Expectations of hook products on large partitions

M. Adler* A. Borodin† P. van Moerbeke‡

Abstract

Given uniform probability on words of length $M = Np + k$, from an alphabet of size p, consider the probability that a word (**i**) contains a subsequence of letters $p, p-1, \ldots, 1$ in that order and (**ii**) that the maximal length of the disjoint union of $p-1$ increasing subsequences of the word is $\leq M - N$. A generating function for this probability has the form of an integral over the Grassmannian of p-planes in $\mathbb{C}^n$. The present paper shows that the asymptotics of this probability, when $N \to \infty$, is related to the k^{th} moment of the χ^2-distribution of parameter $2p^2$. This is related to the behavior of the integral over the Grassmannian $Gr(p, \mathbb{C}^n)$ of p-planes in $\mathbb{C}^n$, when the dimension of the ambient space $\mathbb{C}^n$ becomes very large. A different scaling limit for the Poissonized probability is related to a new matrix integral, itself

*2000 *Mathematics Subject classification.* 60C55, 60F05, 05A05, 05A15, 34M55 .
Key words and phrases. Integrals over Grassmannians, longest increasing sequences in random words, limit theorems, chi-square distribution, Painlevé equations.
Department of Mathematics, Brandeis University, Waltham, Mass 02454, USA. E-mail: adler@brandeis.edu. The support of a National Science Foundation grant # DMS-01-00782 is gratefully acknowledged.

†Department of Mathematics, California Institute of technology, Pasadena, CA 91125, USA and Clay mathematics institute Long-term Prize Fellow. E-mail: borodin@caltech.edu.

‡Department of Mathematics, Université de Louvain, 1348 Louvain-la-Neuve, Belgium and Brandeis University, Waltham, Mass 02454, USA. E-mail: vanmoerbeke@math.ucl.ac.be and @brandeis.edu. The support of a National Science Foundation grant # DMS-01-00782, a Nato, a FNRS and a Francqui Foundation grant is gratefully acknowledged. This work was done while PvM was a member of the Clay Mathematics Institute, One Bow Street, Cambridge, MA 02138, USA. MA and PvM would like to thank Persi Diaconis for interesting conversations regarding section 4.

a solution of the Painlevé IV equation. This is part of a more general set-up related to the Painlevé V equation.

Contents

1 Introduction **2**

2 A differential equation for a matrix integral **10**

3 Hermitian matrix integrals and Painlevé equations **15**

4 A limit theorem for the probability on words and Chi-square **20**

5 A limit theorem for the Poissonized probability on words **22**

6 Painlevé IV as a limit of Painlevé V **26**

7 Appendix: Chazy classes **28**

1 Introduction

Consider the set of words

$$\pi \in S_\ell^p := \{\text{words } \pi \text{ of length } \ell, \text{ built from an alphabet } \{1, ..., p\}\},$$

(with $|S_\ell^p| = p^\ell$) with the uniform probability distribution

$$P^{\ell,p}(\pi) = \frac{1}{p^\ell}. \tag{1.0.1}$$

Let $\mathbb{Y}$ denote the set of all partitions λ. Let $\lambda^\top$ be the dual partition, i.e., obtained by flipping the Young diagram λ about its diagonal. So, $\lambda_1^\top$ is the length of the first column of λ. Moreover h^λ denotes the product of the hook lengths $h_{ij}^\lambda := \lambda_i + \lambda_j^\top - i - j + 1$ and also 1^p denotes the infinite vector

$$1^p := (\overbrace{1, \ldots, 1}^{p}, 0, 0, \ldots).$$

The RSK correspondence between words and pairs of semi-standard and standard tableaux induces a probability measure on partitions

$$\lambda \in \mathbb{Y}_\ell = \{\text{partitions } \lambda \in \mathbb{Y} \text{ of weight } |\lambda| = \ell\}, \tag{1.0.2}$$

given by

$$P^{\ell,p}(\lambda) = \frac{f^\lambda \, \mathbf{s}_\lambda(1^p)}{p^{|\lambda|}}, \tag{1.0.3}$$

having

$$(\text{ support } P^{\ell,p}) \subseteq \mathbb{Y}_\ell^{(p)} := \{\lambda \in \mathbb{Y}_\ell, \text{ such that } \lambda_1^\top \le p\}.$$

The symbol $\mathbf{s}_\lambda$ denotes the Schur polynomial associated with the partition λ. Besides the probability $P^{\ell,p}$, we also consider the corresponding Poissonized measure, depending on the real parameter x,

$$P_{x,p}(\lambda) = e^{-px} \frac{(xp)^{|\lambda|}}{|\lambda|!} P^{|\lambda|,p}(\lambda), \qquad \lambda \in \mathbb{Y}^{(p)}, \tag{1.0.4}$$

having

$$(\text{ support } P_{x,p}) \subseteq \mathbb{Y}^{(p)} := \{\lambda \in \mathbb{Y}, \text{ such that } \lambda_1^\top \le p\}.$$

We discuss very briefly the combinatorics needed in this problem; for more details, see [17, 19, 20]. For the partition λ, define the symbol

$$(n)_\lambda := \prod_i (n+1-i)_{\lambda_i}, \text{ with } (x)_n = x(x+1)\dots(x+n-1), \; x_0 = 1 \tag{1.0.5}$$

Define for $q \ge \lambda_1^\top$, (throughout Δ_q denotes the Vandermonde determinant in q variables)

$$\begin{aligned} f^\lambda = \#\left\{ \begin{array}{l} \text{standard tableaux} \\ \text{of shape } \lambda \end{array} \right\} &= \frac{|\lambda|!}{h^\lambda} = \frac{|\lambda|!}{u^{|\lambda|}} \; \mathbf{s}_\lambda(x)\big|_{\sum_\ell x_\ell^i = \delta_{1i} u} \\ &= |\lambda|! \; \frac{\Delta_q(q+\lambda_1-1, \dots, q+\lambda_q-q)}{\prod_1^q (q+\lambda_i-i)!}, \end{aligned} \tag{1.0.6}$$

and

$$\#\left\{\begin{array}{l}\text{semi-standard tableaux}\\ \text{of shape } \lambda \text{ filled with}\\ \text{numbers from 1 to } q\end{array}\right\} = \mathbf{s}_\lambda(1^q) \;\; = \;\; \frac{\Delta_q(q+\lambda_1-1,\ldots,q+\lambda_q-q)}{\prod_{i=1}^{q-1} i!}$$

$$= \prod_{(i,j)\in\lambda} \frac{j-i+q}{h_{ij}^{\lambda}}$$

$$= \frac{(q)_\lambda}{h^\lambda} \tag{1.0.7}$$

A subsequence σ of the word π is *weakly k-increasing*, if it can be written as

$$\sigma = \sigma_1 \cup \sigma_2 \cup \ldots \cup \sigma_k, \tag{1.0.8}$$

where σ_i are disjoint weakly increasing subsequences of the word π, i.e., possibly with repetitions. The length of the longest increasing/decreasing subsequences is closely related to the shape of the associated partition, via the RSK correspondence :

$$\begin{aligned} d_1(\pi) &= \left\{\begin{array}{c}\text{length of the longest } \textit{strictly}\\ \text{decreasing subsequence of } \pi\end{array}\right\} = \lambda_1^\top \\ i_k(\pi) &= \left\{\begin{array}{c}\text{length of the longest } \textit{weakly}\\ k\text{-increasing subsequence of } \pi\end{array}\right\} = \lambda_1 + \ldots + \lambda_k \end{aligned} \tag{1.0.9}$$

Define the generalized hypergeometric function in terms of the symbol (1.0.5), viewed as a symmetric function in an infinite number of variables x_i

$${}_2F_1^{(1)}(p,q;n;x) := \sum_{\kappa\in\mathbb{Y}} \frac{(p)_\kappa (q)_\kappa}{(n)_\kappa} \frac{\mathbf{s}_\kappa(x)}{h^\kappa}, \tag{1.0.10}$$

which, upon restriction, using power sums and upon using (1.0.6), yields

$$\begin{aligned} {}_2F_1^{(1)}(p,q;n;x)\Big|_{\sum_\ell x_\ell^i = \delta_{1i} u} &= \sum_{\kappa\in\mathbb{Y}} u^{|\kappa|} \frac{(p)_\kappa (q)_\kappa}{(h^\kappa)^2 (n)_\kappa} \\ &= \sum_{k=0}^{\infty} u^k \sum_{\kappa\in\mathbb{Y}_k} \frac{(p)_\kappa (q)_\kappa}{(h^\kappa)^2 (n)_\kappa} \end{aligned} \tag{1.0.11}$$

As a reminder, the chi-square distribution of parameter m is the distribution of

$$Z_m = \sum_1^m X_i^2,$$

where the X_i's are m independent normal $N(0,1)$-random variables.

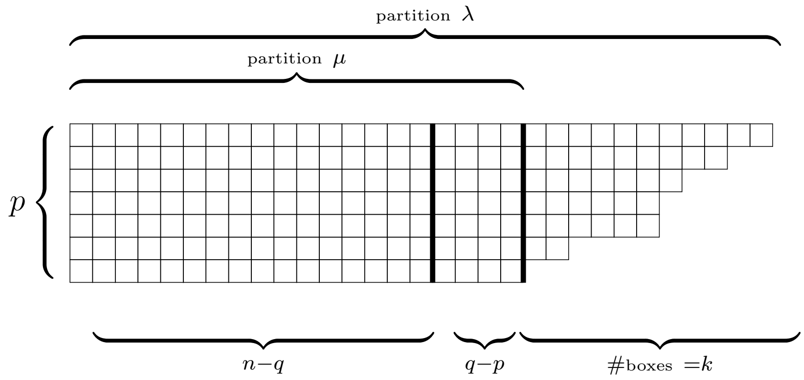

Figure 1

The expectations $E_{x,p}$ and $E^{\ell,p}$ are taken with regard to the probabilities $P_{x,p}$ and $P^{\ell,p}$ defined above. The functions on partitions, of which the expectations are taken, are products of hook lengths restricted to a vertical strip in the partition of width $q-p$, as in Figure 1. In [3], expectations of this type have been studied and linked to integrals over the Grassmannian space Gr(p,$\mathbb{C}^n$) of p-planes in $\mathbb{C}^n$, with regard to the Weyl measure $d\rho(Z)$; besides, these integrals relate to specific solutions of the Painlevé V equation.

We now state the following proposition, established in [3]:

Proposition 1.1 *Given the partition*

$$\mu = (n-p)^p := (\overbrace{n-p, n-p, \ldots, n-p}^{p}). \tag{1.0.12}$$

the mathematical expectation of the product of the hook lengths over the strip of width $q-p$, with regard to the probability (1.0.3), is given by[1]

[1]Remember (support $P^{\ell,p}$) $\subseteq \{\lambda \in \mathbb{Y},$ such that $|\lambda| = \ell,\ \lambda_1^\top \leq p\}$.

$$
\begin{aligned}
& e^{px}\ E_{x,p}\left(I_{\{\lambda\supseteq\mu\}}(\lambda)\prod_{\substack{(i,j)\in\lambda\\ n-q<j\leq n-p}} h^{\lambda}_{(i,j)}\right)\\
&= \sum_{\ell\geq p(n-p)}\frac{(px)^\ell}{\ell!}E^{\ell,p}\left(I_{\{\lambda\supseteq\mu\}}(\lambda)\prod_{\substack{(i,j)\in\lambda\\ n-q<j\leq n-p}} h^{\lambda}_{(i,j)}\right)\\
&= \tilde{\tilde{c}}x^{(n-p)p}\ {}_2F_1^{(1)}(p,q;n;y)\Big|_{\sum_\ell y_\ell^i=\delta_{1i}x}\\
&= \tilde{c}x^{(n-p)p}\int_{Gr(p,\mathbb{C}^n)} e^{x\,\mathrm{Tr}(I+Z^\dagger Z)^{-1}}\det(Z^\dagger Z)^{-(q-p)}d\rho(Z)\\
&= \tilde{\tilde{c}}x^{(n-p)p}\exp\int_0^x\frac{u(y)-p(n-p)+py}{y}dy, \qquad (1.0.13)
\end{aligned}
$$

where[2] $u(x)$ *is the unique solution to the initial value problem:*

$$
\begin{cases}
x^2u'''+xu''+6xu'^2-4uu'+4Qu'-2Q'u+2R=0 \\
\textit{with} \qquad\qquad\qquad\qquad \textbf{(Painlevé V)}\\
u(x)=p(n-p)-\dfrac{p(n-q)}{n}x+\ldots+a_{n+1}x^{n+1}+O(x^{n+2}),\ \textit{near } x=0,
\end{cases}
\qquad (1.0.14)
$$

with a specific coefficient a_{n+1} *and where*

$$
\begin{aligned}
4Q &= -x^2+2(n+2(p-q))x-(n-2p)^2\\
2R &= p(p-q)(x+n-2p). \qquad (1.0.15)
\end{aligned}
$$

This paper is concerned with what happens when $n\to\infty$. This is related to the behavior of the integral over the Grassmannian $Gr(p,\mathbb{C}^n)$ of p-planes in $\mathbb{C}^n$, when the dimension of the ambient space $\mathbb{C}^n$ becomes very large. To be precise:

---

[2] with

$$
\tilde{c}^{-1} := \prod_{i=1}^{p} i!\ (n-q-i)! \quad \text{and} \quad \tilde{\tilde{c}} := \frac{(q-i)!}{(n-i)!}.
$$

Theorem 1.2 *Given the partition $\mu = (n-p)^p$, as in Figure 1, the following expectation behaves, for large n, like the moments of the chi-square distribution of parameter $2pq$:*

$$\lim_{n\to\infty} n^{\frac{p^2-1}{2}} E^{p(n-p)+k,p}\left(I_{\lambda\supseteq\mu}(\lambda) \prod_{n-q<j\le n-p} h_{ij}^{\lambda}\right)$$

$$= \frac{\sqrt{p}\prod_1^p (q-j)!}{(\sqrt{2\pi})^{p-1}k!} E\left(\left(\frac{1}{2}Z_{2pq}\right)^k\right), \qquad (1.0.16)$$

and so the expectation decays, when $n\to\infty$, as

$$E^{k+p(n-p),p}\left(I_{\lambda\supseteq\mu}(\lambda) \prod_{n-q<j\le n-p} h_{ij}^{\lambda}\right) \simeq c_{p,q,k}\; n^{-\frac{p^2-1}{2}},$$

with

$$c_{p,q,k} := \frac{\sqrt{p}\prod_1^p (q-j)!}{(\sqrt{2\pi})^{p-1}} \binom{pq-1+k}{k}.$$

The next statement deals with the special case, where $q = p$. Namely, setting $N := n-p$, what is, asymptotically for large N, the probability $P^{Np+k,p}(\lambda_p \ge N)$?

Corollary 1.3 *Given an alphabet of size p and an integer $k > 0$, we give the behavior of the probability on the set of words of length $Np+k$ for large N. Notice $i_p(\pi) = \{$length of the word$\} = Np+k$ automatically, when $d_1(\pi) = p$. So, $i_{p-1}(\pi)$ is the first non-trivial quantity. We now have:*

$$\lim_{N\to\infty} N^{\frac{p^2-1}{2}} P^{Np+k,p}(\lambda_p \ge N)$$

$$= \lim_{N\to\infty} N^{\frac{p^2-1}{2}} P^{Np+k,p}\left(\begin{array}{c} d_1(\pi) = p \\ \\ i_{p-1}(\pi) \le N(p-1)+k \end{array}\right)$$

$$= \frac{\sqrt{p}\prod_1^p (p-j)!}{(\sqrt{2\pi})^{p-1}k!} E\left(\left(\frac{1}{2}Z_{2p^2}\right)^k\right).$$

Thus the following decay holds for $N \nearrow \infty$,

$$P^{Np+k,p}\left(\begin{array}{c} d_1(\pi)=p \\ \\ i_{p-1}(\pi)\leq N(p-1)+k \end{array}\right) \simeq c_{p,p,k}\left(\frac{1}{N}\right)^{\frac{p^2-1}{2}}.$$

The proofs of Theorem 1.2 and Corollary 1.3 will be given in section 4. In the next statement, we consider the expectation for the Poissonized probability $P_{x,p}$ on Young diagrams $\lambda \in \mathbb{Y}_\ell^{(p)}$, as defined in (1.0.4). Define

$$\mathcal{H}_p = \{p \times p \text{ Hermitian matrices}\},$$

and, for an interval $I \subset \mathbb{R}$,

$$\mathcal{H}_p(I) = \{M \in \mathcal{H}_p \text{ with spectrum in } I\}.$$

Theorem 1.4 *Take $x > 0$, $s \in \mathbb{R}$ and set*

$$n - p = x + s\sqrt{2x}, \tag{1.0.17}$$

we have, upon expressing n in terms of x by means of the rescaling (1.0.17) [3]

$$\begin{aligned}
&\lim_{x\to\infty}\frac{1}{(2x)^{\frac{p(q-p)}{2}}}E_{x,p}\left(I_{\lambda\supseteq\mu}(\lambda)\prod_{\substack{(i,j)\in\lambda\\ n-q<j\leq n-p}} h_{ij}^{\lambda}\right)\\
&= \lim_{x\to\infty}\hat{c}e^{-px}x^{p(n-\frac{p+q}{2})}\int_{Gr(p,\mathbb{C}^n)} e^{x\operatorname{Tr}(I+Z^\dagger Z)^{-1}}\det(Z^\dagger Z)^{-(q-p)}d\rho(Z)\\
&= p!\,\frac{\displaystyle\int_{\mathcal{H}_p[s,\infty)}\det(M-sI)^{q-p}e^{-\operatorname{Tr}M^2}dM}{\displaystyle\int_{\mathcal{H}_p}e^{-\operatorname{Tr}M^2}dM}\\
&= c\ \exp\int_0^s h(y)dy,
\end{aligned} \tag{1.0.18}$$

with $h(y)$ satisfying the Painlevé IV equation:

$$h''' + 6h'^2 - 4(y^2 + 2(q-2p))h' + 4yh - 8(q-p)p = 0. \tag{1.0.19}$$

[3]with $\hat{c} = \tilde{c}2^{\frac{p(p-q)}{2}}$. Remember $\tilde{c}$ from footnote 2.

The proof of Theorem 1.3 will be given in sections 3 and 6. In section 2 we show that the logarithmic derivative of a general multiple integral, involving the square of a Vandermonde, satisfies a third order differential equation, from which we derive, in section 3, that the logarithmic derivative of the matrix integral in (1.0.18) satisfies the Painlevé equation. Section 6 contains a discussion on how the Painlevé IV equation can be obtained from Painlevé V, by means of the rescaling (1.0.17), for large x. The following result shows that a certain general integral satisfies Painlevé V:

Theorem 1.5 *Given the weight*

$$\rho(z) = (z-a)^{\alpha}(b-z)^{\beta}e^{\gamma z}, \quad \textit{with } \alpha, \beta > -1,$$

on the interval $[a,b]$*, the integral*

$$g(x) := \frac{\partial}{\partial x} \log \int_{[a,b]^n} \Delta^2(z) \prod_1^n e^{xz_k}\rho(z_k)dz_k \tag{1.0.20}$$

is a solution to the Painlevé V equation. To be precise,

$$f(y) = n(n+\alpha+\beta) - \frac{y}{b-a}\left(g\Big(\frac{y}{b-a} - \gamma\Big) - na\right) \tag{1.0.21}$$

satisfies a version of the Painlevé V equation,

$$\begin{aligned} f''^2 \quad &+ \frac{4}{P^2}\Big((Pf'^2 + Qf' + R)f' - (P'f'^2 + Q'f' + R')f \\ &+ \frac{1}{2}(P''f' + Q'')f^2 - \frac{1}{6}P'''f^3 - \frac{1}{4}\beta^2 n^2\Big) = 0; \end{aligned} \tag{1.0.22}$$

with

$$\begin{aligned} P(y) &= y \\ 4Q(y) &= -y^2 + 2y(2n+\alpha-\beta) - (\alpha+\beta)^2 \\ 2R(y) &= -\beta n(\alpha+\beta+y). \end{aligned}$$

Note that (1.0.22) is a well known form of Painlevé V, as discussed in the Appendix. Theorem 1.5 and also Theorem 3.1 below (which is exactly the last equality in (1.0.18), can be derived form the work of Forrester and Witte [12, 13].

2 A differential equation for a matrix integral

Proposition 2.1 *Consider a weight $\rho(z)$ on an interval $E \subseteq \mathbb{R}$, with rational logarithmic derivative of the form*

$$-\frac{\rho'(z)}{\rho(z)} = \frac{b_0 + b_1 z + b_2 z^2}{a_0 + a_1 z + a_2 z^2} =: \frac{B(z)}{A(z)}, \tag{2.0.1}$$

and boundary condition

$$A(z)\rho(z)z^k\Big|_{\partial E} = 0, \quad \textit{for all } k = 0, 1, 2, \ldots. \tag{2.0.2}$$

The expression[4]

$$g(x) := \frac{\partial}{\partial x} \log \int_{E^n} \Delta^2(z) \prod_1^n e^{xz_k} \rho(z_k) dz_k \tag{2.0.3}$$

satisfies a third order differential equation, equivalent to Painlevé V, namely:

$$g''' + 6g'^2 + \frac{4a_2(g'' + 2gg')}{a_2 x - b_2} + \frac{2a_2^2 g^2 + P_2 g'}{(a_2 x - b_2)^2} + \frac{P_1 g - nQ_1}{(a_2 x - b_2)^3} = 0, \tag{2.0.4}$$

which can be transformed into the Painlevé V equation. In (2.0.4), the P_i's and Q_1 are polynomials in x,

$$P_2(x) := \left\{ \begin{array}{l} (4a_0a_2 - a_1^2)x^2 + 2(2a_1a_2 n - 2a_0b_2 + a_1b_1 - 2a_2b_0)x \\ -4a_2^2n^2 - 4(2a_1b_2 - a_2b_1)n + 4b_0b_2 - b_1^2 + 2a_2^2 \end{array} \right\}$$

$$P_1(x) := \left\{ \begin{array}{l} (2a_1a_2^2 n + 2a_0a_2b_2 - a_1^2b_2 + a_1a_2b_1 - 2a_2^2b_0)x \\ -4a_2^3n^2 + (-6a_1a_2b_2 + 4a_2^2b_1)n \\ -2a_0b_2^2 + a_1b_1b_2 + 2a_2b_0b_2 - a_2b_1^2 \end{array} \right\}$$

$$Q_1(x) := \left\{ \begin{array}{l} (2a_0a_2^2 n - a_1^2a_2 n + a_0a_1b_2 - 2a_0a_2b_1 + a_1a_2b_0)x \\ +2a_1a_2^2n^2 + (2a_1^2b_2 - a_1a_2b_1 - 2a_2^2b_0)n \\ +a_0b_1b_2 - 2a_1b_0b_2 + a_2b_0b_1 \end{array} \right\}.$$

The proof of Proposition 2.1 hinges on the following Lemma, which we state in its full generality, although only the case $I = E$ will be used; see [1, 23].

[4] When E is a finite interval, the integral always converges, and for a infinite interval, one may have to require $x > \alpha$ or $x < \alpha$, for some $\alpha \in \mathbb{R}$.

Lemma 2.2 *Given a disjoint union of intervals* $I=\bigcup_1^r[c_{2i-1},c_{2i}]\subset E$*, the integral*

$$\tau_n(t,c)=\int_{I^n}\Delta^2(z)\prod_1^n e^{\sum_{i=1}^\infty t_i z_k^i}\rho(z_k)dz_k$$

with $\rho(z)$ *and* E *as in (2.0.1) and (2.0.2) satisfies*

(i) *Virasoro constraints for all* $m\geq -1$*:*

$$\left(-\sum_1^{2r} c_i^{m+1}A(c_i)\frac{\partial}{\partial c_i}+\mathbb{V}_m\right)\tau_n(t,c)=0. \tag{2.0.5}$$

with

$$\mathbb{V}_m:=\sum_{k=0}^{2}\left\{\begin{array}{c} a_k(J^{(2)}_{k+m}+2n\ J^{(1)}_{k+m}+n^2\ J^{(0)}_{k+m}) \\ \\ -b_k(J^{(1)}_{k+m+1}+n\ \delta\ J^{(0)}_{k+m+1}) \end{array}\right\},$$

where

$$\begin{aligned} J_k^{(2)} &= \sum_{i+j=k}\frac{\partial^2}{\partial t_i\partial t_j}+\sum_{-i+j=k} it_i\frac{\partial}{\partial t_j}+\frac{1}{4}\sum_{-i-j=k} it_i\ jt_j \\ J_k^{(1)} &= \frac{\partial}{\partial t_k}+\frac{1}{2}(-k)t_k, \quad J_k^{(0)}=\delta_{k0}. \end{aligned}$$

(ii) *The KP-hierarchy*[5] *(*$k=0,1,2,\ldots$*)*

$$\left(\mathbf{s}_{k+4}\Big(\frac{\partial}{\partial t_1},\frac{1}{2}\frac{\partial}{\partial t_2},\frac{1}{3}\frac{\partial}{\partial t_3},\ldots\Big)-\frac{1}{2}\frac{\partial^2}{\partial t_1\partial t_{k+3}}\right)\tau_n\circ\tau_n=0,$$

of which the first equation reads:

$$\left(\left(\frac{\partial}{\partial t_1}\right)^4+3\left(\frac{\partial}{\partial t_2}\right)^2-4\frac{\partial^2}{\partial t_1\partial t_3}\right)\log\tau_n+6\left(\frac{\partial^2}{\partial t_1^2}\log\tau_n\right)^2=0. \tag{2.0.6}$$

[5] Given a polynomial $p(t_1,t_2,...)$, define the customary Hirota symbol $p(\partial_t)f\circ g:=p(\frac{\partial}{\partial y_1},\frac{\partial}{\partial y_2},...)f(t+y)g(t-y)\Big|_{y=0}$. The $\mathbf{s}_\ell$'s are the elementary Schur polynomials $e^{\sum_1^\infty t_iz^i}:=\sum_{i\geq 0}\mathbf{s}_i(t)z^i$ and for later use, set $\mathbf{s}_\ell(\tilde\partial):=\mathbf{s}_\ell(\frac{\partial}{\partial t_1},\frac{1}{2}\frac{\partial}{\partial t_2},\ldots)$.

Proof: The most transparent way to prove this lemma is via vector vertex operators, for which the β-integrals

$$\tau_n(t,c;\beta) := \int_{E^n} |\Delta_n(x)|^{2\beta} \prod_{k=1}^{n} \left(e^{\sum_1^\infty t_i x_k^i} \rho(x_k) dx_k \right), \quad \text{for } n>0 \tag{2.0.7}$$

are fixed points (see [1]). Another method, more computational, but much less conceptual, is to use a self-similarity argument, as in [1]. Namely, setting

$$d\tau_n(z) := |\Delta_n(z)|^{2\beta} \prod_{k=1}^{n} \left(e^{\sum_1^\infty t_i z_k^i} \rho(z_k) dz_k \right),$$

we have the following variational formula:

$$\frac{d}{d\varepsilon} d\tau_n(z_i \mapsto z_i + \varepsilon A(z_i) z_i^{k+1}) \Big|_{\varepsilon=0} = \sum_{\ell=0}^{\infty} \left(a_\ell \ {}^\beta\mathbb{J}^{(2)}_{k+\ell,n} - b_\ell \ {}^\beta\mathbb{J}^{(1)}_{k+\ell+1,n} \right) d\tau_n, \tag{2.0.8}$$

with

$$\begin{aligned} {}^\beta\mathbb{J}^{(2)}_{k,n}(t,n) &= \beta \ {}^\beta J^{(2)}_k + \Big(2n\beta + (k+1)(1-\beta) \Big) \ {}^\beta J^{(1)}_k + n\Big((n-1)\beta + 1 \Big) \delta_{k0}, \\ {}^\beta\mathbb{J}^{(1)}_{k,n}(t,n) &= {}^\beta J^{(1)}_k + n\delta_{k0}, \end{aligned}$$

where

$$\begin{aligned} {}^\beta J^{(1)}_k &= \frac{\partial}{\partial t_k} + \frac{1}{2\beta}(-k)t_{-k} \\ {}^\beta J^{(2)}_k &= \sum_{i+j=k} \frac{\partial^2}{\partial t_i \partial t_j} + \frac{1}{\beta} \sum_{-i+j=k} i t_i \frac{\partial}{\partial t_j} + \frac{1}{4\beta^2} \sum_{-i-j=k} i t_i j t_j. \end{aligned} \tag{2.0.9}$$

The change of integration variable $z_i \mapsto z_i + \varepsilon A(z_i) z_i^{k+1}$ in the integral (2.0.7) leaves the integral invariant, but it induces a change of limits of integration, given by the inverse of the map above; namely the c_i's in $I = \bigcup_1^r [c_{2i-1}, c_{2i}]$, get mapped as follows

$$c_i \mapsto c_i - \varepsilon A(c_i) c_i^{k+1} + O(\varepsilon^2).$$

Therefore, setting

$$I^\varepsilon = \bigcup_1^r [c_{2i-1} - \varepsilon A(c_{2i-1}) c_{2i-1}^{k+1} + O(\varepsilon^2), c_{2i} - \varepsilon A(c_{2i}) c_{2i}^{k+1} + O(\varepsilon^2)],$$

we find, using (2.0.8) and the fundamental theorem of calculus,

$$\begin{aligned}
0 &= \frac{\partial}{\partial \varepsilon}\int_{(I^{\varepsilon})^{2n}} |\Delta_{2n}(z+\varepsilon A(z)z^{k+1})|^{2\beta}\prod_{i=1}^{n} e^{-V(z_i+\varepsilon A(z_i)z_i^{k+1},t)}d(z_i+\varepsilon A(z_i)z_i^{k+1})\\
&= \left(-\sum_{i=1}^{2r} c_i^{k+1}A(c_i)\frac{\partial}{\partial c_i}+\sum_{\ell=0}^{\infty}\left(a_\ell\ {}^{\beta}\mathbb{J}^{(2)}_{k+\ell,2n}-b_\ell\ {}^{\beta}\mathbb{J}^{(1)}_{k+\ell+1,2n}\right)\right)\tau_n(t,c,\beta).
\end{aligned}$$

For Lemma 2.2 one sets $\beta=1$; also, when $I=E$, the condition (2.0.2) implies that the boundary terms in the formula above are absent. For statement (ii), concerning the KP equation, we refer the reader to [23]. This ends the proof of Lemma 2.2. ■

Proof of Proposition 2.1: Setting $F(t):=F_n(t)=\log\tau_n(t)$, a few of the Virasoro constraints of Lemma 2.2, evaluated along the locus

$$\mathcal{L}:=\{t=(x,0,0,\dots)\},$$

read as follows:

$$\begin{aligned}
\frac{\mathbb{V}_{-1}\tau_n}{\tau_n}\Big|_{\mathcal{L}} &= a_0\left(\sum_{i\geq 2} it_i\frac{\partial F}{\partial t_{i-1}}+nt_1\right)+a_1\left(\sum_{i\geq 1} it_i\frac{\partial F}{\partial t_i}+n^2\right)\\
&\quad + a_2\left(\sum_{i\geq 1} it_i\frac{\partial F}{\partial t_{i+1}}+2n\,\frac{\partial F}{\partial t_1}\right)-\left(b_0 n+b_1\frac{\partial F}{\partial t_1}+b_2\frac{\partial F}{\partial t_2}\right)\Big|_{\mathcal{L}}\\
&= n(a_0x+a_1n-b_0)+(a_1x+2na_2-b_1)\frac{\partial F}{\partial t_1}+(a_2x-b_2)\frac{\partial F}{\partial t_2}\Big|_{\mathcal{L}}=0.
\end{aligned}$$

$$\begin{aligned}
\frac{\mathbb{V}_{0}\tau_n}{\tau_n}\Big|_{\mathcal{L}} &= a_0\left(\sum_{i\geq 1} it_i\frac{\partial F}{\partial t_i}+n^2\right)+a_1\left(\sum_{i\geq 1} it_i\frac{\partial F}{\partial t_{i+1}}+2n\frac{\partial F}{\partial t_1}\right)\\
&\quad + a_2\left(\sum_{i\geq 1} it_i\frac{\partial F}{\partial t_{i+2}}+\frac{\partial^2 F}{\partial t_1^2}+\left(\frac{\partial F}{\partial t_1}\right)^2+2n\frac{\partial F}{\partial t_2}\right)\\
&\quad -\left(b_0\frac{\partial F}{\partial t_1}+b_1\frac{\partial F}{\partial t_2}+b_2\frac{\partial F}{\partial t_3}\right)\Big|_{\mathcal{L}}
\end{aligned}$$

$$
\begin{aligned}
&= a_0 n^2 + (a_0 x + 2na_1 - b_0)\frac{\partial F}{\partial t_1} + a_2\left(\frac{\partial F}{\partial t_1}\right)^2 + a_2\frac{\partial^2 F}{\partial t_1^2} \\
&\quad + (a_1 x + 2na_2 - b_1)\frac{\partial F}{\partial t_2} + (a_2 x - b_2)\frac{\partial F}{\partial t_3}\Big|_{\mathcal{L}} = 0.
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial}{\partial t_1}\frac{\mathbb{V}_{-1}\tau_n}{\tau_n}\Big|_{\mathcal{L}} &= a_0 n + \sum_1^2 a_i \frac{\partial F}{\partial t_i} + (a_1 x + 2na_2 - b_1)\frac{\partial^2 F}{\partial t_1^2} \\
&\quad + (a_2 x - b_2)\frac{\partial^2 F}{\partial t_1 \partial t_2}\Big|_{\mathcal{L}} = 0.
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial}{\partial t_1}\frac{\mathbb{V}_0\tau_n}{\tau_n}\Big|_{\mathcal{L}} &= \sum_{i=0}^2 a_i \frac{\partial F}{\partial t_{i+1}} + a_2\left(\frac{\partial^3 F}{\partial t_1^3} + 2\frac{\partial F}{\partial t_1}\frac{\partial^2 F}{\partial t_1^2}\right) \\
&\quad + (a_0 x + 2na_1 - b_0)\frac{\partial^2 F}{\partial t_1^2} + (a_1 x + 2n - b_1)\frac{\partial^2 F}{\partial t_1 \partial t_2} \\
&\quad + (a_2 x - b_2)\frac{\partial^2 F}{\partial t_1 \partial t_3}\Big|_{\mathcal{L}} = 0.
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial}{\partial t_2}\frac{\mathbb{V}_{-1}\tau_n}{\tau_n}\Big|_{\mathcal{L}} &= 2\sum_{i=0}^2 a_i \frac{\partial F}{\partial t_{i+1}} + (a_1 x + 2na_2 - b_1)\frac{\partial^2 F}{\partial t_1 \partial t_2} \\
&\quad + (a_2 x - b_2)\frac{\partial^2 F}{\partial t_2^2}\Big|_{\mathcal{L}} = 0.
\end{aligned}
$$

These five equations form a linear system in the five unknowns

$$
\frac{\partial F}{\partial t_2}\Big|_{\mathcal{L}}, \frac{\partial^2 F}{\partial t_1 \partial t_2}\Big|_{\mathcal{L}}, \frac{\partial^2 F}{\partial t_2^2}\Big|_{\mathcal{L}}, \frac{\partial F}{\partial t_3}\Big|_{\mathcal{L}}, \frac{\partial^2 F}{\partial t_1 \partial t_3}\Big|_{\mathcal{L}},
$$

which upon solving in terms of $\left(\frac{\partial}{\partial t_1}\right)^k F\Big|_{\mathcal{L}}$ and substituting into the KP-equation (remembering equation (2.0.6))

$$
\left(\frac{\partial^4}{\partial t_1^4} + 3\frac{\partial^2}{\partial t_2^2} - 4\frac{\partial^2}{\partial t_1 \partial t_3}\right)F + 6\left(\frac{\partial^2 F}{\partial t_1^2}\right)^2 = 0
$$

yields a differential equation in F. Pure F never appears in this equation, because the Virasoro constraints only contain partials of F. Therefore, it is a differential equation in $g(x) = \frac{\partial}{\partial t_1} F(t_1, 0, 0, \ldots)\Big|_{t_1=x}$, which one computes has the form (2.0.4); this ends the proof of Proposition 2.1. ■

The proof of Theorem 1.5 will be given at the end of section 3.

3 Hermitian matrix integrals and Painlevé equations

Define

$$\mathcal{H}_n = \{n \times n \text{ Hermitian matrices}\},$$

and, for an interval $I \subset \mathbb{R}$,

$$\mathcal{H}_n(I) = \{M \in \mathcal{H}_n \text{ with spectrum in } I\}.$$

We give a proof of the following theorem due to Forrester and Witte [12, 13], using Proposition 2.1.

Theorem 3.1 *For*[6] $a, b > -1$, *the logarithmic derivatives of the integrals*

$$\begin{aligned} h(s) &:= \frac{d}{ds} \log \int_{\mathcal{H}_n(-\infty,s]} \det(M - sI)^a e^{-\operatorname{Tr} M^2} dM \\ k(s) &:= s\frac{d}{ds} \log \int_{\mathcal{H}_n[0,s]} \det(sI - M)^b \det M^a e^{-\operatorname{Tr} M} dM \end{aligned} \tag{3.0.1}$$

satisfy the Painlevé IV and V equations, respectively:

$$h''' + 6h'^2 - 4(s^2 + 2(a-n))h' + 4sh - 8an = 0, \tag{3.0.2}$$

and

$$\begin{aligned} k''' + \frac{k''}{s} + \frac{6}{s}k'^2 - \frac{4}{s^2}kk' - \left(s^2 - 2s(2n+a-b) + (a+b)^2\right)\frac{k'}{s^2} \\ + (s - 2n - a + b)\frac{k}{s^2} - \frac{bn}{s^2}(s + a + b) = 0. \end{aligned} \tag{3.0.3}$$

[6] For the first integral of (3.0.1), one can choose the intervals $I_1 = (-\infty, s], [s, \infty)$ or $(-\infty, \infty)$. For the second integral of (3.0.1), one may choose the intervals $I_2 = [0, s], [s, \infty)$ or $[0, \infty)$

Proof of Theorem 3.1: Set

$$\tau(x) = \int_{I^n} \Delta^2(z) \prod_1^n e^{xz_k} \rho(z_k) dz_k, \tag{3.0.4}$$

with ρ and I as in (2.0.1) and (2.0.2).

(i) Then, expressed in spectral coordinates and making the substitution $y_i = z_i - s$ for $1 \leq i \leq n$, the first matrix integral in (3.0.1) reads, for $I = (-\infty, s], [s, \infty)$ and $(-\infty, \infty)$,

$$\begin{aligned} \int_{(I_1)^n} \Delta^2(z) \prod_1^n (z_i - s)^a e^{-z_i^2} dz_i &= e^{-ns^2} \int_{(I_1')^n} \Delta^2(y) \prod_1^n y_i^a e^{-y_i^2 - 2sy_i} dy_i \\ &= e^{-ns^2} \tau(-2s) \end{aligned} \tag{3.0.5}$$

with $I_1' = (-\infty, 0], [0, \infty), (-\infty, \infty)$. Here $\tau(x)$ as in (3.0.4) contains $\rho(z) = z^a e^{-z^2}$, for which

$$-\frac{\rho'}{\rho} = \frac{-a + 2z^2}{z}.$$

Thus, setting

$$\begin{aligned} &a_0 = a_2 = 0, \quad a_1 = 1 \\ &b_0 = -a, \quad b_1 = 0, \quad b_2 = 2. \end{aligned} \tag{3.0.6}$$

in the equation (2.0.4), one deduces that

$$g(x) := \frac{\partial}{\partial x} \log \tau(x), \quad \text{for } I = \begin{cases} (-\infty, 0] \\ (-\infty, \infty) \\ (0, \infty) \end{cases},$$

satisfies

$$g''' + 6g'^2 - g' \left(\frac{x^2}{4} + 4n + 2a \right) + \frac{xg}{4} + \frac{n}{2}(n + a) = 0.$$

But we need a differential equation for $e^{-ns^2}\tau(-2s)$, instead of $\tau(x)$. Therefore consider

$$h(s) = \frac{\partial}{\partial s} \log(e^{-ns^2} \tau(-2s)) = -2ns - 2g(x)\Big|_{x=-2s},$$

which relates to $g(x)$ as follows,

$$\begin{aligned} h'(s) &= -2n + 4g'(x)\Big|_{x=-2s} \\ h''(s) &= -8g''(x)\Big|_{x=-2s} \\ h'''(s) &= 16g'''(x)\Big|_{x=-2s}. \end{aligned}$$

Expressing $g(x), g'(x), g''(x), g'''(x)\Big|_{x=-2s}$ in terms of h, h', h'' and h''', and setting $x = -2s$ yield the differential equation (3.0.2), which according to the table in Appendix 1 is a version of Painlevé IV; this establishes the first part of Theorem 3.1.

(ii) Then, making the substitution $y_i = z_i/s$ for $1 \le i \le n$,

$$\begin{aligned} &\int_{(I_2)^n} \Delta^2(z) \prod_1^n (s - z_i)^b z_i^a e^{-z_i} dz_i \\ &= s^{n(n+a+b)} \int_{(I_2')^n} \Delta^2(y) \prod_1^n (1 - y_i)^b y_i^a e^{-sy_i} dy_i \\ &= s^{n(n+a+b)} \tau(-s) \end{aligned} \tag{3.0.7}$$

with $I_2 = [0, s]$, $I_2' = [0, 1]$, for $I_2 = [s, \infty]$, $I_2' = [1, \infty]$ and finally for $I_2 = [0, \infty]$, $I_2' = [0, \infty]$; $\tau(x)$ now corresponds to $\rho(z) = z^a(1 - z)^b$, and so

$$-\frac{\rho'}{\rho} = -\frac{a}{z} + \frac{b}{1 - z} = \frac{a - (a + b)z}{-z + z^2}.$$

Thus

$$\begin{aligned} &a_0 = 0, \quad a_1 = -1, \qquad a_2 = 1 \\ &b_0 = a, \quad b_1 = -(a + b), \quad b_2 = 0. \end{aligned}$$

Setting these special values in the equation (2.0.4), one checks that

$$g(x) := \frac{\partial}{\partial x} \log \int_{I_2'^n} \Delta^2(z) \prod_1^n e^{xz_k} z_k^a (1 - z_k)^b dz_k,$$

satisfies

$$\begin{aligned} & g''' + 6g'^2 + \frac{4}{x}(g'' + 2gg') + 2\left(\frac{g}{x}\right)^2 \\ - & \frac{g'}{x^2}\left(x^2 + 2(2n+a-b)x + (2n+a+b)^2 - 2\right) \\ - & \frac{g}{x^3}\left((2n+a-b)x + (2n+a+b)^2\right) \\ + & \frac{n}{x^3}(n+a)(x+2n+a+b) = 0. \end{aligned} \tag{3.0.8}$$

Since $g(x) = \dfrac{\partial}{\partial x}\log\tau(x)$, we have

$$\begin{aligned} k(s) &= s\,\frac{\partial}{\partial s}\log s^{n(n+a+b)}\tau(-s) \\ &= n(n+a+b) - sg(-s) \end{aligned}$$

and so

$$\begin{aligned} g(-s) &= \frac{1}{s}\left(-k(s) + n(n+a+b)\right) \\ g'(-s) &= \frac{1}{s^2}\left(sk'(s) - k(s) + n(n+a+b)\right) \\ g''(-s) &= \frac{2}{s^3}\left(-s^2k''(s)/2 + sk'(s) - k(s) + n(n+a+b)\right) \\ g'''(-s) &= \frac{3}{s^4}\left(s^3k'''(s)/3 - s^2k''(s) + 2sk'(s) - 2k(s) + 2n(n+a+b)\right). \end{aligned}$$

Substituting these expressions into (3.0.8) yields the differential equation (3.0.3), which again referring to the table in appendix 1 is Painlevé V, ending the proof of Theorem 3.1. ■

Proof of Theorem 1.5: Because of the form (2.0.1) of $\rho(z)dz$, we have for an open set of real constants a_i and b_i,

$$\rho(z) = (z-a)^\alpha(b-z)^\beta e^{\gamma z}, \text{ with } \alpha,\beta > -1 \text{ and } a,b \in \mathbb{R},$$

which, upon making a linear change of variables $z \mapsto y$ in the integral (2.0.3), leads to

$$g(x) = na + \frac{\partial}{\partial x} \log \int_{[0,1]^n} \Delta^2(y) \prod_1^n e^{(x+\gamma)(b-a)y_k} y_k^{\alpha} (1-y_k)^{\beta} dy_k$$

and so

$$\frac{1}{b-a} \left(g\Big(\frac{x'}{b-a} - \gamma \Big) - na \right) = \frac{\partial}{\partial x'} \log \int_{[0,1]^n} \Delta^2(y) \prod_1^n e^{x' y_k} y_k^{\alpha} (1-y_k)^{\beta} dy_k,$$

which is shown to be a solution of Painlevé V in part **(ii)** of the proof of Theorem 3.1 in this section.

To compute the constant in (1.0.22) (that is δ in equation (7.0.2)), multiply the equation (1.0.22) with $P^2/4$, set $y = 0$ and use the explicit expressions for the polynomials $P,\ Q,\ R$ and the function $f(y)$, as in (1.0.21), yielding the identity

$$\delta = -\left[Q f' + R \right] f' + \left[f'^2 + Q' f' + R' \right] f - \frac{1}{2} Q'' f^2 \Big|_{y=0}. \tag{3.0.9}$$

All the expressions can readily be computed, except for $f'(0)$, which requires some argument. One computes

$$\begin{aligned} f'(0) &= \frac{1}{b-a} \left(na - g(-\gamma) \right) \\ &= \frac{n}{b-a} \left(a - \frac{\int_{[a,b]^n} \Delta^2(z) (\frac{1}{n} \sum_1^n z_k)(z_k - a)^{\alpha} (b - z_k)^{\beta} dz_k}{\int_{[a,b]^n} \Delta^2(z) \prod_1^n (z_k - a)^{\alpha} (b - z_k)^{\beta} dz_k} \right) \\ &= \frac{n}{b-a} \left(a - \langle z_1 \rangle_{(a,b)} \right) \\ &= -\frac{n(n+\alpha)}{2n + \alpha + \beta}. \end{aligned}$$

To establish the last equality above, we need the Aomoto extension [5] (see [2], Appendix D) of Selberg's integral:

$$\langle x_1 ... x_m \rangle_{(0,1)} := \frac{\displaystyle\int_{[0,1]^n} x_1 \ldots x_m \, |\Delta(x)|^{2\gamma} \prod_{j=1}^{n} x_j^{\alpha}(1-x_j)^{\beta} dx_j}{\displaystyle\int_{[0,1]^n} |\Delta(x)|^{2\gamma} \prod_{j=1}^{n} x_j^{\alpha}(1-x_j)^{\beta} dx_j}$$

$$= \prod_{j=1}^{m} \frac{\alpha + 1 + (n-j)\gamma}{\alpha + \beta + 2 + (2n-j-1)\gamma}. \tag{3.0.10}$$

In particular, setting $\gamma = 1$, this formula implies

$$\langle z_1 \rangle_{(a,b)} = \langle a + (b-a)x_1 \rangle_{(a,b)} = a + (b-a)\langle x_1 \rangle_{(0,1)} = a + (b-a)\frac{n+\alpha}{2\,n+\beta+\alpha} \tag{3.0.11}$$

Adding up all the pieces in (3.0.9), one finds the value of the constant δ in equation (7.0.2), ending the proof of Theorem 1.5. ■

4 A limit theorem for the probability on words and Chi-square

The purpose of this section is to prove Theorem 1.2 and Corollary 1.3. Given m independent normal $N(0,1)$-random variables X_i, the sum

$$Z_m = \sum_{1}^{m} X_i^2$$

is χ_m^2-distributed, with Fourier transform

$$E(e^{t\,Z_m}) = (1-2t)^{-m/2}.$$

Developing both sides in t yields the moments

$$\frac{1}{k!}E\left(\left(\frac{1}{2}Z_m\right)^k\right) = \left(\begin{array}{c} \frac{m}{2} - 1 + k \\ k \end{array} \right).$$

Proof of Theorem 1.2: From Stirling formula $N! = \sqrt{2\pi N}\ e^{N(\log N - 1)}(1 + \mathbf{O}(\frac{1}{N}))$ and $(N+\alpha)! \simeq N! N^{\alpha}$, we have for large n,

$$\begin{aligned} (pn+k-p^2)! &\simeq (pn)!(pn)^{k-p^2} &&\simeq \frac{(n!\ p^n)^p}{(2\pi n)^{\frac{p-1}{2}}}\sqrt{p}\ (pn)^{k-p^2} \\ \prod_1^p (n-j)! &\simeq \prod_1^p (n!\ n^{-j}) &&\simeq (n!)^p\ n^{-\frac{p(p+1)}{2}} \end{aligned} \tag{4.0.1}$$

meaning here that the ratios between two consecutive expressions tend to 1, when $n \to \infty$. On the one hand, we have, changing the summation index $\ell \to \ell - p(n-p)$,

$$\begin{aligned} &\lim_{n\to\infty} \prod_{j=1}^p \frac{(n-j)!}{(q-j)!} \sum_{\ell \geq p(n-p)} \frac{p^{\ell}(nu)^{\ell - p(n-p)}}{\ell!} E^{\ell,p}\left(I_{\lambda\supseteq\mu}(\lambda) \prod_{n-q<j\leq n-p} h_{ij}^{\lambda}\right) \\ &= \lim_{n\to\infty} \prod_{j=1}^p \frac{(n-j)!}{(q-j)!} \sum_{k=0}^{\infty} \frac{p^{p(n-p)}(pnu)^k}{(k+p(n-p))!} E^{k+p(n-p),p}\left(I_{(\lambda\supseteq\mu)}(\lambda) \prod_{n-q<j\leq n-p} h_{ij}^{\lambda}\right) \\ &= \sum_{k=0}^{\infty} u^k \lim_{n\to\infty} \frac{(pn)^k p^{p(n-p)}}{\prod_1^p (q-j)!} \frac{\prod_1^p (n-j)!}{(pn+k-p^2)!} E^{k+p(n-p),p}\left(I_{\lambda\supseteq\mu}(\lambda) \prod_{n-q<j\leq n-p} h_{ij}^{\lambda}\right) \\ &= \sum_{k=0}^{\infty} u^k \lim_{n\to\infty} \frac{(2\pi)^{\frac{p-1}{2}} n^{\frac{p^2-1}{2}}}{\prod_1^p (q-j)! \sqrt{p}}\ E^{k+p(n-p),p}\left(I_{\lambda\supseteq\mu}(\lambda) \prod_{n-q<j\leq n-p} h_{ij}^{\lambda}\right), \end{aligned} \tag{4.0.2}$$

using (4.0.1) in the last equality.

On the other hand, using (1.0.11), (1.0.5), (1.0.13) and

$$\begin{aligned} \lim_{n\to\infty} \frac{(n)_\lambda}{n^{|\lambda|}} &= \lim_{n\to\infty} \left[\frac{n(n+1)\dots(n+\lambda_1-1)}{n^{\lambda_1}}\right] \left[\frac{(n-1)n\dots(n+\lambda_2-2)}{n^{\lambda_2}}\right] \dots, \\ &= 1 \end{aligned} \tag{4.0.3}$$

we have

$$\lim_{n\to\infty}\prod_{j=1}^{p}\frac{(n-j)!}{(q-j)!}\sum_{\ell\geq p(n-p)}\frac{p^{\ell}(nu)^{\ell-p(n-p)}}{\ell!}E^{\ell,p}\left(I_{\lambda\supseteq\mu}(\lambda)\prod_{n-q<j\leq n-p}h_{ij}^{\lambda}\right)$$

$$\begin{aligned}
&= \lim_{n\to\infty}\ {}_2F_1^{(1)}(p,q;n;x)\Big|_{\sum_\ell x_\ell^i = nu\,\delta_{i1}} \\
&= \lim_{n\to\infty}\sum_{k=0}^{\infty}u^k\sum_{\kappa\in\mathbb{Y}_k}\frac{n^k(p)_\kappa(q)_\kappa}{(h^\kappa)^2(n)_\kappa} \\
&= \sum_{k=0}^{\infty}u^k\sum_{\kappa\in\mathbb{Y}_k}\frac{(p)_\kappa(q)_\kappa}{(h^\kappa)^2} \\
&= \sum_{k=0}^{\infty}u^k\sum_{\kappa\in\mathbb{Y}_k}\mathbf{s}_\kappa(1^p)\mathbf{s}_\kappa(1^q), \quad \text{using (1.0.7)} \\
&= \sum_{\kappa\in\mathbb{Y}}\mathbf{s}_\kappa(u^p)\mathbf{s}_\kappa(1^q) \\
&= (1-u)^{-pq}, \ \text{using the Cauchy identity} \\
&= E\left(e^{\frac{1}{2}u\,Z_{2pq}}\right) \qquad (4.0.4)
\end{aligned}$$

Then comparing the coefficients of u^k in the identical asymptotic expressions (4.0.2) and (4.0.4) yields Theorem 1.2. ■

Proof of Corollary 1.3: Setting $p=q$ and $N=n-p$ in Theorem 1.2, the expectation in (1.0.16) becomes $P^{Np+k,p}(\lambda\supseteq\mu)$ for fixed μ. By RSK and (1.0.9), the condition $\lambda\supseteq\mu$ translates into $d_1(\pi)=\lambda_1^\top=p$ and $i_{p-1}(\pi)=\sum_1^p\lambda_i-\lambda_p=Np+k-\lambda_p$, with $\lambda_p\geq N$. This means $i_{p-1}(\pi)\leq k+(p-1)N$. ■

5 A limit theorem for the Poissonized probability on words

In the next Theorem, we consider the expectation for the Poissonized probability $P_{x,p}$ on Young diagrams $\lambda\in\mathbb{Y}_\ell^{(p)}$, as defined in (1.0.4). Before proving Theorem 1.4, one needs the following proposition:

Proposition 5.1 *([15, 22])For every continuous function g on $\mathbb{R}^p$, we have*

$$\lim_{\ell\to\infty} E^{\ell,p}\left(g\left(\frac{\lambda_1-\ell/p}{\sqrt{2\ell/p}},\ldots,\frac{\lambda_p-\ell/p}{\sqrt{2\ell/p}}\right)\right) \qquad |\lambda|=\ell$$

$$= \; p!\sqrt{\pi p} \int\limits_{\substack{x\in\mathbb{R}^p\\ x_1>\ldots>x_p\\ \sum_1^p x_i=0}} g(x_1,\ldots,x_p)\varphi_p(x_1,\ldots,x_p)dx_1\ldots dx_{p-1}$$

and

$$\lim_{x\to\infty} E_{x,p}\left(g\left(\frac{\lambda_1-x}{\sqrt{2x}},\ldots,\frac{\lambda_p-x}{\sqrt{2x}}\right)\right)$$

$$= \; p! \int\limits_{\substack{x\in\mathbb{R}^p\\ x_1>\ldots>x_p}} g(x_1,\ldots,x_p)\varphi_p(x_1,\ldots,x_p)dx_1\ldots dx_p \; ,$$

where $\varphi_p(x_1,\ldots,x_p)$ is the probability density

$$\varphi_p(x_1,\ldots,x_p)=\frac{1}{Z_p}\Delta_p(x)^2\prod_{j=1}^p e^{-x_j^2},$$

with

$$Z_p=(2\pi)^{p/2}\;\frac{2^{-p^2/2}}{\prod\limits_1^p j!}.$$

Proof of Theorem 1.4: Given partition λ with the probability measure (1.0.4), consider the random variable

$$\varepsilon_i(\lambda):=\frac{\lambda_i-x}{\sqrt{2x}}. \tag{5.0.1}$$

From (1.0.17), we have

$$s=\frac{(n-p)-x}{\sqrt{2x}}. \tag{5.0.2}$$

Consider the hook length $h^\lambda_{(ij)}$ for box $(i,j)\in\lambda$, such that $n-q<j\le n-p$. Define r_{ij} such that

$$\begin{aligned} h^\lambda_{(ij)} &:= \lambda_i-(n-p)+r_{ij}\\ &= \sqrt{2x}(\varepsilon_i-s)+r_{ij}, \end{aligned}$$

upon using formulas (5.0.1) and (5.0.2) in the last equality. The position (i,j) in the partition λ such that $n-q < j \le n-p$ implies that $r_{ij} \le (q-p)+p = q$, from visual inspection of Figure 1. We also have

$$\begin{aligned}
\{\lambda \in \mathbb{Y}^{(p)}, \lambda \supseteq \mu\} &= \{\lambda \in \mathbb{Y}^{(p)}, \lambda_i \ge n-p, \text{all } 1 \le i \le p\} \\
&= \left\{\lambda \in \mathbb{Y}^{(p)}, \frac{\lambda_i - x}{\sqrt{2x}} \ge \frac{(n-p)-x}{\sqrt{2x}}, \text{all } 1 \le i \le p\right\} \\
&= \bigcap_{i=1}^{p} \{\lambda \in \mathbb{Y}^{(p)}, \varepsilon_i(\lambda) \ge s\}.
\end{aligned}$$

On the set $\varepsilon_i(\lambda) \ge s$ and for $\lambda \in \mathbb{Y}^{(p)}$, we have, since $0 \le r_{ij} \le q$,

$$\begin{aligned}
&(2x)^{-\frac{p(q-p)}{2}} \prod_{\substack{(i,j)\in\lambda \\ n-q<j\le n-p}} h^{\lambda}_{(i,j)} - \prod_{1\le i\le p} (\varepsilon_i - s)^{q-p} \\
&= (2x)^{-\frac{p(q-p)}{2}} \prod_{\substack{(i,j)\in\lambda \\ n-q<j\le n-p}} (\sqrt{2x}(\varepsilon_i - s) + r_{ij}) - \prod_{1\le i\le p} (\varepsilon_i - s)^{q-p} \\
&= \prod_{\substack{1\le i\le p \\ n-q<j\le n-p}} \left((\varepsilon_i - s) + \frac{r_{ij}}{\sqrt{2x}}\right) - \prod_{1\le i\le p} (\varepsilon_i - s)^{q-p} \\
&\le \prod_{1\le i\le p} \left((\varepsilon_i - s) + \frac{q}{\sqrt{2x}}\right)^{q-p} - \prod_{1\le i\le p} (\varepsilon_i - s)^{q-p} \\
&= \sum_{\substack{0 \le \ell_i \le q-p \\ 1 \le i \le p \\ \sum \ell_i \ge 1}} \left(\frac{q}{\sqrt{2x}}\right)^{\sum \ell_i} \prod_{i=1}^{p} \binom{q-p}{\ell_i} \prod_{i=1}^{p} (\varepsilon_i - s)^{q-p-\ell_i}
\end{aligned}$$

The positive expression is now estimated as follows:

$$\begin{aligned}
&(2x)^{-\frac{p(q-p)}{2}} E^{\ell,p}\left(I_{\{\lambda\supseteq\mu\}}(\lambda) \prod_{\substack{(i,j)\in\lambda\\ n-q<j\leq n-p}} h^{\lambda}_{(i,j)} \right) \\
&\qquad -E^{\ell,p}\left(\prod_{1\leq i\leq p}(\varepsilon_i - s)^{q-p} I_{\varepsilon_i\geq s}\right) \\
\leq & \sum_{\substack{0\leq k_i\leq q-p\\ 1\leq i\leq p\\ \sum k_i\geq 1}} \left(\frac{q}{\sqrt{2x}}\right)^{\sum k_i} \prod_{i=1}^{p}\binom{q-p}{k_i} E^{\ell,p}\prod_{i=1}^{p}(\varepsilon_i - s)^{q-p-k_i} I_{\varepsilon_i\geq s}
\end{aligned}$$

Note that this estimate holds for all ℓ, with all expressions vanishing when $\ell < p(n-p)$. Using the previous estimate and using (5.0.1), one finds the estimate

$$\begin{aligned}
0 \ \leq\ & \frac{1}{(2x)^{\frac{p(q-p)}{2}}} E_{x,p}\left(I_{\lambda\supseteq\mu}(\lambda) \prod_{\substack{(i,j)\in\lambda\\ n-q<j\leq n-p}} h^{\lambda}_{(i,j)} \right) \\
&\qquad -e^{-px}\sum_{\ell\geq 0}\frac{(px)^{\ell}}{\ell!} E^{\ell,p}\left(\prod_{1\leq i\leq p}(\varepsilon_i - s)^{q-p} I_{\varepsilon_i\geq s}\right) \\
\leq\ & \sum_{\substack{0\leq k_i\leq q-p\\ 1\leq i\leq p\\ \sum k_i\geq 1}} \left(\frac{q}{\sqrt{2x}}\right)^{\sum k_i} \prod_{i=1}^{p}\binom{q-p}{k_i} \\
&\qquad E_{x,p}\left(\prod_{i=1}^{p}\left(\frac{\lambda_i - x}{\sqrt{2x}} - s\right)^{q-p-k_i} I_{\frac{\lambda_i - x}{\sqrt{2x}}\geq s}\right) \\
=\ & O\left(\frac{1}{\sqrt{2x}}\right) \quad \text{for } x\to\infty
\end{aligned}$$

because the expectation $E_{x,p}$ tends to an integral, by Johansson's theorem, applied to the function

$$g(x_1,\ldots,x_p) = \prod_{1}^{p}(x_i - s)^{q-p-k_i} I_{[s,\infty)}(x_i).$$

For $0 \leq \ell < p(n-p)$, the expectation is automatically zero. Then

$$\lim_{x\to\infty} \frac{1}{(2x)^{\frac{p(q-p)}{2}}} E_{x,p}\left(I_{\lambda\supseteq\mu}(\lambda) \prod_{\substack{(i,j)\in\lambda \\ n-q<j\leq n-p}} h^{\lambda}_{(i,j)} \right)$$

$$\begin{aligned}
&= \lim_{x\to\infty} e^{-px} \sum_{\ell\geq 0} \frac{(px)^{\ell}}{\ell!} E^{\ell,p}\left(\prod_{1\leq i\leq p} (\varepsilon_i - s)^{q-p} I_{\varepsilon_i \geq s} \right) \\
&= \lim_{x\to\infty} e^{-px} \sum_{\ell\geq 0} \frac{(px)^{\ell}}{\ell!} E^{\ell,p}\left(\prod_{1\leq i\leq p} \left(\frac{\lambda_i - x}{\sqrt{2x}} - s \right)^{q-p} I_{\frac{\lambda_i - x}{\sqrt{2x}} \geq s} \right) \\
&= \lim_{x\to\infty} E_{x,p}\left(\prod_{1\leq i\leq p} \left(\frac{\lambda_i - x}{\sqrt{2x}} - s \right)^{q-p} I_{\frac{\lambda_i - x}{\sqrt{2x}} \geq s} \right) \\
&= p! \int_{x_1>\ldots>x_p} \prod_{1\leq i\leq p} \left((x_i - s)^{q-p} I_{[s,\infty)}(x_i) \right) \frac{1}{Z_p} \Delta_p(x)^2 \prod_{j=1}^{p} e^{-x_j^2} dx_j,
\end{aligned}$$

by applying Proposition 5.1 to the continuous function

$$g(x_1, \ldots, x_p) = \prod_1^p (x_i - s)^{q-p} I_{[s,\infty)}(x_i).$$

This leads to the ratio of Hermitian matrix integrals in (1.0.18). The numerator is precisely the matrix integral (3.0.1), which according to Theorem 3.1 satisfies the Painlevé equation (3.0.2), which ends the proof of Theorem 1.4. ■

6 Painlevé IV as a limit of Painlevé V

According to Theorem 1.4, the limit (1.0.18) lead to a solution of Painlevé IV; this was established by identifying the limit as a matrix integral and then applying Theorem 3.1. In this section we give an alternative proof of this fact, by directly taking the scaling limit of the Painlevé V equation (1.0.14).

Remember from Proposition 1.1 and Theorem 1.4, we have

$$\prod_{i=1}^{p}\frac{(n-i)!}{(q-i)!}x^{-(n-p)p}e^{px}\ E_{x,p}\left(I_{\{\lambda\supseteq\mu\}}(\lambda)\prod_{\substack{(i,j)\in\lambda\\ n-q<j\leq n-p}}h^{\lambda}_{(i,j)}\right)$$
$$=\exp\int_0^x\frac{u(y)-p(n-p)+py}{y}dy \qquad (6.0.1)$$

and, with the rescaling

$$n-p=x+\sqrt{2x}s, \qquad (6.0.2)$$

we have from section 5, that

$$\lim_{x\to\infty}\frac{1}{(2x)^{\frac{p(q-p)}{2}}}E_{x,p}\left(I_{\lambda\supseteq\mu}(\lambda)\prod_{\substack{(i,j)\in\lambda\\ n-q<j\leq n-p}}h^{\lambda}_{ij}\right)=c\ \exp\int_0^s h(y)dy. \qquad (6.0.3)$$

for an appropriate choice of $h(y)$ and c. Taking into account the inverse of the rescaling (6.0.2), which for large n reads

$$x=n-s\sqrt{2n}+o(1) \qquad (6.0.4)$$

and taking the logarithmic derivatives of both equations (6.0.1) and (6.0.3), we have

$$\begin{aligned}\frac{\partial}{\partial s}\log E_{x,p}\left(I_{\{\lambda\supseteq\mu\}}(\lambda)\prod_{\substack{(i,j)\in\lambda\\ n-q<j\leq n-p}}h^{\lambda}_{(i,j)}\right) &= \frac{u(x)}{x}\frac{\partial x}{\partial s}\\ \frac{p(q-p)}{2x}\frac{\partial x}{\partial s}+\frac{\partial}{\partial s}\log E_{x,p}\left(I_{\{\lambda\supseteq\mu\}}(\lambda)\prod_{\substack{(i,j)\in\lambda\\ n-q<j\leq n-p}}h^{\lambda}_{(i,j)}\right) &\simeq h(s),\end{aligned}$$

for large, but fixed n.

Letting $n\to\infty$ in both equations and keeping the leading terms lead to

$$\begin{aligned}\frac{\partial}{\partial s}\log E_{x,p} &\simeq -\sqrt{\frac{2}{n}}u(n-s\sqrt{2n})\\ \frac{\partial}{\partial s}\log E_{x,p} &\simeq h(s).\end{aligned}$$

So, setting

$$h(s) := -\sqrt{\frac{2}{n}}u(n - s\sqrt{2n}),$$

we have

$$u(x) = -\sqrt{\frac{n}{2}}h\left(\frac{n-x}{\sqrt{2n}}\right)$$

and thus

$$u'(x) = \frac{1}{2}h' \ , \qquad u''(x) = -\frac{1}{2\sqrt{2n}}h'' \ , \qquad u'''(x) = -\frac{1}{4n}h'''.$$

Setting the rescaling (6.0.4) in the polynomials $Q(x)$ and $R(x)$, as defined in (1.0.15), we have

$$\begin{aligned} 4Q &= n\left(-2s^2 + 4(2p-q)\right) + o(n) \\ 4Q' &= 2\sqrt{2n}s + o(\sqrt{n}) \\ 2R &= 2np(p-q) + o(n). \end{aligned}$$

Substituting into (1.0.14), keeping the leading terms, which are of order n, and multiplying by 4, one finds the equation (1.0.19),

$$h''' + 6h'^2 - 4(y^2 + 2(q-2p))h' + 4yh - 8(q-p)p = 0. \qquad (6.0.5)$$

7 Appendix: Chazy classes

In his classification of differential equations

$f''' = F(z, f, f', f'')$, where F is rational in f, f', f'' and locally analytic in z,

subjected to the requirement that the general solution be free of movable branch points, Chazy found thirteen cases, the first of which is given by

$$f''' + \frac{P'}{P}f'' + \frac{6}{P}f'^2 - \frac{4P'}{P^2}ff' + \frac{P''}{P^2}f^2 + \frac{4Q}{P^2}f' - \frac{2Q'}{P^2}f + \frac{2R}{P^2} = 0 \qquad (7.0.1)$$

with arbitrary polynomials $P(z), Q(z), R(z)$ of maximal degree $3, 2, 1$ respectively. Cosgrove and Scoufis [10, 9], (A.3), show that this third order equation has a first integral, which is second order in f and quadratic in f'',

$$\begin{aligned} f''^2 \quad &+ \frac{4}{P^2}\Big((Pf'^2 + Qf' + R)f' - (P'f'^2 + Q'f' + R')f \\ &+ \frac{1}{2}(P''f' + Q'')f^2 - \frac{1}{6}P'''f^3 + \delta\Big) = 0; \qquad (7.0.2) \end{aligned}$$

δ is the integration constant.

P	$4Q$	$2R$	Painlevé eqt
1	$-4(u^2+2(a-n))$	$-8an$	P IV
u	$-u^2+2u(2n+a-b)-(a+b)^2$	$-bn(a+b+u)$	P V.

Equations of the general form

$$f''^2 = G(x, f, f')$$

are invariant under the map

$$x \mapsto \frac{a_1 z + a_2}{a_3 z + a_4} \quad \text{and} \quad f \mapsto \frac{a_5 f + a_6 z + a_7}{a_3 z + a_4}.$$

Using this map, the polynomial $P(z)$ can be normalized to

$$P(z) = z(z-1), \ z, \ \text{or } 1.$$

In this way, Cosgrove shows (7.0.2) is a master Painlevé equation, containing the 6 Painlevé equations. In the cases of PIV and PV, the canonical equations are respectively:

$$g''^2 = -4g'^3 + 4(zg'-g)^2 + A_1 g' + A_2 \qquad \textbf{(Painlevé IV)}$$

$$(zg'')^2 = (zg'-g)\Big(-4g'^2 + A_1(zg'-g) + A_2\Big) + A_3 g' + A_4. \qquad \textbf{(Painlevé V)}$$

References

[1] M. Adler and P. van Moerbeke: *Hermitian, symmetric and symplectic random ensembles: PDE's for the distribution of the spectrum*, Annals of Mathematics, **153**, 149–189, (2001) (with M. Adler). (preliminary versions in: solv-int/9903009 and math-ph/0009001)

[2] M. Adler and P. van Moerbeke: *Integrals over classical groups, random permutations, Toda and Toeplitz lattices*, Comm. Pure Appl. Math., **54**, 153–205, (2000) (arXiv: math.CO/9912143).

[3] M. Adler and P. van Moerbeke: *Integrals over Grassmannians and Random permutations*, Advances in Math, **181** (2003), 190-249 (arXiv: math.CO/0110281)

[4] D. Aldous and P. Diaconis: *Longest increasing subsequences: From patience sorting to the Baik-Deift-Johansson theorem*, Bull. Am. Math. Soc. (new series) **36** (4), 413–432 (1999).

[5] K. Aomoto: *Jacobi polynomials assocaited with Selberg integrals* , SIAM J. Math. Anal. **18** , 545–549 (1987).

[6] A. Borodin, G. Olshanski: *Z-Measures on partitions, Robinson-Schensted-Knuth correspondence, and $\beta = 2$ random matrix ensembles*, "Random matrices and their applications", Mathematical Sciences research Institute Publications #40, Cambridge University Press, (2001) (MathCO/9905189)

[7] A. Borodin, G. Olshanski: *Distributions on partitions, point processes, and the hypergeometric kernel* Comm. Math. Phys. **211**, 335–358 (2001). (math.RT/9904010)

[8] A. Borodin, A. Okounkov, G. Olshanski: *Asymptotics of Plancherel measures for symmetric groups* J. Amer. Math. Soc. (to appear) (math.CO/9905032)

[9] C. M. Cosgrove: *Chazy classes IX-XII of third-order differential equations*, Stud. Appl. Math. **104** ,3, 171–228 (2000).

[10] C. M. Cosgrove, G. Scoufis: *Painlevé classification of a class of differential equations of the second order and second degree* , Studies. Appl. Math. **88** , 25–87 (1993).

[11] E. Dueñez: *Random matrix ensembles associated to compact symmetric spaces*, Thesis, Princeton University 2001.

[12] P.J. Forrester and N. Witte: *Applications of the τ-functions of Painlevé equations to random matrices: P V, P III and the LUE, JUE and CUE*, Comm. Pure and Appl. Math. **LV**, 679–727 (2002).

[13] P.J. Forrester and N. Witte: *Application of the τ-function theory of Painlevé equations to random matrices: P VI, the JUE, CyUE, cJUE and scaled limits* ArXiv: math-ph/0204008

[14] C. Greene: *An extension of Schensted's theorem*, Adv. in Math. **14**, 254–265 (1974)

[15] K. Johansson: *Discrete orthogonal polynomial ensembles and the Plancherel measure*, arXiv:math.CO/9906120.

[16] J. Kaneko : *Selberg integrals and hypergeometric functions associated with Jack polynomials* SIAM J. Math. Anal. **24**, 1086–1110 (1993).

[17] I.G. MacDonald: "Symmetric functions and Hall polynomials", Clarendon Press, 1995.

[18] R.J. Muirhead: "Aspects of Multivariate Statistical Theory", Wiley series in probability and Mathematical statistics, 1982.

[19] B.E. Sagan: The Symmetry Group,Wadsworth & Brooks, Pacific Grove, California, 1991.

[20] R. P. Stanley: Enumerative Combinatorics , Cambridge Studies in Advanced Math. 49, Vols 1 and 2 (1997).

[21] R. P. Stanley: *Some combinatorial properties of Jack symmetric functions*, Adv. in Math., **77**, 76–115 (1989).

[22] C.A. Tracy and H. Widom: *On the distribution of the lengths of the longest monotone subsequences in random words*, math.CO/9904042 , (1999).

[23] P. van Moerbeke: *Integrable lattices: random matrices and random permutations*, "Random matrices and their applications", Mathematical Sciences research Institute Publications #40, Cambridge University Press, pp 321-406, (2001)