

A.J. GUTTMANN (1982), *On the number of lattice animals embedable in the square lattice*, J. Phys. A:Math. Gen. 15 , 1987-1990.

A.J. GUTTMANN (1991), *Planar polygons, regular, convex, almost convex, staircase and row convex*, preprint, Melbourne,.

A.J. GUTTMANN, I.G. ENTING (1988a), *The number of convex polygons on the square and honeycomb lattices*, J. Phys. A: Math. Gen. 21 , L467-L474.

A.J. GUTTMAN and I.G.ENTING (1988b), J. Phys. A : Math. Gene. 21, L165-172.

A.J. GUTTMAN and I.G.ENTING (1988c), J. Phys. A : Math. Gene., 21, L467.

A.J. GUTTMAN and G.S.JOYCE (1972), J. Phys. A : Math. Gene., 5, L81.

V. HAKIM and J.P. NADAL (1983), *Exact result for 2D directed lattice animals on a strip of finite width* , J. Phys. A: Math. Gen. 16 , L 213-L 218.

G.S.JOYCE and A.J. GUTTMAN (1972), in "Padé approximants and their applications", ed. P.R. GRAVES-MORRIS, Academic Press, New-York, pp 163-167.

D. KIM (1988) , *The number of convex polyominoes*, Discrete Maths. 70, 47.

D.A. KLARNER (1965), *Some results concerning polyominoes*, Fibonacci Quart. 3 , 9-20.

D.A. KLARNER (1967), *Cell growth problems*, Canad. J. Math 19, 851-853.

D.A. KLARNER (1981), *My life among polyominoes*, in "The Mathematical Gardner", 243-262, Wadsworth, Belmont CA.

D.A. KLARNER and R.L. RIVEST (1974), *Asymptotic bounds for the number of convex n-ominoes*, Discrete Maths 8 , 31-40.

J.C. LALANNE (1991a), *q-Bessel et polyominos*, in 'Actes du 25ème Séminaire Lotharingien de combinatoire", Publ. IRMA, Université Strasbourg I.

J.C.LALANNE (1991b), *Polyominos parallelogrammes a franges et fonctions de Bessel*, Discrete Maths, to be published.

J.C.LALANNE (1991c), *q-Bessel, arbres et chemins valué* , submitted.

J.C.LALANNE (1991d), *Une involution sur les chemins de Dyck*, submitted.

K.Y. LIN (1990a), J. Phys. A: Math. Gene 23, 4703-4705.

K.Y. LIN (1990b), Int. J. Mod. Phys. B4, 1717-1741.

K.Y. LIN (1991a), Chin. J. Phys. 29, 7-15.

K.Y. LIN (1991b), *Exact solution of the convex polygon perimeter and area generating function*, J. Phys. A: Math. Gene. 24, 2411-2417.

K.Y. LIN and S.J. CHANG (1988), *Rigourous results for the number of convex polygons on the square and honeycomb lattices*, J. Phys. A: Math. Gen. 21, 2635-2642..

K.Y. LIN, S.K.MA, C.H. KAO and S.H CHIU (1987), J. Phys. A: Math. Gene. 20, 1881-1888.

K.Y. LIN and F.Y. WU (1990) J. Phys. A: Math. Gene, 23, 5003.

S.K. MA and K.Y LIN (1988), Int. J. Mod. Phys. B2.

J.P. NADAL, B.DERRIDA and J. VANNIMENUS (1982), *Directed lattices animals in 2 dimension: numerical and exact results*, J. Physique 43 , 1561.

B. NIENHUIS (1982), Phys. Rev. Lett., 49, 1062.

B. NIENHUIS (1984) , J.Stat. Phys., 34, 731

G. POLYA (1969), *On the number of certain lattice polygons*, J. Comb. Theory, 6, 102-105.

J.G. PENAUD (1989), *Une nouvelle bijection pour les animaux dirigés*, Actes du. 22ème Séminaire Lotharingien de Combinatoire, Publi. IRMA, Univ. Strasbourg I.

J.G. PENAUD (1990a), *Animaux dirigés diagonalement convexes et arbres ternaires*, rapport LaBRI n°90-62, Université de Bordeaux I.

J.G. PENAUD (1990b), *Arbres et Animaux*, Mémoire d'Habilitation, Université Bordeaux I.

J.G. PENAUD (1992), *Une preuve combinatoire d'une formule de Touchard-Riordan*, in Proc. 4th Colloque "Formal power series and algebraic combinatorics", Montreal, June 1992, this volume.

V. PRIVMAN and G. FORGACS (1987), *Exact solution of the partially directed compact lattice animal model*, J. Phys. A; Math Gene. 20 L543-L547.

V. PRIVMAN, N. M. ŠVRAKIČ (1988), *Exact generating function for fully directed compact lattice animals*, Physical Review Letters, 60, 12, 1107-1109.

V. PRIVMAN, N. M. ŠVRAKIČ (1989), Directed models of polymers, interfaces and clusters: scaling and finite-size properties, Lecture Notes in Phys. 338 , Springer-Verlag.

R.C. READ (1962), *Contribution to the cell growth problem*, Canad. J. of Math., 14, 1-20.

M.P.SCHÜTZENBERGER (1962),*Certain elementary families of automata*, Proc. Symp. on Mathematical Theory of Automata (Polytechnic Institute of Brooklyn, pp. 139-153.

M.P.SCHÜTZENBERGER (1963), *Context-free langages and pushdown automata*, Information and Control 6 , 246-264.

N.J. SLOANE, *A handbook of integer sequences,* Academic Press, New-York, 1979.

R. STANLEY (1986), Enumerative combinatorics, vol I, (example 4.7.18), Wadsworth & Brooks, Monterey, California.

H.N.V. TEMPERLEY (1952), Proc. Cambridge Phil. Soc., 48, 683-697.

H.N.V. TEMPERLEY (1956), Phys. Rev. 103, 1.

W.J. TZENG and K.Y.LIN (1991), *Exact solution of the row-convex polygon generating function for the checkerboard lattice*, Int. J. Mod. Phys. B, to be published.

M. VAUCHAUSSADE de CHAUMONT and G.VIENNOT (1985), *Enumeration of RNAs Secondary structure by complexity*, in "Mathematics in Medecine and Biology ", Lecture Notes in Biomaths., n 57. eds V.Capasso, E. Grosso and S.L. Paven-Fontana, pp 360-365, Springer-Verlag

G. VIENNOT (1983), *Une théorie combinatoire des polynômes orthogonaux généraux*, Lecture Notes, Université du Québec à Montréal, (reprinted in publication du LACIM, 1991, Montréal).

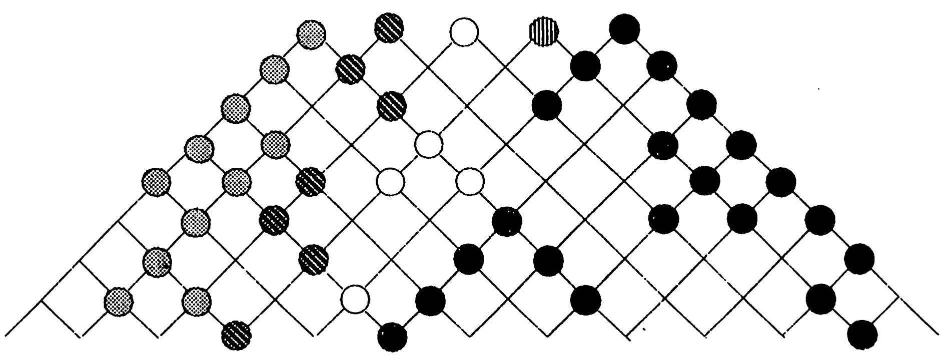
X.G.VIENNOT (1985a), *Problèmes combinatoires posés par la physique statistique*, Séminaire Bourbaki n° 626, 36ème année, in Astérisque n°121-122, (1985) 225-246, Soc. Math. France.

X.G. VIENNOT (1985b), *Enumerative combinatorics and algebraic langages*, Proceedings FCT'85, ed. L. Budach, Lecture Notes in Computer Science n°199, Springer-Verlag, Berlin, , p. 450-464.

X.G. VIENNOT (1986), *Heaps of pieces, I: Basic definitions and combinatorial lemmas*, in "Combinatoire énumérative", eds.G. Labelle, P.Leroux, Lecture Notes in Maths n°1234, p321-350, Springer,Berlin

X.G. VIENNOT (1987), *Bijections for the Roger-Ramanujan reciprocal,* J. Indian Math. Soc. 52, 171-183

E.M.WRIGHT (1968), *Stacks*, Quat. J. Math. Oxford, 2 19, 313-320.



3^n

Zeros of rank-generating functions of Cohen-Macaulay complexes

DAVID G. WAGNER *
Department of Combinatorics and Optimization
University of Waterloo
Waterloo, Ontario, Canada N2L 3G1

Abstract

Many combinatorial polynomials are related to rank-generating functions of Cohen-Macaulay complexes; notable among these are reliability, chromatic, flow, Birkhoff, and order polynomials. We prove two analytic theorems on the location of zeros of polynomials which have direct applications to the rank-generating functions of Cohen-Macaulay complexes and discuss their consequences for each of the aforementioned classes of polynomials.

0 Introduction

The rank-generating functions of Cohen-Macaulay complexes provide a unified setting for a variety of results and conjectures in the literature which concern the values of coefficients or the location of zeros of some combinatorial polynomials. Among these polynomials are the chromatic and flow polynomials of a graph (or matroid), the reliability polynomial of a (cographic) matroid, the order polynomial of a partially ordered set, and the Birkhoff (or characteristic) polynomial of a geometric lattice. Thus the study of these generating functions can be seen as a natural avenue of attack both on Rota's "critical problem" (cf. Chapter 16 of [6]) *via* the location of zeros of Birkhoff polynomials, and on the Read-Hoggar conjecture concerning logarithmic concavity of the coefficients of chromatic polynomials (cf. [11, 7]), as well as on other

*Partially supported by NSERC grant #0105392

more recent conjectures. Our purpose here is to present two new theorems, and to see to what extent these theorems relate to the conjectured behaviour of the polynomials in question.

1 Results

Let Δ be a (finite) simplicial complex with vertex-set V of size n; that is, Δ is a subset of 2^V which contains all singletons and is closed by taking subsets. We naturally identify the singletons with the elements of V. Elements of Δ are called *faces*; maximal faces are called *facets*. If all facets of Δ have the same size then Δ is called *pure*.

The primary numerical invariant of a simplicial complex is its f-vector, or equivalently, its rank-generating function. Let Δ have $f_i(\Delta)$ faces of size i, for each natural number i. The *rank-generating function* of Δ is

$$\mathrm{rgf}(\Delta; x) := \sum_{i \in \mathbb{N}} f_i(\Delta) x^i.$$

Since Δ is finite, $\mathrm{rgf}(\Delta; x)$ is a polynomial in x, and has constant term 1 corresponding to $\emptyset$. The sequence of coefficients $\{f_i(\Delta)\}$ is the *f-vector* of Δ.

A less obvious but even more important numerical invariant is the h-vector of a simplicial complex. Let d be the maximum size of a facet of Δ, so that $\mathrm{rgf}(\Delta; x)$ has degree d. The polynomials $x^i(x+1)^{d-i}$ for $i = 0, ..., d$ form a free basis for the additive group of polynomials of degree at most d in $\mathbb{Z}[x]$. Thus we may expand $\mathrm{rgf}(\Delta; x)$ uniquely in terms of this basis:

$$\mathrm{rgf}(\Delta; x) = \sum_{i=0}^{d} h_i(\Delta) x^i (x+1)^{d-i}. \tag{1}$$

This sequence of coefficients $\{h_i(\Delta)\}$ is the *h-vector* of Δ. For us it will be useful to define the *h-generating function* of Δ to be

$$\mathrm{hgf}(\Delta; x) := \sum_{i=0}^{d} h_i(\Delta) x^i.$$

Clearly $\mathrm{rgf}(\Delta; x)$ and $\mathrm{hgf}(\Delta; x)$ are related as follows:

$$\mathrm{rgf}(\Delta; x) = (x+1)^d \mathrm{hgf}(\Delta; x/(x+1))$$

$$\mathrm{hgf}(\Delta; x) = (1-x)^d \mathrm{rgf}(\Delta; x/(1-x)).$$

Many of the simplicial complexes which occur in nature share a structural condition called shellability. For a face F of a simplicial complex Δ, let $\bar{F}$ be the set of all subsets of F. A simplicial complex Δ is *shellable* when it is pure (of facet-size d, say), and the set of its facets may be ordered $F_1, F_2, \ldots, F_m$ in such a way that for each $j = 2, ..., m$, every facet of the subcomplex $\bar{F}_j \cap (\bar{F}_1 \cup ... \cup \bar{F}_{j-1})$ has size $d-1$. Given such a *shelling order* for the facets of Δ, for each $j = 1, ..., m$ let $\nu(j)$ denote the number of faces of size $d-1$ of $\bar{F}_j \cap (\bar{F}_1 \cup ... \cup \bar{F}_{j-1})$; hence $\nu(1) = 0$ and $\nu(j) > 0$ for all $j = 2, ..., m$. Notice that when $\bar{F}_j$ is adjoined to $\bar{F}_1 \cup ... \cup \bar{F}_{j-1}$ it contributes $x^{\nu(j)}(x+1)^{d-\nu(j)}$ to the rank-generating function of Δ. Consequently, we have the following proposition (a very special case of Theorem 6 of [14]).

PROPOSITION 1.1 *Let Δ be a shellable simplicial complex with facet-size d, and let $F_1, ..., F_m$ be a shelling order for the facets of Δ. Then $h_i(\Delta) = \#\nu^{-1}(i)$ for all $i = 0, ..., d$. In particular, $h_i(\Delta) \geq 0$ for all $i = 0, ..., d$.*

In fact, the nonnegativity of the h-vector follows from a more general (ring-theoretic) condition on the simplicial complex, known as Cohen-Macaulayness, but Proposition 1.1 will suffice for the applications we have in mind. It is also worth noting that Stanley has given a complete characterization of the h-vectors of shellable and Cohen-Macaulay complexes (cf. Theorem 6 of [14]).

We are almost ready to state the main results of this paper. For a natural number j, let $x_{\langle j \rangle} = x(x-1)\cdots(x-j+1)$ denote the j-th *falling factorial polynomial*, and define a linear transformation $S : \mathbb{R}[x] \to \mathbb{R}[x]$ by $Sx_{\langle j \rangle} = x^j$ and linear extension. We call S the *Stirling tansformation* since $x_{\langle j \rangle}$ is the generating function for the Stirling numbers of the first kind and Sx^j is the generating function for the Stirling numbers of the second kind.

THEOREM 1.2 *Let $p \in \mathbb{R}[x]$ be any polynomial, say $p(x) = \sum_{i=0}^{d} c_i x^i (x+1)^{d-i}$. If $c_i \geq 0$ for all $i = 0, ..., d$ then $Sp(x)$ has only real nonpositive zeros.*

THEOREM 1.3 *Let $p \in \mathbb{R}[x]$ be a polynomial such that $p(0) = 0$, say $p(x) = x \sum_{i=0}^{d} c_i x^i (x-1)^{d-i}$. If $c_i \geq 0$ for all $i = 0, ..., d$ then $Sp(x)$ has only real nonpositive zeros.*

Proofs of these theorems are deferred until Section 2. In view of Proposition 1.1 and formula (1) the following corollary is immediate.

COROLLARY 1.4 *Let Δ be a finite simplicial complex with nonnegative h-vector. Then $S\mathrm{rgf}(\Delta;x)$ and $Sx\mathrm{rgf}(\Delta;-x)$ have only real nonpositive zeros.*

Before turning to applications of these results, let's consider some of their ramifications in general. Let $\varepsilon : \mathbb{R}[x] \to \mathbb{R}[x]$ be the $\mathbb{R}$-algebra automorphism defined by $\varepsilon x = -x$ and linear and multiplicative extension: for any $p \in \mathbb{R}[x]$ we have $\varepsilon p(x) = p(-x)$. In particular, $\varepsilon x_{(j)} = (-1)^j x^{\langle j \rangle}$, where $x^{\langle j \rangle} = x(x+1)\cdots(x+j-1)$ is the j-th *rising factorial polynomial.* Conjugating S by ε we obtain the linear transformation $T = \varepsilon S \varepsilon$ defined by $Tx^{\langle j \rangle} = x^j$ and linear extension. In Theorems 1.2, 1.3, and Corollary 1.4 we may conjugate by ε to obtain the following statements.

COROLLARY 1.5 (a) *Let $p \in \mathbb{R}[x]$ be any polynomial, say $p(x) = \sum_{i=0}^d c_i x^i (x-1)^{d-i}$. If $c_i \geq 0$ for all $i = 0, ..., d$ then $Tp(x)$ has only real nonnegative zeros.*
(b) *Let $p \in \mathbb{R}[x]$ be a polynomial such that $p(0) = 0$, say $p(x) = x\sum_{i=0}^d c_i x^i (x+1)^{d-i}$. If $c_i \geq 0$ for all $i = 0, ..., d$ then $Tp(x)$ has only real nonnegative zeros.*
(c) *Let Δ be a finite simplicial complex with nonnegative h-vector. Then $Tx\mathrm{rgf}(\Delta;x)$ and $T\mathrm{rgf}(\Delta;-x)$ have only real nonnegative zeros.*

The condition that a polynomial has only real nonpositive zeros places strong restrictions on the values of its coefficients; see Chapter 8 of [8]. For example, we have the following special case of a theorem of Schoenberg (Theorem 7.1 in Chapter 8 of [8]). Another extension of this result can be found in Theorem 1.3 of [4].

PROPOSITION 1.6 *Let $p(x) = \sum_{i=0}^d c_i x^i$ be such that $c_d > 0$ and $c_0 \neq 0$ and if $p(z) = 0$ then $|\pi - \arg(z)| < \pi/3$. Then $c_i > 0$ for all $i = 0, ..., d$, and $c_i^2 > c_{i-1}c_{i+1}$ for all $i = 1, ..., d-1$.*

One difficulty arises in applying Proposition 1.6 in conjunction with Corollaries 1.4 and 1.5: we are primarily interested in the zeros and coefficients of $\mathrm{rgf}(\Delta;x)$, but our conclusions concern $S\mathrm{rgf}(\Delta;x)$ and $Tx\mathrm{rgf}(\Delta;x)$. Accordingly, let's consider more closely the condition that $Sp(x)$ has only real nonpositive zeros. Proposition 1.7(a) is implicit in Section 4 of [15]; parts (b) and (c) comprise Theorem 4.7 of [4].

PROPOSITION 1.7 *Let $p(x), q(x) \in \mathbb{R}[x]$.*
(a) *Suppose that $Sp(x)$ and $Sq(x)$ both have only real nonpositive zeros. Then $S[p(x)q(x)]$ also has only real nonpositive zeros.*

(b) *Let* $p(x)$ *be a polynomial such that* $x_{(m)}$ *divides* $p(x)$. *If every real zero of* $p(x)$ *is at most* m *and every complex zero of* $p(x)$ *is in the parabolic region*

$$\mathcal{R}_S(m) := \{s + it : 4t^2 \leq 1 + 4m - 4s\}$$

then $Sp(x)$ *has only real nonpositive zeros.*
(c) *A quadratic polynomial* $p(x)$ *without real zeros is such that* $Sp(x)$ *has only real nonpositive zeros if and only if the zeros of* $p(x)$ *lie in* $\mathcal{R}_S(0)$.

Conjugating by ε, we also see that the parabolic regions

$$\mathcal{R}_T(m) := \{s + it : 4t^2 \leq 1 + 4m + 4s\}$$

play the same rôle for the transformation T as the $\mathcal{R}_S(m)$ do for S. In an as yet imprecise way, Proposition 1.7 suggests that if all the zeros of $Sp(x)$ are real and nonpositive then "most" of the zeros of $p(x)$ lie inside the parabolic region $\mathcal{R}_S(m)$, where m is the multiplicity of 0 as a zero of $Sp(x)$. The accompanying figures corroborate this impression. In Figure 1 we plot all zeros other than -1 of the 941 rank-generating functions of Cohen-Macaulay complexes with $d \leq 4$ and $h_1 = 4$, as well as the boundaries of $\mathcal{R}_S(0)$ and $\mathcal{R}_T(1)$. In Figure 2 we plot all zeros other than -1 of the 813 C-M RGFs with $d \leq 5$ and $h_1 = 3$. Figure 3 is a similar plot for the 520 C-M RGFs with $d \leq 8$ and $h_1 = 2$. Of course, a quantitative converse to Proposition 1.7(b) is very much to be desired.

One simple bound on the location of zeros of $\mathrm{rgf}(\Delta; x)$ follows immediately from nonnegativity of the h-vector (cf. Theorem (1,1) of [9]), but in view of our computations it seems that much stronger results are possible.

PROPOSITION 1.8 *Let* $p(x) = \sum_{i=0}^{d} c_i x^i$, *where* $c_i \geq 0$ *for all* $i = 0, \ldots, d$ *and* $c_0 > 0$. *It follows that if* $p(z) = 0$ *then* $|\arg(z)| \geq \pi/d$. *Consequently, if* Δ *is a simplicial complex with nonnegative h-vector and maximum facet-size d then the following hold:*
(a) *all zeros* z *of* $\mathrm{hgf}(\Delta; x)$ *satisfy* $|\arg(z)| \geq \pi/d$, *and*
(b) *all zeros* z *of* $\mathrm{rgf}(\Delta; x)$ *satisfy* $|\arg(z) - \arg(z+1)| \geq \pi/d$.

One can also apply other classical results on location of zeros (cf. Chapter VII of [9]), but the resulting bounds do not seem to adequately describe the observed location of zeros of C-M RGFs.

Finally, consider the following easy example. For $n \geq 1$ let $[n] := \{1, \ldots, n\}$ and let $\Gamma_n := 2^{[n]} \setminus \{[n]\}$. Then Γ_n is a shellable simplicial complex (the boundary

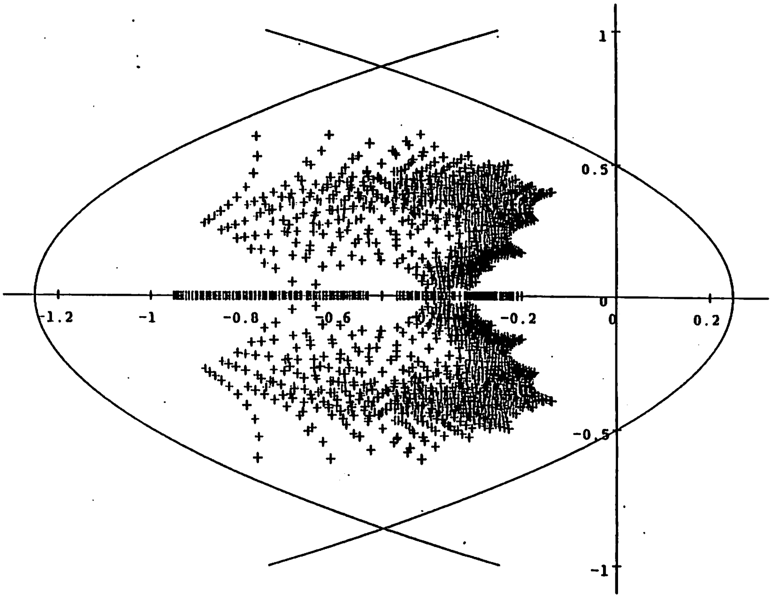


Figure 1: Zeros of C-M RGFs with $d \leq 4$ and $h_1 = 4$.

of a simplex) with $\text{rgf}(\Gamma_n; x) = (x+1)^n - x^n$ and $\text{hgf}(\Gamma_n; x) = 1 + x + \cdots + x^{n-1} = (x^n - 1)/(x-1)$. As $n \to \infty$ the limit distribution of zeros of $\text{hgf}(\Gamma_n; x)$ is uniformly concentrated on the circle $\{z \in \mathbb{C} : |z| = 1\}$. Applying the transformation $z \mapsto z/(1-z)$ takes the zeros of $\text{hgf}(\Gamma_n; x)$ to the zeros of $\text{rgf}(\Gamma_n; x)$; this amounts to inversion in the unit circle centered at 1, followed by reflection in the imaginary axis. It is a simple matter to check that all the zeros of $\text{rgf}(\Gamma_n; x)$ lie on the line $\text{Re}(z) = -1/2$, and that as $n \to \infty$, the proportion of zeros of $\text{rgf}(\Gamma_n; x)$ in the region $\mathcal{R}_S(0) \cap \mathcal{R}_T(1)$ tends to 2/3.

Chromatic and Birkhoff Polynomials

Let $\mathcal{M}$ be a loopless matroid on the set E, with no parallel elements. Then $\mathcal{M}$ is determined by its geometric lattice $\mathcal{L}$ of flats, or closed subsets of E. Let the minimal and maximal elements of $\mathcal{L}$ be denoted by $\hat{0}$ and $\hat{1}$, respectively, let the rank of $p \in \mathcal{L}$ be denoted $r(p)$, and let $d = r(\hat{1})$. The *Birkhoff (characteristic) polynomial* of $\mathcal{L}$ is

$$B(\mathcal{L}; x) := \sum_{p \in \mathcal{L}} \mu(\hat{0}, p) x^{d - r(p)},$$

where $\mu(\cdot, \cdot)$ is the Möbius function of $\mathcal{L}$. When $\mathcal{M}$ is the graphic matroid of a

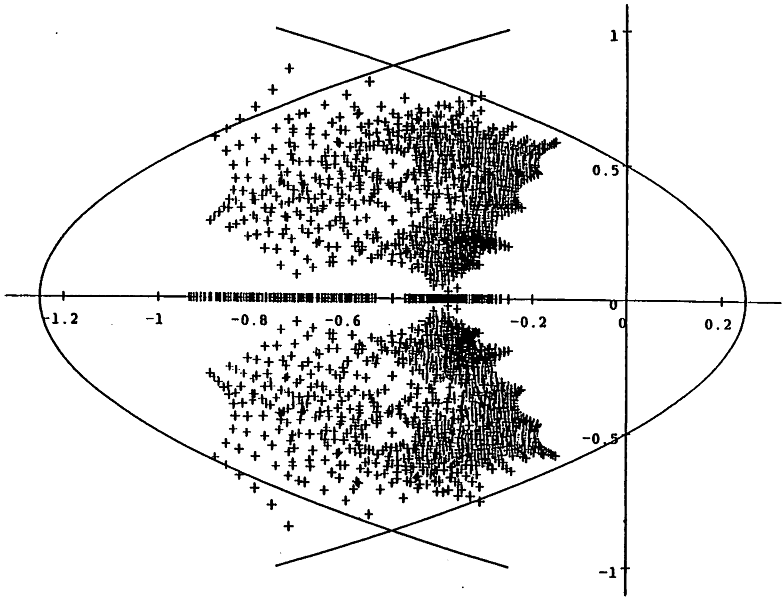


Figure 2: Zeros of C-M RGFs with $d \leq 5$ and $h_1 = 3$.

connected simple graph G, the lattice $\mathcal{L}$ is the lattice of contractions of G, and $P(G;x) := xB(\mathcal{L};x)$ is the *chromatic polynomial* of G (cf. Section 9 of [12]). A theorem of Whitney [18], generalized by Rota (cf. [12], p.359), asserts that for any geometric lattice $\mathcal{L}$ there is a shellable simplicial complex $\mathcal{B}$, called the *broken circuit complex* of $\mathcal{L}$, such that

$$B(\mathcal{L};x) = (-x)^d \mathrm{rgf}(\mathcal{B};-1/x).$$

Thus Corollaries 1.4 and 1.5(c) imply that the zeros of $S(-x)^d B(\mathcal{L};-1/x)$ are all real and nonpositive, and that the zeros of $T(-x)^{1+d}B(\mathcal{L};-1/x)$ are all real and nonnegative. This suggests that "most" of the zeros of $(-x)^d B(\mathcal{L};-1/x)$ lie in $\mathcal{R}_S(0) \cap \mathcal{R}_T(1)$. Replacing x by $-1/x$ amounts to inverting in the unit circle and reflecting in the imaginary axis. Then $\mathcal{R}_S(0)$ is transformed to the exterior of the cardiod $\mathcal{C} := \{re^{i\theta} \in \mathbf{C} : r \geq 2(1-\cos\theta)\}$ and $\mathcal{R}_T(1)$ is transformed to the unbounded region

$$\mathcal{C}' := \{s + it \in \mathbf{C} : 5(s^2+t^2)^2 - 4(s^3 + st^2 + t^2) \geq 0\}.$$

Hence we expect "most" zeros of Birkhoff polynomials of geometric lattices to lie in $\mathcal{C} \cap \mathcal{C}'$.

As seen in [4], there is a great deal of evidence for the following conjecture.

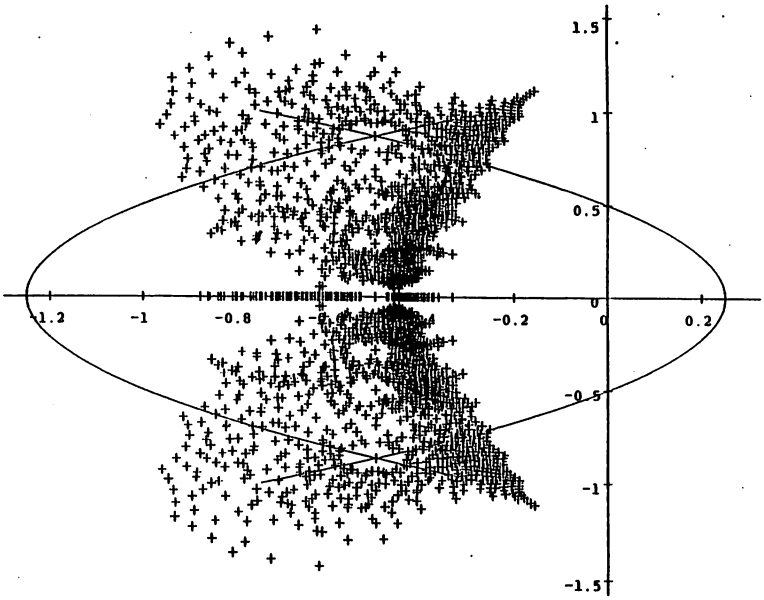


Figure 3: Zeros of C-M RGFs with $d \leq 8$ and $h_1 = 2$.

CONJECTURE 1.9 *Let G be a connected simple graph with chromatic polynomial $P(G;x)$. Then $TP(G;x)$ has only real nonnegative zeros.*

Unfortunately, Corollaries 1.4 and 1.5(c) seem to be working in exactly the wrong direction with regard to this conjecture, as our results show that if G has n vertices then $T(-x)^n P(G;-1/x)$ has only real nonnegative zeros. However, since there are graphs G for which $SP(G;x)$ has nonreal zeros (see [4]), perhaps we should not put much faith in Conjecture 1.9.

Reliability Polynomials

Let $\mathcal{M}$ be a loopless matroid on a set E of size m, and let $\mathcal{I}$ be the collection of independent sets of $\mathcal{M}$; then $\mathcal{I}$ is a shellable simplicial complex (cf. Proposition 4.2 of [2]), and so it is pure, of facet-size d, say. Suppose that each element of E fails independently with probability q; we are interested in the probablility $\mathrm{Rel}(\mathcal{M};q)$ that the set of failing elements is in $\mathcal{I}$. For example, let G be a finite connected loopless multigraph with edge-set E, and let $\mathcal{M}$ be the cographic matroid of G: a set of edges is in $\mathcal{I}$ if and only if its complement induces a connected spanning subgraph of G. Hence in this case $\mathrm{Rel}(\mathcal{M};q)$ is the probablility that G remains connected when the edges fail independently with probability q.

We may partition the event that the set of failing elements X is in $\mathcal{I}$ into its constituent subevents: that $X = F$ for a given face F of $\mathcal{I}$. This leads to the following expansion of the probability $\mathrm{Rel}(\mathcal{M}; q)$:

$$\mathrm{Rel}(\mathcal{M}; q) = \sum_{i \in \mathbb{N}} f_i(\mathcal{I}) q^i (1-q)^{m-i} = (1-q)^m \mathrm{rgf}\left(\mathcal{I}; \frac{q}{1-q}\right) = (1-q)^{m-d} \mathrm{hgf}(\mathcal{I}; q).$$

Thus the probability $\mathrm{Rel}(\mathcal{M}; q)$ is a polynomial function of q, called the *reliability polynomial* of $\mathcal{M}$.

Since $\mathrm{rgf}(\mathcal{I}; x) = (x+1)^m \mathrm{Rel}(\mathcal{M}; x/(x+1))$, Corollaries 1.4 and 1.5(c) imply that all the zeros of $S(x+1)^m \mathrm{Rel}(\mathcal{M}; x/(x+1))$ are real and nonpositive, and that all the zeros of $Tx(x+1)^m \mathrm{Rel}(\mathcal{M}; x/(x+1))$ are real and nonnegative. This suggests that "most" of the zeros of $(x+1)^m \mathrm{Rel}(\mathcal{M}; x/(x+1))$ are in $\mathcal{R}_S(0) \cap \mathcal{R}_T(1)$. Now $\theta \neq -1$ is a zero of $\mathrm{rgf}(\mathcal{I}; x)$ if and only if $\theta/(\theta+1)$ is a zero of $\mathrm{Rel}(\mathcal{M}; x)$. Mapping z to $z/(z+1)$ amounts to inverting in the unit circle centered at -1 and then reflecting in the imaginary axis. Thus $\mathcal{R}_S(0)$ is transformed to the unbounded region $1 - \mathcal{C}'$ and $\mathcal{R}_T(1)$ is transformed to the exterior of the cardiod $1 - \mathcal{C}$. Hence we expect "most" zeros of reliability polynomials other than $q = 1$ to lie in in region $1 - (\mathcal{C} \cap \mathcal{C}')$.

Brown and Colbourne [5] make the following conjecture:

CONJECTURE 1.10 *For any loopless cographic matroid $\mathcal{M}$ all the zeros of the polynomial* $\mathrm{Rel}(\mathcal{M}; q)$ *are in the disc* $\{z \in \mathbb{C} : |z| \leq 1\}$.

Note that this is equivalent to saying that all the zeros of $\mathrm{rgf}(\mathcal{I}; x)$ have real part greater than or equal to $-1/2$. Again, Corollaries 1.4 and 1.5(c) seem to have little to do with this conjecture. This is not surprising, as these results rely merely on the nonnegativity of the h-vector while a proof of the Brown-Colbourne conjecture must make use of the matroid structure since the conjecture fails to hold for all shellable complexes.

Order Polynomials

Let P be a nonempty partially ordered set, and for each positive integer m let $\Omega(P; m)$ denote the number of order-preserving functions $\phi : P \to \mathbf{m}$ from P into the m-element chain $\mathbf{m} = \{1 < 2 < \ldots < m\}$. Then $\Omega(P; m)$ is a polynomial function of m, called the *order polynomial* of P, and has the expansion

$$\Omega(P; x) = \sum_{j \geq 1} e_j(P) \binom{x}{j}$$

in which $e_j(P)$ is the number of order-preserving surjections from P onto a j-element chain (cf. Proposition 13.1 of [13]).

To see the connection with rank-generating functions, define a new partially ordered set $\mathcal{S}$ as follows. The elements of $\mathcal{S}$ are the order-preserving surjections from P onto a nonempty chain, and two surjections $\phi : P \to \mathbf{j}$ and $\psi : P \to \mathbf{k}$ are related by $\phi \leq \psi$ in $\mathcal{S}$ if and only if there is an order-preserving map $\sigma : \mathbf{k} \to \mathbf{j}$ such that $\phi = \sigma\psi$. One can check that this is the same thing as the order complex of the proper part of the finite distributive lattice of order ideals of P. As such, S. Provan showed that $\mathcal{S}$ is a shellable simplicial complex; see Theorem 3.7 and the remarks after Corollary 3.2 of [1]. It is clear that the rank-generating function of $\mathcal{S}$ is $\mathrm{rgf}(\mathcal{S}; x) = \sum_{j\geq 1} e_j(P)x^{j-1}$. Now a simple calculation with geometric series yields the identity

$$\sum_{m\geq 0} \Omega(P; m)t^m = \frac{t}{(1-t)^2}\mathrm{rgf}\left(\mathcal{S}; \frac{t}{1-t}\right) = \frac{t \cdot \mathrm{hgf}(\mathcal{S}; t)}{(1-t)^{2+d}}.$$

Neggers [10] made a conjecture equivalent to the following in 1978. In 1986 Stanley made an analogous conjecture for all labelled posets (private communication). See [3, 17] for recent work on these conjectures.

CONJECTURE 1.11 *For any nonempty poset P, all the zeros of* $\mathrm{rgf}(\mathcal{S}; x)$ *are in the interval* $[-1, 0]$.

If we assume that this conjecture holds then by Proposition 1.7(b) and its conjugate by ε we may conclude that $\mathcal{S}\mathrm{rgf}(\mathcal{S}; x)$ has only real nonpositive zeros, and that $\mathcal{T}x\mathrm{rgf}(\mathcal{S}; x)$ has only real nonnegative zeros. But this we know to be true, by Corollaries 1.4 and 1.5(c); hence these results are consistent with the validity of Conjecture 1.11.

2 Proofs

In order to prove Theorems 1.2 and 1.3 we require a few definitions and lemmas. A polynomial $p(x)$ is called *standard* if either $p(x) = 0$ identically or its leading coefficient is positive. For a subset I of the complex plane, $p(x)$ is called *I-rooted* if either $p(x) = 0$ identically or $p(z) = 0$ implies $z \in I$. Let $p(x)$ and $q(x)$ be two $\mathbb{R}$-rooted polynomials; let the zeros of $p(x)$ be $\xi_1 \leq \xi_2 \leq \cdots \leq \xi_r$ and let the zeros of $q(x)$ be $\theta_1 \leq \theta_2 \leq \cdots \leq \theta_s$. We say that $p(x)$ *interlaces* $q(x)$ when $s = 1 + r$ and

$$\theta_1 \leq \xi_1 \leq \theta_2 \leq \xi_2 \leq \cdots \leq \theta_r \leq \xi_r \leq \theta_{r+1}.$$

Also, we say that $p(x)$ *alternates left of* $q(x)$ when $s = r$ and

$$\xi_1 \leq \theta_1 \leq \xi_2 \leq \cdots \leq \theta_{r-1} \leq \xi_r \leq \theta_r.$$

We will use the notations $p \dagger q$ for p interlaces q, and $p \langle\!\langle q$ for p alternates left of q. Lemmas 2.1 to 2.4 follow easily from the Intermediate Value Theorem. The method of proof is explained in detail in Section 3 of [16]. We use D to denote the differentiation operator d/dx.

LEMMA 2.1 *Let p be a* $\mathbb{R}$*-rooted polynomial. Then $Dp \dagger p$.*

LEMMA 2.2 *Let p, q be nonzero standard* $\mathbb{R}$*-rooted polynomials such that $p \dagger q$. Then $p \dagger p+q$ and $p+q \langle\!\langle q$. Also, $p \dagger q-p$ and $q \langle\!\langle q-p$.*

LEMMA 2.3 *Let p, q be nonzero standard* $\mathbb{R}$*-rooted polynomials such that $p \langle\!\langle q$. Then $p \langle\!\langle p+q$ and $p+q \langle\!\langle q$.*

LEMMA 2.4 *Let $p_1, ..., p_m$ be nonzero* $\mathbb{R}$*-rooted polynomials such that $p_i \langle\!\langle p_{i+1}$ for $1 \leq i \leq m-1$, and $p_1 \langle\!\langle p_m$. Then $p_i \langle\!\langle p_j$ for all $1 \leq i \leq j \leq m$.*

Lemma 2.5 is a simple induction on $\deg p$ (cf. Proposition 5.1(f) of [4]).

LEMMA 2.5 *For any $p \in \mathbb{R}[x]$ and $\alpha \in \mathbb{R}$, $S(x-\alpha)p(x) = (x(1+D)-\alpha)Sp(x)$.*

To prove Theorem 1.2 we consider the polynomials $\phi_i^d := Sx^i(x+1)^{d-i}$, for $i = 0, ..., d$, and $d \geq 0$.

PROPOSITION 2.6 *Each ϕ_i^d is nonzero, standard, and $(-\infty, 0]$-rooted. Furthermore, $\phi_i^d \langle\!\langle \phi_j^d$ for all $0 \leq i \leq j \leq d$.*

PROOF: The proposition is true for $d = 0$ since $\phi_0^0 = 1$, and true for $d = 1$ since $\phi_0^1 = x+1$ and $\phi_1^1 = x$. By induction suppose that the proposition is true for $d-1$. Note that for any $i = 0, ..., d-1$ we have by Lemma 2.5 that

$$\phi_i^d = (x(1+D)+1)Sx^i(x+1)^{d-1-i} = (1+D)x\phi_i^{d-1} \tag{2}$$

and that

$$\phi_{i+1}^d = Sx^{i+1}(x+1)^d = x(1+D)\phi_i^{d-1}. \tag{3}$$

It follows that $\phi_0^d = [(1+D)x]^d 1$ and that $\phi_d^d = [x(1+D)]^d 1$ for all $d \geq 0$, and hence that $\phi_0^d = (1+D)\phi_d^d$. By induction ϕ_i^{d-1} is $(-\infty, 0]$-rooted, so it follows from (3) and Lemmas 2.1 and 2.2 that ϕ_{i+1}^d is also $(-\infty, 0]$-rooted, and that $\phi_i^{d-1} \dagger \phi_{i+1}^d$. Now by (2) we have $\phi_i^d = \phi_{i+1}^d + \phi_i^{d-1}$, and it follows from Lemmas 2.1 and 2.2 that ϕ_i^d is $(-\infty, 0]$-rooted and that $\phi_i^d \langle\!\langle \phi_{i+1}^d$. We have shown that $\phi_0^d \langle\!\langle \ldots \langle\!\langle \phi_d^d$ and that all these polynomials are $(-\infty, 0]$-rooted. Now, since $\phi_0^d = (1+D)\phi_d^d$, Lemmas 2.1 and 2.2 imply that $\phi_0^d \langle\!\langle \phi_d^d$. By Lemma 2.4 this suffices to finish the induction step and the proof. □

PROPOSITION 2.7 *Let $f_0, \ldots, f_d$ be any sequence of nonzero standard* R*-rooted polynomials such that $f_i \langle\!\langle f_j$ for all $0 \leq i \leq j \leq d$. Then for any nonnegative numbers $c_0, \ldots, c_d$ the polynomial $p = c_0 f_0 + \ldots + c_d f_d$ is* R*-rooted and $f_0 \langle\!\langle p \langle\!\langle f_d$.*

PROOF: By induction on d; the basis $d = 0$ is trivial, and the case $d = 1$ is Lemma 2.3. For the induction step let $f_i' = f_i$ for $i < d-1$ and let $f_{d-1}' = c_{d-1}f_{d-1} + c_d f_d$. By Lemma 2.3 we have $f_{d-1} \langle\!\langle f_{d-1}' \langle\!\langle f_d$, and hence by Lemma 2.4 we find that $f_0', \ldots, f_{d-1}'$ satisfy the inductive hypothesis. Putting $c_i' = c_i$ for $i < d-1$ and $c_{d-1}' = 1$ we find that $p = c_0' f_0' + \ldots c_{d-1}' f_{d-1}'$ is R-rooted and that $f_0' \langle\!\langle p \langle\!\langle f_{d-1}'$. But since $f_0' = f_0$, $f_{d-1}' \langle\!\langle f_d$, and $f_0 \langle\!\langle f_d$, we conclude from Lemma 2.4 that $f_0 \langle\!\langle p \langle\!\langle f_d$, which completes the induction step and the proof. □

PROOF OF THEOREM 1.2: We have $Sp(x) = \sum_{i=0}^d c_i \phi_i^d(x)$ and the result follows directly from Propositions 2.6 and 2.7.

To prove Theorem 1.3 we consider the polynomials $\psi_i^d := Sx^i(x-1)^{d-i}$ for $i = 1, \ldots, d$, and $d \geq 1$.

PROPOSITION 2.8 *Each ψ_i^d is nonzero, standard, and $(-\infty, 0]$-rooted. Furthermore, $\psi_j^d \langle\!\langle \psi_i^d$ for all $1 \leq i \leq j \leq d$.*

PROOF: The proposition is true for $d = 1$ since $\psi_1^1 = x$, and true for $d = 2$ since $\psi_1^2 = x^2$ and $\psi_2^2 = x(x+1)$. By induction suppose that the proposition is true for $d-1$. Note that for any $i = 1, \ldots, d-1$ Lemma 2.5 implies that

$$\psi_i^d = (x(1+D)-1)Sx^i(x-1)^{d-1-i} = (x(1+D)-1)\psi_i^{d-1} \tag{4}$$

and that

$$\psi_{i+1}^d = Sx^{i+1}(x+1)^d = x(1+D)\psi_i^{d-1}. \tag{5}$$

Thus we have the relation $\psi_{i+1}^d = \psi_i^d + \psi_i^{d-1}$ for all $i = 1, ..., d-1$. By the induction hypothesis, ψ_i^{d-1} is $(-\infty, 0]$-rooted. Hence, by (5) and Lemmas 2.1 and 2.2, ψ_{i+1}^d $(-\infty, 0]$-rooted and $\psi_i^{d-1} \dagger \psi_{i+1}^d$. It follows from (4) and Lemma 2.2 again that $\psi_i^d = \psi_{i+1}^d - \psi_i^{d-1}$ is $\mathbb{R}$-rooted and that $\psi_{i+1}^d \langle\!\langle \psi_i^d$. We have deduced that $\psi_d^d \langle\!\langle ... \langle\!\langle \psi_1^d$ are all nonzero, standard, and $\mathbb{R}$-rooted.

Now we claim that for any $k \in \mathbb{N}$, $Sx(x-1)^k = xSx^k$. For $k = 0$ this is trivial. By induction we calculate, using Lemma 2.5, that

$$\begin{aligned} Sx(x-1)^k &= (x(1+D)-1)Sx(x-1)^{k-1} \\ &= (x(1+D)-1)xSx^{k-1} \\ &= x^2(1+D)Sx^{k-1} = xSx^k, \end{aligned}$$

as desired. This identity implies that $\psi_1^d = x\psi_{d-1}^{d-1}$ for all $d \geq 2$. Since ψ_{d-1}^{d-1} is $(-\infty, 0]$-rooted by the inductive hypothesis, it follows that all ψ_i^d are $(-\infty, 0]$-rooted. Above, we found that $\psi_{d-1}^{d-1} \dagger \psi_d^d$; since $\psi_1^d = x\psi_{d-1}^{d-1}$ we can now conclude that $\psi_d^d \langle\!\langle \psi_1^d$. By Lemma 2.4 this suffices to finish the induction step and the proof. □

PROOF OF THEOREM 1.3 Theorem 1.3 now follows directly from Propositions 2.7 and 2.8.

References

[1] A. BJÖRNER, Shellable and Cohen-Macaulay partially ordered sets, *Trans. Amer. Math. Soc.*, **260** (1980), 159-183.

[2] A. BJÖRNER, Homology of matroids, in *Matroid Applications* (N. White, ed.), Cambridge Univ. Press, London/New York, to appear.

[3] F. BRENTI, *Unimodal, Log-Concave, and Pólya Frequency Sequences in Combinatorics*, Memoirs Amer. Math. Soc. **413**, Providence, RI, 1989.

[4] F. BRENTI, G.F. ROYLE, and D.G. WAGNER, Location of zeros of chromatic and related polynomials of graphs (preprint).

[5] J.I. BROWN and C.J. COLBOURNE, Roots of the reliability polynomial (preprint).

[6] H.H. CRAPO and G.-C. ROTA, *On the Foundations of Combinatorial Theory: Combinatorial Geometries*, M.I.T. Press, Cambridge, MA, 1970.

[7] S. HOGGAR, Chromatic polynomials and logarithmic concavity, *J. Comb. Theory Ser. B* **16** (1974), 248-254.

[8] S. KARLIN, *Total Positivity, vol.I*, Stanford U.P., Stanford, 1968.

[9] M. MARDEN, *Geometry of Polynomials*, Amer. Math. Soc., Providence, RI, 1966.

[10] J. NEGGERS, Representations of finite partially ordered sets, *J. Comb. Info. System Sci.* **3** (1978), 113-133.

[11] R.C. READ, An introduction to chromatic polynomials, *J. Comb. Theory* **4** (1968), 52-71.

[12] G.-C. ROTA, On the foundations of combinatorial theory I: Theory of Möbius functions, *Z. Wahrscheinlichkeitstheorie* **2** (1964), 340-368.

[13] R.P. STANLEY, *Ordered Structures and Partitions*, Memoirs Amer. Math. Soc. **119**, Providence, RI, 1972.

[14] R.P. STANLEY, Cohen-Macaulay complexes, in *Higher Combinatorics* (M. Aigner, ed.), Riedel, Dordrecht and Boston, 1977.

[15] D.G. WAGNER, The partition polynomial of a finite set system, *J. Comb. Theory Series A* **56** (1991), 138-159.

[16] D.G. WAGNER, Total positivity of Hadamard products, *J. Math. Anal. Appl.* (to appear).

[17] D.G. WAGNER, Enumeration of functions from posets to chains (preprint).

[18] H. WHITNEY, A logical expansion in mathematics, *Bull. Amer. Math. Soc* **38** (1932), 572-579.

Decompositions of Matroids and Exponential Structures

VOLKMAR WELKER [1]
UNIVERSITÄT ERLANGEN
IMMD1
MARTENSSTR.3
DW-8520 ERLANGEN
GERMANY
WELKER@INFORMATIK.UNI-ERLANGEN.DE

Abstract

We associate to a simple matroid (resp. a geometric lattice) a partially ordered set whose upper intervals are set partition lattices. Indeed for some important cases they are exponential structures in the sense of [Sta78]. Our construction includes the partition lattice, the poset of partitions whose size is divisible by a fixed number d, and the poset of direct sum decompositions of a vector space. If we start with a modularly complemented matroid the resulting poset is CL-shellable. This generalizes results of B. Sagan [Sag86] and M. Wachs and settles the open problem of the shellability of the poset of direct sum decompositions. Finally we give a formula for the Möbius number of the poset of direct sum decompositions of a vector space.

1 Matroids, Flats and Exponential Structures

The notation of exponential structures was introduced by Stanley [Sta78]. It provides a general setting for the treatment of a wide class of interesting posets. The poset of set partitions and the poset of partitions whose block sizes are divisible by a fixed number d [Sta78], [Sag86] are well studied examples. Another example, also mentioned by Stanley, is the poset of direct sum decomposition of vector spaces. This example led us to a more general approach to "decompositions" of matroids by which we retrieve some important classes of exponential structures.

For a finite matroid M of rank $r = rank(M)$ we denote by $\mathcal{F}(M)$ the geometric lattice of its flats. We write $\vee$ and $\wedge$ for the join and meet in $\mathcal{F}(M)$. We call a subset $\{F_1, \ldots, F_k\}$ of flats (of rank ≥ 1) of $\mathcal{F}(M)$ a decomposition of M if

$$\text{(A)} \quad \sum_{i=1}^{k} rank(F_i) = r \text{ and}$$

$$\text{(B)} \quad F_1 \vee \cdots \vee F_k = M.$$

[1] Partially supported by the Mittag-Leffler Institute

We order the set $\mathcal{D}(M)$ of decompositions of M by refinement. This means that for two decompositions $F = \{F_1,\ldots,F_k\}$ and $E = \{E_1,\ldots,E_l\}$ of M the inequality $F \leq E$ holds if and only if for all $1 \leq j \leq k$ there exists an $1 \leq i \leq l$ such that $F_j \leq E_i$. Since in general there does not exist a decomposition which refines all the others, we will always add a least element 0 to $\mathcal{D}(M)$ in order to make $\mathcal{D}(M)$ to a bounded poset. It is easy to see (Proposition 1.2) that $\mathcal{D}(M)$ carries for some important cases the properties of elements of exponential structures. Because we also want to cover the poset of set partitions where each block is divisible by a certain number d we define a type selected subposet of $\mathcal{D}(M)$. More precisely we write $\mathcal{D}(M)_d$ for the poset of all elements E of $\mathcal{D}(M)$ such that either $E = 0$ or each block of E is a flat in $\mathcal{F}(M)$ whose rank is divisible by d. For $d = 1$ we simply retrieve the original poset $\mathcal{D}(M)$. Obviously it makes no sense to look for flats of rank greater then r in a matroid of rank r. Therefore we restrict d to the interval $[r] := \{1,\ldots,r\}$. For a $d \in [r]$ which does not divide r the poset $\mathcal{D}(M)$ consists only of the single element 0. Hence we can further restrict our considerations to the case when d divides r.

It turns out that quite a lot of exponential structures investigated so far arise by this construction from a series of matroids. In particular for the projective geometry $\mathcal{P}^r\mathsf{F}_q$ of dimension r over the field with q elements the poset $\mathcal{D}(\mathcal{P}^r\mathsf{F}_q)$ is the poset of direct sum decompositions of the vector space F_p^{r+1}. The shellability of $\mathcal{D}(\mathcal{P}^r\mathsf{F}_q)$ has been left as an open problem in the paper of Sagan [Sag86]. More generally we prove that $\mathcal{D}(M)_d$ is shellable if $\mathcal{F}(M)$ is a modularly complemented geometric lattice. A geometric lattice (resp. a matroid) is called modularly complemented if there is a base of the matroid such that each subset of the base generates a modular element. Modularly complemented geometric lattices have been introduced by Stonesifer [Sto80]. Their classification by Kahn and Kung [KK86] shows that they are actually not too far away from being a modular geometric lattice. But they enlarge the class by some interesting examples such as the Dowling lattices [Dow73].

For a poset P and two elements $a, b \in P$ we denote by $[a,b] = \{c \mid c \in P. a \leq c \leq b\}$ the interval of all elements between a and b. We call a poset P exponential of rank n if

Exp_1) P is graded of rank n and P has a greatest element 1.

Exp_2) For every atom a of P the interval $[a,1]$ is isomorphic to the lattice Π_n of set partitions.

Exp_3) If a is an atom of P and if $b \in P$ is an element of the interval $[a,1] \cong \Pi_n$ then the type $(\lambda_1,\ldots,\lambda_k)$ of b in Π_n does not depend on the choice of a. Furthermore there are graded posets $Q_0,\ldots,Q_{r+1}$ such that Q_i has rank i and if $b \in P$ has type $(\lambda_1,\ldots,\lambda_k)$ then $\{c \in P \mid c \leq b\} \cong Q_{\lambda_1} \times \cdots \times Q_{\lambda_k}$.

According to Stanley we call a sequence $(P_1,\ldots,P_t,\ldots)$ of posets an exponential structure if each P_i is an exponential poset of rank $i+1$ and $Q_i \cong P_i$ for the posets in Exp_3 of the definition of an exponential poset.

Lemma 1.1 *Let M be a finite matroid of rank $r = d \cdot n$ for some divisor d of r. Then $\mathcal{D}(M)_d - \{0\}$ satisfies Exp_1 and Exp_2.*

Proof : From the fact that $r = n \cdot d$ and the fact that $\mathcal{F}(M)$ is a geometric lattice it follows that there exist flats $F_1,\ldots,F_n$ of rank d in $\mathcal{F}(M)$ such that $rank(F_1 \vee \cdots \vee F_n) = d \cdot n$. Moreover since the intervals of the geometric lattice $\mathcal{F}(M)$ are also geometric we can refine each nontrivial element $\{F_1,\ldots,F_k\}$ of $\mathcal{D}(M)_d$ to a decomposition of M into n flats of rank d. Hence all atoms of $\mathcal{D}(M)_d$ are of the form $\{F_1,\ldots,F_n\}$ where each F_i is a flat of rank d in $\mathcal{F}(M)$. Now we fix an atom $F = \{F_1,\ldots,F_n\}$ of $\mathcal{D}(M)_d$. Then the mapping which assigns to each partition $\pi = A_1 + \cdots + A_k$ in Π_n the element $\{\bigvee_{j \in A_1} F_j,\ldots,\bigvee_{j \in A_k} F_j\}$ establishes an isomorphism between Π_n and the interval $[F,1]$. This facts follows immediately from $F_1 \vee \cdots \vee F_n = M$ and $rank(\bigvee_{j \in A_i} F_j) = d \cdot |A_i|$. ∎

Proposition 1.2 *Let M be a finite matroid of rank $r = d \cdot n$ for some divisor d of r. Then $\mathcal{D}(M)_d - \{0\}$ is an exponential poset of rank n if and only if for any fixed i, $1 \leq i \leq n$ the posets $\mathcal{F}(N)_d$ are isomorphic for all $(d \cdot i)$-flats N of M.*

Proof : Let $b = \{B_1, \ldots, B_k\}$ be an element of $\mathcal{D}_d(M) - \{0\}$. Obviously any refinement $a \leq b$ corresponds bijectively to a k-tuple of elements of the decompositions posets $\mathcal{D}_d(B_i) - \{0\}$ of the matroids determined by the flats B_i. Hence the interval $[0, b]$ in $\mathcal{D}(M)_d$ is isomorphic to the direct product $(\mathcal{D}_d(B_1)_d - \{0\}) \times \cdots \times (\mathcal{D}_d(B_k) - \{0\})$.

If b contains only one flat then condition Exp_3 is trivially satisfied by b. If b contains more than one flat then by induction $\mathcal{D}(B_i)_d$ depends up to isomorphy only on the rank of B_i. Hence condition Exp_3 holds.

On the other hand assume that Exp_3 holds. Then assume that there is are flats B and B' of the same rank in $\mathcal{F}(M)$ for which $\mathcal{F}(B)_d$ and $\mathcal{F}(B')_d$ are not isomorphic. Since B and B' are different flats of the same rank their rank is not 0 and not maximal. Hence there exists a decomposition b (resp. b') in $\mathcal{D}(M)_d$ which consist of the block B (resp. B') and some blocks of rank d. By construction the interval $[0, b]$ (resp. $[0, b']$) is isomorphic to $\mathcal{D}(B)_d$ (resp. $\mathcal{D}(B')_d$). But this contradicts Exp_3 and we are done. ∎

Before we show, that if $\mathcal{F}(M)$ is modularly complemented then $\mathcal{D}(M)_d$ is shellable, we give some examples :

i) Let $M = U_{n,r}$ be the uniform matroid of rank r on an n-element set. For this matroid the poset $\mathcal{D}(M)$ is the disjoint union of $\binom{n}{r}$ partition lattices Π_r where the greatest elements are identified and an extra 0 is added. For an integer d dividing $r = k \cdot d$ the poset $\mathcal{D}(U_{r,r})_d$ is isomorphic to the poset $\Pi_r^{(d)}$ of partitions of r whose block sizes are divisible by d.

ii) If $M = \mathcal{P}^r\mathbb{F}_q$ is the projective geometry of dimension r over the field $\mathbb{F}_q$ then $\mathcal{D}(M)$ is isomorphic to the poset of direct sum decompositions of the vector space $\mathbb{F}_q^{r+1}$ ordered by refinement.

iii) If $M = \Pi_r$ is the partition lattice then every nontrivial element $b = \pi_1 / \cdots / \pi_k \in \mathcal{D}(\Pi_r)$ represents a family of partitions such that only $(1 \cdots r)$ is a partition coarser than all π_i. Additionally for all proper subsets of the set $\{\pi_1, \ldots, \pi_k\}$ there is a partition different from $(1 \cdots r)$ which is coarser than all π_i.

In contrast to examples i) and ii) $\mathcal{D}(\Pi_r)$ is not an exponential poset for $r \geq 4$.

We saw that for $r < n$ the proper part of $\mathcal{D}(U_{r,k})$ is a disconnected poset of rank $r - 2$. Hence $\mathcal{D}(U_{r,k})$ is not shellable for $r > k \geq 2$. But we will prove in the next section that $\mathcal{D}(M)_d$ is shellable for a suitably sized class of matroids.

2 Shellable decomposition posets

For this section we assume that M is a modularly complemented matroid of rank r. We assume further that $\mathrm{F} = \{F_1, \ldots, F_r\}$ is a base of M such that each subset of the base generates a modular element. Such a base exists by the definition of modularly complemented. For a subset $J \subseteq [r]$ we set $F^J := \bigvee_{j \in J} F_j$; for $J = \emptyset$ we simply denote by F^J the least element of $\mathcal{F}(M)$. By $\mathcal{A}(M)$ we denote the atoms of $\mathcal{F}(M)$. For our purposes we can restrict ourselves to simple matroids. Therefore the set $\mathcal{A}(M)$ is just the set on which the matroid M is defined.

Having this notation fixed we define a mapping v from the set $\mathcal{A}(M)$ into the set of vectors $\{0,1\}^r$. We set $v(E) := (v_1, \dots, v_r)$ where $v_j = 1$ if and only if $E \not\leq F^{[r]-\{j\}}$. We write $v(E)_l$ to denote the l-th entry v_l in $v(E)$. By the usual linear order $0 < 1$ on $\{0,1\}$ we can define the reverse lexicographic $\leq_l$ order on $\{0,1\}^r$. Via the map v the order $\leq_l$ defines a partial order $\mathcal{A}(M)$. For this order we write also $\leq_*$ and we set $E \leq_* G :\Leftrightarrow v(E) \leq_l v(G)$. Now we order the bases of the k-flats of $\mathcal{F}(M)$ by an arbitrary (but fixed) linear extension $\leq_{\mathcal{B}}$ of the lexicographic order on the k-tuples $(E_1 <_* \cdots <_* E_k)$. We omit an additional index k for the order $\leq_{\mathcal{B}}$.

Furthermore we will use an extension of $\leq_{\mathcal{B}}$ to compare decompositions which correspond to integer partitions $d + d + \cdots + d = r$. Here we associate to a decomposition $\{E_1, \dots, E_k\}$ the integer partition $rank(E_1) + \cdots + rank(E_k) = r$ of r. The necessity of ordering decompositions which correspond to partitions of integers in equal parts of size greater 1 arises since we will have to order the atoms of $\mathcal{D}(M)_d$ also in the case $d \neq 1$. We impose the order in the following way. We order the flats of the same rank simply by comparing the associated least bases by $\leq_{\mathcal{B}}$. In order to keep the notation for this relation consistent with order relation on the atoms (= 1-flats) we will denote the order relation also by $\leq_*$. So far we have transformed a decomposition into parts of rank d into an ordered decomposition. Now we can order decompositions into blocks of rank d by the lexicographic order induced by the order $\leq_{\mathcal{B}}$ on the blocks of the decomposition. For the atoms of $\mathcal{D}(M)$ we simply get the old order. Hence we can use $\leq_{\mathcal{B}}$ unambiguously for the atoms of $\mathcal{D}(\mathcal{F})_d$ for all d dividing the rank of $\mathcal{F}$.

Now the main aim of this section is to show that the order $\leq_{\mathcal{B}}$ induces a recursive atom order for $\mathcal{D}(M)_d$. To verify this we have to prove the following two properties [BW83] :

i) Let E $\leq_{\mathcal{B}}$ G be two atoms of $\mathcal{D}(M)_d$. Let I be another element of $\mathcal{D}(M)_d$ such that E $\leq$ I and G $\leq$ I holds in $\mathcal{D}(M)_d$. Then there is an atom of the interval [G, I] which covers an atom of $\mathcal{D}(M)_d$ preceding the atom G in the order $\leq_{\mathcal{B}}$.

ii) For every atom E of $\mathcal{D}(M)_d$ there is a recursive atom order of [E, 1] such that the atoms of [E, 1] which cover an atom of $\mathcal{D}(M)_d$ preceding E come first.

Our first two lemmas characterize the least base of a k-flat with respect to $\leq_{\mathcal{B}}$.

Lemma 2.1 *Let G be a k-flat in $\mathcal{F}(M)$. Let* E $= \{E_1 <_* \cdots <_* E_k\}$ *be the least base of G with respect to $\leq_{\mathcal{B}}$. For $1 \leq i \leq k$ let j_i be the greatest index for which $v(E_i)_{j_i}$ is nonzero. Then the following two assertions hold :*

i) For $1 \leq l < i \leq k$ the j_l-th entry in $v(E_i)$ is zero.

ii) The inequalities $j_1 < \cdots < j_k$ are strict.

Proof : Let $1 \leq l < i \leq k$ be integers. Now let J be the set

$$[j_i] - \Big(\{j_l\} \cup \{j \mid j < j_l \text{ and } v(E_i)_j = 0\}\Big).$$

The fact that $E_i <_* E_{i+1}$ certainly implies $j_i \leq j_{i+1}$. Hence J does not contain j_i. If the j_l-th entry in $v(E_i)$ is not zero then $E_i \not\leq F^J$. We also know by construction that $E_l \not\leq F^J$. Since $F^J \vee E_i = F^J \vee E_i \vee E_l = F^{J \cup \{j_l\}}$ we deduce from the modularity of F^J the fact that $rank(F^J \wedge (E_i \vee E_l)) = 1$. Hence there is an atom $E_0 \leq F^J$ which is also an atom in $E_{j_l} \vee E_{j_i}$. By construction $v(E_{j_i})_t = 0$ implies $v(E_0)_t = 0$ for $t > j_l$. On the other hand we infer from $E_0 \leq F^J$ that $v(E_0)_{j_l} = 0$. Hence the atom E_0 precedes E_{j_l} in the order $\leq_*$. But from the fact that $E_0 \vee E_i = E_0 \vee E_l$ we conclude that replacing E_{j_l} by E_0 in the base E we can construct another base. Since $E_0 <_* E_{j_l}$ the constructed base of G precedes the base E. But this contradicts the assumptions.

By the fact that $E_i \leq_* E_{i+1}$ implies $j_i \leq j_{i+1}$ (which we have already used in the first part of the proof), the second assertion is an immediate consequence of the first one. ∎

Lemma 2.2 *Let G be a k-flat in $\mathcal{F}(M)$. For a base $\mathrm{E} = \{E_1 <_* \cdots <_* E_k\}$ of F the following four conditions are equivalent :*

i) The base E *is the least base of G with respect to $\leq_{\mathcal{B}}$.*

ii) For all integers d, $2 \leq d \leq k$, and for all all d-element subsets J of $[k]$ the base $\{E_j \mid j \in J\}$ is the least base of E^J with respect to $\leq_{\mathcal{B}}$.

iii) There is an integer d, $2 \leq d \leq k$, such that for all d-element subsets J of $[k]$ the base $\{E_j \mid j \in J\}$ is the least base of E^J with respect to $\leq_{\mathcal{B}}$.

iv) For $i \in [k]$ let j_i be the greatest index such that $v(E_i)_{j_i}$ is nonzero. Then $E_i \leq F^J$ for $J = [j_i] - \{j_1, \ldots, j_{i-1}\}$. Furthermore E_i is the least atom in the flat $F^J \wedge G$.

Proof : First we prove the implication $i) \Rightarrow ii)$. Let J be a d-element subset of $[k]$. Then any base of E^J can be extended to a base of F by the elements E_j for $j \notin J$. Hence any base of E^J preceding the base $\{E_j \mid \in J\}$ would allow the construction of a base of G which precedes $\mathrm{E} = \{E_1, \ldots, E_k\}$. Since E is the least base of F we are done by assumption.

Obviously the implication $ii) \Rightarrow iii)$ is true.

Now consider the implication $iii) \Rightarrow iv)$. Let us take a d-element subset of $[k]$ which contains i. By assumption $\{E_j \mid j \in J\}$ is the least base of E^J. Now we apply Lemma 2.1 to $G = E^J$. Hence $v(E_i)_{j_l} = 0$ for all $l \in J \cap [i-1]$. Since $d \geq 2$ there exists for every pair of elements of $[k]$ a d element subset of $[k]$ which contains it. Now the preceding reasoning shows the first part of the claim. The second part of iv) follows immediately by the arguments used for the implication i) $\Rightarrow$ ii).

So it remains to prove the implication iv) $\Rightarrow$ i). Assume that the least base $\mathrm{I} = \{I_1 <_* \cdots <_* I_k\}$ of G is strictly smaller than the base E with respect to $\leq_{\mathcal{B}}$. Let i be the first index such that $E_i \neq I_i$. Let s be the greatest index for which $v(I_i)_s = 1$. Since $I_i <_* E_i$ the inequality $j_{i-1} < s \leq j_i$ follows from the choice of I_i. Hence $E^{[i]} = E^{[i-1]} \vee E_i$ and $I^{[i]} = I^{[i-1]} \vee I_i$ are contained in $G \wedge F^{[j_i]}$. First we treat the case $s = j_i$. The equality $s = j_i$ implies $I^{[i]} = E^{[i]}$. Since I is the least base of G the truncation $\{I_1, \ldots, I_i\}$ is the least base of $I^{[i]}$. For this conclusion we can argue in the same way as for the implication i) $\Rightarrow$ ii). Hence Lemma 2.1 shows that $v(I_i)_{j_l} = 0$ for $l \in [i-1]$. Since $rank(E_i \vee I_i) = 2$ we deduce from the modularity of $E^{[i-1]} = I^{[i-1]}$ that there is an atom I_0 in the meet $I^{[i-1]} \wedge (E_i \vee I_i)$. From the fact $v(E_i)_{j_l} = v(I_i)_{j_l} = 0$ for $l \in [i-1]$ we infer that I_0 is contained in $F^{[t]-\{j_1,\ldots,j_i\}}$. But then I_0 cannot be contained in $E^{[i-1]} = I^{[i-1]}$. Hence $\{I_1, \ldots, I_i, I_0\}$ is another base of G. By construction $v(I_0)_{j_i} = 0$ and therefore $I_0 <_* I_i$. But this contradicts the fact that $\{I_0, \ldots, I_i\}$ is the least base of $I^{[i]}$. What is still left is the case $j_{i-1} < s < j_i$ where all inequalities are strict. Let l be the least index such that $I^{[i]} \leq E^{[l]}$. Hence by the modularity of E^J for $J = \{i, \ldots, l\}$ we have

$$l = rank(E^{[l]}) = rank(I^{[i]} \vee E^J) =$$

$$= rank(I^{[i]}) + rank(E^J) - rank(I^{[i]} \wedge E^J) = i + (l - i + 1) - rank(I^{[i]} \wedge E^J).$$

This shows that $rank(I^{[i]} \wedge E^J) = 1$. But for every atom E_0 in $I^{[i]}$ the entry $v(E_0)_t$ is 0 for $t > s$ whereas $v(E_0)_t = 1$ for an atom E_0 in E^J for $j_i > t$. So we have deduced a contradiction from the assumption also in the case that $s < j_i$. Therefore E is the least base of G, which shows i). ∎

To deal with the poset $\mathcal{D}(M)_d$ uniformly we introduce the notion of a d-base. Let G be any flat in the lattice $\mathcal{F}(M)$ whose rank $l = d \cdot l_1$ is divisible by d. A set of l_1 flats of rank d is called a d-base of G if their join is G. For $d = 1$ we simply have reintroduced the notion of a base. As indicated at the beginning of the section it proves to be useful to identify each d-flat with its least 1-base. Since 1-bases are ordered by $\leq_{\mathcal{B}}$ their lexicographic order induces a linear order on the d-bases.

Theorem 2.3 *Let M be a finite modularly complemented matroid of rank $r = n \cdot d$. Then the order $<_{\mathcal{B}}$ is a recursive atom order for $\mathcal{D}(M)_d$. In particular the poset $\mathcal{D}(M)_d$ is CL-shellable.*

Proof : By Proposition 1.1 and by the definition of an exponential posets we know that $\mathcal{D}(M)_d$ is a graded poset and that each interval is isomorphic to Π_r.

It is well known that Π_r is a geometric lattice and hence upper semimodular. Since every order on the atoms of an upper semimodular lattice is a recursive atom order [BW83, Theorem 5.1] is suffices to show C_1 for the order $<_{\mathcal{B}}$ [Sag86, Lemma 3].

Let $\mathrm{E} = \{E_1 <_* \cdots <_* E_n\}$ and $\mathrm{G} = \{G_1 <_* \cdots <_* G_n\}$ be two atoms of $\mathcal{D}(M)_d$ for which $\mathrm{E} <_{\mathcal{B}} \mathrm{G}$ holds. Let $\mathrm{I} = \{I_1, \ldots, I_k\}$ be an element of $\mathcal{D}(M)_d$ such that $\mathrm{E} < \mathrm{I}$ and $\mathrm{G} < \mathrm{I}$ (here $<$ is the order relation in $\mathcal{D}(M)$). Since $\mathrm{E} <_{\mathcal{B}} \mathrm{G}$ there is a flat in I for which E contains a d-base strictly preceding its d-base contained in G. We may assume that I_1 is this flat. Hence G does not contain the least d-base for I_1. Now by Lemma 2.2 ii) there is a flat H of rank $2 * d$ contained in I_1 such that G does not contain the least d-base for I_1. Let G_1, G_2 be the d-base of I_1 contained in G. By construction the decomposition $\{H, G_3, \ldots, G_k\}$ is an atom of $[\mathrm{G}, 1]$ which is contained in I. Let H_1, H_2 be the least d-base of H; then $\{H_1, H_2, G_3, \ldots, G_k\}$ is a d-base of M and hence an atom in $\mathcal{D}(M)_d$ which is covered by $\{H, G_3, \ldots, G_k\}$. Again by construction the atom $\{H_1, H_2, G_3, \ldots, G_k\}$ precedes the atom G and hence we have verified C_2 for the order $\leq_{\mathcal{B}}$. ∎

Actually we do not know whether the decomposition poset of a modularly complemented matroid $\mathcal{D}(M)_d$ is EL-shellable or not. In a recent paper by M. Wachs [Wac] it is proved that this is true for the free matroid.

Finally we give an example of a not modularly complemented line configuration for which it is easy to verify that the decomposition poset $\mathcal{D}(M)$ of the associated matroid is shellable (even EL-shellable). Figure 1 depicts the line configuration of a geometric lattice which is not modularly complemented. Obviously the corresponding matroid has rank 3. Hence any base of M consist of 3 points. But there are only two modular hyperplanes (the 3-point lines) in the lattice. Therefore there must be two of the base elements which do not generate a modular line. This shows that the matroid is not modularly complemented.

On the other hand the associated decomposition poset is depicted on the right side of Figure 1. It is easily seen that the poset is bounded of rank 3 and moreover its proper part is connected. This implies that the decomposition poset is shellable. But we would like to remark that the described matroid is supersolvable. Indeed we know of no supersolvable matroid for which the decomposition poset is not shellable and we know of no non supersolvable matroid with shellable decomposition poset.

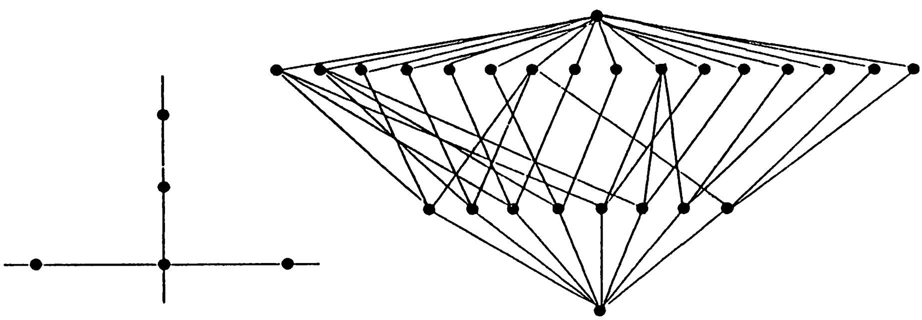


Figure 1

3 The Möbius numbers

For the free matroid M on $r = n \cdot d$ elements the Möbius number $\mu(\mathcal{D}(M))_d = \mu(\Pi_{nd}^{(d)})$ equals the number of permutations with descent set $\{d, 2d, \dots, nd\}$ [Sta78]. But in the case of projective geometries $\mathcal{P}^r\mathbb{F}_q$ it seems that the Möbius number of $\mathcal{D}(\mathcal{P}^r\mathbb{F}_q)$ has not been investigated so far. We will denote the Möbius number of $\mathcal{D}(\mathcal{P}^r\mathbb{F}_q)$ by $\mu_{r+1}(q)$. The shift in the dimension is for the sake of a more convenient formulation of the formulas. Moreover we will provide some general tools which can be useful for the determination of the Möbius numbers for arbitrary decomposition posets which arise from a modularly complemented matroid.

It is well known that the conditions given Proposition 1.2 are true for $\mathcal{P}^r\mathbb{F}_q$ and therefore $\mathcal{D}(\mathcal{P}^r\mathbb{F}_q)$ is an exponential structure. Hence we can apply the results of Stanley [Sta78] for an analysis of the exponential generating function of $\mu_{r+1}(q)$. It will turn out that a transformation of the generating series leads to a nice formula for the logarithm of the q-hypergeometric series $\sum_{r=0}^{\infty} \frac{q^{r^2}}{\prod_{i=0}^{r}(1-q^i)} z^i$. Additionally we would like to point out that for a prime number p congruent 1 modulo q the Möbius number of $\mathcal{D}(\mathcal{P}^{r-1}\mathbb{F}_q)$ is the same as the Möbius number of the poset $S_p(GL(r,q))$ of nontrivial p-subgroups of $GL(r,q)$. Since the order complex of $\mathcal{D}(\mathcal{P}^{r-1}\mathbb{F}_q)$ is even homotopy equivalent to $S_p(GL(r,q))$ [Qui78, Theorem 12.4] we obtain by $|\mu_r(q)|$ the dimension of the $(r-1)$-th homology group of the order complex of $S_p(GL(r,q))$ which is regarded as an analog of the Steinberg module [AS91]. Of course an analysis of the homotopy equivalence eventually shows that also the $GL(r,q)$-module structure is preserved. But the determination of this representation will not be done here. In his proof D. Quillen shows that $S_p(GL(r,q))$ is homotopy equivalent to a subposet which is Cohen-Macaulay. The subposet is given by the elementary abelian p-subgroups. But although by Theorem 2.3 the poset $\mathcal{D}(\mathcal{P}^{r-1}\mathbb{F}_q)$ is CL-shellable it remains open whether the poset of elementary abelian p-subgroups of this $GL(n,q)$ is itself shellable.

Lemma 3.1 *The Möbius number $\mu_r(q)$ is a polynomial of degree $r(r-1)$ in q. The leading coefficient is $\frac{(-1)^r}{r}$.*

Proof : The assertion is trivially fulfilled for $r = 1$. In this case we have $\mu_r(q) = -1$. Now assume $r > 1$. By definition we have

$$\mu_r(q) = - \sum_{\lambda=(\lambda_1 \geq \cdots \geq \lambda_k \geq 1) \in P_r - \{(r)\}} \frac{(q^r - 1)\cdots(q^r - q^{r-1})}{\prod_{i=1}^k((q^{\lambda_i} - 1)\cdots(q^{\lambda_i} - q^{\lambda_i - 1})) \cdot j_1^\lambda! \cdots j_r^\lambda!} \cdot (-1)^{k-1} \cdot \prod_{i=1}^k \mu_{\lambda_i}(q).$$

Here we denote by P_r the integer partitions of r and for an integer partition $\lambda \in P_r$ the number j_i^λ is the number of blocks of size i in λ. The fraction (first factor of each summand) counts the number of direct sum decompositions of F_q^r into blocks of sizes $\lambda_1, \ldots, \lambda_k$. The last two factors stem from the direct decomposition (Exp_3) of the lower intervals in $\mathcal{D}(M) - \{0\}$. The factor $(-1)^{k-1}$ comes up by the fact that the decomposition only exists if the least element is removed. First all factors are easily seen to be polynomials. After the cancellation of the all $(q-1)$'s the first factor has degree $2\binom{r}{2} - 2\sum_{i=1}^k \binom{\lambda_i}{2}$ and the second part has by induction degree $2 \cdot \sum_{i=1}^k \binom{\lambda_i}{2}$. Summing up the degree we obtain as the maximal total degree $r(r-1)$. It remains to show that the coefficient is correct. The first factor has leading coefficient $\frac{1}{j_1! \cdots j_r!}$. The second one has by induction leading coefficient $\frac{(-1)^k}{\prod_{i=1}^k \lambda_i}$. Hence the coefficient for $q^{\binom{r}{2}}$ is

$$\sum_{\lambda=(\lambda_1 \geq \cdots \geq \lambda_k \geq 1) \in P_r - \{(r)\}} \frac{(-1)^k}{j_1! \cdots j_r! \cdot \lambda_1 \cdots \lambda_k}.$$

Substituting in the power series expansion of $exp(\sum_{i=1}^\infty \frac{x_i}{i})$ the variable x_i by $-z^i$ we obtain

$$\sum_{(\lambda_1 \geq \cdots \geq \lambda_k \geq 1) \in P_r} \frac{(-1)^k}{j_1! \cdots j_r! \cdot \lambda_1 \cdots \lambda_k}$$

as the coefficient of z^n. Since

$$exp(\sum_{i=1}^\infty \frac{(-z)^i}{i}) = exp(log(1-z)) = 1 - z$$

the assertion follows immediately for $r > 1$. ∎

Now we will show that for the analysis of $\mu(\mathcal{D}(M))$ for an modularly complemented matroid it suffices to investigate the $(0-1)$-matrices $(v(E_i)_j)_{(i,j)}$ associated to bases $\mathrm{E} = \{E_1 <_* \cdots <_* E_r\}$ of M. Of course $v(E_i)$ denotes the $(0-1)$-vector introduced in the second section. In the sequel we mean by a descending chain in the decomposition poset $\mathcal{D}(M)$ of a modularly complemented matroid M a maximal chain which has descending labels [BW83] in an arbitrary but fixed CL-shelling induced by $\leq_{\mathcal{B}}$.

Before we can give a more detailed analysis of the Möbius number $\mu_r(q)$ we will derive some general properties of the recursive atom order presented in Theorem 2.3 and properties of CL-shellings induced by that atom order.

Lemma 3.2 *Let* $\mathrm{E} = \{E_1 <_* \cdots <_* E_r\}$ *be a base of a modularly complemented matroid* M*. Then the number of descending chains in* $\mathcal{D}(M)$ *passing through* E *depends only on* $(v(E_i)_j)_{(i,j)}$*. The atoms of the interval* $[\mathrm{E}, 1]$ *which cover an atom of* $\mathcal{D}(M)$ *preceding* E *in* $\leq_{\mathcal{B}}$ *correspond to the lines for which* E *does not contain the least base.*

Proof : By the definition of $\leq_{\mathcal{B}}$ it remains to prove that the number of descending chains passing through E does not depend on the choice of the linear extension of the lexicographic order on the bases induced by $\leq_*$. Assume that E $\leq_{\mathcal{B}}$ F are two bases for which $(v(E_i)_j)_{(i,j)} = (v(F_i)_j)_{(i,j)}$. Assume that there is an atom of $[E,1]$ which covers F but no other atom of $\mathcal{D}(M)$ preceding E in $\leq_{\mathcal{B}}$. Hence there is a 2-flat for which F contains not the least base with respect to $\leq_{\mathcal{B}}$. This statement actually includes the second assertion of the lemma. Hence one of the equivalent conditions of Lemma 2.2 is violated. But this conditions depend only on the entries of the matrix $(v(F_i)_j)_{(i,j)}$. Therefore by assumption E does also not contain the least base for this 2-flat. Replacing the two base elements of this 2-flat in E by its least base gives a base preceding E in the order $\leq_{\mathcal{B}}$ which is also covered by the crucial atom of $[F,1]$. Hence the set of atoms of $[E,1]$ meeting condition ii) on the recursive atom order is independent on the choice of the linear extension. Therefore the construction of a CL-labeling derived from the recursive atom order shows that the number of descending chains starting in E is also independent of the choice of the linear extension. ∎

In the next step we investigate the configurations in a matrix $(v(E_i)_j)_{(i,j)}$ which are conducive to descending chains for a certain induced CL-shelling. To do this we need some more insight into the structure of descending chains generated by different recursive atom orders of the partition lattice. Here we would like to remind the reader of the trivial fact that an atom in the lattice Π_r of set partitions of the set $[r]$ consists of a 2-block and $r-2$ blocks of size 1. Here we say that a set of atoms E of a geometric lattice generates the chain $\pi_1 < \cdots < \pi_t$ if all π_i are joins of atoms from E. At first we give a lemma which applies even to all geometric lattices.

Lemma 3.3 *Let $\{E_1,\ldots,E_k\}$ be a set of atoms of a geometric lattice L. Let $\{E_1,\ldots,E_k\}$ be an initial segment of a recursive atom order of L. Then the number of descending chains generated by $\{E_1,\ldots,E_k\}$ is independent of the linear order within the set.*

Proof : Let L' be the set of all elements of L which are joins of elements of $\{E_1,\ldots,E_k\}$. Let E be the greatest element of L'. If E is not the greatest element of L then no maximal chain in L is generated by $\{E_1,\ldots,E_k\}$. Obviously this statement is independent of the linear order on $\{E_1,\ldots,E_k\}$.

Now assume that E is the maximal element of L. Therefore L' is another geometric lattice. Since all atom orders of a geometric lattice are recursive atom orders the linear order on $\{E_1,\ldots,E_k\}$ is a recursive atom order for L'. The number of descending chains counts the absolute value of the Möbius number [Sta86]. Hence the number of descending chains generated by $\{E_1,\ldots,E_k\}$ in L' is independent of the actual linear order. Now it remains to show that the descending chains in L' generated by $\{E_1,\ldots,E_k\}$ correspond to the descending chains generated by $\{E_1,\ldots,E_k\}$ in L. But the labeling induced by a recursive atom order on an edge in L which actually lies in L' does only depend on elements of L'. This follows from that fact that the labeling is determined by the position of the top element of an edge relative to the elements prior to the bottom element in the recursive atom order. ∎

The next step is to figure out how we can describe the descending chains generated by a subset $\{E_1,\ldots,E_k\}$ of the set of atoms of the partition lattice Π_r. Here we would like to remind the reader of the fact that the modular elements of the partition lattice Π_r are just the partitions with at most one nontrivial block. An element a of a geometric lattice L is called modular if for all $b \in L$ the equation $rank(a \vee b) + rank(a \wedge b) = rank(a) + rank(b)$ holds.

Lemma 3.4 *Let $\{E_1,E_2,\ldots,E_k\}$ be a set of atoms of Π_r. Let $i \in [r]$ be a number such that there exists a modular coatom of Π_r which is generated by $\{E_1,\ldots,E_k\}$ and for which i is not contained in the nontrivial block of this coatom. Then there exists a CL-shelling of Π_r such that a descending chain generated by $E_1,\ldots,E_k$ satisfies :*

i) Each element of the chain is a modular element of Π_r.

ii) The nontrivial block of every member of the chain contains i.

Proof : If $E_1 \vee \cdots \vee E_k \neq (1 \cdots r)$ then no maximal and in particular no descending chain can be generated by $E_1, \ldots, E_k$. Therefore we may assume that $E_1 \vee \cdots \vee E_k = (1 \cdots r)$. But then there exists at least one maximal chain of modular elements generated by $\{E_1, \ldots, E_k\}$. After a suitable renumbering we may assume that $(1) \cdots (r) < (12)(3) \cdots (r) < \cdots < (1 \cdots r-1)(r) < (1 \cdots r)$ is a chain of modular elements generated by $\{E_1, \ldots, E_k\}$. It is well known that a maximal chain of modular elements in a graded lattice induces an EL-shelling. Indeed in this case the lattice is supersolvable. Since an EL-shelling is also a CL-shelling there is a recursive atom order corresponding to that CL-shelling. Now we order the atoms $\{E_1, \ldots, E_k\}$ according to this recursive atom order. We put the atoms $\{E_1, \ldots, E_k\}$ in the described order as the initial segment of another recursive atom order of Π_r. Now a recursive argument will prove that every descending chain in an induced CL-shelling is a chain of modular elements whose nontrivial block contains r. Hence the descending chains generated by $\{E_1, \ldots, E_k\}$ in the chosen atom order correspond to chains of modular elements generated by $\{E_1, \ldots, E_k\}$ whose nontrivial block contains r. Hence the assertion is fulfilled for $i = r$.
∎

Now we can give a preliminary description of the descending chains in a specific CL-shelling induced by $\leq_{\mathcal{B}}$. We will give a more detailed description in Proposition 3.7.

Lemma 3.5 *Let M be a modularly complemented matroid. Then there is a CL-shelling of the poset $\mathcal{D}(M)$ such that a descending chain $0 < G_1 < \cdots < G_r$ satisfies :*

Let $\{F_1, \ldots, F_k\}$ be the atoms of $[G_1, 1]$ which cover an atom of $\mathcal{D}(M)$ preceding $G_1 = \{E_1, \ldots, E_r\}$ in the order $\leq \mathcal{B}$. Let i be the greatest index such that there is a modular coatom of $[G_1, 1]$ generated by a subset of $\{F_1, \ldots, F_k\}$ for which E_i is not contained in its nontrivial block.

i) Then $G_1 < \cdots < G_r$ is a chain of modular elements of the interval $[G_1, 1]$ which is generated by $\{F_1, \ldots, F_k\}$.

ii) The element E_i is contained in the nontrivial block of every decomposition G_i.

Proof : In a CL-shelling induced by $\leq_{\mathcal{B}}$ the chain $0 < G_1 < G_2$ is descending if and only if G_2 covers an atom of $\mathcal{D}(M)$ preceding G_1 in $\leq_{\mathcal{B}}$. Now the assertion follows immediately from the fact that $[G_1, 1] \cong \Pi_r$ and Lemma 3.4. ∎

Finally we can classify the bases E of $\mathcal{D}(M)$ such that there are descending chains in the CL-shelling constructed in the preceding lemma which pass through E.

Lemma 3.6 *There is a CL-shelling of $\mathcal{D}(M)$ for a modularly complemented matroid M such that for a base $\mathrm{E} = \{E_1 <_* \cdots <_* E_r\}$ of M the following conditions are equivalent :*

i) There is a descending chain passing through E.

ii) For all $2 \leq i \leq r$ there is an index $t \neq i$ such that $\{E_t, E_i\}$ is not the least base for $E_i \vee E_t$.

iii) Let j_i be the greatest index such that the entry $v(E_i)_{j_i}$ in the vector $v(E_i)$ is nonzero. Then for all $2 \leq i \leq r$ there is an index $t \neq i$ such that the entry $v(E_i)_{j_t}$ is nonzero.

Proof : The equivalence *ii)* $\Leftrightarrow$ *iii)* follows immediately from Lemma 2.1. The equivalence of ii) and iii) can also be deduced along the lines of the proof of Lemma 2.1 ∎

Now we give a final necessary condition on a chains in to be descending. We would like to remark here that it is possible to give a condition which exactly characterizes the descending chains.

Proposition 3.7 *Let M be a modularly complemented matroid. Then there is a CL-shelling of $\mathcal{D}(M)$ induced by $\leq_{\mathcal{B}}$ such that a descending chain $0 < G_1 < \cdots < G_r$ satisfies the following conditions :*

i) The base $G_1 = \{E_1 <_* \cdots <_* E_r\}$ satisfies one of the conditions of Lemma 3.6.

ii) All decompositions G_i are modular elements of the interval $[G_1, 1] \cong \Pi_r$.

iii) Let j be the greatest index such that for each index $1 \leq i \leq r$ and $i \neq j$ there is an index $t \neq i, j$ for which that $\{E_i, E_j\}$ is not the least base of $E_i \vee E_j$. Then E_j is contained in the nontrivial block of every decomposition G_i.

Proof : One easily sees that CL-shelling referred to in Lemma 3.5 is the same as in Lemma 3.6. By the construction of j in condition iii) of the assertion there is a modular coatom in $[G_1, 1]$ such that E_j is not contained in its nontrivial block. Now the result follows immediately from Lemma 3.5 and Lemma 3.6. ∎

Having control over the descending chains allows us to identify the absolute value of the Möbius number of $\mathcal{D}(\mathcal{P}^{r-1}\mathsf{F}_q)$ with combinatorial objects.

Theorem 3.8 *The Möbius number $\mu_r(q)$ factors in a monic polynomial $f_r(q)$ of degree $\binom{r}{2}$ with positive integral coefficients and the polynomial $\frac{(-1)^r}{r}\prod_{i=1}^{r-1}(q^i - 1)$.*

Proof : (Sketch !) Since $\leq_{\mathcal{B}}$ is a recursive atom order the absolute value of the Möbius number is given by the number of descending chains in $\mathcal{D}(\mathcal{P}^{r-1}\mathsf{F}_q)$ in a CL-shelling induced $\leq_{\mathcal{B}}$ [Sta86].

In the sequel we will be concerned with a CL-shelling induced by $\leq_{\mathcal{B}}$ which fulfills the conditions of Proposition 3.7.

Now let $0 < G_1 < \cdots < G_r$ be a descending chain, Then $G_1 = \{E_1 <_* \cdots <_* E_r\}$ is a base of M. Assume j is the index defined in condition iii) of Proposition 3.7. Now the nontrivial blocks B_i of the decompositions G_i determine a sequence $j = t_1, \ldots, t_r$ for which $B_i = E_{j_1} \vee \cdots \vee E_{j_i}$. The crucial point is to show that by some projective transformations which preserve $E_j = E_{t_1}$ one generates $\prod_{i=1}^{r-1}(q^i - 1)$ other descending chains. Counting the multiplicity of the occurrence of each descending chain in this enumeration explains the factor $\frac{1}{r}$. Finally Lemma 3.1 proves the assertions on the degree of $f_r(q)$ and the fact that $f_r(q)$ is monic. ∎

In the following corollary we deduce a surprising relation between a series involving the polynomials $f_r(q)$ and the logarithm of one of the basic q-hypergeometric series. For the formulation of the corollary we abbreviate $\prod_{i=1}^{r}(1 - q^i)$ as usual by $(q, q)_r$.

Corollary 3.9 *The following equation holds*

$$\sum_{r=1}^{\infty} \frac{(-1)^r}{r} \cdot \frac{q^r}{1 - q^r} \cdot q^{\binom{r}{2}} f_r(1/q) \cdot z^r = -log\Big(\sum_{r=0}^{\infty} \frac{q^{r^2}}{(q, q)_r} z^r\Big).$$

Proof : By a general result on the Möbius functions of exponential structures [Sta78, (8)] it follows that

$$\sum_{r=1}^{\infty} \frac{\mu_r(q) \cdot z^r}{r! \cdot M(r)_q} = -log\Big(1 + \sum_{r=1}^{\infty} \frac{z^r}{r! \cdot M(r)_q}\Big).$$

Here $M(r)_q = \frac{\prod_{i=0}^{r-1}(q^r - q^i)}{r! \cdot (q-1)^r}$ is the number of atoms in $\mathcal{D}(\mathcal{P}^{r-1}\mathsf{F}_q)$. Replacing z by $u := (q-1)z$ we easily obtain

$$\sum_{r=1}^{\infty} \frac{q^{\binom{r}{2}} \cdot \mu_r(q) \cdot u^r}{\prod_{i=1}^{r}(q^i - 1)} = -log\Big(\sum_{r=0}^{\infty} \frac{q^{\binom{r}{2}} \cdot u^r}{\prod_{i=1}^{r}(q^i - 1)}\Big).$$

Now an application of Theorem 3.8 shows

$$\sum_{r=1}^{\infty} \frac{(-1)^r}{r} \cdot \frac{q^{\binom{r}{2}} \cdot f_r(q) \cdot u^r}{q^r - 1} = -log\Big(\sum_{r=0}^{\infty} \frac{q^{\binom{r}{2}} \cdot u^r}{\prod_{i=1}^{r}(q^i - 1)}\Big).$$

Replacing q by $1/q$ proves the desired equation after some easy calculations. ∎

We are indebted to D. Stanton for providing an easy transformation which proves that the exponential of $\sum_{r=1}^{\infty} \frac{1}{r} f_r(q)$ equals the Rogers-Ramanujan continued fraction [Sta90] and hence the exponential of the generating series of the q-Catalan numbers $C_n(q)$ [Sta90]. R. Stanley pointed out to us that by a theorem [Sta86, Proposition 4.7.11] about the generating series of very pure monoids this also allows us a more explicit combinatorial interpretation of the coefficients of $f_r(q)$. Vice versa going the other way round the described argumentation would give another proof of Theorem 3.8. But this would certainly not enable us to characterize the descending chains in $\mathcal{D}(\mathcal{P}^r\mathsf{F}_q)$.

4 Acknowledgment

Finally we would like to thank the Mittag-Leffler Institute for providing the extremely fruitful environment in which most of the work was done. We also would like to thank R. Stanley for pointing out an error in our original definition of exponential poset and we are grateful to G. Ziegler and V. Strehl for a lot of discussions and invaluable hints.

References

[AS91] M. Aschbacher and S. Smith. On Quillen's conjecture for the p-subgroups complex. *preprint*, 1991.

[BW83] A. Björner and M. Wachs. Lexicographic shellable posets. *Transactions of the AMS*, 277(4):323–342, 1983.

[Dow73] T.A. Dowling. A class of geometric lattices bases on finite groups. *Journal of Combinatorial Theory (B)*, 14:61–86, 1973.

[KK86] J. Kahn and P.S. Kung. A classification of modularly complemented geometric lattices. *European Journal of Combinatorics*, 7:243–248, 1986.

[Qui78] D. Quillen. Homotopy properties of the poset of nontrivial p-subgroups of a group. *Advances in Mathematics*, 28:101–128, 1978.

[Sag86] B. Sagan. Shellability of exponential structures. *Order*, 3:47–54, 1986.

[Sta78] R. Stanley. Exponential structures. *Studies in Applied Mathematics*, 59:73–82, 1978.

[Sta86] R. Stanley. *Enumerative Combinatorics I.* Wadsworth & Brooks, Monterey, 1986.

[Sta90] D. Stanton. Unimodality of Young's lattice. *Journal of Combinatorial Theory (A)*, 54:41–53, 1990.

[Sto80] J.R. Stonesifer. Modularly complemented geometric lattices. *Discrete Mathematics*, 32:85–88, 1980.

[Wac] M. Wachs. A basis for the homology of the d-divisible partition lattice. preprint.

Extracting Combinatorics from Discrete Applied Geometry.

Walter Whiteley†
Champlain Regional College,
Centre de recherches mathematiques, Université de Montréal,
and LACIM, Université de Québec à Montréal

For many discrete applied geometry problems, the analysis separates into layers, described by Felix Klein's classical hierarchy of geometries. This hierarchy focuses on the properties invariant under various groups (categories) of transformations. Combinatorics, and combinatorial topology, appear as several of these layers in the lattice of groups and invariants. Other important layers include projective geometry, affine geometry, and Euclidean geometry. How does the combinatorics appear within this applied geometry? How does the geometry illustate and expand the combinatorics? What common patterns of results/unsolved problems are emerging?

We present a common pattern for extracting those combinatorial layers which are implicit in recent work on

- the rigidity of frameworks (civil engineering);
- correct pictures of spatial polyhedral (computer vision);
- multivariate splines (approximation theory);
- rigidity of polytopial skeletons (h-vectors of polytopes).

In each of these examples, the intial form of the problem is:

(i) an abstract cell complex (e.g a graph, a simplicial polytope, an abstract polyhedral surface, etc.) is given, with a clear combinatorial topology;

(ii) a set of 'geometric realizations' is given for these cell complexes, in $\mathbb{R}^d$ (for example, the vertices of each i-dimensional face span an i-dimensional affine subspace);

(iii) The geometric problem is recorded by a matrix, whose kernal or co-kernal represent the objects of study: the matrix has zero and non-zero entries controlled by the cell complex, (a modification the 'oriented incidence' matrix of the cell complex), and the non-zero entries are polynomials in the coordinates of the realization;

As first clarifying steps, the problem is transformed as follows.

(iv) This basic matrix is re-interpreted as a boundary operator and extended to a chain complex of vector spaces, indexed by the combinatorial faces in each each dimension.

† This work was supported, in part, by grants from FCAR (Québec) and NSERC (Canada).

The top homology, or top two homologies (and the corresponding cohomologies) represent the original geometric properties. (This geometric chain complex modifies the usual chain complex for the reduced homology of the underlying abstract cell complex.)

(vi) The generic rank of the top homology is determined by the combinatorics of the original cell complex: specifically, the rank is determined by the Euler characteristic of the chain complex (which is combinatorial) and the lower homologies (which are also combinatorial for generic realizations). Ideally, we can show that the lower Betti numbers are zero for interesting classes of structures and realizations, so that the top homology is determined by the Euler characteristic of the complex.

(vii) The relevant combinatorics for these generic Betti numbers builds on the matroid for the underlying homology, by a process of matroid union, followed by algebraic specialization (sometimes equivalent to matroid truncation). The known combinatorial algorithms for these problems are built from homology algorithms (disjoint spanning trees, maximal acyclic subcomplexes etc.).

(viii) The residual geometry of the non-generic realizations is captured in the maximal non-zero minors of the generic matrices, representing the 'determinant of the complex', if the sequence of chains and boundary maps is exact.

This underlying pattern is illustrated in explicit form for three of these geometric studies. While the methods apply to singular geometric homology of CW complexes, for simplicity we present only the simplicial theory, beginning with the standard simplicial homology.

Acknowledgements. This talk is based, in part, on continuing joint work with Henry Crapo; with Peter Alfeld and Larry Schumaker; and with Tiong-Seng Tay and Neil White.

Homology of simplicial complexes

We recall the classical reduced homology of a simplicial cell complex, presented by chains and boundary operators (for example [Munkres, 1984]). An *abstract simplicial complex* is a collection Δ of finite sets (simplices), such that if A is in Δ, then all subsets of A are in Δ. (We shall assume that Δ is a finite collection.) The *d-simplices*, of cardinality $d+1$, are denoted $\Delta^{(d)}$ and the vertices (0-simplices) are $V = \Delta^{(0)}$.

We choose an arbitrary orientation $[v_o, \ldots, v_i \ldots, v_p]$ for each simplex σ. A *p-chain* is a map $\mathbf{c}$ from the oriented p-simplices to the reals, such that $\mathbf{c}(\sigma) = -\mathbf{c}(\sigma')$ if σ and

σ' are opposite orientations of the same simplex. The set of all p-chains is a real vector space $C_p(\Delta)$, and an arbitrary element is written: $\sum_{\sigma\in\Delta^p} c_\sigma\sigma$.

For an oriented p-simplex $\sigma = [v_o, v_1, \ldots, v_p]$ the *boundary* operator is a map from the p-chain σ to the $(p-1)$-chains:

$$\partial_p(\sigma) = \sum(-1)^i[v_o, \ldots, v_{i-1}, \hat{v}_i \ldots, v_p] = \sum_{\{\tau\in\Delta^{(p-1)}|\tau\subset\sigma\}} Sign\,(\tau,\sigma)\tau.$$

This map is extended linearly to all p-chains. We note that $\partial_{p-1}\partial_p(c) = 0$ for any p-chain. (We also include the empty set as $\Delta^{(-1)}$, generating $\mathbb{R}$ as the -1-chains.)

The kernal of ∂_p, $Z_p(\Delta)$, is called the *p-cycles*. The matrix for operator ∂_p, with Z_p as cokernal, is the *oriented incidence matrix*, $M(\Delta^p)$, recording the incidences of the oriented p-simplices and the oriented $p-1$-simplices, with appropriate entries of 0, 1 and -1. The image of ∂_{p+1}, $B_p(\Delta)$ is called the p-boundaries. The row space of the matrix $M(\Delta^p)$ is the $p-1$-boundaries, $B_{p-1}(\Delta)$. With $B_p(\Delta) \subset Z_p(\Delta)$, the vector space $\tilde{H}_p(\Delta) = Z_p(\Delta)/B_p(\Delta)$ is called the *p-th homology* of the complex.

Example 1. For a graph, a complex of dimension 1, this is the usual matrix representation of the graph (leaving the zero entries blank):

$M(\Delta^p)$	λ	a	b	c	d
ab	1	-1	1		
ac	0	-1		1	
ad	-1	-1			1
bc	1		-1	1	
bd	0		-1		1
cd	1			-1	1

The second column gives a row dependence i.e. the coefficients of a 1-cycle, or a polygon. The rank of the matrix is the size of a maximal subgraph with no cycles - a maximal forest.

Example 2. For a set of triangles, we have, for example:

$M(\Delta^p)$	λ	ab	ac	ad	bc	bd	cd
abc	1	1	-1		1		
abd	-1	1		-1		1	
acd	1		1	-1			1
bcd	-1				1	-1	1

.

Again, the row dependence is a 2-cycle, and the rows are the 1-boundaries.

If Δ has dimension d, the *augmented simplicial chain complex* $\tilde{\mathcal{H}}$ is the entire sequence of vector spaces and maps

$$\tilde{\mathcal{H}}:\ 0 \longrightarrow C_d(\Delta) \xrightarrow{\partial_d} C_{d-1}(\Delta) \xrightarrow{\partial_{d-1}} \dots \xrightarrow{\partial_1} C_0(\Delta) \xrightarrow{\partial_0} \mathbb{R} \longrightarrow 0.$$

There is Euler-Poincaré identity for any such sequence of vector spaces and maps, with $\partial_{p-1}\partial_p = 0$:

$$\chi(\Delta) = \sum_{i=-1}^{i=d} (-1)^i \dim(C_i(\Delta)) = \sum_{i=-1}^{i=d} (-1)^i \dim(\tilde{H}_i(\Delta)).$$

Since $\dim(C_i(\Delta)) = |\Delta^{(i)}|$, the first sum is the a combinatorially defined number called the *Euler characteristic* of the simplicial complex:

$$\chi(\Delta) = \sum_{i=-1}^{i=d} (-1)^i |\Delta^{(i)}|.$$

There are combinatorial criteria under which $\dim(\tilde{H}_i(\Delta)) = 0$ for $i \leq n-2$. (For example, if the complex is a topological wedge of spheres.) When this holds, we have the simple combinatorial identity:

$$\dim(\tilde{H}_n(\Delta)) - \dim(\tilde{H}_{n-1}(\Delta)) = \sum_{i=0}^{i=d} (-1)^{n+i} |\Delta^{(i)}|.$$

Example 3. Consider the standard graph $G = (V, E)$ of graph theory and matroid theory. This gives the chain complex:

$$\tilde{\mathcal{H}}:\ 0 \longrightarrow \mathbb{R}^{|E|} \xrightarrow{\partial_1} \mathbb{R}^{|V|} \xrightarrow{\partial_0} \mathbb{R} \longrightarrow 0.$$

The 1-cycles are generated by the polygons. This complex has Euler characteristic $|E| - |V| + 1$. If the graph is connected, then $\tilde{H}_0 = 0$, and $\tilde{H}_1 = |E| - (|V| - 1)$. The maximal acyclic subcomplex, with $\tilde{H}_1 = 0$, is a polygonal-free subgraph, or a forest, which satisfies: $|E| \leq |V| - 1$.

For graphs, there are simple, polynomial time algorithms for finding the rank of the 'homology matrix', i.e. the size of the maximal forest. This can also be expressed as a polynomial algorithm for the graphic matroid, or the matroid defined by the submodular function on the edges

$$f(E') = |V(E')| - 1$$

where $V(E)$ is the vertices of the subset E'.

We are not aware of similar polynomial time algorithms for the rank of the matrices $M(\Delta)$, $p \geq 2$. (For a general complex, this matroid is not defined by a submodular function.) This issue of algorithms for homology is important in the applications.

There is the corresponding cohomology of the simplicial complex. In fact, the kernal of the matrix $M(\Delta^p)$ is the cocyles, $Z^p(\Delta)$ for the corresponding cohomology:

$$\tilde{\mathcal{H}}:\ \mathbf{0} \longleftarrow C_d(\Delta) \stackrel{\delta_d}{\longleftarrow} C_{d-1}(\Delta) \stackrel{\delta_{d-1}}{\longleftarrow} \ldots C_0(\Delta) \stackrel{\delta_0}{\longleftarrow} \mathbb{R} \longleftarrow \mathbf{0}.$$

Rigidity of frameworks

The theory of static and infinitesimal rigidity of frameworks [Maxwell, 1864] has recently experienced a upsurge of research (Laman, 1970, Asimow & Roth, 1978, Whiteley, 1984 etc.). The underlying abstract structure of a framework is its graph $G = (V, E)$, with an arbitrary orientation of all edges. For frameworks in n-space, the *Euclidean realizations* are maps $\mathbf{p} : V \to \mathbb{R}^n$. For convenience $\mathbf{p}(v_i) = \mathbf{p}(i) = \vec{p}_i$. The realized graph $G(\mathbf{p})$ is called a *bar framework in n-space* provided $\vec{p}_i \neq \vec{p}_j$ for all $\{i, j\} \in E$.

This classical theory studies the kernal and cokernal of the *rigidity matrix* $R_1(\Delta)$. For example:

$R_1(\Delta)$	a	b	c	d
ab	$\vec{b}-\vec{a}$	$\vec{a}-\vec{b}$		
ac	$\vec{c}-\vec{a}$		$\vec{a}-\vec{c}$	
ad	$\vec{d}-\vec{a}$			$\vec{a}-\vec{d}$
bc		$\vec{c}-\vec{b}$	$\vec{b}-\vec{c}$	
bd		$\vec{d}-\vec{b}$		$\vec{b}-\vec{c}$
cd			$\vec{d}-\vec{c}$	$\vec{c}-\vec{d}$

The row dependencies of this matrix are the static *self-stresses* and the column dependencies are the *infinitesimal motions* of the framework. This matrix represents a basic boundary operator for a chain complex on the graph:

The 1-chains are $\mathcal{R}_1 = C_1 = \mathbb{R}^{|E|}$ as defined for homology.

The set of 0-chains is $\mathcal{R}_0 = \oplus_{v_i \in V}\mathbb{R}^n = \mathbb{R}^{n|V|}$.

The boundary map ∂_1 is defined by: $\partial_1(i,j) = -(\vec{p}_i - \vec{p}_j)v_i + (\vec{p}_i - \vec{p}_j)v_j$.

The set of -1-chains is $\mathbb{R}^{\binom{n+1}{2}}$.

∂_0 is defined by: $(\ldots, (\vec{p}_i - \vec{p}_j), \ldots, [(\vec{p}_i)_h(\vec{p}_j)_k - (\vec{p}_i)_k(\vec{p}_j)_h], \ldots)$, where the second part runs over all pairs of distinct coordinates $1 \leq h < k \leq n$.

The complete chain complex is then

$$\mathcal{R}_d : 0 \longrightarrow \oplus_{e\in E}\mathbb{R} \xrightarrow{\partial_1} \oplus_{v\in V}\mathbb{R}^n \xrightarrow{\partial_0} \mathbb{R}^{\binom{n+1}{2}} \longrightarrow 0.$$

The Euler characteristic of this complex is $|E| - n|V| + \binom{n+1}{2}$. If the set of vertices spans at least a hyperplane of $\mathbb{R}^n$, ∂_0 is onto and $\tilde{H}_{-1}(\mathcal{R})) = 0$. If all 0-cycles are 0-boundaries, that is if $\tilde{H}_0 = 0$, the framework is called *statically rigid*.

The dual cohomology is:

$$\mathcal{R}^d : 0 \longleftarrow 2\oplus_{e\in E}\mathbb{R} \xleftarrow{\delta_1} \oplus_{v\in V}\mathbb{R}^n \xleftarrow{\delta_0} \mathbb{R}^{\binom{n+1}{2}} \longleftarrow 0.$$

The 0-cochains (the kernal of $R_1(G)$) are the *infinitesimal motions*, and the 0-coboundaries are called *trivial motions* (they are the derivatives of Euclidean motions). If all infinitesimal motions are trivial motions, that is if $H^0(\mathcal{R}^d(G)) = 0$, the framework is *infinitesimally rigid*. Since the homology and cohomology are isomorphic in these situations, infinitesimal rigidity is equivalent to static rigidity.

Example 4. Rigidity on the line gives the chain complex: $0 \rightarrow \mathbb{R}^{|E|} \xrightarrow{\partial_1} \mathbb{R}^{|V|} \xrightarrow{\partial_0} \mathbb{R} \rightarrow 0$. The boundary operator gives: $\partial_1([i,j]) = (p_j - p_i)v_i - (p_j - p_i)v_j$. This only differs from the homology map by the non-zero constant $(p_j - p_i)$. The cycle spaces are isomorphic, so this is the simplicial homology of the graph.

Example 5. For higher n, the rigidity matrix retains the appearance of the oriented incidence matrix. Each entry has been replaced by an n-vector, with 0 replaced by $\mathbf{0}$ and the two non-zero entries of a row, $+1$ and -1, replaced by the vectors $\vec{p}_i - \vec{p}_j$ and $-(\vec{p}_i - \vec{p}_j)$. If we had picked a truly new vector $\vec{b}_{ij}$ for each row, this would be the matroid union of graphic matroids (Whiteley 1988a). The maximum possible rank would be $n(|V| - 1)$ - and the actual rank could be determined by repeated use of the algorithms for maximal forests.

For higher space, the matrix for the boundary map retains the superficial appearance of a matroid union of copies of the homology matroid of the graph. Each entry in the homology matroid has been replaced by a vector, with $\vec{0}$ replacing 0, and the two non-zero entries of a row $+1$ and -1 replaced by the vectors $\vec{p_i} - \vec{p_j}$ and $-(\vec{p_i} - \vec{p_j})$. If we had picked a truly new vector $\vec{b}_{ij}$ for each row, this would be the matroid union. The maximum possible rank would then be rank would then $n(|V| - 1)$ - and the actual rank could be determined by repeated use of the algorithms for maximal forests.

However, the actual form of the vector entries imposes a relationship among the entries, and the maximum possible rank is reduced to $n|V| - \binom{n+1}{2}$. What is the impact of this 'specialization' of the entries in the matroid union? For $n = 2$, the maximum rank of the rigidity matrix is $2|V| - 3$, and the matroid is a generic truncation of the matroid union - resulting in precise combinatorial algorithms for the rank (Laman, 1970, Whiteley, 1988a). For $n = 3$, the precise combinatorics are unknown.

However, important partial results support several conjectures on complex combinatorial characterizations (Tay & Whiteley, 1985). In particular, the 1-skeleton of a simplicial convex d-polytope is infinitesimally in d-space (Whiteley, 1984). Because of the underlying projective geometry (see below), crucial tools for these proofs include coning and projections.

Skeletal Rigidity

Recent works of Lee [1991], Filliman [1991] and Tay, White & Whiteley [1992a] have extracted matrices resembling the matrix for rigidity of frameworks, from the face ring of a simplicial complex. Their common goal is to analyse the combinatorics and geometry of the g-theorem of polyhedral combinatorics - a theorem which characterizes the f-vector of a simplicial polytope in n-space:

$$f(\Delta) = (f_o, f_1, \ldots, f_n) = (|\Delta^{(0)}|, |\Delta^{(10}|, \ldots, |\Delta^{(n-1)}|).$$

The g-theorem is a combinatorial theorem with a complex analytic proof, based on the homology of a corresponding toric variety and the hard (or strong) Lefschetz theorem. Major efforts are underway to give a direct combinatorial or geometric proof of the combinatorial result. For g_2, the desired result is

$$0 \leq g_2 = |\Delta^{(1)}| - d|\Delta^{(0)} + \binom{d+1}{2} = |E| - d|V| + \binom{d+1}{2}.$$

This inequality follows from the infiniteismal rigidity of the 1-skeleton of the simplicial complex (Kalai, 1988).

Tay, White & Whiteley [1992b] extends this 'rigidity style' analysis to a complete chain complex. As always, the rank of the highest matrix is expressed by the combinatorics of the complex, and the lower homologies (which should be zero under appropriate assumptions).

We begin with a reexamination of rigidity of bar frameworks. Surprisingly, the ranks of the framework rigidity matrices, or equivalently, the framework matroid, is invariant under projective transformations of the underlying points. The matrices and the chain complex can also be written in a projective form. Let $\tilde{x} = (x_1, \ldots, x_n, 1)$ represent the affine coordinates of the point $\vec{x}$. Then the classical exterior (or Grassman) product of two such points is:

$$\tilde{x} \vee \tilde{y} = (\ldots, x_h y_k - x_k y_h, \ldots) \quad 1 \leq h < k \leq n+1).$$

This extends to the full exterior algebra of a set of points in projective space. All products of d points generate a real vector space $V^{(d)}$. For an oriented simplex $\sigma = [v_0, \ldots, v_d]$, with $\tilde{\sigma} = \tilde{p}_0 \vee \ldots \vee \tilde{p}_n \in V^{(d+1)}$, we have a equivalence relation on elements of $V^{(r)}$

$$P \stackrel{ker\,\sigma}{=} Q \iff P \vee \tilde{\sigma} = Q \vee \tilde{\sigma}.$$

These equivalence classes, $V^{(r)}/ker\,\sigma$ form the coefficients of σ in our chain complex.

Example 6. Consider again the framework rigidity. Since $\tilde{x} \vee \tilde{y} = \mathbf{0}$, $\tilde{p}_i \vee \tilde{p}_j$ is precisely the desired image for $\partial_0(\tilde{p}_i \cdot v_j)$. It is a simple exercise to show that the chain complex can be rewritten as:

$$\mathcal{R}_d(\Delta^1): \ \mathbf{0} \rightarrow \bigoplus_{\{i,j\} \in E} \mathbb{R} \xrightarrow{\partial_0} \bigoplus_{a_i \in \Delta^0} V_d^{(r)}/ker\,a_i \xrightarrow{\partial_0} V_d^{(r+1)} \rightarrow \mathbf{0}.$$

This chain complex has the same homologies as the more usual rigidity chain complex (Crapo & Whiteley 1982).

For a simplicial complex Δ realized in projective n-space, the *r-skeletal chain complex* is

$$\mathcal{R}_n(\Delta^r): \ \mathbf{0} \rightarrow \bigoplus_{\rho \in \Delta^{(r-1)}} V_d^{(0)} \xrightarrow{\partial_{r-1}} \bigoplus_{\sigma \in \Delta^{(r-2)}} V_d^{(1)}/ker\,\sigma \xrightarrow{\partial_{r-2}} \bigoplus_{\tau \in \Delta^{(r-3)}} V_d^{(2)}/ker\,\tau \xrightarrow{\partial_{r-3}} \ldots$$
$$\ldots \xrightarrow{\partial_1} \bigoplus_{a_i \in \Delta^{(0)}} V_d^{(r-1)}/ker\,a_i \xrightarrow{\partial_0} V_d^{(r)} \rightarrow \mathbf{0}.$$

For an elementry d-chain $\tilde{P}\sigma$ the boundary map is defined by:

$$\partial_d(\tilde{P}\sigma) = \partial_d(\tilde{P}[a_0, a_1, \ldots, a_d]) = \sum_{\{j|a_j \in \sigma\}} \tilde{P} \vee \tilde{a}_j \cdot [a_0, a_1, \ldots, \hat{a}_j, \ldots, a_d].$$

This is then extended linearly to all d-chains.

Again, the matrices corresponding to the top operators are formed from the oriented incidence matrices, replacing non-zero entries by appropriate non-zero vectors, and zero entries by corresponding zero vectors. In particular if $r = n + 1$, this is simplicial homology.

For shellable complexes in n-space, and $n \leq 2r + 1$, the lower homologies ($k < r - 2$) are zero, and the study concentrates on showing that $H_{r-2}(\mathcal{R}(\Delta^k)) = 0$ (i.e the r-skeleton is *r-rigid*). We note that, as usual, there are geometric special positions, which depend on the underlying projective geoemtry.

Of course we have the dual *r-skeletal cochain complex* is

$$\mathcal{R}^n(\Delta^r):\ 0 \leftarrow \bigoplus_{\rho \in \Delta^{(r-1)}} V_d^{(0)} \xleftarrow{\delta_{r-1}} \bigoplus_{\sigma \in \Delta^{(r-2)}} V_d^{(1)}/ker\,\sigma \xleftarrow{\delta_{r-2}} \bigoplus_{\tau \in \Delta^{(r-3)}} V_d^{(2)}/ker\,\tau \xleftarrow{\delta_{r-3}} \ldots$$
$$\ldots \xleftarrow{\delta_1} \bigoplus_{a_i \in \Delta^{(0)}} V_d^{(r-1)}/ker\,a_i \xleftarrow{\delta_0} V_d^{(r)} \leftarrow 0.$$

Again, the cohomologies are isomorphic to homologies, but this cohomology gives additional insights into the geometry of the simplicial complex.

Multivariate splines

The theory of mulitivariate splines studies the vector space of all piecwise polynomial (of maximal degree k) globally C^r-functions over a simplicial (or more general polyhedral) decomposition of the domain Ω. Billera [1988] expresses this through a chain complex, as follows. We denote the space of polynomials of degree at most d by P_d. A realized simplex τ generates an ideal I_τ of polynomials which are zero on all vertices of τ, and $(I_\tau)^k$ is the ideal generated by all kth powers.

Consider two polynomial functions f, g of n variables, defined on opposite sides of the hyperplane of a facet τ in $\mathbb{R}^n$ with equation $L(x_1, \ldots, x_n) = 0$. The functions meet with continuity C^r over the simplex τ if and only if:

$$f - g = \beta(x_1, \ldots, x_n)[L(x_1, \ldots, x_n)]^{r+1} \iff f - g \in I_\tau^{r+1}.$$

Assume that the cell complex forms a topological manifold. The functions defined by polynomials of degree at most d over the cells, meeting in pairs over the facets with C^r

continuity, are then the elements of the vector space $\oplus_{\sigma\in\Delta}P_d$ such that, if $\sigma\cap\sigma' = \tau$ for a facet τ, then $f_\sigma - f'_\sigma \in I_\tau^{r+1}$. Thus the splines are cycles for the boundary map: $\partial_{n-1}(f_\sigma\sigma) = f_\sigma\partial_{n-1}(\sigma)$ where $\partial_{n-1}(\sigma)$ is the usual boundary map of homology and the image is taken in $\oplus_{\tau\in\Delta}P_d/(P_d\cap(I_\tau)^{r+1})$.

This boundary operator, expressed in matrix form, looks like a matroid union of the oriented incidence matrix, but, again, the entries in various rows are linked, leaving a specialization of the matroid union.

This basic operation is extended to a chain complex:

$$\mathcal{S}_d^r(\Delta):\ 0 \xrightarrow{\partial_n} \oplus_{\sigma\in\Delta}P_d \xrightarrow{\partial_{n-1}} \oplus_{\tau\in\Delta}P_d/P_d\cap(I_\tau)^{r+1}\ldots \xrightarrow{\partial_0} \oplus_{v\in\Delta}P_d/P_d\cap(I_v)^{r+1} \xrightarrow{\partial_{-1}} 0.$$

Using the Euler characteristic, we find that:

$$\dim(H_n(\mathcal{C}(\Delta)) = \sum_i(|\Delta^{(i)}|) - \sum_{i\leq n-1} H_i(\mathcal{C}(\Delta)).$$

Again the goal is to prove that the lower homologies are trivial, so that this gives the dimension of the space of splines. This has been shown to hold for C_d^1 over generic simplicial manifolds in the plane [Billera, 1988] and in 3-space [Alfeld, Schumaker & Whiteley, 1991].

In practice, this chain complex for splines is also analyzed by means of a short exact sequence of chain complexes - and the resulting long-exact sequence of homologies [Billera, 1988]). Without giving the details, we note that the matrices for all these related chain complexes are formed from the oriented incidence matrix, by inserting appropriate matrices for the non-zero entries across a row (following the other signs) and similarly shaped zero matrices for the zero entries.

In this setting, the corresponding cohomology has not been directly analysed. However it may become a crucial tool for understanding the behaviour of special geometric realizations of the cell complex. We note that, once more, the underlying geometry is projective (Whiteley 1991b). This geometry is important in the use of such tools as coning and projection in the proof of the results.

Polyhedral pictures

The theory of polyhedral pictures studies a plane realization of a set of polygons, together with the vector space of *spatial liftings* which lift the underlying vertices into space, and keeps each of the polygons flat. Without going into detail, we note:

(a) The problem of polyhedral pictures is isomoprphic the the problem of piecewise linear, globally continuous functions, C_1^0 splines, if the polygons do not overlap.

(b) Classical results of James Clerk Maxwell [1864] connect polyhedral pictures with the self-stresses of a planar graph (see also Whiteley, 1982, Crapo & Whiteley, 1986).

(c) These correspondences extend to pictures of 'polytopes' in higher dimensions connecting to splines and skeletal rigidity.

These particular chain complexes are defined only if the underlying cell complex is a manifold (with boundary). However there are also extended theories:

(d) There is an explicit theory of such polyhedral scenes, via a 'geometric' homology theory, elaborated by Crapo & Ryan (1986a,b). This theory is based on a theory of chains and boundary operators different from the standard homology.

(e) Crapo's homology and cohomology is intimately related to the analysis of C^0 splines by sheaves and cohomology (Yuzvinski,1991).

For both this general theory and the manifold theories there are explicit combinatorial algorithms for the lifitings of a generic realization, derived from truncations of a matroid union of appropriate incidence matrices (Sugihara, 1986, Whiteley, 1988b).

Summary.

The oriented incidence matrices of cell complexes, and the corresponding chain complexes, are the combinatorial foundation of an important class of problems in discrete applied geometry. The combinatorial analysis of these extensions draws on everything we know about the homology of the cell complex - and poses additional combinatorial questions for the structures, and their matroids.

The analysis of each of these geometric chain complexes includes the following steps:
(i) A proof that the 'lower homologies' (all but the top two) are zero, under appropriate assumptions (e.g. a strongly connected manifold with boundary). This involves simple arguments, and induction from step (ii), below, for structures in lower dimensions.
(ii) A proof (or conjecture) that the matrix for the top boundary operation has 'maximum rank' (defined from the Euler characteristic), possibly under additional assumptions, so that only the top homology is non-zero. This involves more direct combinatorial analysis of the matrix, and the underlying cell complex.

For example, framework rigidity is analysed by graph theoretic inductions, in ways that apply to all triangulated manifolds ($n \geq 3$) (Tay & Whiteley, 1985, Whiteley, 1984, 1991a). The results actually apply to a larger class of 'minimal homology cycles' - sets of simplices which are a cycle in the homology of an underlying complex, and are minimal with this property (Fogelsanger, 1989).

(iii) For each problem, the chain complexes for related structures in adjacent dimensions are connected by geometric or combinatorial constructions. For example, cones of structures in lower dimensions connect the homologies of the original structure in dimension $n-1$ and the cone in dimension n (Whiteley, 1983, 1984, Alfeld, Schumaker & Whiteley, 1991). This coning reflects the underlying projective invariance for each of the examples (Crapo & Whiteley, 1982, Whiteley, 1991b).

(iv) If the chain complex is exact for generic realizations (the homologies are all zero) then the geometry of non-generic realizations is captured in the 'determinant' of the complex (Gelfand, Kapranov & Zelivinski, 1991), generalizing the pure conditions of rigidity theory (White & Whiteley 1984).

Some common combinatorial techniques have been developed for these examples. Other basic combinatorial and algorithmic questions remain to be solved. Continuing combinatorial research on each of these problems should expand the results in all these related fields.

Finally, although we have talked about the role of combinatorics in understanding these geometric problems, we should not ignore the role of geometry as a tool to prove combinatorial results. One feature of the the work on the g-theorem is the role of geometric tools (either toric varieties or perhaps skeletal rigidity) in proofs of the combinatorial results. The interaction of combinatorics and geometry is a two-way street.

References.

P. Alfeld, L.L. Schumaker, and W. Whiteley 1991; The generic dimension of the space of C^1 splines of degree ≥ 8 on tetrahedral decompositions; *SIAM J. Numerical Analysis*, to appear.

L. Asimow and B. Roth, 1978; Rigidity of graphs; *Trans. Amer. Math. Soc.* 245, 279-289.

L. Billera, 1988; Homology of smooth splines: generic triangulations and a conjecture by Strang; Trans. A.M.S. **310**, 325–340.

H. Crapo & J. Ryan, 1986a; Scene analysis and geometric homology; Proceedings of the Second ACM Symposium on Computational Geometry.

H. Crapo & J. Ryan, 1986b; Spatial realizations of linear scenes; *Structural Topology* 13, 33-69.

H. Crapo and W. Whiteley, 1982; Stresses in frameworks and motions of panel structures: a projective introduction; *Structural Topology*, 43-82.

H. Crapo and W. Whiteley, 1986; Plane stresses and projected polyhedra I: the basic pattern; to appear, *Structural Topology*.

P. Filliman, 1991; Face numbers of pl-spheres; MSI technical report 91-4, Cornell University, Ithaca New York.

A. Fogelsanger, 1989; Generic rigidity of minimla homology cycles; PhD Thesis, Cornell University, Ithaca New York.

I. Gelfand, M. Kapranov, and A. Zelevinski, 1991; Determinants of complexes; preprint chapter for their book 'Resultants and Discriminants'.

G. Kalai, 1987; Rigidity and the lower bound theorem I; *Invent. Math* 88, 125-151.

F. Klein, 1872; **Vergleichende Betrachtungen über neure geometrische**; Forschungen, Erlangen, (*Math. Annalen* 43, p63-?).

G. Laman, 1970; On graphs and the rigidity of plane skeletal structures; *J. Engineering Math.* 4, 331-340.

C. Lee, 1991; Generalized stress and convex polytopes; preprint University of Kentucky, Lexington, Kentucky.

J.C. Maxwell, 1864; On reciprocal figures and diagrams of forces; *Phil. Mag.* series 4 27, 250-261.

J. E. Munkres, 1984; Elements of Algebraic Topology; Addison-Wesley, Reading, Massachusetts.

K. Sugihara, 1986; **Machine Interpretation of Line Drawings**; MIT press Cambridge MA.

T-S. Tay, N, White and W. Whiteley, 1992a; Skeletal rigidity of simplicial complexes; Mittag-Lefler technical report 1991-92 #20.

T-S. Tay, N, White and W. Whiteley, 1992b; A homological interpretation of skeletal rigidity; preprint Champlain Regional College.

T-S. Tay and W. Whiteley, 1985; Generating all isostatic frameworks; *Structural Topology* 11, 21-69.

N. White and W. Whiteley, 1984; The algebraic geometry of stresses in frameworks; *SIAM Journal of Algebraic and Discrete Methods* 4, 481-511.

W. Whiteley, 1982; Motions, stresses and projected polyhedra; *Structural Topology* 7, 13-38.

W. Whiteley, 1983; Cones infinity and one-story buildings; *Structural Topology* 8, 57-70.

W. Whiteley, 1984; Infinitesimally rigid polyhedra I: statics of frameworks; *Trans. Amer. Math. Soc.* 285, 431-465.

W. Whiteley, 1988a; The union of matroids and the rigidity of frameworks; *SIAM J. Discrete Methods* 1, 237-255 .

W. Whiteley, 1988b; Some matroids on hypergraphs with applications to scene analysis and geometry; *Disc. and Comp. Geometry* 4, 75-95.

W. Whiteley, 1991a; Vertex splitting in isostatic frameworks; *Structural Topology* 16, 23-30.

W. Whiteley, 1991b; The combinatorics of bivariate splines; in *Applied Geometry and Discrete Mathematics, Victor Klee Festschrift*, P. Gritzmann and B. Sturmfels (eds.), DIMACS series, AMS Press, 567–608.

W. Whiteley, 1992; Matroids and rigidity; in **Matroid Applications**, Neil White (ed.), Cambridge University Press, 1-53.

S. Yuzvinski, 1991; Modules of splines on polyhedral complexes; preprint, University of Oregon Eugene, OR 97403.

Address

Walter Whiteley
Champlain Regional College
900 Riverside Drive,
St. Lambert, Québec,
J4P-3P2 Canada.
e-mail whiteley@ere.umontreal.ca

The Plethystic Inverse of a Formal Power Series (Extended Abstract)

Julia S. Yang*

Abstract

A bijective definition of Littlewood's plethysm of symmetric functions in terms of unlabelled colored combinatorial structures is introduced. The plethystic inverse is understood by examining associated partially ordered sets and using Möbius inversion techniques.

The combinatorial theory of distribution and occupancy (also called "balls in boxes") provides a classical interpretation of symmetric functions. In this view, one counts the number of ways to place distinct balls in distinct boxes such that certain conditions are met (e.g., there is no more than one ball per box). Alternately, the balls can be viewed as "labelled nodes" of a graph, and the boxes can be seen as additional colors assigned to the nodes. Symmetric functions then count the number of colored labelled structures by keeping track of how many use a particular color scheme.

The first systematic development of this viewpoint was given by Doubilet. [4] More recently, Bonetti, Rota, Senato, and Venezia provided set-theoretic interpretations for infinite sums, infinite products, and exponentiation of symmetric functions; this permitted bijective interpretations for many classical symmetric function identities. [2] In this paper, the equivalence classes of colored, labelled structures under group action are examined. This leads to a bijective definition of Littlewood's notion of plethysm of symmetric functions. This theory of symmetric functions is then extended to include posets, and an interpretation of the plethystic inverse is given.

Intuitively, the equivalence classes under the group action may be visualized as unlabelled graphs with colored nodes. This is an extension of types and analytic functors, or "Joyal labellings".[5][6] The combinatorial operations of addition, multiplication, and exponentiation can be readily extended to these unlabelled, colored structures, which will be called colored types.

In addition, the appropriate composition of colored types corresponds to Littlewood's plethysm of the generating functions. Usually, formal power series in infinitely many variables count the number of structures built on objects like partitions of labelled sets, and the resulting interpretations of plethysm can be rather complicated. However, when attention is turned to colored types, the construction of a plethysm becomes remarkably simple. Given two families of colored types M and N, the plethysm of N into M yields M-structures whose colored nodes have been replaced with the appropriate N-structures; two "nodes"

*Partially supported by a National Science Foundation Graduate Fellowship. This material formed part of the author's doctoral dissertation.

have the same "color" when the associated N-structures are the "same". This plethysm is a generalization of the usual combinatorial interpretation for Littlewood's plethysm.

The theory of symmetric functions is then extended to include posets. Furthermore, the plethystic inverse is understood by applying Möbius inversion techniques to suitable lattice structures based on colored combinatorial objects.

In Section 1, background material from the theory of symmetric functions and the theory of polynomial species is reviewed. In Section 2, the plethysm is introduced. Then several examples are provided; for example, the plethysm of a symmetric function into a homogeneous symmetric function is shown to correspond to assemblies (sets) of colored types. In Section 3, the plethystic inverse is addressed by examining appropriate lattice-theoretic structures.

1 Background Material

1.1 Symmetric Functions and Littlewood's Plethysm

Let $X = \{x_i\}_{i \in I}$ be a finite or infinite set of variables with index set, I. A *symmetric function* is a formal power series in X with integer coefficients which is invariant under permutation of the variables and of bounded degree. The symmetric functions form a ring denoted by $\Lambda(X)$.

Let $\bigoplus_I \mathbb{N}$ denote the set of sequences of non-negative integers indexed by I with only finitely many non-zero entries. For $\alpha = (\alpha_i) \in \bigoplus_I \mathbb{N}$, x^α denotes the monomial $\prod_i x_i^{\alpha_i}$.

Let $F(X) = \sum_{k \geq 0} f_k(X)$ and $G(X) = \sum_{k \geq 1} g_k(X)$ be two sums of symmetric functions in the variable set X where $f_k(X)$ and $g_k(X)$ are symmetric functions of degree k (these sums are called generating functions of symmetric functions). Suppose that the index set is countably infinite. A special kind of multiplication for these generating functions, called the *plethysm* or *composition*, can be defined. The basic idea is that if $G(X)$ is expressed as a sum of terms of the form x^α, then the plethysm $F * G(X)$ is obtained by substituting the terms of $G(X)$ into the variables x_i in $F(X)$. Formally, write $G(X)$ as the sum of monomials: $G(X) = \sum_\alpha w_\alpha x^\alpha$ where the sum ranges over all $\alpha \in \bigoplus_I \mathbb{N}$.

Define a set of dummy variables, y_i, such that

$$\prod_i (1 + y_i z) = \prod_\alpha (1 + x^\alpha z)^{w_\alpha} .$$

To simplify notation, suppose that the index set, I, is linearly ordered. Littlewood's plethysm is defined as

$$F * G(x_1, x_2, \ldots) = F(y_1, y_2, \ldots).$$

When $F(X)$ is homogeneous of degree n and $G(X)$ is homogeneous of degree k, then $F*G(X)$ is homogeneous of degree nk.

See [9] for more about symmetric functions and plethysm.

1.2 Polynomial Species

The theory of polynomial species provides set-theoretic interpretations of operations on symmetric functions. In this subsection, the definitions for polynomial species and the combinatorial operations of addition, multiplication, and exponentiation will be reviewed.

Let **Ens** be the category of sets and functions, and let **B** be the category of finite sets and bijections. Define a covariant functor $Hom : \mathbf{B} \rightarrow \mathbf{Ens}$ as $Hom[E] = \{f : E \rightarrow X\}$ where E is a finite set, and f is a function.

Let $M : \mathbf{B} \rightarrow \mathbf{Ens}$ be a functor (frequently, M is a (Joyal) species $M : \mathbf{B} \rightarrow \mathbf{B}$, and the image of a finite set E under M is called the set of all M-structures with label set E). Let R be a subfunctor of $M \times Hom$. Given a R-structure (m, f), the function f may be visualized as "coloring" the labels (or nodes) of the M-structure with elements from the variable set, X. A bijection $g : E \rightarrow E'$ of finite sets induces the morphism

$$\begin{array}{llcl} R[g]E: & R[E] & \rightarrow & R[E'] \\ & (m, f : E \rightarrow X) & \mapsto & (M[g]m, f \circ g^{-1}). \end{array}$$

The subfunctor R is called a *polynomial species* when for any finite subset $F \subseteq X$, the set $\{(m,f) \in R[E] : \text{im } f \subseteq F\}$ is finite; im f denotes the set of variables in the image of f. This technical condition ensures that the corresponding generating functions have well-behaved coefficients.

A polynomial species is said to be *symmetric* when for any bijection $u : X \rightarrow X$ and for any $(m, f) \in R[E]$, $(m, u \circ f) \in R[E]$.

Given any polynomial species R and any non-negative integer n, let $R_{[n]}$ be the polynomial species where $R_{[n]}[E] = R[E]$ if $|E| = n$, and $R_{[n]}[E] = \emptyset$ otherwise. Let R_0 denote the polynomial species where $R_0[E] = R[E]$ if E is nonempty, and $R_0[E] = \emptyset$ otherwise.

The generating functions associated with a polynomial species will now be covered.

For a finite set E and an arbitrary function $f : E \rightarrow X$, let $\text{gen}(f) = \prod_{e \in E} f(e) = x^\alpha$ where $\alpha = (\alpha_i) \in \bigoplus_I \mathbb{N}$, and α_i is the cardinality of the set $\{f^{-1}(x_i)\}$. Given any R-structure (m, f), call f the *coloring*, and call α or x^α the *color scheme*. Set $\text{gen}(m, f) = \text{gen}(f)$. Define

$$\text{gen}(R[E]) = \sum_{(m,f) \in R[E]} \text{gen}(m, f).$$

Note that every monomial in the sum has a finite coefficient by the definition of a polynomial species. Also, if R is symmetric, then $\text{gen}(R[n])$ is a symmetric function of degree n where $R[n] = R[\{1, 2, \cdots, n\}]$.

The *generating function* of a polynomial species, R, is defined as the formal power series

$$\text{Gen}(R) = \sum_{n \geq 0} \text{gen}(R[n])/n!.$$

Polynomial species form a category whose morphisms are natural maps $\tau : R \rightarrow S$ such that for any finite set, E, $\tau_E : R[E] \rightarrow S[E]$ is a bijection where $\tau_E(m, f) = (m', f)$ for $(m, f) \in R[E]$. Write $R = S$ when τ is a natural isomorphism. R and S are said to be *equipotent* ($R \equiv S$) if $\text{gen}(R[n]) = \text{gen}(S[n])$ for all non-negative integers n. This is a weaker condition.

The combinatorial operations of addition, multiplication, and exponentiation will now be covered. Let $R \subseteq M_1 \times Hom$ and $S \subseteq M_2 \times Hom$ be two polynomial species. Let E be a finite set.

The *sum* and *product* of R and S are given by

$$R + S[E] = R[E] \bigcup S[E]$$
$$R \cdot S[E] = \textstyle\sum_{E_1 + E_2 = E} \{((m, m'), f) : (m, f \mid E_1) \in R[E_1], (m', f \mid E_2) \in S[E_2]\}.$$

Infinite sums and infinite products may also be defined.

Suppose that a polynomial species, S, is without constant term ($S[0] = \emptyset$). An *assembly* of S-structures with associated partition π is an ordered pair (a, f) where $(m_B, f_B) \in S[B]$ for every $B \in \pi$, $a = \{m_B\}$, and $f|B = f_B$ for all $B \in \pi$. The *k-th divided power* $\gamma_k(S)$ is the polynomial species of assemblies of S-structures where the associated partition contains k blocks. The *exponential* $\exp(S)$ is the polynomial species of all assemblies of S-structures.

Note that when $|X| = 1$, and M_1 and M_2 are Joyal species, these operations reduce to the corresponding operations on species.

The generating functions for these operations behave as desired.

Theorem 1.1 *Let R and S be polynomial species. Then*

$$\begin{aligned} Gen(R+S) &= Gen(R) + Gen(S) \\ Gen(R \cdot S) &= Gen(R) \cdot Gen(S). \end{aligned}$$

When S is a polynomial species without constant term,

$$\begin{aligned} Gen(\gamma_k(S)) &= Gen(S)^k/k! \\ Gen(\exp(S)) &= \exp(Gen(S)). \end{aligned}$$

2 Colored Types and Plethysm

In the standard combinatorial approach to Littlewood's plethysm (which will be phrased in the language of species), given two Joyal species M and N where $N[0] = \emptyset$, the plethysm corresponds to unlabelled $M \circ N$-structures whose nodes are colored by the variable set X in all ways.[1]

In this section, a more refined plethysm is introduced where we may restrict which color schemes appear. More specifically, let R be a symmetric polynomial species. The group $E!$ of permutations on E acts on $R[E]$ by the morphism of structures. The orbits of $R[E]$ under this group action may be identified with *unlabelled* R-structures. This is a generalization of *types* which correspond to unlabelled structures, and of analytic functors or *Joyal labellings* which correspond to unlabelled structures colored in all ways. Call the unlabelled R-structures *colored types.*

The plethysm of colored types has a remarkably simple combinatorial interpretation. Let R and S be two symmetric polynomial species. Associate a color in X to each unlabelled S-structure. Then the operation of plethysm takes an unlabelled R-structure and replaces each colored node with an unlabelled S-structure of the same associated color.

This operation is related to the plethysm of symmetric functions. While polynomial species are associated with exponential generating functions, orbits of polynomial species (and hence, colored types) correspond to ordinary generating functions. Let $orb(S[E])$ denote the set of orbits of $S[E]$ under the group action. Set $\text{gen}(\underline{q}) = \text{gen}(q)$ where q is a representative of the orbit $\underline{q}$. The generating function for the orbits of S-structures is defined as

$$\text{Gen}(orb(S)) = \sum_{n \geq 0} \sum_{\underline{q} \in orb(S[n])} \text{gen}(\underline{q}).$$

[1] Recall that $M \circ N$ consists of all pairs (a, m) where a is an assembly of N-structures and m is an M-structure on the associated partition of a.

Note that this generating function equals the (ordinary) generating function for the colored types of S where each unlabelled S-structure contributes its color scheme to the sum. It will be shown that the plethysm of the generating functions for the colored types of R and S equals the generating function for the plethysm of the colored types as defined above.

These notions will now be covered rigorously.

2.1 Colored Types

In practice, it is inconvenient to work directly with orbits in these kinds of arguments. For example, let $M : \mathbf{B} \to \mathbf{B}$ be a Joyal species. In the standard approach to Littlewood's plethysm, one begins with the cycle index series $Z_M \in \mathbb{Q}[x_1, x_2, \ldots]$ for M, and the usual plethysm or *wreath product* for formal power series in infinitely many variables (not Littlewood's plethysm) commutes with the combinatorial operation of substitution. [5] Then the results for symmetric functions are obtained by replacing x_n with the power sum symmetric function $p_n(X)$. Combinatorially, this corresponds to taking a labelled structure with automorphism σ and then coloring its nodes by *independently* coloring the cycles of σ; this is equivalent to examining unlabelled structures colored in all ways. In [1], F. Bergeron extends this method by examining structures based on permutations.

In this section, a different approach is taken where we work with the symmetric functions directly rather than going through the intermediate step of power sum substitution. Combinatorially, this corresponds to beginning with a colored labelled structure, and then examining an automorphism of the structure which preserves the coloring. The advantage of this method is that the colorings do not depend upon an underlying permutation of the labelled structure; this results in more control over the coloring schemes. For example, the elementary symmetric functions can be interpreted in this context.

From this point forward, assume that all polynomial species are symmetric unless otherwise specified, and let $X = \{x_i\}_{i \in I}$ be a countably infinite variable set.

Definition 2.1 Let S be a polynomial species. The (polynomial) species of colored types $\tilde{S}$ is defined as follows:

$$\tilde{S}[E] = \{(\sigma, m, f) : (m, f) \in S[E],\ \sigma \in E!, \text{ and } \sigma(m, f) = (m, f)\}$$

for E a finite set. In other words, a $\tilde{S}$-structure is a pair (σ, s) where σ is an automorphism of the S-structure, s. ■

By Burnside's Lemma, the exponential generating function of the polynomial species $\tilde{S}$ is equal to the ordinary generating function for the orbits of S.

Before defining the combinatorial operation of plethysm, colored types will be assigned colors in the variable set X. For now, assume that S is nontrivial.

Theorem 2.1 *Any symmetric polynomial species S may be decomposed into its orbits*

$$S = \sum_{i \in I} S_i$$

where the S_i are polynomial species (not necessarily symmetric), and the sum ranges over the index set, I, of the variable set, X. More formally, for $(m, f) \in S_i[E]$, and $(m', f') \in S_j[E]$, (m, f) and (m', f') are in the same orbit iff $i = j$.

Let x_i be the *associated color* of an S_i-structure (m, f), and let y_i be the color scheme used by this structure ($y_i = \text{gen}(f)$).

The decomposition of S into its orbits induces a decomposition of colored types:

$$\widetilde{S} = \sum_i \widetilde{S}_i.$$

As before, let x_i be the associated color of a $\widetilde{S}_i$-structure. The monomial y_i is simply the color scheme of any $\widetilde{S}_i$-structure; the variables, y_i, will be called the *dummy variables* for $\widetilde{S}$. By Burnside's Lemma, $\text{Gen}(\widetilde{S}_i) = y_i$.

Example 2.1 Let A be the *elementary symmetric species* of uniquely colored sets. $A[0] = \{(\emptyset, f_\emptyset)\}$, and for a nonempty finite set E, $A[E] = \{(E, f) : f \text{ is injective}\}$. $\widetilde{A}$ may be identified with distinctly colored unlabelled sets. Let $\bigoplus_I\{0,1\}$ denote the set of $0-1$ sequences indexed by I with only finitely many non-zero entries. Since two A-structures with the same subjacent set and same coloring scheme are in the same orbit, $\widetilde{A} = \sum A_\alpha$ where the sum ranges over all $\alpha \in \bigoplus_I\{0,1\}$, and $y_\alpha = x^\alpha$. In addition, $\text{Gen}(\widetilde{A})$ equals the generating function for the elementary symmetric functions. ∎

Example 2.2 Let U be the *uniform symmetric species* of colored sets. For a finite set, E, $U[E] = \{(E, f) : f : E \to X\}$. $\widetilde{U}$ corresponds to unlabelled, colored sets. Since two U-structures with the same subjacent set and the same coloring scheme are in the same orbit, $\widetilde{U} = \sum U_\alpha$ where the sum ranges over all $\alpha \in \bigoplus_I \mathbb{N}$, and $y_\alpha = x^\alpha$. $\text{Gen}(\widetilde{U})$ is equal to the generating function for the homogeneous symmetric functions. ∎

Example 2.3 Let L be the *linear symmetric species* of monochromatic linear orders. $L[0] = \emptyset$, and for a nonempty finite set, E, $L[E] = \{(l, f) : l \text{ linear order on } E, |\text{im } f| = 1\}$. $\text{Gen}(\widetilde{L})$ is equal to the generating function for the power sum symmetric functions. ∎

2.2 Plethysm of Polynomial Species

Let R and S be nontrivial polynomial species where S is without constant term. Recall that the species of colored type, $\widetilde{S}$, can be decomposed ($\widetilde{S} = \sum \widetilde{S}_i$). The monomials y_i (where y_i is the color scheme of an $\widetilde{S}_i$-structure) are the dummy variables for $\text{Gen}(\widetilde{S})$.

A rigorous definition for the plethysm of polynomial species of colored type may now be given.

Definition 2.2 Let $\widetilde{R}$ and $\widetilde{S}$ be two polynomial species of colored type. The plethysm $\widetilde{R} * \widetilde{S}[E]$ consists of structures of the form $(\sigma, a, f, \hat{\sigma}, m, \hat{f})$ where

1. (a, f) is an assembly of S-structures with automorphism σ i.e., $(\sigma, a, f) \in \widetilde{\exp(S)}[E]$. Let π be the associated partition.

2. The decomposition of S induces a coloring $\hat{f} : \pi \to X$; namely, if $(s, h) \in S_i[B]$ is a member of the assembly built on the block $B \in \pi$, then $\hat{f}(B) = x_i$. Furthermore, since the automorphism σ transports an S-structure in the assembly to itself or to a neighboring S-structure of the same type, σ induces a permutation $\hat{\sigma} : \pi \to \pi$ of the associated partition such that $\hat{f} \cdot \hat{\sigma} = \hat{f}$. Specifically, $\hat{\sigma}(B) = B'$ when there exists representatives $b \in B$, $b' \in B'$ such that $\sigma(b) = b'$.

3. The $\tilde{R}$-structure $(\hat{\sigma}, m, \hat{f})$ is an element of $R[\pi]$ where the coloring $\hat{f}$ and the automorphism $\hat{\sigma}$ are as specified above.

■

Theorem 2.2 *Let R and S be polynomial species. Then*

$$Gen(\tilde{R} * \tilde{S}) \quad = \quad Gen(\tilde{R}) * Gen(\tilde{S}).$$

2.3 Examples

Theorem 2.3 *$\tilde{A}_{[1]}$ acts as a two-sided identity for any polynomial species $\tilde{S}$ (without constant term).*

$$\tilde{S} * \tilde{A}_{[1]} = \tilde{A}_{[1]} * \tilde{S} = \tilde{S}.$$

Theorem 2.4 *Let R, S, and T be polynomial species where T is without constant term, and let n and k be positive integers. Then*

1. $\tilde{L}_{[n]} * \tilde{L}_{[k]} = \tilde{L}_{[k]} * \tilde{L}_{[n]} = \tilde{L}_{[n \cdot k]}$.
2. $\tilde{T} * \tilde{L}_{[n]} \equiv \tilde{L}_{[n]} * \tilde{T}$.
3. $(\tilde{R} + \tilde{S}) * \tilde{T} = \tilde{R} * \tilde{T} + \tilde{S} * \tilde{T}$.
4. $(\tilde{R} \cdot \tilde{S}) * \tilde{T} = (\tilde{R} * \tilde{T}) \cdot (\tilde{S} * \tilde{T})$.

Let S be any non-trivial polynomial species such that $S[0] = \emptyset$. Intuitively, $\tilde{A} * \tilde{S}$ corresponds to unlabelled assemblies of S-structures where no two members are the "same".

Example 2.4 For any positive integer k, $\tilde{A} * \tilde{A}_{[k]}$ may be associated with unlabelled assemblies of distinctly colored k-sets where no two sets use the same color scheme. Furthermore, $a_n * a_k(X)$, which equals the symmetric function of degree nk in the corresponding generating function, counts the number of unlabelled assemblies (as specified above) with n members. ■

Example 2.5 $\tilde{A} * \tilde{U}_{[k]}$ corresponds to unlabelled assemblies of colored k-sets where no two sets use the same color scheme. Also, $a_n * h_k(X)$, which equals the symmetric function of degree nk in the associated generating function, counts the number of unlabelled assemblies of colored k-sets where there are n sets in the assembly, and no two sets use the same color scheme. ■

Let S be any polynomial species without constant term. The polynomial species $\widetilde{\exp(S)}$ is isomorphic to $\tilde{U} * \tilde{S}$. Intuitively, $\widetilde{\exp(S)}$ corresponds to unlabelled assemblies of S-structures, and therefore the plethysm of a symmetric function into a homogeneous symmetric function counts unlabelled assemblies.

Example 2.6 The polynomial species $\tilde{U} * \tilde{A}_{[k]}$ corresponds to unlabelled assemblies of distinctly colored k-sets. Also, $h_n * a_k(X)$, which equals the symmetric function of degree nk in the associated generating function, counts the number of unlabelled assemblies of distinctly colored k-sets where the assembly contains n members. ■

Example 2.7 For any positive integer k, $\tilde{U} * \tilde{U}_{[k]}$ corresponds to unlabelled assemblies of colored k-sets. Furthermore, $h_n * h_k(X)$, which equals the symmetric function of degree nk in the associated generating function, counts the number of unlabelled assemblies of colored k-sets with n members. ∎

3 Plethystic Inverse

Let $F(X) = \sum f_k(X)$ be a generating function of symmetric functions in the variable set X where each $f_k(X)$ contains only non-negative integer coefficients. The function $a_1(X)$ acts as a two-sided identity for $F(X)$:

$$F(X) * a_1(X) = a_1(X) * F(X) = F(X).$$

When $f_0(X) = 0$ and $f_1(X) = a_1(X)$, $F(X)$ has an associated generating function $G(X)$ which acts as its plethystic inverse:

$$G(X) * F(X) = a_1(X).$$

Chen and Read have examined inverses of formal power series under the wreath product. [3][11] Joyal and G. Labelle have studied inversion of species in the context of virtual species, and inversion of indicator series for species under the wreath product. [6][7][8] In this section, the plethystic inverse of a generating function of symmetric functions is given a combinatorial interpretation by examining appropriate lattice-theoretic structures and using Möbius inversion techniques. [12] The general method used here for interpreting inverses was introduced in [10]. First of all, the notion of a polynomial species is extended by taking as the base category not families of colored structures, but families of colored finite partially ordered sets with unique minima and maxima. The generating function for such a *Möbius polynomial species* is computed by replacing the notion of cardinality with the evaluation of the Möbius function. This permits interpretation of symmetric functions with negative terms.

Then given a (symmetric) polynomial species $\tilde{R}$, a Möbius polynomial species $\tilde{R}^{[-1]}$ is associated to it which is, in a natural way, its inverse. A partial order is defined on assemblies of R-structures provided with automorphisms by first defining something like a monoid on these structures. This partial order leads to a Möbius polynomial species whose generating function is the plethystic inverse of the generating function for $\tilde{R}$.

A summary of this procedure is given below. For a fuller exposition, see [14].

3.1 Möbius Polynomial Species

In this subsection, polynomial species are extended to colored families of partially ordered sets by combining polynomial species with Möbius species. [2][10]

A Möbius species is a functor $M : \mathbf{B} \rightarrow \mathbf{Int}$ from the category of finite sets and bijections to the category of finite families of finite posets with $\hat{0}$ and $\hat{1}$ where the morphisms are bijective functions $f : A \rightarrow B$ such that if $f(I) = I'$, then I and I' are isomorphic as posets. A *Möbius polynomial species* is defined as a subfunctor $R \subseteq M \times Hom$ such that for any finite subset F of the variable set, X, the set $\{(m, f) \in R[E] : \text{im } f \subseteq F\}$ is finite. R is said to be symmetric when for any bijection $u : X \to X$, and for any structure $(m, f) \in R[E]$, $(m, u \circ f) \in R[E]$. From now on, assume that a Möbius polynomial species is symmetric.

Set $\text{gen}(m, f) = \mu(m)\text{gen}(f)$ where (m, f) is an R-structure, and $\mu(m)$ is the Möbius function of the interval m evaluated at the two extremes. Define

$$\text{gen}(R[E]) = \sum_{(m,f)\in R[E]} \text{gen}(m, f).$$

Since R is symmetric, $\text{gen}(R[n])$ is a symmetric function of degree n with (possibly negative) integer coefficients.

The *Möbius generating function* for R is defined as

$$\text{Gen}(R) = \sum_{n\geq 0} \text{gen}(R[n])/n!.$$

Addition and multiplication from the theory of polynomial species can be extended to Möbius polynomial species; the corresponding generating function identities behave as desired since the Möbius valuation acts with respect to the sum and product of partially ordered sets in the same way that set cardinality does for ordinary sets. Furthermore, the additive and multiplicative inverses of polynomial species can be interpreted. [13]

3.2 $\mathcal{A}_*$ and the Partial Order

First of all, a little terminology. Recall that any nontrivial (symmetric) polynomial species S may be decomposed into its orbits: $S = \sum S_i$. In subsequent discussions, any polynomial species will come with an implicit decomposition. When S is without constant term, an assembly of S-structures (a, f) with associated partition π has an *associated coloring* $\hat{f} : \pi \to X$ where $\hat{f}(B) = x_i$ for $B \in \pi$ when the corresponding member of the assembly built on B is an S_i-structure (refer to the definition of plethysm for the relevance of $\hat{f}$). Let σ be an automorphism of the assembly (a, f). The *associated partition* of (σ, a, f) is defined as the associated partition of (a, f). The structure (σ, a, f) has an *induced permutation* $\hat{\sigma} : \pi \to \pi$ of the associated partition where $\hat{\sigma}(B) = B'$ when there exists $b \in B$ and $b' \in B'$ with $\sigma(b) = b'$.

For any polynomial species R, let $el(R)$ denote the category whose objects are R-structures, and whose morphisms are isomorphisms of R-structures. Let $\mathcal{A}_*$ be the family whose members are of the form $el(\widetilde{U}_0 * \widetilde{R})$ where R is a (symmetric) polynomial species such that $\text{gen}(R[0]) = 0$ and $\text{gen}(R[1]) = a_1(X)$. For $el(\widetilde{U}_0 * \widetilde{R})$ and $el(\widetilde{U}_0 * \widetilde{S})$ in $\mathcal{A}_*$, define a product $\otimes$ as

$$el(\widetilde{U}_0 * \widetilde{R}) \otimes el(\widetilde{U}_0 * \widetilde{S}) = el(\widetilde{U}_0 * (\widetilde{S} * \widetilde{R})).$$

(Actually, $\mathcal{A}_*$ is a special kind of category, and $\otimes$ is a special kind of bifunctor, but details have been ommitted for brevity.)

Definition 3.1 A *c-monoid* in $\mathcal{A}_*$ is a triple (A, c, I) consisting of a nonvacuous element $A \in A$, a functor $c : A \otimes A \to A$, and an $I \subseteq A$ such that

1. (a) (associativity) The following diagram commutes.

$$\begin{array}{ccccc} (A\otimes A)\otimes A & \xrightarrow{=} & A\otimes(A\otimes A) & \xrightarrow{id\otimes c} & A\otimes A \\ \downarrow{\scriptstyle c\otimes id} & & & & \downarrow{\scriptstyle c} \\ A\otimes A & & \xrightarrow{\quad c \quad} & & A \end{array}$$

Note that $A \otimes A$ is a subcategory of the product category $A \times A$. Write $a_1 \cdot a_2$ for $c(a_1, a_2)$.

(b) (left cancellation) For any $a_1, a_2, a_2' \in A$ such that $(a_1, a_2), (a_1, a_2') \in A \otimes A$, if $a_1 \cdot a_2 = a_1 \cdot a_2'$ then $a_2 = a_2'$.

2. (a) (right identity) The product c induces a natural isomorphism $\gamma : A \otimes I \rightarrow A$. In particular, for any $a \in A$, $\gamma(a, i_R(a)) = a$ iff $a \cdot i_R(a) = a$.

(b) (left identity) The product c induces a natural isomorphism $\beta : I \otimes A \rightarrow A$. In particular, for any $a \in \mathcal{A}$, $\beta(i_L(a), a') = a$ iff a' is isomorphic to a in A and $i_L(a) \cdot a' = a$.

(c) (no proper divisors of unity) If there is an $i \in I$ and $(a_1, a_2) \in A \otimes A$ such that $a_1 \cdot a_2 = i$ then $a_1 = i$.

(d) (graded) Let $|a|$ equal the number of blocks of the associated partition of any $a \in A$. Then for any $a_1, a_2, a_2' \in A$ where $a_1 \cdot a_2 = a_3$, $|a_2| = |a_3|$.

■

Example 3.1 Recall that L is the linear symmetric species of monochromatic linear orders. $(el(\widetilde{U}_0 * \widetilde{L}), c, I)$ is a c-monoid where given $(\sigma, a, f, \hat{\sigma}, l, \hat{f}) \in \widetilde{L} * \widetilde{L}[E]$, the product concatenates the linear orders in the assembly a in the order specified by l. Note that all members of (a, f) are linear orders of the same length and color since $\hat{f}$ is a constant function. Also, the automorphisms σ and $\hat{\sigma}$ must be identity permutations. It is simple enough to generalize the product to $\widetilde{U}_0 * \widetilde{L} * \widetilde{L}$-structures. ■

The identity elements for c-monoids are easily described.

Lemma 3.1 *Let $(el(\widetilde{U}_0 * \widetilde{R}), c, I)$ be a c-monoid in $\mathcal{A}_*$. Then for any $(\sigma, a, f) \in \widetilde{U}_0 * \widetilde{R}[E]$,*

1. *$i_L(\sigma, a, f)$ is the unique element of $\widetilde{\gamma_{|E|}(R)}[E]$ with coloring f and automorphism σ.*

2. *$i_R(\sigma, a, f)$ is the unique element of $\widetilde{\gamma_{|\pi|}(R)}[\pi]$ with coloring $\hat{f}$ and automorphism $\hat{\sigma}$ where π is the associated partition of (σ, a, f), $\hat{f} : \pi \rightarrow X$ is the induced coloring of (σ, a, f), and $\hat{\sigma} : \pi \rightarrow \pi$ is the induced permutation of σ.*

A c-monoid $(el(\widetilde{U}_0 * \widetilde{R}), c, I)$ in $\mathcal{A}_*$ induces a partial order on assemblies of R-structures provided with automorphisms by setting

$$(\sigma, a_1, f) \leq_c (\sigma, a_3, f) \text{ iff there exists an } (\hat{\sigma}, a_2, \hat{f}) \text{ with } c((\sigma, a_1, f), (\hat{\sigma}, a_2, \hat{f})) = (\sigma, a_3, f).$$

Note that comparable elements use the same coloring and the same automorphism.

For any finite set E, coloring $f : E \rightarrow X$, and permutation $\sigma : E \rightarrow E$ with $f \circ \sigma = f$, let $P_R[\sigma, E, f]$ be the poset

$$P_R[\sigma, E, f] = \langle \{(\sigma, a, f) \in \widetilde{U}_0 * \widetilde{R}[E]\}, \leq_c \rangle$$

where $\leq_c$ is the restriction of $\langle el(\widetilde{U}_0 * \widetilde{R}), \leq_c \rangle$ to the elements of $P_R[\sigma, E, f]$.

The following properties will be crucial in defining a Möbius polynomial species $\widetilde{R}^{[-1]}$ with the desired properties.

Theorem 3.1 *The family of posets $\{P[\sigma,E,f]\}$ indexed by the triples (σ,E,f) where E is a finite set, $f:E\rightarrow X$, and $\sigma:E\rightarrow E$ is a permutation with $f\circ\sigma=f$ satisfy the following properties:*

1. *For any bijection $g:E\rightarrow F$ between finite sets, the map*
$$\begin{array}{rccc} P_R[g]: & P[\sigma,E,f] & \rightarrow & P[g\circ\sigma\circ g^{-1},F,f\circ g^{-1}] \\ & r & \mapsto & \widetilde{U}_0 * \widetilde{R}[g]r \end{array}$$
is an order isomorphism.

2. *$P[\sigma,E,f]$ has a $\hat{0}$ equal to the unique element of $\widetilde{\gamma_{|E|}(R)}[E]$ with coloring f and automorphism σ.*

3. *For any $p\in P[\sigma,E,f]$, the upper order ideal $I_p=\{p'\in P[\sigma,E,f]:p\leq_c p'\}$ is isomorphic to $P[\hat{\sigma},\pi,\hat{f}]$ where π is the associated partition of p, $\hat{f}$ is the associated coloring, and $\hat{\sigma}$ is the induced permutation.*

3.3 Plethystic Inverse of a Polynomial Species

Given an R-structure (m,f), let $\{(m,f)\}$ denote the assembly containing only the R-structure (m,f). For an $\widetilde{R}$-structure (σ,m,f), let $\{(\sigma,m,f)\}$ denote the assembly $\{(m,f)\}$ provided with the automorphism σ.

Definition 3.2 Let $(el(\widetilde{U}_0*\widetilde{R}),c,I)$ be a c-monoid in $\mathcal{A}_*$. The inverse Möbius polynomial species $\widetilde{R}^{[-1]}$ is defined as

$$\begin{array}{lcl} \widetilde{R}^{[-1]}[E] & = & \{([\hat{0},\{(\sigma,m,f)\}],f):(\sigma,m,f)\in\widetilde{R}[E],[\hat{0},\{(\sigma,m,f)\}]\subseteq P[\sigma,E,f]\} \\ \widetilde{R}^{[-1]}[g]([\hat{0},\{r\}],f) & = & ([\hat{0},\{\widetilde{R}[g]r\}],f\circ g^{-1}) \end{array}$$

where $g:E\rightarrow F$ is a bijection of finite sets. ∎

Theorem 3.2 *Let $(el(\widetilde{U}_0*\widetilde{R}),c,I)$ be a c-monoid in $\mathcal{A}_*$. Then*

$$Gen(\widetilde{R}^{[-1]})*Gen(\widetilde{R})=a_1(X).$$

Proof: Use Möbius inversion arguments and the previous theorem. ∎

3.4 Examples

Example 3.2 (power sum) Recall that $\widetilde{U}_0*\widetilde{L}$ has a c-monoid structure where the product is concatenation. Let $(\sigma,l)\in\widetilde{L}[n]$ where σ must be the trivial permutation and l is a monochromatic linear order. Any element of the interval $[\hat{0},\{(\sigma,l)\}]$ is of the form (σ,a) where the assembly a is "obtained" by "cutting" the linear order l into continuous segments of the same length. Therefore, the interval is isomorphic to the lattice of divisors of n, and the Möbius valuation $\mu(\hat{0},\{(\sigma,l)\})=\mu(n)$ where $\mu(n)$ is the classical Möbius function of number theory. Let $p_n(X)=\sum_i x_i^n$. Then

$$\mathrm{Gen}(\widetilde{L})=\sum_{n\geq1}p_n(X)\quad\text{and}\quad\mathrm{Gen}(\widetilde{L}^{[-1]})=\sum_{n\geq1}\mu(n)p_n(X).$$

∎

Example 3.3 (homogeneous) Let π be a partition of a set E, and let τ be a partition of π. The coinduced partition $ind_\pi(\tau)$ is a partition of E whose blocks are the subsets C of E of the form $C = \bigcup\{B : B \in D\}$ as D ranges over the blocks of τ. For example, if $\pi = \{B_1, B_2, B_3\}$ and $\tau = \{\{B_1, B_2\}, \{B_3\}\}$ where $B_1 = \{1,2\}$, $B_2 = \{3,4\}$, and $B_3 = \{5\}$, then $ind_\pi(\tau) = \{\{1,2,3,4\},\{5\}\}$.

$(el(\tilde{U}_0 * \tilde{U}_0), c, I)$ is a c-monoid in $\mathcal{A}_*$ where given $(\sigma, \pi, f, \hat{\sigma}, \tau, \hat{f}) \in \tilde{U}_0 * \tilde{U}_0 * \tilde{U}_0[E]$,

$$c(\sigma, \pi, f, \hat{\sigma}, \tau, \hat{f}) = (\sigma, ind_\pi(\tau), f).$$

When σ is the identity permutation, the interval $[\hat{0}, \{(\sigma, E, f)\}]$ obeys the usual ordering for partitions of E. Note that different automorphisms will result in different intervals. For example, $\sigma_1 = (1)(2)(3)$ and $\sigma_2 = (12)(3)$ are automorphisms of the U-structure $r = (\{1,2,3\}, f)$ where $f(1) = f(2) = x_1$, and $f(3) = x_2$. The interval $[\hat{0}, \{(\sigma_1, r)\}]$ is isomorphic to the lattice of partitions of $\{1,2,3\}$ while $[\hat{0}, \{(\sigma_2, r)\}]$ is isomorphic to a chain of length 2. ■

Example 3.4 (multicolored linear orders) Let R be the symmetric polynomial species of (nontrivial) linear orders colored in all ways. $el(\tilde{U}_0 * \tilde{R})$ has a c-monoid structure where the product is concatenation. Let $(\sigma, l) \in \tilde{R}[n]$ where σ can only be the trivial permutation and l is a colored linear order. Any element of the interval $[\hat{0}, \{(\sigma, l)\}]$ is of the form (σ, a) where the assembly a is obtained by "cutting" l into continuous segments. Since specifying when l is cut completely determines (σ, a), and since more cuts result in a structure lower in the partial order, $[\hat{0}, \{(\sigma, l)\}]$ is isomorphic to the Boolean poset B_{n-1} Therefore,

$$\mathrm{Gen}(\tilde{R}) = \sum_\lambda (-1)^{|\lambda|-1} \frac{|\lambda|!}{\prod \lambda_i!} m_\lambda(X)$$

where the sum ranges over all number partitions $\lambda = (\lambda_1, \lambda_2, \ldots)$ with $|\lambda| > 0$, and $m_\lambda(X)$ denotes the monomial symmetric function. ■

Example 3.5 (monochromatic sets) Let R be the symmetric polynomial species of (nontrivial) monochromatic sets. The induced copartition may be used to define a c-monoid structure.

Consider the interval $[\hat{0}, \{(\sigma, E, f)\}]$ where $(\sigma, E, f) \in \tilde{R}[E]$. Any element of this interval is a triple (σ, π, f) where π is a partition of E with blocks of the same size, f is a constant function, and σ is an automorphism of π. Since the generating function for $\tilde{R}$ equals the generating function for the power sum symmetric functions, this is another (more complicated) way to treat the plethystic inverse for the power sum. ■

Example 3.6 (rooted trees) Let A denote the Joyal species of rooted trees. Given $(a, t_1) \in A_0(A_0)[E]$, define a product $t = p(a, t_1)$ where $t \in A_0[E]$ contains all of the edges in a plus a few more defined as follows: for every pair of trees m_1 and m_2 in a such that the associated blocks form an edge of t_1, insert an edge between the roots of m_1 and m_2. The root of t will be the root of the tree in a whose associated block is the root of t_1.

The polynomial species $\widetilde{A_0 \times Hom}$ corresponds to unlabelled rooted trees colored in all ways; it has a c-monoid where given $(\sigma, a, f, \hat{\sigma}, t, \hat{f}) \in (\widetilde{A_0 \times Hom}) * (\widetilde{A_0 \times Hom})[E]$,

$$c(\sigma, a, f, \hat{\sigma}, t, \hat{f}) = (\sigma, p(a, t), f).$$

A similar c-monoid can be defined for monochromatic rooted trees by restricting the product to the appropriate monochromatic structures.

A *fiber* of a tree vertex is the set of the vertex's children. For a Joyal species $M : \mathbf{B} \to \mathbf{B}$, a tree is said to be M-enriched if each of its fibers is provided with an M-structure (not forgetting the empty fibers). A c-monoid structure in $\mathcal{A}_*$ can be constructed for many kinds of enriched (colored) trees such as trees enriched with linear orders (i.e., ordered trees), trees where each vertex has a multiple of k children (k a positive integer), partition-enriched trees, permutation-enriched trees, and even trees enriched by ordered trees.

For example, let A_L denote the species of ordered trees. An ordered tree is of the form $(T, \{l_x\}_{x\in E})$ where $T \in A[E]$ is a rooted tree and l_x is a linear order on the fiber of the vertex x. Consider an element

$$(a, T_L) = (\{(T_B, \{l_x\}_{x\in B}) : B \in \pi\}, (T_1, \{l_B\}_{B\in\pi})) \in A_L(A_L)[E]$$

where π is the associated partition of the forest of ordered trees, a. Define a product $p(a, T_B) = (T_2, \{l'_x\}_{x\in B})$ where T_2 is obtained by taking the products of the rooted forests $\{T_B : B \in \pi\}$ and $\{T_1\}$ as defined above. Suppose that x is the root of some tree T_B, the linear order $l_B = B_1B_2\cdots B_k$, and y_i is the root of tree T_{B_i}. Then l'_x equals the concatenation of l_x with $y_1\cdots y_k$. When x is not the root of some tree T_B, then $l'_x = l_x$.

Then as done earlier in this example, the product p can be used to define a c-monoid structure for ordered trees colored in all ways. ■

Example 3.7 (graphs) Let G be the Joyal species of simple connected graphs ($G[0] = \emptyset$). Define a product p of $(\{g_B : B \in \pi\}, g_1) \in G(G)[E]$ as the graph $g \in G[E]$ where $\{x, y\}$ is an edge of g if $\{x, y\}$ is an edge of some graph g_B, or the associated blocks of π containing x and y form an edge in g_1. Use the product p to define a c-monoid structure for $G \times \widetilde{Hom}$. ■

Note that in the last few examples, a product p on labelled structures was extended to a c-monoid for the appropriate colored structures. This is a powerful technique for generating examples; that is, products for labelled structures (which correspond to c-monoids for the compositional inverses of exponential generating functions as described in [10]) can be extended to c-monoids in $\mathcal{A}_*$.

Many more examples such as vertebrates (a tree with an ordered pair of not necessarily distinct vertices), bushes (rooted trees with no grandchildren), parenthesizations, and pointed simple connected graphs (a simple connected graph with a special vertex–similar to the root of a tree) fall into this framework.

References

[1] F. Bergeron, A Combinatorial Outlook on Symmetric Functions, *Journal of Combinatorial Theory, Series A*, **50** (1989) 226-234.

[2] F. Bonetti, G.-C. Rota, D. Senato, and A. Venezia, On the Foundations of Combinatorial Theory X: A Categorical Setting for Symmetric Functions, *Studies in Applied Math*, **86** (1992) 1-29.

[3] W. Chen, On the Combinatorics of Plethysm, Ph.D. Thesis, Masssachusetts Institute of Technology, June 1991.

[4] P. Doubilet, On the Foundations of Combinatorial Theory VII: Symmetric Functions through the Theory of Distribution and Occupancy, *Studies in Applied Math.*, **37** (1981) 193-202.

[5] A. Joyal, Une Theorie Combinatoire des Series Formelles, *Advan. in Math.*, **42** (1981) 1-82.

[6] A. Joyal, Foncteurs Analytiques et Espèces de Structeres, in G. Labelle and P. Leroux, eds., Combinatoire Enumérative, Proceedings, Montréal, Québec, 1985, Lecture Notes in Mathematics No. 1234 (Springer-Verlag, Berlin 1986) 126-159.

[7] G. Labelle, Some New Computational Methods in the Theory of Species, in G. Labelle and P. Leroux, eds., Combinatoire Enumérative, Proceedings, Montréal, Québec, 1985, Lecture Notes in Mathematics No. 1234 (Springer-Verlag, Berlin 1986) 192-209.

[8] G. Labelle, On the Generalized Iterates of Yeh's Combinatorial K-Species, *Journal of Combinatorial Theory, Series A*, **50** (1989) 235-258.

[9] I.G. MacDonald, "Symmetric Functions and Hall Polynomials," Oxford Univ. Press (Clarendon), London, 1979.

[10] M. Mendez and J. Yang, Möbius Species, *Advan. in Math.*, **85** (1991) 83-128.

[11] R. C. Read, Enumeration with Cycle Index Sums: First Steps towards a Simplified Method, unpublished manuscript.

[12] G.-C. Rota, On the Foundations of Combinatorial Theory I: Theory of Möbius Functions, *Z. Wahrscheinlichkeitstheorie,* **2** (1964) 340-368.

[13] D. Senato, A. Venezia, and J. Yang, Möbius Polynomial Species, unpublished manuscript.

[14] J. Yang, Symmetric Functions, Plethysm, and Enumeration, Ph.D. Thesis, Massachusetts Institute of Technology, June 1991.

Function Composition and Automatic Average case Analysis

Paul Zimmermann*
zimmermann@inria.inria.fr

Abstract

We define the composition of functions defined over extended context-free languages. We show that this composition is automatically computable. It enables the automatic analysis of complex problems with *small* input descriptions, for example repeated differentiation or iterated automata on regular languages.

Keywords: analysis of algorithms, average-case complexity, automatic analysis, program transformation.

1 Introduction

In the field of automatic complexity analysis, the length of the problem description is often a limitation: writing a long program is not only boring, but may introduce some errors. Thus we need some powerful constructs to describe algorithms, with the necessary constraint that these constructs allow an *automatic* analysis.

We are particularly interested here in the average case analysis of programs including some *compositions* of functions. To our knowledge, none of the existing systems, including METRIC [3], COMPLEXA [7] or Lambda-Upsilon-Omega [2], is able to analyze the composition of functions. The main reason may be the following: in these systems, the analysis of statements like $f(x)$ relies on the fact that all required data types are defined, either implicitly like in METRIC and COMPLEXA where all data structures are lists, or explicitly like in Lambda-Upsilon-Omega. But in the statement $f(g(y))$, the difficulty is to get a formal description of the object $g(y)$, which is not known *a priori.*

As an example, suppose we have written a function *diff* performing the differentiation of symbolic expressions with respect to one variable, and now we would like to analyze the two-fold differentiation by just defining

$$\mathit{diff2}(e) \stackrel{\text{def}}{=} \mathit{diff}(\mathit{diff}(e))$$

*Domaine de Voluceau, B.P. 105, 78153 Le Chesnay Cedex. This research was partially supported by the ESPRIT II Basic Research Actions Program of the EC under contract No. 3075 (project ALCOM).

instead of having to write the entire body of the function *diff2*. We show in this paper that this shorthand is possible. More precisely, we define a class of programs including function compositions, such that every program can be *automatically* expanded into another one without any composition, and equivalent to the original one in what concerns complexity analysis. This result allows to define and to analyze large problems by short description programs.

The paper contains two main sections: in Section 2, we define a class of programs including compositions in terms of the Lambda-Upsilon-Omega system language, and we state the main result (Theorem 1): these programs are automatically reducible to other equivalent programs, the latter lying in a class where automatic analysis methods are already known. In Section 3, we explain how the implementation of Theorem 1 helped us to guess a conjecture concerning iterated differentiation, and to obtain an interesting result about the Collatz conjecture; it is also shown that using composition may drastically reduce the size of the description program.

2 A class of programs with composition

In this section, we will introduce the composition of functions in a language called ADL (Algorithm Description Language), especially designed for automatic average case analysis in the Lambda-Upsilon-Omega system [1, 2]. Instead of giving a formal description of this language, we prefer to show as example an ADL program performing the differentiation of symbolic expressions.

An ADL program has three parts: the data type specification, the definition of one or more procedures, and the declaration of complexity measures. The data type specification looks like a formal language grammar; symbolic expressions constructed with $0, 1$, the variable x and the binary operators $+$ and $\times$ are defined for example as follows:

```
type expression = zero | one | x
                | plus(expression,expression)
                | times(expression,expression);
     plus,times,zero,one,x = atom(1);
```

where the last line defines $+, \times, 0, 1, x$ as atoms (or terminals) of *size* 1. The size is additive, thus the size of an expression as defined above is the number of atoms it contains. For example,

$$E = \mathtt{plus(times(x,x),one)}$$

is of type `expression` and of size 5. The function computing the derivative of such expressions with respect to x is written like this:

```
function diff(e : expression) : expression;
begin
   case e of
      plus(e1,e2)       : plus(diff(e1),diff(e2));
      times(e1,e2)      : plus(times(diff(e1),copy(e2)),
                                  times(copy(e1),diff(e2)));
      zero              : zero;
```

```
        one                : zero;
        x                  : one
    end;
end;
```

where the function copy, which simply makes a carbon copy of one expression, is also defined in the same manner. If we apply this function *diff* on the above expression E, we obtain

$$diff(E) = \texttt{plus(plus(times(one, x), times(x, one)), zero)}$$

(the *diff* function performs no simplification). Now if we want to define the cost of the function as the number of atoms in the output of *diff*, it suffices to define the cost of each atom as 1:

```
measure plus,times,zero,one,x : 1;
```

With this declaration, the cost of *diff*(E) is 9. If we analyze the *diff* function in the Lambda-Upsilon-Omega system, we will get the following average cost for expressions of size n:

$$\tau\overline{diff}_n = \frac{1}{4}\sqrt{2\pi}n^{3/2} + O(n).$$

More details about Lambda-Upsilon-Omega or ADL will be found in [2] or [5].

Now we allow the use of the composition in ADL programs, that is statements of the form $f(g(y))$ where f and g are two functions defined in the program, and y is a local variable. For example, the second order differentiation is defined as follows

```
function diff2(e : expression) : expression;
begin
    diff(diff(e))
end;
```

Definition 1 *The* composition graph *associated to an* ADL *program is the graph whose vertices are the function names, and for each composition $f(g(\ldots))$ in the body of a function h, there is an arrow from h to all functions on which f and g depend.*[1]

Theorem 1 *If the composition graph of an* ADL *program is acyclic, then the program translates into an equivalent program without composition.*

Proof: [Sketch] If the composition graph is acyclic, it is possible to totally order the functions, say $h_1, \ldots, h_k$, such that all arrows starting from a function go to functions of smaller index. Then we *expand* the body of the functions in increasing index order, that is we compute an equivalent body without any composition. Expanding h_1 is trivial because there is no composition in the body of h_1. Suppose we have already expanded $h_1, \ldots, h_{j-1}$. For each composition $f(g(\ldots))$ appearing in the body of h_j, the functions f and g are necessary of smaller index, therefore they have already been expanded.

Thus the only difficulty is to expand a call $f(g(x))$ where f and g have already been expanded. To do this, we replace this call by $k(x)$, where k is a new function name. We put as body for k the body of g where every returned expression y is replaced by $f(y)$. We

[1]The relation "depends on" is the reflexive and transitive closure of the relation "has in its body".

then simplify the expression $f(y)$ according to the body of f while it is possible. During this simplification, some compositions $f' \circ g'$ may appear. In that case either f' (same for g') has been created during the expansion process (thus is already expanded), or there is necessary an arrow from h_j to f', therefore f' has already been expanded. As the number of such new compositions that may appear is bounded (by the number of functions f' on which f depends by the number of functions g' on which g depends), the expansion of the body of h_j will eventually terminate. ∎

As the average case analysis of ADL programs *without* composition is already known to be automatic [2, 5], the above theorem implies directly the following result:

Corollary 1 *The average case analysis of* ADL *programs with an acyclic composition graph is possible automatically.*

An example of ADL program whose composition graph has a cycle is the following:

```
type integer = one | one integer;
      one = atom(1);

function f (i : integer) : integer;
begin
   case i of
      one : one;
      (one,j) : f(f(j))
   end;
end;
```

The expansion process described above will not terminate, though the function f simply returns one for all inputs.

3 Automatic analysis of programs with composition

In this section, we present two research problems where the implementation of the expansion process on a computer allowed us to discover some results which would have been very difficult to find at hand. We also prove an interesting result: the expansion process may produce exponentially large programs, with respect to the initial one.

3.1 Analysis of kth order differentiation

The expansion process described in the proof of Theorem 1 has been encoded in an experimental version (V1.4) of the system Lambda-Upsilon-Omega. When we analyze the function `diff2` as defined in Section 2, the system displays with "printlevel" 3 the expanded form of the function body:

```
function diff_of_diff (e : expression) : expression;
begin
   case e of
```

```
      (plus,(e1,e2)) : plus(diff_of_diff(e1),diff_of_diff(e2));
      (times,(e1,e2)) : plus(plus(times(diff_of_diff(e1),copy_of_copy(e2)),
                                  times(copy_of_diff(e1),diff_of_copy(e2))),
                             plus(times(diff_of_copy(e1),copy_of_diff(e2)),
                                  times(copy_of_copy(e1),diff_of_diff(e2))));
      zero : zero;
      one : zero;
      x : zero;
   end;
end;
```

Three other new functions have been introduced, namely `diff_of_copy`, `copy_of_diff` and `copy_of_copy` (the function copy is not initially known by the system). The system then proceeds in the usual way (Algebraic Analysis, Solver, Analytic Analysis) described in [2] and gives the final result:

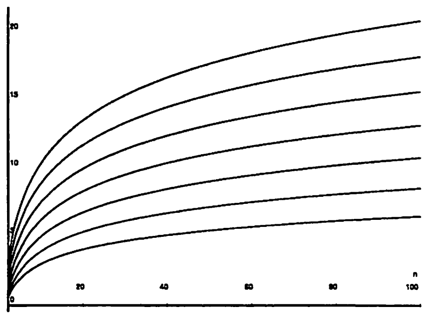
```
Average cost for diff2 on random inputs of size n is:

                                                 2          3/2
                                        (1/2 n ) + (O(n     ))

for n mod 2 = 1, and 0 otherwise.
```

In this way, by just adding one more call to the function `diff`, we were able to analyze the k-fold iterated differentiation until $k = 7$. To figure out the difficulty of the task, just try to write the body of the function `diff3` without any error! We obtained the following figures.

	average cost
`diff`	$\frac{1}{4}\sqrt{2\pi}n^{3/2} + O(n)$
`diff2`	$\frac{1}{2}n^2 + O(n^{3/2})$
`diff3`	$\frac{3}{16}\sqrt{2\pi}n^{5/2} + O(n^2)$
`diff4`	$\frac{1}{2}n^3 + O(n^{5/2})$
`diff5`	$\frac{15}{64}\sqrt{2\pi}n^{7/2} + O(n^3)$
`diff6`	$\frac{3}{4}n^4 + O(n^{7/2})$
`diff7`	$\frac{105}{256}\sqrt{2\pi}n^{9/2} + O(n^4)$



(The graph on the right hand side shows the logarithm of the complexity of the kth order derivative, for $1 \leq k \leq 7$). These figures gave us the idea to conjecture an average cost of

$$\frac{\Gamma(k/2+1)}{2^{k/2}} n^{k/2+1} + O(n^{(k+1)/2}) \tag{1}$$

for the kth order differentiation. The equation (1) is indeed the correct expansion. To prove this, we translate the body of diff_k into an equation for its cost generating function according to the rules given in [2][2]:

$$\begin{aligned}\tau\,\mathit{diff}_k(z) &= zE(z)^2 + 2zE(z)\tau\,\mathit{diff}_k(z)\\ &+ (2^{k+1}-1)zE(z)^2 + zE(z)\sum_{i=0}^{k}\binom{k}{i}(\tau\,\mathit{diff}_i(z) + \tau\,\mathit{diff}_{k-i}(z))\\ &+ 3z.\end{aligned}$$

The right hand side in the first line reflects the kth order derivative of a sum `plus(e1,e2)`: the term $zE(z)^2$ corresponds to the atom `plus`, and the other term to the recursive calls `diff_k(e1)` and `diff_k(e2)`. The kth order derivative of a product `times(e1,e2)` looks like a complete binary tree of height $k+1$, whose nodes of depth 0 to $k-1$ are `plus` atoms, whereas the level of depth k contains only `times` atoms, and the 2^{k+1} leaves are all kth compositions of either `diff` or `copy` on either `e1` or `e2`. The first term in the second line is the cost of writing all $2^{k+1}-1$ internal nodes; the second term is the cost of the recursive calls, with the convention that $\tau\,\mathit{diff}_0 = \tau\,\mathit{copy}$. From the above equation, we derive the following recurrence relation for $\tau\,\mathit{diff}_k$:

$$\tau\,\mathit{diff}_k(z) = \frac{3z + 2^{k+1}zE(z)^2 + 2zE(z)\sum_{i=0}^{k-1}\binom{k}{i}\tau\,\mathit{diff}_i(z)}{1-4zE(z)}. \tag{2}$$

To get an asymptotic expansion of the coefficient of z^n in $\tau\,\mathit{diff}_k(z)$, we first have to compute a local expansion around its main singularities, here the roots of $24z^2 = 1$. For $k \geq 1$, the dominant contribution comes from the term corresponding to $i = k-1$ in the sum of (2), thus we get a simple recurrence leading to

$$\tau\,\mathit{diff}_k(z) \sim \sqrt{6}\frac{k!}{2^{k+1}}\frac{1}{(1-24z^2)^{\frac{k+1}{2}}}.$$

We transfer this expansion into coefficients, using the relation[3] $[z^n](1-z)^{-\alpha} \sim n^{\alpha-1}/\Gamma(\alpha)$, and we divide by the asymptotic expansion of the number of expressions, to get the following average cost

$$\tau\,\overline{\mathit{diff}}_k(z) \sim \sqrt{\pi}\frac{k!}{2^{3k/2}\Gamma(\frac{k+1}{2})}n^{k/2+1}.$$

This expansion is in fact the same as (1) because $2^k\Gamma(k/2+1)\Gamma(k/2+1/2) = k!\sqrt{\pi}$. This first example shows how powerful the expansion process is: starting from a program with composition of length $O(k)$ (the kth order derivation), it produces a program without composition of length $\Omega(2^k)$!

---

[2] If $\mathcal{A}$ is a set of combinatorial structures, the (counting) generating function of $\mathcal{A}$ is $A(z) = \sum_{a\in\mathcal{A}} z^{|a|}$, where $|\cdot|$ denotes the size function. If P is a procedure taking inputs in $\mathcal{A}$, the (cost) generating function associated to P is $\tau P(z) = \sum_{a\in\mathcal{A}} \tau P\{a\} z^{|a|}$, where $\tau P\{a\}$ is the cost of the evaluation of P on a. Thus the average cost of P over inputs of size n is simply $\tau P_n/A_n$ where f_n denotes the coefficient of z^n in the Taylor expansion of f around $z = 0$.

[3] As usually, $[z^n]f(z)$ denotes the coefficient of z^n in the Taylor expansion of f around $z = 0$.

3.2 Regular languages and the Collatz conjecture

In this section, we show that function composition, used jointly with analysis of functions with a finite number of return values [6], helps to compute grammars of sets derived from regular languages. To illustrate this, we take as example the Collatz conjecture: "starting from a positive integer, the iteration of the function

$$f(n) = \begin{cases} n/2 & \text{if } n \text{ is even,} \\ 3n+1 & \text{if } n \text{ is odd,} \end{cases} \tag{3}$$

ultimately reaches 1". For example, we obtain the following chain for the number 13:

$$13 \to 40 \to 20 \to 10 \to 5 \to 16 \to 8 \to 4 \to 2 \to 1.$$

In [4], David Wilson introduced the sets S_k, where the index k denotes the number of times the function $3n+1$ is applied before 1 is reached. In the above example, the function $3n+1$ is applied two times (from 13 to 40 and from 5 to 16), thus 13 belongs to S_2. The first sets begin like this:

$$\begin{aligned} S_0 &= \{1,2,4,8,16,32,64,128,256,512,\ldots\} \\ S_1 &= \{5,10,20,21,40,42,80,84,85,160,\ldots\} \\ S_2 &= \{3,6,12,13,24,26,48,52,53,96,\ldots\} \\ S_3 &= \{17,34,35,68,69,70,75,136,138,140,\ldots\}. \end{aligned}$$

D. Wilson has shown that the base-two string expressions of any S_k form a regular language (accepted by a finite automaton and writable as a regular expression). For instance,

$$\begin{aligned} S_0 &\to 1\ 0^* \\ S_1 &\to 101\ (01)^*\ 0^*. \end{aligned}$$

This result implies that the number of n-bit integers in S_k is easily computable: it is the coefficient of z^n in a rational function derived from the regular expression of S_k, for example $z^3/(1-z^2)/(1-z)$ for S_1.

Function composition will enable us to compute a grammar for S_k *automatically*, with a description file of *linear* length with respect to k. From this grammar, we easily derive a regular expression. At the end of this section, we obtain a regular expression for S_2 and S_3.

Let us introduce the function g dividing its input by two as long as possible, then applying *one time* the function $3n+1$:

$$g(n) = \begin{cases} g(n/2) & \text{if } n \text{ is even,} \\ 0 & \text{if } n = 1, \\ 3n+1 & \text{otherwise.} \end{cases} \tag{4}$$

We have for instance $g(13) = 40$, $g(40) = 16$ and $g(16) = 0$, and the function g gives a characterization of S_k:

$$S_k = \{n \mid g^{(k)}(n) \text{ is a power of two }\}. \tag{5}$$

Therefore to construct an ADL program recognizing integers in S_k, we have to encode the function g, and a function recognizing powers of two. For this purpose, we represent integers in base two:

```
type integer = nil | bit integer;
     bit = zero | one;
     zero, one = atom(1);
     nil = atom(0);
```

The function g is written using a function called `three_x_plus_1`, whose input is the base-two representation of an integer n, and which outputs the base-two representation of $3n + 1$:

```
function three_x_plus_1 (i : integer) : integer;
begin
   case i of
      nil : product(one,nil);
      (zero,j) : product(one,three_x(j));
      (one,j) : product(zero,three_x_plus_2(j));
   end;
end;
```

The other functions `three_x` and `three_x_plus_2` are defined similarly. With the functions g and `is_a_power_of_two`, according to equation (5), we write the function `is_in_S3` to recognize integers in S_3:

```
function g (i : integer) : integer;
begin
   case i of
      nil : nil;
      (zero,j) : g(j);
      (one,nil) : nil;
      otherwise : three_x_plus_1(i);
   end
end;

function is_a_power_of_two (i : integer) : boolean;
begin
   case i of
      nil : false;                          % 0 is not a power of 2 %
      (zero,j) : is_a_power_of_two(j);
      (one,j) : is_zero(j)
   end;
end;
```

```
function is_in_S3 (i : integer) : boolean;
begin
   is_a_power_of_two(g(g(g(i))))
end;
```

With these functions, we could compute automatically the probability for an integer of n bits to be in S_3, with a procedure like this:

```
procedure main (i : integer);
begin
   if is_in_S3(i) then count
end;
measure count : 1;
```

But our goal is to get a regular expression for S_k like those obtained by Wilson for S_0 and S_1. During the analysis of the procedure `main`, the system prints some messages, for example (among other lines):

```
Computing composition of is_a_power_of_two and g : f2
Computing composition of f2 and g : f8
Computing composition of f8 and g : f26
Introducing the new type T84 for which function f26 returns true
Introducing the new type T142 for which function f26 returns false
```

At the first line, the system computes the body of the composition of `is_a_power_of_two` and g, which is the new function `f2`. The message `Computing composition of f and g` means that the expansion process of theorem 1 is currently being applied. Thus the body of `f2` contains no composition, and looks like the body of `diff_of_diff` in section 3.1. Then it computes the composition of `f2` and g, and calls it `f8` (second line). It also computes the composition of `f8` with g, namely `f26`, which is therefore equivalent to `is_in_S3`.

At this stage, we have constructed a set of ADL functions without any composition, containing the function `f26` equivalent to `is_in_S3`. For such a set, it is possible to derive automatically a grammar of the data structures for which each function with a finite number of possible outputs (in particular a boolean function like `f26`) returns a given value [6]. For example, as explained by the last lines in the above messages, the system introduced two new data types `T84` and `T142`, which stand for the integers in S_3 and not in S_3 respectively. Like for the expansion process, a complete grammar for `T84` and `T142` was in fact generated, starting from the grammar of the type `integer`.

Due to the form of the rules used (cf [6]), this grammar is unambiguous because so was the grammar of `integer`. The raw grammar we get has 58 non-terminals, among them 27 do not derive any finite string. After some simplifications by hand (they took longer than the automatic construction of the grammar!), we got the following regular expression for S_3:

$$S_3 \rightarrow ((\epsilon \mid (10010111101101000 0)^* 1001011\ (\epsilon \mid 1 \mid 1101 \mid 110110011 \mid 11011010000 \mid 1101101000011))\ (100011)^* 1000 \mid (100101111011010000)^* 100101\ (\epsilon \mid 11101100 \mid 1110110100000))(\epsilon \mid 1)\ (10)^* 10^*.$$

Similarly, we computed with the help of the Lambda-Upsilon-Omega system the following regular expression of the set S_2, starting from a grammar with 22 non-terminals:

$$S2 \rightarrow (1 \mid 11100\ (011100)^*\ (0 \mid 01))\ (10)^*\ 1\ 0^*$$

4 Conclusion

We have shown that some kinds of function compositions are well suited for an automatic average case analysis. The main idea is the following: a program including compositions first translates into an similar program without composition (*expansion process* of Theorem 1), then this last program is analyzed by already known techniques [2].

Composition of functions is not only useful in the description of algorithms, but in some cases it is *necessary* to use it, otherwise the description would be too long, as the examples of Section 3 prove it. In these cases, the long description is generated by the computer, therefore it contains no error (we hope it!).

References

[1] P. Flajolet, B. Salvy, and P. Zimmermann. Lambda–Upsilon–Omega: The 1989 Cookbook. Rapport de recherche 1073, Institut National de Recherche en Informatique et en Automatique, August 1989. 116 pages.

[2] P. Flajolet, B. Salvy, and P. Zimmermann. Automatic Average–case Analysis of Algorithms. *Theoretical Computer Science*, 79(1):37–109, February 1991.

[3] B. Wegbreit. Mechanical Program Analysis. *Communications of the ACM*, 18(9):528–539, 1975.

[4] D. W. Wilson. Transaction ⟨1778@cvbnetPrime.COM⟩ of usenet.sci.math, 1991.

[5] P. Zimmermann. *Séries génératrices et analyse automatique d'algorithmes.* Thèse de doctorat, École Polytechnique, Palaiseau, 1991.

[6] P. Zimmermann. Analysis of functions with a finite number of return values. Research Report 1625, Institut National de Recherche en Informatique et en Automatique, 1992.

[7] W. Zimmermann. *Automatische Komplexitätsanalyse funktionaler Programme.* PhD thesis, Fakultät für Informatik der Universität Karlsruhe, June 1990. Also available in the collection *Informatik Fachberichte*, number 261, Springer Verlag.

Liste des arbitres

M. Barr
R. Bédard
J. Bétréma
S. Brlek
G. Butler
C. Choffrut
H. Crapo
C. Cummins
M. Delest
C. Delorme
G. Duchamp
S. Dulucq
J.-M. Fedou
P. Flajolet
W. F. de la Vega
D. Foata
D. Ford
P. Y. Gaillard
D. Goyou-Beauchamps
L. Habsieger
M. Haiman
G. Hahn
P. Hanlon
D. Jackson
G. Jacob
M. Jambu
D. Krob
J. Labelle
C. Lam
F. Lamnabhi-Lagarrigue
M. Lassalle
P. Leroux
D. Loeb
G. Melançon
J. Néraud
M. Okada
J. Opatrny
J.-G. Penaud
C. Reutenauer
D. Robbins
G.-C. Rota
B. Sagan
R. Simion
M. Soria
R. Stanley
D. Stanton
J. Stembridge
J.-M. Steyaert
V. Strehl
L. Tao
D. Thérien
J.-Y. Thibon
T. Walsh
J. West
D. Zeilberger
P. Zimmerman
A. K. Zvonkin