Arctic circles, domino tilings and square Young tableaux

Dan Romik

Department of Mathematics, UC Davis

October 23, 2010 (Revised version)

Abstract

The arctic circle theorem of Jockusch, Propp, and Shor asserts that uniformly random domino tilings of an Aztec diamond of high order are frozen with asymptotically high probability outside the "arctic circle" inscribed within the diamond. A similar arctic circle phenomenon has been observed in the limiting behavior of random square Young tableaux. In this paper, we show that random domino tilings of the Aztec diamond are asymptotically related to random square Young tableaux in a more refined sense that looks also at the behavior inside the arctic circle. This is done by giving a new derivation of the limiting shape of the height function of a random domino tiling of the Aztec diamond that uses the large-deviation techniques developed for the square Young tableaux problem in a previous paper by Pittel and the author. The solution of the variational problem that arises for domino tilings is almost identical to the solution for the case of square Young tableaux by Pittel and the author. The analytic techniques used to solve the variational problem provide a systematic, guess-free approach for solving problems of this type which have appeared in a number of related combinatorial probability models.

Key words and phrases: Domino tiling, Young tableau, alternating sign matrix, Aztec diamond, arctic circle, large deviations, variational problem, combinatorial probability, Hilbert transform.

ASM 2000 Subject classifications: 60C05, 60K35, 60F10.

Contents

1 Introduction

1.1 Domino tilings and the arctic circle theorem

A domino in $\mathbb{R}^2$ is a $\mathbb{Z}^2$-translate of either of the two sets $[0,1] \times [0,2]$ or $[0,2] \times [0,1]$. If $S \subset \mathbb{R}^2$ is a region comprised of a union of $\mathbb{Z}^2$-translates of $[0,1]^2$, a ***domino tiling*** of S is a representation of S as a union of dominoes with pairwise disjoint interiors. Domino tilings, or equivalently the dimer model on a square lattice, are an extensively studied and well-understood lattice model in statistical physics and combinatorics. Their rigorous analysis dates back to Kasteleyn [21] and Temperley-Fisher [31], who independently derived the formula

$$\prod_{j=1}^{m} \prod_{k=1}^{n} \left| 2\cos\left(\frac{\pi j}{m+1}\right) + 2\sqrt{-1}\cos\left(\frac{\pi k}{n+1}\right) \right|^{1/2}$$

for the number of domino tilings of an $n \times m$ rectangular region. About thirty years later, a different family of regions was found to have a much simpler formula for the number of its domino tilings: if we define the ***Aztec diamond*** of order n to be the set

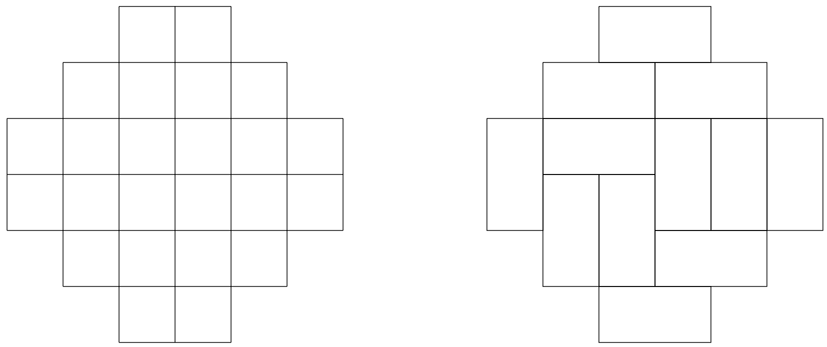
$$\mathrm{AD}_n = \bigcup_{i=-n}^{n-1} \bigcup_{j=\max(-n-i-1,-n+i)}^{\min(n+i,n-i-1)} [i,i+1] \times [j,j+1]$$

(see Figure 1), then Elkies, Kuperberg, Larsen and Propp [7] proved that AD_n has exactly

$$2^{\binom{n+1}{2}}$$

domino tilings. This can be proved by induction in several ways, but is perhaps best understood via a connection to ***alternating sign matrices***.

One of the best-known results on domino tilings is the ***arctic circle theorem*** due to Jockusch, Propp and Shor [17], which describes the asymptotic behavior of uniformly random domino tilings of the Aztec diamond. Roughly, the theorem states that the so-called ***polar regions***, which are the four contiguous regions adjacent to the four corners of the Aztec diamond in which the tiling behaves in a predictable brickwork pattern, cover a region that is approximately equal to the area that lies outside the circle inscribed in the diamond. See Figure 2, where the outline of the so-called "arctic" circle can be clearly discerned. The precise statement is the following.



(a) (b)

Figure 1: The Aztec diamond of order 3 and one of its 64 tilings by dominoes.

Theorem 1 (The arctic circle theorem [17])**.** *Fix $\epsilon > 0$. For each n, consider a uniformly random domino tiling of AD_n scaled by a factor $1/n$ in each axis to fit into the limiting diamond*

$$AD_\infty := \{|x| + |y| \le 1\},$$

and let $P_n^\circ \subset n^{-1}AD_n$ be the image of the polar regions of the random tiling under this scaling transformation. Then as $n \to \infty$ the event that

$$\left\{(x,y) \in AD_\infty : x^2 + y^2 > \frac{1}{2} + \epsilon\right\} \cap (n^{-1}\, AD_n) \\ \subset P_n^\circ \subset \left\{(x,y) \in AD_\infty : x^2 + y^2 > \frac{1}{2} - \epsilon\right\}$$

holds with probability that tends to 1.

In later work, Cohn, Elkies and Propp [2] derived more detailed asymptotic information about the behavior of random domino tilings of the Aztec diamond, that gives a quantitative description of the behavior of the tiling inside the arctic circle. They proved two main results (which are roughly equivalent, if some technicalities are ignored), concerning the ***placement probabilities*** (the probabilities to observe a given type of domino in a given

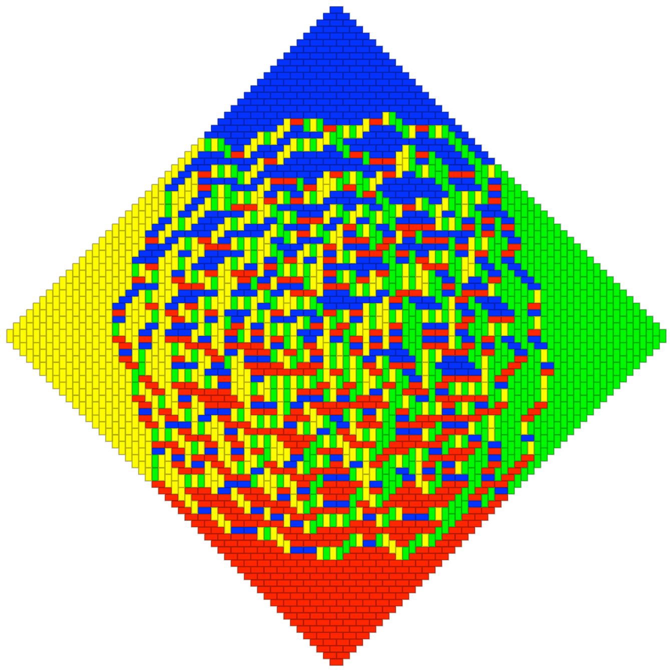


Figure 2: The arctic circle theorem: in a random domino tiling of AD_{50}, the circle-like shape is clearly visible. Here, dominoes are colored according to their type and parity.

position in the diamond) and the ***height function*** of the tiling (which, roughly speaking, encodes a weighted counting of the number of dominoes of different types encountered while travelling from a fixed place to a given position in the diamond—see Section 6 for the precise definition).

A main goal of this paper is to give a new proof of the Cohn-Elkies-Propp limit shape theorem for the height function of a uniformly random domino tiling of the Aztec diamond—see Theorem 12 in Section 6. Our proof is based on a large deviations analysis, and so gives some information that the proof in [2] (which is based on generating functions) does not: a large deviation principle for the height function. Perhaps more importantly, it highlights a surprising connection between the domino tilings model and another, seemingly unrelated, combinatorial probability model, namely that of ***random square Young tableaux***.

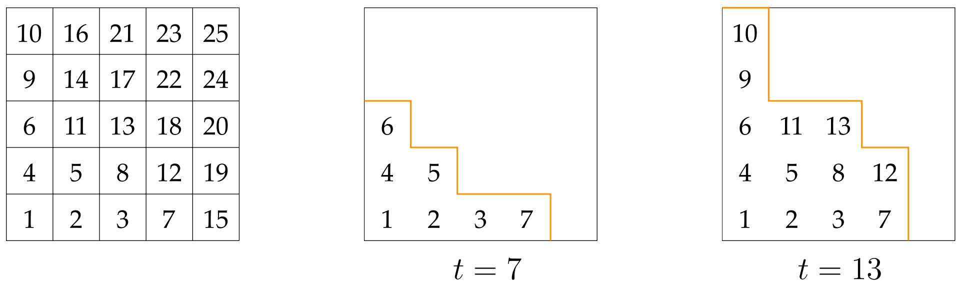


Figure 3: A square Young tableau of order 5 (shown in the "French" coordinate system), and the wall whose construction the tableau encodes at various stages of its construction.

1.2 Random square Young tableaux

Recall that a square (standard) Young tableau of order n is an array $(t_{i,j})_{i,j=1}^{n}$ of integers whose entries consist of the integers $1, 2, \ldots, n^2$, each one appearing exactly once, and such that each row and column are arranged in increasing order. One can think of a square Young tableau as encoding a sequence of instructions for constructing an $n \times n$ wall of square bricks leaning against the $x-$ and $y-$ axes by laying bricks sequentially, where the rule is that each brick can be placed only in a position which is supported from below and from the left by existing bricks or by the axes. In this interpretation, the number $t_{i,j}$ represents the time at which a brick was added in position (i, j); see Figure 3. The number of square Young tableaux of order n is known (via the hook-length formula of Frame-Thrall-Robinson) to be

$$\frac{(n^2)!}{\prod_{i,j=1}^{n}(i+j-1)}.$$

In [27], Boris Pittel and the author solved the problem of finding the ***limiting growth profile***, or ***limit shape***, of a randomly chosen square Young tableau of high order. In other words, the question is to find the growth profile of the square wall "constructed in the most random way". This can be expressed either in terms of the limit in probability $L(x, y)$ of the scaled tableau entries $n^{-2}t_{i,j}$, where $(x, y) \in [0, 1]^2$ and $i = i(n)$ and

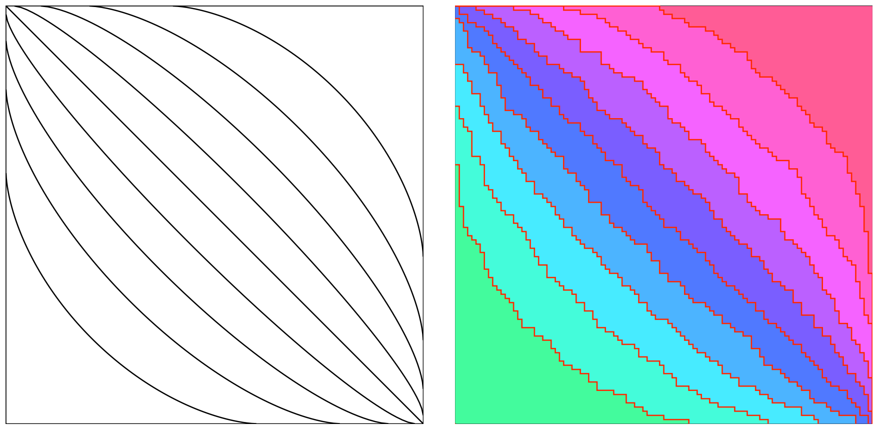


Figure 4: The limiting growth profile of a random square Young tableau and the profile of a randomly sampled tableau of order 100. The curves shown correspond to (scaled) times $t = i/10$, $i = 1, 2, \ldots, 9$.

$j = j(n)$ are some sequences such that $i/n \to x$ and $j/n \to y$ as $n \to \infty$; or alternatively in terms of the limiting shape of the family of scaled "sublevel sets" $\{n^{-1}(i,j) : t_{i,j} \leq \alpha \cdot n^2\}$ for each $\alpha \in (0,1)$ (which in the "wall-building" metaphor represents the shape of the wall at various times, and thus can be thought of as encoding the growth profile of the wall). Figure 4 shows the limiting growth profile found by Pittel and Romik and the corresponding profile of a randomly sampled square Young tableau of order 100.

For the precise definition of the limiting growth profile, see [27]. Here, we mention only the following fact which will be needed in the next subsection: If $L : [0,1] \times [0,1] \to [0,1]$ is the limit shape function mentioned above, then its values along the boundary of the square are given by

$$L(0,t) = L(t,0) = \frac{1 - \sqrt{1-t^2}}{2}, \qquad (0 \leq t \leq 1), \tag{1}$$

$$L(1,t) = L(t,1) = \frac{1 + \sqrt{t(2-t)}}{2}, \qquad (0 \leq t \leq 1). \tag{2}$$

Also note that according to the limit shape theorem, the convergence of $n^{-2} t_{i,j}$ to $L(i/n, j/n)$ as $n \to \infty$ is uniform in i and j (this follows easily from monotonicity considerations).

1.3 An arctic circle theorem for square Young tableaux

While it is not immediately apparent from the description of this limit shape result, it follows as a simple corollary of it that random square Young tableaux also exhibit an "arctic circle"-type phenomenon. That is, there is an equivalent way of visualizing the random tableau in which a spatial phase transition can be seen occurring along a circular boundary, where outside the circle the behavior is asymptotically deterministic (the "frozen" phase) and inside the circle the behavior is essentially random (the "disordered" or "temperate" phase). This fact, overlooked at the time of publication of the paper [27], was observed shortly afterwards by Benedek Valkó [32]. In fact, deducing the arctic circle result is easy and requires only the facts (1), (2) mentioned above, which contain only a small part of the information of the limit shape.

To see how the arctic circle appears, we consider a different encoding of the information contained in the tableau via a system of particles on the integer lattice $\mathbb{Z}$. In this encoding we have n particles numbered $1, 2, \ldots, n$, where initially, each particle with index k is in position k. The particles are constrained to remain in the interval $[1, 2n]$. At discrete time steps, particles jump one step to the right, provided that the space to their right is empty (and provided that they do not leave the interval $[1, 2n]$). At each time step, exactly one particle jumps.

It is easy to see that after exactly n^2 steps, the system will terminate when it reaches the state in which each particle k is in position $n + k$, and no further jumps can take place. We call the instructions for evolving the system of particles from start to finish a ***jump sequence***. We can now add a probabilistic element to this combinatorial model by considering the uniform probability measure on the set of all jump sequences of order n, and name the resulting probability model the ***jump process*** of order n. But in fact, this is nothing more than a thinly disguised version of the random square Young tableaux model, since jump sequences are in a simple bijection with square Young tableaux: given a square tableau, think of the sequence of numbers in row k of the tableau as representing the sequence of times during which particle $n + 1 - k$ jumps to the right. This is illustrated in Fig. 5. We leave to the reader the easy verification that this gives the desired bijection.

With these definitions, it is now natural to consider the asymptotic behavior of this system of particles as $n \to \infty$. Figure 6 shows the result for a simulated system with $n = 40$. Here we see a circle-like shape appearing

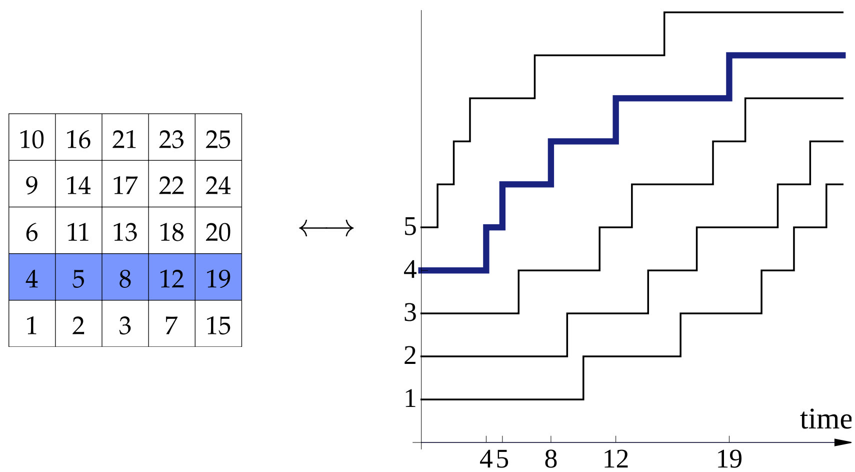


Figure 5: The bijection between square Young tableaux and jump sequences: Each row in the tableau encodes the sequence of times at which a given particle jumps. As an example, the highlighted trajectory on the right-hand side corresponds to the highlighted row on the left-hand side.

again. To formulate precisely what is happening, given a jump process of order n, for each $1 \leq k \leq 2n$, let $\tau_n^-(k)$ and $\tau_n^+(k)$ denote respectively the first and last times at which a particle k jumped from or to position k. Define the ***frozen time-period*** in position k to be the union of the two intervals

$$[0, \tau_n^-(k)] \cup [\tau_n^+(k), n^2].$$

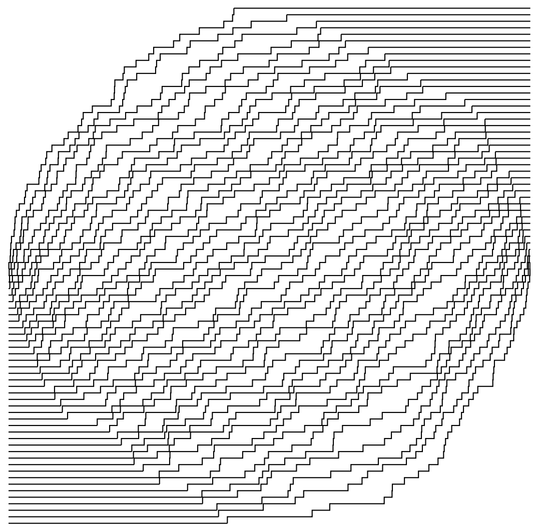


Figure 6: A jump process with 40 particles.

Theorem 2 (The arctic circle theorem for random square Young tableaux)**.** *Fix any* $\epsilon > 0$*. Denote*

$$\varphi_{\pm}(x) = \frac{1}{2} \pm \sqrt{x(1-x)}.$$

As $n \to \infty$*, the event*

$$\left\{ \max_{1 \le k \le 2n} \left| n^{-2} \tau_n^-(k) - \varphi_-\left(k/2n\right) \right| < \epsilon \right\} \cap \left\{ \max_{1 \le k \le 2n} \left| n^{-2} \tau_n^+(k) - \varphi_+\left(k/2n\right) \right| < \epsilon \right\}$$

holds with probability that tends to 1*. In other words, if the space-time diagram of the trajectories in a random jump process is mapped to the unit square* $[0,1] \times [0,1]$ *by scaling the time axis by a factor* $1/n^2$ *and scaling the position axis by a factor of* $1/2n$*, then for large* n *the frozen time-periods will occupy approximately the part of the space-time diagram that lies in the complement of the disc*

$$\left\{ (x,y) \in \mathbb{R}^2 : (x-1/2)^2 + (y-1/2)^2 \le 1/2 \right\}$$

inscribed in the square.

Proof. First, note the following simple observations that express the times $\tau_n^-(k)$ and $\tau_n^+(k)$ in terms of the Young tableau $(t_{i,j})_{i,j=1}^n$:

(i) For $1 \le k \le n$ we have $\tau_n^-(k) = t_{n+1-k,1}$.

(ii) For $n+1 \le k \le 2n$ we have $\tau_n^-(k) = t_{1,k-n}$.

(iii) For $1 \le k \le n$ we have $\tau_n^+(k) = t_{n,k}$.

(iv) For $n+1 \le k \le 2n$ we have $\tau_n^+(k) = t_{2n+1-k,n}$.

For example, the first statement is based on the fact that when $1 \le k \le n$, the time $\tau_n^-(k)$ is simply the first time at which the particle starting at position k (which corresponds to row $n+1-k$ in the tableau) jumps. The three remaining cases are equally simple and may be easily verified by the reader.

Combining these observations with (1) and (2) and the limit shape theorem, we now see that after scaling the times $\tau_n^-(k)$ and $\tau_n^+(k)$ by a factor of n^{-2}, we get quantities that converge in the limit, uniformly in k, to values determined by the appropriate substitution of boundary values in the limit shape function $L(x,y)$. For example, to deal with case (i) above, when $1 \le k \le n$, using (1) we have that

$$\begin{aligned} n^{-2}\tau_n^-(k) &= n^{-2} t_{n+1-k,1} \approx L\left(0, 1 - \frac{k-1}{n}\right) \\ &= \frac{1 - \sqrt{1 - \left(1 - \frac{k-1}{n}\right)^2}}{2} = \frac{1 - \sqrt{\frac{k-1}{n}\left(1 - \frac{k-1}{n}\right)}}{2} \\ &= \varphi_-\left(\frac{k-1}{2n}\right) \approx \varphi_-(k/2n), \end{aligned}$$

uniformly in $1 \le k \le n$. Similarly, the other three cases each imply that $n^{-2}\tau_n^{\pm}(k)$ is uniformly close to $\varphi_{\pm}(k/2n)$ in the appropriate range of values of k; we omit the details. Combining these four cases gives exactly that the event in Theorem 2 holds with asymptotically high probability as $n \to \infty$. $\square$

1.4 Similarity of the models and the analytic technique

Apart from giving a new proof of the limit shape theorem of Cohn, Elkies and Propp, another main goal of this paper is to show that the two models described in the preceding sections (random domino tilings of the Aztec diamond and random square Young tableaux) exhibit similar behavior on a more detailed level than that of the mere appearance of the arctic circle, and that in fact they are almost equivalent in an asymptotic sense. Our new proof of the limit shape theorem for the height function will use the same techniques developed in [27] for the case of random square Young tableaux: we first derive a large deviations principle, not for domino tilings but for a related model of random ***alternating sign matrices***, then solve the resulting problem in the calculus of variations using an analysis that parallels, to a remarkable (and, in our opinion, rather surprising) level of similarity, the analysis of the variational problem in [27]. The resulting formulas for the solution of the variational problem are almost identical to the formulas for the limiting growth profile of random square Young tableaux. Up to some trivial scaling factors related to the choice of coordinate system, the formulas for the two limit shapes can be written in such a way that the only difference between them is a single minus sign.

Another important aspect of our results lies not in the results themselves but in the techniques used. We use the methods first presented in [27] to solve another variational problem belonging to a class of problems previously thought to be difficult to analyze, due to a lack of a systematic framework that enables one to derive the solution in a relatively mechanical way (as opposed to having to guess it using some deep analytic insight) and then rigorously verify its claimed extremal properties. This justifies to some extent the claim from [27] that the analytic techniques of that paper provide a systematic approach for dealing with such problems, which seem to appear frequently in the analysis of combinatorial probability models (see [4, 24, 27, 33, 34]), and are also strongly related to classical variational problems arising in electrostatics and in random matrix theory.

The rest of the paper is organized as follows. In Section 2 we recall some facts about alternating sign matrices, and study the problem of finding the limiting height matrix of an alternating sign matrix chosen randomly according to ***domino measure***, which is a natural (non-uniform) probability measure on the set of alternating sign matrices of order n. In Section 3 we derive a large deviation principle for this model. This problem is solved in Section 4. In Section 5 we prove a limiting shape theorem for the height matrix of an alternating sign matrix chosen according to domino measure. In Section 6 we deduce from the previous results the Cohn-Elkies-Propp limiting shape theorem for the height function of uniformly random domino tilings of

$$
\begin{pmatrix}
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & -1 & 0 & 1 & 0 \\
1 & -1 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 & -1 & 1 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0
\end{pmatrix}
\qquad
\begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 1 & 1 \\
0 & 0 & 1 & 1 & 1 & 2 & 2 \\
0 & 1 & 1 & 1 & 2 & 3 & 3 \\
0 & 1 & 2 & 2 & 3 & 3 & 4 \\
0 & 1 & 2 & 3 & 4 & 4 & 5 \\
0 & 1 & 2 & 3 & 4 & 5 & 6
\end{pmatrix}
$$

(a) (b)

Figure 7: (a) An ASM of order 6; (b) its height matrix.

the Aztec diamond. Section 7 has some final remarks, including a discussion on the potential applicability of our methods to attack the well-known open problem of the limit shape of uniformly random alternating sign matrices.

2 Alternating sign matrices

An ***alternating sign matrix*** (often abbreviated as ***ASM)*** of order n is an $n \times n$ matrix with entries in $\{0, -1, 1\}$ such that in every row and every column, the sum of the entries is 1 and the non-zero numbers appear with alternating signs. See Fig. 7(a) for an example. Alternating sign matrices were first defined and studied in the early 1980's by David Robbins and Howard Rumsey in connection with their study [30] of Charles Dodgson's condensation method for computing determinants and of the λ-***determinant***, a natural generalization of the determinant that arises from the condensation algorithm. Later, Robbins, Rumsey and William Mills published several intriguing theorems and conjectures about them [25], tying them to the study of plane partitions and leading to many later interesting developments, some of which are described, e.g., in [1, 29].

Denote by $\mathcal{A}_n$ the set of ASM's of order n. For a matrix $M \in \mathcal{A}_n$, denote by $N_+(M)$ the number of its entries equal to 1. An important formula proved by Mills, Robbins and Rumsey states that

$$\sum_{M \in \mathcal{A}_n} 2^{N_+(M)} = 2^{\binom{n+1}{2}}. \tag{3}$$

This is sometimes referred to as the "2-enumeration" of ASM's. The reader may note that the right-hand side is equal to the number of domino tilings of

AD_n mentioned at the beginning of the introduction; indeed, a combinatorial explanation for (3) in terms of domino tilings was found by Elkies, Kuperberg, Larsen and Propp [7]. In Section 6 we will say more about this connection and how to make use of it, but for now, we rewrite (3) more probabilistically as

$$2^{-\binom{n+1}{2}} \sum_{M \in \mathcal{A}_n} 2^{N_+(M)} = 1,$$

and consider this as the basis for defining a probability measure on $\mathcal{A}_n$, which we call ***domino measure*** (thus named since it is closely related to the uniform measure on domino tilings of AD_n; see Section 6), given by the expression

$$\mathbb{P}^n_{\mathrm{DOM}}(M) = 2^{N_+(M) - \binom{n+1}{2}}, \qquad (M \in \mathcal{A}_n).$$

Our first goal will be to study the asymptotic behavior of large random ASM's chosen according to domino measure, and specifically the limit shape of their ***height matrix***. The height matrix of an ASM $M = (m_{i,j})_{i,j=1}^n \in \mathcal{A}_n$ is defined to be the new matrix $H(M) = (h_{i,j})_{i,j=0}^n$ of order $(n+1) \times (n+1)$ whose entries are given by

$$h_{i,j} = \sum_{p \le i} \sum_{q \le j} m_{p,q}.$$

The matrix $H(M)$ is also sometimes referred to as the ***corner sum matrix*** of M. It satisfies the following conditions:

$$h_{0,k} = h_{k,0} = 0 \qquad \text{for all } 0 \le k \le n, \tag{H1}$$

$$h_{n,k} = h_{k,n} = k \qquad \text{for all } 0 \le k \le n, \tag{H2}$$

$$0 \le h_{i+1,j} - h_{i,j}, h_{j,i+1} - h_{j,i} \le 1 \quad \text{for all } 0 \le i < n, 0 \le j \le n. \tag{H3}$$

See Fig. 7(b) for an example. (In fact, it is not too difficult to see that the correspondence $M \to H(M)$ defines a bijection between the set of ASM's of order n and the set of matrices satisfying conditions (H1)–(H3)—see [30, Lemma 1]—but we will not need this fact here). In particular, the "Lipschitz"-type condition (H3) means that the height matrix can be thought of as a discrete version of a two-dimensional surface, and is therefore a natural candidate for which to try and prove a limit shape result.

The basis for our analysis of $\mathbb{P}^n_{\mathrm{DOM}}$-random ASM's is a formula which will give the probability distribution (under the measure $\mathbb{P}^n_{\mathrm{DOM}}$) of the k-th row of the height matrix, for each $1 \le k \le n$. To describe this, first, as usual, denote the Vandermonde function by

$$\Delta(u_1, \dots, u_m) = \prod_{1 \le i < j \le m} (u_j - u_i).$$

Second, for an ASM $M \in \mathcal{A}_n$ and some $1 \le k \le n$, let $X_k(1) < X_k(2) < \dots < X_k(k)$ be the unique ***ascents*** of the k-th row of the height matrix $H(M)$, namely those column indices such that

$$h_{k,X_k(i)} - h_{k,X_k(i)-1} = 1, \quad (i = 1, 2, \dots, k).$$

Note that the conditions (H1)–(H3) guarantee that the ascents exist, that there are exactly k of them, and that the original k-th row of $H(M)$ can be recovered from them.

Theorem 3. *If integers $1 \le x_1 < x_2 < \dots < x_k \le n$ are given, and if $y_1 < y_2 < \dots < y_{n-k}$ are the numbers in $\{1, 2, \dots, n\} \setminus \{x_1, \dots, x_k\}$ arranged in increasing order, then, in the notation above, we have*

$$\mathbb{P}^n_{\mathrm{DOM}}\Big[M \in \mathcal{A}_n : (X_k(1), \dots, X_k(k)) = (x_1, \dots, x_k) \Big]$$
$$= \frac{2^{\binom{k+1}{2}} 2^{\binom{n-k+1}{2}}}{2^{\binom{n+1}{2}}} \cdot \frac{\Delta(x_1, \dots, x_k) \Delta(y_1, \dots, y_{n-k})}{\Delta(1, 2, \dots, k) \Delta(1, 2, \dots, n-k)}. \quad (4)$$

To prove Theorem 3, we use another well-known combinatorial bijection relating ASM's to ***monotone triangles***. A monotone triangle of order n is a triangular array $(t_{i,j})_{1 \le i \le n, 1 \le j \le i}$ of integers satisfying the inequalities

$$t_{i,j} < t_{i,j+1}, \quad t_{i,j} \le t_{i-1,j} \le t_{i,j+1} \qquad (2 \le i \le n, 1 \le j \le i-1).$$

A ***complete*** monotone triangle of order n is a monotone triangle whose bottom row consists of the numbers $(1, 2, \dots, n)$. It is well-known that alternating sign matrices of order n are in bijection with complete monotone triangles of order n. In our terminology, the bijection assigns to an ASM $M = (m_{i,j})_{i,j=1}^n$ the monotone triangle

$$T = (t_{i,j})_{1 \le i \le n, 1 \le j \le i} = \varphi_{\mathrm{ASM} \to \mathrm{CMT}}(M)$$

whose k-th row $(t_{k,j})_{1 \le j \le k}$ consists for each $1 \le k \le n$ of the ascents of the k-th row of the height matrix $H(M)$, arranged in increasing order. See

$$
\begin{array}{ccccccccccc}
 & & & & & 3 & & & & & \\
 & & & & 2 & & 5 & & & & \\
 & & & 1 & & 4 & & 5 & & & \\
 & & 1 & & 2 & & 4 & & 6 & & \\
 & 1 & & 2 & & 3 & & 4 & & 6 & \\
1 & & 2 & & 3 & & 4 & & 5 & & 6
\end{array}
\qquad\qquad
\begin{array}{ccccccccccc}
1 & & 2 & & 3 & & 4 & & 5 & & 6 \\
 & 1 & & 2 & & 4 & & 5 & & 6 & \\
 & & 1 & & 3 & & 4 & & 6 & & \\
 & & & 2 & & 3 & & 6 & & & \\
 & & & & 3 & & 5 & & & & \\
 & & & & & 5 & & & & &
\end{array}
$$

(a) (b)

Figure 8: (a) The complete monotone triangle corresponding to the ASM in Figure 7; (b) its dual, shown "standing on its head".

Figure 8(a) for an example. More explicitly, it is easy to check that this means that an index j will be present in the k-th row of T if and only if

$$\sum_{i=1}^{k} m_{i,j} = 1$$

holds.

Another notion that will prove useful is that of the ***dual*** of a complete monotone triangle. If T is a complete monotone triangle of order n, and M is the ASM in $\mathcal{A}_n$ such that $T = \varphi_{\mathrm{ASM}\to\mathrm{CMT}}(M)$, then the dual T^* of T is the complete monotone triangle of order n that corresponds via the same bijection to the matrix W, defined as the vertical reflection of M, i.e., the matrix such that $w_{i,j} = m_{n+1-i,j}$ for all i, j (clearly it, too, is an ASM). See Figure 8(b), where the dual triangle is drawn reflected vertically.

The following simple observation describes more explicitly the connection between a monotone triangle and its dual.

Lemma 4. *If $T = (t_{i,j})_{1\le i\le n, 1\le j\le i}$ is a complete monotone triangle of order n, then for each $1 \le k \le n-1$, the $(n-k)$-th row of the dual triangle T^* consists of the numbers in the complement*

$$\{1, 2, \ldots, n\} \setminus \{t_{k,1}, t_{k,2}, \ldots, t_{k,k}\}$$

of the k-th row of T, arranged in increasing order.

Proof. Let $M = (m_{i,j})_{i,j} \in \mathcal{A}_n$ be such that $T = \varphi_{\mathrm{ASM}\to\mathrm{CMT}}(M)$. As mentioned above, $1 \le j \le n$ appears in the k-th row of T if and only if

$\sum_{i=1}^{k} m_{i,j} = 1$. Similarly, from the definition of T^* we see that j appears in the $(n-k)$-th row of T^* if and only if $\sum_{i=k+1}^{n} m_{i,j} = 1$. But from the definition of an alternating sign matrix, one and only one of these conditions must hold. $\square$

As the last step in the preparation for proving Theorem 3, we note that if $M \in \mathcal{A}_n$ and $T = \varphi_{\text{ASM}\to\text{CMT}}(M)$, then it is easy to see that $N_+(M)$, the number of $+1$ entries in M, can be expressed in terms of T as the number of entries $t_{i,j}$ in T that do not appear in the preceding row (including, vacuously, the singleton element in the top row). We denote this quantity also by $N_+(T)$; note that it is defined more generally also for non-complete monotone triangles. We furthermore recall the following formula proved by Mills, Robbins and Rumsey in [25, Th. 2] (see also [7, Eq. (7), Section 4], and see [12] for a recent alternative proof and some generalizations):

Lemma 5. *If $k \geq 1$ and $x_1 < x_2 < \ldots < x_k$ are integers, then the sum of $2^{N_+(T)}$ over all monotone triangles T of order k with bottom row $(x_1, \ldots, x_k)$ is equal to*

$$2^{\binom{k+1}{2}} \prod_{1 \leq i < j \leq k} \frac{x_j - x_i}{j - i}.$$

Proof of Theorem 3. Denote by $\mathcal{T}_n(x_1, \ldots, x_k)$ the set of complete monotone triangles of order n whose k-th row is equal to $(x_1, \ldots, x_k)$. From the remarks above, it follows that the left-hand side of (4) is equal to

$$2^{-\binom{n+1}{2}} \sum_{T \in \mathcal{T}_n(x_1, \ldots, x_k)} 2^{N_+(T)}.$$

In addition, for a monotone triangle $T \in \mathcal{T}_n(x_1, \ldots, x_k)$, define T_{top} and T_{bottom} as the two monotone triangles, of orders k and $n-k$, respectively, where T_{top} is comprised of the top k rows of T, and T_{bottom} is comprised of the top $n-k$ rows of the dual triangle T^*. From Lemma 4, it follows that the correspondence

$$T \to (T_{\text{top}}, T_{\text{bottom}})$$

defines a bijection between $\mathcal{T}_n(x_1, \ldots, x_k)$ and the cartesian product $\mathcal{A} \times \mathcal{B}$, where $\mathcal{A}$ is the set of monotone triangles with bottom row $(x_1, \ldots, x_k)$ and $\mathcal{B}$ is the set of monotone triangles with bottom row $(y_1, \ldots, y_{n-k})$ (in the notation of Theorem 3). This corresponence furthermore has the property that

$$N_+(T) = N_+(T_{\text{top}}) + N_+(T_{\text{bottom}})$$

(since $N_+(T_{\text{top}})$ counts the number of $+1$ entries in the first k rows of the ASM corresponding to T, whereas $N_+(T_{\text{bottom}})$ counts the number $+1$'s in the last $n-k$ rows), or equivalently that $2^{N_+(T)} = 2^{N_+(T_{\text{top}})}2^{N_+(T_{\text{bottom}})}$. Combining these last observations, we get that the left-hand side of (4) is equal to

$$2^{-\binom{n+1}{2}} \sum_{T_{\text{top}}\in\mathcal{A}} 2^{N_+(T_{\text{top}})} \sum_{T_{\text{bottom}}\in\mathcal{B}} 2^{N_+(T_{\text{bottom}})},$$

which by Lemma 5 is equal exactly to the right-hand side of (4). □

We remark that an equivalent version of Theorem 3, phrased in the language of domino tilings and certain so-called zig-zag paths defined in terms of them, is proved by Johansson in [18] (see Proposition 5.14 in that paper and eq. (5.16) following it). See also the subsequent papers [19, 20] where Johansson proves many interesting results about random domino tilings of the Aztec diamond by combining a variant of (4) with ideas from the theory of orthogonal polynomials and the theory of determinantal point processes.

3 A large deviation principle

We now turn from combinatorics to analysis, with the goal in mind being to use Theorem 3 as the starting point for a large deviation analysis of the behavior of $\mathbb{P}^n_{\text{DOM}}$-random ASM's. First, we define the space of functions on which our analysis takes place. Fix $0 < y < 1$. We wish to understand the behavior of the k-th row of the height matrix of a $\mathbb{P}^n_{\text{DOM}}$-random ASM of order n for values of k satisfying $k \approx y \cdot n$, when n is large.

Define the space of ***y-admissible*** functions to be the set

$$\mathcal{F}_y = \Big\{ f : [0,1] \to [0,1] : f \text{ is monotone nondecreasing, 1-Lipschitz,} \\ \text{and satisfies } f(0) = 0, f(1) = y \Big\}.$$

Define the space of ***admissible functions*** as the union of all the y-admissible function spaces:

$$\mathcal{F} = \bigcup_{y\in[0,1]} \mathcal{F}_y.$$

$$\mathbf{u} = (0, 0, 1, 2, 2, 2, 3) \quad \longleftrightarrow$$

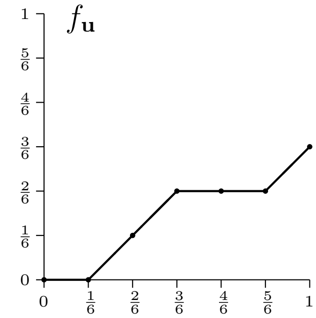


Figure 9: A $(6, 3)$-admissible sequence $\mathbf{u}$ and the corresponding function $f_{\mathbf{u}}$.

We also define a discrete analogue of the admissible functions. Given integers $0 \leq k \leq n$, a sequence $\mathbf{u} = (u_0, u_1, \ldots, u_n)$ of integers is called an (n, k)***-admissible sequence*** if it satisfies

$$u_0 = 0, \quad u_n = k, \quad \text{and } u_{i+1} - u_i \in \{0, 1\} \text{ for all } 0 \leq i \leq n - 1,$$

Note that (n, k)-admissible sequences are exactly those that can appear as the k-th row of a height matrix $H(M)$ of an ASM $M \in \mathcal{A}_n$. We embed the (n, k)-admissible sequences in the space $\mathcal{F}_y$ for $y = k/n$, in the following way: For each (n, k)-admissible sequence $\mathbf{u}$, define a function $f_{\mathbf{u}} : [0, 1] \to [0, 1]$ as the unique function having the values

$$f_{\mathbf{u}}(j/n) = u_j/n, \qquad 0 \leq j \leq n,$$

and on each interval $[j/n, (j + 1)/n]$ for $0 \leq j \leq n - 1$ being defined as the linear interpolation of the values on the endpoints of the interval; see Figure 9. Clearly, $f_{\mathbf{u}}$ is a (k/n)-admissible function. In fact, it is easy to see that the admissible functions are precisely the limits of such functions in the uniform norm topology.

With these definitions, we can now formulate the large deviation principle.

Theorem 6 (Large deviation principle for $\mathbb{P}^n_{\mathrm{DOM}}$-random ASM's)**.** *Let* $0 \le k \le n$*, and let* $\boldsymbol{u} = (u_0, u_1, \dots, u_n)$ *be an* (n,k)*-admissible sequence. Let* $H(M)_k$ *denote the* k*-th row of a height matrix* $H(M)$*. Then*

$$\mathbb{P}^n_{\mathrm{DOM}}\Big[M \in \mathcal{A}_n : H(M)_k = \boldsymbol{u}\Big] \\ = \exp\Big(-(1+o(1))n^2(I(f_{\boldsymbol{u}}) + \theta(k/n))\Big), \quad (5)$$

where we define

$$\begin{aligned} \theta(y) &= \frac{1}{2}y^2 \log y + \frac{1}{2}(1-y)^2 \log(1-y) + \frac{2\log 2 - 3}{2} y(1-y) + \frac{3}{2}, \\ I(f) &= -\int_0^1 \int_0^1 \log|s-t| f'(s)(f'(t)-1)\, ds\, dt, \qquad (f \in \mathcal{F}). \end{aligned}$$

The $o(1)$ *error term in* (5) *is uniform over all* $0 \le k \le n$ *and all* (n,k)*-admissible sequences* $\boldsymbol{u}$*, as* $n \to \infty$*.*

Proof. Let $1 \le x_1 < x_2 < \ldots < x_k \le n$ be the positions of the k ascents in the sequence $(u_0, u_1, \dots, u_n)$ (in the same sense defined before, namely that $u_{x_i} - u_{x_{i-1}} = 1$), and let $1 \le y_1 < \ldots < y_{n-k} \le n$ be the numbers in the complement $\{1, \dots, n\} \setminus \{x_1, \dots, x_k\}$ arranged in increasing order.

By (4), we have

$$\begin{aligned} & n^{-2} \log \mathbb{P}^n_{\mathrm{DOM}}\Big[M \in \mathcal{A}_n : H(M)_k = \mathbf{u}\Big] \\ &= n^{-2}\left(\binom{k+1}{2} + \binom{n-k+1}{2} - \binom{n+1}{2}\right)\log 2 \\ &\quad -n^{-2} \sum_{1 \le i < j \le k} \log(j-i) - n^{-2} \sum_{1 \le i < j \le n-k} \log(j-i) \\ &\quad +n^{-2} \sum_{1 \le i < j \le k} \log(x_j - x_i) + n^{-2} \sum_{1 \le i < j \le n-k} \log(y_j - y_i). \qquad (6) \end{aligned}$$

We estimate each of the summands. First, we have

$$\begin{aligned} n^{-2}&\left(\binom{k+1}{2}+\binom{n-k+1}{2}-\binom{n+1}{2}\right)\log 2 \\ &= \frac{\log 2}{2}\left(\frac{k}{n}\right)^2+\frac{\log 2}{2}\left(1-\frac{k}{n}\right)^2-\frac{\log 2}{2}+o(1) \\ &= -\log 2\cdot\frac{k}{n}\left(1-\frac{k}{n}\right)+o(1). \qquad (7)\end{aligned}$$

Secondly, the sum $n^{-2}\sum_{1\le i<j\le k}\log(j-i)$ can be rewritten as

$$\begin{aligned} n^{-2}\sum_{1\le i<j\le k}\log(j-i) &= n^{-2}\sum_{d=1}^{k-1}(k-d)\log d \\ &= n^{-2}\sum_{d=1}^{k-1}(k-d)\log k+\left(\frac{k}{n}\right)^2\sum_{d=1}^{k-1}\left(1-\frac{d}{k}\right)\log\frac{d}{k}\cdot\frac{1}{k} \\ &= \frac{k(k-1)}{2n^2}\log k+\left(\frac{k}{n}\right)^2\int_0^1(1-t)\log t\,dt+o(1) \\ &= \frac{1}{2}\left(\frac{k}{n}\right)^2\log k-\frac{3}{4}\left(\frac{k}{n}\right)^2+o(1), \qquad (8)\end{aligned}$$

where the error term $o(1)$ is uniform in k as $n\to\infty$ (the estimate for this sum is essentially the leading-order asymptotic expansion for the so-called ***Barnes G-function***, related also to the ***hyperfactorial***; for more detailed asymptotics of these special functions, see [9, Sec. 2.15, p. 135]). Similarly, replacing k by $n-k$ we get that

$$n^{-2}\sum_{1\le i<j\le n-k}\log(j-i)=\frac{1}{2}\left(1-\frac{k}{n}\right)^2\log(n-k)-\frac{3}{4}\left(1-\frac{k}{n}\right)^2+o(1). \quad (9)$$

Finally, we estimate the terms in (6) that depend directly on the sequence $\mathbf{u}$. The idea is to replace each term $n^{-2}\log(x_j-x_i)$ with an integral of the form $\iint\log(t-s)f'_{\mathbf{u}}(t)f'_{\mathbf{u}}(s)\,ds\,dt$ over a certain region. Observe that for $X>1$ we have the (easily verifiable) identity

$$\begin{aligned}\int_0^1\int_X^{X+1}&\log(v-u)\,dv\,du \\ &=\log X+\left(\frac{1}{2}(X^2+1)\log\left(\frac{X^2-1}{X^2}\right)+X\log\left(\frac{X+1}{X-1}\right)-\frac{3}{2}\right).\end{aligned}$$

When X is large, this behaves like $\log X + O\left(\frac{1}{X}\right)$. The integral is also defined and finite when $X = 1$. So we can write

$$\begin{aligned}
&n^{-2} \sum_{1 \le i < j \le k} \log(x_j - x_i) \\
&= n^{-2} \sum_{1 \le i < j \le k} \int_{x_i - 1}^{x_i} \int_{x_j - 1}^{x_j} \log(v-u)\, dv\, du + O\left(\sum_{1 \le i < j \le k} \frac{n^{-2}}{x_j - x_i} \right) \\
&= \sum_{1 \le i < j \le k} \int_{x_i - 1}^{x_i} \int_{x_j - 1}^{x_j} \log(v-u) \frac{dv\, du}{n^2} + O\left(\frac{\log n}{n}\right). \\
&= \sum_{1 \le i < j \le k} \int_{x_i - 1}^{x_i} \int_{x_j - 1}^{x_j} \log\left(\frac{v-u}{n}\right) \frac{dv\, du}{n^2} + \log n \cdot \frac{k(k-1)}{2n^2} \\
&\quad + O\left(\frac{\log n}{n}\right).
\end{aligned}$$

Now observe that $f'_{\mathbf{u}}(x)$ is equal to 1 if $(x_i - 1)/n < x < x_i/n$ for some i, or to 0 otherwise; so this last expression can be rewritten as

$$\iint_{R_n} \log(t-s) f'_{\mathbf{u}}(s) f'_{\mathbf{u}}(t)\, ds\, dt + \frac{1}{2}\left(\frac{k}{n}\right)^2 \log n + O\left(\frac{\log n}{n}\right), \tag{10}$$

where the integral is over the region

$$R_n = \bigcup_{1 \le i < j \le n} \left[\frac{i-1}{n}, \frac{i}{n}\right] \times \left[\frac{j-1}{n}, \frac{j}{n}\right].$$

The region of integration in (10) can be replaced with the slightly larger region

$$R = \{(s,t) \in [0,1] \times [0,1] : s < t\},$$

at the cost of an additional error which can be bounded in absolute value by

$$\int_0^1 dy \int_{y-1/n}^{y} |\log(y-x)|\, dx = \left| \int_0^{1/n} \log t dt \right| = O\left(\frac{\log n}{n}\right).$$

To summarize, after this change in the region of integration and, in addition, after symmetrizing the region of integration for convenience, we have shown that

$$
\begin{aligned}
&n^{-2} \sum_{1\le i<j\le k} \log(x_j - x_i) \\
&\quad = \frac{1}{2}\int_0^1 \int_0^1 \log|t-s| f_{\mathbf{u}}'(s) f_{\mathbf{u}}'(t)\, ds\, dt + \frac{1}{2}\left(\frac{k}{n}\right)^2 \log n + O\left(\frac{\log n}{n}\right). \quad (11)
\end{aligned}
$$

Symmetrically, following exactly the same reasoning for the last sum in (6) we get the similar estimate

$$
\begin{aligned}
&n^{-2} \sum_{1\le i<j\le n-k} \log(y_j - y_i) \\
&\quad = \frac{1}{2}\int_0^1 \int_0^1 \log|t-s| (1-f_{\mathbf{u}}'(s))(1-f_{\mathbf{u}}'(t))\, ds\, dt + \frac{1}{2}\left(1-\frac{k}{n}\right)^2 \log n \\
&\qquad + O\left(\frac{\log n}{n}\right). \quad (12)
\end{aligned}
$$

It remains to plug the estimates (7), (8), (9), (11) and (12) into (6), and simplify. Denoting $y = k/n$, and using the integral evaluation

$$
\frac{1}{2}\int_0^1 \int_0^1 \log|t-s| ds\, dt = -\frac{3}{4},
$$

this gives that the left-hand side of (6) is equal to

$$
\begin{aligned}
&-\frac{3}{4} + \frac{3}{4}y^2 + \frac{3}{4}(1-y)^2 - \log 2 \cdot y(1-y) - \frac{1}{2}y^2 \log y - \frac{1}{2}(1-y)^2 \log(1-y) \\
&+ \int_0^1 \int_0^1 \log|t-s| f_{\mathbf{u}}'(s) f_{\mathbf{u}}'(t) ds\, dt - \int_0^1 \int_0^1 \log|t-s| f_{\mathbf{u}}'(s) ds\, dt + o(1) \\
&= -\theta(y) - I(f_{\mathbf{u}}) + o(1), \quad (13)
\end{aligned}
$$

as claimed. □

4 The variational problem and its solution

Fix $0 < y < 1$. Motivated by Theorem 6, we now turn our attention to the problem of minimizing the integral functional $I(f)$ over the appropriate class of y-admissible functions. In the next section we will show how this implies a limit shape result for $\mathbb{P}^n_{\mathrm{DOM}}$-random ASM's.

The precise variational problem that we will solve is the following:

Variational Problem 1. *For a given $0 < y < 1$, find the function f_y^* that minimizes $I(f)$ over all functions $f \in \mathcal{F}_y$.*

Variational Problem 1 is a variant of a class of variational problems that have appeared in several random combinatorial models; see, e.g., [4, 24, 27, 33, 34]. Such problems bear a strong resemblance to classical physical problems of finding the distribution of electrostatic charges subject to various constraints in a one-dimensional space, as well as to problems of finding limiting eigenvalue distributions in random matrix theory. However, the variational problems arising from combinatorial models usually have non-physical constraints that make the analysis trickier. In particular, in several of the works cited above, the presence of such constraints required the authors to first (rather ingeniously) *guess* the solution. Once the solution was conjectured, it was possible to verify that it is indeed the correct one using fairly standard techniques. Cohn, Larsen and Propp, who derived the limit shape of a random boxed plane partition, ask (see Open Question 6.3 in [4]) whether there exists a method of solution for their problem that does not require guessing the solution.

In [27], it was argued however that when dealing with such problems, it is not necessary to guess the solution, since a well-known formula in the theory of singular integral equations for inverting a Hilbert transform on a finite interval actually enables mechanically *deriving* the solution rather than guessing it, once certain intuitively plausible assumptions on the form of the solution are made. Here, we demonstrate again the use of this more systematic approach by using it to solve our variational problem. As an added bonus, the solution rather elegantly turns out to be nearly identical to the solution of the variational problem for the square Young tableaux case (although we see no *a priori* reasons why this should turn out to be the case), and we are able to make use of certain nontrivial computations that appeared in [27], which further simplifies the analysis.

Our goal in the rest of this section will be to prove the following theorem.

Theorem 7. *Define*

$$
\begin{aligned}
Z(x,y) \;=\; & \frac{2}{\pi}\left[(x-1/2)\arctan\left(\frac{\sqrt{\frac{1}{4}-(x-1/2)^2-(y-1/2)^2}}{1/2-y}\right)\right. \\
& +\frac{1}{2}\arctan\left(\frac{2(x-1/2)(1/2-y)}{\sqrt{\frac{1}{4}-(x-1/2)^2-(y-1/2)^2}}\right) \\
& \left.-(1/2-y)\arctan\left(\frac{x-1/2}{\sqrt{\frac{1}{4}-(x-1/2)^2-(y-1/2)^2}}\right)\right].
\end{aligned}
\tag{14}
$$

For $0<y<1/2$*, the solution* f_y^* *to Variational Problem 1 is given by*

$$
f_y^*(x) = \begin{cases} 0 & 0 \le x \le \frac{1-2\sqrt{y(1-y)}}{2}, \\ \frac{y}{2}+\frac{1}{2}Z(x,y) & \frac{1-2\sqrt{y(1-y)}}{2} < x < \frac{1+2\sqrt{y(1-y)}}{2}, \\ y & \frac{1+2\sqrt{y(1-y)}}{2} \le x \le 1. \end{cases}
\tag{15}
$$

For $y=1/2$*, the solution is given by*

$$
f_{1/2}^*(x) = \frac{x}{2}.
$$

For $y>1/2$ *the solution is expressed in terms of the solution for* $1-y$ *by*

$$
f_y^* = x - f_{1-y}^*.
$$

Moreover, for all $0<y<1$ *we have*

$$
I(f_y^*) = -\theta(y).
$$

As a first step, for convenience we reformulate the variational problem slightly to bring it to a more symmetric form, by replacing each $f \in \mathcal{F}_y$ by the function

$$
g(x) = 2f(x) - x. \tag{16}
$$

It is easy to check how the class of y-admissible functions and the functional $I(\cdot)$ transform under this mapping. The result is the following equivalent form of our variational problem.

Variational Problem 2. *For $0 < y < 1$, define the space of functions*

$$\mathcal{G}_y = \Big\{ g : [0,1] \to [-1,1] : g(0) = 0, \;\; g(1) = 2y - 1, \;\; \textit{and } g \textit{ is 1-Lipschitz} \Big\}$$

and the integral functional

$$J(g) = -\int_0^1 \int_0^1 g'(s) g'(t) \log |s - t| \, ds \, dt.$$

Find the function $g_y^* \in \mathcal{G}_y$ that minimizes the functional J over all functions $g \in \mathcal{G}_y$.

The reader may verify that if $f \in \mathcal{F}_y$ and $g \in \mathcal{G}_y$ are related by (16), then the integral functionals I and J are related by

$$I(f) = \frac{1}{4} J(g) - \frac{3}{8}.$$

This implies that the following theorem is an equivalent version of Theorem 7.

Theorem 7′. *For $0 < y < 1/2$, the solution g_y^* to Variational Problem 2 is given by*

$$g_y^*(x) = \begin{cases} -x & 0 \le x \le \frac{1 - 2\sqrt{y(1-y)}}{2}, \\ y - x + Z(x,y) & \frac{1 - 2\sqrt{y(1-y)}}{2} < x < \frac{1 + 2\sqrt{y(1-y)}}{2}, \\ 2y - x & \frac{1 + 2\sqrt{y(1-y)}}{2} \le x \le 1, \end{cases}$$

where $Z(x,y)$ is defined in (14). For $y = 1/2$, the solution is given by $g_{1/2}^*(x) \equiv 0$. For $y > 1/2$ the solution is expressed in terms of the solution for $1 - y$ by $g_y^* = -g_{1-y}^*$. Moreover, for all $0 < y < 1$ we have

$$J(g_y^*) = -4\theta(y) + \frac{3}{2}.$$

We now concentrate our efforts on proving Theorem 7′. First, in the following lemma we recall some basic facts about the space $\mathcal{G}_y$ and the functional J. We omit the proofs, since they are relatively simple and essentially the same claims, with minor differences in the coordinate system, were proved in [27]. (See also [4] where similar facts are proved.)

Lemma 8. *(i) The space $\mathcal{G}_y$ is compact in the uniform norm.*
(ii) The functional J on $\mathcal{G} = \cup_{0<y<1}\mathcal{G}_y$ is a quadratic functional which can be written as

$$J(g) = \langle g, g\rangle,$$

where $\langle\cdot,\cdot\rangle$ is defined by

$$\langle g, h\rangle = -\int_0^1\int_0^1 g'(s)h'(t)\log|s-t|\,ds\,dt.$$

The bilinear form $\langle\cdot,\cdot\rangle$ is defined for any two Lipschitz functions g, h, is continuous on $\mathcal{G}$ with the uniform norm, and is positive semidefinite in the sense that $\langle g, g\rangle \geq 0$ for any Lipschitz function g. The restriction of $\langle\cdot,\cdot\rangle$ to $\mathcal{G}_y$ is positive-definite.
(iii) J is strictly convex on $\mathcal{G}_y$. Therefore, a minimizer g_y^* exists and is unique.

The lemma already solves the problem in the case $y = 1/2$, where clearly $g_{1/2}^* \equiv 0$ is the minimizer for J among *all* Lipschitz functions, and in particular on $\mathcal{G}_{1/2}$. It is also easy to see that a function g is the minimizer for J on $\mathcal{G}_y$ if and only if $-g$ is the minimizer on $\mathcal{G}_{1-y}$. So we may assume for the rest of the discussion that $y < 1/2$.

With these preparations, we can start the analysis. We need to minimize $J(g)$ under the constraints $g \in \mathcal{G}_y$, which we rewrite as

(i) $g(0) = 0$,

(ii) g is differentiable almost everywhere and g' satisfies

$$-1 \leq g' \leq 1, \tag{17}$$

(iii) $\int_0^1 g'(x)\,dx = 2y - 1$.

To address the third constraint, we consider J as being defined on the larger space $\mathcal{G}$ and form the Lagrangian

$$\mathcal{L}(g,\lambda) = J(g) - \lambda\int_0^1 g'(x)\,dx,$$

where λ is a Lagrange multiplier. Minimizing J under this constraint leads, via the usual recipe for constrained optimization, to the equation

$$W(s) := -2\int_0^1 g'(t)\log|s-t|\,dt - \lambda = 0. \tag{18}$$

The reason for this is that, informally, $W(s)$ as defined above can be thought of as "the partial derivative of $\mathcal{L}$ with respect to $g'(s)$" (where we think of $\mathcal{L}$ as a function of the uncountably many variables $\big(g'(s)\big)_{s\in[0,1]}$, which is a standard point of view in the variational calculus).

The relation (18) should hold whenever $g'(s)$ is defined and is in $(-1,1)$. However, because of the constraint (17), the condition will be different when $g' = -1$ or $g' = 1$. The correct condition (the so-called "complementary slackness" condition) is given by the following lemma.

Lemma 9. *If $g \in \mathcal{G}_y$ and for some real number λ the function $W(s)$ defined in* (18) *satisfies*

$$W(s) \text{ is } \begin{cases} = 0 & \text{if } g'(s) \in (-1,1), \\ \geq 0 & \text{if } g'(s) = -1, \\ \leq 0 & \text{if } g'(s) = 1, \end{cases} \tag{19}$$

then $g = g_y^*$ is the minimizer for J in $\mathcal{G}_y$.

Proof. We copy the proof almost verbatim from [27, Lemma 7]. If $h \in \mathcal{G}_y$, then in particular h is 1-Lipschitz, so

$$(h'(s) - g'(s))W(s) \geq 0$$

for all s for which this is defined. So

$$\int_0^1 h'(s)W(s)\,ds \geq \int_0^1 g'(s)W(s)\,ds,$$

or in other words

$$2\langle g, h\rangle - \lambda(2y-1) \geq 2\langle g, g\rangle - \lambda(2y-1),$$

which shows that $\langle g, h\rangle \geq \langle g, g\rangle$. Therefore we get, using Lemma 8(ii), that

$$\langle h, h\rangle = \langle g, g\rangle + 2\langle g, h-g\rangle + \langle h-g, h-g\rangle \geq \langle g, g\rangle,$$

as claimed. $\square$

Having established a *sufficient* condition (comprised of the three separate conditions in (19)) for a function to be a minimizer, we first try to satisfy the condition (18), and save the other conditions for later. Based on intuition that comes from the problem's connection to the combinatorial model, we make the assumption that the minimizer g is piecewise smooth and satisfies

$$g'(s) \in (-1,1) \quad \text{if } s \in \left[\frac{1-\beta}{2}, \frac{1+\beta}{2}\right], \tag{20}$$

$$g'(s) = -1 \qquad \text{if } s \notin \left[\frac{1-\beta}{2}, \frac{1+\beta}{2}\right], \tag{21}$$

where

$$\beta = 2\sqrt{y(1-y)}.$$

Note that $g'(s) = -1$ translates (via (16)) to $f'(s) = 0$ in the original space $\mathcal{F}_y$ of y-admissible functions, which corresponds to having no ascents (or very few ascents) in the vicinity of the scaled position (s, y) in the height matrix of the ASM. Our knowledge of the endpoints of the interval in which $g'(s) > -1$ is related to our foreknowledge of the arctic circle theorem, and one might raise the criticism that this constitutes a "guess". However, the analysis in [27] shows that it would be possible to complete the solution even without knowing this function in advance; here, we guess its value (which actually can be easily guessed based on empirical evidence) so as to simplify the analysis slightly.

Substituting this new knowledge about g into (18) gives the equation

$$\begin{aligned} -\int_{\frac{1-\beta}{2}}^{\frac{1+\beta}{2}} g'(t)\log|s-t|\,dt &= \tfrac{1}{2}\lambda - s\log s - (1-s)\log(1-s) \\ &\quad + \left(s - \frac{1-\beta}{2}\right)\log\left(s - \frac{1-\beta}{2}\right) \\ &\quad + \left(\frac{1+\beta}{2} - s\right)\log\left(\frac{1+\beta}{2} - s\right) - \beta, \qquad s \in \left(\frac{1-\beta}{2}, \frac{1+\beta}{2}\right). \end{aligned}$$

Differentiating with respect to s then gives

$$\begin{aligned} -\int_{\frac{1-\beta}{2}}^{\frac{1+\beta}{2}} \frac{g'(t)}{s-t}\,dt &= -\log s + \log(1-s) \\ &\quad + \log\left(s - \frac{1-\beta}{2}\right) - \log\left(\frac{1+\beta}{2} - s\right). \end{aligned} \tag{22}$$

So, just like in the analysis in [27], we have reached the problem of inverting a Hilbert transform on a finite interval (the so-called ***airfoil equation***).

Moreover, the function whose inverse Hilbert transform we want to compute is very similar to the one that appeared in [27]—in fact, up to scaling factors only the signs of some of the terms are permuted, and in [27] there is an extra term equal to the Lagrange multiplier λ.

Now recall that in fact the general form of the solution of equations of this type is known. The following theorem appears in [8, Sec. 3.2, p. 74] (see also [28, Sec. 9.5.2]):

Theorem 10. *The general solution of the airfoil equation*

$$\frac{1}{\pi}\int_{-1}^{1}\frac{h(u)}{u-v}\,du = p(v), \qquad |v|<1$$

with the integral understood in the principal value sense, and h satisfying a Hölder condition, is given by

$$h(v) = \frac{1}{\pi}\frac{1}{\sqrt{1-v^2}}\int_{-1}^{1}\frac{\sqrt{1-u^2}p(u)}{v-u}\,du + \frac{c}{\sqrt{1-v^2}}$$

for some c.

Now set

$$h(v) = g'((1+\beta v)/2). \tag{23}$$

This function should satisfy

$$\int_{-1}^{1}\frac{h(u)}{u-v}\,du = \log\left(\frac{1-\beta u}{2}\right) - \log\left(\frac{1+\beta u}{2}\right) + \log(1+u) - \log(1-u),$$

so, applying Theorem 10, we get the equation

$$\begin{aligned} h(v) &= \frac{1}{\pi^2}\frac{1}{\sqrt{1-v^2}}\int_{-1}^{1}\frac{\sqrt{1-u^2}}{v-u}\left[\log\left(\frac{1+u}{1-u}\right) + \log\left(\frac{1-\beta u}{1+\beta u}\right)\right] du \\ &\quad + \frac{c}{\sqrt{1-v^2}}. \end{aligned}$$

Where c is an arbitrary constant. This can be written as

$$h(v) = \frac{1}{\pi^2\sqrt{1-v^2}}(\mathcal{I}(v,1/\beta) + \mathcal{I}(-v,1/\beta)) + \frac{c}{\sqrt{1-v^2}}, \tag{24}$$

where $\mathcal{I}$ is defined by

$$\mathcal{I}(\xi,\gamma) = \int_{-1}^{1} \frac{\sqrt{1-\eta^2}}{\xi-\eta} \log\left(\frac{1+\eta}{\gamma+\eta}\right) d\eta$$

and is evaluated in [27, Lemma 8] as

$$\mathcal{I}(\xi,\gamma) = \pi \Bigg[1-\gamma+\sqrt{\gamma^2-1}-\xi \operatorname{arccosh}(\gamma) \\ -2\sqrt{1-\xi^2}\arctan\sqrt{\frac{(\gamma-1)(1-\xi)}{(\gamma+1)(1+\xi)}}\Bigg]$$

Therefore we get that

$$\begin{aligned} h(v) &= \frac{1}{\pi\sqrt{1-v^2}}\left(c+\frac{\beta-1+\sqrt{1-\beta^2}}{\beta}\right) \\ &\quad -\frac{2}{\pi}\left(\arctan\sqrt{\frac{(\beta^{-1}-1)(1-v)}{(\beta^{-1}+1)(1+v)}}+\arctan\sqrt{\frac{(\beta^{-1}-1)(1+v)}{(\beta^{-1}+1)(1-v)}}\right). \end{aligned}$$

Since c is an arbitrary constant, we see that the only sensible choice that will allow h to be a bounded function on the interval $(-1,1)$ is that of $c = -(\beta-1+\sqrt{1-\beta^2})/\beta$. So we have

$$h(v) = -\frac{2}{\pi}\left(\arctan\sqrt{\frac{(\beta^{-1}-1)(1-v)}{(\beta^{-1}+1)(1+v)}}+\arctan\sqrt{\frac{(\beta^{-1}-1)(1+v)}{(\beta^{-1}+1)(1-v)}}\right).$$

At this point, it is worth pointing out that in (24), if we had the *difference* of the two $\mathcal{I}$ integrals instead of their sum, we would get at the end (up to some trivial scaling factors that are due to the use of different coordinate systems) exactly the function from the paper [27] that solves the variational problem for random square Young tableaux! (Compare with eq. (36) in [27] and subsequent formulas). Thus, while the variational problems arising from these two combinatorial models are not exactly isomorphic (which would be perhaps less surprising), they are in some sense *nearly* equivalent. It would be interesting to understand if this phenomenon has a conceptual explanation of some sort, but we do not see one at present.

Simplifying the expression for h using the sum-of-arctangents identity

$$\arctan X + \arctan Y = \arctan\frac{X+Y}{1-XY}$$

gives

$$h(v) = -\frac{2}{\pi} \arctan \sqrt{\frac{1-\beta^2}{\beta^2 - \beta^2 v^2}}.$$

Going back to the original function g related to h via (23), we get that

$$\begin{aligned}
g'(s) &= h((2s-1)/\beta) = -\frac{2}{\pi} \arctan \sqrt{\frac{\frac{1}{4} - y(1-y)}{s(1-s) + y(1-y) - \frac{1}{4}}} \\
&= -\frac{2}{\pi} \arctan \left(\frac{1/2 - y}{\sqrt{\frac{1}{4} - (y-1/2)^2 - (s-1/2)^2}} \right) \\
&= \frac{2}{\pi} \arctan \left(\frac{\sqrt{\frac{1}{4} - (y-1/2)^2 - (s-1/2)^2}}{1/2 - y} \right) - 1
\end{aligned}$$

for $s \in \left(\frac{1-\beta}{2}, \frac{1+\beta}{2}\right)$. From this, we can now get g by integration. First, from (21) we obtain that

$$g(s) = -s \qquad \text{if } \ 0 \le s \le \frac{1-\beta}{2}.$$

Next, in the interval $\left(\frac{1-\beta}{2}, \frac{1+\beta}{2}\right)$ we can integrate g' using the identity

$$\begin{aligned}
&\int_0^t \arctan \sqrt{a - u^2}\, du = t \arctan \sqrt{a - t^2} \\
&\quad + \sqrt{1+a} \arctan \left(\frac{t}{\sqrt{1+a}\sqrt{a-t^2}} \right) - \arctan \left(\frac{t}{\sqrt{a - t^2}} \right), \quad (t^2 < a),
\end{aligned}$$

and obtain without much difficulty that

$$\begin{aligned}
g(s) &= g\left(\frac{1-\beta}{2}\right) + \int_{\frac{1-\beta}{2}}^{s} g'(x)\,dx = \\
&= y - s + \frac{2}{\pi}\left[(s-1/2)\arctan\left(\frac{\sqrt{\frac{1}{4}-(s-1/2)^2-(y-1/2)^2}}{1/2-y}\right)\right. \\
&\qquad + \frac{1}{2}\arctan\left(\frac{2(s-1/2)(1/2-y)}{\sqrt{\frac{1}{4}-(s-1/2)^2-(y-1/2)^2}}\right) \\
&\qquad \left. -(1/2-y)\arctan\left(\frac{s-1/2}{\sqrt{\frac{1}{4}-(s-1/2)^2-(y-1/2)^2}}\right)\right]
\end{aligned}$$

for $s \in \left(\frac{1-\beta}{2}, \frac{1+\beta}{2}\right)$.

Finally, from this last equation it is easy to check that

$$g\left(\frac{1+\beta}{2}\right) = \lim_{s\uparrow\frac{1+\beta}{2}} g(s) = 2y - \frac{1+\beta}{2},$$

so, for $s > \frac{1+\beta}{2}$, again because of (21) we get that $g(s) = 2y-s$. In particular, g satisfies the conditions $g(0) = 0,\ g(1) = 2y - 1$, and it is also 1-Lipschitz, so $g \in \mathcal{G}_y$.

To summarize, we have recovered as a candidate minimizer exactly the function from Theorem 7′. We also verified that it is in $\mathcal{G}_y$. Furthermore, by the derivation and the use of Theorem 10, we know that it satisfies (22), or in other words that $W'(s) \equiv 0$ on $\left(\frac{1-\beta}{2}, \frac{1+\beta}{2}\right)$. We wanted to show that $W(s) \equiv 0$ on this interval. But looking at the definition of $W(s)$ in (18), we see that we are still free to choose the Lagrange multiplier λ, which starting from eq. (22) has disappeared from the analysis! So, taking $\lambda = -2\int_0^1 g'(t)\log|t-1/2|\,dt$ ensures that (18) holds on $\left(\frac{1-\beta}{2}, \frac{1+\beta}{2}\right)$, which is one of the sufficient conditions in Lemma 9.

All that remains to finish the proof that $g = g_y^*$ is the minimizer is to verify the second and third conditions in (19), which we have not considered until now. The third condition is irrelevant, since g' is never equal to 1, so we need to prove that $W(s)$, which we will now re-denote by $W(s, y)$ to emphasize its dependence on y, is nonnegative when $s \notin \left[\frac{1-\beta}{2}, \frac{1+\beta}{2}\right]$. Since

g' is an even function, it follows that $W(\cdot, y)$ is also even, so it is enough to check this when $s > \frac{1+\beta}{2}$.

Once again, our argument follows closely in the footsteps of the analogous part of the proof in [27]. Fix $1/2 < s \le 1$, and let $\hat{y} = \frac{1-\sqrt{1-s^2}}{2}$, so that $\beta(\hat{y}) = s$. We know from (18) that $W(s, \hat{y}) = 0$. To finish the proof, it is enough to show that

$$\frac{\partial W(s,y)}{\partial y} \le 0 \quad \text{for } 0 \le y \le \hat{y}.$$

Denote $G(x,y) = g_y^*(x)$. Then

$$\frac{\partial W(s,y)}{\partial y} = -2 \int_0^1 \frac{\partial^2 G(t,y)}{\partial t \partial y} \log|s-t|\, dt + 2 \int_0^1 \frac{\partial^2 G(t,y)}{\partial t \partial y} \log|t - 1/2|\, dt.$$

A computation shows that if $t \in \left(\frac{1-\beta(y)}{2}, \frac{1+\beta(y)}{2}\right)$ then

$$\frac{\partial^2 G(t,y)}{\partial t \partial y} = \frac{\partial}{\partial y} {g_y^*}'(x) = \frac{2}{\pi} \cdot \frac{1}{\sqrt{\frac{1}{4} - (x-1/2)^2 - (y-1/2)^2}},$$

and otherwise $\partial^2 G(t,y)/\partial t \partial y$ is clearly 0, so that

$$\frac{\partial W(s,y)}{\partial y} = \frac{4}{\pi} \int_{(1-\beta)/2}^{(1+\beta)/2} \frac{\log|t-1/2| - \log(s-t)}{\sqrt{\frac{1}{4} - (t-1/2)^2 - (y-1/2)^2}}\, dt.$$

Now use the two standard integral evaluations

$$\begin{aligned} \int_{-1}^1 \frac{\log|x|}{\sqrt{1-x^2}}\, dx &= -\pi \log(2), \\ \int_{-1}^1 \frac{\log(a-x)}{\sqrt{1-x^2}}\, dx &= \pi \log\left(\frac{a+\sqrt{a^2-1}}{2}\right), \qquad (a>1), \end{aligned}$$

(see [16, eq. 4.241-7, p. 533], and [16, eq. 4.292-3, p. 553]) to conclude that

$$\frac{\partial W(s,y)}{\partial y} = -4 \log\left(\frac{s - 1/2 + \sqrt{(s-1/2)^2 - (\beta/2)^2}}{\beta/2}\right).$$

Since we assumed that $y \le \hat{y}$, or in other words that $s \ge \frac{1+\beta(y)}{2}$, it follows that

$$\frac{\partial W(s,y)}{\partial y} \le -4 \log\left(\frac{s-1/2}{\beta/2}\right) \le 0,$$

as claimed.

We have proved Theorem 7′ (hence also Theorem 7), except the claim about the value of the integral functional J at the minimizer g_y^*. This value could be computed in a relatively straightforward way, as was done for the analogous claim in [27]. We omit this computation, since, as was pointed out in [27], this can also be proved indirectly by using the large deviation principle to conclude that the infimum of the large deviations rate functional $I(f) + \theta(y)$ over the space $\mathcal{F}_y$ must be equal to 0. Therefore the proof of Theorem 7′ is complete. $\square$

5 The limit shape of $\mathbb{P}^n_{\mathbf{Dom}}$-random ASM's

We now apply the results from the previous sections to prove a limit shape result for the height matrix of random ASM's chosen according to the measure $\mathbb{P}^n_{\mathrm{DOM}}$.

Theorem 11. *Let $F(x,y) = f_y^*(x)$, where for each $0 \le y \le 1$, f_y^* is the function defined in* (15)*. For each n let M_n be a $\mathbb{P}^n_{\mathrm{DOM}}$-random ASM of order n, and let $H_n = H(M_n) = (h^n_{i,j})^n_{i,j=0}$ be its associated height matrix. Then as $n \to \infty$ we have the convergence in probability*

$$\max_{0 \le i,j \le n} \left| \frac{h^n_{i,j}}{n} - F(i/n, j/n) \right| \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

Proof. Fix $\epsilon > 0$. We want to show that

$$A^n_\epsilon = \left\{ \max_{0 \le i,j \le n} \left| \frac{h^n_{i,j}}{n} - F(i/n, j/n) \right| > \epsilon \right\}$$

satisfies $\mathbb{P}^n_{\mathrm{DOM}}(A^n_\epsilon) \to 0$ as $n \to \infty$. We start by showing a weaker statement, namely that if $y \in (0,1)$ is given, then $\mathbb{P}^n_{\mathrm{DOM}}(B^n_{\epsilon,y}) \to 0$ as $n \to \infty$, where

$$B^n_{\epsilon,y} = \left\{ \max_{0 \le j \le n} \left| \frac{h^n_{\lfloor ny \rfloor, j}}{n} - F(y, j/n) \right| > \epsilon/2 \right\}$$

(and $\lfloor x \rfloor$ denotes as usual the integer part of a real number x). To prove this, note that

$$B_{\epsilon,y}^n \subseteq \bigcup_{\mathbf{u}} \left\{ M \in \mathcal{A}_n : H(M)_{\lfloor ny \rfloor} = \mathbf{u} \right\},$$

where the union is over all (n,k)-admissible sequences $\mathbf{u}$ (with $k = \lfloor ny \rfloor$) such that

$$||f_{\mathbf{u}} - f_y^*||_\infty = \max_{0 \le x \le 1} |f_{\mathbf{u}}(x) - f_y^*(x)| > \epsilon/2$$

(here, $||\cdot||_\infty$ denotes the supremum norm on continuous functions on $[0,1]$). The number of such sequences is bounded by the total number of (n,k)-admissible sequences, which is equal to $\binom{n}{k} \le 2^n$ (since an (n,k)-admissible sequence is determined by the positions of its k ascents), and for each such $\mathbf{u}$, by Theorem 6 we have

$$\mathbb{P}_{\mathrm{DOM}}^n \left(M \in \mathcal{A}_n : H(M)_{\lfloor ny \rfloor} = \mathbf{u} \right) \le C \exp\Big(-(1+o(1)) c(\epsilon,y) n^2 \Big),$$

where C is a universal constant, and

$$c(\epsilon, y) = \inf \left\{ I(f) + \theta(y) : f \in \mathcal{F}_y, \quad ||f - f_y^*||_\infty \ge \epsilon/2 \right\}. \tag{25}$$

If the infimum in the definition of $c(\epsilon,y)$ were taken over all $f \in \mathcal{F}_y$, it would be equal to 0 by Theorem 7. Note however that the set of $g \in \mathcal{G}_y$ that correspond via (16) to some $f \in \mathcal{F}_y$ participating in the infimum in (25) is a closed subset (in the uniform norm topology) of $\mathcal{G}_y$ that does not contain the minimizer g_y^*. Therefore by Theorem 7$'$ and Lemma 8 we get that in fact $c(\epsilon,y) > 0$. Combining these last observations, we see that indeed $\mathbb{P}_{\mathrm{DOM}}^n(B_{\epsilon,y}^n) \to 0$ as $n \to \infty$.

Next, we claim that the event A_ϵ^n is contained in the union of a finite number (that depends on ϵ but not on n) of events B_{ϵ,y_j}^n, so if $\mathbb{P}_{\mathrm{DOM}}^n(B_{\epsilon,y}^n) \to 0$ for all y then also $\mathbb{P}_{\mathrm{DOM}}^n(A_\epsilon^n) \to 0$. This follows because of the Lipschitz property of the height matrix and of the limit shape function F, which means that proximity to the limit at a sufficiently dense set of values of y implies proximity to the limit everywhere. The details are simple, so we leave to the reader to check that taking $y_j = \lfloor j\epsilon/8 \rfloor$ for $j = 1, 2, \ldots, \lfloor 8/\epsilon \rfloor$ is in fact sufficient to guarantee that

$$A_\epsilon^n \subset \bigcup_{j=1}^{\lfloor 8/\epsilon \rfloor} B_{\epsilon,y_j}^n,$$

as required. $\square$

$$\begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 2 & 3 & 2 & 3 & 4 & 5 \\ 2 & 3 & 2 & 3 & 4 & 3 & 4 \\ 3 & 2 & 3 & 4 & 3 & 2 & 3 \\ 4 & 3 & 2 & 3 & 2 & 3 & 2 \\ 5 & 4 & 3 & 2 & 1 & 2 & 1 \\ 6 & 5 & 4 & 3 & 2 & 1 & 0 \end{pmatrix}$$

Figure 10: The symmetrized height matrix of the ASM from Figure 7.

In the next section we will use a connection between uniformly random domino tilings of the Aztec diamond and $\mathbb{P}^n_{\text{DOM}}$-random ASM's to prove a limit shape theorem for the height function of the random domino tiling. It will be helpful to consider for this purpose a variant of the height matrix of an ASM M, which we call the ***symmetrized height matrix*** (it is sometimes referred to as the ***skewed summation of*** M). If $M \in \mathcal{A}_n$, we define this as the matrix $H_{\text{SYM}}(M) = (h^*_{i,j})^n_{i,j=0}$ with entries given by

$$h^*_{i,j} = i + j - 2H(M)_{i,j}, \qquad (M \in \mathcal{A}_n, \ \ 0 \le i, j \le n),$$

where $H(M)_{i,j}$ is the (i,j)-th entry of the (ordinary) height matrix of M. See Figure 10 for an example. The following theorem is an equivalent version of Theorem 11 formulated for these matrices.

Theorem 11′. *Let $G(x,y) = x + y - 2F(x,y)$, where F is defined in Theorem 11. For each n let M_n be a $\mathbb{P}^n_{\text{DOM}}$-random ASM of order n, and let $H^*_n = H_{\text{SYM}}(M_n) = ({h^*_{i,j}}^n)^n_{i,j=0}$ be its associated symmetrized height matrix. Then as $n \to \infty$ we have the convergence in probability*

$$\max_{0 \le i,j \le n} \left| \frac{{h^*_{i,j}}^n}{n} - G(i/n, j/n) \right| \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

We remark that it would have been possible to work with symmetrized height matrices right from the beginning. In that case the large deviation analysis would have lead directly to Variational Problem 2 without going first through Variational Problem 1. (Note that the limiting symmetrized height function $G(x,y)$ can also be written as $G(x,y) = y - g^*_y(x)$, where g^*_y is the solution to Variational Problem 2.)

6 Back to domino tilings

We now recall some basic facts from [7] about domino tilings of the Aztec diamond AD_n, their height functions, and their connection to alternating sign matrices and their height matrices. This will enable us to use our previous results to reprove the Cohn-Elkies-Propp limit shape result for the height function of a uniformly random domino tiling of AD_n as $n \to \infty$.

Let $\mathcal{G} = \mathcal{G}(\mathrm{AD}_n)$ be the directed graph whose vertex set is

$$V(\mathrm{AD}_n) = \Big\{ (i,j) \in \mathbb{Z}^2 : |i| + |j| \le n+1 \Big\},$$

and where the adjacency relations are

$$(i_1, j_1) \to (i_2, j_2) \iff \begin{array}{c} j_1 = j_2 \text{ and } i_1 - i_2 = (-1)^{n+i_1+j_1}, \\ \text{or} \\ i_1 = i_2 \text{ and } j_1 - j_2 = (-1)^{n+i_1+j_1+1}. \end{array}$$

We call $\mathcal{G}(\mathrm{AD}_n)$ the ***Aztec diamond graph***. Note that its adjacency structure is the standard nearest-neighbor graph structure induced from $\mathbb{Z}^2$, where in addition edges are directed according to a checkerboard parity rule, namely, that if a checkerboard coloring is imposed on the squares $[n, n+1] \times [m, m+1]$ in the lattice dual to $\mathbb{Z}^2$, then the nearest-neighbor edges $u \to v$ are all directed such that a traveller crossing the directed edge will see a black square on her left; see Figure 11(a).

Define a ***height function*** to be any function η on $V(\mathrm{AD}_n)$ such that for any edge $u \to v$ in $\mathcal{G}(\mathrm{AD}_n)$ we have

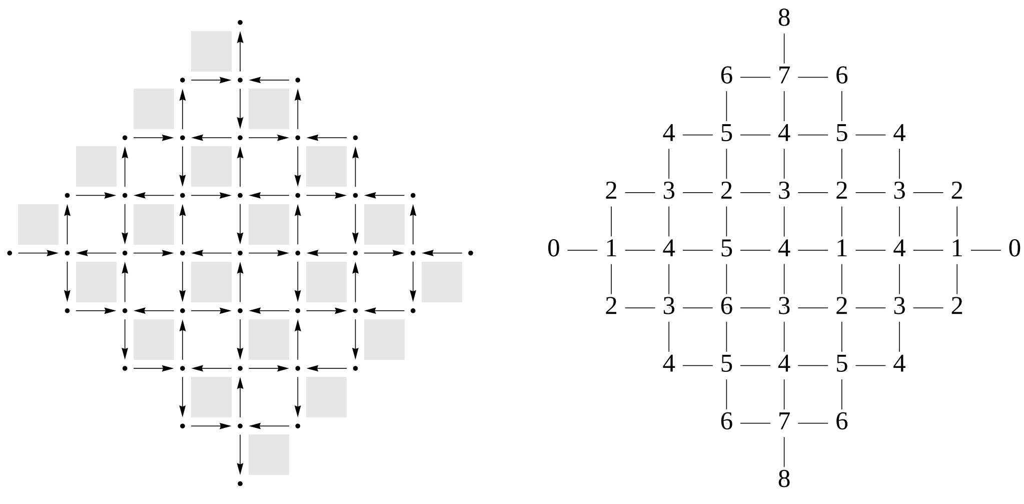
$$\eta(u) - \eta(v) = 1 \text{ or } -3,$$

and such that $\eta(u) - \eta(v) = 1$ whenever $u \to v$ is one of the boundary edges. A height function η on $V(\mathrm{AD}_n)$ is called ***normalized*** if $\eta(-n, 0) = 0$.

It is known that any domino tiling T of AD_n determines a unique normalized height function η_T by the requirement that for any directed edge $u \to v$ we have

$$\eta_T(u) - \eta_T(v) = \begin{cases} -3 & \text{the segment (u,v) crosses a domino tile in } T, \\ 1 & \text{otherwise.} \end{cases}$$

Conversely, any normalized height function η is of the form η_T for some domino tiling. See Figure 11(b).



(a) (b)

Figure 11: (a) The Aztec diamond graph of order 3; (b) The normalized height function corresponding to the tiling from Figure 1.

Another important fact concerns the beautiful connection, discovered by Elkies, Kuperberg, Larsen and Propp [7], between height functions of domino tilings of AD_n and height matrices of ASM's: each normalized height function η on $V(\mathrm{AD}_n)$ is essentially comprised of the superposition of two (symmetrized) height matrices $H_{\mathrm{SYM}}(A), H_{\mathrm{SYM}}(B)$ where A is an ASM of order n and B is an ASM of order $n+1$. More precisely, $H_{\mathrm{SYM}}(A)$ and $H_{\mathrm{SYM}}(B)$ can be recovered from η by

$$H_{\mathrm{SYM}}(A)_{i,j} = \frac{\eta(-n+1+i+j,-i+j)-1}{2}, \tag{26}$$

$$H_{\mathrm{SYM}}(B)_{i,j} = \frac{\eta(-n+i+j,-i+j)}{2}, \tag{27}$$

(note the slight difference from the formulas in [7] due to a difference in the center of the coordinate system used). This correspondence defines a one-to-one mapping from the set of domino tilings of AD_n to the set of pairs (A, B) where $A \in \mathcal{A}_n$ and $B \in \mathcal{A}_{n+1}$. The pairs (A, B) which are obtained via this mapping are exactly the so-called ***compatible*** pairs defined by Robbins and Rumsey [30]: A and B are called compatible if the (non-symmetrized) height matrices $H(A), H(B)$ satisfy the conditions

$$\begin{aligned} H(B)_{i,j} &\leq H(A)_{i,j}, \\ H(B)_{i+1,j+1} - 1 &\leq H(A)_{i,j}, \\ H(A)_{i,j} &\leq H(B)_{i+1,j}, \\ H(A)_{i,j} &\leq H(B)_{i,j+1}. \end{aligned}$$

It was also shown in [30] that for a given ASM $A \in \mathcal{A}_n$, the number of $B \in \mathcal{A}_{n+1}$ that are compatible with A is equal to $2^{N_+(A)}$. Combined with the formula for the number of domino tilings of AD_n, this implies that if T is a uniformly random domino tiling of AD_n, and (A, B) is the associated pair of compatible ASM's, then the random ASM A is distributed according to the domino measure $\mathbb{P}^n_{\text{DOM}}$ (of course, this provides the explanation for our choice of name for this measure).

We now combine Theorem 11′ with the above discussion to easily obtain the following result, originally proved in [2].

Theorem 12. *For each $n \geq 1$, let T_n be a uniformly random domino tiling of AD_n, and let $\eta_n = \eta_{T_n}$ be its associated height function. Then as $n \to \infty$ we have the convergence in probability*

$$\max_{(i,j)\in V(AD_n)} \left| \frac{1}{n}\eta_n(i,j) - R(i/n, j/n) \right| \xrightarrow[n\to\infty]{\mathbb{P}} 0,$$

where

$$R(u,v) = 2G\left(\frac{u-v+1}{2}, \frac{u+v+1}{2}\right), \qquad (|u| + |v| \leq 1),$$

and G is defined in Theorem 11′.

Proof. For pairs $(i,j) \in V(\mathrm{AD}_n)$ for which $i + j + n$ is odd, the proximity of $n^{-1}\eta_n(i,j)$ to $R(i/n, j/n)$ follows from (26). For other pairs (i,j), apply the previous observation to any pair (i', j') adjacent to (i,j) and use the facts that $|\eta_n(i,j) - \eta(i',j')| \leq 3$ and that R is a continuous function. $\square$

7 Concluding remarks

7.1 Relation to the arctic circle theorem

Theorem 12 implies a weak form of the arctic circle theorem (Theorem 1): First, since inside the arctic circle the limit shape function $R(u,v)$ is not a linear function, it follows that the frozen region cannot extend in the limit into the arctic circle, which is "half" of the theorem. In the other direction, we get only a weaker statement that outside the arctic circle we can have in the limit at most $o(n^2)$ "non-frozen" dominoes, since that is what the linearity of the limiting height function in that region implies.

It is interesting to contrast this with the square Young tableaux problem. There, too, the large deviation approach gave only a bound in one direction on the behavior of the square Young tableau along the boundary of the square. However, Pittel and Romik managed to prove the other direction using an additional combinatorial argument (inspired by a method of Vershik and Kerov [34]). It would be interesting to see whether one can emulate this approach in the present case to get a new proof of the arctic circle theorem. A similar question applies to the problem of random boxed plane partitions studied by Cohn, Larsen and Propp [4], where again the limit shape theorem for the height function does not imply an arctic circle result in its strong form.

7.2 Other arctic circles and more general arctic curves

In this paper we have shown that two so-called arctic circle phenomena, namely those appearing in the contexts of random domino tilings of the Aztec diamond and of random square Young tableaux, are closely related, in the sense that the limit shape results underlying them can be given a more or less unified treatment using the techniques of large deviation theory and the calculus of variations, and that the derivations in both cases result in nearly identical computations and formulas. Note that these are not the only combinatorial models in which arctic circles appear. Other examples known to the author include the shape of a uniformly random boxed plane partition derived by Cohn, Larsen and Propp [4] and the arctic circle theorem for random groves, due to Petersen and Speyer [26]. One might therefore wish to extend the insights of the present paper to these other models. The treatment of boxed plane partitions in [4] is already based on a large deviations analysis, and in fact the variational problem studied there seems to be quite closely related to the variational problems studied here and in [27]. Therefore, it should be relatively straightforward to use the techniques presented here to give a new derivation of the solution to the variational problem from [4] (which in particular would provide a fully satisfactory answer to Open Question 6.3 from that paper).

The analysis of random groves, on the other hand, is based on generating function techniques, and it is not clear how to apply the ideas presented here to that setting.

It is also worth mentioning that there is a large literature on the subject of limit shapes of various classes of random combinatorial objects, and tiling models in particular, where one encounters in many cases a spatial phase transition between a "frozen" and a "temperate" region. The equations governing such limit shapes can in general lead to a much more diverse family of non-circular "arctic curves" describing the shape of the interface between the frozen and temperate regions. For details, see for example the papers [3], [22], [23].

7.3 Uniformly random ASM's

One reason why the methods and ideas presented in this paper may be considered worthy of attention is somewhat speculative in nature. It pertains to the potential future applicability of these methods and ideas to a well-known open problem on alternating sign matrices: that is, the problem of finding the limiting shape of a *uniformly* random ASM of high order. Here, "limit shape" is usually taken to refer to the shape of the region in which the nonzero entries cluster (the "temperate region"), although one could also ask (as we have done here in the case of $\mathbb{P}^n_{\mathrm{DOM}}$-random ASM's) about the limiting shape of the height matrix, which also contains useful information about the behavior of the ASM inside the temperate region.

Important progress on this question was made recently by Colomo and Pronko [6], who conjectured the explicit formula

$$x^2 + y^2 + |xy| = |x| + |y|$$

for the limit shape of the boundary of the temperate region in a uniformly random ASM (Fig. 12), and provided a heuristic derivation of this conjectured formula based on certain natural, but still conjectural, analytic assumptions.

In view of this state of affairs, it is worth noting that the ideas presented in this paper seem to be rather suitable for attacking this challenging open problem. There is only one main "missing piece" (albeit possibly a very substantial one) in our understanding. The idea is to replace Theorem 3, which is the combinatorial observation which lies at the heart of the large

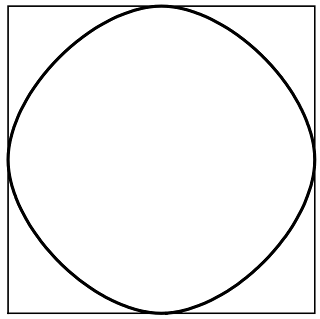


Figure 12: The Colomo-Pronko conjectured limit shape for uniformly random alternating sign matrices.

deviations analysis, with an analogous statement that holds for the uniform measure on the set $\mathcal{A}_n$ of ASM's of order n. This statement is given in the following theorem, whose proof follows similar lines to the proof of Theorem 3 and is omitted.

Theorem 13. *Let $\mathbb{P}_{\text{UNIF}}$ denote the uniform measure on the set of ASM's of order n. For a positive integer k and integers $x_1 < x_2 < \ldots < x_k$, denote by $\alpha_k(x_1, \ldots, x_k)$ the number of monotone triangles of order k with bottom row $(x_1, \ldots, x_k)$. Then, in the notation of Theorem 3, we have*

$$\mathbb{P}_{\text{UNIF}}\Big[M \in \mathcal{A}_n : (X_k(1), \ldots, X_k(k)) = (x_1, \ldots, x_k)\Big] \\ = \frac{1}{|\mathcal{A}_n|}\alpha_k(x_1, \ldots, x_k)\, \alpha_{n-k}(y_1, \ldots, y_{n-k})$$

Unfortunately, while a formula for $|\mathcal{A}_n|$ is known (see [1]), the function α_k seems much more difficult to understand (and in particular, to derive asymptotics for) than the Vandermonde function Δ, and this is the piece that is missing when one tries to duplicate our analysis to the setting of uniformly random ASM's. Nevertheless, the function α_k has recently been the subject of several very fruitful studies. Fischer [10] derived the following beautiful "operator formula" for α_k:

$$\alpha_k(x_1, \ldots, x_k) = \left[\prod_{1 \le i < j \le k} (\mathrm{Id} + E_i D_j)\right] \frac{\Delta(x_1, \ldots, x_k)}{\Delta(1, \ldots, k)}. \tag{28}$$

Here, Δ is the Vandermonde function as before, and Id, E_j and D_i are operators acting on the ring of polynomials $\mathbb{C}[x_1, \ldots, x_k]$: Id is the identity operator, E_j is the shift operator in the variable x_j (that substitutes x_j+1 for each occurrence of x_j in a polynomial), and $D_i = E_i - \mathrm{Id}$ is the (right-)differencing operator in the variable x_i.

Fischer then showed in several subsequent papers that it is possible to use (28) to get highly non-trivial information on the enumeration of alternating sign matrices: In [11] she obtained a new proof of the celebrated Refined Alternating Sign Matrix Theorem (see [1] for the statement and fascinating history of this result); in [15] she and the author obtained additional results concerning a "doubly-refined" enumeration of ASM's; and in [13] and [14] she extended these results further to a "multiply-refined" enumeration. Thus, it seems quite conceivable that additional study of α_k may eventually lead to a deeper understanding of this function, that, in combination with Theorem 13 and the techniques of this paper, could provide a basis for a successful attack on the limit shape problem for uniformly random ASM's.

References

[1] D. M. Bressoud. Proofs and Confirmations: The Story of the Alternating Sign Matrix Conjecture. Cambridge University Press, 1999.

[2] H. Cohn, N. Elkies, J. Propp. Local statistics for random domino tilings of the Aztec diamond. *Duke Math. J.* 85 (1996), 117–166.

[3] H. Cohn, R. Kenyon, J. Propp. A variational principle for domino tilings. *J. Amer. Math. Soc.* 14 (2001), 297–346.

[4] H. Cohn, M. Larsen, J. Propp. The Shape Of A Typical Boxed Plane Partition. *New York J. Math.* 4 (1998), 137–165.

[5] F. Colomo, A. G. Pronko. The arctic circle revisited. In: Integrable System and Random Matrices: In Honor of Percy Deift, Eds. J. Baik, T. Kriecherbauer, L.-C. Li, K. T.-R. McLaughlin, C. Tomei, p. 361–370. Contemporary Mathematics Vol. 458, American Mathematical Society, 2008.

[6] F. Colomo, A. G. Pronko. The limit shape of large alternating sign matrices. To appear in *SIAM J. Disc. Math.*

[7] N. Elkies, G. Kuperberg, M. Larsen, J. Propp. Alternating sign matrices and domino tilings. *J. Algebraic Combin.* 1 (1992), 111–132; 219–234.

[8] R. Estrada, R. P. Kanwal. Singular Integral Equations. Birkhäuser, 2000.

[9] S. R. Finch. Mathematical Constants. Encyclopedia of Mathematics and Its Applications vol. 94, Cambridge University Press, 2003.

[10] I. Fischer. The number of monotone triangles with presecribed bottom row. *Adv. Appl. Math.* 37 (2006), 249–267.

[11] I. Fischer. A new proof of the refined alternating sign matrix theorem. *J. Comb. Theory Ser. A* 114 (2007), 253–264.

[12] I. Fischer. The operator formula for monotone triangles—simplified proof and three generalizations. *J. Combin. Theory Ser. A* 117 (2010), 1143–157.

[13] I. Fischer. Refined enumerations of alternating sign matrices: Monotone (d,m)-trapezoids with prescribed top and bottom rows. To appear in *J. Alg. Combin.*

[14] I. Fischer. Linear relations of refined enumerations of alternating sign matrices. Preprint (2010), `arXiv:1008.0527v1`.

[15] I. Fischer, D. Romik. More refined enumerations of alternating sign matrices. *Adv. Math.* 222 (2009), 2004–2035.

[16] I. S. Gradshteyn, I. M. Ryzhik. Table of Integrals, Series and Products, 6th ed. Academic Press, 2000.

[17] W. Jockusch, J. Propp, P. Shor. Random domino tilings and the arctic circle theorem. Preprint (1995), `arXiv:math/9801068`.

[18] K. Johansson. Discrete orthogonal polynomial ensembles and the Plancherel measure. *Ann. Math.* 153 (2001), 259–296.

[19] K. Johansson. Non-intersecting paths, random tilings and random matrices. *Probab. Theory and Related Fields* 123 (2002), 225–280.

[20] K. Johansson. The arctic circle boundary and the Airy process. *Ann. Probab.* 33 (2005), 1–30.

[21] P. W. Kasteleyn. The statistics of dimers on a lattice. I. The number of dimer arrangements on a quadratic lattice. *Physica* 27 (1961), 1209–1225.

[22] R. Kenyon, A. Okounkov. Limit shapes and the complex burgers equation. *Acta Math.* 199 (2007), 263–302.

[23] R. Kenyon, A. Okounkov, S. Sheffield. Dimers and amoebae. *Ann. Math.* 163 (2006), 1019–1056.

[24] B. F. Logan, L. A. Shepp. A variational problem for random Young tableaux. *Adv. Math.* 26 (1977), 206–222.

[25] W. H. Mills, D. P. Robbins, H. Rumsey. Alternating sign matrices and descending plane partitions. *J. Comb. Theory Ser. A* 34 (1983) 340–359.

[26] T. K. Petersen, D. Speyer. An arctic circle theorem for groves. *J. Comb. Theory Ser. A* 111 (2005), 137–164.

[27] B. G. Pittel, D. Romik. Limit shapes for random square Young tableaux. *Adv. Appl. Math.* 38 (2007), 164–209.

[28] D. Porter, D. S. G. Stirling. Integral Equations. Cambridge University Press, 1990.

[29] J. Propp. The many faces of alternating sign matrices. *Discr. Math Theor. Comp. Sci.* Proceedings of **DM-CCG**, Conference Volume **AA** (2001), 43–58.

[30] D. P. Robbins, H. Rumsey. Determinants and alternating sign matrices. *Adv. Math.* 62 (1986), 169–184.

[31] H. Temperley, M. Fisher. The dimer problem in statistical mechanics — an exact result. *Phil. Mag.* 6 (1961), 1061–1063.

[32] B. Valkó. Private communication, 2006.

[33] A. M. Vershik, S. V. Kerov. Asymptotics of the Plancherel measure of the symmetric group and the limiting shape of Young tableaux. *Soviet Math. Dokl.* 18 (1977), 527–531.

[34] A. M. Vershik, S. V. Kerov. The asymptotics of maximal and typical dimensions of irreducible representations of the symmetric group. *Funct. Anal. Appl.* 19 (1985), 21–31.

Dan Romik
Department of Mathematics
University of California
Davis, CA 95616
USA

email: romik@math.ucdavis.edu