Asymptotic Enumeration of Dense 0-1 Matrices with Equal Row Sums and Equal Column Sums

E. Rodney Canfield*
Department of Computer Science
University of Georgia
Athens, GA 30602, USA
erc@cs.uga.edu

Brendan D. McKay†
Department of Computer Science
Australian National University
Canberra ACT 0200, Australia
bdm@cs.anu.edu.au

Submitted: Dec 22, 2004; Accepted: Jun 11, 2005; Published: Jun 19, 2005
Mathematics Subject Classifications: 05A16, 05C30, 62H17

Abstract

Let s, t, m, n be positive integers such that $sm = tn$. Let $B(m, s; n, t)$ be the number of $m \times n$ matrices over $\{0, 1\}$ with each row summing to s and each column summing to t. Equivalently, $B(m, s; n, t)$ is the number of semiregular bipartite graphs with m vertices of degree s and n vertices of degree t. Define the density $\lambda = s/n = t/m$. The asymptotic value of $B(m, s; n, t)$ has been much studied but the results are incomplete. McKay and Wang (2003) solved the sparse case $\lambda(1-\lambda) = o\big((mn)^{-1/2}\big)$ using combinatorial methods. In this paper, we use analytic methods to solve the problem for two additional ranges. In one range the matrix is relatively square and the density is not too close to 0 or 1. In the other range, the matrix is far from square and the density is arbitrary. Interestingly, the asymptotic value of $B(m, s; n, t)$ can be expressed by the same formula in all cases where it is known. Based on computation of the exact values for all $m, n \le 30$, we conjecture that the same formula holds whenever $m + n \to \infty$ regardless of the density.

1 Introduction

Let s, t, m, n be positive integers such that $sm = tn$. Let $B(m, s; n, t)$ be the number of $m \times n$ matrices over $\{0, 1\}$ with each row summing to s and each column summing to t. Equivalently, $B(m, s; n, t)$ is the number of semiregular bipartite graphs with m vertices of degree s and n vertices of degree t. The *density* $\lambda = s/n = t/m$ is the fraction of entries in the matrix which are 1.

*Research supported by the NSA Mathematical Sciences Program
†Research supported by the Australian Research Council

We are concerned in this paper with the asymptotic value of $B(m,s;n,t)$. Historically, the first significant result was that of Read [20], who obtained the asymptotic behavior for $s=t=3$. This was extended by Everett and Stein [8] to the case where s and t are arbitrary constants, not necessarily equal. The first result to allow s and t to increase was that of O'Neil [18], who permitted $s,t=O\bigl((\log n)^{1/4-\epsilon}\bigr)$. This was improved by Mineev and Pavlov [17] to permit $s=t\le\gamma(\log n)^{1/2}$ for fixed $\gamma<1$ and also for $1<s\le(t-1)^{-1}\gamma(\log n)^{1/4}$.

McKay [13] obtained $B(m,s;n,t)$ asymptotically whenever $s,t=o\bigl((sm)^{1/4}\bigr)$. This was improved by McKay and Wang [14] to the case $st=o\bigl((mn)^{1/2}\bigr)$.

All the prior work so far mentioned considers matrices for which the density is quite small. Obviously $B(m,n-s;n,m-t)=B(m,s;n,t)$ by complementation, so the very dense case is also handled. The intermediate range of densities, such as constant density, is considerably harder to deal with and until the present paper no exact asymptotics had been determined. Ordentlich and Roth [19] proved that, without any conditions except $ms=nt$,

$$B(m,s;n,t)\ge\binom{m}{t}^{n}\binom{n}{s}^{m}\bigl(\lambda^{\lambda}(1-\lambda)^{1-\lambda}\bigr)^{mn},$$

and that this bound is low by at most $\exp\bigl(O(n+\log m)\bigr)$ uniformly over λ if $\lambda(1-\lambda)m$ exceeds some absolute constant. More recently, Litsyn and Shpunt [11] determined an upper bound on $B(m,s;n,t)$ when $m=\Theta(n)$ and $\lambda=t/m=s/n$ is constant that, together with Ordentlich and Roth's lower bound, gives that

$$B(m,s;n,t)=\bigl(\lambda^{\lambda}(1-\lambda)^{1-\lambda}\bigr)^{-mn}\bigl(2\pi\lambda(1-\lambda)\bigr)^{-m/2-n/2}m^{-n/2}n^{-m/2}e^{O(n^{\epsilon})}$$

for any $\epsilon>0$.

Without giving more than a heuristic justification, Good and Crook [9] suggested the approximation

$$B(m,s;n,t)\approx\frac{\binom{n}{s}^{m}\binom{m}{t}^{n}}{\binom{mn}{\lambda mn}}.$$

We will see below that this is remarkably accurate, being within a constant of the correct value over a wide range and perhaps always.

In the present paper, we will focus on two quite different cases, using analytic methods inspired by [15]. In one case, the matrix is relatively square and the density is not too close to 0 or 1. (This includes the range considered by Litsyn and Shpunt.) In the other case, the matrix is much wider than high (or vice-versa) but the density is arbitrary. In both cases, we obtain precise asymptotics.

Remarkably, both the results we establish in this paper and the earlier results in the sparse case can be expressed using the same formula.

Theorem 1. *Consider a sequence of 4-tuples of positive integers m,s,n,t such that $ms=nt$ and $1\le t\le m-1$. Define $\lambda=s/n=t/m$ and $A=\frac{1}{2}\lambda(1-\lambda)$. Suppose that $\epsilon>0$*

is sufficiently small and that one of the following conditions holds (perhaps with m, n and s, t interchanged):
(a) $m, n \to \infty$ *and* $st = o\big((mn)^{1/2}\big)$;
(b) $m, n \to \infty$ *with* $n \le m = o(A^2 n^{1+\epsilon})$ *and, for some constant* $\gamma < \frac{3}{2}$,
$(1 - 2\lambda)^2 m \le \gamma A n \log n$;
(c) $n \to \infty$ *with* $2 \le m = O\Big(\big(t(m-t)n\big)^{1/4-\epsilon}\Big)$.
Then

$$B(m, s; n, t) = \frac{\binom{n}{s}^m \binom{m}{t}^n}{\binom{mn}{\lambda mn}} \Big(\frac{m-1}{m}\Big)^{(m-1)/2} \Big(\frac{n-1}{n}\Big)^{(n-1)/2} \exp\big(\tfrac{1}{2} + o(1)\big). \tag{1.1}$$

Proof. Part (a) was established by McKay and Wang [14]. Part (b) will be proved in Sections 2–4; specifically, it follows from (2.2) and Theorems 2 and 3. Part (c) follows from Theorem 4 in Section 5. ◻

Note that

$$\Big(\frac{N-1}{N}\Big)^{(N-1)/2} = \exp\big(-\tfrac{1}{2} + O(N^{-1})\big)$$

as $N \to \infty$, so one or both such terms in (1.1) can be simplified depending on which of m, n tend to ∞.

In Section 6 we show how $B(m, s; n, t)$ can be computed exactly for small m, n and show how the values for $m, n \le 30$ suggest the following conjecture.

Conjecture 1. *Consider a sequence of 4-tuples of positive integers m, s, n, t such that $ms = nt$. Then (1.1) holds uniformly over $1 \le t \le m - 1$ whenever $m + n \to \infty$.*

Calculations of the exact values for all $m, n \le 30$ show excellent agreement with Conjecture 1. There is less than 10% discrepancy between the exact value and the conjectured asymptotic value in all cases computed and less than 1% discrepancy whenever $m+n \ge 35$. More precisely, write the quantity indicated by "$o(1)$" in (1.1) as $\Delta(m, s; n, t)/(Amn)$. Our experiments, including the exact values mentioned above and many numerical estimates described in Section 6, suggest that $\Delta(m, s; n, t)$ always lies in the interval $(-\frac{1}{12}, 0)$. From [14], (see [10, Corollary 5.1]), we know that $\Delta(m, s; n, t) \to -\frac{1}{12}$ as $m, n \to \infty$ with $st = o\big((mn)^{1/5}\big)$. At the upper end, the greatest value we know is $\Delta(4, 2; 4, 2) \approx -0.0171$.

In a future paper we will allow the row sums, and similarly the column sums, to be unequal within limits. For the case of sparse matrices, the best result is by Greenhill, McKay and Wang [10]. We also plan to address the issue of matrices over $\{0, 1, 2, \ldots\}$ with equal row sums and equal column sums.

2 An integral for $B(m,s;n,t)$

Our proof of Theorem 1(b) occupies this section and the following two. We express $B(m,s;n,t)$ as an integral in $(m+n)$-dimensional complex space then estimate its value by the saddle-point method.

It is clear that $B = B(m,s;n,t)$ is the coefficient of $x_1^s \cdots x_m^s\, y_1^t \cdots y_n^t$ in

$$\prod_{j=1}^{m} \prod_{k=1}^{n} (1 + x_j y_k).$$

Applying Cauchy's Theorem we have

$$B = \frac{1}{(2\pi i)^{m+n}} \oint \cdots \oint \frac{\prod_{j,k}(1+x_j y_k)}{x_1^{s+1} \cdots x_m^{s+1} y_1^{t+1} \cdots y_n^{t+1}}\, dx_1 \cdots dx_m\, dy_1 \cdots dy_n, \tag{2.1}$$

where each contour circles the origin once in the anticlockwise direction.

It will suffice to take the contours to be circles; specifically, we will put $x_j = re^{i\theta_j}$ and $y_k = re^{i\phi_k}$ for each j,k, where

$$r = \sqrt{\frac{\lambda}{1-\lambda}}\,.$$

This gives

$$B = \frac{1}{(2\pi)^{m+n}\bigl(\lambda^\lambda(1-\lambda)^{1-\lambda}\bigr)^{mn}}\, I(m,n), \tag{2.2}$$

where

$$I(m,n) = \int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} \frac{\prod_{j,k}\bigl(1+\lambda(e^{i(\theta_j+\phi_k)}-1)\bigr)}{e^{is\sum_j \theta_j + it\sum_k \phi_k}}\, d\boldsymbol{\theta}\, d\boldsymbol{\phi}, \tag{2.3}$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$ and $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_n)$.

In equation (2.3) it is to be noted that the integrand is invariant under the two substitutions $\theta_j \leftarrow \theta_j + 2\pi$ and $\phi_k \leftarrow \phi_k + 2\pi$. In analyzing the magnitude of this integrand, it is often necessary to consider what might be called the "wrap-around" neighborhood of a point $\theta \in [-\pi, +\pi]$. This neighborhood consists of the union of two half-open intervals $[-\pi, -\pi+\delta)$ and $(\pi-\delta, \pi]$. To avoid numerous awkward expressions such as this, we find it convenient to think of θ_j and ϕ_k as points on the unit circle. To this end, we let C be the real numbers modulo 2π, which we can interpret as points on a circle in the usual fashion. Let z be the canonical mapping from C to the real interval $(-\pi, \pi]$; that is, if x lies on the unit circle, then $z(x)$ is its signed arc length from the point 1. An *open half-circle* is $C_t = (t-\pi/2, t+\pi/2) \subseteq C$ for some t. With this notion of half-circle, we may define an important subset of the Cartesian product C^N; namely, define $\hat{C}^N$ to be the subset of vectors $\boldsymbol{x} = (x_1, \ldots, x_N) \in C^N$ such that $x_1, \ldots, x_N$ all lie in a single open half-circle (where that open half-circle can depend on $\boldsymbol{x}$).

If $\boldsymbol{x} = (x_1, \ldots, x_N) \in C_0^N$ then define

$$\bar{\boldsymbol{x}} = z^{-1}\Big(\frac{1}{N}\sum_{j=1}^{N} z(x_j)\Big).$$

More generally, if $\boldsymbol{x} \in C_t^N$ then define $\bar{\boldsymbol{x}} = t + \overline{(x_1 - t, \ldots, x_N - t)}$. It is easy to see that the function $\boldsymbol{x} \mapsto \bar{\boldsymbol{x}}$ is well-defined and continuous for $\boldsymbol{x} \in \hat{C}^N$.

3 The principal part of the integral

To estimate the integral $I(m,n)$, we show that it is concentrated in a rather small region, then we expand the integrand inside that region.

For some sufficiently small $\epsilon > 0$, let $\mathcal{R}$ denote the set of vector pairs $\boldsymbol{\theta}, \boldsymbol{\phi} \in \hat{C}^m \times \hat{C}^n$ such that

$$\begin{aligned} |\bar{\boldsymbol{\theta}} + \bar{\boldsymbol{\phi}}| &\le (mn)^{-1/2+2\epsilon} \\ |\hat{\theta}_j| &\le n^{-1/2+\epsilon}, 1 \le j \le m \\ |\hat{\phi}_k| &\le m^{-1/2+\epsilon}, 1 \le k \le n, \end{aligned}$$

where $\hat{\theta}_j = \theta_j - \bar{\boldsymbol{\theta}}$ and $\hat{\phi}_k = \phi_k - \bar{\boldsymbol{\phi}}$. In this definition, values are considered in C.

Let $I_{\mathcal{R}}(m,n)$ denote the integral $I(m,n)$ restricted to the region $\mathcal{R}$. In the following section, we will show that $I(m,n) \sim I_{\mathcal{R}}(m,n)$. In the present section, we will estimate $I_{\mathcal{R}}(m,n)$.

Our calculations are guided by the similar problem solved in [15]. In particular, we will use the following result which can be proved from a special case of [15, Lemma 3]. Let $\mathrm{Im}(z)$ denote the imaginary part of z.

Lemma 1. *Let ϵ and ϵ' be such that $0 < \epsilon' < 2\epsilon < \frac{1}{12}$. Let $\hat{A} = \hat{A}(N)$ be a real-valued function such that $N^{-\epsilon'} \le \hat{A}(N) \le N^{\epsilon'}$ for sufficiently large N. Let $\hat{B} = \hat{B}(N)$, $\hat{C} = \hat{C}(N)$, $\hat{E} = \hat{E}(N)$, $\hat{F} = \hat{F}(N)$ be complex-valued functions such that the ratios $\hat{B}/\hat{A}, \hat{C}/\hat{A}, \hat{E}/\hat{A}, \hat{F}/\hat{A}$ are bounded. Suppose that, for some $\delta > 0$,*

$$f(\boldsymbol{z}) = \exp\big(-\hat{A}N\xi_2 + \hat{B}N\xi_3 + \hat{C}\xi_1\xi_2 + \hat{E}N\xi_4 + \hat{F}\xi_2^2 + O(N^{-\delta})\big)$$

is integrable for $\boldsymbol{z} = (z_1, z_2, \ldots, z_N) \in U_N$, where $\xi_t = \sum_{j=1}^{N} z_j^t$ for $t = 1, 2, 3, 4$ and

$$U_N = \big\{\boldsymbol{z} \;\big|\; |z_j| \le N^{-1/2+\epsilon} \text{ for } 1 \le j \le N\big\}.$$

Then, provided the $O(\,)$ term in the following converges to zero,

$$\int_{U_N} f(\boldsymbol{z})\, d\boldsymbol{z} = \Big(\frac{\pi}{\hat{A}N}\Big)^{N/2} \exp\Big(\frac{3\hat{E} + \hat{F}}{4\hat{A}^2} + \frac{15\hat{B}^2 + 6\hat{B}\hat{C} + \hat{C}^2}{16\hat{A}^3} + O\big((N^{-1/2+12\epsilon} + N^{-\delta})\hat{Z} + \hat{A}^{-1}N^{-\frac{1}{4}+3\epsilon}\big)\Big),$$

where

$$\hat{Z} = \exp\Bigl(\frac{15\,\mathrm{Im}(\hat{B})^2 + 6\,\mathrm{Im}(\hat{B})\,\mathrm{Im}(\hat{C}) + \mathrm{Im}(\hat{C})^2}{16\hat{A}^3}\Bigr).$$

Proof. Lemma 3 of [15] implies a result that is the same except that the condition $N^{-\epsilon'} \le \hat{A}(N) \le N^{\epsilon'}$ is replaced by the stronger condition $N^{-\epsilon'} \le \hat{A}(N) = O(1)$ and the condition $\epsilon < \frac{1}{24}$ is is replaced by the weaker condition $\epsilon < \frac{1}{12}$. Moreover, the error term is

$$O\bigl((N^{-1/2+6\epsilon} + N^{-\delta})\hat{Z} + N^{-1+12\epsilon} + \hat{A}^{-1}N^{-\Delta}\bigr)$$

for any Δ satisfying $0 < \Delta < \frac{1}{4} - \frac{1}{2}\epsilon$. Clearly this covers the case $N^{-\epsilon'} \le \hat{A}(N) \le 1$ of the present lemma, on taking $\Delta = \frac{1}{4} - \epsilon$.

For the remaining case, where $1 \le \hat{A}(N) < N^{\epsilon'}$, apply the transformation $z_j \mapsto N^{-\epsilon'/2}z_j$, then invoke Lemma 3 of [15] again, using $\Delta = \frac{1}{4} - \epsilon$ as before. □

In the following, we assume that $m, n \to \infty$. A word of explanation about the symbol ϵ as used in the paper is in order. It represents a definite positive constant. Whenever an assertion is made which the reader can confirm only by knowing the value of ϵ, s/he should note that the assertion is correct as long as ϵ is small enough. There being only finitely many statements in the paper, there is some positive value for ϵ small enough for all of them. In short, all equations and inequalities should be read with an understood "for m, n sufficiently large and ϵ sufficiently small".

The following lemma will be needed soon. We use the notation $\mathcal{R}^{\mathrm{c}}$ for the complement of a region $\mathcal{R}$. Recall that $A = \frac{1}{2}\lambda(1-\lambda)$.

Lemma 2. *Let $m, n \to \infty$ be integers, $x_1, \ldots, x_m$ variables, $M_2 = \sum_{j=1}^m x_j^2$, and $\mathcal{K}$ the region of m-space defined by*

$$\mathcal{K} = \Bigl\{\boldsymbol{x} \Bigm| \frac{m}{2An}(1 - m^{-1/4}) \le M_2 \le \frac{m}{2An}(1 + m^{-1/4})\Bigr\}.$$

Then,

$$\int_{\mathcal{K}^{\mathrm{c}}} \exp(-AnM_2)\,d\boldsymbol{x} = O(1)\Bigl(\frac{\pi}{An}\Bigr)^{m/2}\exp\bigl(-\tfrac{1}{5}m^{1/2}\bigr).$$

Proof. We'll be brief, because the idea is very much the same as found in the proof of Lemma 1, which can be consulted for details in [15]. Recalling the formula for the surface area of the ball of radius ρ in m-space, we have

$$\int_{M_2 \in [a,b]} \exp(-AnM_2) = \frac{2\pi^{m/2}}{\Gamma(m/2)} \int_{a^{1/2}}^{b^{1/2}} e^{-An\rho^2}\rho^{m-1}d\rho.$$

Case (i): $a = 0$, $b = (m/(2An))(1 - m^{-1/4})$. Using

$$e^{-An(b-x)^2}(b-x)^{m-1} \le e^{-Anb^2 - Anx^2}b^{m-1},\ 0 \le x \le b,$$

and Stirling's formula for the Gamma function,

$$\int_{M_2\in[0,b]} \exp(-AnM_2) = O(1)\left(\frac{\pi}{An}\right)^{m/2} \exp\bigl(-\tfrac{1}{5}m^{1/2}\bigr).$$

Case (ii): $a = (m/(2An))(1+m^{-1/4})$, $b=\infty$. Using

$$e^{-An(a+x)^2}(a+x)^{m-1} \le e^{-Ana^2-Anx^2}a^{m-1},\ x\ge 0,$$

we find the same bound for the integral over $M_2 \in [a,\infty)$ as in Case (i). Combining the two cases completes the proof of the Lemma. □

Let T_1 be the transformation which expresses the original $m+n$ variables θ_j, ϕ_k (see (2.3)) in terms of $\mu = \bar{\boldsymbol\theta} + \bar{\boldsymbol\phi}$, $\delta = \bar{\boldsymbol\theta} - \bar{\boldsymbol\phi}$, $\hat\theta_j$ $(1 \le j \le m-1)$, and $\hat\phi_k$ $(1 \le k \le n-1)$. Explicitly,

$$\theta_j = \tfrac{1}{2}(\mu+\delta) + \hat\theta_j,\ \phi_k = \tfrac{1}{2}(\mu-\delta) + \hat\phi_k,$$

where here and hereafter we use the abbreviations

$$\hat\theta_m = -\sum_{j=1}^{m-1}\hat\theta_j,\ \hat\phi_n = -\sum_{k=1}^{n-1}\hat\phi_k.$$

We have

$$I_{\mathcal{R}}(m,n) = 2\pi mn\, J(m,n),$$

where

$$J(m,n) = \int_{\mathcal{S}} G(\mu, \hat{\boldsymbol\theta}, \hat{\boldsymbol\phi})\, d\hat{\boldsymbol\theta}\, d\hat{\boldsymbol\phi}\, d\mu.$$

Here, the function G is the composition $F \circ T_1$, which is easily seen to be independent of the difference $\delta = \bar{\boldsymbol\theta} - \bar{\boldsymbol\phi}$. The region of integration $\mathcal{S} = T_1^{-1}(\mathcal{R})$ is defined by virtually the same inequalities as was $\mathcal{R}$ with these two notes: we now write the first inequality as $|\mu| \le (mn)^{-1/2+2\epsilon}$; and, second, neither $\hat\theta_m$ nor $\hat\phi_n$ is a variable of integration, but the definition of $\mathcal{S}$ includes the inequalities

$$\Bigl|\sum_{j=1}^{m-1}\hat\theta_j\Bigr| \le n^{-1/2+\epsilon},\quad \Bigl|\sum_{k=1}^{n-1}\hat\phi_k\Bigr| \le m^{-1/2+\epsilon}$$

arising from the $\mathcal{R}$-inequalities $|\hat\theta_m| \le n^{-1/2+\epsilon}$ and $|\hat\phi_n| \le m^{-1/2+\epsilon}$. The factor of $2\pi mn$ comes from the integration over δ (which has a range of 4π) and the Jacobian $mn/2$ of transformation T_1.

In this section we prove

Theorem 2. *Suppose $m, n \to \infty$ with $\lambda = \lambda(m,n)$, such that $m \geq n$ and*

$$m = o(A^2 n^{1+\epsilon}). \tag{3.1}$$

Suppose further that, for some constant $\gamma < \frac{3}{2} - \frac{45}{2}\epsilon - 6\epsilon^2$,

$$(1-2\lambda)^2 m \leq \gamma A n \log n. \tag{3.2}$$

Then,

$$J(m,n) = (mn)^{-1/2} \exp\Bigl\{-\frac{1}{2} - \frac{1-2A}{24A}\Bigl(\frac{m}{n} + \frac{n}{m}\Bigr) + O(D)\Bigr\}$$
$$\times \Bigl(\frac{\pi}{Amn}\Bigr)^{1/2} \Bigl(\frac{\pi}{An}\Bigr)^{(m-1)/2} \Bigl(\frac{\pi}{Am}\Bigr)^{(n-1)/2},$$

where

$$D = n^{-1/4+\gamma/24+4\epsilon+o(1)} + n^{-1/2+\gamma/3+15\epsilon/2+2\epsilon^2}.$$

Proof. The assumption $m \geq n$ has been made only to avoid frequent use of the expressions $\max(m,n)$ and $\min(m,n)$. Two easy consequences of (3.1) will be used without repeatedly citing that equation:

$$A^{-1} \leq A^{-1}\frac{m}{n} = o(An^\epsilon),\; m = o(An^{1+\epsilon}).$$

For future reference we establish:

$$\log n = o(An^\epsilon), \quad \log m = o(Am^\epsilon). \tag{3.3}$$

Indeed, for the first, $\log^2 n = o(A^{-1} \cdot An^\epsilon)$, and $A^{-1} = O(An^\epsilon)$. The second then follows since $\log m = O(\log n)$ and $m \geq n$. In particular, both Am^ϵ, An^ϵ become infinite.

For $|x|$ small, see [15],

$$1 + \lambda(e^{ix} - 1) = \exp\bigl(\lambda ix - Ax^2 - iA_3x^3 + A_4x^4 + O(A|x|^5)\bigr)$$

with

$$A = \tfrac{1}{2}\lambda(1-\lambda),\; A_3 = \tfrac{1}{6}\lambda(1-\lambda)(1-2\lambda),\; A_4 = \tfrac{1}{24}\lambda(1-\lambda)(1-6\lambda+6\lambda^2).$$

Uniformly in the region $\mathcal{S}$, where all $|\mu + \hat\theta_j + \hat\phi_k|$ are small,

$$G = \exp\Bigl\{-A\sum\nolimits_{j,k}(\mu + \hat\theta_j + \hat\phi_k)^2 - iA_3\sum\nolimits_{j,k}(\mu + \hat\theta_j + \hat\phi_k)^3$$
$$+ A_4\sum\nolimits_{j,k}(\mu + \hat\theta_j + \hat\phi_k)^4 + O\bigl(A\sum\nolimits_{j,k}|\mu + \hat\theta_j + \hat\phi_k|^5\bigr)\Bigr\}.$$

Here and below, the undelimited summation over j,k runs over $1 \leq j \leq m$, $1 \leq k \leq n$, and we continue to use the abbreviations $\hat\theta_m = -\sum_{j=1}^{m-1}\hat\theta_j$, $\hat\phi_n = -\sum_{k=1}^{n-1}\hat\phi_k$.

We now proceed to a second change of variables, $(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\phi}}) = T_2(\boldsymbol{\sigma}, \boldsymbol{\tau})$ given by

$$\hat{\theta}_j = \sigma_j + c\mu_1, \quad \hat{\phi}_k = \tau_k + d\nu_1,$$

where, for $1 \le h \le 4$, μ_h and ν_h denote the power sums $\sum_{j=1}^{m-1} \sigma_j^h$ and $\sum_{k=1}^{n-1} \tau_k^h$, respectively. The scalars c and d are chosen to eliminate the second-degree cross-terms $\sigma_{j_1}\sigma_{j_2}$ and $\tau_{k_1}\tau_{k_2}$, and thus diagonalize the quadratic in $\boldsymbol{\sigma}, \boldsymbol{\tau}$. Suitable choices for c, d are

$$c = -\frac{1}{m + m^{1/2}}, \quad d = -\frac{1}{n + n^{1/2}},$$

and we find the following:

$$\begin{aligned}
\textstyle\sum_{j,k} (\mu + \hat{\theta}_j + \hat{\phi}_k)^2 &= mn\mu^2 + n\mu_2 + m\nu_2 \\
\textstyle\sum_{j,k} (\mu + \hat{\theta}_j + \hat{\phi}_k)^3 &= mn\mu^3 + 3\mu(n\mu_2 + m\nu_2) + n(\mu_3 + 3c\mu_2\mu_1 - c_2\mu_1^3) \\
&\quad + m(\nu_3 + 3d\nu_2\nu_1 - d_2\nu_1^3) \\
\textstyle\sum_{j,k} (\mu + \hat{\theta}_j + \hat{\phi}_k)^4 &= mn\mu^4 + 6\mu_2\nu_2 + n(\mu_4 + 4c\mu_3\mu_1 + 6c^2\mu_2\mu_1^2 + c_3\mu_1^4) \\
&\quad + m(\nu_4 + 4d\nu_3\nu_1 + 6d^2\nu_2\nu_1^2 + d_3\nu_1^4) + 6\mu^2(n\mu_2 + m\nu_2) \\
&\quad + 4\mu\big(n(\mu_3 + 3c\mu_2\mu_1 - c_2\mu_1^3) + m(\nu_3 + 3d\nu_2\nu_1 - d_2\nu_1^3)\big) \\
\textstyle\sum_{j,k} |\mu + \hat{\theta}_j + \hat{\phi}_k|^5 &= O(mn^{-3/2+5\epsilon} + nm^{-3/2+5\epsilon}),
\end{aligned}$$

in which we have introduced the additional abbreviations

$$\begin{aligned}
c_2 &= \frac{1}{m^{1/2}(m^{1/2}+1)^2}, \qquad & c_3 &= \frac{m^{1/2}+3}{m(m^{1/2}+1)^3}, \\
d_2 &= \frac{1}{n^{1/2}(n^{1/2}+1)^2}, \qquad & d_3 &= \frac{n^{1/2}+3}{n(n^{1/2}+1)^3}.
\end{aligned}$$

The determinant of the matrix T_2 is $(mn)^{-1/2}$, and so

$$J(m, n) = (mn)^{-1/2} \int_{T_2^{-1}(\mathcal{S})} E_1,$$

where $E_1 = \exp(L_1)$, and

$$\begin{aligned}
L_1 = {} & \mu^4(A_4mn) + \mu^3(-iA_3mn) + \mu^2(-Amn + 6A_4n\mu_2 + 6A_4m\nu_2) \\
& + \mu\big(-3iA_3n\mu_2 - 3iA_3m\nu_2 + 4A_4n(\mu_3 + 3c\mu_2\mu_1 - c_2\mu_1^3) \\
& \qquad + 4A_4m(\nu_3 + 3d\nu_2\nu_1 - d_2\nu_1^3)\big) \\
& - An\mu_2 - Am\nu_2 + 6A_4\mu_2\nu_2 \\
& - iA_3n(\mu_3 + 3c\mu_2\mu_1 - c_2\mu_1^3) - iA_3m(\nu_3 + 3d\nu_2\nu_1 - d_2\nu_1^3) \\
& + A_4n(\mu_4 + 4c\mu_3\mu_1 + 6c^2\mu_2\mu_1^2 + c_3\mu_1^4) + A_4m(\nu_4 + 4d\nu_3\nu_1 + 6d^2\nu_2\nu_1^2 + d_3\nu_1^4) \\
& + O(Amn^{-3/2+5\epsilon} + Am^{-3/2+5\epsilon}n).
\end{aligned}$$

To complete the evaluation of the integral, we need to consider a number of different regions within the space of the variables μ, σ_j, τ_k, as well as a number of different integrands. Let us introduce all of these at the outset. Define $\rho_\sigma, \rho_\tau > 0$ by

$$\rho_\sigma^2 = \frac{m}{2An}, \quad \rho_\tau^2 = \frac{n}{2Am}.$$

The regions we shall use, in addition to $T_2^{-1}(\mathcal{S})$, are these:

$$\begin{aligned}
\mathcal{Q} &= \big\{\, |\sigma_j| \le n^{-1/2+\epsilon},\ j = 1, \ldots, m{-}1 \,\big\} \cap \big\{\, |\tau_k| \le m^{-1/2+\epsilon},\ k = 1, \ldots, n{-}1 \,\big\} \\
&\qquad \cap \big\{\, |\mu| \le (mn)^{-1/2+2\epsilon} \,\big\} \\
\mathcal{M} &= \big\{\, |\mu_1| \le m^{1/2} n^{-1/2+\epsilon} \,\big\} \cap \big\{\, |\nu_1| \le n^{1/2} m^{-1/2+\epsilon} \,\big\} \\
\mathcal{B} &= \big\{\, (1 - m^{-1/4})\rho_\sigma^2 \le \mu_2 \le (1 + m^{-1/4})\rho_\sigma^2 \,\big\} \\
&\qquad \cap \big\{\, (1 - n^{-1/4})\rho_\tau^2 \le \nu_2 \le (1 + n^{-1/4})\rho_\tau^2 \,\big\}.
\end{aligned}$$

As integrands we will use three functions $E_h = \exp(L_h)$, $h = 1, 2, 3$. The definition of L_1 has appeared already. The function L_2 consists of some of the summands found in L_1:

$$\begin{aligned}
L_2 = {} & -Amn\mu^2 + 6A_4\mu_2\nu_2 + A_4n\mu_4 + A_4m\nu_4 - 3iA_3n\mu\mu_2 - 3iA_3m\mu\nu_2 \\
& - An\mu_2 - Am\nu_2 - iA_3n\mu_3 - iA_3m\nu_3 - 3iA_3cn\mu_2\mu_1 - 3iA_3dm\nu_2\nu_1.
\end{aligned}$$

The third function L_3 equals $\operatorname{Re}(L_2)$, the real part of L_2:

$$L_3 = -Amn\mu^2 + 6A_4\mu_2\nu_2 + A_4n\mu_4 + A_4m\nu_4 - An\mu_2 - Am\nu_2.$$

For convenience we define two expressions in m, n that recur in our big-oh expressions,

$$\begin{aligned}
H_1 &= Am^{1/2+2\epsilon}n^{-1+5\epsilon} + An^{1/2+2\epsilon}m^{-1+5\epsilon} \\
H_2 &= A(mn)^{2\epsilon} + Amn^{-1+4\epsilon} + Am^{-1+4\epsilon}n.
\end{aligned}$$

Having made all the necessary definitions, the next step is to establish a few relationships among the regions and functions just defined. Summing for $1 \le j \le m-1$ the equation $\hat\theta_j = \sigma_j + c\mu_1$, and inserting the value of c, we find

$$m^{-1/2}\,\mu_1 = \sum_{j=1}^{m-1} \hat\theta_j.$$

In the region $\mathcal{S}$ we have $|\sum_{j=1}^{m-1}\hat\theta_j| \le n^{-1/2+\epsilon}$, and so in $T_2^{-1}(\mathcal{S})$ we have

$$|\mu_1| \le m^{1/2}n^{-1/2+\epsilon}.$$

Similarly, $|\nu_1| \le n^{1/2}m^{-1/2+\epsilon}$; using these, the reader can check that

$$\tfrac{1}{2}\mathcal{Q} \cap \mathcal{M} \subseteq T_2^{-1}(\mathcal{S}) \subseteq \tfrac{3}{2}\mathcal{Q} \cap \mathcal{M}.$$

We also have the following bounds in $\frac{3}{2}\mathcal{Q}$:

$$\begin{aligned}\sigma_j &= O(n^{-1/2+\epsilon})\\ \mu_2 &= O(mn^{-1+2\epsilon})\\ \mu_3 &= O(mn^{-3/2+3\epsilon})\\ \mu_4 &= O(mn^{-2+4\epsilon}).\end{aligned}$$

Similar bounds, but with m and n interchanged, hold in $\frac{3}{2}\mathcal{Q}$ for τ_k, ν_2, ν_3, and ν_4. These estimates, along with $A_3, A_4 = O(A)$, $c = O(m^{-1})$, $\mu = O\big((mn)^{-1/2+2\epsilon}\big)$, bounds for c_2, c_3, d, d_2, d_3, and the definition of $\mathcal{M}$, allow us to conclude

$$L_1 = L_2 + O\big(H_1\big), \quad (\mu, \boldsymbol{\sigma}, \boldsymbol{\tau}) \in \tfrac{3}{2}\mathcal{Q} \cap \mathcal{M}.$$

We also record

$$\begin{aligned}|E_2| &= E_3, \quad \text{and}\\ L_3 &= -Amn\mu^2 - An\mu_2 - Am\nu_2 + O(H_2), \quad (\mu, \boldsymbol{\sigma}, \boldsymbol{\tau}) \in \tfrac{3}{2}\mathcal{Q}.\end{aligned}$$

Our strategy for evaluating the integral is presented in the next four equations, and summarized in equation (3.4) below. The principles underlying these equations are familiar: (1) Split an integrand into a principal part and a negligible part; (2) Integrate a positive integrand over a larger region if it helps and only an upper bound is needed; (3) Split a region into two subregions, on one of which the integrand simplifies, and the other of which is negligible; (4) Strive towards integrals which can be evaluated by separating the variables.

$$\begin{aligned}\int_{T_2^{-1}(\mathcal{S})} E_1 &= \int_{T_2^{-1}(\mathcal{S})} E_2 + O(H_1) \int_{\frac{3}{2}\mathcal{Q}} E_3\\ \int_{T_2^{-1}(\mathcal{S})} E_2 &= \int_{\frac{1}{2}\mathcal{Q}\cap\mathcal{M}} E_2 + O(1) \int_{\frac{3}{2}\mathcal{Q}-\frac{1}{2}\mathcal{Q}} E_3\\ \int_{\frac{1}{2}\mathcal{Q}\cap\mathcal{M}} E_2 &= \int_{\frac{1}{2}\mathcal{Q}\cap\mathcal{M}\cap\mathcal{B}} E_2 + O(1) \int_{\mathcal{B}^c\cap\mathcal{Q}} E_3\\ \int_{\frac{1}{2}\mathcal{Q}\cap\mathcal{M}\cap\mathcal{B}} E_2 &= \int_{\frac{1}{2}\mathcal{Q}} E_2 + O(1) \int_{\mathcal{B}^c\cap\mathcal{Q}} E_3 + O(1) \int_{\mathcal{M}^c\cap\mathcal{B}\cap\frac{1}{2}\mathcal{Q}} E_3.\end{aligned}$$

Altogether,

$$\begin{aligned}\int_{T_2^{-1}(\mathcal{S})} E_1 = \int_{\frac{1}{2}\mathcal{Q}} E_2 &+ O(H_1) \int_{\frac{3}{2}\mathcal{Q}} E_3\\ &+ O(1)\bigg(\int_{\frac{3}{2}\mathcal{Q}-\frac{1}{2}\mathcal{Q}} E_3 + \int_{\mathcal{B}^c\cap\mathcal{Q}} E_3 + \int_{\mathcal{M}^c\cap\mathcal{B}\cap\frac{1}{2}\mathcal{Q}} E_3\bigg).\end{aligned} \tag{3.4}$$

Let us now analyze each of the four integrals of E_3 arising in (3.4): over $\frac{3}{2}\mathcal{Q}$, $\frac{3}{2}\mathcal{Q}-\frac{1}{2}\mathcal{Q}$, $\mathcal{B}^{\mathrm{c}}\cap\mathcal{Q}$, and $\mathcal{M}^{\mathrm{c}}\cap\mathcal{B}\cap\frac{1}{2}\mathcal{Q}$. We can integrate E_3 over $\mathcal{Q}$ because the variables almost completely split. Using

$$\left(\int_{|x|\le n^{-1/2+\epsilon}} e^{-Anx^2}\bigl(1+6A_4\nu_2x^2+A_4nx^4+O(A_4^2n^2m^{-2+4\epsilon}x^4+A_4^2n^2x^8)\bigr)\,dx\right)^{m-1}$$

for integration with respect to the σ's, and a similar formula for integration with respect to the τ's, we find

$$\int_{\mathcal{Q}} E_3=\exp\Bigl\{\frac{3A_4}{2A^2}+\frac{3A_4}{4A^2}\Bigl(\frac{m}{n}+\frac{n}{m}\Bigr)+O\bigl(m^{-1+4\epsilon}+n^{-1+4\epsilon}\bigr)\Bigr\}$$
$$\times\Bigl(\frac{\pi}{Amn}\Bigr)^{1/2}\Bigl(\frac{\pi}{An}\Bigr)^{(m-1)/2}\Bigl(\frac{\pi}{Am}\Bigr)^{(n-1)/2}.$$

It is immediate that the same result is obtained for integration over either $\frac{1}{2}\mathcal{Q}$ or $\frac{3}{2}\mathcal{Q}$. Reviewing the previous derivation, we see that if one of the σ_j were restricted to the range

$$\tfrac{1}{2}n^{-1/2+\epsilon}\le|\sigma_j|\le\tfrac{3}{2}n^{-1/2+\epsilon},$$

then the exponent $(m-1)/2$ above would be replaced by $(m-2)/2$, and a new factor would be introduced. To see what this new factor is, we use the inequality

$$\int_{\frac{1}{2}n^{-1/2+\epsilon}}^{\frac{3}{2}n^{-1/2+\epsilon}} e^{-Anx^2}dx\le(An^{1/2+\epsilon})^{-1}\exp\bigl(-\tfrac{1}{4}An^{2\epsilon}\bigr),$$

and note that in the latter interval of integration

$$-Anx^2+6A_4\nu_2x^2+A_4nx^4=-Anx^2\bigl(1+O(m^{-1+2\epsilon}+n^{-1+4\epsilon})\bigr).$$

It follows (using a similar argument if one of the $|\tau_k|$ exceeds $\frac{1}{2}m^{-1/2+\epsilon}$ or if $|\mu|$ exceeds $(mn)^{-1/2+2\epsilon}/2)$, that

$$\int_{\frac{3}{2}\mathcal{Q}-\frac{1}{2}\mathcal{Q}} E_3=O(1)\bigl(e^{-An^{2\epsilon}/4}+e^{-Am^{2\epsilon}/4}\bigr)$$
$$\times\Bigl(\frac{\pi}{Amn}\Bigr)^{1/2}\Bigl(\frac{\pi}{An}\Bigr)^{(m-1)/2}\Bigl(\frac{\pi}{Am}\Bigr)^{(n-1)/2}.$$

To bound the integral of E_3 over $\mathcal{B}^{\mathrm{c}}\cap\mathcal{Q}$, we apply Lemma 2. Recalling that H_2 is the bound for how much L_3 differs from $-Amn\mu^2-An\mu_2-Am\nu_2$ in $\mathcal{Q}$, and noting that $H_2=o(m^{1/2})$ and $H_2=o(n^{1/2})$, we find

$$\int_{\mathcal{B}^{\mathrm{c}}\cap\mathcal{Q}} E_3=O(1)\bigl(e^{-m^{1/2}/6}+e^{-n^{1/2}/6}\bigr)$$
$$\times\Bigl(\frac{\pi}{Amn}\Bigr)^{1/2}\Bigl(\frac{\pi}{An}\Bigr)^{(m-1)/2}\Bigl(\frac{\pi}{Am}\Bigr)^{(n-1)/2}.$$

We now turn to the integral of E_3 over $\mathcal{M}^c \cap \mathcal{B} \cap \frac{1}{2}\mathcal{Q}$. Define κ by

$$\kappa^2 = n^{-\epsilon}.$$

We wish to replace $\frac{1}{2}\mathcal{Q}$ with the smaller $\kappa\mathcal{Q}$, which can be justified in the same manner that we treated the region $\frac{3}{2}\mathcal{Q} - \frac{1}{2}\mathcal{Q}$ a few lines earlier. Because $A_4 n x^4$ is uniformly $o(1)$ in the interval of integration, and because $A\kappa n^{1/2+\epsilon} \to \infty$, we have

$$\int_{\kappa n^{-1/2+\epsilon}}^{\frac{1}{2}n^{-1/2+\epsilon}} \exp\bigl(-Anx^2 + A_4 n x^4\bigr)\,dx = o(1)e^{-A\kappa^2 n^{2\epsilon}}.$$

In $\mathcal{B}$ we have $A_4\mu_2\nu_2 = O(A^{-1})$; moreover,

$$\exp\bigl(O(A^{-1})\frac{m}{n}\bigr)(An)^{1/2} \le e^{A\kappa^2 n^{2\epsilon}/2},$$

since $\log(n) = o(A\kappa^2 n^{2\epsilon})$ by (3.3), and $A^{-1}m/n = o(An^{\epsilon})$. This clears the way to proceed as we did in bounding $\int_{\frac{3}{2}\mathcal{Q}-\frac{1}{2}\mathcal{Q}} E_3$ to find

$$\int_{\mathcal{M}^c\cap\mathcal{B}\cap\frac{1}{2}\mathcal{Q}} E_3 = \int_{\mathcal{M}^c\cap\mathcal{B}\cap\kappa\mathcal{Q}} E_3 + O(1)\bigl(e^{-A\kappa^2 n^{2\epsilon}/2} + e^{-A\kappa^2 m^{2\epsilon}/2}\bigr) \times \Bigl(\frac{\pi}{Amn}\Bigr)^{1/2}\Bigl(\frac{\pi}{An}\Bigr)^{(m-1)/2}\Bigl(\frac{\pi}{Am}\Bigr)^{(n-1)/2}. \qquad (3.5)$$

In $\mathcal{B} \cap \kappa\mathcal{Q}$ we have, in addition to $A_4\mu_2\nu_2 = O(A^{-1})$,

$$A_4 n\mu_4 = O(A)n\mu_2(\kappa n^{-1/2+\epsilon})^2 = O(A\mu_2\kappa^2 n^{2\epsilon}) = O(\kappa^2 m n^{-1+2\epsilon})$$

and a similar bound for $A_4 m\nu_4$; thus,

$$\int_{\mathcal{M}^c\cap\mathcal{B}\cap\kappa\mathcal{Q}} E_3 \le \exp\bigl(O(A^{-1} + \kappa^2 m n^{-1+2\epsilon} + \kappa^2 m^{-1+2\epsilon} n)\bigr) \times \int_{\mathcal{M}^c\cap\mathcal{B}} \exp\bigl(-Amn\mu^2 - An\mu_2 - Am\nu_2\bigr). \qquad (3.6)$$

The complement of $\mathcal{M}$ is the union of

$$\bigl\{\, |\mu_1| \ge m^{1/2} n^{-1/2+\epsilon} \,\bigr\}$$

and

$$\bigl\{\, |\nu_1| \ge n^{1/2} m^{-1/2+\epsilon} \,\bigr\}.$$

Let's assume the first condition holds; the argument is entirely similar if it is the second. The region described by the assumed condition is contained in the region

$$\Bigl|\sum_{j=1}^{m-1} \frac{\sigma_j}{(m-1)^{1/2}}\Bigr| \ge n^{-1/2+\epsilon}.$$

The summation on the left side of the previous is of the form $|\vec{\zeta} \cdot \boldsymbol{\sigma}|$, where $\vec{\zeta}$ is a unit vector. Since the region $\mathcal{B}$ is spherically symmetric, the integral of $\exp(-An\mu_2)$ over $\mathcal{B} \cap \{|\vec{\zeta} \cdot \boldsymbol{\sigma}| \geq \cdots\}$ is independent of the unit vector $\vec{\zeta}$. If we replace $\vec{\zeta}$ by the vector $(1, 0, \ldots, 0)$, then we may integrate over $\mathcal{B} \cap \{|\sigma_1| \geq n^{-1/2+\epsilon}\}$. Throughout the latter region, the integrand on the right of (3.6) is bounded above by

$$\exp(-An^{2\epsilon}) \exp\Bigl(-Amn\mu^2 - An \sum_{j=2}^{m-1} \sigma_j^2 - Am\nu_2\Bigr).$$

Since, in $\mathcal{B}$, $\sigma_1^2 \leq \mu_2 = O(m/An)$, we have $\mathcal{B} \subseteq \bigl\{\, |\sigma_1| = O((m/An)^{1/2}) \,\bigr\}$, and so

$$\begin{aligned}
&\int_{\mathcal{B}} \exp\Bigl(-Amn\mu^2 - An \sum_{j=2}^{m-1} \sigma_j^2 - Am\nu_2\Bigr) \\
&\qquad = O(1) \Bigl(\frac{m}{An}\Bigr)^{1/2} \Bigl(\frac{\pi}{Amn}\Bigr)^{1/2} \Bigl(\frac{\pi}{An}\Bigr)^{(m-2)/2} \Bigl(\frac{\pi}{Am}\Bigr)^{(n-1)/2}.
\end{aligned}$$

Summarizing, with H_3 an abbreviation for $A^{-1} + \kappa^2 mn^{-1+2\epsilon} + \kappa^2 m^{-1+2\epsilon} n$, and noting $H_3 = o(An^{2\epsilon})$ and $H_3 = o(Am^{2\epsilon})$ (because $A^{-1}m/n = o(An^{\epsilon})$),

$$\begin{aligned}
&\int_{\mathcal{M}^c \cap \mathcal{B} \cap \kappa\mathcal{Q}} E_3 \\
&\qquad \leq \exp\bigl(O(H_3)\bigr) \int_{\mathcal{M}^c \cap \mathcal{B}} \exp\bigl(-Amn\mu^2 - An\mu_2 - Am\nu_2\bigr) \\
&\qquad \leq \exp\bigl(O(H_3)\bigr) \int_{\{|\sigma_1| \geq n^{1/2+\epsilon}\} \cap \mathcal{B}} \exp\bigl(-Amn\mu^2 - An\mu_2 - Am\nu_2\bigr) \\
&\qquad\qquad + \exp\bigl(O(H_3)\bigr) \int_{\{|\tau_1| \geq m^{1/2+\epsilon}\} \cap \mathcal{B}} \exp\bigl(-Amn\mu^2 - An\mu_2 - Am\nu_2\bigr) \\
&\qquad \leq \exp\bigl(-An^{2\epsilon}/2\bigr) \int_{\mathcal{B}} \exp\Bigl(-Amn\mu^2 - An \sum_{j=2}^{m-1} \sigma_j^2 - Am\nu_2\Bigr) \\
&\qquad\qquad + \exp\bigl(-Am^{2\epsilon}/2\bigr) \int_{\mathcal{B}} \exp\Bigl(-Amn\mu^2 - An\mu_2 - Am \sum_{k=2}^{n-1} \tau_k^2\Bigr) \\
&\qquad = O(1) \bigl(e^{-An^{2\epsilon}/3} + e^{-Am^{2\epsilon}/3}\bigr) \Bigl(\frac{\pi}{Amn}\Bigr)^{1/2} \Bigl(\frac{\pi}{An}\Bigr)^{(m-1)/2} \Bigl(\frac{\pi}{Am}\Bigr)^{(n-1)/2}.
\end{aligned}$$

Combining this with (3.5), we have altogether

$$\begin{aligned}
\int_{\mathcal{M}^c \cap \mathcal{B} \cap \frac{1}{2}\mathcal{Q}} E_3 &= O(1)\bigl(e^{-A\kappa^2 n^{2\epsilon}/2} + e^{-A\kappa^2 m^{2\epsilon}/2}\bigr) \\
&\qquad \times \Bigl(\frac{\pi}{Amn}\Bigr)^{1/2} \Bigl(\frac{\pi}{An}\Bigr)^{(m-1)/2} \Bigl(\frac{\pi}{Am}\Bigr)^{(n-1)/2}.
\end{aligned}$$

Looking back at equation (3.4), we have now bounded all four of the error terms – the four integrals of E_3 over various regions – appearing on the right side of that equation. We have

$$\begin{aligned}\frac{3A_4}{2A^2} &+ \frac{3A_4}{4A^2}\Big(\frac{m}{n}+\frac{n}{m}\Big) + O\big(m^{-1+4\epsilon}+n^{-1+4\epsilon}\big)\\ &= -\tfrac{1}{2}+\tfrac{1}{8}A^{-1}(1-2\lambda)^2+\tfrac{1}{16}A^{-1}(1-6\lambda+6\lambda^2)\Big(\frac{m}{n}+\frac{n}{m}\Big)+o(1)\\ &= o(An^{\epsilon}).\end{aligned}$$

Also, by (3.3),

$$\log H_1^{-1} = O(1)\big(\log A^{-1}+\log m\big) = o(A\kappa^2 n^{2\epsilon}).$$

It follows, recalling $An^{\epsilon} = A\kappa^2 n^{2\epsilon}$, that the last three error terms in (3.4) are all little-oh of the first, $O(H_1)\int_{\frac{3}{2}\mathcal{Q}} E_3$. This allows us to conclude

$$\begin{aligned}\int_{T_2^{-1}(\mathcal{S})} E_1 = \int_{\frac{1}{2}\mathcal{Q}} & E_2\\ &+ O(H_1)\exp\Big\{-\tfrac{1}{2}+\tfrac{1}{8}A^{-1}(1-2\lambda)^2+\tfrac{1}{16}A^{-1}(1-6\lambda+6\lambda^2)\Big(\frac{m}{n}+\frac{n}{m}\Big)\Big\}\\ &\times\Big(\frac{\pi}{Amn}\Big)^{1/2}\Big(\frac{\pi}{An}\Big)^{(m-1)/2}\Big(\frac{\pi}{Am}\Big)^{(n-1)/2}. \qquad (3.7)\end{aligned}$$

It remains to compute the integral of E_2 over $\frac{1}{2}\mathcal{Q}$. We proceed in three stages, starting with integration with respect to μ. For the latter, the first step is to replace the limits of integration with $\pm\infty$:

$$\begin{aligned}&\int_{|\mu|\le(mn)^{-1/2+2\epsilon}} \exp\big(-Amn\mu^2-3iA_3(n\mu_2+m\nu_2)\mu\big)\,d\mu\\ &\qquad= \int_{-\infty}^{+\infty}\langle\text{same}\rangle + O(1)\int_{|\mu|\ge(mn)^{-1/2+2\epsilon}} e^{-Amn\mu^2}d\mu\\ &\qquad= \int_{-\infty}^{+\infty}\langle\text{same}\rangle + O(1)\,\big(A(mn)^{1/2+2\epsilon}\big)^{-1}\,\exp\big(-A(mn)^{4\epsilon}\big).\end{aligned}$$

To integrate over the real line, we use the formula (for β real)

$$\int_{-\infty}^{+\infty}\exp\big(-Amn\mu^2-i\beta\mu\big)\,d\mu = \sqrt{\frac{\pi}{Amn}}\,\exp\Big(-\frac{\beta^2}{4Amn}\Big).$$

Since

$$\frac{\big(A_3(n\mu_2+m\nu_2)\big)^2}{4Amn} = O\big(Amn^{-1+4\epsilon}+Anm^{-1+4\epsilon}\big) = o\big(A(mn)^{4\epsilon}\big),$$

integration with respect to μ of E_2 equals

$$\Big(\frac{\pi}{Amn}\Big)^{1/2}\exp\Big(\frac{-9A_3^2(n\mu_2+m\nu_2)^2}{4Amn}+o\big(e^{-A(mn)^{4\epsilon}/2}\big)\Big).$$

The second step is to integrate with respect to $\boldsymbol{\sigma}$ the integrand

$$\exp\Big(-An\mu_2 + \Big(6A_4 - \frac{9A_3^2}{2A}\Big)\nu_2\mu_2 - \Big(\frac{9A_3^2 n}{4Am}\Big)\mu_2^2 \\ + A_4 n\mu_4 - iA_3 n\mu_3 - 3iA_3 cn\mu_2\mu_1 + o\big(e^{-A(mn)^{4\epsilon}/2}\big)\Big).$$

This is accomplished by an appeal to Lemma 1. We apply the latter with $N = m-1$, $\delta = \frac{3}{4}$, say, and

$$\begin{aligned}
\hat{A} &= A\frac{n}{m-1}\Big(1 - \frac{6A_4\nu_2}{An} + \frac{9A_3^2\nu_2}{2A^2 n}\Big) &&= A\frac{n}{m}(1 + O(m^{-1+2\epsilon})) \\
\hat{B} &= -\frac{iA_3 n}{m-1} &&= -iA_3\frac{n}{m}(1 + O(m^{-1})) \\
\hat{C} &= -3iA_3 cn &&= 3iA_3\frac{n}{m}(1 + O(m^{-1/2})) \\
\hat{E} &= \frac{A_4 n}{m-1} &&= A_4\frac{n}{m}(1 + O(m^{-1})) \\
\hat{F} &= -\frac{9A_3^2 n}{4Am} &&= -\frac{9A_3^2}{4A}\frac{n}{m}.
\end{aligned}$$

We need

$$\frac{3\hat{E} + \hat{F}}{4\hat{A}^2} = \frac{m}{n}\Big(\frac{3A_4}{4A^2} - \frac{9A_3^2}{16A^3}\Big) + O(A^{-1}n^{-1} + m^{2\epsilon}n^{-1})$$

and

$$\frac{15\hat{B}^2 + 6\hat{B}\hat{C} + \hat{C}^2}{16\hat{A}^3} = -\frac{3A_3^2 m}{8A^3 n} + O(A^{-1}m^{1/2}n^{-1}).$$

Then, integration with respect to the σ_j contributes a τ-free factor

$$\Big(\frac{\pi}{An}\Big)^{(m-1)/2}\exp\Big\{\frac{m}{n}\Big(\frac{3A_4}{4A^2} - \frac{15A_3^2}{16A^3}\Big)\Big\}$$

and for the final integrand we are left with

$$\exp\Big\{-Am\nu_2 + \Big(\frac{3A_4 m}{An} - \frac{9A_3^2 m}{4A^2 n}\Big)\nu_2 - \frac{9A_3^2 m}{4An}\nu_2^2 + A_4 m\nu_4 \\ - iA_3 m\nu_3 - 3iA_3 dm\nu_2\nu_1 + O\big(m^{-1/2+12\epsilon}\hat{Z} + A^{-1}m^{3/4+3\epsilon}n^{-1}\big)\Big\},$$

where

$$\hat{Z} = \exp\Big\{\frac{(1+o(1))(1-2\lambda)^2}{24}A^{-1}\frac{m}{n}\Big\}.$$

Again, we make use of Lemma 1. This time we take $N = n-1$ and we claim that we may take $\delta = \frac{1}{4} - 4\epsilon$. To justify this claim, we must check that both $A^{-1}m^{3/4+3\epsilon}n^{-1}$ and $m^{-1/2+12\epsilon}\hat{Z}$ are $O(n^{-1/4+4\epsilon})$. The first follows from $A^{-1}m/n = O(An^{\epsilon})$. For the second,

$$m^{-1/2+12\epsilon}\hat{Z} \le n^{-1/2+12\epsilon+\gamma/24+o(1)},$$

and the latter is $O(n^{-1/4+4\epsilon})$ by our condition on γ. This justifies the claim that δ may be taken to be $1/4 - 4\epsilon$.

After calculations similar to the previous, we find the third and final factor, from integration with respect to the τ_k's, is equal to

$$\Big(\frac{\pi}{Am}\Big)^{(n-1)/2} \exp\Big\{ \frac{n}{m}\Big(\frac{3A_4}{4A^2} - \frac{15A_3^2}{16A^3}\Big) + \frac{3A_4}{2A^2} - \frac{9A_3^2}{8A^3} + O\big(n^{-\delta} Z_{\text{final}} + A^{-1}n^{3/4+3\epsilon}m^{-1}\big)\Big\},$$

where

$$Z_{\text{final}} = \exp\Big\{ \frac{(1+o(1))(1-2\lambda)^2}{24} A^{-1}\frac{n}{m} \Big\}.$$

We calculate this time that

$$n^{-\delta} Z_{\text{final}} \le n^{-1/4+4\epsilon+\gamma/24+o(1)},$$

and that $A^{-1}n^{3/4+3\epsilon}m^{-1}$ is negligible in comparison. When we multiply the three factors, and perform the algebra

$$\frac{3A_4}{2A^2} - \frac{9A_3^2}{8A^3} = -\frac{1}{2}$$
$$\frac{3A_4}{4A^2} - \frac{15A_3^2}{16A^3} = -\frac{1-2A}{24A},$$

we find

$$\int_{\frac{1}{2}\mathcal{Q}} E_2 = \exp\Big\{-\frac{1}{2} - \frac{1-2A}{24A}\Big(\frac{m}{n} + \frac{n}{m}\Big) + O\big(n^{-1/4+4\epsilon+\gamma/24+o(1)}\big)\Big\} \times \Big(\frac{\pi}{Amn}\Big)^{1/2}\Big(\frac{\pi}{An}\Big)^{(m-1)/2}\Big(\frac{\pi}{Am}\Big)^{(n-1)/2}.$$

To obtain the formula for $J(m,n)$ stated in the theorem, we combine the previous equation with (3.7). Start with the algebraic calculation

$$\tfrac{1}{24}(1-2A) + \tfrac{1}{16}(1-6\lambda+6\lambda^2)s = \tfrac{5}{48}(1-2\lambda)^2,$$

and the estimate

$$\tfrac{1}{8}A^{-1}(1-2\lambda)^2 + \tfrac{5}{48}A^{-1}(1-2\lambda)^2\Big(\frac{m}{n} + \frac{n}{m}\Big) = A^{-1}(1-2\lambda)^2\,\frac{m}{n}\,\Big(\frac{5}{48} + \frac{n}{8m} + \frac{5n^2}{48m^2}\Big) \le \tfrac{1}{3}\gamma\log n.$$

Then,

$$H_1 \exp\Big\{\tfrac{1}{8}A^{-1}(1-2\lambda)^2 + \tfrac{5}{48}A^{-1}(1-2\lambda)^2\Big(\frac{m}{n} + \frac{n}{m}\Big)\Big\} \le 2Am^{1/2+2\epsilon}n^{-1+5\epsilon}n^{\gamma/3} = O\big(n^{-1/2+\gamma/3+15\epsilon/2+2\epsilon^2}\big).$$

Thus, the sum of $\int_{\frac{1}{2}\mathcal{Q}} E_2$ and

$$O(H_1)\exp\Bigl\{-\tfrac{1}{2}+\tfrac{1}{8}A^{-1}(1-2\lambda)^2+\tfrac{1}{16}A^{-1}(1-6\lambda+6\lambda^2)\Bigl(\frac{m}{n}+\frac{n}{m}\Bigr)\Bigr\} \times\Bigl(\frac{\pi}{Amn}\Bigr)^{1/2}\Bigl(\frac{\pi}{An}\Bigr)^{(m-1)/2}\Bigl(\frac{\pi}{Am}\Bigr)^{(n-1)/2}$$

is

$$\exp\Bigl\{-\frac{1}{2}-\frac{1-2A}{24A}\Bigl(\frac{m}{n}+\frac{n}{m}\Bigr)\Bigr\}\bigl(1+O(n^{-1/4+\gamma/24+4\epsilon+o(1)})+O(n^{-1/2+\gamma/3+15\epsilon/2+2\epsilon^2})\bigr) \times\Bigl(\frac{\pi}{Amn}\Bigr)^{1/2}\Bigl(\frac{\pi}{An}\Bigr)^{(m-1)/2}\Bigl(\frac{\pi}{Am}\Bigr)^{(n-1)/2},$$

which equals

$$\exp\Bigl\{-\frac{1}{2}-\frac{1-2A}{24A}\Bigl(\frac{m}{n}+\frac{n}{m}\Bigr)+O(n^{-1/4+\gamma/24+4\epsilon+o(1)}+n^{-1/2+\gamma/3+15\epsilon/2+2\epsilon^2})\Bigr\} \times\Bigl(\frac{\pi}{Amn}\Bigr)^{1/2}\Bigl(\frac{\pi}{An}\Bigr)^{(m-1)/2}\Bigl(\frac{\pi}{Am}\Bigr)^{(n-1)/2}.$$

This completes the proof of Theorem 2. □

4 Concentration of the integral

In this section we will complete the estimation of $I(m,n)$ by establishing the following.

Theorem 3. *Define I_0 by*

$$I_0=(mn)^{1/2}\Bigl(\frac{\pi}{Amn}\Bigr)^{1/2}\Bigl(\frac{\pi}{An}\Bigr)^{(m-1)/2}\Bigl(\frac{\pi}{Am}\Bigr)^{(n-1)/2}\exp\Bigl\{-\frac{1-2A}{24A}\Bigl(\frac{m}{n}+\frac{n}{m}\Bigr)\Bigr\}. \qquad (4.1)$$

For sufficiently small $\epsilon>0$, if $m=o(A^2n^{1+2\epsilon})$, and $n=o(A^2m^{1+2\epsilon})$, then

$$I(m,n)=I_{\mathcal{R}}(m,n)+O(n^{-1})I_0.$$

To motivate the definition of I_0, recall that it was shown in the previous section to be within a constant of $I_{\mathcal{R}}(m,n)$ under stronger conditions than we wish to assume in the present section.

We begin with two technical lemmas whose proofs are omitted.

Lemma 3. *The absolute value of the integrand of $I(m,n)$ is*

$$F(\boldsymbol{\theta},\boldsymbol{\phi})=\prod_{j,k} f(\theta_j+\phi_k),$$

where

$$f(z)=\sqrt{1-4A(1-\cos z)}.$$

Moreover, for all real z,

$$0\le f(z)\le \exp\bigl(-Az^2+\tfrac{1}{12}Az^4\bigr).\quad\square$$

Lemma 4. *For all* $c > 0$,

$$\int_{-\pi/10}^{\pi/10} \exp\bigl(c(-x^2 + \tfrac{9}{4}x^4)\bigr)\, dx \le \sqrt{\pi/c}\, \exp(2/c). \quad \square$$

Proof of Theorem 3. Our approach will be to bound $\int F(\boldsymbol{\theta}, \boldsymbol{\phi})$ over a variety of regions whose union covers $C^{m+n} \setminus \mathcal{R}$.

Take any small $\delta > 0$. By the pigeon hole principle, there is some interval $[x, x+\delta]$ that contains at least $\delta m/2\pi$ values of θ_j. Let $\mathcal{S}_1(x)$ be the set of $(\boldsymbol{\theta}, \boldsymbol{\phi})$ such that $\theta_j \in [x, x+\delta]$ for at least $\delta m/2\pi$ values of j and $\phi_k \notin [-x - 2\delta, -x + \delta]$ for at least n^ϵ values of k. By Lemma 3, $F(\boldsymbol{\theta}, \boldsymbol{\phi}) \le \exp(-c_1 A m n^\epsilon)$ for some $c_1 > 0$ and so the contribution from $\mathcal{S}_1$ is at most

$$\int_{\mathcal{S}_1(x)} F(\boldsymbol{\theta}, \boldsymbol{\phi}) \le (2\pi)^{m+n} \exp(-c_1 A m n^\epsilon).$$

Next define $S_2(x)$ to be the set of $(\boldsymbol{\theta}, \boldsymbol{\phi})$ such $\theta_j \notin [x - 2\delta, x + 3\delta]$ for at least m^ϵ values of j. By the same argument as before with the roles of $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ reversed,

$$\int_{\mathcal{S}_1(x)\cap\mathcal{S}_2(x)} F(\boldsymbol{\theta}, \boldsymbol{\phi}) \le (2\pi)^{m+n} \exp(-c_2 A m^\epsilon n) \tag{4.2}$$

for some $c_2 > 0$.

If we subtract x from each θ_j and add x to each ϕ_k the integrand $F(\boldsymbol{\theta}, \boldsymbol{\phi})$ is unchanged. Thus we can assume that $x = 0$ from now on, after multiplying (4.2) by 2π to cover all possible x. We will also fix $\delta = \pi/300$. Define $\mathcal{R}_1$ to be the set of $(\boldsymbol{\theta}, \boldsymbol{\phi})$ such that $|\theta_j| > \pi/100$ for at most m^ϵ values of j, and $|\phi_k| > \pi/100$ for at most n^ϵ values of k. Under our just-made assumption, we have proved that

$$\int_{C^n\setminus\mathcal{R}_1} F(\boldsymbol{\theta}, \boldsymbol{\phi}) \le (2\pi)^{m+n}\bigl(\exp(-c_3 A m n^\epsilon) + \exp(-c_3 A m^\epsilon n)\bigr) \tag{4.3}$$

for some $c_3 > 0$.

Assume $(\boldsymbol{\theta}, \boldsymbol{\phi}) \in \mathcal{R}_1$. Define $S_0 = S_0(\boldsymbol{\theta})$, $S_1 = S_1(\boldsymbol{\theta})$ and $S_2 = S_2(\boldsymbol{\theta})$ to be the indices j such that $|\theta_j| \le \frac{1}{100}\pi$, $\frac{1}{100}\pi < |\theta_j| \le \frac{1}{20}\pi$, and $|\theta_j| > \frac{1}{20}\pi$, respectively. Similarly define $T_0 = T_0(\boldsymbol{\phi})$, $T_1 = T_1(\boldsymbol{\phi})$ and $T_2 = T_2(\boldsymbol{\phi})$.

The value of $F(\boldsymbol{\theta}, \boldsymbol{\phi})$ can now be bounded using

$$f(\theta_j + \phi_k) \le \begin{cases} \exp\bigl(-A(\theta_j+\phi_k)^2 + \frac{1}{12}A(\theta_j+\phi_k)^4\bigr) & \text{if } (j,k) \in (S_0 \cup S_1) \times (T_0 \cup T_1) \\ \sqrt{1 - 4A(1 - \cos(\frac{1}{25}\pi))} \le e^{-A/6} & \text{if } (j,k) \in (S_0 \times T_2) \cup (S_2 \times T_0) \\ 1 & \text{otherwise.} \end{cases}$$

Let $I_2(m_2, n_2)$ be the contribution to $\int_{\mathcal{R}_1} F(\boldsymbol{\theta}, \boldsymbol{\phi})$ of those $(\boldsymbol{\theta}, \boldsymbol{\phi})$ with $|S_2| = m_2$ and $|T_2| = n_2$. Recall that $|S_0| \ge m - m^\epsilon$ and $|T_0| \ge n - n^\epsilon$. We have

$$|I_2(m_2, n_2)| \le \binom{m}{m_2}\binom{n}{n_2}(2\pi)^{m_2+n_2} \exp\bigl(-\tfrac{1}{6}A(n - n^\epsilon)m_2 - \tfrac{1}{6}A(m - m^\epsilon)n_2\bigr) I_2'(m_2, n_2),$$

where

$$I_2'(m_2,n_2) = \int_{-\pi/20}^{\pi/20} \cdots \int_{-\pi/20}^{\pi/20} \exp\bigl(-A\sum\nolimits'' (\theta_j+\phi_k)^2 + \tfrac{1}{12}A\sum\nolimits''(\theta_j+\phi_k)^4\bigr)\, d\boldsymbol{\theta}'' d\boldsymbol{\phi}'',$$

and the double-primes denote restriction to $j \in S_0 \cup S_1$ and $k \in T_0 \cup T_1$. Write $m' = m - m_2$ and $n' = n - n_2$ and define $\bar{\boldsymbol{\theta}} = \frac{1}{m'}\sum'' \theta_j$, $\hat\theta_j = \theta_j - \bar{\boldsymbol{\theta}}$ for all j, $\bar{\boldsymbol{\phi}} = \frac{1}{n'}\sum'' \phi_k$, $\hat\phi_k = \phi_k - \bar{\boldsymbol{\phi}}$, $\mu = \bar{\boldsymbol{\phi}} + \bar{\boldsymbol{\theta}}$ and $\nu = \bar{\boldsymbol{\phi}} - \bar{\boldsymbol{\theta}}$. Change variables from $(\boldsymbol{\theta}'', \boldsymbol{\phi}'')$ to $\{\hat\theta_j \mid j \in S_3\} \cup \{\hat\phi_k \mid k \in T_3\} \cup \{\mu, \nu\}$, where S_3 is some subset of $m'-1$ elements of $S_0 \cup S_1$ and T_3 is some subset of $n'-1$ elements of $T_0 \cup T_1$. From the previous section we know that the determinant of this transformation is $2/(m'n')$. The integrand of I_2' can now be bounded using

$$\sum\nolimits''(\theta_j+\phi_k)^2 = m'n'\mu^2 + n'\sum\nolimits_j'' \hat\theta_j^2 + m'\sum\nolimits_k'' \hat\phi_k^2$$

and

$$\sum\nolimits''(\theta_j+\phi_k)^4 \le 27m'n'\mu^4 + 27n'\sum\nolimits_j'' \hat\theta_j^4 + 27m'\sum\nolimits_k'' \hat\phi_k^4.$$

For an upper bound we can restrict the sums to $j \in S_3$ and $k \in T_3$, since $-x^2 + \frac{9}{4}x^4 < 0$ for $|x| \le \frac{1}{10}\pi$. The integral now separates over the new variables and Lemma 4 gives that

$$I_2'(m_2,n_2) = O(1)\frac{\pi^{(m'+n')/2}}{A^{(m'+n'-1)/2}(m')^{n'/2-1}(n')^{m'/2-1}} \exp\bigl(O(m'/(An') + n'/(Am'))\bigr).$$

Applying (4.1), we find that

$$\sum_{\substack{m_2=0\\ m_2+n_2\ge 1}}^{m^\epsilon} \sum_{n_2=0}^{n^\epsilon} |I_2(m_2,n_2)| \le O\bigl(e^{-c_4Am} + e^{-c_4An}\bigr) I_{\mathcal{R}}(m,n)$$

for some $c_4 > 0$.

Finally we consider the case where $m_2 = n_2 = 0$ in the previous calculation. That is, we have that $|\theta_j| \le \pi/20$ and $|\phi_k| \le \pi/20$ for all j and k. Apply the same transformation as before and bound it by a separable integral as before. The total value of the separable bound is

$$\frac{\pi^{(m+n)/2}}{A^{(m+n-1)/2}m^{n/2-1}n^{m/2-1}} \exp\bigl(O(m/(An) + n/(Am))\bigr).$$

Since $-x^2 + \frac{9}{4}x^4$ is unimodal in $[-\pi/10, \pi/10]$, we easily see that the value is multiplied by a factor of $O(e^{-A(mn)^{4\epsilon}/2})$ by the restriction $\mu > (mn)^{-1/2+2\epsilon}$. Similarly, restricting any θ_j to $|\hat\theta_j| > m^{-1/2+\epsilon}$ multiplies the value by a factor of $O(e^{-Am^{2\epsilon}/2})$ (Choose the transformation such that $\hat\theta_j$ is one of those integrated over.), and restricting any ϕ_k to $|\hat\phi_k| > m^{-1/2+\epsilon}$ multiplies the value by a factor of $O(e^{-An^{2\epsilon}/2})$.

In summary, the integral of $F(\boldsymbol{\theta}, \boldsymbol{\phi})$ over $[-\pi,\pi]^{m+n} \setminus \mathcal{R}$ is

$$O\bigl(e^{-c_5Am^{2\epsilon}} + e^{-c_5An^{2\epsilon}}\bigr) I_0$$

for some $c_5 > 0$. This completes the proof. □

5 Highly oblong matrices

In the case that m is much smaller than n, or vice-versa, we can use a similar but much simpler calculation to estimate $B(m,s;n,t)$.

To be precise, we will assume that for some sufficiently small $\epsilon > 0$ we have that $1 \le t \le m-1$ and

$$m = O\big((t(m-t)n)^{1/4-\epsilon}\big). \tag{5.1}$$

Unlike in the previous calculation, all values of λ except 0 and 1 are permitted.

For a vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_m)$, define the scaled elementary symmetric function

$$\varphi_t(\boldsymbol{x}) = \binom{m}{t}^{-1} \sum_{1\le j_1<j_2<\cdots<j_t\le m} x_{j_1}x_{j_2}\cdots x_{j_t}.$$

Then $B(m,s;n,t)$ is clearly the coefficient of $x_1^s \cdots x_m^s$ in $\binom{m}{t}^n \varphi_t(\boldsymbol{x})^n$.

Applying Cauchy's Theorem with $x_j = e^{i\theta_j}$ for all j, we have

$$B(m,s;n,t) = \binom{m}{t}^n (2\pi)^{-m} K(m,n), \tag{5.2}$$

where

$$K(m,n) = \int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} \frac{\varphi_t\big(e^{i\theta_1}, e^{i\theta_2}, \ldots, e^{i\theta_m}\big)^n}{e^{is\sum_j \theta_j}}\, d\boldsymbol{\theta}, \tag{5.3}$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$.

In Lemma 6 we will estimate $K_{\mathcal{U}}(m,n)$, which is the contribution to $K(m,n)$ of those $\boldsymbol{\theta}$ inside a small region $\mathcal{U}$, then in Lemma 7 we will show that the contributions from the other regions are negligible in comparison.

First we will prove a technical lemma that will be needed soon. For $k \ge 1$, and vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_m)$, define the symmetric functions

$$\varrho_{t,k}(\boldsymbol{x}) = \binom{m}{t}^{-1} \sum_{1\le j_1<j_2<\cdots<j_t\le m} (x_{j_1} + x_{j_2} + \cdots + x_{j_t})^k$$

$$\pi_k(\boldsymbol{x}) = \sum_{j=1}^{m} x_j^k.$$

Lemma 5. *For $m \ge 4$ and $1 \le t \le m-1$, let the real vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_m)$ be such that $\sum_j x_j = 0$ and $\max_j |x_j| = 1$. Then the following hold.*
(a)

$$\varrho_{t,2}(\boldsymbol{x}) = \frac{t(m-t)}{m(m-1)}\pi_2(\boldsymbol{x})$$

$$\varrho_{t,3}(\boldsymbol{x}) = \frac{t(m-t)(m-2t)}{m(m-1)(m-2)}\pi_3(\boldsymbol{x})$$

$$\varrho_{t,4}(\boldsymbol{x}) = \frac{t(m-t)(m^2+m-6tm+6t^2)}{m(m-1)(m-2)(m-3)}\pi_4(\boldsymbol{x}) + \frac{3t(t-1)(m-t)(m-t-1)}{m(m-1)(m-2)(m-3)}\pi_2(\boldsymbol{x})^2$$

(b)

$$\frac{t(m-t)}{(m-1)^2} \le \varrho_{t,2}(\boldsymbol{x}) \le \frac{t(m-t)}{m-1}$$
$$|\varrho_{t,3}(\boldsymbol{x})| \le \frac{t(m-t)|m-2t|}{4(m-1)(m-2)}$$
$$\varrho_{t,4}(\boldsymbol{x}) \le \frac{t(m-t)\big(3t(m-t)-2m\big)}{(m-1)(m-3)}$$
$$\varrho_{t,2}(\boldsymbol{x})^2 \le \varrho_{t,4}(\boldsymbol{x}) \le \frac{m^2}{t(m-t)}\varrho_{t,2}(\boldsymbol{x})^2$$

Proof. Since $\varrho_{t,k}(\boldsymbol{x})$ is a symmetric polynomial of total degree k, the fundamental theorem of symmetric functions tells us that identities of the form given in (a) must exist, recalling that $\pi_1(\boldsymbol{x}) = 0$. The coefficients can be determined by choosing one or two values of $\boldsymbol{x}$.

In light of part (a), the first line of (b) requires maximum and minimum values of $\pi_2(\boldsymbol{x})$. If $x_j \ne x_k$, then $\big(\partial/\partial x_j - \partial/\partial x_k\big)\pi_2(\boldsymbol{x})$ has the same sign as $x_j - x_k$. Thus, $\pi_2(\boldsymbol{x})$ is decreased if x_j and x_k are moved slightly towards each other, which can be done within the constraints on $\boldsymbol{x}$ unless $|x_\ell| = 1$ for exactly one value of ℓ, and the other entries of $\boldsymbol{x}$ are equal. This therefore locates the minimum of $\pi_2(\boldsymbol{x})$. The location of the maximum can be similarly identified, but it is easier to just note that $\pi_2(\boldsymbol{x}) \le m$ trivially.

For the second line of (b), we work similarly with $\pi_3(\boldsymbol{x})$. If $x_j \ne x_k$, then $\big(\partial/\partial x_j - \partial/\partial x_k\big)\pi_3(\boldsymbol{x})$ has the same sign as $x_j^2 - x_k^2$. This shows that the maximum occurs when ℓ of the entries equal 1 and the other $m-\ell$ are equal, for some ℓ. The value of $\pi_3(\boldsymbol{x})$ in this case is maximized when $\ell = \lfloor m/3 \rfloor$ or $\ell = \lceil m/3 \rceil$.

The same method also works for the third line of (b). If $j \ne k$, then

$$\begin{aligned}&\big(\partial/\partial x_j - \partial/\partial x_k\big)\varrho_{t,4} \\ &\qquad = 4(x_j - x_k)\big((m^2 - m - 6tm + 6t^2)(x_j^2 + x_j x_k + x_k^2) + 3t(m-t-1)\pi_2(\boldsymbol{x})\big).\end{aligned}$$

The quadratic form multiplying $x_j - x_k$ has non-negative eigenvalues, so we have that $\big(\partial/\partial x_j - \partial/\partial x_k\big)\varrho_{t,4}$ is zero or has the same sign as $x_j - x_k$. Thus, the maximum occurs if the entries of $\boldsymbol{x}$ are evenly divided between -1 and 1, with one zero value for odd m. This gives the desired bound.

The left side of the last line of (b) is just Cauchy's inequality. For the right side, from (a) we know that it suffices to bound $\pi_4(\boldsymbol{x})/\pi_2(\boldsymbol{x})^2$. Either maximum or minimum is required, depending on the sign of the coefficient of $\pi_4(\boldsymbol{x})$. Also, we can ignore the constraint $\max_j |x_j| = 1$ because $\pi_4(\boldsymbol{x})/\pi_2(\boldsymbol{x})^2$ is independent of scale. For distinct i, j, k, the operator $\nabla_{ijk} = (x_k - x_j)\partial/\partial x_i + (x_i - x_k)\partial/\partial x_j + (x_j - x_i)\partial/\partial x_k$ gives 0 when applied to $\pi_1(\boldsymbol{x})$ or $\pi_2(\boldsymbol{x})$. Applying it to $\pi_4(\boldsymbol{x})$ gives $-4(x_k - x_j)(x_i - x_k)(x_j - x_i)(x_i + x_j + x_k)$. If $\boldsymbol{x}$ has four or more distinct entries, we can choose three of them that don't sum to 0 and choose to either increase or decrease $\pi_4(\boldsymbol{x})$ by slight movements. Thus, the maximum and minimum both occur with at most three distinct values, and if there are three they must sum to 0. In the latter case, we can move one x_j of each value without changing $\pi_4(\boldsymbol{x})$ then increase or decrease $\pi_4(\boldsymbol{x})$ as before. Therefore, both the minimum and maximum occur when there are only two distinct values. By direct computation, we now find that $\pi_4(\boldsymbol{x})/\pi_2(\boldsymbol{x})^2$ is minimized when half the entries are equal and positive while half are equal and negative, and minimized when one entry is positive and the rest are equal and negative (or vice-versa). This gives

$$\frac{1}{m} \le \frac{\pi_4(\boldsymbol{x})}{\pi_2(\boldsymbol{x})^2} \le \frac{m^2-3m+3}{m(m-1)}.$$

The required inequality now follows. □

Define $\mathcal{U}$ to be the set of vectors $\boldsymbol{\theta} \in \hat{C}^m$ such that

$$|\hat{\theta}_j| \le (An)^{-1/2+\epsilon/4}, \quad 1 \le j \le m,$$

where $\hat{\theta}_j = \theta_j - \bar{\boldsymbol{\theta}}$ and $A = \frac{1}{2}\lambda(1-\lambda)$ as before.

Lemma 6. *If $m \ge 4$ and condition (5.1) holds, then*

$$K_{\mathcal{U}}(m,n) = \bigl(1+O(n^{-\epsilon})\bigr)(2\pi)^{(m+1)/2} m^{m/2} \left(\frac{m-1}{t(m-t)n}\right)^{(m-1)/2}.$$

Proof. In the integral (5.3), change variables from $\boldsymbol{\theta}$ to $(\bar{\boldsymbol{\theta}}, \hat{\theta}_1, \ldots, \hat{\theta}_{m-1})$. This transformation, which has Jacobian m, produces an integrand independent of $\bar{\boldsymbol{\theta}}$, so we can integrate over $\bar{\boldsymbol{\theta}}$ by multiplying by 2π the integral over $(\hat{\theta}_1, \ldots, \hat{\theta}_{m-1})$ with $\bar{\boldsymbol{\theta}} = 0$.

As before, we use $\hat{\theta}_m$ as an abbreviation for $-\sum_{j=1}^{m-1} \hat{\theta}_j$ even though it is not one of the variables of integration. For $\boldsymbol{\theta} \in \mathcal{U}$, we find that the integrand of $K(m,n)$ has value

$$\exp\Bigl(n \log\bigl(1 - \tfrac{1}{2}\varrho_{t,2}(\hat{\boldsymbol{\theta}}) - \tfrac{1}{6} i \varrho_{t,3}(\hat{\boldsymbol{\theta}}) + O(1)\varrho_{t,4}(\hat{\boldsymbol{\theta}})\bigr)\Bigr) = \exp\bigl(-Q(\hat{\boldsymbol{\theta}}) + O(n^{-\epsilon})\bigr)$$

by Lemma 5(b), where

$$Q(\hat{\boldsymbol{\theta}}) = \frac{t(m-t)n}{m(m-1)} \Biggl(\sum_{j=1}^{m-1} \hat{\theta}_j^2 + \tfrac{1}{2} \sum_{\substack{j,k=1 \\ j \ne k}}^{m-1} \hat{\theta}_j \hat{\theta}_k \Biggr). \tag{5.4}$$

Since the quadratic form $Q(\hat{\boldsymbol{\theta}})$ is real, we have

$$\int_{\mathcal{U}} \exp\bigl(-Q(\hat{\boldsymbol{\theta}}) + O(n^{-\epsilon})\bigr) = \bigl(1+O(n^{-\epsilon})\bigr) \int_{\mathcal{U}} \exp\bigl(-Q(\hat{\boldsymbol{\theta}})\bigr).$$

To complete the proof of the lemma, we only need to note that the integral on the right differs from the same integral over $\mathbb{R}^m$ by a negligible amount. Apart from normalization, $\exp\bigl(-Q(\hat{\boldsymbol{\theta}})\bigr)$ is the density of an $(m-1)$-dimensional Gaussian whose covariance matrix Σ is the inverse of twice the matrix defining Q; that is,

$$\Sigma = \frac{m(m-1)}{2t(m-t)n}\bigl(2I_{m-1} - \tfrac{2}{m}J_{m-1}\bigr),$$

where I_{m-1} and J_{m-1} are the identity matrix and the matrix of all ones, respectively. The variance of $\hat{\theta}_j$ for $1 \leq j \leq m-1$ is the j-th diagonal element of Σ, while the variance of $\hat{\theta}_m = -\sum_{j=1}^{m-1} \hat{\theta}_j$ is $(1, 1, \ldots, 1)\Sigma(1, 1 \ldots, 1)^T$. These turn out to be the same, namely

$$\operatorname{Var}(\hat{\theta}_j) = \frac{(m-1)^2}{t(m-t)n}, \quad 1 \leq j \leq m.$$

Using the assumption (5.1), we find that the constraints defining $\mathcal{U}$ occur at more than $n^{\epsilon/3}$ standard deviations, so far more than the necessary fraction of $\exp\bigl(-Q(\hat{\boldsymbol{\theta}})\bigr)$ lies inside $\mathcal{U}$.

Finally, we note that the determinant of Q is

$$m\biggl(\frac{t(m-t)n}{2m(m-1)}\biggr)^{m-1}.$$

The lemma now follows. □

Lemma 7. *If $m \geq 4$ and (5.1) holds, then*

$$K(m, n) = K_{\mathcal{U}}(m, n)\bigl(1 + O(n^{-4\epsilon})\bigr).$$

Proof. Define

$$z_1 = (An)^{-1/2+\epsilon/4}, \quad z_2 = m^{3/2}\bigl(t(m-t)n\bigr)^{-1/2+\epsilon}.$$

We wish to concentrate the integral in a box of size z_1, but first we will achieve the box $\mathcal{V} \subseteq \hat{C}^m$ defined by $|\hat{\theta}_j| \leq z_2$ for $1 \leq j \leq m$. Note that $z_2 = o(1)$.

The absolute value of the integrand in (5.3) is

$$F(\boldsymbol{\theta}) = \binom{m}{t}^{-n} \Bigl(\sum_{S,S'} \cos\bigl(\Sigma_S - \Sigma_{S'}\bigr)\Bigr)^{n/2}, \tag{5.5}$$

where the sum is over all subsets S, S' of $\{1, 2, \ldots, m\}$ of cardinality t, and $\Sigma_S = \sum_{j\in S} \theta_j$.

If $\boldsymbol{\theta} \notin \mathcal{V}$, then two of the θ_j differ by at least z_2. Without loss of generality, suppose $|\theta_2 - \theta_1| > z_2$ where the difference is measured mod 2π. Let T be a subset of $\{3, 4, \ldots, m\}$ of cardinality $t-1$. Then

$$\cos\bigl(\Sigma_{T\cup\{1\}} - \Sigma_{S'}\bigr) + \cos\bigl(\Sigma_{T\cup\{2\}} - \Sigma_{S'}\bigr)$$

is maximized over S' when $\Sigma_{S'} = \Sigma_T + \frac{1}{2}(\theta_1 + \theta_2)$ or $\Sigma_{S'} = \Sigma_T + \frac{1}{2}(\theta_1 + \theta_2) + \pi$. There are $\binom{m-2}{t-1}$ choices for T, so we have that

$$\begin{aligned} F(\boldsymbol{\theta}) &< \biggl(1 - \frac{2t(m-t)}{m(m-1)}\bigl(1 - \cos(\tfrac{1}{2}z_2)\bigr)\biggr)^{n/2} \\ &< \exp\Bigl(-\tfrac{1}{9}\bigl(t(m-t)n\bigr)^{2\epsilon} m\Bigr) \end{aligned}$$

for n sufficiently large. Multiplying by the total volume, which is less than $(2\pi)^m$, we find that such $\boldsymbol{\theta}$ contribute $O(e^{n^{-\epsilon}})K_{\mathcal{U}}(m, n)$ to $K(m, n)$.

Since $\mathcal{V} \subseteq \hat{C}^m$, and the value of $F(\boldsymbol{\theta})$ is independent of $\bar{\boldsymbol{\theta}}$ in that case, we continue the investigation in $(\bar{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}})$-space with the assumption that $\bar{\boldsymbol{\theta}} = 0$. Choose an arbitrary fixed $\hat{\boldsymbol{\theta}}$ such that $\bar{\boldsymbol{\theta}} = 0$ and $\max_j |\hat{\theta}_j| = 1$, and define $f(r) = F(r\hat{\boldsymbol{\theta}})$. Using $\cos(\alpha) \le 1 - \alpha^2/2 + \alpha^4/4$, which is valid for all real α, we find

$$\begin{aligned}\binom{m}{t}^{-2} \sum_{S,S'} \cos(\Sigma_S - \Sigma_{S'}) &\le 1 - 2ar^2 + 2br^4 \\ &\le \exp\bigl(-2ar^2 + 2br^4\bigr),\end{aligned}$$

where

$$\begin{aligned} a = a(\hat{\boldsymbol{\theta}}) &= \tfrac{1}{2}\varrho_{t,2}(\hat{\boldsymbol{\theta}}) \\ 0 \le b = b(\hat{\boldsymbol{\theta}}) &= \tfrac{1}{4}\varrho_{t,4}(\hat{\boldsymbol{\theta}}) \le \frac{a^2m^2}{t(m-t)}, \end{aligned} \tag{5.6}$$

and the last inequality comes from Lemma 5(b). Referring to (5.5), after raising both sides to the $n/2$ power we conclude that

$$f(r) \le \exp\bigl(-anr^2 + bnr^4\bigr).$$

Now consider an infinitesimally small piece of solid angle $d\Theta$ in the direction of $\hat{\boldsymbol{\theta}}$. The contribution this makes to $\int F(\hat{\boldsymbol{\theta}})$ is

$$d\Theta \int_0^z r^{m-2} f(r)\, dr,$$

where $z = z_1$ for $\mathcal{U}$ and $z = z_2$ for $\mathcal{V}$. (The power of r is $m-2$ due to the constraint $\bar{\boldsymbol{\theta}} = 0$.)

Define $f_0(r) = \exp(-anr^2)$. Then

$$\int_0^\infty r^{m-2} f_0(r)\, dr = \Theta(1) \frac{a^{1/2}n^{1/2}}{m-1} \Bigl(\frac{m}{2aen}\Bigr)^{m/2},$$

and $r^{m-2}f_0(r)$ has its maximum at

$$r_0 = \sqrt{\frac{m-2}{2an}}.$$

We will bound $\int_0^{z_2} r^{m-2}\bigl(f(r) - f_0(r)\bigr)\, dr$ by breaking it into two parts at $r = r_1 = n^{\epsilon/2} r_0$.

From (5.6) we find that $bnr^4 = o(1)$, and so $\exp(bnr^4) \le 1 + 2bnr^4$, for $r \le r_1$. Thus

$$\begin{aligned}\int_0^{r_1} r^{m-2}\bigl(f(r) - f_0(r)\bigr)\, dr &\le 2bn \int_0^\infty r^{m+2} f_0(r)\, dr \\ &= O(1) \frac{bm^2}{a^2 n} \int_0^\infty r^{m-2} f_0(r)\, dr \\ &= O(n^{-4\epsilon}) \int_0^\infty r^{m-2} f_0(r)\, dr. \end{aligned} \tag{5.7}$$

For $r_1 \le r \le z_2$ we find that $r^{m-2}f(r)$ is decreasing, so

$$\begin{aligned}\int_{r_1}^{z_2} r^{m-2}f(r)\,dr &\le z_2 r_1^{m-2} f(r_1)\\ &= \exp\bigl(-\tfrac{1}{2}mn^{\epsilon} + O(m\log n)\bigr)\int_0^\infty r^{m-2}f_0(r)\,dr. \qquad (5.8)\end{aligned}$$

At this point we can notice that $f_0(r)$ is in fact the same quadratic form that was called $\exp\bigl(-Q(\hat{\boldsymbol{\theta}})\bigr)$ in Lemma 6. Bounds (5.7) and (5.8) thus imply that

$$\int_{\mathcal{V}} F(\hat{\boldsymbol{\theta}}) \le \bigl(1 + O(n^{-4\epsilon})\bigr)\int_{\mathbb{R}^{m-1}} \exp\bigl(-Q(\hat{\boldsymbol{\theta}})\bigr)$$

Since $bnr^4 = O(n^{-1/3})$ for $0 \le r \le z_1$, we also have

$$\begin{aligned}\int_{\mathcal{U}} F(\hat{\boldsymbol{\theta}}) &= \bigl(1 + O(n^{-1/3})\bigr)\int_{\mathcal{U}} \exp\bigl(-Q(\hat{\boldsymbol{\theta}})\bigr)\\ &= \bigl(1 + O(n^{-1/3})\bigr)\int_{\mathbb{R}^{m-1}} \exp\bigl(-Q(\hat{\boldsymbol{\theta}})\bigr),\end{aligned}$$

where the last step is proved in the proof of Lemma 6. It follows that

$$\int_{\mathcal{V}-\mathcal{U}} F(\hat{\boldsymbol{\theta}}) = O(n^{-4\epsilon})\int_{\mathbb{R}^{m-1}} \exp\bigl(-Q(\hat{\boldsymbol{\theta}})\bigr),$$

and the lemma now follows from Lemma 6. □

Theorem 4. *Under condition (5.1), for sufficiently small $\epsilon > 0$,*

$$B(m,s;n,t) = \binom{m}{t}^n \left(\frac{m-1}{2\pi t(m-t)n}\right)^{(m-1)/2} m^{m/2}\bigl(1 + O(n^{-\epsilon})\bigr).$$

Proof. For $m \le 3$, we can verify the claim directly using the exact value $B(m, n/m; n, 1) = n!/(n/m)!^m$. For $m \ge 4$, it follows from (5.3) and the two lemmas just proved. □

Theorem 4 can be seen as a particular Central Limit Theorem result, and this can be taken further. Define $X_1, X_2, \ldots, X_n$ to be *iid* random values taking values in $\{0,1\}^{m-1}$. The common distribution is that $X_j = (x_1, x_2, \ldots, x_{m-1})$ with probability $\binom{m}{t}^{-1}$ if $\sum x_i \in \{t-1, t\}$ and with probability 0 otherwise. The values taken by X_j can be interpretted as the first $m-1$ entries in the j-th column of an $m \times n$ 0-1 matrix with each column having sum t. From this is is clear that

$$B(m,s;n,t) = \binom{m}{t}^n \operatorname{Prob}\Bigl(\sum_{j=1}^n X_j = (s,s,\ldots,s)\Bigr). \qquad (5.9)$$

Theorem 4 now follows in the case of constant m, t from the CLT. In, fact, under the same conditions, there is an asymptotic expansion for $B(m,s;n,t)$.

Theorem 5. *Let m, t be fixed integers with $1 \leq t \leq m-1$. Then there are values $h_1(m,t), h_2(m,t), \ldots$ depending only on m, t such that*

$$B(m,s;n,t) = \binom{m}{t}^n \left(\frac{m-1}{2\pi t(m-t)n}\right)^{(m-1)/2} m^{m/2}\Bigl(1 + \sum_{i=1}^{q} h_i(m,t)n^{-i} + o(n^{-q})\Bigr).$$

for any integer $q > 0$, as $n \to \infty$ through integer multiples of m/t.

Proof. The theorem is an example of an Edgeworth expansion. We need the case of multivariable lattice distributions, such as Corollary 22.3 of [3]. Our theorem follows from that Corollary by calculation. The only additional observation required is that (in the notation of [3]) the functions $P_r(-\phi_{0,V} : \{\chi_\nu\})$ are odd in the case that r is odd, implying that we don't have terms of order $n^{-1/2}, n^{-3/2}, \ldots$. □

Computation of the values $h_j(m,t)$ is quite tedious, but we have established that

$$h_1(m,t) = \frac{m-1}{12(m-2)}\left(m + 2 + \frac{(m-1)m^2}{t(m-t)}\right). \tag{5.10}$$

This implies that the value $\Delta(m,s;n,t)$ defined in the Introduction converges to

$$-\frac{m-1}{12(m-2)} + \frac{t(m-t)(5m-2)}{24(m-2)m^2} \in (-\tfrac{1}{12}, -\tfrac{1}{32}]$$

as $n \to \infty$ with bounded m, in accordance with our conjecture.

Unfortunately, it does not appear that the existing theory of Edgeworth expansions includes error bounds explicit enough that we can increase m as n increases.

6 Exact values and estimates

In this section, we will explain how we computed the exact values of $B(m,s;n,t)$ for many values of the parameters.

It is clear that $B(m,s;n,t)$ is the constant term in

$$G(\boldsymbol{x},\boldsymbol{y}) = x_1^{-s}\cdots x_m^{-s} y_1^{-t}\cdots y_n^{-t} \prod_{j=1}^{m}\prod_{k=1}^{n}(1 + x_j y_k).$$

For small values of m and n, we can extract the constant term of G by using a method of summing over roots of unity. A technique of this nature was given by Good and Crook [9] and improved by McKay [12]. We will further improve it in this paper.

Let q_1 and q_2 be integers such that $q_1 \geq m - t + 1$ and $q_2 \geq s + 1$. Consider any field $\mathcal{F}$ which contains elements α, β of multiplicative order q_1 and q_2, respectively. Let $\langle\alpha\rangle$ be the multiplicative subgroup of $\mathcal{F}$ generated by α and define $\langle\alpha\rangle^n = \langle\alpha\rangle \times \cdots \times \langle\alpha\rangle$ (n factors). Similarly define $\langle\beta\rangle$ and $\langle\beta\rangle^m$. As explained in [12], if we sum $G(\boldsymbol{x},\boldsymbol{y})$ over $\boldsymbol{x} \in \langle\beta\rangle^m$ and $\boldsymbol{y} \in \langle\alpha\rangle^n$, the contributions of the terms of the expansion of $G(\boldsymbol{x},\boldsymbol{y})$ are zero except for those terms where each x_j has degree divisible by q_2 and each y_k has degree divisible by q_1. This includes the constant term of G, but otherwise no terms with any x_j of negative degree or any y_k of positive degree (by the constraints on q_1 and q_2). However, the total $\boldsymbol{x}$-degree of each term equals the total $\boldsymbol{y}$-degree, so the only term giving non-zero contribution is the constant term. Since the constant term is independent of $\boldsymbol{x}, \boldsymbol{y}$, we have that

$$\begin{aligned} B(m,s;n,t) &= q_2^{-m} q_1^{-n} \sum_{\boldsymbol{x}\in\langle\beta\rangle^m} \sum_{\boldsymbol{y}\in\langle\alpha\rangle^n} G(\boldsymbol{x},\boldsymbol{y}) \\ &= q_2^{-m} q_1^{-n} \sum_{\boldsymbol{x}\in\langle\beta\rangle^m} \prod_{j=1}^{m} x_j^{-s} \biggl(\sum_{y\in\langle\alpha\rangle} y^{-t} \prod_{j=1}^{m} (1+x_j y) \biggr)^{n}. \end{aligned}$$

The outside sum, which has q_2^m terms, can be computed more quickly by noting that the summand is a symmetric function in $\boldsymbol{x}$. Using m_i to denote the number of x_j's equal to β^i for $0 \le i \le q_2 - 1$, we have

$$\begin{aligned} B(m,s;n,t) &= q_2^{-m} q_1^{-n} \\ &\times \sum_{m_0+m_1+\cdots+m_{q_2-1}=m} \binom{m}{m_0, m_1, \ldots, m_{q_2-1}} \prod_{i=1}^{q_2-1} \beta^{-i m_i s} \biggl(\sum_{y\in\langle\alpha\rangle} y^{-t} \prod_{i=1}^{q_2-1} (1+\beta^i y)^{m_i} \biggr)^{n}. \end{aligned} \tag{6.1}$$

Note that (6.1) is evaluated in the field $\mathcal{F}$ and the left side is whatever the constant term of $G(\boldsymbol{x},\boldsymbol{y})$ is when it is expanded in that field. In principle we could take $\mathcal{F}$ to be the field of complex numbers, and so obtain the normal integer value of $B(m,s;n,t)$ directly, but this poses numerical difficulties. In practice it is better to take $\mathcal{F} = \mathrm{GF}(p)$ for various primes p, then the normal integer value of $B(m,s;n,t)$ can be recovered using the Chinese Remainder Theorem. The roots α and β exist so long as $p-1$ is divisible by both q_1 and q_2.

By this means we computed all values of $B(m,s;n,t)$ for $m,n \le 30$. As an example,

$$\begin{aligned} B(30,15;30,15) = 75\,&51081\,53829\,51405\,59732\,48475\,93800\,76934\,94252\,92103 \\ &89151\,81695\,05028\,07370\,84462\,72734\,38430\,42892\,52001 \\ &35264\,46320\,41706\,98298\,25720\,80514\,93000\,44864\,24346 \\ &92361\,36642\,96667\,59160\,41398\,51347\,38588\,32514\,94564 \\ &17934\,76366\,44171\,38875\,08829\,26548\,61238\,27200. \end{aligned}$$

All the values we computed are available to interested researchers on the Internet [5]. For these parameters, the accuracy of the estimates derived in this paper is excellent, as explained earlier with the statement of Conjecture 1.

Beyond the point to which exact values are readily computed, they can nevertheless be estimated to good accuracy using sampling methods. The best approach of which we are aware is due to Chen, Diaconis, Holmes and Liu [6]. We will describe the method here, since it admits of some streamlining in the case of constant column sums.

Let $\boldsymbol{p} = (p_1, p_2, \ldots, p_m)$ be a vector of real numbers in $[0,1]$. Let $X_1, X_2, \ldots, X_m$ be independent Bernoulli random variables, with $\operatorname{Prob}(X_i = 1) = p_i$ and $\operatorname{Prob}(X_i = 0) = q_i = 1 - p_i$ for each i. For an integer t, $0 \le t \le m$, the conditional random variable $(X_1, X_2, \ldots, X_m \mid \sum_i X_i = t)$ has a *conditional-Poisson* distribution $Z(\boldsymbol{p}, t)$. If $\boldsymbol{u} = (u_1, u_2, \ldots, u_m) \in \{0,1\}^m$ has $\sum_i u_i = t$, and $u_i = p_i$ whenever $p_i \in \{0,1\}$, then the probability $\operatorname{Prob}_{\boldsymbol{p},t}(\boldsymbol{u})$ of $\boldsymbol{u}$ in this distribution satisfies

$$\frac{1}{\operatorname{Prob}_{\boldsymbol{p},t}(\boldsymbol{u})} = [x^t] \prod_{i \mid u_i = 0} \Bigl(1 + \frac{p_i}{q_i} x\Bigr) \prod_{i \mid u_i = 1} \Bigl(\frac{q_i}{p_i} + x\Bigr), \tag{6.2}$$

where $[x^t]$ denotes extraction of the coefficient of x^t. This is easily proved using the probability generating function of $\sum X_i$.

We can now describe Chen et al.'s method for estimating $B(m, s; n, t)$. Define $\boldsymbol{p}^{(n)} = (s/n, s/n, \ldots, s/n)$, then, for $k = n, n-1, \ldots, 2$, compute

$$\begin{aligned} \boldsymbol{u}^{(k)} &:= \text{a uniformly random vector in } Z(\boldsymbol{p}^{(k)}, t); \\ N^{(k)} &:= 1/\operatorname{Prob}_{\boldsymbol{p}^{(k)},t}(\boldsymbol{u}^{(k)}); \\ \boldsymbol{p}^{(k-1)} &:= (k\boldsymbol{p}^{(k)} - \boldsymbol{u}^{(k)})/(k-1). \end{aligned}$$

Theorem 6. [6] *The expected value of $N^{(n)} N^{(n-1)} \cdots N^{(2)}$ is $B(m, s; n, t)$.* □

Chen et al. note that Theorem 6 does not depend on the particular distribution $Z(\boldsymbol{p}, t)$ being used, and is in fact true for any distribution having the same support. However, $Z(\boldsymbol{p}, t)$ is suggested on the grounds of statistical efficiency, since $Z(\boldsymbol{p}, t)$ is the distribution of the first column of a randomly chosen $m \times n$ matrix with row sums np_i and first column sum t. Hopefully this is similar to the distribution $Z'(\boldsymbol{p}, t)$ of the first column subject to all the column sums being t, which would be the ideal choice. In our application, with $\sum p_i = t$ and constant column sums, we have additional point of similarity that the marginal distributions of X_i for $Z(\boldsymbol{p}, t)$ and $Z'(\boldsymbol{p}, t)$ are the same, which is not always true if the column sums vary.

Note that, contrary to the more general case considered by Chen et al., the algorithm cannot get stuck in our case due to $\boldsymbol{u}^{(k)}$ not existing. We can also simplify and accelerate the implementation using the property that $\sum_{i=1}^m p^{(k)} = t$ for all k. This means that it is efficient to generate $\boldsymbol{u}^{(k)}$ by repeatedly generating $X_1, X_2, \ldots, X_m$ until their sum is t. The expected number of repetitions is $O(t^{1/2})$, giving an overall expected time less than any of the sampling methods described by Chen and Liu [7]. Clearly (6.2) can be evaluated in time $O(tm)$, so the total expected time for one estimate of $B(m, s; n, t)$ is $O\bigl(mnt\bigr)$. Of course we can combine many estimates to obtain a more accurate estimate.

As an example, a million trials gave a 99% confidence interval for $B(30, 15; 30, 15)$ of $(7.5525 \pm 0.0042) \times 10^{221}$ in about 5 minutes, comfortably enclosing the exact value given above.

If we define $\boldsymbol{u}^{(1)} = \boldsymbol{p}^{(1)}$, then the matrix with columns $\boldsymbol{u}^{(1)}, \boldsymbol{u}^{(2)}, \ldots, \boldsymbol{u}^{(n)}$ has all column sums t and row sums s. It is not a uniform sample from this class of matrices, but nevertheless the method can be extended to estimate any statistic on the class. See [6] for details.

To further test Conjecture 1, we used this method to obtain accurate estimates of $B(m, s; n, t)$ for all cases with $\max\{m, n\} \in \{50, 100\}$. In every instance, we obtained 99% confidence intervals for $\Delta(m, s; t, n)$ lying inside the interval $(-\frac{1}{12}, 0)$ mentioned at the end of the Introduction.

We also estimated a variety of values with $\max\{m, n\} = 1000$ to approximately 4 digits accuracy, in each case obtaining an answer consistent with the conjecture.

References

[1] A. Békéssy, P. Békéssy and J. Komlós, Asymptotic enumeration of regular matrices, *Studia Sci. Math. Hungar.*, **7** (1972) 343–353.

[2] E. A. Bender, The asymptotic number of nonnegative integer matrices with given row and column sums, *Discrete Math.*, **10** (1974) 345–353.

[3] R. N. Bhattacharya and R. R. Rao, Normal Approximation and Asymptotic Expansions, John Wiley & Sons (NY, 1976).

[4] B. Bollobás and B. D. McKay, The number of matchings in random regular graphs and bipartite graphs, *J. Combinatorial Theory (Series B)*, **41** (1986) 80–91.

[5] E. R. Canfield and B. D. McKay, Table of the numbers of semiregular 0-1 matrices of order $m \times n$ for $m, n \leq 30$, and a related Maple program. Available at `http://cs.anu.edu.au/~bdm/data/semiregular.html`

[6] Y. Chen, P. Diaconis, S. P. Holmes and J. S. Liu, Sequential Monte Carlo methods for statistical analysis of tables, *J. Amer. Statist. Assoc.*, **100** (2005) 109–120.

[7] S. X. Chen and J. S. Liu, Statistical Applications of the Poisson-Binomial and conditional Bernoulli distributions, *Statistica Sinica*, **7** (1997) 875–892.

[8] C. J. Everett, Jr., and P. R. Stein, The asymptotic number of integer stochastic matrices, *Discrete Math.*, **1** (1971) 33–72.

[9] I. J. Good and J. F. Crook, The enumeration of arrays and a generalization related to contingency tables, *Discrete Math.*, **19** (1977) 23–45.

[10] C. S. Greenhill, B. D. McKay and X. Wang, Asymptotic enumeration of sparse irregular bipartite graphs, *J. Combinatorial Theory, Ser. A*, to appear.

[11] S. Litsyn and A. Shpunt, On the capacity of two-dimensional weight-constrained memories, to appear in "Algebraic Coding Theory and Information Theory", DIMACS Series in Discrete Mathematics and Theoretical Computer Science, Amer. Math. Soc., 2005.

[12] B. D. McKay, Applications of a technique for labelled enumeration, *Congressus Numerantium*, **40** (1983) 207–221.

[13] B. D. McKay, Asymptotics for 0-1 matrices with prescribed line sums, *in* Enumeration and Design, (Academic Press, 1984) 225–238.

[14] B. D. McKay and X. Wang, Asymptotic enumeration of 0-1 matrices with equal row sums and equal column sums, *Linear Alg. Appl.*, **373** (2003) 273–288.

[15] B. D. McKay and N. C. Wormald, Asymptotic enumeration by degree sequence of graphs of high degree. *European J. Combinatorics*, **11** (1990) 565–580.

[16] B. D. McKay and N. C. Wormald, Asymptotic enumeration by degree sequence of graphs with degrees $o(n^{1/2})$, *Combinatorica*, **11** (1991) 369–382.

[17] M. P. Mineev and A. I. Pavlov, On the number of (0,1)-matrices with prescribed sums of rows and columns, *Doc. Akad. Nauk SSSR 230*, **2** (1976) 1276–1282.

[18] P. E. O'Neil, Asymptotics and random matrices with row-sum and column-sum restrictions, *Bull. Amer. Math. Soc.*, **75** (1969) 1276–1282.

[19] E. Ordentlich and R. M. Roth, Two-dimensional weight-constrained codes through enumeration bounds, *IEEE Trans. Inf. Theory*, **46** (2000) 1292–1301.

[20] R. C. Read, Some enumeration problems in graph theory, *Doctoral Thesis*, *London University*, (1958).