CRM-3347 (2015)

THE GENERATING FUNCTION ENUMERATING WORDS IN n LETTERS WITHOUT INCREASING SUBSEQUENCES OF LENGTH d AND WITH EACH LETTER OCCURRING r TIMES

FERENC BALOGH

Abstract. Gessel's famous Bessel determinant formula gives the generating function to enumerate permutations without increasing subsequences of a given length. Ekhad and Zeilberger recently proposed the challenge to find a suitable generalization to count words of length rn in an alphabet consisting of n letters in which each letter appears exactly r times and which have no increasing subsequences of length d.

In this paper we present a generating function for arbitrary values of r expressible as multiple integrals of Gessel-type Toeplitz determinants multiplied by the exponentiated cycle index polynomial of the symmetric group on r letters.

1. Introduction and statement of results

For integers $r \geq 1$ and $d \geq 2$, let $A_{d,r}(n)$ denote the number of words in the alphabet $\{1, 2, \ldots, n\}$ of length rn in which each letter appears exactly r times and which have no increasing subsequences of length d. As it was pointed out[1] by Ekhad and Zeilberger in [EZ], the Robinson–Schensted–Knuth (RSK) correspondence (cf. [St]) implies that $A_{d+1,r}(n)$ can be written as a sum over partitions λ of weight $|\lambda|$ equal to rn and of length $\ell(\lambda)$ at most d in the following form:

$$A_{d+1,r}(n) = \sum_{\substack{|\lambda|=rn \\ \ell(\lambda)\leq d}} f_\lambda g_\lambda^{(r)}, \tag{1}$$

where f_λ is the number of standard Young tableaux of shape λ and $g_\lambda^{(r)}$ is the number of semi-standard Young tableaux of shape λ whose content is $1^r 2^r \cdots n^r$.

For the special case $r = 1$ (each letter appears exactly once) we have $g_\lambda^{(1)} = f_\lambda$ and hence

$$A_{d+1,1}(n) = \sum_{\substack{|\lambda|=n \\ \ell(\lambda)\leq d}} f_\lambda^2. \tag{2}$$

Gessel [Ge] found the following remarkable generating function for the sequence $\{A_{d+1,1}(n)\}_n$:

$$\sum_{n=0}^{\infty} \frac{A_{d+1,1}(n)}{(n!)^2} x^{2n} = \det_{1\leq i,j\leq d} \left(I_{|i-j|}(2x) \right), \tag{3}$$

where the entries of the $d \times d$ Toeplitz determinant are *modified Bessel functions* $I_m(x)$ with integer parameters, given by the generating function (cf. [DLMF], 10.35.1)

$$\sum_{m=-\infty}^{\infty} I_m(x) z^m = \exp\left(\frac{x}{2} \left(\frac{1}{z} + z \right) \right). \tag{4}$$

Date: June 25, 2015.

[1] The formula in [EZ] had a misprint which was found and corrected by Chapuy in [Ch].

For the sake of comparison with the general formulae to be presented, one can write Gessel's generating function (3) in a slightly different but equivalent Toeplitz determinant form:

$$\sum_{n=0}^{\infty} \frac{A_{d+1,1}(n)}{(n!)^2} |v|^{2n} = \det_{1\le i,j\le d} \left(\varphi_{i-j}(v,\bar{v})\right), \tag{5}$$

where the entries $\varphi_m(v,\bar{v})$ are the Laurent series coefficients at $z=0$ of the Toeplitz symbol

$$\varphi(v,\bar{v};z) = \sum_{m=-\infty}^{\infty} \varphi_m(v,\bar{v}) z^m = \exp\left(\frac{v}{z} + \overline{v} z\right), \tag{6}$$

where $\overline{v}$ denotes the complex conjugate complex-valued book-keeping parameter v and $|v|$ is its absolute value.

Gessel's Toeplitz determinant formula played a crucial role in the subsequent developments of the asymptotic analysis of large permutations, most notably the famous Baik–Deift–Johansson theorem [BDJ]. For several related results in this direction see [AD, BOO, ITW1, ITW2, vM, TW] and references therein.

Recently Ekhad and Zeilberger [EZ] proposed the challenge to generalize Gessel's determinant formula (3) for multiple occurrences $r > 1$. This goal is fully accomplished in the main theorem of this paper.

Before stating the main result, recall that the *cycle index polynomial* of the symmetric group S_r is the multivariate polynomial

$$\mathrm{Cyc}_r(t_1,\ldots,t_r) = \frac{1}{r!} \sum_{|\mu|=r} |C_\mu| \prod_{i=1}^{\ell(\mu)} t_{\mu_i} \tag{7}$$

in the indeterminates $t_1,\ldots,t_r$, where the sum is over partitions μ of weight $|\mu| = r$, $C_\mu \subset S_r$ is the conjugacy class consisting of permutations of cycle type μ and $\ell(\mu)$ stands for the length (number of parts) of the partition μ. For example, for $r = 1,2,3$ the cycle index polynomials are

$$\begin{aligned} \mathrm{Cyc}_1(t_1) &= t_1, \\ \mathrm{Cyc}_2(t_1,t_2) &= \frac{1}{2}t_1^2 + \frac{1}{2}t_2, \\ \mathrm{Cyc}_3(t_1,t_2,t_3) &= \frac{1}{6}t_1^3 + \frac{1}{2}t_1^2 t_2 + \frac{1}{3}t_3. \end{aligned}$$

Note that the cycle index polynomial Cyc_r is a weighted homogeneous polynomial of degree r in the natural grading $\deg(t_j) = j$.

Theorem 1. *The sequence $\{A_{d+1,r}(n)\}_{n=0}^{\infty}$ has the formal series generating function*

$$\sum_{n=0}^{\infty} \frac{A_{d+1,r}(n)}{(rn)!n!} |v|^{2rn} = \frac{1}{\pi^r} \underbrace{\int_{\mathbb{C}} \cdots \int_{\mathbb{C}}}_{r} \det_{1\le i,j\le d} \left(\varphi_{i-j}(v,\mathbf{t})\right) \cdot e^{\bar{v}^r \overline{\mathrm{Cyc}_r(\mathbf{t})}} \prod_{j=1}^{r} e^{-|t_j|^2} dA(t_j), \tag{8}$$

in the indeterminates v and $\bar{v}$, where $\varphi(v,\mathbf{t};z)$ is the Toeplitz symbol

$$\varphi(v,\mathbf{t};z) = \sum_{m=-\infty}^{\infty} \varphi_m(v,\mathbf{t}) z^m = \exp\left(\frac{v}{z} + \sum_{j=1}^{r} t_j z^j\right), \tag{9}$$

$\mathrm{Cyc}_r(\mathbf{t})$ *is the cycle index polynomial*

$$\mathrm{Cyc}_r(\mathbf{t}) = \frac{1}{r!}\sum_{|\mu|=r}|C_\mu|\prod_{i=1}^{\ell(\mu)} t_{\mu_i} \tag{10}$$

of the symmetric group S_r and dA stands for the area measure in the complex plane.

Remark 1. The indeterminates v and $\bar{v}$ are considered independent book-keeping variables in terms of which the Toeplitz determinant $\det(\varphi_{i-j}(v,\mathbf{t}))$ and the exponentiated cycle index polynomial $e^{\overline{v}^r\overline{\mathrm{Cyc}_r(\mathbf{t})}}$ can be expanded into formal power series simultaneously and independently. The coefficients of the resulting product series

$$\det(\varphi_{i-j}(v,\mathbf{t}))\, e^{\overline{v}^r\overline{\mathrm{Cyc}_r(\mathbf{t})}} = \sum_{k,l=0}^{\infty} F_k(\mathbf{t})\overline{G_l(\mathbf{t})}v^k\bar{v}^l$$

are products of weighted homogeneous polynomials in $\mathbf{t}$ and $\overline{\mathbf{t}}$ and therefore their integrals with respect to multiple planar Gaussian measures in (8) are well-defined.

Remark 2. Note that[2]

$$\begin{aligned}\frac{1}{\pi^r}\underbrace{\int_{\mathbb{C}}\cdots\int_{\mathbb{C}}}_{r} t_1^{k_1}\cdots t_r^{k_r}\cdot\overline{t_1^{l_1}\cdots t_r^{l_r}}\prod_{j=1}^{r} e^{-|t_j|^2}dA(t_j) &= \prod_{j=1}^{r}\delta_{k_jl_j}k_j! \\ &= \langle t_1^{k_1}\cdots t_r^{k_r}, t_1^{l_1}\cdots t_r^{l_r}\rangle,\end{aligned} \tag{11}$$

where $\langle\cdot,\cdot\rangle$ is Macdonald's scalar product on symmetric functions ([Mac], Eq. (4.7)), written in terms of the normalized power sum variables $t_i = p_i/i$. Therefore we can rephrase (8) as

$$\sum_{n=0}^{\infty}\frac{A_{d+1,r}(n)}{(rn)!n!}|v|^{2rn} = \left\langle \det_{1\le i,j\le d}(\varphi_{i-j}(v,\mathbf{t}))\,, e^{\bar{v}^r\,\mathrm{Cyc}_r(\mathbf{t})}\right\rangle. \tag{12}$$

Remark 3. The last integral with respect to t_r in (8) can be replaced by the direct evaluation

$$\bar{t}_r = \frac{\bar{v}^r}{r} \tag{13}$$

as a consequence of the integral identity

$$\frac{1}{\pi}\int_{\mathbb{C}} P(t_r)e^{\frac{\bar{v}^r}{r}\bar{t}_r-|t_r|^2}dA(t_r) = P\left(\frac{\bar{v}^r}{r}\right) \qquad \text{for any polynomial } P(t). \tag{14}$$

The $1/r$ factor in (13) comes from

$$[t_r]\,\mathrm{Cyc}_r(\mathbf{t}) = \frac{1}{r!}\left|C_{(r)}\right| = \frac{(r-1)!}{r!} = \frac{1}{r}. \tag{15}$$

This shows that for $r=1$ the generating function (8) reduces to Gessel's Toeplitz determinant (5).

Remark 4. For the case of double occurrences $r=2$, there is a *quadrature identity*

$$\frac{1}{\pi^2}\int_{\mathbb{C}}\int_{\mathbb{C}} Q(t_1,t_2)e^{\frac{\bar{v}^2}{2}(\bar{t}_1^2+\bar{t}_2)-|t_1|^2-|t_2|^2}dA(t_1)dA(t_2) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} Q\left(\bar{v}x, \frac{\bar{v}^2}{2}\right)e^{-\frac{x^2}{2}}dx \tag{16}$$

[2] I am thankful for M. Bertola for this observation.

for polynomials $Q(t_1, t_2)$, assuming that $\left|\bar{v}^2/2\right| < 1$. Therefore, for $r = 2$ our general formula (8) reduces to

$$\sum_{n=0}^{\infty} \frac{A_{d+1,2}(n)}{(2n)!n!} |v|^{4n} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \det_{1\le i,j\le d} \left(\varphi_{i-j}(v, \bar{v}, x)\right) e^{-\frac{x^2}{2}} dx \tag{17}$$

with Toeplitz symbol

$$\varphi(v, \bar{v}, x; z) = \sum_{m=-\infty}^{\infty} \varphi_m(v, \bar{v}, x) z^m = \exp\left(\frac{v}{z} + \bar{v}xz + \frac{\bar{v}^2}{2} z^2\right). \tag{18}$$

By setting $\bar{v} = 1$ we can further simplify to

$$\sum_{n=0}^{\infty} \frac{A_{d+1,2}(n)}{(2n)!n!} v^{2n} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \det_{1\le i,j\le d} \left(\varphi_{i-j}(v, x)\right) e^{-\frac{x^2}{2}} dx \tag{19}$$

where

$$\varphi(v, x; z) = \exp\left(\frac{v}{z} + xz + \frac{1}{2} z^2\right). \tag{20}$$

Remark 5. The numbers $A_{d+1,r}(n)$ can be given a standard probabilistic interpretation described as follows. Consider the uniform measure on the set of words of length rn in n letters with each letter appearing exactly r times, and let $L_{n,r}(w)$ denote the length of the longest increasing subsequence in a random word w, interpreted itself as a random variable. By the definition of $A_{d+1,r}(n)$ we have

$$\operatorname{Prob}\left(L_{n,r}(w) \le d\right) = \frac{A_{d+1,r}(n)}{\frac{(rn)!}{(r!)^n}}, \tag{21}$$

and hence a slightly adapted version of (8) can be interpreted as the "Poissonization" of the distributions of $L_{n,r}$:

$$\begin{aligned} e^{-\frac{|v|^{2r}}{r!}} \sum_{n=0}^{\infty} \operatorname{Prob}\left(L_{n,r}(w) \le d\right) \frac{1}{n!} \left(\frac{|v|^{2r}}{r!}\right)^n \\ = \frac{e^{-\frac{|v|^{2r}}{r!}}}{\pi^r} \underbrace{\int_{\mathbb{C}} \cdots \int_{\mathbb{C}}}_{r} \det_{1\le i,j\le d} \left(\varphi_{i-j}(v, \mathbf{t})\right) \cdot e^{\bar{v}^r \overline{\mathrm{Cyc}_r(\mathbf{t})}} \prod_{j=1}^{r} e^{-|t_j|^2} dA(t_j). \end{aligned} \tag{22}$$

In the probabilistic context (22) may be useful in a subsequent asymptotic analysis of the probabilities $\operatorname{Prob}\left(L_{n,r}(w) \le d\right)$ as $n \to \infty$ and $d \to \infty$ simultaneously, possibly via the Toeplitz–Fredholm techniques of [BOO] combined with the de-Poissonization lemma of [J].

Outline. In Sec. 2 an integral representation of $g_\lambda^{(r)}$ is given in terms of the Schur polynomial s_λ associated to the partition λ, which in turn leads to an integral formula for $A_{d+1,r}(n)$. Gessel's general Toeplitz determinant for length-restricted Cauchy–Littlewood-type sums over partitions is recalled in Sec. 3 and it is employed to complete the proof of Theorem 1. In Sec. 4 we illustrate how the abstract generating function can be used in practice to calculate $A_{d+1,r}(n)$ for small values of d, r and n by using standard symbolic computer algebra packages.

Acknowledgements. The author is grateful to Marco Bertola for his interest in this project and helpful discussions, and, in particular, for pointing out the alternative scalar product interpretation (12). Thanks also to Doron Zeilberger for his encouragement and kind feedback on a previous version of this manuscript, along with a donation to the OEIS in the name of the author. The present work is supported in part by the *Fonds de recherche du Québec - Nature* *et technologies* (FRQNT) via the research grant *Matrices Aléatoires, Processus Stochastiques et Systèmes Intégrables.* The author acknowledges the excellent working conditions at Concordia University and at the Centre des Recherches Mathématiques (CRM).

2. The Kostka coefficients $g_\lambda^{(r)}$ as integrals of Schur polynomials

Following the standard approach in the theory of symmetric functions [Mac], one defines the Schur polynomials $s_\lambda \in \mathbb{C}[t_1, t_2, t_3, \dots]$ in the normalized power sum variables $\mathbf{t} = (t_1, t_2, \dots)$ via the Jacobi-Trudi formula

$$s_\lambda(\mathbf{t}) = \det_{1 \le i,j \le \ell(\lambda)} \left(h_{\lambda_i - i + j}(\mathbf{t}) \right), \tag{23}$$

where the polynomials $h_k(\mathbf{t})$ (the complete symmetric polynomials) are obtained from the generating function

$$\sum_{k=0}^{\infty} h_k(\mathbf{t}) z^k = \exp\left(\sum_{j=1}^{\infty} t_j z^j \right) \tag{24}$$

with $h_k(\mathbf{t}) = 0$ for negative values of k. The polynomial $h_k(\mathbf{t})$ is a weighted homogeneous polynomial of degree k in the natural grading

$$\deg(t_j) = j. \tag{25}$$

Therefore $s_\lambda(\mathbf{t})$ is also a weighted homogeneous polynomial of degree $|\lambda|$ and hence it has the scaling property

$$s_\lambda\left(at_1, a^2 t_2, a^3 t_3, \dots\right) = a^{|\lambda|} s_\lambda(t_1, t_2, t_3, \dots). \tag{26}$$

It is well-known [Mac] that the number of standard Young tableaux of shape λ is given by the special evaluation

$$f_\lambda = |\lambda|! s_\lambda\left(1, 0, 0, \dots\right). \tag{27}$$

As it is pointed out in [Ch], $g_\lambda^{(r)}$ is the Kostka number $K_{\lambda,(r^n)}$ associated to the pair of partitions λ and (r^n), that is, $g_\lambda^{(r)}$ is the coefficient of $m_{(r^n)}$ in the expansion of s_λ in the monomial symmetric function basis $\{m_\mu\}_\mu$ of $\mathbb{C}[\mathbf{t}]$ (see [Mac]):

$$g_\lambda^{(r)} = \left[x_1^r x_2^r \cdots x_n^r\right] s_\lambda(t_1(\mathbf{x}), t_2(\mathbf{x}), \dots), \tag{28}$$

where

$$t_j(\mathbf{x}) = \frac{1}{j} \sum_{i=1}^{n} x_i^j \qquad j = 1, 2, \dots \tag{29}$$

are the *normalized power sum symmetric polynomials* in the variables $x_1, \dots, x_n$.

In what follows, the coefficient representation (28) of $g_\lambda^{(r)}$ will be rewritten as an r-fold integral of the Schur polynomial associated to λ in terms of the normalized power sum variables. To proceed in this direction we need the following basic but important

Proposition 1. *Let $P(t_1, t_2, \dots, t_r)$ a polynomial in the indeterminates $t_1, \dots, t_r$ and consider the polynomial*

$$P\left(t_1(\mathbf{x}), t_2(\mathbf{x}), \dots, t_r(\mathbf{x})\right) \in \mathbb{C}[x_1, \dots, x_n] \tag{30}$$

in the indeterminates $x_1, \dots, x_n$, where $t_k(\mathbf{x})$ are the normalized power sum symmetric polynomials (29) in $\mathbf{x}$. The following integral identity holds:

$$(31)\quad [x_1^r x_2^r \cdots x_n^r] P\left(t_1(\mathbf{x}), t_2(\mathbf{x}), \dots, t_r(\mathbf{x})\right) = \frac{1}{\pi^r} \underbrace{\int_{\mathbb{C}} \cdots \int_{\mathbb{C}}}_{r} P(u_1, \dots, u_r) \overline{h_r\left(u_1, \frac{u_2}{2}, \dots, \frac{u_r}{r}, 0, 0, \dots\right)}^n \prod_{j=1}^{r} e^{-|u_j|^2} dA(u_j),$$

where $h_r\left(u_1, \frac{u_2}{2}, \dots, \frac{u_r}{r}, 0, 0, \dots\right)$ is the complete symmetric polynomial of degree r evaluated at $\left(u_1, \frac{u_2}{2}, \dots, \frac{u_r}{r}, 0, 0, \dots\right)$ and dA is the area measure in the complex plane.

Remark 6. A multiple integral formula for $g_\lambda^{(r)}$ involving h_r^n was found in [Ch], expressed in terms of the original indeterminates x_j. Our identity (31), expressed in terms of the normalized power sum symmetric polynomials $t_k(\mathbf{x})$, has essentially the same combinatorial content as the one presented in [Ch]. As it will be clear below, our choice of coordinates turns out to be more advantageous in finding a generating function for the numbers $A_{d+1,r}(n)$.

Proof of Lemma 1. For any r-tuple of non-negative integers $k_1, k_2, \dots, k_r$ let

$$(32)\quad a_{k_1,\dots,k_r}^{n,r} := [x_1^r x_2^r \cdots x_n^r] t_1(\mathbf{x})^{k_1} \cdots t_r(\mathbf{x})^{k_r}.$$

The exponential generating function of the numbers $a_{k_1,\dots,k_r}^{n,r}$ can be calculated as follows:

$$(33)\quad \sum_{k_1,\dots,k_r=0}^{\infty} a_{k_1,\dots,k_r}^{n,r} \prod_{j=1}^{r} \frac{u_j^{k_j}}{k_j!} = [x_1^r x_2^r \cdots x_n^r] \sum_{k_1,\dots,k_r=0}^{\infty} \prod_{j=1}^{r} \frac{(t_j(\mathbf{x}) u_j)^{k_j}}{k_j!}$$

$$(34)\quad = [x_1^r x_2^r \cdots x_n^r] \exp\left(\sum_{j=1}^{r} t_j(\mathbf{x}) u_j\right)$$

$$(35)\quad = [x_1^r x_2^r \cdots x_n^r] \prod_{i=1}^{n} \exp\left(\sum_{j=1}^{r} \frac{x_i^j}{j} u_j\right)$$

$$(36)\quad = \prod_{i=1}^{n} [x_i^r] \exp\left(\sum_{j=1}^{r} \frac{x_i^j}{j} u_j\right)$$

$$(37)\quad = \left(h_r\left(u_1, \frac{u_2}{2}, \dots, \frac{u_r}{r}, 0, 0, \dots\right)\right)^n.$$

Since

$$(38)\quad \frac{1}{\pi} \int_{\mathbb{C}} z^k \overline{z}^l e^{-|z|^2} dA(z) = \delta_{kl} k!$$

we have

$$(39)\quad \frac{1}{\pi^r} \underbrace{\int_{\mathbb{C}} \cdots \int_{\mathbb{C}}}_{r} u_1^{k_1} \cdots u_r^{k_r} \overline{\left(h_r\left(u_1, \frac{u_2}{2}, \dots, \frac{u_r}{r}, 0, 0, \dots\right)\right)^n} \prod_{j=1}^{r} e^{-|u_j|^2} dA(u_j) = a_{k_1,\dots,k_r}^{n,r} = [x_1^r x_2^r \cdots x_n^r] t_1(\mathbf{x})^{k_1} \cdots t_r(\mathbf{x})^{k_r}$$

for all $k_1, k_2, \dots, k_r$, and hence (31) follows by taking linear combinations. $\square$

Corollary 1. *For a partition of weight $|\lambda| = rn$ we have*

(40)
$$g_\lambda^{(r)} = \frac{1}{\pi^r} \underbrace{\int_{\mathbb{C}} \cdots \int_{\mathbb{C}}}_{r} s_\lambda(t_1, \dots, t_r, 0, \dots) \overline{\left(h_r\left(t_1, \frac{t_2}{2}, \dots, \frac{t_r}{r}, 0, \dots\right)\right)^n} \prod_{j=1}^{r} e^{-|t_j|^2} dA(t_j)$$

Note that for a pair of weighted homogeneous polynomials P and Q,

(41)
$$\frac{1}{\pi^r} \underbrace{\int_{\mathbb{C}} \cdots \int_{\mathbb{C}}}_{r} P(t_1, \dots, t_r) \overline{Q(t_1, \dots, t_r)} \prod_{j=1}^{r} e^{-|t_j|^2} dA(t_j) = 0 \qquad \text{unless} \quad \deg P = \deg Q.$$

Since s_λ and h_r^m are weighted homogeneous polynomials with $\deg(s_\lambda) = |\lambda|$ and $\deg(h_r^n) = rm$ we have

(42)
$$\frac{1}{\pi^r} \underbrace{\int_{\mathbb{C}} \cdots \int_{\mathbb{C}}}_{r} s_\lambda(t_1, \dots, t_r, 0, \dots) \overline{\left(h_r\left(t_1, \frac{t_2}{2}, \dots, \frac{t_r}{r}, 0, \dots\right)\right)^m} \prod_{j=1}^{r} e^{-|t_j|^2} dA(t_j) = \delta_{|\lambda|, rm} g_\lambda^r.$$

Remark 7. For example, for $r = 1, 2, 3$ we have

$$h_1(u_1, 0, \dots) = u_1$$
$$h_2\left(u_1, \frac{u_2}{2}, 0, \dots\right) = \frac{1}{2}u_1^2 + \frac{1}{2}u_2$$
$$h_3\left(u_1, \frac{u_2}{2}, \frac{u_3}{3}, 0, \dots\right) = \frac{1}{6}u_1^3 + \frac{1}{2}u_1^2 u_2 + \frac{1}{3}u_3.$$

The general formula is (cf. [Mac], (2,14'))

(43)
$$h_r\left(u_1, \frac{u_2}{2}, \dots, \frac{u_r}{r}, 0, \dots\right) = \sum_{|\mu|=r} \frac{1}{z_\mu} \prod_{i=1}^{\ell(\mu)} u_{\mu_i},$$

where z_μ is the standard constant factor

(44)
$$z_\mu = \prod_{i \geq 1} i^{m_i} \cdot m_i!$$

in which $m_i = m_i(\mu)$ is number of parts of the partition μ equal to i. Note that

(45)
$$|C_\mu| = \frac{r!}{z_\mu} \qquad |\mu| = r$$

where C_μ is the conjugacy class of permutations of cycle type μ in the symmetric group S_r. Hence $h_r\left(u_1, \frac{u_2}{2}, \dots, \frac{u_r}{r}, 0, 0, \dots\right)$ can be written alternatively as the cycle index polynomial of S_r:

(46)
$$h_r\left(u_1, \frac{u_2}{2}, \dots, \frac{u_r}{r}, 0, \dots\right) = \frac{1}{r!} \sum_{|\lambda|=r} |C_\lambda| \prod_{i=1}^{\ell(\lambda)} u_{\lambda_i} = \mathrm{Cyc}_r(u_1, u_2, \dots, u_r).$$

By taking into account (42) and (46) we obtain the following

Corollary 2.

(47)
$$\frac{1}{\pi^r} \underbrace{\int_{\mathbb{C}} \cdots \int_{\mathbb{C}}}_{r} s_\lambda(t_1, \dots, t_r, 0, \dots) e^{\bar{v}^r \overline{\mathrm{Cyc}_r(t_1, \dots, t_r)}} \prod_{j=1}^{r} e^{-|t_j|^2} dA(t_j) = \begin{cases} \frac{\bar{v}^{rn}}{n!} g_\lambda^{(r)} & |\lambda| = rn \\ 0 & \textit{otherwise.} \end{cases}$$

By the RSK formula this leads to an integral representation of $A_{d+1,r}(n)$:

Lemma 1. *The following identity holds for all meaningful values of d, r and n:*

$$\frac{A_{d+1,r}(n)}{(rn)!n!}|v|^{2rn} = \frac{1}{\pi^r}\underbrace{\int_{\mathbb{C}}\cdots\int_{\mathbb{C}}}_{r}\left[\sum_{\substack{|\lambda|=rn\\ \ell(\lambda)\le d}} s_\lambda(v,0,\dots)s_\lambda(t_1,\dots,t_r,0,\dots)\right]\times \tag{48}$$
$$\times\, e^{\bar{v}^r\overline{\mathrm{Cyc}_r(t_1,\dots,t_r)}}\prod_{j=1}^{r} e^{-|t_j|^2}dA(t_j).$$

Proof. By using the elementary identity

$$\frac{f_\lambda}{|\lambda|!}v^\lambda = s_\lambda(v,0,0,\dots) \tag{49}$$

we obtain

$$\frac{A_{d+1,r}(n)}{(rn)!n!}|v|^{2rn} \tag{50}$$

$$= \sum_{\substack{|\lambda|=rn\\ \ell(\lambda)\le d}} \frac{f_\lambda}{|\lambda|!}v^{|\lambda|}\cdot\frac{g_\lambda^{(r)}}{n!}\bar{v}^{rn}, \tag{51}$$

$$= \sum_{\substack{|\lambda|=rn\\ \ell(\lambda)\le d}} s_\lambda(v,0,\dots)\cdot\frac{1}{\pi^r}\underbrace{\int_{\mathbb{C}}\cdots\int_{\mathbb{C}}}_{r} s_\lambda(t_1,\dots,t_r,0,\dots)e^{\bar{v}^r\overline{\mathrm{Cyc}_r(t_1,\dots,t_r)}}\prod_{j=1}^{r} e^{-|t_j|^2}dA(t_j) \tag{52}$$

$$= \frac{1}{\pi^r}\underbrace{\int_{\mathbb{C}}\cdots\int_{\mathbb{C}}}_{r}\left[\sum_{\substack{|\lambda|=rn\\ \ell(\lambda)\le d}} s_\lambda(v,0,\dots)s_\lambda(t_1,\dots,t_r,0,\dots)\right]e^{\bar{v}^r\overline{\mathrm{Cyc}_r(t_1,\dots,t_r)}}\prod_{j=1}^{r} e^{-|t_j|^2}dA(t_j). \tag{53}$$

□

3. Gessel's Toeplitz determinant identity and the proof of the main theorem

Recall Gessel's determinant formula [Ge]: the length-restricted Cauchy–Littlewood-type sum of products of Schur functions can be written as

$$\sum_{\ell(\lambda)\le d} s_\lambda(\mathbf{t})s_\lambda(\mathbf{s}) = \det_{1\le i,j\le d}\left(\varphi_{i-j}(\mathbf{t},\mathbf{s})\right) \tag{54}$$

where $\mathbf{t} = (t_1, t_2, \dots), \mathbf{s} = (s_1, s_2, \dots)$ and the symbol φ associated to the above Toeplitz determinant is

$$\varphi(\mathbf{t},\mathbf{s};z) = \exp\left(\sum_{k=1}^{\infty} t_k z^k + \sum_{k=1}^{\infty} s_k z^{-k}\right) = \sum_{m=-\infty}^{\infty}\varphi_m(\mathbf{t},\mathbf{s})z^m. \tag{55}$$

This "master identity", combined with the integral representation for $A_{d+1,r}(n)$ is sufficient to prove the main result of this paper.

Proof of Theorem 1. Gessel's determinantal formula (54) with

$$\mathbf{t} = (t_1,\dots,t_r,0,,0\dots) \quad \text{and} \quad \mathbf{s} = (v,0,0,\dots) \tag{56}$$

gives

$$\sum_{\ell(\lambda)\le d} s_\lambda(v,0,\dots)s_\lambda(t_1,\dots,t_r,0,\dots) = \det_{1\le i,j\le d}\left(\varphi_{i-j}(v,\mathbf{t})\right) \tag{57}$$

with Toeplitz symbol

$$\varphi(v,\mathbf{t};z) = \sum_{m=-\infty}^{\infty} \varphi_m(v,\mathbf{t})z^m = \exp\left(\frac{v}{z} + \sum_{j=1}^{r} t_j z^j\right). \tag{58}$$

Therefore

$$\frac{1}{\pi^r}\underbrace{\int_{\mathbb{C}}\cdots\int_{\mathbb{C}}}_{r} \det_{1\le i,j\le d}\left(\varphi_{i-j}(v,\mathbf{t})\right)\cdot e^{\bar{v}^r\overline{\mathrm{Cyc}_r(\mathbf{t})}}\prod_{j=1}^{r} e^{-|t_j|^2}dA(t_j) \tag{59}$$

$$= \frac{1}{\pi^r}\underbrace{\int_{\mathbb{C}}\cdots\int_{\mathbb{C}}}_{r}\left[\sum_{\ell(\lambda)\le d} s_\lambda(v,0,\dots)s_\lambda(t_1,\dots,t_r,0,\dots)\right]\cdot e^{\bar{v}^r\overline{\mathrm{Cyc}_r(\mathbf{t})}}\prod_{j=1}^{r} e^{-|t_j|^2}dA(t_j) \tag{60}$$

$$= \frac{1}{\pi^r}\underbrace{\int_{\mathbb{C}}\cdots\int_{\mathbb{C}}}_{r}\sum_{n=0}^{\infty}\left[\sum_{\substack{\ell(\lambda)\le d\\ |\lambda|=rn}} s_\lambda(v,0,\dots)s_\lambda(t_1,\dots,t_r,0,\dots)\right]\cdot e^{\bar{v}^r\overline{\mathrm{Cyc}_r(\mathbf{t})}}\prod_{j=1}^{r} e^{-|t_j|^2}dA(t_j), \tag{61}$$

where in the last step we used the simple observation that (47) implies that a partition λ may give a non-trivial contribution to the integral only if its weight $|\lambda|$ is divisible by r. The proof is concluded by comparing (61) with the integral representation (48) of $A_{d+1,r}(n)$. □

4. Calculation of $A_{d+1,r}(n)$ using symbolic computer algebra

The purpose of this section is to illustrate that the generating function (8) can be used to calculate $A_{d+1,r}(n)$ for small values of d, r and n using basic symbolic computer algebra. There was no attempt made to optimize the algorithms used in the symbolic computation, our goal was simply to demonstrate that the generating function is not only a compact and abstract formula, but it can also be put to work in practice. This also provides numerical evidence that the series (8) does give the integer numbers it is supposed to, which, after all, is the most rewarding part of finding a generating function.

Below we use the short-hand notation $\mathbf{t}_r = (t_1,\dots,t_r)$. It is easy to see that the Laurent series coefficients of the Toeplitz symbol $\varphi(v,\mathbf{t}_r;z)$ are

$$\varphi_m(v,\mathbf{t}_r) = \sum_{k=0}^{\infty} h_{k+m}(\mathbf{t}_r)\frac{v^k}{k!} \quad m=0,1,2,\dots \tag{62}$$

$$\varphi_{-m}(v,\mathbf{t}_r) = \sum_{k=m}^{\infty} h_{k-m}(\mathbf{t}_r)\frac{v^k}{k!} \quad m=1,2,\dots, \tag{63}$$

where $h_k(\mathbf{t}_r)$ are the complete symmetric polyomials in $\mathbf{t}_r$. We can define the truncated Toeplitz entries

$$\varphi_m^{(N)}(v,\mathbf{t}_r) = \sum_{k=0}^{rN} h_{k+m}(\mathbf{t}_r)\frac{v^k}{k!} \quad m = 0,1,2,\dots \tag{64}$$

$$\varphi_{-m}^{(N)}(v,\mathbf{t}_r) = \sum_{k=m}^{rN} h_{k-m}(\mathbf{t}_r)\frac{v^k}{k!} \quad m = 1,2,\dots \tag{65}$$

so that

$$\varphi_m(v,\mathbf{t}_r) = \varphi_m^{(N)}(v,\mathbf{t}_r) + \mathcal{O}\left(v^{rN+1}\right). \tag{66}$$

For small values of d the Toeplitz determinant with $\mathcal{O}\left(v^{rN+1}\right)$-truncated entries can be evaluated, and it can be further simplified by eliminating the terms of $\mathcal{O}\left(v^{rN+1}\right)$. We note that the evaluation of the determinant can be done much more efficiently by employing the division-free algorithm suggested in [ESZ].

The second factor in (8) $e^{\overline{v}^r \operatorname{Cyc}_t(\bar{t}_1,\dots,\bar{t}_r)}$ has the truncated expansion in $\overline{v}$ of the form

$$T_r^{(N)}(\bar{t}_1,\bar{t}_2/2,\dots,\bar{t}_r/r;\bar{v}) = \sum_{k=0}^{N} h_r(\bar{t}_1,\bar{t}_2/2,\dots,\bar{t}_r/r,0,0)^k\frac{\bar{v}^{kr}}{k!} \tag{67}$$

for which

$$e^{\overline{v}^r \operatorname{Cyc}_t(\bar{t}_1,\dots,\bar{t}_r)} = T_r^{(N)}(\bar{t}_1,\bar{t}_2/2,\dots,\bar{t}_r/r;\bar{v}) + \mathcal{O}\left(\bar{v}^{rN+r}\right). \tag{68}$$

Therefore by combining the truncated factors of the integrand we find

$$\begin{aligned}\frac{1}{\pi^r}\underbrace{\int_{\mathbb{C}}\cdots\int_{\mathbb{C}}}_{r} \det_{1\le i,j\le d}\left(\varphi_{i-j}^{(N)}(v,\mathbf{t}_r)\right) T_r^{(N)}(\bar{t}_1,\bar{t}_2/2,\dots,\bar{t}_r/r;\bar{v})\prod_{j=1}^{r} e^{-|t_j|^2}dA(t_j)\\ = \sum_{n=0}^{N}\frac{A_{d+1,r}(n)}{(rn)!n!}|v|^{2rn} + \mathcal{O}\left(|v|^{2rN+2r}\right).\end{aligned} \tag{69}$$

The computation of area integrals can be implemented easily by using polar coordinates

$$\frac{1}{\pi}\int_{\mathbb{C}} F(t)e^{-|t|^2}dA(t) = 2\int_0^{\infty}\left[\frac{1}{2\pi i}\oint_{|u|=1} F(ru)\frac{du}{u}\right]e^{r^2}rdr \tag{70}$$

and evaluate the contour integrals using a residue calculation at $u=0$ ($F(u)$ is a polynomial in u for any variable u in our setting.

This simple method was implemented as a Maple code and we obtained the values shown in Tables 1-3 These all coincide with the known values presented in various appendices of the journal entry [EZ].

$d\backslash n$	1	2	3	4	5	6	7	8	9	10
1	1	1	1	1	1	1	1	1	1	1
2	1	6	43	352	3114	29004	280221	2782476	28221784	291138856
3	1	6	90	1879	47024	1331664	41250519	1367533365	47808569835	1744233181074
4	1	6	90	2520	102011	5176504	307027744	20472135280	1496594831506	117857270562568

TABLE 1. The calculated values of $A_{d+1,2}(n)$

$d\backslash n$	1	2	3	4	5
1	1	1	1	1	1
2	1	20	374	8124	190893
3	1	20	1680	173891	21347262
4	1	20	1680	369600	117392909

Table 2. The calculated values of $A_{d+1,3}(n)$

$d\backslash n$	1	2	3	4	5
1	1	1	1	1	1
2	1	70	3199	173860	10203181
3	1	70	34650	16140983	8854463421
4	1	70	34650	63063000	142951955371

Table 3. The calculated values of $A_{d+1,4}(n)$

5. Conclusion and outlook

In this paper we presented the generating function (8) to solve the challenge posed by Ekhad and Zeilberger [EZ] to generalize Gessel's Bessel determinant formula (3) for arbitrary values of the number of occurrences r. The key step in constructing these generating functions was to express the coefficient $g_\lambda^{(r)}$ as a use Gessel's "master identity" to sum up the resulting Cauchy–Littlewood-type terms under the integrals.

We expect that the general formula (8) may be useful to find the associated linear differential equation with polynomial coefficients for the generating function, or equivalently, to find the recurrence relation with polynomial coefficients for $A_{d+1,r}(n)$ for arbitrary values of r in a closed form.

Moreover, one can easily construct generating functions analogous to (8) for words with different types of multiple occurrences in its letters.

It may also be interesting to investigate the asymptotic distribution obtained from a fine-tuned scaling limit of the probability $\text{Prob}\,(L_{n,r}(w) \leq d)$ as $n, d \to \infty$. It is then natural to ask which universality class the limiting probability density belongs to, either a to a known one or to one of a new type that has not yet been characterized.

References

[AD] D. Aldous, P. Diaconis, Longest increasing subsequences: from patience sorting to the Baik-Deift-Johansson theorem. *Bull. Amer. Math. Soc.* **36** (4) 413-432 (1999).

[BDJ] J. Baik, P. Deift, K. Johansson, On the distribution of the length of the longest increasing subsequence of random permutations. *J. Amer. Math. Soc.* **12** (4) 1119-1178 (1999)

[BOO] A. Borodin, A. Okounkov, G. Olshanski, Asymptotics of Plancherel measures for symmetric groups. *J. Amer. Math. Soc.* **13** (3) 481-515 (2000)

[Ch] G. Chapuy, The asymptotic number of 12..d-avoiding words with r occurrences of each letter $1, 2, ..., n$. arXiv:1412.6070 (2014)

[DLMF] NIST Digital Library of Mathematical Functions. http://dlmf.nist.gov/, Release 1.0.9 of 2014-08-29.

[EZ] S. B. Ekhad and D. Zeilberger. The Generating Functions Enumerating 12..d-Avoiding Words with r occurrences of each of $1, 2, ..., n$ are D-finite for all d and all r. *The Personal Journal of Shalosh B. Ekhad and Doron Zeilberger, Dec. 5, 2014*, arXiv:1412.2035 (2014)

[ESZ] S. B. Ekhad, N. Shar, D. Zeilberger, The number of 1...d-avoiding permutations of length d+r for SYMBOLIC d but numeric r. *The Personal Journal of Shalosh B. Ekhad and Doron Zeilberger, April 9, 2015*, arXiv:1504.02513 (2015)

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

[ITW1] A. Its, C. Tracy, H. Widom, Random words, Toeplitz determinants, and integrable systems. I. Random matrix models and their applications, 245-258, *Math. Sci. Res. Inst. Publ.* **40** Cambridge Univ. Press (2001)

[ITW2] A. Its, C. Tracy, H. Widom, Random words, Toeplitz determinants and integrable systems. II. Advances in nonlinear mathematics and science. *Phys. D* **152/153** 199-224 (2001)

[J] K. Johansson, The longest increasing subsequence in a random permutation and a unitary random matrix model. *Math. Res. Lett.* **5** (1-2) 63-82 (1998)

[Mac] I. Macdonald, Symmetric functions and Hall polynomials, Oxford Mathematical Monographs. The Clarendon Press Oxford University Press, New York. Second edition (1995)

[vM] P. van Moerbeke, Random and integrable models in mathematics and physics. Random matrices, random processes and integrable systems, 3-130, *CRM Ser. Math. Phys.,* Springer, New York (2011)

[St] R.P. Stanley, Enumerative Combinatorics. Vol. 2. *Cambridge Studies in Advanced Mathematics* **62**, Cambridge University Press (1999)

[TW] C. Tracy, H. Widom, On the distributions of the lengths of the longest monotone subsequences in random words. *Probab. Theory Related Fields* **119** (3) 350-380 (2001)

Department of Mathematics and Statistics
Concordia University
1455 de Maisonneuve Blvd. West
Montréal, Québec, Canada
H3G 1M8
and
Centre de recherches mathématiques
Université de Montréal
C. P. 6128, succ. centre ville
Montréal, Québec, Canada
H3C 3J7

E-mail address: ferenc.balogh@concordia.ca