Schur functions and alternating sums

Marc A. A. van Leeuwen

Université de Poitiers, Département de Mathématiques,
UFR Sciences SP2MI, Téléport 2, BP 30179, 86962 Futuroscope Chasseneuil Cedex, France
Marc.van-Leeuwen@math.univ-poitiers.fr
http://www-math.univ-poitiers.fr/~maavl/

Dedicated to Richard Stanley on the occasion of his 60^{th} ***birthday***

Submitted: Apr 18, 2005; Accepted: Feb 13, 2006; Published: Feb 22, 2006
Mathematics Subject Classifications: 05E05, 05E10

Abstract

We discuss several well known results about Schur functions that can be proved using cancellations in alternating summations; notably we shall discuss the Pieri and Murnaghan-Nakayama rules, the Jacobi-Trudi identity and its dual (Von Nägelsbach-Kostka) identity, their proofs using the correspondence with lattice paths of Gessel and Viennot, and finally the Littlewood-Richardson rule. Our our goal is to show that the mentioned statements are closely related, and can be proved using variations of the same basic technique. We also want to emphasise the central part that is played by matrices over $\{0,1\}$ and over $\mathbf{N}$; we show that the Littlewood-Richardson rule as generalised by Zelevinsky has elegant formulations using either type of matrix, and that in both cases it can be obtained by two successive reductions from a large signed enumeration of such matrices, where the sign depends only on the row and column sums of the matrix.

§0. Introduction.

Many of the more interesting combinatorial results and correspondences in the basic theory of symmetric functions involve Schur functions, or more or less equivalently the notions of semistandard tableaux or horizontal strips. Yet the introduction of these notions, in any of the many possible ways, is not very natural when considering only symmetric functions, cf. [Stan, 7.10]. One way the importance of Schur functions can be motivated is by representation theory: interpreting symmetric functions in the representation theory either of the symmetric groups or the general linear groups, the Schur functions correspond to the irreducible representations. However, there is another way of motivating it: if one broadens the scope slightly from symmetric functions to alternating polynomials, then Schur functions do arise quite naturally as quotients of alternants. It is this point of view, which could also be reached from representation theory if use is made only of Weyl's character formula, that we shall take in this paper; from this perspective it is not so surprising that proofs of basic identities involving Schur functions should involve alternating summations and cancellations.

The main point we would like to make in this paper is that the use of the definition of Schur functions as quotients of alternants can be limited to the deduction of a single simple formula (lemma 2.2), which describes the multiplication by an arbitrary symmetric function in the basis of Schur functions; after this, alternating polynomials need not be considered any more. In general the formula produces an alternating sum of Schur functions; in various particular cases, one can obtain classical results from it (the Pieri and Murnaghan-Nakayama rules, Jacobi and Von Nägelsbach-Kostka identities, and the Littlewood-Richardson rule) by judiciously applying combinatorially defined cancellations. Our presentation is nearly self-contained, but we do use an enumerative identity that follows from the RSK-correspondence; we shall omit its well known elementary combinatorial proof, which is not directly related to the theme of this paper.

Our paper is structured as follows. In §1 we give the basic definitions concerning symmetric functions, alternating polynomials and (skew) Schur functions. In §2 we introduce our basic lemma, and its most elementary applications giving the Pieri and Murnaghan-Nakayama rules. In §3 we first establish the duality of the bases of complete and minimal symmetric functions (this is where the RSK-correspondence is used). This allows us to interpret (skew) Schur functions as generating series of semistandard tableaux (which elsewhere is often used as their definition), and to deduce the Cauchy, Jacobi and Von Nägelsbach-Kostka identities. In §4 we discuss cancellations defined for intersecting families of lattice paths, in the style of Gessel and Viennot, and relate them to identities derived from the Pieri rules. These considerations lead to natural encodings of families of lattice paths, and of the semistandard tableaux that correspond to non-intersecting families of paths, by matrices with entries in $\mathbf{N}$ or in $\{0,1\}$; these encodings are also important in the sequel. In §5 we give a final application of our basic lemma to derive the Littlewood-Richardson rule. Formulating (Zelevinsky's generalisation of) that rule in terms of binary or integral matrices reveals in both cases an unexpected symmetry. We also exhibit an equally symmetrical doubly alternating expressions for the same numbers, in which no tableaux appear at all. We close by raising a question inspired by these expressions, which will be taken up in a sequel to this paper.

§1. Preliminaries and definitions.

Studying symmetric functions involves the use of various combinatorial objects; we start with some general considerations concerning those. We shall make much use of sequences (vectors) and matrices, of which the entries will almost always be natural numbers. In some cases the entries are restricted to be either 0 or 1, in which case we shall refer to the objects as "binary". While all objects we shall encounter can be specified using finite information, we shall consider vectors and matrices as associations of entries to indices, without restricting those indices to a finite set (just like for polynomials one usually does not give an a priori bound for the degrees of their monomials). Thus vectors and matrices are "finitely supported", in that the entries are zero outside a finite range of indices; finite vectors and matrices are identified with infinite ones obtained by extension with null entries. This convention notably allows addition of vectors or matrices without concern about their sizes.

When displaying matrices we shall as usual let the first index increase downwards and the second to the right, and the same convention will be used whenever subsets of $\mathbf{N} \times \mathbf{N}$ are displayed, such as Young diagrams (in the sequel to this paper we shall in fact encounter Young diagrams in the role of subsets of indices in matrices). Some objects, notably tableaux, are defined as sequences of vectors; in this case the indices for the sequence are written as parenthesised superscripts to avoid confusion with the subscripts indexing individual vectors.

We always start indexing at 0, in particular this applies to sequences, rows and columns of matrices and tableaux, and entries of tableaux. Hence in the situation where a sequence of objects is determined by the intervals between members of another sequence (such as horizontal strips in a semistandard tableau, which are given by successive members of a sequence of shapes), the index used for an interval is the same as that of the *first* of the members bounding it. Our standard n-element set is $[n] = \{\, i \in \mathbf{N} \mid i < n \,\}$. For the set theoretic difference $S \setminus T$ we shall write $S - T$ when it is known that $T \subseteq S$.

We shall frequently use the "Iverson symbol": for any Boolean expression *condition* one puts

$$[\, condition \,] = \begin{cases} 1 & \text{if } condition \text{ is satisfied,} \\ 0 & \text{otherwise.} \end{cases}$$

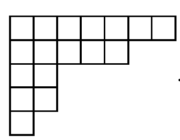
This notation, proposed in [GKP, p. 24], and taken from the programming language APL by K. Iverson, generalises the Kronecker delta symbol: instead of $\delta_{i,j}$ one can write $[\, i = j \,]$. Among other uses, this notation allows us to avoid putting complicated conditions below summations to restrict their range: it suffices to multiply their summands by one or more instances of $[\, condition \,]$. By convention, in a product containing such a factor, the factors to its right are evaluated only if the condition holds; if it fails, the product is considered to be 0 even if some remaining factor should be undefined.

1.1. *Compositions and partitions.*

The most basic combinatorial objects we shall use are finitely supported sequences of natural numbers $\alpha = (\alpha_i)_{i \in \mathbf{N}}$. The entries α_i are called the parts of α, and the main statistic on such sequences is the sum of the parts, written $|\alpha| = \sum_{i\in\mathbf{N}} \alpha_i$. The systematic name for such sequences α with $|\alpha| = d$ would be infinite weak compositions of d, but we shall simply call them just *compositions* of d. The set of compositions of d will be denoted by $\mathcal{C}_d$ (this set is infinite when $d > 0$), and $\mathcal{C} = \bigcup_{d\in\mathbf{N}} \mathcal{C}_d$ denotes the set of all compositions. In order to denote specific compositions, we shall specify an initial sequence of their parts, which are implicitly extended by zeroes. When the parts of a composition are restricted to lie in $\{0,1\} = [2]$, it will be called a *binary composition*; we define $\mathcal{C}^{[2]}_d = \{\, \alpha \in \mathcal{C}_d \mid \forall i \in \mathbf{N}\colon \alpha_i \in [2] \,\}$ and $\mathcal{C}^{[2]} = \bigcup_{d\in\mathbf{N}} \mathcal{C}^{[2]}_d$. Binary compositions of d correspond to d-element subsets of $\mathbf{N}$, while arbitrary compositions of d correspond to multisets of size d on $\mathbf{N}$. Among other uses, compositions parametrise monomials; if $X_{\mathbf{N}} = \{\, X_i \mid i \in \mathbf{N} \,\}$ is a countable set of commuting indeterminates, then the monomial $\prod_{i\in\mathbf{N}} X_i^{\alpha_i}$ will be denoted by X^α.

We shall consider permutations of indeterminates, and correspondingly of the parts of compositions. The group that acts is the group $\mathbf{S}_\infty$ of permutations of $\mathbf{N}$ that fix all but finitely many numbers. The permutation $\sigma \in \mathbf{S}_\infty$ acts by simultaneously substituting $X_i := X_{\sigma(i)}$ for all indeterminates, and therefore operates on compositions by permuting their parts: $\sigma(\alpha) = (\alpha_{\sigma^{-1}(i)})_{i\in\mathbf{N}}$. Obviously $|\sigma(\alpha)| = |\alpha|$, and the orbit of α contains a unique composition whose parts are weakly decreasing, which will be denoted by α^+; for instance for $\alpha = (0,5,2,0,0,1,7,0,2)$ one has $\alpha^+ = (7,5,2,2,1)$. For $d \in \mathbf{N}$ we define the finite set $\mathcal{P}_d = \{\, \lambda \in \mathcal{C}_d \mid \forall i \in \mathbf{N}\colon \lambda_i \geq \lambda_{i+1} \,\}$, whose elements are called *partitions* of d; then $\alpha^+ \in \mathcal{P}_d$ for any $\alpha \in \mathcal{C}_d$. We also put $\mathcal{P} = \bigcup_{d\in\mathbf{N}} \mathcal{P}_d$. All binary compositions of d form a single orbit under permutations of their parts, so there is just a single binary partition of d: it is the partition $[\, i \in [d] \,]_{\, i\in\mathbf{N}}$ whose d initial parts are 1 and the rest 0, and we shall denote it by $1^{(d)}$.

We shall usually denote compositions by Greek letters $\alpha, \beta, \ldots$, but for partitions we use Greek letters further on in the alphabet: λ, μ, ν, and sometimes κ. Apart from listing its nonzero parts, a partition $\lambda \in \mathcal{P}$ can also be specified by drawing its *diagram* $[\lambda] = \{\, (i,j) \in \mathbf{N}^2 \mid j \in [\lambda_i] \,\}$. Elements of the diagram are drawn (and usually referred to) as squares, so that for instance the diagram of $\lambda = (7,5,2,2,1)$ would be drawn as



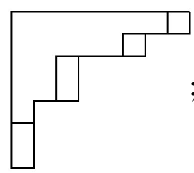
The transpose partition of $\lambda \in \mathcal{P}$, which will be denoted by λ^t, is the one whose parts give the lengths of the columns of $[\lambda]$, so that $[\lambda^t]$ is the transpose diagram $[\lambda]^t$; one has $\lambda^t_j = \#\{\, i \in \mathbf{N} \mid j \in [\lambda_i] \,\}$.

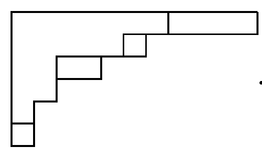
We shall be considering several relations defined between partitions; we collect their definitions here. The most fundamental relation is the partial ordering '$\subseteq$' defined by inclusion of diagrams: $\mu \subseteq \lambda$ means that $[\mu] \subseteq [\lambda]$ or equivalently that $\mu_i \leq \lambda_i$ for all $i \in \mathbf{N}$. Note that if $\mu \subseteq \lambda$ then $\lambda - \mu$ and $\lambda^t - \mu^t$ are compositions. The relation '$\subseteq$' will be used mostly implicitly via the notion of a *skew shape* λ/μ, which denotes the interval from μ to λ in the poset $(\mathcal{P}, \subseteq)$; the corresponding skew diagram is $[\lambda/\mu] = [\lambda] - [\mu]$, and we define $|\lambda/\mu| = |\lambda| - |\mu|$. Several relations refining '$\subseteq$' will be used; for the ones in the following definition it will be convenient to define them on the set of all compositions, although they will never hold unless both arguments are actually partitions.

1.1.1. Definition. *The relations '$\leftharpoonup$' and '$\leftharpoondown$' on $\mathcal{C}$ are defined as follows. To have either $\mu \leftharpoonup \lambda$ or $\mu \leftharpoondown \lambda$, it is necessary that λ/μ be a skew shape (in other words $\lambda, \mu \in \mathcal{P}$, and $\mu \subseteq \lambda$). If this is the case, then $\mu \leftharpoonup \lambda$ holds if and only if $\lambda - \mu \in \mathcal{C}^{[2]}$, in which case λ/μ is called a vertical strip; similarly $\mu \leftharpoondown \lambda$ holds if and only if $\lambda_{i+1} \leq \mu_i \leq \lambda_i$ for all $i \in \mathbf{N}$, in which case λ/μ is called a horizontal strip.*

Note that the final condition for $\mu \leftharpoondown \lambda$ already implies that λ/μ is a skew shape; in addition it means that $[\lambda/\mu]$ has at most one square in any column. Similarly, for a skew shape λ/μ, the condition $\mu \leftharpoonup \lambda$ means that $[\lambda/\mu]$ has at most one square in any row. Therefore $\mu \leftharpoondown \lambda$ is equivalent to $\mu^t \leftharpoonup \lambda^t$ when $\lambda, \mu \in \mathcal{P}$. To denote the opposite relations we shall rotate rather than reflect the symbol, so $\lambda \rightharpoonup \mu$ means the same as $\mu \leftharpoondown \lambda$, while $\lambda \rightharpoondown \mu$ means the same as $\mu \leftharpoonup \lambda$.

We illustrate concrete instances of these relations graphically by superimposing the contours of the diagrams of the two partitions involved:

$(7,5,2,2,1) \leftharpoonup (8,6,3,3,1,1,1)$:  ;

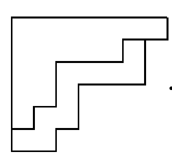
$(7,5,2,2,1) \leftharpoondown (11,6,4,2,1,1)$:  .

For the following definition we use the partitioning of $\mathbf{N} \times \mathbf{N}$ into *diagonals* D_d, for $d \in \mathbf{Z}$:

$$D_d = \{\, (i,j) \in \mathbf{N}^2 \mid j - i = d \,\} \qquad \text{for } d \in \mathbf{Z}. \tag{1}$$

1.1.2. Definition. *For $k > 0$, a relation '$\prec_{\mathrm{r}(k)}$' on $\mathcal{P}$ is defined as follows: $\mu \prec_{\mathrm{r}(k)} \lambda$ means that λ/μ is a skew shape with $|\lambda/\mu| = k$, for which the k squares of $[\lambda/\mu]$ lie on k consecutive diagonals. In this case we call the shape λ/μ a k-ribbon. The height $\mathrm{ht}(\lambda/\mu)$ of a k-ribbon λ/μ is the difference between the initial (row) coordinates of the squares of $[\lambda/\mu]$ on the first and the last of those k diagonals.*

Again we give a graphic illustration in the same style as before, for $\lambda/\mu = (7,6,6,3,3,2)/(7,5,2,2,1)$:

$(7,5,2,2,1) \prec_{\mathrm{r}(10)} (7,6,6,3,3,2)$:  .

One has $[\lambda/\mu] = \{(5,0),(5,1),(4,1),(4,2),(3,2),(2,2),(2,3),(2,4),(2,5),(1,5)\}$, which diagram has its squares on the 10 consecutive diagonals D_d for $-5 \leq d < 5$; moreover, we see that $\mathrm{ht}(\lambda/\mu) = 5 - 1 = 4$.

Finally we shall need the dominance partial ordering on each set $\mathcal{P}_d$ separately.

1.1.3. Definition. *For any fixed $d \in \mathbf{N}$ a relation '$\leq$' on $\mathcal{P}_d$, called the dominance ordering, is defined by $\mu \leq \lambda$ if and only if for every $k \in \mathbf{N}$ one has $\sum_{i\in[k]} \mu_i \leq \sum_{i\in[k]} \lambda_i$.*

1.2. *Matrices and tableaux.*

We shall use the two-dimensional counterparts of compositions: finitely supported matrices with entries in $\mathbf{N}$. Like for compositions the binary case, where entries are restricted to $[2] = \{0, 1\}$, will be of special interest. The statistic given by the sum of all entries can be refined by taking sums separately either of rows or of columns; in either case the result is a composition.

1.2.1. Definition. *Let $\mathcal{M}$ denote the set of matrices $(M_{i,j})_{i,j\in\mathbf{N}}$ with entries $M_{i,j}$ in $\mathbf{N}$, of which only finitely many are nonzero, and let $\mathcal{M}^{[2]}$ denote its subset of binary matrices, those of which all entries lie in $\{0, 1\}$. Let $\mathrm{row}\colon \mathcal{M} \to \mathcal{C}$ be the map $M \mapsto (\sum_{j\in\mathbf{N}} M_{i,j})_{i\in\mathbf{N}}$ that takes row sums, and $\mathrm{col}\colon \mathcal{M} \to \mathcal{C}$ the map $M \mapsto (\sum_{i\in\mathbf{N}} M_{i,j})_{j\in\mathbf{N}}$ that takes column sums; put $\mathcal{M}_{\alpha,\beta} = \{\, M \in \mathcal{M} \mid \mathrm{row}(M) = \alpha, \mathrm{col}(M) = \beta \,\}$ and $\mathcal{M}^{[2]}_{\alpha,\beta} = \mathcal{M}_{\alpha,\beta} \cap \mathcal{M}^{[2]}$ for $\alpha, \beta \in \mathcal{C}$.*

These matrices can be used to record sequences of (binary) compositions with finite support, either by rows or by columns. We shall denote row i of M by M_i, and column j by M^t_j. We shall also need sequences of partitions, but these will be subject to the condition that adjacent terms differ by horizontal or vertical strips, and the condition of finite support is replaced by the sequence becoming ultimately stationary. This gives rise to the notion of semistandard tableau, and some variants of it.

1.2.2. Definition. *Let λ/μ be a skew shape, and $\alpha \in \mathcal{C}$. A semistandard tableau of shape λ/μ and weight α is a sequence of partitions $(\lambda^{(i)})_{i\in\mathbf{N}}$ with $\lambda^{(i)} \leftharpoondown \lambda^{(i+1)}$ and $|\lambda^{(i+1)}/\lambda^{(i)}| = \alpha_i$ for all $i \in \mathbf{N}$, $\lambda^{(0)} = \mu$, and $\lambda^{(N)} = \lambda$ for any N that is so large that $\alpha_i = 0$ for all $i \geq N$. The weight of a tableau T is denoted by $\mathrm{wt}(T)$, and the set of all semistandard tableaux of shape λ/μ by $\mathrm{SST}(\lambda/\mu)$; we also put $\mathrm{SST}(\lambda/\mu, \alpha) = \{\, T \in \mathrm{SST}(\lambda/\mu) \mid \mathrm{wt}(T) = \alpha \,\}$. A transpose semistandard tableau of shape λ/μ and weight α is a sequence of partitions defined similarly, with $\lambda^{(i)} \leftharpoonup \lambda^{(i+1)}$ replacing $\lambda^{(i)} \leftharpoondown \lambda^{(i+1)}$.*

We shall reserve the qualification "Young tableau" to the case $\mu = (0)$, in which case λ/μ will be abbreviated to λ in the notations just introduced. There are maps from semistandard tableaux to transpose semistandard tableau and vice versa, defined by transposing each partition in the sequence; under these maps the shape of the tableau is transposed while the weight is preserved. Another variation on the notion of semistandard tableau is to replace the relations $\lambda^{(i)} \leftharpoondown \lambda^{(i+1)}$ or $\lambda^{(i)} \leftharpoonup \lambda^{(i+1)}$ by their opposite relations $\lambda^{(i)} \rightharpoonup \lambda^{(i+1)}$ respectively $\lambda^{(i)} \rightharpoondown \lambda^{(i+1)}$. This gives the notions of reverse (transpose) semistandard tableaux, which will occur in the sequel to this paper; their shape λ/μ and weight α are such that the sequence starts at $\lambda = \lambda^{(0)}$ and ultimately becomes μ, while $|\lambda^{(i)}/\lambda^{(i+1)}| = \alpha_i$ for all $i \in \mathbf{N}$.

The traditional way to display a semistandard tableau is to draw the diagram of its shape filled with numbers, which identify for each square the horizontal strip to which it belongs. We shall label with an entry i the squares of $[\lambda^{(i+1)}/\lambda^{(i)}]$. The entries will then increase weakly along rows, and increase strictly down columns, and for this reason semistandard tableaux are also called column-strict tableaux (and transpose semistandard tableaux are then called row-strict tableaux). Thus the semistandard tableau $T = \big((4,1) \leftharpoondown (5,2) \leftharpoondown (5,3,2) \leftharpoondown (6,3,3,1) \leftharpoondown (6,4,3,2) \leftharpoondown (7,5,4,3) \leftharpoondown (9,5,5,3,1) \leftharpoondown (9,8,5,5,3) \leftharpoondown (9,8,5,5,3) \leftharpoondown \cdots\big)$, which is of shape $(9,8,5,5,3)/(4,1)$ and weight $(2,3,3,2,4,4,7)$, will be displayed as

$$T = \begin{array}{ccccccccc} & & & & 0 & 2 & 4 & 5 & 5 \\ & 0 & 1 & 3 & 4 & 6 & 6 & 6 & \\ 1 & 1 & 2 & 4 & 5 & & & & \\ 2 & 3 & 4 & 6 & 6 & & & & \\ 5 & 6 & 6 & & & & & & \end{array} . \tag{2}$$

More important in our paper than this display will be two ways of representing tableaux by matrices. Simply recording the partitions forming a tableau T in the rows or columns of a matrix does not give a finitely supported matrix, but we can obtain one by recording the differences between successive partitions. We shall call the matrix so obtained an *encoding* of T, but one should realise that decoding the matrix to reconstruct T requires knowledge of at least one of the partitions forming the (skew) shape of T. Various ways are possible to record horizontal strips $\lambda^{(i+1)}/\lambda^{(i)}$: one may either record the differences $\lambda^{(i+1)} - \lambda^{(i)} \in \mathcal{C}$ or the differences between the transpose shapes $(\lambda^{(i+1)})^t - (\lambda^{(i)})^t \in \mathcal{C}^{[2]}$, and one may record these compositions either in the rows or the columns of the matrix. From the four possible combinations we choose the two for which one has a correspondence either between the rows of the tableau and the rows of the matrix, or between the columns of the tableau and the columns of the matrix.

1.2.3. Definition. *Let $T = (\lambda^{(i)})_{i\in\mathbf{N}}$ be a semistandard tableau. The integral encoding of T is the matrix $M \in \mathcal{M}$ defined by $M_{i,j} = (\lambda^{(j+1)} - \lambda^{(j)})_i$, and the binary encoding of T is the matrix $M' \in \mathcal{M}^{[2]}$ defined by $M'_{i,j} = ((\lambda^{(i+1)})^t - (\lambda^{(i)})^t)_j$. The sets of integral and binary encodings of semistandard tableaux $T \in \mathrm{SST}(\lambda/\mu)$ will be denoted by $\mathrm{Tabl}(\lambda/\nu)$ and $\mathrm{Tabl}^{[2]}(\lambda/\mu)$, respectively.*

For instance for the tableau T of (2), one finds the integral and binary encodings

$$M = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 2 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 3 \\ 0 & 2 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 & 1 & 2 \end{pmatrix} \quad \text{and} \quad M' = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \end{pmatrix}, \tag{3}$$

which finite matrices must be thought of as extended indefinitely by zeroes. To reconstruct from either of these matrices the tableau T or the other matrix, one must in addition know at least that $\mu = (4,1)$ or that $\lambda = (9,8,5,5,3)$ for the shape λ/μ of T. Each entry $M_{i,j}$ counts the number of entries j in row i of the displayed form of T, while entry $M'_{i,j}$ counts the number (at most one) of entries i in column j. Therefore the row M_i records the weight of row i of the display of T, while the column $(M')^t_j$ records the weight of its column j. One has $\mathrm{row}(M) = \lambda - \mu$, $\mathrm{col}(M') = \lambda^t - \mu^t$, and $\mathrm{col}(M) = \mathrm{row}(M') = \mathrm{wt}(T)$.

1.3. *Symmetric functions.*

There are several equivalent ways to define the ring Λ of symmetric functions. Following [Stan], we shall realise Λ as a subring of the ring $\mathbf{Z}[[X_\mathbf{N}]]$ of power series in infinitely many indeterminates. The elements f of this subring are characterised by the fact that the coefficients in f of monomials X^α, X^β are the same whenever $\alpha^+ = \beta^+$ (so f is stable under the action of $\mathbf{S}_\infty$), and that the degree of monomials with nonzero coefficients in f is bounded. Elements $f \in \Lambda$ are called *symmetric functions*. Since the indicated subring of $\mathbf{Z}[[X_\mathbf{N}]]$ is just one realisation of Λ, we make a notational distinction between occurrences of a symmetric function f that are independent of any realisation of Λ (for instance in identities internal to Λ), and occurrences where the realisation inside $\mathbf{Z}[[X_\mathbf{N}]]$ is essential (because indeterminates X_i occur explicitly in the same equation); in the latter case we shall write $f[X_\mathbf{N}]$ instead of f. If the nonzero coefficients of $f[X_\mathbf{N}]$ only occur for monomials of degree d, then f is called homogeneous of degree d; due to the required degree bound, this makes Λ into a graded ring.

Another realisation of Λ is via its images in polynomial rings in finite sets of indeterminates. This is for instance the point of view taken in [Macd]; for us this realisation is important in order to be able to consider alternating expressions, which is hard to do for infinitely many indeterminates. For any $n \in \mathbf{N}$, let $X_{[n]} = \{\, X_i \mid i \in [n] \,\}$ be the set of the first n indeterminates. There is a ring morphism $\mathbf{Z}[[X_\mathbf{N}]] \to \mathbf{Z}[[X_{[n]}]]$ defined by setting $X_i := 0$ for all $i \geq n$, and the image of the subring Λ under this morphism is the subring of the symmetric polynomials in $\mathbf{Z}[X_{[n]}]$, those invariant under all permutations of the indeterminates; we shall denote this image by $\Lambda_{[n]}$. For $f \in \Lambda$, the image in $\Lambda_{[n]}$ of $f[X_\mathbf{N}] \in \mathbf{Z}[[X_\mathbf{N}]]$ will be denoted by $f[X_{[n]}]$. Thus each $f \in \Lambda$ gives rise to a family $(f[X_{[n]}])_{n\in\mathbf{N}}$ of elements $f[X_{[n]}] \in \Lambda_{[n]}$ of bounded degree, which family is coherent with respect to the projections $\Lambda_{[n+1]} \to \Lambda_{[n]}$ defined by the substitution $X_n := 0$. We shall write this final property as $f[X_{[n+1]}][X_n := 0] = f[X_{[n]}]$ for all $n \in \mathbf{N}$. Conversely each family $(f_n)_{n\in\mathbf{N}}$ with $f_n \in \Lambda_{[n]}$ for all $n \in \mathbf{N}$ that satisfies $f_{n+1}[X_n := 0] = f_n$ for all n, and for which $\deg f_n$ is bounded, forms the set of images of a unique element $f \in \Lambda$. In other words, one can realise Λ as the inverse limit in the category of graded rings of the system $(\Lambda_{[n]})_{n\in\mathbf{N}}$ relative to the given projections $\Lambda_{[n+1]} \to \Lambda_{[n]}$.

For any $\alpha \in \mathcal{P}_d$, the sum $m_\alpha[X_\mathbf{N}] = \sum_{\beta\in\mathcal{C}_d} [\, \alpha^+ = \beta^+ \,]\, X^\beta$ of all distinct monomials in the permutation orbit of X^α is a symmetric function. Since no nonempty proper subset of its nonzero terms defines a symmetric function, we shall call m_α a *minimal symmetric function* (we avoid the more traditional term “monomial” symmetric function since the set of all m_α is not closed under multiplication). The set $\{\, m_\lambda \mid \lambda \in \mathcal{P}_d \,\}$ is a basis of the additive group of homogeneous symmetric functions of degree d.

The *elementary symmetric functions* e_d for $d \in \mathbf{N}$ are instances of minimal symmetric functions: they are defined as $e_d = m_{1^{(d)}}$. One can write more explicitly

$$e_d[X_{\mathbf{N}}] = \sum_{\alpha \in \mathcal{C}_d^{[2]}} X^\alpha = \sum_{i_1,\ldots,i_d \in \mathbf{N}} [\, i_1 < \cdots < i_d \,]\, X_{i_1} \cdots X_{i_d}. \tag{4}$$

The *complete (homogeneous) symmetric functions* h_d for $d \in \mathbf{N}$ are defined by $h_d = \sum_{\lambda \in \mathcal{P}_d} m_\lambda$. Like the elementary symmetric functions, they can be written more explicitly

$$h_d[X_{\mathbf{N}}] = \sum_{\alpha \in \mathcal{C}_d} X^\alpha = \sum_{i_1,\ldots,i_d \in \mathbf{N}} [\, i_1 \leq \cdots \leq i_d \,]\, X_{i_1} \cdots X_{i_d}. \tag{5}$$

The *power sum symmetric functions* p_d for $d > 0$ are defined by $p_d = m_{(d)}$, so $p_d[X_{\mathbf{N}}] = \sum_{i \in \mathbf{N}} X_i^d$. These families of symmetric functions have the following generating series, expressed in $\mathbf{Z}[[X_{\mathbf{N}}, T]]$.

$$\sum_{d \in \mathbf{N}} e_d[X_{\mathbf{N}}] T^d = \prod_{i \in \mathbf{N}} (1 + X_i T), \tag{6}$$

$$\sum_{d \in \mathbf{N}} h_d[X_{\mathbf{N}}] T^d = \prod_{i \in \mathbf{N}} \left(\sum_{k \in \mathbf{N}} (X_i T)^k \right) = \prod_{i \in \mathbf{N}} \frac{1}{1 - X_i T}, \tag{7}$$

$$\sum_{k > 0} p_k[X_{\mathbf{N}}] T^k = \sum_{i \in \mathbf{N}} \left(\sum_{k > 0} (X_i T)^k \right) = \sum_{i \in \mathbf{N}} \frac{X_i T}{1 - X_i T}. \tag{8}$$

For any $\alpha \in \mathcal{C}$ we define $e_\alpha = \prod_{i \in \mathbf{N}} e_{\alpha_i}$ and $h_\alpha = \prod_{i \in \mathbf{N}} h_{\alpha_i}$; since $e_0 = h_0 = 1$ the infinite products converge, and it is clear by commutativity that $e_\alpha = e_{\alpha^+}$ and $h_\alpha = h_{\alpha^+}$. The products e_α and h_α can be expanded into monomials combinatorially, in terms of binary respectively integral matrices: by multiplying together copies of the first equality in (4) respectively in (5), one finds

$$e_\beta[X_{\mathbf{N}}] = \sum_{\alpha \in \mathcal{C}} \#\mathcal{M}^{[2]}_{\alpha,\beta} X^\alpha \tag{9}$$

$$h_\beta[X_{\mathbf{N}}] = \sum_{\alpha \in \mathcal{C}} \#\mathcal{M}_{\alpha,\beta} X^\alpha \tag{10}$$

We can obtain generating series in $\mathbf{Z}[[X_{\mathbf{N}}, Y_{\mathbf{N}}]]$ in which all e_β or all h_β appear, either from the preceding equations, or by substituting $T := Y_j$ into copies of (6) or (7) for $j \in \mathbf{N}$ and multiplying them, giving

$$\sum_{\beta \in \mathcal{C}} e_\beta[X_{\mathbf{N}}] Y^\beta = \sum_{M \in \mathcal{M}^{[2]}} X^{\mathrm{row}(M)} Y^{\mathrm{col}(M)} = \prod_{i,j \in \mathbf{N}} (1 + X_i Y_j) \tag{11}$$

$$\sum_{\beta \in \mathcal{C}} h_\beta[X_{\mathbf{N}}] Y^\beta = \sum_{M \in \mathcal{M}} X^{\mathrm{row}(M)} Y^{\mathrm{col}(M)} = \prod_{i,j \in \mathbf{N}} \frac{1}{1 - X_i Y_j}. \tag{12}$$

1.4. *Alternating polynomials and Schur functions.*

Now fix $n \in \mathbf{N}$, and let $A_{[n]}$ denote the additive subgroup of $\mathbf{Z}[X_{[n]}]$ of alternating polynomials, i.e., of polynomials p such that for all permutations $\sigma \in \mathbf{S}_n$ the permutation of indeterminates given by σ operates on p as multiplication be the sign $\varepsilon(\sigma)$. Multiplying an alternating polynomial by a symmetric polynomial gives another alternating polynomial, so if we view $\mathbf{Z}[X_{[n]}]$ as a module over its subring $\Lambda_{[n]}$, then it contains $A_{[n]}$ as a submodule. Like for symmetric polynomials, the condition of being an alternating polynomial can be expressed by comparing coefficients of monomials in the same permutation orbit: a polynomial $\sum_{\alpha\in\mathbf{N}^n} c_\alpha X^\alpha$ is alternating if and only if for every $\alpha \in \mathbf{N}^n$ and $\sigma \in \mathbf{S}_n$ one has $c_{\sigma\cdot\alpha} = \varepsilon(\sigma)c_\alpha$. In particular this implies that $c_\alpha = 0$ whenever α is fixed by any odd permutation, which happens as soon as $\alpha_i = \alpha_j$ for some pair $i \neq j$. In the contrary case, α is not fixed by any non-identity permutation, and the alternating orbit sum $a_\alpha[X_{[n]}] = \sum_{\sigma\in\mathbf{S}_n} \varepsilon(\sigma)X^{\sigma\cdot\alpha}$ is an alternating polynomial that is minimal in the sense that its nonzero coefficients are all ±1 and no nonempty proper subset of its nonzero terms defines an alternating polynomial. The element $a_\alpha[X_{[n]}]$ is called an *alternant*, and can be written as a determinant

$$a_\alpha[X_{[n]}] = \det\bigl(X_j^{\alpha_j}\bigr)_{i,j\in[n]} = \begin{vmatrix} X_0^{\alpha_0} & X_0^{\alpha_1} & \cdots & X_0^{\alpha_{n-1}} \\ X_1^{\alpha_0} & X_1^{\alpha_1} & \cdots & X_1^{\alpha_{n-1}} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n-1}^{\alpha_0} & X_{n-1}^{\alpha_1} & \cdots & X_{n-1}^{\alpha_{n-1}} \end{vmatrix}; \tag{13}$$

we define $a_\alpha[X_{[n]}]$ by the same expression even when α is fixed by some transposition, but in that case it is 0. The set of alternants generates $A_{[n]}$ as an additive group, but to obtain a $\mathbf{Z}$-basis one must remove the null alternants, and for all other orbits of compositions choose one of the two opposite alternants associated to it. Thus one finds the $\mathbf{Z}$-basis $\{\, a_\alpha[X_{[n]}] \mid \alpha \in \mathbf{N}^n; \alpha_0 > \cdots > \alpha_{n-1} \,\}$ of $A_{[n]}$. Our convention of interpreting finite vectors by extension with zeroes as finitely supported ones, allows us to view $\mathbf{N}^n$ as a subset of $\mathcal{C}$. Then putting $\delta_n = (n-1, n-2, \ldots, 1, 0) \in \mathbf{N}^n$ the above basis of $A_{[n]}$ can be written as $\{\, a_{\delta_n+\lambda}[X_{[n]}] \mid \lambda \in \mathcal{P} \cap \mathbf{N}^n \,\}$.

Put $\Delta_n = a_{\delta_n}[X_{[n]}]$; in other words, $\Delta_n \in A_{[n]}$ is the Vandermonde determinant, which evaluates to $\prod_{0\leq i<j<n}(X_i - X_j)$. Alternating polynomials are all divisible by each factor $X_i - X_j$, and therefore by Δ_n. So viewing $A_{[n]}$ as an $\Lambda_{[n]}$-module, it is cyclic with generator Δ_n. The map $\Lambda_{[n]} \to A_{[n]}$ of multiplication by Δ_n is a $\mathbf{Z}$-linear bijection, so one can apply its inverse to the basis of $A_{[n]}$ consisting of elements $a_{\delta_n+\lambda}[X_{[n]}]$. Thus defining

$$s_\lambda[X_{[n]}] = \frac{a_{\delta_n+\lambda}[X_{[n]}]}{\Delta_n} \in \Lambda_{[n]} \tag{14}$$

for $\lambda \in \mathcal{P}$ with $\lambda_n = 0$, the set $\{\, s_\lambda[X_{[n]}] \mid \lambda \in \mathcal{P}; \lambda_n = 0 \,\}$ forms a $\mathbf{Z}$-basis of $\Lambda_{[n]}$. It is useful to define $s_\alpha[X_{[n]}]$ for arbitrary $\alpha \in \mathbf{N}^n$ by the same formula. Doing so does not introduce any new symmetric functions, since one has $s_\alpha[X_{[n]}] = 0$ unless the n components of $\delta_n + \alpha$ are all distinct, and in that case one has $s_\alpha[X_{[n]}] = \varepsilon s_\lambda[X_{[n]}]$, where $\lambda \in \mathcal{P}$ with $\lambda_n = 0$ is determined by the condition $(\delta_n + \alpha)^+ = \delta_n + \lambda$, and ε is the sign of the (unique) permutation $\sigma \in \mathbf{S}_n$ such that $\sigma(\delta_n+\alpha)=\delta_n+\lambda$. Finally we extend this definition to any (infinite) composition $\alpha \in \mathcal{C}$, by defining $s_\alpha[X_{[n]}]=0$ whenever $\alpha \notin \mathbf{N}^n$.

Here are some examples illustrating these definitions. One has

$$\begin{aligned}
s_{(3,1)}[X_{[2]}] &= \frac{a_{(4,1)}[X_{[2]}]}{\Delta_2} = \frac{X_0^4X_1 - X_0X_1^4}{X_0-X_1} = X_0^3X_1 + X_0^2X_1^2 + X_0X_1^3 \\
&= m_{(3,1)}[X_{[2]}] + m_{(2,2)}[X_{[2]}] \\
s_{(1,2)}[X_{[2]}] &= \frac{a_{(2,2)}[X_{[2]}]}{\Delta_2} = 0 \\
s_{(0,4)}[X_{[2]}] &= \frac{a_{(1,4)}[X_{[2]}]}{\Delta_2} = \frac{X_0X_1^4 - X_0^4X_1}{X_0-X_1} = -s_{(3,1)}[X_{[2]}] \\
s_{(3,1)}[X_{[3]}] &= \frac{a_{(5,2,0)}[X_{[3]}]}{\Delta_3} = \frac{X_0^5X_1^2 - X_0^2X_1^5 - X_0^5X_2^2 + X_1^5X_2^2 + X_0^2X_2^5 - X_1^2X_2^5}{(X_0-X_1)(X_0-X_1)(X_1-X_2)} \\
&= X_0^3X_1 + X_0^3X_2 + X_0^2X_1^2 + 2X_0^2X_1X_2 + X_0^2X_2^2 \\
&\quad + X_0X_1^3 + 2X_0X_1^2X_2 + 2X_0X_1X_2^2 + X_0X_2^3 + X_1^3X_2 + X_1^2X_2^2 + X_1X_2^3 \\
&= m_{(3,1)}[X_{[3]}] + m_{(2,2)}[X_{[3]}] + 2m_{(2,1,1)}[X_{[3]}]
\end{aligned}$$

1.4.1. Proposition. *For all $n \in \mathbf{N}$ and all $\alpha \in \mathcal{C}$ one has $s_\alpha[X_{[n+1]}][X_n := 0] = s_\alpha[X_{[n]}]$.*

For instance one sees in the example above that $s_{(3,1)}[X_{[3]}][X_2 := 0] = s_{(3,1)}[X_{[2]}]$: from the definition one has $m_\alpha[X_{[n+1]}][X_n := 0] = m_\alpha[X_{[n]}]$ for all $\alpha \in \mathcal{C}$, while $m_\alpha[X_{[n]}]=0$ unless $\alpha \in \mathbf{N}^n \subset \mathcal{C}$, so in particular $m_{(2,1,1)}[X_{[2]}]=0$.

Proof. The value $s_\alpha[X_{[n+1]}][X_n := 0]$ can be computed by applying the substitution $X_n := 0$ separately to the numerator $a_{\delta_{n+1}+\alpha}[X_{[n+1]}]$ and the denominator Δ_{n+1} in the definition of $s_\alpha[X_{[n+1]}]$, provided that the latter substitution yields a nonzero value; this is the case since $\Delta_{n+1}[X_n := 0] = X_0 \cdots X_{n-1}\Delta_n$. We may assume that $\alpha \in \mathbf{N}^{n+1}$ holds, since otherwise both $s_\alpha[X_{[n+1]}]$ and $s_\alpha[X_{[n]}]$ are zero by definition. Now put $\beta = \delta_{n+1}+\alpha = (n+\alpha_0,\ldots,1+\alpha_{n-1},\alpha_n) \in \mathbf{N}^{n+1}$, so that the mentioned numerator is $a_\beta[X_{[n+1]}]$. If one has $\alpha \notin \mathbf{N}^n$, so that $s_\alpha[X_{[n]}]=0$ by definition, then $\alpha_n \neq 0$, and all of the (first $n+1$) components of β are nonzero; in this case the substitution $X_n := 0$ kills all terms of the numerator, so that $s_\alpha[X_{[n+1]}][X_n := 0] = 0$. On the other hand if $\alpha \in \mathbf{N}^n$ then also $\beta \in \mathbf{N}^n$, and in this case $a_\beta[X_{[n+1]}][X_n := 0] = a_\beta[X_{[n]}] = X_0 \cdots X_{n-1} a_{\delta_n+\alpha}[X_{[n]}]$; after simplification of the substituted numerator and denominator by $X_0 \cdots X_{n-1}$, one obtains $s_\alpha[X_{[n+1]}][X_n := 0] = s_\alpha[X_{[n]}]$ as desired. □

Thus for fixed $\alpha \in \mathcal{C}$, the families $(s_\alpha[X_{[n]}])_{n\in\mathbf{N}}$ of symmetric polynomials are coherent with respect to the projections $\Lambda_{[n+1]} \to \Lambda_{[n]}$ defined by the substitution $X_n := 0$. This property allows the following definition, which we already anticipated in our notation.

1.4.2. Definition. *For $\alpha \in \mathcal{C}$, the symmetric function s_α is the unique element of Λ whose image in $\Lambda_{[n]}$ under the substitutions $X_i := 0$ for all $i \geq n$ is $s_\alpha[X_{[n]}] = [\,\alpha \in \mathbf{N}^n\,]\,\frac{a_{\delta_n+\alpha}[X_{[n]}]}{a_{\delta_n}[X_{[n]}]}$, for all $n \in \mathbf{N}$.*

The set $\{ s_\lambda \mid \lambda \in \mathcal{P} \}$ forms a $\mathbf{Z}$-basis of Λ, whose elements are called *Schur functions*. They are the central subject of this paper, and we shall now introduce several notations to facilitate their study. Firstly we shall denote by $\langle \cdot \mid \cdot \rangle$ the scalar product on Λ for which the basis of Schur functions is orthonormal. Thus one has

$$f = \sum_{\lambda \in \mathcal{P}} \langle f \mid s_\lambda \rangle s_\lambda \qquad \text{for any } f \in \Lambda. \tag{15}$$

The operation of multiplication by a fixed Schur function s_μ has an adjoint operation s_μ^* for this scalar product, i.e., which satisfies $\langle s_\mu^*(f) \mid g \rangle = \langle f \mid s_\mu g \rangle$ for all $f, g \in \Lambda$; in terms of this the *skew Schur functions* are defined by

$$s_{\lambda/\mu} = s_\mu^*(s_\lambda). \tag{16}$$

They typically arise when one expresses the multiplication by a fixed symmetric function in the basis of Schur functions, as skew Schur functions are characterised by

$$\langle s_\mu f \mid s_\lambda \rangle = \langle f \mid s_{\lambda/\mu} \rangle \qquad \text{for all } \lambda, \mu \in \mathcal{P} \text{ and } f \in \Lambda. \tag{17}$$

For $f = h_\alpha$ and $f = e_\alpha$ these scalar products are of particular interest, and are called *Kostka numbers*.

1.4.3. Definition. *For $\mu, \lambda \in \mathcal{P}$ and $\alpha \in \mathcal{C}$, we set $K_{\lambda/\mu,\alpha} = \langle h_\alpha \mid s_{\lambda/\mu} \rangle$ and $K'_{\lambda/\mu,\alpha} = \langle e_\alpha \mid s_{\lambda/\mu} \rangle$.*

When $\mu = (0)$, we abbreviate $K_{\lambda/\mu,\alpha}$ to $K_{\lambda,\alpha}$ and $K'_{\lambda/\mu,\alpha}$ to $K'_{\lambda,\alpha}$. Then as a special case of (15) one has

$$h_\alpha = \sum_{\lambda \in \mathcal{P}} K_{\lambda,\alpha} s_\lambda \qquad \text{and} \qquad e_\alpha = \sum_{\lambda \in \mathcal{P}} K'_{\lambda,\alpha} s_\lambda. \tag{18}$$

We shall later give combinatorial descriptions of the Kostka numbers (corollary 2.6), from which it will become clear that they are non-negative (this also follows from representation theoretic considerations), and are related by $K'_{\lambda/\mu,\alpha} = K_{\lambda^t/\mu^t,\alpha}$.

Computing scalar products $\langle f \mid s_\lambda \rangle$ directly from the definition is usually quite hard. But if $f = s_\alpha$ with $\alpha \in \mathcal{C}$, then f can be considered as slightly generalised Schur function: either $f = 0$, or $f = \pm s_\lambda$ for some $\lambda \in \mathcal{P}$. Setting

$$\varepsilon(\alpha, \lambda) = \langle s_\alpha \mid s_\lambda \rangle \qquad \text{for } \alpha \in \mathcal{C} \text{ and } \lambda \in \mathcal{P}, \tag{19}$$

one has $\varepsilon(\alpha, \lambda) \in \{-1, 0, 1\}$, and given α there is at most one λ with $\varepsilon(\alpha, \lambda) \neq 0$; this symbol will be used as a signed variant of the Iverson symbol. Any expression in terms of such s_α can be converted to one in terms of Schur functions using

$$s_\alpha = \sum_{\lambda \in \mathcal{P}} \varepsilon(\alpha, \lambda) s_\lambda, \tag{20}$$

which removes null terms, and replaces the remaining terms s_α by the appropriate $\pm s_\lambda$ with $\lambda \in \mathcal{P}$. This process, which is the main source of alternating sums in this paper, will be called normalisation.

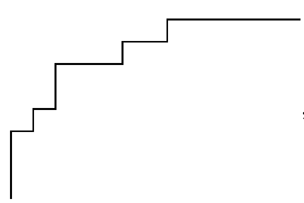
1.5. *Diagram boundaries.*

The partition λ with $\varepsilon(\alpha, \lambda) \neq 0$, if any, is characterised by the condition $\delta_n + \lambda = (\delta_n + \alpha)^+$, where n is so large that $\alpha \in \mathbf{N}^n$, and $\varepsilon(\alpha, \lambda)$ is the sign of the permutation $\sigma \in \mathbf{S}_n$ such that $\delta_n + \lambda = \sigma(\delta_n + \alpha)$. The given condition gives rise to an equivalent one when n is increased, and the permutation involved does not change either, if each $\mathbf{S}_n$ is considered as a subgroup of $\mathbf{S}_\infty$. Nonetheless, it is convenient to have a description that does not involve n at all. If we subtract n from each of the n entries of $\delta_n + \alpha = (n - 1 - i + \alpha_i)_{i \in [n]}$, then coefficient i becomes $\alpha_i - 1 - i$, and therefore independent of $n > i$; moreover this transformation is compatible with the action of $\mathbf{S}_n$ by permutation of the coefficients. Now increasing n allows to associate with α a unique infinite sequence of numbers; note however that it will contain negative entries, and will no longer be finitely supported. We shall denote this sequence by $\alpha[\,] = (\alpha[i])_{i \in \mathbf{N}}$, where

$$\alpha[i] = \alpha_i - 1 - i \qquad \text{for } \alpha \in \mathcal{C} \text{ and } i \in \mathbf{N}. \tag{21}$$

Thus if $\varepsilon(\alpha, \lambda)$ is nonzero, it is the sign of the unique permutation $\sigma \in \mathbf{S}_\infty$, that transforms $\alpha[\,]$ into $\lambda[\,]$.

Instead of using sequences $\alpha[\,]$, one could understand $\varepsilon(\alpha, \lambda)$ using an alternative way for the group $\mathbf{S}_\infty$ to act on finitely supported sequences of integers: defining $\sigma \cdot \alpha = \bigl(\alpha_{\sigma^{-1}(i)} + i - \sigma^{-1}(i)\bigr)_{i \in \mathbf{N}}$, one may interpret $\varepsilon(\alpha, \lambda)$ as the sign of the unique permutation $\sigma \in \mathbf{S}_\infty$, if any, such that $\sigma \cdot \alpha = \lambda$. However, considering sequences $\alpha[\,]$ has the advantage that they can be interpreted directly in a graphical manner, in particular when α is a partition. Each part λ_i of a partition λ corresponds to a row of its diagram $[\lambda]$, and therefore to a unique vertical segment of the bottom right boundary of $[\lambda]$, namely the segment that delimits row i of $[\lambda]$ to the right. In case $\lambda_i = 0$, row i of $[\lambda]$ is empty, but we can still consider it as delimited to the right by the segment on the vertical axis that crosses row i. Now, whereas the sequence λ records the horizontal coordinates of these vertical segments (taken from top to bottom), the sequence $\lambda[\,]$ records their "diagonal coordinates", where a segment has diagonal coordinate d if it connects points on diagonals d and $d+1$. For instance, for $\lambda = (7, 5, 2, 2, 1)$ the shape of the boundary can be drawn as

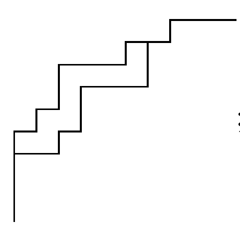


and the sequence of the diagonal coordinates of its vertical segments, taken from top to bottom, is $(6, 3, -1, -2, -4, -6, -7, \ldots) = \lambda[\,]$. For general compositions α one could define a "diagram" in the same way as for partitions (although we did not), and while their boundaries are more ragged, reading off the diagonal coordinates of the vertical segments from top to bottom still gives the sequence $\alpha[\,]$.

Let us make more precise some terms that were used above. Both squares and points are formally elements of $\mathbf{N} \times \mathbf{N}$, but we shall treat them as different types of objects.

The square (i,j) has four corners, which are the points in the set $\{i,i+1\}\times\{j,j+1\}$, so the square has the same coordinates as (and is formally speaking identified with) its top left corner; this is in keeping with our convention to index by i the interval from i to $i+1$. Here we shall treat the diagonals D_d (which were defined in (1)) as sets of points, which should not be too confusing since a square "lies on the same diagonal" as its top left corner. The boundary associated to $\lambda\in\mathcal{P}$ consists of a set of points, one on each diagonal D_d, which are connected by edges between the point on D_d and the point on D_{d+1} for each $d\in\mathbf{Z}$. Using the fact that each diagonal is totally ordered under the coordinate-wise partial ordering of points, the point on D_d can be characterised as the minimal element of the difference set $D_d\setminus[\lambda]$, in other words the top-leftmost point on D_d that is not the top-left corner of any square of $[\lambda]$. The edges of this boundary are also indexed by numbers $d\in\mathbf{Z}$ (their diagonal coordinate), as described above. Thus the vertical edge associated to the part λ_i of λ is the one between the points (i,λ_i) and $(i+1,\lambda_i)$, which has as diagonal coordinate the index $\lambda_i-(i+1)=\lambda[i]$ of the diagonal containing the latter point. In the example above the vertical segments are $(0,7)$–$(1,7)$, $(1,5)$–$(2,5)$, $(2,2)$–$(3,2)$, $(3,2)$–$(4,2)$, $(4,1)$–$(5,1)$, $(5,0)$–$(6,0)$, $(6,0)$–$(7,0)$, $\ldots$

We did not use the horizontal segments of the boundary associated to $\lambda\in\mathcal{P}$. However, since the sequence $\lambda[\,]$ is strictly decreasing, it is entirely determined by the *set* $\{\,\lambda[i]\mid i\in\mathbf{N}\,\}\subset\mathbf{Z}$, and the set of diagonal coordinates of the horizontal segments of the boundary forms the complement of that set in $\mathbf{Z}$, which can be written as $\{\,-1-\lambda^t[j]\mid j\in\mathbf{N}\,\}$. It is often useful to think of λ as represented by the characteristic function of the set $\{\,\lambda[i]\mid i\in\mathbf{N}\,\}$, i.e., the function $\mathbf{Z}\to\{0,1\}$ that takes the value 1 at d if d is the diagonal coordinate of a vertical segment of the boundary, and the value 0 if d is the diagonal coordinate of a horizontal segment. In turn that function can be thought of as a doubly infinite sequence of bits (the edge sequence of λ), which describes the form of the boundary, traversed from bottom left to top right. The relations $\mu\leftharpoonup\lambda$, $\mu\leftharpoondown\lambda$ and $\mu\prec_{\mathrm{r}(k)}\lambda$ can be easily understood in terms of transpositions of bits in edge sequences; notably $\mu\prec_{\mathrm{r}(k)}\lambda$ means that the edge sequence of λ is obtained from that of μ by transposing a single bit 1 with a bit 0 that is k places to its right. This is graphically obvious if we draw the diagram of the k by superimposing the boundaries associated to the partitions involved:

$(7,5,2,2,1)\prec_{\mathrm{r}(10)}(7,6,6,3,3,2)$:  ;

the transposition involves the underlined bits in $\cdots 111\underline{1}0101100001\underline{0}01000\cdots$. Moreover $\mathrm{ht}(\lambda/\mu)$ is the sum of the bits that are being "jumped over" during the transposition, which is 4 in the example.

1.5.1. Proposition. *For $k>0$ and $\mu,\lambda\in\mathcal{P}$, the relation $\mu\prec_{\mathrm{r}(k)}\lambda$ holds if and only if there exist indices $i_0,i_1\in\mathbf{Z}$ such that $\{\,\mu[i]\mid i\in\mathbf{N}-\{i_0\}\,\}=\{\,\lambda[i]\mid i\in\mathbf{N}-\{i_1\}\,\}$ and $\mu[i_0]+k=\lambda[i_1]$. In this case one has moreover* $\mathrm{ht}(\lambda/\mu)=i_0-i_1$.

Proof. By definition, $\mu \prec_{\mathrm{r}(k)} \lambda$ means that k consecutive points on the boundary associated to μ must be moved one place down their diagonal to reach the boundary associated to λ. This means that only the two segments that link a point of this subset of k points and a point of its complement change orientation, and their diagonal coordinates differ by k. In terms of the set $\{\, \mu[i] \mid i \in \mathbf{N} \,\}$, this means that one element $\mu[i_0]$ is replaced by $\mu[i_0] + k = \lambda[i_1]$, and since i_0 and i_1 are the row numbers of the vertical boundary segments with diagonal coordinates $\mu[i_0]$ and $\lambda[i_1]$, respectively, one has $\mathrm{ht}(\lambda/\mu) = i_0 - i_1$. □

§2. The Pieri and Murnaghan-Nakayama rules.

In this section we shall compute the product of Schur polynomials by respectively elementary, power-sum and complete symmetric functions, expressing the result in the basis of Schur functions. These lead to nice combinatorial formulae known as the Pieri and Murnaghan-Nakayama rules, in which one encounters the notions of vertical and horizontal strips, and of ribbons.

Before starting our computations, we consider the validity of summations involving the values s_α for all $\alpha \in \mathcal{C}_n$ at once. Although each of these is either zero or up to a sign equal to one of the finitely many Schur functions s_λ with $\lambda \in \mathcal{P}_n$, this does not prove that only finitely many s_α are nonzero, and such summations therefore well defined. The following proposition shows that this is nevertheless the case.

2.1. Proposition. *If $s_\alpha \neq 0$ for some $\alpha \in \mathcal{C}_n$, then $\alpha \in \mathbf{N}^n$, where $\mathbf{N}^n$ is considered as subset of $\mathcal{C}$.*

Proof. If $s_\alpha \neq 0$, then there exists $\lambda \in \mathcal{P}_n \subseteq \mathbf{N}^n$ with $\varepsilon(\alpha, \lambda) \neq 0$. Now the sequence $\alpha[\,]$ is a permutation of $\lambda[\,]$, and therefore in particular it contains all values $-1-i$ for $i \geq n$ as entries. If α had any part $\alpha_i > 0$ with $i \geq n$, then for the largest such i, the value $-1-i$ would not occur in the sequence $\alpha[\,]$, which would give a contradiction. Therefore one must have $\alpha \in \mathbf{N}^n$. □

The starting point of our approach is a very simple formula describing the multiplication between a Schur function and an arbitrary symmetric function.

2.2. Lemma. *Let $f \in \Lambda$, and write $f[X_{\mathbf{N}}]$ as a sum of monomials $f[X_{\mathbf{N}}] = \sum_{\alpha \in \mathcal{C}} c_\alpha X^\alpha$, then*

$$f s_\beta = \sum_{\alpha \in \mathcal{C}} c_\alpha s_{\alpha+\beta} \qquad \text{for all } \beta \in \mathcal{C}.$$

Proof. First, we must assure that the summation is effectively finite (or stated more fancily: that it converges in Λ for the discrete topology), so that the right hand side is well defined. For this it suffices to observe that in order to have $c_\alpha \neq 0$ one must have $|\alpha| \leq \deg f$, and that in order to have $s_{\alpha+\beta} \neq 0$, one must have $\alpha + \beta \in \mathbf{N}^{|\alpha|+|\beta|}$ by proposition 2.1, and therefore certainly $\alpha \in \mathbf{N}^{|\alpha|+|\beta|}$, which leaves only finitely many possibilities for α. With this point settled, it will suffice to prove that the image $f[X_{[n]}] s_\beta[X_{[n]}] = \sum_{\alpha \in \mathbf{N}^n} c_\alpha s_{\alpha+\beta}[X_{[n]}]$ of the identity in $\Lambda_{[n]}$ holds for all $n \in \mathbf{N}$. We may assume $\beta \in \mathbf{N}^n$; it is then obvious that $f[X_{[n]}] X^{\delta_n + \beta} = \sum_{\alpha \in \mathbf{N}^n} c_\alpha X^{\delta_n + \alpha + \beta}$. To this equation we apply every $\sigma \in \mathbf{S}_n$, acting by permutation of the indeterminates, and take the alternating sum by $\varepsilon(\sigma)$ of the resulting equations. Since $f[X_{[n]}]$ is invariant under such permutations, we obtain the equation $f[X_{[n]}]a_{\delta_n+\beta}[X_{[n]}] = \sum_{\alpha\in\mathbf{N}^n} c_\alpha a_{\delta_n+\alpha+\beta}[X_{[n]}]$ in $A_{[n]}$. Now division by Δ_n gives the desired identity. □

This lemma can be paraphrased as follows: when multiplying symmetric functions by Schur functions, one may behave as if $X^\alpha s_\beta$ were equal to $s_{\alpha+\beta}$ (although of course it is not). This lemma will be our main tool for this paper. It allows formulae for products of symmetric functions by Schur functions to be obtained very easily, but the resulting expression on the right hand side is not normalised, even when $\beta \in \mathcal{P}$. After normalisation one gets an alternating summation with possibly many internal cancellations; the remaining task is to perform such cancellations explicitly, and describe the surviving terms in a combinatorial way. This often involves a certain amount of arbitrariness: in general an alternating sum neither uniquely determines the set of terms that will survive the cancellation, nor the way in which the other terms are paired up so as to cancel (in so far as they are not already null initially). In our applications however, a set of candidates for survival that are already normalised will usually present itself, and the arbitrariness is limited to the way in which the remaining terms cancel.

Here is a concrete example of a computation using lemma 2.2, namely that of the decomposition of the product $m_{(3,1)}s_{(2,1)}$ into Schur functions. We traverse the $\mathbf{S}_\infty$-orbit of $(3,1)$ in decreasing lexicographic order $((3,1), (3,0,1), (3,0,0,1), \ldots, (1,3), (1,0,3), \ldots, (0,3,1), \ldots, (0,1,3), \ldots, (0,0,3,1), \ldots)$, but omit those elements α for which $((2,1)+\alpha)[\,]$ has repeated entries. After normalisation there are no cancellations, but in two cases a pair of terms is combined into one:

$$\begin{aligned} m_{(3,1)}s_{(2,1)} &= \sum_{\alpha\in\mathbf{S}_\infty\cdot(3,1)} s_{(2,1)+\alpha} \\ &= s_{(5,2)} + s_{(5,1,1)} + s_{(3,1,3)} + s_{(3,1,0,0,3)} + s_{(2,4,1)} \\ &\quad + s_{(2,2,0,3)} + s_{(2,2,0,0,3)} + s_{(2,1,3,1)} + s_{(2,1,1,0,0,3)} + s_{(2,1,0,0,3,1)} \\ &= s_{(5,2)} + s_{(5,1,1)} - s_{(3,2,2)} + s_{(3,1,1,1,1)} - s_{(3,3,1)} \\ &\quad - 2s_{(2,2,2,1)} + s_{(2,2,1,1,1)} + 2s_{(2,1,1,1,1,1)}. \end{aligned}$$

In the remainder of this section we shall apply the lemma three times, in each case for a symmetric function f all of whose nonzero coefficients c_α are equal to 1. The first and easiest application will be for $f = e_d$, where the monomials X^α with nonzero coefficients are those for binary compositions α of d.

2.3. Proposition. *For $d \in \mathbf{N}$ and $\mu \in \mathcal{P}_n$ one has $e_d s_\mu = \sum_{\lambda\in\mathcal{P}_{n+d}} [\,\mu \leftharpoondown \lambda\,]\, s_\lambda$.*

Proof. Applying lemma 2.2 gives $e_d s_\mu = \sum_{\alpha\in\mathcal{C}_d^{[2]}} s_{\mu+\alpha}$. Now $\mu+\alpha$ can only fail to be a partition if for some index i one has $\mu_i = \mu_{i+1}$ while $(\alpha_i, \alpha_{i+1}) = (0,1)$; in that case one has $s_{\mu+\alpha} = 0$ since $(\mu+\alpha)[i] = (\mu+\alpha)[i+1]$. The remaining terms are already normalised, and the corresponding values of $\mu+\alpha$ run through the set $\{\,\lambda \in \mathcal{P}_{n+d} \mid \mu \leftharpoondown \lambda\,\}$ by the definition (1.1.1) of the relation '$\leftharpoondown$'. □

As a first consequence of this proposition one finds for $\mu = (0)$ that $s_{1^{(d)}} = e_d$, since $s_{(0)} = 1$ and $\{\, \lambda \in \mathcal{P}_d \mid 0 \leftharpoondown \lambda \,\} = \mathcal{P}_d \cap \mathcal{C}^{[2]} = \{1^{(d)}\}$; this fact can also be easily deduced directly from the definition of Schur functions. Another instance of the proposition gives $e_1 s_\mu = \sum_{\lambda \in \mathcal{P}_{n+1}} [\, \mu \subseteq \lambda \,]\, s_\lambda$ for $\mu \in \mathcal{P}_n$.

In this first application of lemma 2.2, it sufficed to discard the null terms in the resulting summation. Our second application, which will be for $f = p_k$, is slightly more complicated: it still suffices to discard the null terms (there are no cancellations of pairs of terms), but the remaining terms are not all normalised, and in normalising them some signs do remain (for $k \geq 2$). This will in fact be the only case in our paper where we obtain a result that is not a positive sum of Schur functions. The monomials occurring in p_k are just the powers X_i^k, which are the monomials X^α where α is a permutation of the composition (k), in other words a composition of k with just one nonzero part $\alpha_i = k$ for some $i \in \mathbf{N}$.

2.4. Proposition. *For $k > 0$ and $\mu \in \mathcal{P}_n$ one has $p_k s_\mu = \sum_{\lambda \in \mathcal{P}_{n+k}} [\, \mu \prec_{\mathrm{r}(k)} \lambda \,]\, (-1)^{\mathrm{ht}(\lambda/\mu)} s_\lambda$.*

Proof. Applying lemma 2.2 gives $p_k s_\mu = \sum_{i \in \mathbf{N}} s_{\mu + \alpha^{(i)}}$ where $\alpha^{(i)} = ([\, i = j \,]\, k)_{j \in \mathbf{N}}$ is the composition of k with a unique nonzero part at position i. If for any $j < i$ one has $\mu[i] + k = \mu[j]$, then the corresponding term $s_{\mu+\alpha^{(i)}}$ vanishes, because the sequence $(\mu + \alpha^{(i)})[\,]$ has equal entries at positions i and j. We discard such terms, and henceforth suppose that no such j exists. Then the sequence $(\mu + \alpha^{(i)})[\,]$ can be transformed into a strictly decreasing sequence by successively transposing the entry $(\mu + \alpha^{(i)})[i] = \mu[i] + k$ with its preceding entry some number of times (possibly zero), and the resulting sequence is of the form $\lambda[\,]$ for some $\lambda \in \mathcal{P}_{n+k}$. If $j \leq i$ is the final position of the term $\mu[i] + k$, so that $\mu[i] + k = \lambda[j]$, then the sign of the permutation that has been applied is $(-1)^{i-j} = \varepsilon(\mu + \alpha^{(i)}, \lambda)$. Now by proposition 1.5.1 one has $\mu \prec_{\mathrm{r}(k)} \lambda$ and $\mathrm{ht}(\lambda/\mu) = i - j$, and the proposition follows. ∎

One obtains for instance, for $\mu = (0)$, an expression of p_k as an alternating sum of "hook shape" Schur functions: $p_k = \sum_{i \in [k]} (-1)^i s_{\nu^{(i)}}$ where $\nu^{(i)} = (k - i, 1^{(i)})$ is the partition whose nonzero parts consist of one part $k - i$ followed by i parts 1; a concrete example is $p_4 = s_{(4)} - s_{(3,1)} + s_{(2,1,1)} - s_{(1,1,1,1)}$.

Using (15), one sees that this proposition is equivalent to the statement that for $\lambda \in \mathcal{P}_n$, $k > 0$, and $f \in \Lambda$ of degree $n - k$, one has

$$\langle\, p_k f \mid s_\lambda \,\rangle = \sum_{\mu \in \mathcal{P}_{n-k}} [\, \mu \prec_{\mathrm{r}(k)} \lambda \,]\, (-1)^{\mathrm{ht}(\lambda/\mu)} \langle\, f \mid s_\mu \,\rangle. \tag{22}$$

Since it is known that a scalar product of the form $\langle\, p_{k_1} \cdots p_{k_l} \mid s_\lambda \,\rangle$ can be interpreted as the value of the symmetric group character parametrised by λ on the class of permutations with cycles of lengths $k_1, \ldots, k_l$, this gives a recurrence relation for symmetric group characters. For instance, the value of the $\mathbf{S}_{17}$-character parametrised by $\lambda = (7,5,2,2,1)$ on a permutation of cycle type $(4,3,3,2,2,1,1,1)$ is $\langle\, p_4\, p_3^2\, p_2^2\, p_1^3 \mid s_\lambda \,\rangle$, which can be computed either as $-\langle\, p_3^2\, p_2^2\, p_1^3 \mid s_{(4,4,2,2,1)} \,\rangle + \langle\, p_3^2\, p_2^2\, p_1^3 \mid s_{(7,5,1)} \,\rangle$ using the recurrence relation for $k=4$, or as $\langle\, p_4\, p_3\, p_2^2\, p_1^3 \mid s_{(7,2,2,2,1)} \,\rangle - \langle\, p_4\, p_3\, p_2^2\, p_1^3 \mid s_{(7,5,2)} \,\rangle$ using the recurrence relation for $k=3$ (one could also use $k=2$ or $k=1$, but that gives expressions with 3 and 4 terms, respectively); if the character tables for $\mathbf{S}_{13}$ or $\mathbf{S}_{14}$ have been previously determined, this can be evaluated to give the value $-3+-2=-4-1=-5$. This recurrence relation is known as the Murnaghan-Nakayama rule, and appears in [Mur] and [Nak]. That name is also given to the combinatorial expression for $\langle\, p_{k_1}\cdots p_{k_l} \mid s_\lambda \,\rangle$ in terms of chains of partitions related by '$\prec_{\mathrm{r}(k)}$' for varying k, that can be obtained by expanding the recurrence relation recursively; however, in that form, which can already be found in [LiRi], the rule is rather less practical for computations, due to the large number of terms it often gives (which in addition depends heavily on the order in which the factors p_{k_i} are extracted).

Our third application of lemma 2.2 will be for $f=h_d$; here the coefficients c_α are nonzero for all compositions $\alpha \in \mathcal{C}_d$. Thus the summation obtained from the lemma gives many more terms than in the previous applications, and we must perform some real cancellations. Yet, surprisingly, the resulting linear combination of Schur functions will be quite similar to the one obtained in proposition 2.3, notably all nonzero coefficients are equal to 1. In fact the only difference is the replacement of vertical strips by horizontal strips. That proposition and the following one are known as the Pieri rules (or Pieri formulae).

2.5. Proposition. *For $d \in \mathbf{N}$ and $\mu \in \mathcal{P}_n$ one has $h_d s_\mu = \sum_{\lambda\in\mathcal{P}_{n+d}} [\,\mu \leftharpoondown \lambda\,]\, s_\lambda$.*

Proof. Application of lemma 2.2 gives $h_d s_\mu = \sum_{\alpha\in\mathcal{C}_d} s_{\mu+\alpha}$. In this summation we must cancel all the terms with $\mu \not\leftharpoondown \mu+\alpha$, in other words those terms with $(\mu+\alpha)_{i+1} > \mu_i$ for some index $i \in \mathbf{N}$. That inequality is equivalent to $(\mu+\alpha)[i+1] \geq \mu[i]$, which means that $(\mu+\alpha)[i+1]$ lies outside the "safe interval" $I_{i+1} = \{\, j \in \mathbf{Z} \mid \mu[i+1] \leq j < \mu[i] \,\}$. Setting in addition to this $I_0 = \{\, j \in \mathbf{Z} \mid \mu[0] \leq j \,\}$, the intervals I_i form a partition of $\mathbf{Z}$, so α determines a unique map $f\colon \mathbf{N} \to \mathbf{N}$ such that $(\mu+\alpha)[i] \in I_{f(i)}$ for all i. Moreover, one clearly has $f(i) \leq i$ for all i, so if f is injective, it must be the identity, which is equivalent to $\mu \leftharpoondown \mu+\alpha$. We must then cancel the terms for those α for which f is not injective. Indeed these terms can be cancelled, since if $f(i_0)=f(i_1)$ for $i_0 \neq i_1$, then the transposition of entries i_0 and i_1 of the sequence $(\mu+\alpha)[\,]$ gives a sequence of the form $(\mu+\alpha')[\,]$, with $\alpha' \in \mathcal{C}_d$ and $s_{\mu+\alpha} + s_{\mu+\alpha'} = 0$; either these are two distinct terms that cancel each other, or if $\alpha=\alpha'$ they are twice the same term, which is then null already. It remains to show that cancelling terms can be paired up effectively, in other words that a concrete pair of indices (i_0,i_1) can be chosen for each $\alpha \in \mathcal{C}_d$ with $\mu \not\leftharpoondown \mu+\alpha$, in such a way that for the resulting α' the same pair (i_0,i_1) will be chosen. This is easy: since the function f is the same for any α' that would cancel with α, any choice that depends only on f will do; for instance one can take the minimal possible pair (i_0,i_1) in lexicographic order. □

Similarly to what we saw for proposition 2.3, a first consequence of proposition 2.5 is that $s_{(d)} = h_d$, which is not quite as easy to obtain directly from definition 1.4.2 as $s_{1^{(d)}} = e_d$ is. As an illustration we choose a rather modest example, since the number of terms in the summation tends to grow rather rapidly: one has $h_4 s_{(4,2,2,1)} = s_{(9,2,2,1)} + s_{(8,3,2,1)} + s_{(8,2,2,2)} + s_{(8,2,2,2,1)} + s_{(7,4,2,1)} + s_{(7,3,2,2)} + s_{(7,3,2,1,1)} + s_{(7,2,2,2,1)} + s_{(6,4,2,2)} +$

$s_{(6,4,2,1,1)} + s_{(6,3,2,2,1)} + s_{(5,4,2,2,1)}$. Although a dozen terms remain, the number of nonzero terms produced by application of lemma 2.2 before normalisation is considerably larger, namely 178. To illustrate the cancellation, consider the cancelled term $s_{(4,2,4,2,2)}$ that arises for $\alpha = (0,0,2,1,2)$. Putting $\mu = (4,2,2,1)$, and comparing $(\mu+\alpha)[\,] = (4,2,4,2,2)[\,] = (3,0,1,-2,-3,-6,-7\ldots)$ with $\mu[\,] = (3,0,-1,-3,-5,-6,-7\ldots)$, we see that the entries of the former sequence that lie outside their respective save intervals are $(\mu+\alpha)[2] = 1 \notin I_2 = \{-1\}$ and $(\mu+\alpha)[4] = -3 \notin I_4 = \{-5,-4\}$. Since in fact $1 \in I_1 = \{0,1,2\}$ and $-3 \in I_3 = \{-3,-2\}$, the values $f(i) \neq i$ are $f(2) = 1$ and $f(4) = 3$. Therefore a term $s_{\mu+\alpha'}$ cancelling $s_{\mu+\alpha}$ can be obtained either by transposing the entries 0 and 1 at indices 1 and 2 of $(\mu+\alpha)[\,]$ or by transposing the entries -2 and -3 at indices 3 and 4. The result can be written as $(4,3,3,2,2)[\,]$ respectively as $(4,2,4,1,3)[\,]$, and indeed one has $s_{(4,3,3,2,2)} = s_{(4,2,4,1,3)} = -s_{(4,2,4,2,2)}$, while the occurring compositions can be written as $\mu+\alpha'$ for $\alpha' = (0,1,1,1,2), (0,0,2,0,3)$, respectively, so that both $s_{(4,3,3,2,2)}$ and $s_{(4,2,4,1,3)}$ occur in the summation. The rule of choosing the lexicographically first pair of indices would in fact make the term $s_{(4,2,4,2,2)}$ cancel against $s_{(4,3,3,2,2)}$.

As with proposition 2.4, propositions 2.3 and 2.5 lead to recurrence relations for certain scalar products. In particular we obtain such relations for the Kostka numbers:

$$K'_{\lambda/\mu,\alpha} = \langle\, s_\mu e_\alpha \mid s_\lambda \,\rangle = \sum_{\mu' \in \mathcal{P}_{|\mu|+\alpha_0}} [\, \mu \leftharpoonup \mu' \,]\, K'_{\lambda/\mu',(\alpha_1,\alpha_2,\ldots)}, \tag{23}$$

$$K_{\lambda/\mu,\alpha} = \langle\, s_\mu h_\alpha \mid s_\lambda \,\rangle = \sum_{\mu' \in \mathcal{P}_{|\mu|+\alpha_0}} [\, \mu \leftharpoondown \mu' \,]\, K_{\lambda/\mu',(\alpha_1,\alpha_2,\ldots)}, \tag{24}$$

for all $\lambda,\mu \in \mathcal{P}$ and $\alpha \in \mathcal{C}$. One can recursively apply these recurrence relations; using the basic case $K'_{\lambda/\lambda,\alpha} = K'_{\lambda/\lambda,\alpha} = [\,\alpha = (0)\,]$ and comparing with definition 1.2.2, one finds the promised combinatorial description of Kostka numbers.

2.6. Corollary. *For $\lambda,\mu \in \mathcal{P}$ and $\alpha \in \mathcal{C}$, the Kostka number $K_{\lambda/\mu,\alpha}$ equals the number $\#\,\mathrm{SST}(\lambda/\mu,\alpha)$ of semistandard tableaux of shape λ/μ and weight α, and $K'_{\lambda/\mu,\alpha}$ is the number of transposed semistandard tableaux of shape λ/μ and weight α. As a consequence the two are related by $K'_{\lambda/\mu,\alpha} = K_{\lambda^t/\mu^t,\alpha}$.* □

This description shows that one has $\#\,\mathrm{SST}(\lambda/\mu,\alpha) = \#\,\mathrm{SST}(\lambda/\mu,\sigma(\alpha))$ for any $\sigma \in \mathbf{S}_\infty$, and it turns equation (18) into a combinatorial expression of h_α and e_α in terms of Schur functions. In fact, we may consider h_α and e_α as generating polynomials in Λ of semistandard (respectively transposed semistandard) Young tableaux of weight α by shape, in the basis of Schur functions:

$$h_\alpha = \sum_{\lambda \in \mathcal{P}_{|\alpha|}} \sum_{T \in \mathrm{SST}(\lambda)} [\, \mathrm{wt}(T) = \alpha \,]\, s_\lambda, \tag{25}$$

$$e_\alpha = \sum_{\lambda \in \mathcal{P}_{|\alpha|}} \sum_{T \in \mathrm{SST}(\lambda^t)} [\, \mathrm{wt}(T) = \alpha \,]\, s_\lambda. \tag{26}$$

Besides the combinatorial description of corollary 2.6, one also has the following expressions for $K_{\lambda/\mu,\beta}$ and $K'_{\lambda/\mu,\beta}$ that can be obtained directly from lemma 2.2, using (10) and (9):

$$K_{\lambda/\mu,\beta} = \sum_{\alpha\in\mathcal{C}} \#\mathcal{M}_{\alpha,\beta}\langle\, s_{\mu+\alpha} \mid s_\lambda \,\rangle = \sum_{M\in\mathcal{M}} [\,\mathrm{col}(M)=\beta\,]\,\varepsilon(\mu+\mathrm{row}(M),\lambda), \qquad (27)$$

$$K'_{\lambda/\mu,\beta} = \sum_{\alpha\in\mathcal{C}} \#\mathcal{M}^{[2]}_{\alpha,\beta}\langle\, s_{\mu+\alpha} \mid s_\lambda \,\rangle = \sum_{M\in\mathcal{M}^{[2]}} [\,\mathrm{col}(M)=\beta\,]\,\varepsilon(\mu+\mathrm{row}(M),\lambda). \quad (28)$$

These expressions produce the Kostka numbers by an alternating enumeration of matrices rather than by a direct enumeration of tableaux, and are therefore rather inefficient as a means of computing those numbers. However it will prove useful to have, as a complement to corollary 2.6, these descriptions that do not involve semistandard tableaux, or horizontal or vertical strips.

§3. The Cauchy, Jacobi-Trudi and Von Nägelsbach-Kostka identities.

We have expressed products of elementary, of power sum, and of complete symmetric functions as $\mathbf{Z}$-linear combinations of Schur functions. However we have not yet expressed Schur functions themselves in terms of anything else, not even in terms of monomials, other than via the polynomial division used in their definition. In this section we shall provide such expressions of (skew) Schur functions, in terms of minimal symmetric functions (which implies an expression in terms of monomials), in terms of products h_α of complete symmetric functions, and in terms of products e_α of elementary symmetric functions. The latter two expressions will eventually be obtained via another application of lemma 2.2, but the expression in terms of minimal symmetric functions will be obtained in an indirect manner, by establishing that the dual basis of $\{\, m_\lambda \mid \lambda\in\mathcal{P}\,\}$ with respect to the scalar product on Λ is in fact the set $\{\, h_\lambda \mid \lambda\in\mathcal{P}\,\}$ (with the obvious correspondence between their elements, in other words $\langle\, m_\lambda \mid h_\mu \,\rangle = [\,\lambda=\mu\,]$). Another way of formulating that property is that scalar products $\langle\, f \mid h_\lambda \,\rangle$ can be interpreted as the coefficient of m_λ in the expression of f as linear combination of minimal symmetric functions, and hence that $\langle\, f \mid h_\alpha \,\rangle$ can be interpreted as the coefficient of X^α in the expansion of $f[X_\mathbf{N}]$ into monomials. This will imply in particular that $K_{\lambda/\mu,\alpha}$ can be interpreted as the coefficient of X^α in $s_{\lambda/\mu}[X_\mathbf{N}]$.

Let us first show that $\{\, h_\lambda \mid \lambda\in\mathcal{P}\,\}$ is a $\mathbf{Z}$-basis of Λ in the first place. Since its elements are precisely the monomials that can be formed from the set $\{\, h_i \mid i>0\,\}$, this amounts to showing that those h_i are algebraically independent generators of Λ, which can then be viewed as a polynomial ring $\mathbf{Z}[h_1,h_2,\ldots]$. One has $h_\mu = \sum_{\lambda\in\mathcal{P}_n} K_{\lambda,\mu} s_\lambda$ for $\mu\in\mathcal{P}_n$, while $\{\, s_\lambda \mid \lambda\in\mathcal{P}\,\}$ is known to be a $\mathbf{Z}$-basis of Λ, so we can obtain our goal by showing that the matrix $(K_{\lambda,\mu})_{\lambda,\mu\in\mathcal{P}_n}$ of Kostka numbers is invertible for every n. That will follow from the fact that this matrix is unitriangular for the dominance ordering on $\mathcal{P}_n$.

3.1. Proposition. *For $\lambda\in\mathcal{P}_n$ and $\alpha\in\mathcal{C}_n$, $K_{\lambda,\alpha}\neq 0$ implies $\alpha^+\leq\lambda$; moreover $K_{\lambda,\alpha}=1$ if $\alpha^+=\lambda$.*

Proof. Since $K_{\lambda,\alpha} = K_{\lambda,\alpha^+}$ we may assume $\alpha = \alpha^+ \in \mathcal{P}$. We use the description of $K_{\lambda,\alpha}$ in corollary 2.6. For any Young tableau $(\lambda^{(i)})_{i\in\mathbf{N}} \in \mathrm{SST}(\lambda)$, the diagram $[\lambda^{(k)}]$ is contained in the topmost k rows of $[\lambda]$, for all $k \in \mathbf{N}$. The existence of $(\lambda^{(i)})_{i\in\mathbf{N}} \in \mathrm{SST}(\lambda,\alpha)$ therefore implies $\sum_{i\in[k]} \alpha_i = |\lambda^{(k)}| \le \sum_{i\in[k]} \lambda_i$. If $\alpha = \lambda$, we see that $[\lambda^{(k)}]$ must coincide with the topmost k rows of $[\lambda]$, whence $\#\,\mathrm{SST}(\lambda,\lambda) = 1$. □

A combinatorial proof of this proposition for general $\alpha \in \mathcal{C}$, without using $K_{\lambda,\alpha} = K_{\lambda,\alpha^+}$, can also be given, but is slightly more awkward to formulate (one basically argues that the number of squares contributed by any column to the first k rows of $[\lambda]$ is at least as large as the contribution, in any $T \in \mathrm{SST}(\lambda,\alpha)$, of the k largest parts of α to that column). But we shall only use the case $\alpha \in \mathcal{P}$.

The proposition implies that the matrix $(K_{\lambda,\mu})_{\lambda,\mu\in\mathcal{P}_n}$ is upper unitriangular, and hence invertible, if $\mathcal{P}_n$ is ordered in such a way that λ comes before μ whenever $\lambda \ge \mu$, for instance by decreasing lexicographic order. Since, by corollary 2.6, $(K'_{\lambda,\mu})_{\lambda,\mu\in\mathcal{P}_n}$ is obtained from that matrix by a permutation of its rows (the permutation being given by the map $\lambda \mapsto \lambda^t\colon \mathcal{P}_n \to \mathcal{P}_n$), we may conclude from $e_\mu = \sum_{\lambda\in\mathcal{P}_n} K'_{\lambda,\mu} s_\lambda$ that Λ is also a polynomial ring $\mathbf{Z}[e_1, e_2, e_3 \ldots]$, a fact that is sometimes called the fundamental theorem of symmetric functions, and that is usually proved without using Schur functions.

To prove that the basis $\{\, h_\lambda \mid \lambda \in \mathcal{P} \,\}$ is dual to $\{\, m_\lambda \mid \lambda \in \mathcal{P} \,\}$, one may establish, for f running through some basis of Λ, that $f = \sum_{\lambda\in\mathcal{P}} \langle\, f \mid h_\lambda \,\rangle m_\lambda$ or equivalently that $f[X_\mathbf{N}] = \sum_{\alpha\in\mathcal{C}} \langle\, f \mid h_\alpha \,\rangle X^\alpha$. It is not practical to use the basis of Schur functions for this, since in that case those expressions are precisely what we would like to conclude from duality. However for $f = e_\beta$ or for $f = h_\beta$ we already know their expressions in terms of monomials, which are given in equations (9) and (10), while the scalar products that should coincide with the coefficients in those equations are easily computed. Indeed using equations (18) to develop these elements in the orthonormal basis of Schur functions one finds $\langle\, e_\beta \mid h_\alpha \,\rangle = \sum_{\lambda\in\mathcal{P}} K'_{\lambda,\beta} K_{\lambda,\alpha}$ and $\langle\, h_\beta \mid h_\alpha \,\rangle = \sum_{\lambda\in\mathcal{P}} K_{\lambda,\beta} K_{\lambda,\alpha}$. Thus we need to show the following.

3.2. Proposition. *For every $\alpha, \beta \in \mathcal{C}$ one has*

$$\#\mathcal{M}_{\alpha,\beta} = \sum_{\lambda\in\mathcal{P}} K_{\lambda,\alpha} K_{\lambda,\beta} \tag{29}$$

and

$$\#\mathcal{M}^{[2]}_{\alpha,\beta} = \sum_{\lambda\in\mathcal{P}} K_{\lambda,\alpha} K'_{\lambda,\beta}. \tag{30}$$

Proof. This is the one place where we import a result from elsewhere. The proposition expresses the enumerative consequences of the bijections between integer matrices and pairs of semistandard tableaux of equal shape defined in [Knu], the first of which is known as the RSK-correspondence. □

3.3. Corollary. *The $\mathbf{Z}$-bases $\{\, h_\lambda \mid \lambda \in \mathcal{P} \,\}$ and $\{\, m_\lambda \mid \lambda \in \mathcal{P} \,\}$ of Λ are dual to each other with respect to $\langle\,\cdot \mid \cdot\,\rangle$: for every $f \in \Lambda$ one has $f = \sum_{\lambda\in\mathcal{P}} \langle\, h_\lambda \mid f \,\rangle m_\lambda$, and consequently $f[X_\mathbf{N}] = \sum_{\alpha\in\mathcal{C}} \langle\, h_\alpha \mid f \,\rangle X^\alpha$.*

Proof. Since we have shown that $\{ h_\lambda \mid \lambda \in \mathcal{P} \}$ and $\{ e_\lambda \mid \lambda \in \mathcal{P} \}$ are $\mathbf{Z}$-bases of Λ, either of the special cases $f = h_\beta$ or $f = e_\beta$ that are given by proposition 3.2 suffice to prove the general case $f \in \Lambda$. $\square$

Applying this corollary to $f = s_{\lambda/\mu}$ gives the expressions that we were after, of (skew) Schur functions in terms of minimal symmetric function and in terms of monomials:

$$s_{\lambda/\mu} = \sum_{\nu \in \mathcal{P}} K_{\lambda/\mu,\nu} m_\nu, \tag{31}$$

$$s_{\lambda/\mu}[X_{\mathbf{N}}] = \sum_{\alpha \in \mathcal{C}} K_{\lambda/\mu,\alpha} X^\alpha = \sum_{T \in \mathrm{SST}(\lambda/\mu)} X^{\mathrm{wt}(T)}. \tag{32}$$

For $\mu = (0)$, the final expression describes $s_\lambda[X_{\mathbf{N}}]$ as the generating series in $\mathbf{Z}[[X_{\mathbf{N}}]]$ of semistandard Young tableaux of shape λ by weight. This is often taken as the definition of Schur functions (for instance in [Stan]), but in our setting it is a non-trivial statement that provides an alternative combinatorial interpretation of semistandard Young tableaux $T \in \mathrm{SST}(\lambda, \alpha)$: rather than corresponding to constituent Schur functions s_λ of h_α, they now correspond to constituent monomials X^α of $s_\lambda[X_{\mathbf{N}}]$. One may ask whether this result can be proved in a bijective manner. The way in which we obtained it does not qualify as a bijective proof: although the main ingredients (lemma 2.2 and proposition 3.2) are based on explicit correspondences, we did need some (linear) algebraic reasoning. The most direct possible bijective proof would be to give, for $f = \sum_{\alpha \in \mathbf{N}^n} \sum_{T \in \mathrm{SST}(\lambda,\alpha)} X^\alpha \in \mathbf{Z}[X_{[n]}]$, a correspondence between terms that establishes $f\Delta_n = a_{\delta_n+\lambda}[X_{[n]}]$. That would require a fairly complicated correspondence, but there is a simpler possibility that nevertheless is more or less equivalent: taking $f = \sum_{T \in \mathrm{SST}(\lambda)} X^{\mathrm{wt}(T)}$ and $\beta = (0)$ in lemma 2.2, its right hand side $\sum_{T \in \mathrm{SST}(\lambda)} s_{\mathrm{wt}(T)}$ should reduce by cancellations to s_λ, and one may seek to describe such a cancellation process explicitly. For instance for $\lambda = (2, 1)$ the right hand side, restricted to $\mathrm{wt}(T) \in \mathbf{N}^3$ thanks to proposition 2.1, gives $s_{(2,1)} + s_{(2,0,1)} + s_{(1,2)} + 2s_{(1,1,1)} + s_{(1,0,2)} + s_{(0,2,1)} + s_{(0,1,2)}$; here the terms $s_{(2,0,1)}$, $s_{(1,2)}$, and $s_{(0,1,2)}$ are null, and the remaining terms except the initial one can be cancelled using $0 = s_{(1,1,1)} + s_{(1,0,2)} = s_{(1,1,1)} + s_{(0,2,1)}$. We shall see below, when we consider products of Schur functions, that an explicit and general description of such a cancellation can indeed be given.

Now that we have identified the generating series of semistandard Young tableaux of a given shape by weight, we may turn the equations of proposition 3.2 into generating series identities by multiplying them by $X^\alpha Y^\beta$ and summing over $\alpha, \beta \in \mathcal{C}$; one obtains

$$\prod_{i,j \in \mathbf{N}} \frac{1}{1 - X_i Y_j} = \sum_{\lambda \in \mathcal{P}} s_\lambda[X_{\mathbf{N}}] s_\lambda[Y_{\mathbf{N}}] \tag{33}$$

and

$$\prod_{i,j \in \mathbf{N}} (1 + X_i Y_j) = \sum_{\lambda \in \mathcal{P}} s_\lambda[X_{\mathbf{N}}] s_{\lambda^t}[Y_{\mathbf{N}}]. \tag{34}$$

The first of these is called the Cauchy identity (although it is difficult to justify the attribution; for details see [Stan, p. 397]), and the second the dual Cauchy identity. From (12) and (33) one deduces (after combining terms in the left hand side of the former with a common value $\beta^+ = \lambda$) that

$$\sum_{\lambda\in\mathcal{P}} h_\lambda[X_\mathbf{N}]m_\lambda[Y_\mathbf{N}] = \sum_{\lambda\in\mathcal{P}} s_\lambda[X_\mathbf{N}]s_\lambda[Y_\mathbf{N}] \tag{35}$$

in $\mathbf{Z}[[X_\mathbf{N}, Y_\mathbf{N}]]$, which is an alternative formulation of the duality of corollary 3.3. Similarly (11) and (34) give

$$\sum_{\lambda\in\mathcal{P}} e_\lambda[X_\mathbf{N}]m_\lambda[Y_\mathbf{N}] = \sum_{\lambda\in\mathcal{P}} s_{\lambda^t}[X_\mathbf{N}]s_\lambda[Y_\mathbf{N}]. \tag{36}$$

This equation does not express a duality like the previous one, but one can relate the two equations using the $\mathbf{Z}$-linear involution $\omega\colon\Lambda\to\Lambda$ that maps $s_\lambda\mapsto s_{\lambda^t}$ for all $\lambda\in\mathcal{P}$. By (18), the fact that $K'_{\lambda,\alpha} = K_{\lambda^t,\alpha}$ implies $\omega(h_\alpha) = e_\alpha$ for all $\alpha\in\mathcal{C}$, so ω coincides with the ring morphism $\Lambda\to\Lambda$ that sends $h_i\mapsto e_i$ for all $i>0$ (which exists since $\Lambda = \mathbf{Z}[h_1, h_2, \ldots])$. Then (36) can be obtained from (35) by applying, to the subring of $\mathbf{Z}[[X_\mathbf{N}, Y_\mathbf{N}]]$ that is the the image of $\Lambda\otimes\Lambda$ via $f\otimes g\mapsto f[X_\mathbf{N}]g[Y_\mathbf{N}]$, the ring automorphism corresponding to the automorphism $\omega\otimes\mathbf{1}$ of $\Lambda\otimes\Lambda$. (Actually that is not quite true: one needs to take the closure of the subring in the topology of power series rings, and extend the automorphism to that closure by continuity.)

We now proceed to express skew Schur functions as $\mathbf{Z}$-linear combinations of elements h_α. Since duality of bases is a symmetric notion, the coefficient of h_ν in the expression of any $f\in\Lambda$ in the basis $\{\, h_\lambda \mid \lambda\in\mathcal{P}\,\}$ can be computed as $\langle\, m_\nu \mid f\,\rangle$. Applying this for $f = s_{\lambda/\mu}$, we see that its coefficient of h_ν is $\langle\, m_\nu \mid s_{\lambda/\mu}\,\rangle = \langle\, s_\mu m_\nu \mid s_\lambda\,\rangle$. This number can be determined by applying lemma 2.2 for m_ν (a case that we did not yet exploit before), which gives as result the expression $\sum_{\alpha\in\mathcal{C}}[\,\alpha^+ = \nu\,]\,\varepsilon(\mu+\alpha,\lambda)$. This shows that

$$s_{\lambda/\mu} = \sum_{\nu\in\mathcal{P}}\sum_{\alpha\in\mathcal{C}}[\,\alpha^+ = \nu\,]\,\varepsilon(\mu+\alpha,\lambda)h_\nu = \sum_{\alpha\in\mathcal{C}}\varepsilon(\mu+\alpha,\lambda)h_\alpha, \tag{37}$$

since $h_\nu = h_\alpha$ when $\alpha^+ = \nu$. This achieves the expression of skew Schur functions in terms of complete symmetric functions, but we shall proceed to make the expression more transparent. The process will be somewhat different from what we have seen before, as we shall not actually cancel any pairs of terms, nor even group together terms for values of h_α that are equal.

The nonzero terms of the final summation of (37) can be found by permuting the terms of $\lambda[\,]$ in all possible ways such that each term of the resulting sequence $\sigma(\lambda[\,])$ is at least as large as the corresponding term of $\mu[\,]$ (in formula, $\lambda[\sigma^{-1}(i)]\geq\mu[i]$ for all $i\in\mathbf{N}$), which condition we shall abbreviate as $\sigma(\lambda[\,])\geq\mu[\,]$; then $\alpha = \sigma(\lambda[\,]) - \mu[\,]$ gives a contribution $\varepsilon(\sigma)h_\alpha$. From the proof of proposition 2.1 we see that if $\lambda\in\mathbf{N}^l$, then one needs to consider only permutations $\sigma \in \mathbf{S}_l$. For instance for $\lambda/\mu = (5,4,2)/(3,1)$, only permutations of the 3 initial terms of $\lambda[\,] = (4,2,-1,-4,\ldots)$ need to be considered, and only the permutations giving $(4,2,-1)$, $(2,4,-1)$, $(4,-1,2)$, and $(2,-1,4)$ satisfy $\sigma(\lambda[\,]) \geq \mu[\,] = (2,-1,-3,\ldots)$; for this case (37) therefore gives $s_{(5,4,2)/(3,1)} = h_{(2,3,2)} - h_{(0,5,2)} - h_{(2,0,5)} + h_{(0,0,7)} = h_{(3,2,2)} - 2h_{(5,2)} + h_7$. We can now write (37) as

$$s_{\lambda/\mu} = \sum_{\sigma \in \mathbf{S}_l} [\,\sigma(\lambda[\,]) \geq \mu[\,]\,]\ \varepsilon(\sigma)\, h_{\sigma(\lambda[\,])-\mu[\,]}. \tag{38}$$

With the convention that $h_i = 0$ when $i < 0$ (which will take care of the factor $[\,\sigma(\lambda[\,]) \geq \mu[\,]\,]$), we can write this expression as the determinant of a matrix whose entries are complete symmetric functions:

$$s_{\lambda/\mu} = \det\bigl(h_{\lambda[i]-\mu[j]}\bigr)_{i,j\in[l]} \qquad \text{when } \lambda \in \mathbf{N}^l. \tag{39}$$

This identity is called the Jacobi-Trudi identity. For the example above, it gives

$$s_{(5,4,2)/(3,1)} = \begin{vmatrix} h_2 & h_5 & h_7 \\ h_0 & h_3 & h_5 \\ 0 & h_0 & h_2 \end{vmatrix} = h_{(3,2,2)} - 2h_{(5,2)} + h_7.$$

To obtain an expression of skew Schur functions in terms of elementary symmetric functions, we may apply the automorphism ω to equation (37), giving

$$s_{\lambda^t/\mu^t} = \sum_{\alpha \in \mathcal{C}} \varepsilon(\mu+\alpha,\lambda) e_\alpha, \tag{40}$$

from which one deduces as above, again with the convention that $e_i = 0$ when $i < 0$, that

$$s_{\lambda^t/\mu^t} = \det\bigl(e_{\lambda[i]-\mu[j]}\bigr)_{i,j\in[l]} \qquad \text{when } \lambda \in \mathbf{N}^l, \tag{41}$$

or equivalently

$$s_{\lambda/\mu} = \det\bigl(e_{\lambda^t[i]-\mu^t[j]}\bigr)_{i,j\in[r]} \qquad \text{when } \lambda_0 \leq r. \tag{42}$$

This final equation is sometimes referred to as the dual Jacobi-Trudi identity, but we prefer to call it the Von Nägelsbach-Kostka identity (following Littlewood, see [Litt, p. 87], but restoring the spelling that had been lost in translation). We note that Kostka proved the identity in [Kos] by an ingenious manipulation of determinants representing alternants, see [Muir, p. 155]; see also [Stan, p. 397].

§4. The Gessel-Viennot correspondences.

Let us compare equation (27), which expresses the Kostka number $K_{\lambda/\mu,\beta}$ as an alternating sum over integral matrices with column sums given by β, with the description of $K_{\lambda/\mu,\beta}$ as counting semistandard tableaux of shape λ/μ and weight β. The integral encodings M of such tableaux (given in definition 1.2.3) are also matrices with $\mathrm{col}(M)=\beta$, and since moreover $\mathrm{row}(M)=\lambda-\mu$, these matrices contribute $+1$ to the alternating sum in (27). Therefore these integral encodings give a subset of its terms that completely accounts for the value of the alternating sum; apparently all remaining terms cancel out. We can combine the observed identity for all β into a generating series identity

$$\sum_{\beta\in\mathcal{C}} K_{\lambda/\mu,\beta}X^\beta = \sum_{M\in\mathcal{M}} \varepsilon(\mu+\mathrm{row}(M),\lambda)X^{\mathrm{col}(M)} = \sum_{M\in\mathrm{Tabl}(\lambda/\mu)} X^{\mathrm{col}(M)}, \tag{43}$$

where the second member is obtained from (27), and the third member from corollary 2.6 by taking integral encodings. In this section we shall discuss the reduction of the second member to the third member by cancellations, and a similar question for an identity derived from equation (28).

We know from (32) that the first and last members of (43) equal $s_{\lambda/\mu}[X_\mathbf{N}]$, and the cancellation that we consider is often associated with the Jacobi-Trudi identity. Indeed, if one takes equation (10), slightly rewritten as $h_\alpha[X_\mathbf{N}]=\sum_{M\in\mathcal{M}}[\,\mathrm{row}(M)=\alpha\,]\,X^{\mathrm{col}(M)}$, and substitutes it into (37), then one obtains

$$s_{\lambda/\mu}[X_\mathbf{N}] = \sum_{M\in\mathcal{M}} \varepsilon(\mu+\mathrm{row}(M),\lambda)X^{\mathrm{col}(M)}, \tag{44}$$

whose second member matches that of (43). We note however that we did not need (32) or (37) to obtain (43).

In fact, the way we obtained (43) suggests a way to understand how to cancel terms in it. Its second member was originally obtained from an application of lemma 2.2 for multiplication by h_β, while its third member was obtained by repeated application of proposition 2.5 for individual multiplications by factors h_{β_j}, each of which uses lemma 2.2 followed by a pruning of the summation. It suffices to postpone such intermediate pruning. Stated differently, for each term of the second member that needs to be cancelled, one needs to find the factor h_{β_j} responsible for this fact, and cancel the term correspondingly.

Let $M\in\mathcal{M}$ with $\mathrm{col}(M)=\beta$. Each column M_j^t represents the exponent of a term in the factor h_{β_j} of h_β, so that the entire matrix represents the monomial $X^{\mathrm{row}(M)}$ in the expansion of h_β. If we have $M\in\mathrm{Tabl}(\lambda/\mu)$, which means that M is the integral encoding of some semistandard tableau T of shape λ/μ, then the sequence of partitions $(\lambda^{(l)})_{l\in\mathbf{N}}$ that defines T can be found by starting with μ and successively adding the compositions M_j^t for increasing values of j: $\lambda^{(l)}=\mu+\sum_{j<l}M_j^t$. Therefore the condition that $M\in\mathrm{Tabl}(\lambda/\mu)$ amounts to the requirement that $\mu+\sum_{j<l}M_j^t \leftharpoondown \mu+\sum_{j\leq l}M_j^t$ holds for all $l\in\mathbf{N}$, and that moreover one has $\mu+\mathrm{row}(M)=\lambda$.

Now suppose that $M \notin \mathrm{Tabl}(\lambda/\mu)$, so that M defines a term of the second member of (43) that has to be cancelled. Should only the final part $\mu + \mathrm{row}(M) = \lambda$ of the condition $M \in \mathrm{Tabl}(\lambda/\mu)$ fail, then $\mu + \mathrm{row}(M)$ is a partition distinct from λ (recall that $\alpha \leftharpoondown \alpha'$ implies that $\alpha, \alpha' \in \mathcal{P}$), which means that $\varepsilon(\mu + \mathrm{row}(M), \lambda) = 0$, and the term defined by M is null (this case only arises because (43) just deals with the coefficient $\langle\, h_\beta \mid s_{\lambda/\mu} \,\rangle = \langle\, s_\mu h_\beta \mid s_\lambda \,\rangle$ rather than with all of $s_\mu h_\beta$).

Putting that case aside, there must be an index l for which $\mu + \sum_{j<l} M_j^t \not\leftharpoondown \mu + \sum_{j\leq l} M_j^t$, and we choose the minimal such l; this means that in the computation where s_μ is successively multiplied by the factors h_{β_j}, the term determined by M is cancelled after the multiplication by h_{β_l}. At that point the term is s_α with $\alpha = \mu + \sum_{j\leq l} M_j^t$, and it is cancelled against a term $s_{\alpha'}$ defined similarly, but possibly with a different composition instead of the final term M_l^t of the summation (as usual we do not exclude the possibility $\alpha' = \alpha$, if already $s_\alpha = 0$). From the proof of proposition 2.5 we recall that α' is such that the sequence $\alpha'[\,]$ is obtained from $\alpha[\,]$ by interchanging its entries at appropriately chosen distinct indices i_0 and i_1. Those indices were chosen in a manner that ensures that the value that should replace M_l^t in the expression for α' is a composition, i.e., it has no negative coefficients. That composition is in fact obtained from M_l^t by a transfer between the entries $M_{i_0,l}$ and $M_{i_1,l}$, by which we mean the replacement of those entries by some pair of natural numbers with the same sum.

We have considered M truncated to its columns M_j^t with $j \leq l$, and found another such partial matrix such that their contributions to $s_\mu h_{\beta_0} \cdots h_{\beta_l}$ cancel. It is clear algebraically that the total contribution to $s_\mu h_\beta$ from matrices that extend the first partial matrix cancels the total contribution from matrices that extend the other partial matrix. However, to describe the cancellation in (43) combinatorially, we have to match up the term for M with a specific term cancelling it, taking into account also the columns M_j^t with $j > l$. The obvious way to do this is to exchange, in addition to the transfer between $M_{i_0,l}$ and $M_{i_1,l}$, the entire remainders of rows i_0 and i_1 to the right of those entries; in this way it is guaranteed that the values of $(\mu + \mathrm{row}(M))[i_0]$ and $(\mu + \mathrm{row}(M))[i_1]$ are interchanged between a pair of matching matrices.

Here is a simple example to illustrate the construction of cancelling matrices. Let $\lambda/\mu = (7,5,4)/(1)$ and

$$M = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 2 \\ 3 & 1 & 0 \end{pmatrix},$$

so that $\beta = \mathrm{col}(M) = (6,4,5)$. Then one already has $\mu + \sum_{j<l} M_j^t \not\leftharpoondown \mu + \sum_{j\leq l} M_j^t$ for $l = 0$, the concrete values being $\mu = (1) \not\leftharpoondown (2,2,3) = \mu + M_0^t$. Setting $\alpha = \mu + M_0^t$ one has $\alpha[\,] = (1,0,0,-4,\ldots)$, and the indices (i_0, i_1) can be chosen in various ways. For instance for $(i_0,i_1) = (0,1)$ one finds the sequence $\alpha'[\,] = (0,1,0,-4,\ldots)$ for $\alpha' = (1,3,3)$, which can be realised by replacing M_0^t by $(0,3,3)$. But one can also take $(i_0,i_1) = (1,2)$, which permutes two identical entries of $\alpha[\,]$ and therefore gives $\alpha'' = \alpha$, and an unchanged column M_0^t (and indeed one has $s_{(2,2,3)} = 0$). Performing the indicated changes to M, we find that the contribution to $s_\mu h_\beta$ from M could be cancelled by either one of the

contributions from the matrices

$$M' = \begin{pmatrix} 0 & 1 & 2 \\ 3 & 2 & 3 \\ 3 & 1 & 0 \end{pmatrix} \qquad \text{or} \qquad M'' = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 0 \\ 3 & 1 & 2 \end{pmatrix}.$$

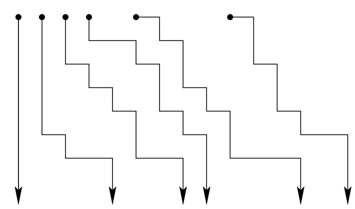
Indeed one has $s_{(7,5,4)} = -s_{(4,8,4)} = -s_{(7,3,6)}$ (the subscripts were respectively computed as $\mu+\mathrm{row}(M)$, $\mu+\mathrm{row}(M')$, and $\mu+\mathrm{row}(M'')$), which shows that the contribution of M to the second member of (43) could cancel against that of M' or of M''; the rule of choosing the minimal possible (i_0, i_1) would in fact select M'. Apart from the operations involved in the cancellation, the example illustrates the fact that when we cancel s_α against $s_{\alpha'}$ with $\alpha \neq \alpha'$, one might still have $s_\alpha = 0$; it also shows that cancelling a null term s_α in $s_\mu h_{\beta_0} \cdots h_{\beta_l}$ "against itself" might lead to the cancellation of two distinct terms in $s_\mu h_\beta$.

There is a beautiful graphical way of visualising this cancellation, and the terms that survive it, due to Gessel and Viennot [GeVi]. Set $\alpha^{(j)} = \mu + \sum_{j'<j} M^t_{j'}$ for $j \in \mathbf{N}$, so that M identifies the term $s_{\alpha^{(j)}}$ in the development of $s_\mu h_{\beta_0} \cdots h_{\beta_{j-1}}$. The idea is to trace for each fixed index i the evolution of the number $\alpha^{(j)}[i]$, as j increases, by means of a lattice path: a path connecting points of $\mathbf{Z}^2$ using unit steps in two orthogonal directions, in the current case down and to the right. One has a separate path for each $i \in \mathbf{N}$; path i passes through all of the points $\bigl(j, \alpha^{(j)}[i]\bigr)$ for $j \in \mathbf{N}$. These points are connected as follows: after arriving at $\bigl(j, \alpha^{(j)}[i]\bigr)$, the path takes $M_{i,j}$ horizontal steps (so the second coordinate becomes $\alpha^{(j)}[i] + M_{i,j} = \alpha^{(j+1)}[i]$), followed by a vertical step to arrive at $\bigl(j+1, \alpha^{(j+1)}[i]\bigr)$. If $i > 0$, this portion of path i remains disjoint from path $i-1$ if and only if $\alpha^{(j+1)}[i] < \alpha^{(j)}[i-1]$, in other words if the addition of $M_{i,j}$ to $\alpha^{(j)}[i]$ leaves it inside what we called the safe interval I_i in the proof of proposition 2.5. Thus all paths will remain disjoint when taking into account column M^t_j if and only if $\alpha^{(j)} \leftharpoondown \alpha^{(j)} + M^t_j = \alpha^{(j+1)}$; they will remain disjoint forever if and only if $M \in \mathrm{Tabl}(\lambda'/\mu)$ for some $\lambda' \in \mathcal{P}$.

For any M, path i starts at $(0, \mu[i])$ and eventually runs along the vertical line $\bigl(*, (\mu + \mathrm{row}(M))[i]\bigr)$. The term for M in the second member of (43) is nonzero if and only if these vertical lines, for $i \in \mathbf{N}$, form a permutation of the lines $(*, \lambda[i])_{i\in\mathbf{N}}$, whose sign then gives the coefficient of the term. If moreover $\mu + \mathrm{row}(M)$ is a partition, which must necessarily be λ given the previous condition, then this permutation will be the identity, so that path i runs from the point $(0, \mu[i])$ to the line $(*, \lambda[i])$. This is in particular (but not exclusively) the case whenever the paths are disjoint, so that $M \in \mathrm{Tabl}(\lambda/\mu)$.

As an example of a term that survives the cancellation, and of the family of disjoint paths that corresponds to it, we recall from §1 the integral encoding $M \in \mathrm{Tabl}(\lambda/\mu)$ of the tableau $T \in \mathrm{SST}(\lambda/\mu)$ of (2), where $\lambda/\mu = (9,8,5,5,3)/(4,1)$. To facilitate recognition of the compositions (in fact partitions) $\alpha^{(j)}$, we have written the coefficients of μ along the left border of M. The paths, taken from right to left, correspond to the rows of the matrix from top to bottom, and to the rows of T in the same order.

$$T:\quad \begin{array}{ccccccccc} &&&&0&2&4&5&5\\ &0&1&3&4&6&6&6&\\ 1&1&2&4&5&&&&\\ 2&3&4&6&6&&&&\\ 5&6&6&&&&&& \end{array} \qquad \begin{array}{c} \mu\\ 4\\1\\0\\0\\0\\0 \end{array} \begin{array}{c} M\\ \begin{pmatrix} 1&0&1&0&1&2&0&0\\ 1&1&0&1&1&0&3&0\\ 0&2&1&0&1&1&0&0\\ 0&0&1&1&1&0&2&0\\ 0&0&0&0&0&1&2&0\\ 0&0&0&0&0&0&0&0 \end{pmatrix} \end{array}$$



On the other hand, for the terms that must be cancelled, we have that whenever two paths of the associated family touch, then a family of paths with opposite contribution can be formed by exchanging these two paths after the point of contact; if the term is nonzero, this exchange will change the sign of the permutation defined by the family. This exchange corresponds to the type of cancellation we considered above. Indeed if these are paths i_0 and i_1, and they have point (j, m) in common, then

$$\max(\alpha^{(j)}[i_0], \alpha^{(j)}[i_1]) \leq m \leq \min(\alpha^{(j+1)}[i_0], \alpha^{(j+1)}[i_1]), \tag{45}$$

and the exchange corresponds to transferring the difference $\alpha^{(j+1)}[i_0] - \alpha^{(j+1)}[i_1]$ from $M_{i_0,j}$ to $M_{i_1,j}$ (which leaves both entries ≥ 0 due to the inequality between the outer members of (45)), followed by the interchange for all $j' > j$ of the entries $M_{i_0,j'}$ and $M_{i_1,j'}$.

Looking back at proposition 2.5 from this path point of view, we see that in its proof there is even more freedom in matching cancelling terms than we used: the existence of a common point on paths i_0 and i_1 is all that is needed in order to obtain a cancellation by a transfer between parts i_0 and i_1, while we used the stricter requirement $f(i_0) = f(i_1)$. (To make this clear, recall that the function f was defined relative to the multiplication by a single factor h_j and under the assumption that $\alpha^{(j)}$ is a partition, by the condition $\alpha^{(j+1)}[i] \in I_{f(i)}$, where $I_0 = \{\, k \in \mathbf{Z} \mid \alpha^{(j)}[0] \leq k \,\}$ and $I_i = \{\, k \in \mathbf{Z} \mid \alpha^{(j)}[i] \leq k < \alpha^{(j)}[i-1] \,\}$ for $i > 0$. Then for $i_1 > i_0$ there is a common point with vertical coordinate j on paths i_0 and i_1 if and only if $i_0 \geq f(i_1)$, which is certainly the case if $f(i_1) = f(i_0)$.)

This extra freedom does not increase the number of terms that cancel, it just gives more possibilities of choosing (i_0, i_1) for those terms that do allow cancellation. Of course, to complete a proof that the second member of (43) reduces to the third member, one must give up this freedom and make a concrete choice that defines a sign-reversing involution on the set of cancelling terms. A fairly natural choice to do so is to choose the top-leftmost point that lies on more than one path, in other words take j minimal in the discussion above and then $m = \alpha^{(j)}[i_0]$ where $i_0 = \max\{\, i \mid \alpha^{(j+1)}[i+1] \geq \alpha^{(j)}[i] \,\}$;

then $i_1 = i_0 + 1$, and only paths i_0 and i_1 pass through (j, m), which makes it easy to show that after exchange of paths this rule finds the same point (j, m) and the same pair (i_0, i_1). Indeed the fact that we took some pair (i_0, i_1) with $f(i_1) = f(i_0)$ in the proof of proposition 2.5 seems a bit artificial in retrospect (regardless of exactly which pair was chosen); in particular there may be no such pair with $i_1 = i_0 + 1$, so the common point used need not be the first one where path i_1 meets another path.

One can proceed analogously for equation (28), comparing it with the description of $K'_{\lambda/\mu,\beta}$ as counting transpose semistandard tableaux of shape λ/μ and weight β. We have not defined binary encodings of such tableaux T (or, for that matter, integral encodings), but if we take as columns differences of successive partitions of T, we obtain a binary matrix M whose transpose is the binary encoding of the transpose of T. The situation is in fact quite similar to the one considered above for (27), as we have $\mathrm{col}(M) = \beta$ and $\mathrm{row}(M) = \lambda - \mu$, whence M contributes $+1$ to the summation in (28), and all such matrices account for the entire value of the summation. Forming a generating series by β, we obtain

$$\sum_{\beta\in\mathcal{C}} K'_{\lambda/\mu,\beta} X^\beta = \sum_{M\in\mathcal{M}^{[2]}} \varepsilon(\mu + \mathrm{row}(M), \lambda) X^{\mathrm{col}(M)} = \sum_{M^t\in\mathrm{Tabl}^{[2]}(\lambda^t/\mu^t)} X^{\mathrm{col}(M)}. \tag{46}$$

The first equality is related to the Von-Nägelsbach-Kostka identity, since the first member of the equation equals $s_{\lambda^t/\mu^t}[X_{\mathbf{N}}]$ by (32), and the second member can be obtained by substituting (9) into (40).

Before we consider the reduction of the second member of (46) to its third member by cancellations, we transform the equation slightly by taking transposes of tableaux and matrices, so that we will be dealing with ordinary semistandard tableaux and their binary encodings. At the same time we replace λ/μ by λ^t/μ^t so that the members of the equation represent $s_{\lambda/\mu}[X_{\mathbf{N}}]$ rather than $s_{\lambda^t/\mu^t}[X_{\mathbf{N}}]$; we get

$$\sum_{M\in\mathcal{M}^{[2]}} \varepsilon(\mu^t + \mathrm{col}(M), \lambda^t) X^{\mathrm{row}(M)} = \sum_{M\in\mathrm{Tabl}^{[2]}(\lambda/\mu)} X^{\mathrm{row}(M)}. \tag{47}$$

Let $M \in \mathcal{M}^{[2]}$ with $\mathrm{row}(M) = \alpha$. Each row M_i represents a monomial X^{M_i} in the factor e_{α_i} of e_α, and the entire matrix represents an occurrence of the monomial $X^{\mathrm{col}(M)}$ in the expansion of e_α. Put $\beta^{(i)} = \mu^t + \sum_{i'<i} M_{i'}$ for $i \in \mathbf{N}$, then $M \in \mathrm{Tabl}^{[2]}(\lambda/\mu)$ means that $\beta^{(i)} \in \mathcal{P}$ for all i (the remaining requirements for the relations $\beta^{(i)} \leftharpoondown \beta^{(i+1)}$ are automatically satisfied) and that $\beta^{(i)} = \lambda^t$ for i sufficiently large (in other words $\mu^t + \mathrm{col}(M) = \lambda^t$). As i increases, M identifies a term $s_{\beta^{(i)}}$ in the development of $s_{\mu^t} e_{\alpha_0} \cdots e_{\alpha_{i-1}}$, until possibly at some point $\beta^{(i)} \notin \mathcal{P}$, in which case $s_{\beta^{(i)}} = 0$. If this never happens, then M also identifies a term $s_{\mu^t+\mathrm{col}(M)}$ in the expression of the final value $s_{\mu^t} e_\alpha$ in the basis of Schur functions. In this case M gives equal contributions to both members of (47), which contributions are nonzero if and only if $\mu^t + \mathrm{col}(M) = \lambda^t$ or equivalently $M \in \mathrm{Tabl}^{[2]}(\lambda/\mu)$ (the contributions are zero when $M \in \mathrm{Tabl}^{[2]}(\lambda'/\mu)$ for some partition $\lambda' \neq \lambda$; as before this case needs no special consideration).

Now suppose to the contrary that one does have $s_{\beta^{(i)}} = 0$ for some i, and fix i to the minimal such value. We see that the situation is somewhat simpler than before, in that there is no need to modify row M_i to find a matching term that accounts for the cancelling of this contribution to $s_{\mu^t} e_{\alpha_0} \cdots e_{\alpha_{i-1}}$. However the contribution $s_{\mu^t + \mathrm{col}(M)}$ to $s_{\mu^t} e_\alpha$ might still be nonzero, and in order to find a matrix with a cancelling contribution (whose rows up to and including M_i will be the same as those of M), one needs to choose a column index j that "causes" the vanishing of $s_{\beta^{(i)}}$. Concretely, the fact that $s_{\beta^{(i)}} = 0$, while $\beta^{(i)}$ was obtained by adding the binary composition M_i to the partition $\beta^{(i-1)}$, implies that the sequence $\beta^{(i)}[\,]$ has equal adjacent entries at some position (perhaps at several different positions). So choose some j for which $\beta^{(i)}[j] = \beta^{(i)}[j+1]$, for instance the minimal one. Then to find a matrix whose term cancels that of M, we may interchange the parts of columns j and $j+1$ beyond row M_i; the cancellation is assured by the fact that this modification causes the interchange of the entries at positions j and $j+1$ of the sequence $(\mu^t + \mathrm{col}(M))[\,]$.

A simple example will make this more concrete. We take $\lambda/\mu = (7,5,2)/(2)$ and

$$M = \begin{pmatrix} 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 \end{pmatrix},$$

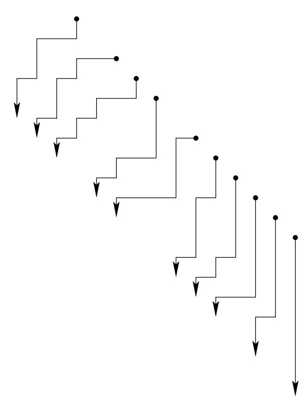
so that $\alpha = \mathrm{row}(M) = (2,4,2,4)$. One has $\beta^{(0)} = \mu^t = (1,1)$ and $\beta^{(1)} = (1,1,1,1)$ which is still a partition, but $\beta^{(2)} = (2,1,2,1,1,0,1)$ is not, and indeed $s_{\beta^{(2)}} = 0$. As cause for the vanishing of $s_{\beta^{(2)}}$ one can indicate either of the two repetitions in the sequence $\beta^{(2)}[\,] = (+1,-1,-1,-3,-4,-6,-6,-8,-9,\ldots)$, and interchanging either columns 1 and 2 or columns 5 and 6 beyond row 1 leads to two matrices whose contribution could cancel against that of M, namely

$$M' = \begin{pmatrix} 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 1 & 0 \end{pmatrix} \qquad \text{and} \qquad M'' = \begin{pmatrix} 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 \end{pmatrix}.$$

Indeed one has $s_{(3,3,2,1,3,1,1)} = -s_{(3,1,4,1,3,1,1)} = -s_{(3,3,2,1,3,0,2)}$ (the subscripts were respectively computed as $\mu^t + \mathrm{col}(M)$, $\mu^t + \mathrm{col}(M')$, and $\mu^t + \mathrm{col}(M'')$), which shows that the contribution of M to the left hand side of (47) could cancel against that of M' or of M''; since $\varepsilon((3,3,2,1,3,1,1), \lambda^t) = -1$, the former contribution is in fact $-X^{(2,4,2,4)}$, and the latter are both $+X^{(2,4,2,4)}$. The rule suggested above of choosing j would pair up M with M'.

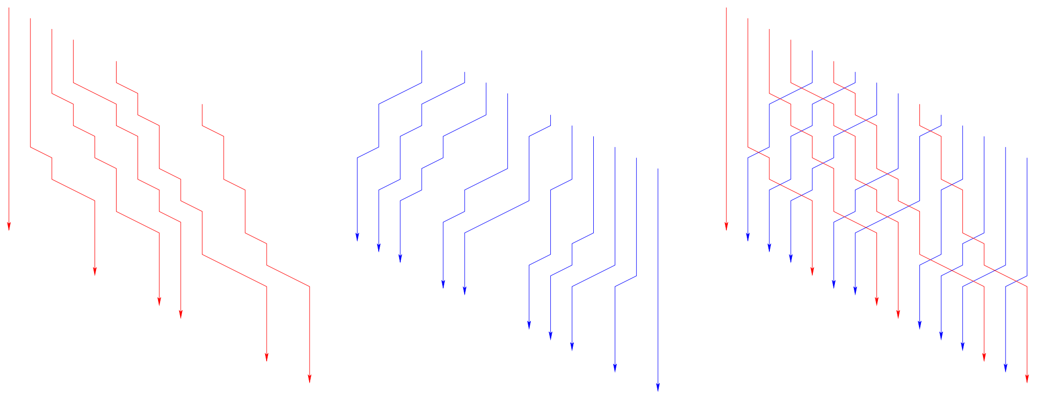
Again we can visualise the cancellation using lattice paths that trace the evolution of the individual terms of the sequence $\beta^{(i)}[\,]$ as i increases. In order that all possible paths (subject only to the condition of eventually running in a prescribed direction) can be obtained, each bit $M_{i,j}$ should directly control the direction of segment i of path j, choosing between two possible directions (note the difference with the correspondence above for integral matrices, where $M_{i,j}$ determined the number of steps to the right on path i at horizontal level j). For reasons that will become clear below we choose the possible directions in the binary case to be downwards (for bits 0) and leftwards (for bits 1). We want the points on all paths that represent terms of the same sequence $\beta^{(i)}[\,]$ to line up, and given the directions in which the paths progress, this alignment must be along diagonals in $\mathbf{Z}^2$. Thus we arrive at the following rule: path j starts at the point $(-\mu^t[j], -\mu^t[j])$ on the main diagonal, and passes through the point $(i-\beta^{(i)}[j], -\beta^{(i)}[j])$, which lies i diagonals below the main diagonal, for each $i \in \mathbf{N}$. If for $j \neq j'$ paths j and j' have a point in common, then this happens for the same value of i, and it means that $\beta^{(i)}[j] = \beta^{(i)}[j']$. All paths will eventually run straight downwards, with path j running along the vertical line $(*, -(\mu^t + \mathrm{col}(M))[j])$. The matrix M will give a nonzero contribution to the left hand side of (47) if and only if those lines for $j \in \mathbf{N}$ are a permutation of the lines $(*, -\lambda^t[j])$, and M will then give a nonzero contribution to the right hand side of (47) if and only if the paths remain disjoint, in which case the permutation must be the identity, so that $\mu^t + \mathrm{col}(M) = \lambda^t$. We illustrate such a family of disjoint paths corresponding to the same semistandard tableau T that was used before, together with its binary encoding M. Note that columns of T correspond to those of M, and to the paths, all in the same order from left to right.

$$T: \quad \begin{array}{ccccccccc}
 & & & & \boxed{0} & \boxed{2} & \boxed{4} & \boxed{5} & \boxed{5} \\
 & \boxed{0} & \boxed{1} & \boxed{3} & \boxed{4} & \boxed{6} & \boxed{6} & \boxed{6} & \\
\boxed{1} & \boxed{1} & \boxed{2} & \boxed{4} & \boxed{5} & & & & \\
\boxed{2} & \boxed{3} & \boxed{4} & \boxed{6} & \boxed{6} & & & & \\
\boxed{5} & \boxed{6} & \boxed{6} & & & & & &
\end{array}
\qquad
\begin{array}{cc}
\mu^t & \begin{array}{cccccccccc} 2 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \\
M & \begin{pmatrix}
0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\
0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}
\end{array}$$



For families of paths that are not disjoint, cancellation can be realised as before by choosing a pair of paths that meet, and exchanging their parts beyond the first point of contact; for the matrices this means exchanging the lower parts of the corresponding columns.

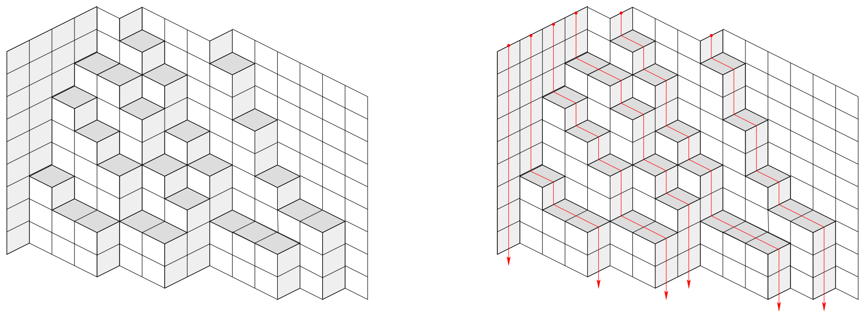
Now that we have two families of lattice paths related to any semistandard tableau T, a natural question is how the two are related. It turns out that with a slight transformation the two families can be aligned so that there is exactly one crossing of members of the two families for every square of $[\lambda/\mu]$. We show the two families shown above after transformation, first individually and then in superposition.



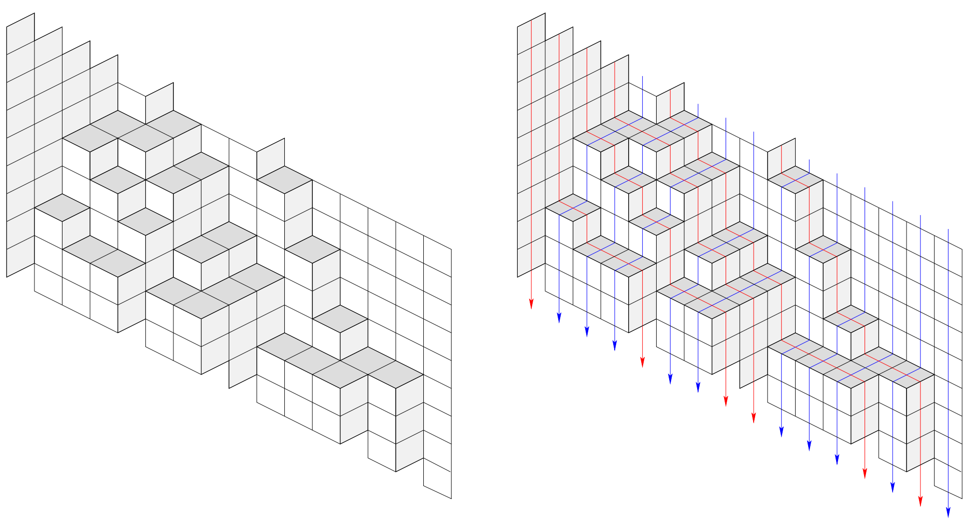
We have in fact, for aesthetic reasons, prefixed vertical segments of one unit to all paths of the first family, and of half a unit to all paths of the second family. The combinatorial explanation for this nice superposition is that, if T is given by the sequence of partitions $(\lambda^{(k)})_{k\in\mathbf{N}}$, then the first family shows the evolution of the sets $\{\,\lambda^{(k)}[i] \mid i \in \mathbf{N}\,\}$ as k traverses $\mathbf{N}$, and the second family shows the evolution of the complementary sets $\{\,-1-(\lambda^{(k)})^t[j] \mid j \in \mathbf{N}\,\}$ (the term -1 needed for complementarity was left out of the coordinates of the second family for simplicity, but is taken into account in the superposition). Indeed one finds for matrices that are integral and binary encodings of a semistandard tableau $T = (\lambda^{(k)})_{k\in\mathbf{N}}$ of shape λ/μ, that in our descriptions above $\alpha^{(j)} = \lambda^{(j)}$, respectively that $\beta^{(i)} = (\lambda^{(i)})^t$.

For the first family of paths, the values $\lambda^{(k)}[i]$ are most clearly read off at the middle of the vertical segment from $(k-1, \lambda^{(k)}[i])$ to $(k, \lambda^{(k)}[i])$, while for the second family of paths, the values of $-1-(\lambda^{(k)})^t[j]$ can be read off on a diagonal passing *through* the vertices; this explains the necessity of half a unit of vertical displacement between the triangular lattices that are the images of the lattice $\mathbf{Z}^2$ under the two transformations, in order to superimpose the lines that correspond to the same sequence $\lambda^{(k)}[\,]$.

This section would not be complete without a *visual* explanation of the superposition of families of paths. The form of path i of the first family describes the evolution of row i, with i fixed, of the Young diagrams $[\lambda^{(k)}]$ for $k \in \mathbf{N}$, so it has the same shape as the path along the boundary of a single "row" (i.e., subset with first coordinate fixed) of the 3-dimensional "block diagram" corresponding to the semistandard tableau, namely $\{\,(i,j,k) \in \mathbf{N}^3 \mid (i,j) \in [\lambda^{(k)}]\,\}$ in which each element is represented by a cube. To match the way we drew the lattice paths, the coordinate k must be made to increase downwards, and the coordinate j to the bottom right. With the coordinate i increasing to the bottom left, we obtain, for the tableau T of our example, the display below. The first picture shows just the block diagram, with the top of the "pillars" for the squares of the diagram $[\mu] = [(4,1)]$ omitted, so that the visible tops of pillars correspond to the squares of $[\lambda/\mu]$. In the second picture we have added the paths along the "row boundaries", which trace the boundary of the block diagram at fixed value of i.



This display has as defects that the relative position of the paths is not correct (so cases where paths should have a common point would still appear disjoint), and the paths along "column boundaries" (with j fixed, so running from top right to bottom left) do not resemble the second family. Both points can be remedied simultaneously, by raising (in the sense of decreasing k) each "row" i (the plane of cubes with that value of i) by i units. This amounts to using, instead of the block diagram of the semistandard tableau $(\lambda^{(k)})_{k\in\mathbf{N}}$, the set $\{\,(i,j,k-i)\mid(i,j)\in[\lambda^{(k)}]\,\}\subseteq\mathbf{N}^3$. We show the resulting diagram, and then add the two superimposed families of paths.



§5. The Littlewood-Richardson rule.

We shall finally apply lemma 2.2 for $f = s_\lambda$ and even for $f = s_{\lambda/\kappa}$, so as to obtain an expression for multiplication by (skew) Schur functions in the basis of Schur functions. As usual, the sport will be to find a cancellation of the resulting alternating sum that leaves only positive contributions (from representation theoretic interpretations of Schur functions it is clear that the decomposition of the product in the basis of Schur functions cannot involve negative coefficients). We shall thus find a combinatorial description of the decomposition of a product $s_\lambda s_\mu$ into Schur functions that is equivalent to the Littlewood-Richardson rule. In fact we shall obtain a generalisation that was given in [Zel], which describes the scalar product of an arbitrary pair of skew Schur functions. Our approach naturally leads to two different formulations of this result, one in terms of binary matrices and another in terms of matrices with coefficients in $\mathbf{N}$. Both are equivalent to a more traditional formulation in terms of Littlewood-Richardson tableaux, but the relation with cancellations in alternating sums gives a perspective somewhat different from for instance the one we took in [vLee], where the Littlewood-Richardson rule was approached via Robinson's correspondence and jeu de taquin. We would like to note however that coplactic operations provide a very elegant means to unify the two points of view, as we shall indicate in a sequel to this paper.

As an introduction we shall discuss the question that was evoked after equation (32), of proving bijectively that Schur functions are the generating series of semistandard Young tableaux of given shape by weight. The suggested proof was to apply lemma 2.2 for $f[X_\mathbf{N}] = \sum_{T \in \mathrm{SST}(\lambda)} X^{\mathrm{wt}(T)}$ and $\beta = (0)$, and prove that its right hand side reduces by cancellations to s_λ. If we pretend not to know already that this generating series $f[X_\mathbf{N}]$ is a Schur function, then there is in fact another obligation, namely to prove that $f[X_\mathbf{N}]$ is a symmetric function, so that lemma 2.2 can be applied in the first place. Proving this symmetry can, surprisingly, be done in almost the same way as proving the mentioned cancellation

$$\sum_{T \in \mathrm{SST}(\lambda)} s_{\mathrm{wt}(T)} = s_\lambda. \tag{48}$$

The symmetry of $f[X_\mathbf{N}]$ can be established by giving for each m an involution on $\mathrm{SST}(\lambda)$ that interchanges components m and $m+1$ of the weight, which means, in terms of the display of the tableaux that the frequencies of the entries m and $m+1$ must be interchanged, while preserving the shape. Since there is no reason to change other entries than m and $m+1$, this involution can be viewed as operating only on the component shape $\lambda^{(m+1)}$ of tableaux $(\lambda^{(i)})_{i \in \mathbf{N}}$ in $\mathrm{SST}(\lambda)$, which shape is constrained by the relation $\lambda^{(m)} \leftharpoondown \lambda^{(m+1)} \leftharpoondown \lambda^{(m+2)}$ in which the shapes $\lambda^{(m)}$ and $\lambda^{(m+2)}$ are fixed.

As for the cancellation (48), the one tableau T in the summation whose term survives the cancellation will be the one whose display has only entries i in any row i, and which therefore satisfies $\mathrm{wt}(T) = \lambda$. For any other $T \in \mathrm{SST}(\lambda)$, let i be minimal such that row i of the display of T has some entry other than i (by column strictness, such entries must be greater than i), and let the final entry of that row, in column $j = \lambda_i - 1$, have value $m+1$ (so $m \geq i$). We shall consider this entry as the one that triggers the cancellation of the term for T, and in constructing $T' \in \mathrm{SST}(\lambda)$ whose terms cancels that of T, we shall leave that entry intact, as well as all rows $i' < i$, so that the same entry $m+1$ also triggers the cancellation for T'. In fact, we shall only change entries m and $m+1$ in the display of T, if any, with the goal of interchanging the values of $\mathrm{wt}(T)[m]$ and of $\mathrm{wt}(T)[m+1]$. The entries that may be changed are in rows $i' \geq i$ and in columns $j' < j$ of the display of T, and one can treat the subtableau of entries $m, m+1$ in that region in isolation: whatever subtableau of entries $m, m+1$ should replace them, it will be compatible with the entry $m+1$ at position (i, j). Moreover, since that single entry causes the contribution to $\mathrm{wt}(T)[\,]$ of the part outside the specified region to have equal values at indices m and $m+1$, the goal must be to interchange the frequencies of entries m and $m+1$ in the mentioned subtableau. Thus, the problem is basically the same one as for proving symmetry of the generating series $f[X_{\mathbf{N}}]$.

The simplest involutions establishing that symmetry are those that were introduced in [BeKn], and which we shall call the Bender-Knuth involutions. They were in fact introduced in the combinatorial proof of a lemma that essentially states our equation (48) (lemma 2 of the cited paper), although an application to the symmetry of f immediately follows it. The idea is straightforward: one ignores any columns containing both entries m and $m+1$ (such entries cannot be changed), after which all rows can be treated independently, and in each of them one separately interchanges the frequencies of m and $m+1$; we refer to the proof of [Stan, theorem 7.10.2] for details.

A simple example will illustrate these involutions and their application to establishing (48). Take the semistandard Young tableau displayed as follows

$$\begin{array}{ccccccc}
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 2 & 2 & 2 & 4 & 4 & 4 \\
3 & 3 & 3 & 4 & 5 & 6 & 6 \\
4 & 5 & 5 & 6 & 6 & & \\
6 & 7 & 7 & & & &
\end{array}$$

The Bender-Knuth involution that interchanges the components 5 and 6 of the weight, operates separately on the singleton sequence $[6]$ in row 4, on the sequence $[5, 5, 6]$ in columns $1, 2, 3$ of row 3, and on the sequence $[6, 6]$ at the end of row 2 (the entries 5 and 6 in column 4 are frozen); the sequences are respectively turned into $[5]$, $[5, 6, 6]$ and $[5, 5]$. As for the cancellation (48), it is triggered by the final entry 4 of row 1, and therefore operates on entries with values 3 and 4; that final entry does not participate in the exchange, nor do the frozen entries in column 0, so $[3, 3, 4]$ in row 2 and the initial $[4, 4]$ in row 1 are respectively changed into $[3, 4, 4]$ and $[3, 3]$. The tableaux resulting for these two operations are

$$\begin{array}{ccccccc}
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 2 & 2 & 2 & 4 & 4 & 4 \\
3 & 3 & 3 & 4 & 5 & 5 & 5 \\
4 & 5 & 6 & 6 & 6 & & \\
5 & 7 & 7 & & & &
\end{array}
\quad \text{and} \quad
\begin{array}{ccccccc}
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 2 & 2 & 2 & 3 & 3 & 4 \\
3 & 3 & 4 & 4 & 5 & 6 & 6 \\
4 & 5 & 5 & 6 & 6 & & \\
6 & 7 & 7 & & & &
\end{array}.$$

One sees that the weights of these tableaux have the required relation to the weight $\alpha = (7,1,3,3,5,3,5,2)$ of the initial tableau, notably the weight $\alpha' = (7,1,3,4,4,3,5,2)$ of the final tableau satisfies $s_\alpha + s_{\alpha'} = 0$.

Now let us turn to more useful applications of lemma 2.2, for which we do not know the result beforehand. With respect to the application in the previous example one could replace the choice $\beta = (0)$ by one taking for β an arbitrary partition μ, which results in a formula for products $s_\lambda s_\mu$ of Schur functions, or one could replace the choice $f = s_\lambda$ by the more general $f = s_{\lambda/\kappa}$, resulting in a formula for decomposing skew Schur functions into ordinary ones. In view of the relation $\langle\, s_\lambda s_\mu \mid s_\nu \,\rangle = \langle\, s_\lambda \mid s_{\nu/\mu} \,\rangle$ the two problems are equivalent, but it will be interesting to perform both generalisations at once, which in fact produces no additional complications. So we apply lemma 2.2 with $f = s_{\lambda/\kappa}$ (and hence $f[X_\mathbf{N}] = \sum_{T \in \mathrm{SST}(\lambda/\kappa)} X^{\mathrm{wt}(T)}$) and $\beta = \mu \in \mathcal{P}$, which gives

$$s_{\lambda/\kappa} s_\mu = \sum_{T \in \mathrm{SST}(\lambda/\kappa)} s_{\mu + \mathrm{wt}(T)}. \tag{49}$$

We can obtain a somewhat more symmetric formulation by taking coefficients of some Schur function s_ν on both sides, which amounts to taking the scalar product with it; using $\langle\, s_{\lambda/\kappa} s_\mu \mid s_\nu \,\rangle = \langle\, s_{\lambda/\kappa} \mid s_{\nu/\mu} \,\rangle$ this gives

$$\langle\, s_{\lambda/\kappa} \mid s_{\nu/\mu} \,\rangle = \sum_{T \in \mathrm{SST}(\lambda/\kappa)} \varepsilon(\mu + \mathrm{wt}(T), \nu). \tag{50}$$

We shall now describe a cancellation that can be applied to the right hand side of (49), which was proposed in [Stem]. Contrary to what we saw for (48) there will in general be more than one term that survives the cancellation, but for those tableaux T whose term does cancel, an entry $m+1$ in its display will be indicated that triggers the cancellation, and a tableau T' whose terms cancels that of T will be constructed by possibly modifying entries m and $m+1$, in much the same way as before. In particular all changes will be strictly to the left and weakly below the entry triggering the cancellation.

The condition for cancellation is stated in terms of the weights of individual columns of T. Define the weight $\mathrm{cwt}_j(T)$ of column j of T to be the binary composition α whose part $\alpha_i \in \{0,1\}$ tells whether or not there is an entry i in column j of the display of T, in other words, $\mathrm{cwt}_j(T)$ is the column $(M')^t_j$ of the binary encoding M' of T. For later use also define the weight $\mathrm{rwt}_i(T)$ of row i of T as the composition β whose part β_j gives the number of entries j in row i of the display of T; thus $\mathrm{rwt}_i(T)$ equals row M_i of the integral encoding M of T. One obviously has $\mathrm{wt}(T) = \sum_{j \in \mathbf{N}} \mathrm{cwt}_j(T) = \sum_{i \in \mathbf{N}} \mathrm{rwt}_i(T)$.

The terms in the right hand side of (49) that will survive the cancellation are those for the tableaux T with the property that for all $l \in \mathbf{N}$ one has $\mu + \sum_{j \geq l} \mathrm{cwt}_j(T) \in \mathcal{P}$; in particular $\mu + \mathrm{wt}(T)$ is a partition, so that the term for such T is already normalised. So once the cancellation is established, it will prove

$$s_{\lambda/\kappa} s_\mu = \sum_{T \in \mathrm{SST}(\lambda/\kappa)} [\, \forall l \in \mathbf{N} \colon \mu + \sum_{j \geq l} \mathrm{cwt}_j(T) \in \mathcal{P} \,]\, s_{\mu + \mathrm{wt}(T)}, \tag{51}$$

which is a generalisation of the Littlewood-Richardson rule. The tableaux that satisfy the condition are those that we termed μ-dominant in [vLee, §1.4] (although it requires a bit of work to see this), and for $\mu = (0)$ these are the traditional Littlewood-Richardson tableaux; as one can see, these describe the decomposition of $s_{\lambda/\kappa}$ in the basis of Schur functions. By taking instead $\kappa = (0)$, one obtains a description of the decomposition of $s_\lambda s_\mu$ in that basis, in terms of μ-dominant tableaux of shape λ.

For any T that fails the condition in (51), there is a maximal l such that $\mu + \sum_{j\geq l} \mathrm{cwt}_j(T) \notin \mathcal{P}$; we fix this l and put $\alpha = \mu + \sum_{j\geq l} \mathrm{cwt}_j(T)$. Since $\alpha \notin \mathcal{P}$ while $\mu + \sum_{j>l} \mathrm{cwt}_j(T) = \alpha - \mathrm{cwt}_l(T) \in \mathcal{P}$, and $\mathrm{cwt}_l(T)$ is a binary composition, it must be that there exist some m with $\alpha_{m+1} = \alpha_m + 1$, which can also be written as $\alpha[m] = \alpha[m+1]$. We choose the minimal such m, which then necessarily also has $\mathrm{cwt}_l(T)_m = 0$ and $\mathrm{cwt}_l(T)_{m+1} = 1$, so column l of the display of T contains no entry m, but it does contain an entry $m+1$. That entry will be the one that is considered to trigger the cancellation of the term for T.

Since the entries in columns $j \geq l$ of the display of T are involved in triggering the cancellation and determining the values of l and m, they will not be altered during the construction of a tableau T' whose term cancels that of T; this will ensure that the cancellation is triggered identically for T' as it was for T. As before the subtableau of entries $m, m+1$ in columns $j < l$ can be treated in isolation, since the frozen entry $m+1$ in column l imposes no constraint. One has $\mu + \mathrm{wt}(T) = \alpha + \sum_{j<l} \mathrm{cwt}_j(T)$ and for T' one will similarly have $\mu + \mathrm{wt}(T') = \alpha + \sum_{j<l} \mathrm{cwt}_j(T')$; since $\alpha[m] = \alpha[m+1]$, the desired cancellation $s_{\mu+\mathrm{wt}(T)} + s_{\mu+\mathrm{wt}(T')} = 0$ can be assured by having $\sum_{j<l} \mathrm{cwt}_j(T')$ differ from $\sum_{j<l} \mathrm{cwt}_j(T)$ by the interchange of the parts indexed by m and $m+1$. Therefore T' can be obtained from T by applying the Bender-Knuth involution for $m, m+1$ to the subtableau in columns $j < l$.

As an example consider the tableau T of (2) with $\mu = (7,6,5,4,2)$; one finds $l = 5$, for which $\alpha = \mu + (0,0,1,0,1,2,3) = (7,6,6,4,3,2,3)$ and $m = 5$, and the entry $m+1 = 6$ at position $(1,5)$ triggers the cancellation. The tableau T' is found by applying the Bender-Knuth involution for interchanging 5 and 6 to the subtableau formed by the 5 leftmost columns of T, which transforms

$$T = \begin{array}{ccccccccc} &&&&0&2&4&5&5\\ &0&1&3&4&6&6&6&\\ 1&1&2&4&5&&&&\\ 2&3&4&6&6&&&&\\ 5&6&6&&&&&& \end{array} \qquad \text{into} \qquad T' = \begin{array}{ccccccccc} &&&&0&2&4&5&5\\ &0&1&3&4&6&6&6&\\ 1&1&2&4&5&&&&\\ 2&3&4&5&6&&&&\\ 5&5&6&&&&&& \end{array}. \tag{52}$$

Indeed one has $s_{\mu+\mathrm{wt}(T)} + s_{\mu+\mathrm{wt}(T')} = s_{(9,9,8,6,6,4,7)} + s_{(9,9,8,6,6,6,5)} = 0$.

The condition for μ-dominance that appears in (51) may appear somewhat strange at first sight, but it has a remarkable resemblance to the condition that characterises binary encodings of tableaux. To bring out this resemblance, we shall replace the tableau T in the summation by its binary encoding M, so that M^t_j will replace $\mathrm{cwt}_j(T)$, and $\mathrm{row}(M)$ will replace $\mathrm{wt}(T)$. Also we shall apply our cancellations to the right hand side of (50) rather than of (49). One has $\mathrm{col}(M) = \lambda^t - \kappa^t$ for any $M \in \mathrm{Tabl}^{[2]}(\lambda/\kappa)$, and those M giving nonzero terms after cancellation satisfy $\mathrm{row}(M) = \nu - \mu$. Spelling out explicitly the condition that $M \in \mathrm{Tabl}^{[2]}(\lambda/\kappa)$, we therefore obtain the following statement, equivalent to (51).

5.1. Theorem. *(generalised Littlewood-Richardson rule, binary formulation) For all $\kappa, \lambda, \mu, \nu \in \mathcal{P}$:*

$$\langle\, s_{\lambda/\kappa} \mid s_{\nu/\mu} \,\rangle = \sum_{M \in \mathcal{M}^{[2]}_{\nu-\mu,\lambda^t-\kappa^t}} [\,\forall k\colon \kappa^t + \sum_{i<k} M_i \in \mathcal{P}\,]\,[\,\forall l\colon \mu + \sum_{j \geq l} M^t_j \in \mathcal{P}\,]\,. \qquad \square$$

There is another formulation of the condition of μ-dominance than the one that appears in (51), which uses row weights rather than column weights, and which is actually closer to the way the notion was defined in [vLee]. One can determine independently for each value of m whether any entry $m+1$ in the display of T is a candidate for triggering cancellation. Although in case of multiple candidates, only the topmost one in the leftmost column containing any candidates will be selected, that rule has no influence on the set of tableaux that will survive cancellation, namely those for which no candidates exist at all. In order for an entry $m+1$ in column l to be candidate for triggering cancellation, one must have $\alpha[m] = \alpha[m+1]$ for $\alpha = \mu + \sum_{j \geq l} \mathrm{cwt}_j(T)$, and l must be the largest value for which this condition holds. Then there will not be any entry m above that entry $m+1$ in the same column l of T, and this means that if the entry $m+1$ is in row k, one also has $\alpha'[m] = \alpha'[m+1]$ for $\alpha' = \mu + \sum_{i<k} \mathrm{rwt}_i(T) + \beta$, where β is the weight of the part of row k of T consisting of the candidate entry and all entries to its right.

Conversely any entry $m+1$ for which this condition in terms of α' holds, and for which k is minimal, cannot have an entry m in the same column, and is therefore a candidate for triggering cancellation. Then, assuming there are no such candidate entries $m+1$ in rows $i<k$, one may test if there is *any* such entry $m+1$ in row k, by considering $\alpha'' = \mu + \sum_{i<k} \mathrm{rwt}_i(T) + \beta'$ where β' the weight of the part of row k consisting of entries exceeding m (in other words β' is obtained from $\mathrm{rwt}_k(M)$ by clearing its parts at indices $\leq m$): there will be such a candidate if and only if $\alpha''_{m+1} > \alpha''_m$. This use of weights of partial rows can be avoided in the formulation, by comparing parts of two different compositions: if one defines $\alpha^{(i_0)} = \mu + \sum_{i<i_0} \mathrm{rwt}_i(T)$ for $i_0 \in \mathbf{N}$, then the final condition $\alpha''_{m+1} > \alpha''_m$ is equivalent to $\alpha^{(k)}_m < \alpha^{(k+1)}_{m+1}$. For instance in our example above, the fact that some entry 6 in row 1 is a candidate for triggering cancellation, can be deduced from $\alpha^{(1)}_5 < \alpha^{(2)}_6$ where $\alpha^{(1)} = \mu + (1,0,1,0,1,2) = (8,6,6,4,3,\underline{2})$ and $\alpha^{(2)} = \alpha^{(1)} + (1,1,0,1,1,0,3) = (9,7,6,5,4,2,\underline{3})$ (only the underlined entries are relevant).

Now let us return to formulating μ-dominance, the condition stating that the term for T is not cancelled at all. This means that the condition just stated fails for all k and for all m. The failure for given k and for all m simply becomes $\alpha^{(k)} \leftharpoondown \alpha^{(k+1)}$, which means that we can reformulate (51) as follows:

$$s_{\lambda/\kappa} s_\mu = \sum_{T \in \mathrm{SST}(\lambda/\kappa)} [\,\forall k \in \mathbf{N}\colon \mu + \sum_{i<k} \mathrm{rwt}_i(T) \leftharpoondown \mu + \sum_{i \leq k} \mathrm{rwt}_i(T)\,]\, s_{\mu+\mathrm{wt}(T)}. \tag{53}$$

Note that, somewhat surprisingly, the sequence $(\alpha^{(i)})_{i\in\mathbf{N}}$ of compositions actually is a semistandard tableau when T survives the cancellation; its shape is $\mu+\mathrm{wt}(T)/\mu$, and in the terminology of [vLee] it is the companion tableau of T witnessing the μ-dominance of T. If this time we replace T in the summation by its integral encoding, we obtain a symmetry in the formulation that is even more striking than the one in theorem 5.1. Again we take (50) as a starting point, and obtain the following statement, which is equivalent to (53), and therefore to (51), like the previous theorem. This time the condition that selects the matrices whose term survives the cancellation is exactly like the condition for being an integral encoding in the first place, except for the interchange of rows and columns; therefore we add a succinct reformulation of the conditions, using the notation introduced in definition 1.2.3.

5.2. Theorem. *(generalised Littlewood-Richardson rule, integral formulation) For all* $\kappa,\lambda,\mu,\nu\in\mathcal{P}$

$$\begin{aligned}\langle\, s_{\lambda/\kappa} \mid s_{\nu/\mu}\,\rangle &= \sum_{M\in\mathcal{M}_{\lambda-\kappa,\nu-\mu}} [\,\forall l\colon \kappa+\sum_{j<l}M_j^t \leftharpoondown \kappa+\sum_{j\le l}M_j^t\,]\,[\,\forall k\colon \mu+\sum_{i<k}M_i \leftharpoondown \mu+\sum_{i\le k}M_i\,]\\ &= \sum_{M\in\mathcal{M}} [\,M\in\mathrm{Tabl}(\lambda/\kappa)\,]\,[\,M^t\in\mathrm{Tabl}(\nu/\mu)\,]\,. \qquad\square\end{aligned}$$

We have completed the description of our final application of lemma 2.2, and of the result of applying cancellations to the expression it yields. There is however one more variation that we would like to present. In theorems 5.1 and 5.2 we have given expressions for $\langle\, s_{\lambda/\kappa} \mid s_{\nu/\mu}\,\rangle$ in terms of matrices that treat rows and columns similarly. But our starting point for these expressions, equation (50), does not exhibit such a symmetry when we replace the semistandard tableaux in the summation by their binary or integral encoding matrices. We would like to give expressions involving alternating sums with the same kind of symmetry as the mentioned theorems, and from which those theorems can be obtained by cancellations. This will require that we proceed backwards, introducing a second layer of signs by a kind of inverse cancellation process. In fact we shall use equations (27) and (28) to translate the Kostka numbers, which are implicit in the summation over semistandard tableaux, into alternating sums over matrices.

Concretely, we write the right hand side of (50) as $\sum_{\beta\in\mathcal{C}} K_{\lambda/\kappa,\beta}\varepsilon(\mu+\beta,\nu)$, and then we can apply (27), which gives

$$\langle\, s_{\lambda/\kappa} \mid s_{\nu/\mu}\,\rangle = \sum_{M\in\mathcal{M}} \varepsilon(\kappa+\mathrm{row}(M),\lambda)\varepsilon(\mu+\mathrm{col}(M),\nu), \qquad (54)$$

or we can apply (28) to $K_{\lambda/\kappa,\beta}=K'_{\lambda^t/\kappa^t,\beta}$, which gives (transposing the binary matrices M)

$$\langle\, s_{\lambda/\kappa} \mid s_{\nu/\mu}\,\rangle = \sum_{M\in\mathcal{M}^{[2]}} \varepsilon(\kappa^t+\mathrm{col}(M),\lambda^t)\varepsilon(\mu+\mathrm{row}(M),\nu). \qquad (55)$$

It is clear from the way we have obtained these equations, that one can go from them to theorems 5.2 and 5.1, respectively, by applying two successive phases of cancellation. The first phase is a Gessel-Viennot cancellation of terms with opposite values of $\varepsilon(\kappa+\mathrm{row}(M),\lambda)$ respectively of $\varepsilon(\kappa^t+\mathrm{col}(M),\lambda^t)$, as in (43) and (47), which gives us back (50) with tableaux replaced by their integral or binary encodings. The second cancellation phase cancels within the remaining terms pairs with opposite values of the remaining factor $\varepsilon(\mu+\mathrm{col}(M),\nu)$, using the cancellation based on the Bender-Knuth involutions that was given following (51), translated in terms of integral or binary encodings.

It is interesting to compare the way in which these two cancellation phases proceed. The Gessel-Viennot cancellations search for a failure of the condition $M\in\mathrm{Tabl}(\lambda/\kappa)$ respectively $M\in\mathrm{Tabl}^{[2]}(\lambda^t/\kappa^t)$ by successively computing the compositions $\alpha^{(i)}$ that would define the shapes (or their transposes in the binary case) forming the semistandard tableau encoded by M if one exists; as soon as one finds $\alpha^{(i)}\not\twoheadrightarrow\alpha^{(i+1)}$ respectively $\alpha^{(i)}\not\rightharpoonup\alpha^{(i+1)}$, the cancellation of the term for M is detected, and the construction of a cancelling matrix M' starts. Up to this point the second cancellation phase proceeds in quite the same way, differing only in whether weights of rows or columns of M are successively added to the compositions $\alpha^{(i)}$, and in the fact that for the binary case these columns are taken in the opposite order.

But after detecting cancellation and selecting a pair of indices responsible for it, the Gessel-Viennot cancellations construct M' from M by simply interchanging entries for those indices in the rows or columns that had not yet been considered; in the integral case there is also a transfer at those indices in the column in which cancellation was detected itself, which can be interpreted as interchange of those parts of the values of the entries concerned that remain after the cancellation has been triggered (the parts of their paths after the first point of contact). Such a simple rule would not work in the Bender-Knuth case: in the binary case weak increase along rows would be violated, while in the integral case strict increase in columns would be violated (in either case this is the only one of the two conditions that can be violated, as the other one is assured by the way tableaux are encoded by matrices).

What the Bender-Knuth involutions actually do in terms of integral and binary encodings is not easily described directly, although one can say in general that the transfers between entries are more limited than what they would be for a Gessel-Viennot cancellation. This can be seen by considering the encodings of the tableaux T,T' in (52). The changes only involve entries 5 and 6, so we just show the changes to columns 5 and 6 of the integral encoding and to rows 5 and 6 of the binary encoding:

$$\begin{pmatrix}2&0\\0&\underline{3}\\1&0\\0&2\\1&2\end{pmatrix}\leftrightarrow\begin{pmatrix}2&0\\0&\underline{3}\\1&0\\1&1\\2&1\end{pmatrix}\quad\text{and}\quad\begin{matrix}\begin{pmatrix}1&0&0&0&1&0&0&1&1\\0&1&1&1&1&\underline{1}&1&1&0\end{pmatrix}\\ \updownarrow\\ \begin{pmatrix}1&1&0&1&1&0&0&1&1\\0&0&1&0&1&\underline{1}&1&1&0\end{pmatrix}\end{matrix}.\qquad(56)$$

The entries where cancellation is triggered have been underlined (in the integral case it is the final unit of the entry 3, i.e., a value 2 at the same position would not trigger cancellation). One sees that the parts of the columns below that entry (in the integral case) or of the rows to the left of that entry (in the binary case) have not been completely interchanged. The precise changes depend on the shape $\lambda^{(5)} = (7, 5, 4, 3)$ in the tableau T to which the entries 5 in its display are added, and in terms of the matrices this shape depends both on μ and on the columns to the left of those that were shown, respectively the rows above those that were shown.

One may wonder, both for (54) and for (55), if it is possible to find a pair of cancellations that reduces them to theorem 5.2 respectively to theorem 5.1, in a manner that better respects the symmetry present both in the initial and in the final expressions. In particular one would like to be able to apply the two cancellations in either order, with the cancellation being applied last operating only on the terms that survived the first cancellation. This requires that either cancellation respects the condition expressing survival under the other cancellation: it will never cancel a survivor of the other cancellation against a non-survivor. Yet while the involutions defining such cancellations should allow restriction to the survivors of the other cancellation, their definition should not be limited to those survivors (as is the case for the Bender-Knuth involutions, which are only defined for semistandard tableaux): they should be as general as the Gessel-Viennot cancellations. We shall show in a forthcoming sequel to this paper that both for (54) and for (55) such a pair of cancellations does indeed exist. Moreover the operations used to define these cancellations are quite interesting in their own right, and they give rise to ramifications that relate to Knuth correspondences and to jeu de taquin.

References.

[BeKn] E. A. Bender and D. E. Knuth, "Enumeration of plane partitions", *J. Combin. Theory, Ser. A* **13**, (1972), 40–54.

[Cau] A. L. Cauchy, "Mémoire sur les fonctions qui ne peuvent obtenir que deux valeurs égales et de signes contraires par suite des transpositions opérées entre les variables qu'elles renferment", *J. École polythechnique* **10**, (1815), 29–112. *Œuvres*, sér. 2, vol. 1, pp. 91–169.

[GeVi] I. Gessel and G. Viennot, "Binomial determinants, paths, and hook length formulae", *Advances in Math.* **58**, (1985), 300–321.

[GKP] R. L. Graham, D. E. Knuth, and O. Patashnik, *Concrete mathematics, a foundation for computer science*, Addison-Wesley publishing company, 1989.

[Jac] C. G. Jacobi, "De functionibus alternantibus", *Crelle's Journal* **22**, (1841), 360–71. Also in "Werke" **3**, 439–52.

[Kos] C. F. A. Kostka, "Über die Bestimmung von symmetrischen Functionen der Wurzeln einer algebraischen Gleichung durch deren Coefficienten", *Crelle's J.* **81**, (1875), 281–289.

[Knu] D. E. Knuth, "Permutations, matrices and generalized Young tableaux", *Pacific J. of Math.* **34**, (1970), 709–727.

[vLee] M. A. A. van Leeuwen, "The Littlewood-Richardson rule, and related combinatorics", pp. 93–145 in *Interaction of Combinatorics and Representation Theory*, Math. Soc. of Japan Memoirs 11, 2001.

[Litt] D. E. Littlewood, *The theory of group characters*, 2nd edition, Oxford University Press, 1950.

[LiRi] D. E. Littlewood and A. R. Richardson, "Group characters and algebra", *Phil. Trans. A* **233**, (1934), 99–141.

[Macd] I. G. Macdonald, *Symmetric Functions and Hall Polynomials*, Oxford Mathematical Monographs, Clarendon press, Oxford, 1979.

[Mur] F. D. Murnaghan, "The characters of the symmetric group", *Am. J. of Math.* **59**, (1937), 739–753.

[Muir] T. Muir, *The theory of determinants in the historic order of development, volume III*, Dover Publications, New York, 1960. Originally published in 1920.

[vNäg] H. E. von Nägelsbach, "Über eine Classe symmetrischer Functionen", *Sch.-Prog.*, (1871), 42 pp.

[Nak] T. Nakayama, "On some modular properties of irreducible representations of a symmetric group, I and II", *Japanese J. of Math.* **17**, (1940), 165–184, 411–423.

[Stan] R. P. Stanley, *Enumerative combinatorics, Volume 2*, Cambridge studies in advanced mathematics 62, Cambridge University press, 1999.

[Stem] J. R. Stembridge, "A concise proof of the Littlewood-Richardson rule", *Electronic J. of Combinatorics* **9 (no 1)**, N5, (2002), 4 pp.

[Tru] N. Trudi, "Intorno ad un determinante più generale di quello che suol dirsi determinante delle radici di una equazione, ed alla funzioni simmetriche complete di questi radici", *Rendic. dell'Accac. (Napoli)*, (1864), 121–134.

[Zel] A. V. Zelevinsky, "A Generalisation of the Littlewood-Richardson Rule and the Robinson-Schensted-Knuth Correspondence", *J. of Algebra* **69**, (1981), 82–94.