

The Limit Shape of Convex Lattice Polygons*

I. Bárány

Mathematical Institute of the Hungarian Academy of Sciences,
 POB 127, 1364 Budapest, Hungary

Abstract. It is proved here that, as $n \rightarrow \infty$, almost all convex $(1/n)\mathbb{Z}^2$ -lattice polygons lying in the square $[-1, 1]^2$ are very close to a fixed convex set.

1. Introduction and Results

This paper partially answers a beautiful and inspiring question of Vershik. Following a result of Arnold [A], he asked about 10 years ago whether a “limit shape” to some set of convex lattice polygons exists. We prove here that the answer is yes for \mathcal{P}_n , the set of all convex lattice polygons lying in the square $[-n, n]^2$. Here a convex lattice (or \mathbb{Z}^2 -lattice) polygon is a convex lattice polygon whose vertices belong to \mathbb{Z}^2 .

For the precise statement a normalization is needed. Define

$$\mathcal{P}_n^0 = \{(1/n)P : P \in \mathcal{P}_n\}$$

so that \mathcal{P}_n^0 is the set of all $(1/n)\mathbb{Z}^2$ -lattice polygons lying in the square $[-1, 1]^2$. Set, further,

$$L = \{(x, y) \in \mathbb{R}^2 : \sqrt{1 - |x|} + \sqrt{1 - |y|} \geq 1\}.$$

This is a convex set in $[-1, 1]^2$. It is bounded by four parabola-arcs, each of them touching consecutive edges of the square at their midpoints. When $P \in \mathcal{P}_n^0$ we write $\chi_P(x)$ for the characteristic function of P , i.e., $\chi_P(x) = 1$ if $x \in P$ and 0 if $x \notin P$.

*This research was partially supported by Hungarian Science Foundation Grants 1907 and 1909.

The limit-shape theorem says that “almost all” polygons in \mathcal{P}_n^0 are very close to L . One way of formulating this is:

Theorem 1. *The average of $\chi_P(x)$ over all $P \in \mathcal{P}_n^0$ tends to 1 if $x \in \text{int } L$ and to 0 if $x \notin L$.*

Another way of stating this result is to work with convex lattice curves (instead of polygons). A convex lattice curve is a convex polygonal path with vertices from \mathbb{Z}^2 . Write \mathcal{E}_n for the set of all convex lattice curves connecting $(0, n)$ to $(n, 0)$ within the triangle with vertices $(0, n)$, $(0, 0)$, $(n, 0)$ and let \mathcal{E}_n^0 be what you get from \mathcal{E}_n after normalization: $\mathcal{E}_n^0 = \{(1/n)C : C \in \mathcal{E}_n\}$. Define further

$$\Gamma = \{(x, y) \in \mathbb{R}^2 : \sqrt{x} + \sqrt{y} = 1\}.$$

Γ is an arc of the parabola that touches the axes at $(1, 0)$ and $(0, 1)$. Let $U_\varepsilon(\Gamma)$ denote the ε -neighborhood of Γ . The limit-shape theorem for convex lattice curves is as follows.

Theorem 2. *For every $\varepsilon > 0$ there is an $n_0(\varepsilon)$ such that for $n > n_0(\varepsilon)$ the number of $C \in \mathcal{E}_n^0$ not lying in $U_\varepsilon(\Gamma)$ is less than $\varepsilon |\mathcal{E}_n^0|$.*

The proof will show that it can be taken that $n_0(\varepsilon) = c\varepsilon^{-1/3} \log 1/\varepsilon$.

It is perhaps worth mentioning here that there are other limit-shape theorems in the literature. For instance, Vershik and Kerov [VK1], [VK2] and Logan and Schepp [LS] proved that there is a limit shape to the Young tableaux on n elements equipped with the Plancherel measure. Further, Vershik [VK1], [VK2] showed that the same holds when the uniform measure is used. In fact, this latter result is implicit in the work of Turán and Szalay [TS] but the formulation there lacks the natural appeal and beauty of the limit-shape phenomenon.

I thank A. M. Vershik for posing the question, for the many discussions and correspondence we had on the topic of this paper, and also for his contribution to the proof of Theorem A below. In fact, shortly after I found Theorem 2 he also proved the limit-shape theorem. His arguments appear in [V]. I also thank G. Halász and I. Z. Ruzsa for useful and illuminating discussions and explanations.

After the results of this paper had been obtained, Ya. G. Sinai found a different way of proving the limit-shape theorem. His approach which is based on probability theory gives more than the methods presented here, for instance, the average number of edges in \mathcal{E}_n^0 , speed of convergence to Γ , central limit theorems, etc. His arguments appear in [S].

2. Auxiliary Results

We need to know the size of \mathcal{E}_n^0 or, what is the same, of \mathcal{E}_n . In fact we need a little more. Write $p(m, n)$ for the number of convex lattice polygons connecting $(0, 0)$ to $(m, n) \in \mathbb{Z}^2$ within the rectangle $\text{conv}\{(0, 0), (0, n), (m, n), (m, 0)\}$.

Theorem A. *As $m, n \rightarrow \infty$ so that $c < m/n < 1/c$ where $c \in (0, 1)$ is any fixed constant,*

$$p(m, n) = \exp \left\{ 3 \sqrt[3]{\frac{\zeta(3)}{\zeta(2)}} (nm)^{1/3} (1 + o(1)) \right\}.$$

It is convenient to introduce the notation $\zeta = \sqrt[3]{\zeta(3)/\zeta(2)}$ for the constant in the exponent.

Next, let $a, b \in \mathbb{Z}^2$ be primitive vectors with $a \neq \pm b$. Assume $\alpha, \beta > 0$ (we think of a, b as fixed and α, β large) and consider the parallelogram

$$T = \text{conv}\{0, \alpha a, \beta b, \alpha a + \beta b\}.$$

Write $p(T)$ for the number of convex lattice curves connecting 0 to $\alpha a + \beta b$ within T . Here edges parallel with a or b are allowed, and $\alpha a + \beta b \in \mathbb{Z}^2$.

We denote the area of parallelogram T by the same letter T . This should not cause any confusion. Theorem A says that when T is the square $[0, n]^2$, then $p(T) = \exp\{3\zeta T^{1/3}(1 + o(1))\}$. The proof of the limit-shape theorem would be simpler if the following estimate were true for every parallelogram T : as $T \rightarrow \infty$,

$$p(T) \leq \exp\{3\zeta T^{1/3}(1 + o(1))\}. \tag{2.1}$$

This estimate is valid for a large class of parallelograms but an argument of Halász [H] shows that it does not hold in general. A sketch of Halász's proof, with his kind permission, is given in a remark following the proof of Theorem A.

The failure of this general estimate forces one to look for "conditional" bounds. Define

$$\det T = |\det(a, b)|, \quad \text{and} \quad \eta(T) = \max\left(\frac{T}{\alpha^3}, \frac{T}{\beta^3}\right) = \det T \max\left(\frac{\beta}{\alpha^2}, \frac{\alpha}{\beta^2}\right),$$

since $T = \alpha\beta \det T$. Notice that both parameters are invariant under lattice-preserving affine transformations. The conditional estimate is as follows.

Theorem B. *There is a constant c_1 such that*

$$p(T) \leq T \exp\{3\zeta T^{1/3}(1 + c_1 \det T \eta^{1/4}(T))\}.$$

The important thing here is that the constant ζ is the same as in Theorem A. We also need the following "unconditional" bound which was proved in [BP] and in [BV].

Theorem C. *There is a constant c_2 such that*

$$p(T) \leq \exp\{c_2 T^{1/3}\}.$$

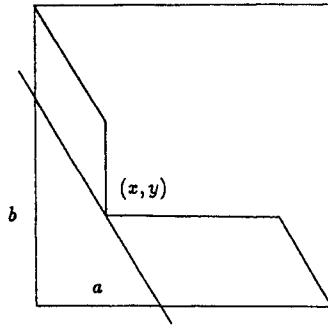


Fig. 1

Remark 1. Quite often we are interested in the set of convex lattice curves in the triangle $\text{conv}\{0, \alpha a, \alpha a + \beta b\}$. Notice that their number is just $(p(T) + 1)/2$. For our purposes, this is the same as $p(T)$.

We also need some simple facts from elementary geometry. Assume $0 \leq x, y \leq 1$ and $0 < a, b \leq 1$ with $x/a + y/b = 1$. Let T_1, T_2 be the parallelograms with vertices $(x, y), (a, 0), (1, 0), (x + 1 - a, y)$ and $(x, y), (0, b), (0, 1), (x, 1 + y - b)$. (See Fig. 1.)

Theorem D.

$$T_1^{1/3} + T_2^{1/3} \leq 1 - \frac{1}{6}(1 - a - b)^2.$$

Moreover, equality holds here if and only if $a + b = 1$ and $\sqrt{x} + \sqrt{y} = 1$.

The proof of Theorem D is simple and is therefore omitted. It is worth mentioning, however, that the inequality $T_1^{1/3} + T_2^{1/3} \leq 1$ was proved some 150 years ago (see [B]).

3. Further Results

The following theorem is due to Vershik [V]. Given a convex curve K in the plane and $\varepsilon > 0$, write $\mathcal{E}_n^0(K, \varepsilon)$ for the set of all convex $(1/n)\mathbb{Z}^2$ -lattice polygons that lie in the ε -neighborhood of K .

Theorem [V]. Assume K is a \mathcal{C}^2 convex curve whose curvature κ satisfies $c < \kappa < 1/c$ everywhere on K (for some fixed $c > 0$). Then

$$\lim_{\varepsilon \rightarrow 0} \lim_{n \rightarrow \infty} n^{-2/3} \log |\mathcal{E}_n^0(K, \varepsilon)| = 3\zeta^3 \sqrt{\frac{1}{4}} \int_K \kappa^{1/3} ds.$$

This theorem may provide an alternative way of proving the limit-shape theorem. It says that there are essentially $\exp\{\text{const } n^{2/3} \int_K \kappa^{1/3} ds\}$ convex lattice curves in the vicinity of K . Observe that $\int \kappa^{1/3} ds$ is a constant multiple of the affine length (see [B]) of the curve K . Note that, among all convex curves connecting $(0, 1)$ to $(1, 0)$ within the triangle $\text{conv}\{0, (0, 1), (1, 0)\}$, Γ has the largest affine length. This is in accordance with Theorem 2. Vershik's theorem is useful when one wants to find, in a set of convex curves, the curve with the maximal number of lattice polygons in its vicinity, since it leads to variational problems and differential equations.

Theorem 1 raises the following question. Given a convex compact set $S \subset \mathbb{R}^2$, is there a limit shape to the set of all convex $(1/n)\mathbb{Z}^2$ -lattice polygons contained in S ? It is not difficult to see that if there is a limit shape, it must be affinely invariant, and the previous theorem suggests that the answer must be that convex subset of S that has the maximal affine perimeter. This is indeed so; details will appear elsewhere.

Another question of interest is this. Which convex set $S \subset \mathbb{R}^2$ of area 1 contains the largest number of convex $(1/n)\mathbb{Z}^2$ -lattice polygons? Vershik's theorem and the affine isoperimetric inequality suggest that the answer must be an ellipsoid. It is, however, not clear which one. It may happen that this ellipsoid is not well determined and, as n tends to infinity, its diameter will tend to infinity as well.

The proof of the limit-shape theorem is based on the following simple but important observation. The convex lattice curves connecting $(0, 0)$ to (m, n) within the triangle $\text{conv}\{(0, 0), (m, 0), (m, n)\}$ are in one-to-one correspondence with the multipartitions of (m, n) into nonparallel, nonnegative, and integral summands. This is easy to see: the summands of any such multipartition can be ordered by slope, and in this order they define a unique convex lattice curve from $(0, 0)$ to (m, n) and vice versa. Alternatively, writing v_1, \dots, v_k for the summands in the multipartition, the zonotope defined as the Minkowski sum of the segments $[0, v_i]$ may be considered, this zonotope contains $(0, 0)$ and (m, n) and is contained in $[0, m] \times [0, n]$. With this formulation a limit-shape theorem is obtained for multipartitions with the property that "parallel summands are not allowed". This leads to two different generalizations.

First, other properties of multipartitions may be considered such as (1) only primitive vectors are allowed, (2) the summands are different, (3) anything is allowed, etc. As it turns out from the proof of Theorem 2, if the multipartitions with the property in question satisfy Theorems A and B with the same constant in the exponent, and Theorem C with some other (larger) constant, then they have a limit shape. In fact, the limit shape is the same as the one with nonparallel summands. The second generalization is probably more interesting, it is higher dimensions. The methods presented here do extend to the statistics of multipartitions of $(n_1, \dots, n_d) \in \mathbb{R}^d$, the geometry part (Theorem D) gets more involved. The outcome is that there is a limit shape to multipartitions (with nonparallel summands, say). This will be the topic of a subsequent paper.

However, the question whether there is a limit shape to all convex $(1/n)\mathbb{Z}^d$ -lattice polytopes contained in the unit cube $[0, 1]^d$ remains open for $d > 2$. It seems likely that the answer for this question will require a completely different approach.

4. Proof of Theorem 2

The proof which uses Theorems A–D is based on first partitioning the curves $C \in \mathcal{E}_n^0$ not lying in $U_\varepsilon(\Gamma)$ into “few” classes, and, second, showing that each such class contains “few” elements.

Define \mathbb{P} to be the set of primitive vectors from \mathbb{Z}^2 . Let N be a large integer satisfying

$$\frac{3}{\sqrt{\varepsilon}} < N < \frac{4}{\sqrt{\varepsilon}}. \tag{4.1}$$

Write

$$W(N) = \{(u, v) \in \mathbb{P} : 0 < u, v \leq N\}.$$

Consider a line l of the form

$$\frac{x}{a} + \frac{y}{b} = 1, \tag{4.2}$$

where $0 < a, b$. Notice that l is tangent to Γ if and only if $a + b = 1$. Using this it is a simple matter to prove:

Claim 1. *If $C \in \mathcal{E}_n^0$ does not lie between $(1 - \varepsilon)\Gamma$ and $(1 + \varepsilon)\Gamma$, then it has a tangent of the form (4.2) with (a, b) parallel to some $(u, v) \in W(N)$ such that $|1 - a - b| \geq \varepsilon/2$.*

Now fix $(u, v) \in W(N)$ and a point $z_0 \in (1/n)\mathbb{Z}^2$ from the unit square. Let l be the line of the form (4.2) with (a, b) parallel to (u, v) and passing through z_0 . Write $\mathcal{E}_n(u, v, z_0)$ for the set of convex lattice curves from \mathcal{E}_n^0 that touch l at the point z_0 . (This is the empty set if l separates $(0, 1)$ from $(1, 0)$, so assume this is not the case.)

Claim 2. *Assume $|1 - a - b| \geq \varepsilon/2$. Then, if n is large enough,*

$$|\mathcal{E}_n(u, v, z_0)| \leq \exp\left\{3\zeta n^{2/3}\left(1 - \frac{\varepsilon^2}{30}\right)\right\}.$$

Let us see first how Claim 2 implies Theorem 2. Notice that both $(1 - \varepsilon)\Gamma$ and $(1 + \varepsilon)\Gamma$ lie in $U_\varepsilon(\Gamma)$. Then, by Claim 1, every $C \in \mathcal{E}_n^0$ not lying in $U_\varepsilon(\Gamma)$ is in one of the classes $\mathcal{E}_n(u, v, z_0)$. The number of such classes is at most n^2N^2 . So the number of convex lattice curves not lying in $U_\varepsilon(\Gamma)$ is at most

$$n^2N^2 \exp\{3\zeta n^{2/3}(1 - \varepsilon^2/30)\}$$

which is, by Theorem A, much less than $|\mathcal{E}_n^0|$.

Proof of Claim 2. Observe that, using the notation of Fig. 1,

$$|\mathcal{E}_n(u, v, z_0)| = p(nT_1)p(nT_2),$$

where nT_i are blown up copies of T_i with vertex (x_0, y_0) placed at the origin. We estimate $p(nT_i)$ using Theorems B and C. The parameters of the parallelogram nT_1 are

$$a_1 = (1, 0), \quad b_1 = (u, -v), \quad \alpha_1 = n(1 - \alpha), \quad \beta_1 = \frac{ny_0}{v}, \quad \det(nT_1) = v,$$

$$\text{area}(nT_1) = n^2(1 - a)y_0, \quad \eta(nT_1) = \frac{1}{n} \max\left(\frac{(1 - a)v^3}{y_0^2}, \frac{y_0}{(1 - a)^2}\right).$$

Similarly, the parameters of nT_2 are

$$a_2 = (0, 1), \quad b_2 = (-u, v), \quad \alpha_2 = n(1 - b), \quad \beta_2 = \frac{nx_0}{u}, \quad \det(nT_2) = v,$$

$$\text{area}(nT_2) = n^2(1 - b)x_0, \quad \eta(nT_2) = \frac{1}{n} \max\left(\frac{(1 - b)u^3}{x_0^2}, \frac{x_0}{(1 - b)^2}\right).$$

We have to consider a few cases now so define the following two conditions where Δ is a small positive number, for instance $\Delta = \varepsilon^7$ or $c\varepsilon^6$ will do if c is small enough:

$$y_0 \geq \Delta \quad \text{and} \quad 1 - a \geq \Delta, \tag{C1}$$

$$x_0 \geq \Delta \quad \text{and} \quad 1 - b \geq \Delta. \tag{C2}$$

Notice that if (Ci) holds, then $\eta(nT_i) \leq N^3\Delta^{-2}n^{-1}$.

If both (C1) and (C2) hold, then, by Theorems B and D,

$$\begin{aligned} |\mathcal{E}_n(u, v, z_0)| &= p(nT_1)p(nT_2) \\ &\leq n^4 T_1 T_2 \exp\{3\zeta(T_1^{1/3} + T_2^{1/3})n^{2/3}(1 + c_1 N^2 \Delta^{-1/2} n^{-1/4})\} \\ &\leq n^4 \exp\left\{3\zeta\left(1 - \frac{(1 - a - b)^2}{6}\right)n^{2/3}(1 + c_1 N^2 \Delta^{-1/2} n^{-1/4})\right\} \\ &\leq \exp\left\{3\zeta n^{2/3}\left(1 - \frac{\varepsilon^2}{30}\right)\right\}, \end{aligned}$$

if n is large enough (depending only on ε).

If both (C1) and (C2) fail, then, by Theorem C,

$$\begin{aligned} |\mathcal{E}_n(u, v, z_0)| &= p(nT_1)p(nT_2) \\ &\leq \exp\{c_2(T_1^{1/3} + T_2^{1/3})n^{2/3}\} \\ &\leq \exp\{2c_2 \Delta^{1/3}n^{2/3}\}, \end{aligned}$$

proving the claim in this case. Finally, when (C1) fails and (C2) holds (say)

$$\begin{aligned} |\mathcal{E}_n(u, v, z_0)| &= p(nT_1)p(nT_2) \\ &\leq n^2 T_2 \exp\{c_2 T_1^{1/3} n^{2/3} + 3\zeta T_2^{1/3} (1 + c_1 N^2 \Delta^{-1/2} n^{-1/4}) n^{2/3}\} \\ &\leq n^2 \exp\{(c_2 \Delta^{1/3} + 3\zeta (T_1^{1/3} + T_2^{1/3}) (1 + c_1 N^2 \Delta^{-1/2} n^{-1/4})) n^{2/3}\} \\ &\leq \exp\left\{3\zeta n^{2/3} \left(1 - \frac{\varepsilon^2}{30}\right)\right\}, \end{aligned}$$

if n is large enough (depending only on ε , again). □

Remark 2. We proved that most convex lattice curves lie between $(1 - \varepsilon)\Gamma$ and $(1 + \varepsilon)\Gamma$. Since the region between these two curves is much thinner around the endpoints of Γ than $U_\varepsilon(\Gamma)$, what we have is a stronger statement than Theorem 2. Actually, every convex lattice curve in \mathcal{E}_n^0 contains $(0, 1)$ and $(1, 0)$, so one can even shorten $(1 + \varepsilon)\Gamma$ at its endpoints.

Remark 3. In this proof we do not use the full strength of Theorem B, we only need the case when α, β go to infinity with their ratio bounded.

Remark 4. Notice that the proof is based on Theorems A–D, and so it works for any set of multipartitions that satisfy these theorems with the same constant in the exponent for Theorems A and B.

We mention further that the same method gives the following result which is needed for Theorem 1. Assume $0 < s, t$, and write $\mathcal{E}_n(s, t)$ for the set of all convex $(1/n)\mathbb{Z}^2$ -lattice curves connecting $(0, t)$ to $(s, 0)$ within the triangle $\text{conv}\{(0, t), (0, 0), (s, 0)\}$. Let $\Gamma(s, t)$ denote the curve with equation $\sqrt{x/s} + \sqrt{y/t} = 1$.

Theorem 3. *Assume $c < s/t < 1/c$ for some small positive constant c . Then, for every $\varepsilon > 0$, there is $n_0(\varepsilon, c)$ such that for $n > n_0$ the number of convex curves from $\mathcal{E}_n(s, t)$ not lying in $U_\varepsilon(\Gamma(s, t))$ is less than $\varepsilon |\mathcal{E}_n(s, t)|$.*

5. Proof of Theorem 1

This is quite simple. For a convex lattice polygon P in \mathcal{P}_n^0 , there is a smallest rectangle of the form $[x_1, x_2] \times [y_3, y_4]$ containing it. P has a vertex on each side of

this rectangle, we call it (x_i, y_i) , $i = 1, 2, 3, 4$. These points belong to $(1/n)\mathbb{Z}^2$. Collect all convex lattice polygons from \mathcal{P}_n with the same set of points $(x_1, y_1), \dots, (x_4, y_4)$ to form a class. The number of classes is bounded by n^8 . Using Theorems B and C it can easily be seen that such a class contains “many” elements if the four points are near the midpoints of the edges of the square, and “few” elements otherwise. According to Theorem 2 (or rather 3), almost all polygons in the large classes are very close to L .

6. Proof of Theorem A

Write $\mathcal{E}(m, n)$ for the set of convex lattice curves connecting $(0, 0)$ to (m, n) within the triangle $\text{conv}\{(0, 0), (m, 0), (m, n)\}$ where horizontal and vertical edges are not allowed. Set $q(m, n) = |\mathcal{E}(m, n)|$. Denote by \mathbb{P}^+ the set of primitive vectors with positive components. It is easy to find the generating function of $q(m, n)$:

$$\begin{aligned} f(X, Y) &= 1 + \sum_{m, n=1}^{\infty} q(m, n) X^m Y^n \\ &= \prod_{(i, j) \in \mathbb{P}^+} (1 + X^i Y^j + X^{2i} Y^{2j} + X^{3i} Y^{3j} + \dots) \\ &= \prod_{(i, j) \in \mathbb{P}^+} (1 - X^i Y^j)^{-1}. \end{aligned} \tag{6.1}$$

Now define $X = e^{-x}$ and $Y = e^{-y}$ and set

$$g(x, y) = \log f(X, Y). \tag{6.2}$$

We use a Tauberian theorem, Satz 1 of [M]. It says, in this particular case, that if $q(m, n)$ is monotone nondecreasing and

$$\log g(x, y) = \frac{A}{xy} (1 + o(1)),$$

where $o(1)$ is understood as $x, y \rightarrow 0$ so that x/y and y/x are bounded and $A > 0$ is a constant, then

$$\log q(m, n) = \sqrt[3]{Amn} (1 + o(1)),$$

where $o(1)$ is understood as $m, n \rightarrow \infty$ so that m/n and n/m are bounded.

Let us see first that q is monotone.

Claim 3. $q(m, n) \leq q(m + 1, n)$ and $q(m, n) \leq q(m, n + 1)$.

Proof. We prove the first inequality. We exhibit a map $\mathcal{E}(m, n) \rightarrow \mathcal{E}(m + 1, n)$ that maps distinct curves to distinct curves. Any $C \in \mathcal{E}(m, n)$ can be given by the

sequence of consecutive vertices

$$z_0 = (0, 0), \quad z_1, z_2, \dots, z_k = (m, n).$$

This curve is mapped to the one in $\mathcal{E}(m + 1, n)$ with vertices

$$z_0 = (0, 0), \quad z_1 + (1, 0), z_2 + (1, 0), \dots, z_k + (1, 0) = (m + 1, n). \quad \square$$

We need to know the behavior of g near $(0, 0)$. This is given in:

Lemma 1. *Let $0 < c < 1$ and assume $x, y \rightarrow 0$ so that $c < x/y < 1/c$. then*

$$g(x, y) = \frac{\zeta(3)}{\zeta(2)xy}(1 + o(1)).$$

We prove this lemma at the end of the section.

Meinardus’s theorem applies now giving

$$q(m, n) = \exp\left\{3\zeta^3\sqrt{mn}(1 + o(1))\right\}$$

as $m, n \rightarrow \infty$ with m/n and n/m bounded.

For the proof of Theorem A we have to consider $p(m, n)$, the number of all convex lattice polygons connecting $(0, n)$ to $(m, 0)$ within the triangle, horizontal and vertical edges are allowed. Clearly,

$$q(m, n) \leq p(m, n) = \sum_{m_1, n_1=0}^{m, n} q(m_1, n_1) \leq (m + 1)(n + 1)q(m, n),$$

since q is monotone. □

Proof of Lemma 1. This is a special case of Lemma 2 from the next section, but parts of this proof are used there. Assuming $x, y > 0$ we have

$$\begin{aligned} g(x, y) &= \log \sum_{(i, j) \in \mathbb{P}^+} (1 - e^{-ix - jy}) = \sum_{\mathbb{P}^+} \sum_{r=1}^{\infty} \frac{1}{r} e^{-(ix+jy)r} \\ &= \sum_{r=1}^{\infty} \frac{1}{r} \sum_{\mathbb{P}^+} e^{-(ix+jy)r} = \sum_{r=1}^{\infty} \frac{1}{r} \sum_{\mathbb{Z}_+^2} \sum_{d=1}^{\infty} \mu(d) e^{-(ix+jy)dr} \\ &= \sum_{r=1}^{\infty} \sum_{d=1}^{\infty} \frac{\mu(d)}{r} \frac{1}{e^{rdx} - 1} \frac{1}{e^{rdy} - 1}, \end{aligned} \tag{6.3}$$

where $\mu(d)$ is the Möbius function (see, e.g., [HW]). Notice that

$$\sum_{r=1}^{\infty} \sum_{d=1}^{\infty} \frac{\mu(d)}{d^2 r^3} = \sum_{r=1}^{\infty} \frac{1}{r^3} \sum_{d=1}^{\infty} \frac{\mu(d)}{d^2} = \frac{\zeta(3)}{\zeta(2)}, \tag{6.4}$$

see also [HW]. Then

$$\begin{aligned}
 g(x, y) - \frac{1}{xy} \frac{\zeta(3)}{\zeta(2)} &= \sum_{d,r} \frac{\mu(d)}{r} \left[\frac{1}{(e^{rdx} - 1)(e^{r dy} - 1)} - \frac{1}{d^2 r^2 xy} \right] \\
 &= \sum_{rd < M} \frac{\mu(d)}{r} [\dots] + \sum_{rd \geq M} \frac{\mu(d)}{r} [\dots] =: S(< M) + S(\geq M).
 \end{aligned}
 \tag{6.5}$$

Here M is chosen to be $\lfloor (x^2 + y^2)^{-1/2} \rfloor$. Using the inequality $e^t - 1 > t$ we infer

$$\begin{aligned}
 |S(\geq M)| &\leq \sum_{dr \geq M} \frac{1}{r} \left[\frac{1}{(drx)(dry)} + \frac{1}{d^2 r^2 xy} \right] \\
 &= \frac{2}{xy} \left(\sum_{r=1}^{M-1} \sum_{d=\lceil M/r \rceil}^{\infty} \frac{1}{d^2 r^3} + \sum_{r=M}^{\infty} \sum_{d=1}^{\infty} \frac{1}{d^2 r^3} \right) \\
 &\leq \frac{2}{xy} \left(\sum_{r=1}^{M-1} \frac{1}{r^3} \frac{1}{\lceil M/r \rceil} + \sum_{r=M}^{\infty} \frac{1}{r^3} \zeta(2) \right) \\
 &\leq \frac{2}{xy} \left(\sum_{r=1}^{M-1} \frac{1}{r^3} \frac{2r}{M} + \frac{\zeta(2)}{M} \sum_{r=M}^{\infty} \frac{1}{r^2} \right) \frac{4}{xyM} \sum_{r=1}^{\infty} \frac{1}{r^2} \\
 &= \frac{4\zeta(2)}{Mxy}.
 \end{aligned}
 \tag{6.6}$$

To estimate the other term we use the Taylor expansion

$$e^t = 1 + t + \frac{e^{\bar{t}} t^2}{2},
 \tag{6.7}$$

where $0 \leq \bar{t} \leq t$. When $t = drx$ or dry , $\bar{t} = dr\bar{x}$ or $dr\bar{y}$ with \bar{x}, \bar{y} depending on rd but we suppress this dependence in the notation. We get

$$\begin{aligned}
 &\frac{1}{(e^{rdx} - 1)(e^{r dy} - 1)} - \frac{1}{xy d^2 r^2} \\
 &= \frac{1}{(drx + e^{dr\bar{x}}(drx)^2/2)(dry + e^{dr\bar{y}}(dry)^2/2)} - \frac{1}{xy d^2 r^2} \\
 &= \frac{1}{xy d^2 r^2} \left(\frac{1}{(1 + e^{dr\bar{x}} drx/2)(1 + e^{dr\bar{y}} dry/2)} - 1 \right) \\
 &= \frac{1}{xy d^2 r^2} \frac{-(e^{dr\bar{x}} drx + e^{dr\bar{y}} dry)/2 - e^{dr\bar{x} + dr\bar{y}} d^2 r^2 xy/4}{(1 + e^{dr\bar{x}} drx/2)(1 + e^{dr\bar{y}} dry/2)}.
 \end{aligned}$$

Since $dr < M$ we have

$$\begin{aligned} \left| \frac{1}{e^{r dx} - 1} \frac{1}{e^{r dy} - 1} - \frac{1}{xy d^2 r^2} \right| &< \frac{1}{2xy dr} \left(e^{dr \bar{x}} x + e^{dr \bar{y}} y + \frac{e^{dr(\bar{x} + \bar{y})} dr xy}{2} \right) \\ &< \frac{e}{2xy dr} \left(x + y + \frac{Mxy}{2} \right) < \frac{3\sqrt{x^2 + y^2}}{xy dr}. \end{aligned} \tag{6.8}$$

Consequently,

$$\begin{aligned} |S(< M)| &< \frac{3\sqrt{x^2 + y^2}}{xy} \sum_{dr < M} \frac{1}{dr^2} < \frac{3\sqrt{x^2 + y^2}}{xy} \sum_{d=1}^M \frac{1}{d} \sum_{r=1}^{\infty} \frac{1}{r^2} \\ &< \frac{3\sqrt{x^2 + y^2}}{xy} (1 + \log M) \zeta(2). \end{aligned} \tag{6.9}$$

This, together with (4.7) proves Lemma 1 in a stronger form, namely,

$$\left| g(x, y) - \frac{\zeta(3)}{\zeta(2)xy} \right| \leq \text{const} \frac{\sqrt{x^2 + y^2}}{xy} \log(x^2 + y^2)^{-1/2}. \quad \square \tag{6.10}$$

Remark 5. Ruzsa showed [R], using the properties of the Möbius function, that the error term in the last statement is $o(\sqrt{x^2 + y^2})/xy$.

Remark 6. Theorem A can be proved with the saddle-point method as well (see [W] for a similar application). However, the technical difficulties arising here are lengthy and tiresome due to the presence of the Möbius function. When using this method, $g(x, y)$ should be computed or estimated when $x = \sqrt[3]{\zeta(3)n/\zeta(2)m^2}$ and $y = \sqrt[3]{\zeta(3)m/\zeta(2)n^2}$. Equation (6.10) shows, then, that (2.1) cannot fail as long as m/n^2 and n/m^2 tend to zero. It is not difficult to see, using the properties of ordinary partitions, that (2.1) holds, again, if $m > cn^2$ or if $n > cm^2$ for some large enough c . Thus (2.1) can fail only in the range $n = \text{const } m^2$.

Here is a sketch of Halász’s argument showing that (2.1) cannot hold in general. It works for rectangles $\text{conv}\{0, (0, n), (m, n), (m, 0)\}$ when $n = \varepsilon m^2$ for some small but fixed $\varepsilon > 0$. It is based on an application of the saddle-point method that is unusual as one of the circles of integration has radius $\approx \sqrt[3]{\varepsilon}$, i.e., does not tend to zero. Again, $g(x, y)$ has to be estimated when $x \approx \sqrt[3]{\varepsilon}$ and $y \rightarrow 0$, and the point in the argument is to show that $\lim_{y \rightarrow 0} xy g(x, y)$ is larger than $\zeta(3)/\zeta(2)$ for some small but fixed positive x . To see this first replace, in the last line of (6.3), $(e^{dxy} - 1)^{-1}$ by $(dry)^{-1}$, the change in $g(x, y)$ is small. The remaining function is

$$h(x) = \sum_{d, r \geq 1} \frac{\mu(d)}{dr^2} \frac{1}{e^{drx} - 1} = \sum_{k=1}^{\infty} e^{-kx} \sum_{rd|k} \frac{\mu(d)}{dr^2}.$$

The function $G(k) = \sum_{r|k} (\mu(d)/dr^2)$ is multiplicative. It follows that

$$\sum_{k \geq 1} \frac{G(k)}{k^s} = \frac{\zeta(s)\zeta(s+2)}{\zeta(s+1)}.$$

So we get

$$h(x) = \frac{1}{2\pi i} \int_{(\sigma)} \frac{\Gamma(s)\zeta(s)\zeta(s+2)}{x^{\frac{1}{2}\zeta(s+1)}} ds,$$

where $\Gamma(\cdot)$ is the gamma-function and the integral is taken along any vertical line with the real part $\sigma > 1$ in the complex plane. The function under the integral sign has a simple pole at $s = 1$ with residue $x^{-\frac{1}{2}\zeta(3)}/\zeta(2)$. At $s = 0$ the singularities cancel. Under the Riemann hypothesis, the line $\Re s = -\frac{1}{2}$ contains the next set of singularities, but even without the Riemann hypothesis, this line contains infinitely many singularities. It follows then that the function $h(x) - x^{-\frac{1}{2}\zeta(3)}/\zeta(2)$ oscillates as $x \rightarrow 0$, moreover, it is infinitely often larger than $x^{1/2+\delta}$ and infinitely often smaller than $-x^{1/2+\delta}$ as $x \rightarrow 0$ for any fixed positive δ . This shows that, for a suitable positive x , $h(x) > x^{-\frac{1}{2}\zeta(3)}/\zeta(2) + x^{1/2+\delta}$, proving that (2.1) does not hold in general. We omit the details.

7. Proof of Theorem B

Write $q_{a,b}(m, n) = q(m, n)$ for the number of convex lattice curves connecting $(0, 0)$ to (m, n) within the triangle $\text{conv}\{0, \alpha a, \alpha a + \beta b\}$ where edges parallel with a, b are not allowed and $(m, n) = \alpha a + \beta b \in \mathbb{Z}^2$. We drop the suffix referring to a, b whenever possible. The generating function for q is, just as in (6.1),

$$\begin{aligned} f(X, Y) &= 1 + \sum_{(m, n) \in \mathbb{P} \cap \text{pos}(a, b)} q(m, n) X^m Y^n \\ &= \prod_{(i, j) \in \mathbb{P} \cap \text{pos}(a, b)} (1 + X^i Y^j + (X^i Y^j)^2 + \dots) \\ &= \prod_{\mathbb{P} \cap \text{pos}(a, b)} (1 - X^i Y^j)^{-1}, \end{aligned}$$

where $\text{pos}(a, b) = \{x = sa + tb : s, t > 0\}$.

Observe that any $(i, j) \in \mathbb{P} \cap \text{pos}(a, b)$ can be written uniquely as

$$(i, j) = sa + tb + c, \quad \text{where } s, t = 0, 1, \dots,$$

and $c \in Q$ where

$$Q = \mathbb{Z}^2 \cap \{x = \gamma a + \delta b : 0 < \gamma, \delta \leq 1\}.$$

It is clear that $|Q| = |\det(a, b)|$. Further, every $c \in Q$ can be written uniquely as

$$c = \gamma(c)a + \delta(c)b,$$

where $\{\gamma(c): c \in Q\} = \{\delta(c): c \in Q\} = \{1/|Q|, 2/|Q|, \dots, 1\}$. This is well known and easy to see as well.

Now set $Z = (X, Y)$. With the new notation

$$X^i Y^j = Z^{(i, j)} = Z^{sa+tb+c} = (Z^a)^s (Z^b)^t Z^c.$$

Then, as in (6.3),

$$\begin{aligned} \log f(X, Y) &= \log \sum_{\mathbb{P} \cap \text{pos}(a, b)} (1 - X^i Y^j)^{-1} \\ &= \sum_{\mathbb{P} \cap \text{pos}(a, b)} \sum_{r=1}^{\infty} \frac{1}{r} (X^i Y^j)^r \\ &= \sum_{r=1}^{\infty} \frac{1}{r} \sum_{Z^2 \cap \text{pos}(a, b)} \sum_{d=1}^{\infty} \mu(d) (X^i Y^j)^{dr} \\ &= \sum_{d, r} \frac{\mu(d)}{r} \sum_{s=0}^{\infty} \sum_{t=0}^{\infty} \sum_{c \in Q} [(Z^a)^s (Z^b)^t Z^c]^{dr} \\ &= \sum_{d, r} \frac{\mu(d)}{r} \frac{\sum_{c \in Q} Z^{c dr}}{(1 - Z^{a dr})(1 - Z^{b dr})}. \end{aligned}$$

Now set $Z^a = e^{-x}$, $Z^b = e^{-y}$, and $g(x, y) = \log f(X, Y)$. Writing $z(c) = \gamma(c)x + \delta(c)y$ we have

$$g(x, y) = \sum_{d, r} \frac{\mu(d)}{r} \frac{\sum_{c \in Q} e^{-z(c) dr}}{(1 - e^{-drx})(1 - e^{-dry})}.$$

Lemma 2. For all $x, y > 0$ sufficiently small,

$$\left| g(x, y) - \frac{|Q| \zeta(3)}{xy \zeta(2)} \right| < \frac{50|Q|^2}{xy} \sqrt{x^2 + y^2} \log(x^2 + y^2)^{-1/2}.$$

Proof. Set $M = [(x^2 + y^2)^{-1/2}]$. In view of (6.5),

$$\begin{aligned} g(x, y) - \frac{|Q| \zeta(3)}{xy \zeta(2)} &= \sum_{d, r} \frac{\mu(d)}{r} \left[\frac{\sum_{c \in Q} e^{-z(c) dr}}{(1 - e^{-drx})(1 - e^{-dry})} - \frac{|Q|}{xy d^2 r^2} \right] \\ &= \sum_{dr < M} + \sum_{dr \geq M} =: S(< M) + S(\geq M). \end{aligned}$$

Observe that

$$\begin{aligned} \frac{e^{-\gamma(c)kx}}{1 - e^{-kx}} &= \frac{1}{1 - e^{-x}} \frac{e^{-\gamma(c)kx}}{1 + e^{-x} + \dots + e^{-(k-1)x}} \\ &\leq \frac{x}{1 - e^{-x}} \frac{1}{x} \frac{1}{e^{\gamma(c)kx} + e^{(\gamma(c)k-1)x} + \dots + e^{(\gamma(c)k - \lfloor \gamma(c)k \rfloor)x}} \leq \frac{2}{\gamma(c)kx}. \end{aligned}$$

Using this we get, just as in (6.6),

$$\begin{aligned} S(\geq M) &\leq \sum_{dr \geq M} \frac{1}{r} \left[\sum_Q \frac{2}{\gamma(c)} \frac{2}{\delta(c)} \frac{1}{d^2 r^2 xy} + \frac{|Q|}{d^2 r^2 xy} \right] \\ &= \frac{|Q|}{xy} \sum_{dr \geq M} \frac{1}{d^2 r^3} \left(1 + \frac{4}{|Q|} \sum_Q \frac{1}{\gamma(c)\delta(c)} \right) \\ &< \frac{16\zeta^2(2)|Q|^2}{xyM}, \end{aligned} \tag{7.1}$$

since $\sum_Q (\gamma(c)\delta(c))^{-1} \leq (\sum_Q \gamma(c)^{-2} \sum_Q \delta(c)^{-2})^{1/2} = \sum_Q \gamma(c)^{-2} \leq |Q|^2 \zeta(2)$.

For $S(< M)$ we use Taylor expansion (6.7) and the same notation as there (with $0 \leq \bar{z}(c) \leq z(c)$).

$$\begin{aligned} &\frac{\sum_{c \in Q} e^{-z(c)dr}}{(1 - e^{-drx})(1 - e^{-dry})} - \frac{|Q|}{xy d^2 r^2} \\ &= \frac{\sum_Q [1 - z(c)dr + e^{-\bar{z}(c)dr} (z(c)dr)^2 / 2]}{[drx - e^{-dr\bar{x}}(drx)^2 / 2][dry - e^{-dr\bar{y}}(dry)^2 / 2]} - \frac{|Q|}{d^2 r^2 xy} \\ &= \frac{|Q|}{(d^2 r^2 xy)(1 - e^{-dr\bar{x}} drx / 2)(1 - e^{-dr\bar{y}} dry / 2)} \\ &\quad \times \left[-\frac{dr}{|Q|} \sum_Q z(c) + \frac{d^2 r^2}{2Q} \sum_Q e^{-\bar{z}(c)dr} \bar{z}^2(c) \right. \\ &\quad \left. + \frac{dr}{2} (e^{-dr\bar{x}} \bar{x} + e^{-dr\bar{y}} \bar{y}) - \frac{d^2 r^2}{4} e^{-dr(\bar{x} + \bar{y})} xy \right]. \end{aligned}$$

Here $\sum_Q z(c) = (|Q| + 1)(x + y)/2$. Moreover, $\sum_Q z^2(c) \leq 3|Q|(x^2 + y^2)$ follows from the Cauchy-Schwarz inequality. Then for the square bracket in the last

expression we have

$$\begin{aligned} |[\dots]| &\leq dr \left[\left(\frac{1}{2} + \frac{|Q|+1}{2|Q|} \right) (x+y) + \frac{dr}{4}(x^2+y^2) + \frac{dr}{4}xy \right] \\ &\leq dr \left[3\sqrt{\frac{(x^2+y^2)}{2}} + dr(x^2+y^2) \right] \leq 4dr\sqrt{x^2+y^2}. \end{aligned}$$

Consequently, we have

$$\begin{aligned} |S(<M)| &= \left| \sum_{dr < M} \frac{\mu(d)}{r} [\dots] \right| \leq \frac{4|Q|\sqrt{x^2+y^2}}{xy} \sum_{dr < M} \frac{1}{dr^2} \\ &\leq \frac{4|Q|\sqrt{x^2+y^2}}{xy} \zeta(2)(1 + \log M), \end{aligned}$$

as we saw in (6.9). This, together with (7.1) proves the lemma. □

The usual method now follows. Recall that $\alpha a + \beta b = (m, n)$. Set

$$x = \sqrt[3]{\frac{|Q|\zeta(3)\beta}{\zeta(2)\alpha^2}} \quad \text{and} \quad y = \sqrt[3]{\frac{|Q|\zeta(3)\alpha}{\zeta(2)\beta^2}}$$

that give explicit values for X, Y . It is easy to see that, for $(i, j) \in \mathbb{Z}^2 \cap \text{pos}(a, b)$, $X^i Y^j$ is between 0 and 1 so that the generating function converges at (X, Y) . Then

$$\log q(m, n) X^m Y^n \leq g(x, y) \leq \frac{|Q|\zeta(3)}{xy\zeta(2)} \left(1 + 50|Q|\sqrt{x^2+y^2} \log(x^2+y^2)^{-1/2} \right).$$

Further, $X^m Y^n = (X^a)^\alpha (Y^b)^\beta = e^{-\alpha x - \beta y}$. This gives, finally,

$$\begin{aligned} \log q(m, n) &\leq \alpha x + \beta y + \frac{|Q|\zeta(3)}{xy\zeta(2)} \left(1 + 50|Q|\sqrt{x^2+y^2} \log(x^2+y^2)^{-1/2} \right) \\ &\leq 3\sqrt[3]{\frac{|Q|\zeta(3)\alpha\beta}{\zeta(2)}} (1 + 50|Q|\eta^{1/4}(T)) \\ &= 3\zeta\sqrt[3]{T} (1 + 50\det(T)\eta^{1/4}(T)). \end{aligned}$$

Finally, we estimate $p(m, n)$ where, when counting the convex lattice polygons, edges parallel to a or b are allowed. The proof method of Claim 3 shows that $q(m_1, n_1) \leq q(m, n)$ whenever $(m - m_1, n - n_1) = sa + tb$ for some $s = 0, 1, \dots$,

$[\alpha]$, $t = 0, 1, \dots, [\beta]$. Consequently,

$$\begin{aligned} p(m, n) &= \sum_{(m-m_1, n-n_1)=sa+tb} q(m-m_1, m-n_1) \\ &\leq \alpha\beta \max\{q(m_1, n_1): (m_1, n_1) \in \mathbb{Z}^2 \cap \{(m, n) - Q\}\} \\ &\leq T \exp\left\{3\xi^3\sqrt{T}(1 + 100 \det(T)\eta^{1/4}(T))\right\}. \quad \square \end{aligned}$$

References

- [A] V. I. Arnold, Statistics of integral lattice polytopes, *Funktsional Anal. Appl.* **14** (1980), 1–3.
- [BP] I. Bárány, J. Pach, On the number of convex lattice polygons, *Combin. Probab. Comput.* **1** (1992), 295–302.
- [BV] I. Bárány, A. M. Vershik, On the number of convex lattice polytopes, *GAFA J.* **2** (1992), 381–393.
- [B] W. Blaschke, *Vorlesungen über Differenzialgeometrie II. Affine Differenzialgeometrie*, Springer-Verlag, Berlin, 1923.
- [H] G. Halász, Private communication (1993).
- [HW] Hardy, E. M. Wright, *An Introduction to the Theory of Numbers*, Oxford University Press, Oxford, 1938.
- [LS] B. Logan, L. Schepp, A variational problem for random Young tableaux, *Adv. in Math.* **26** (1977), 206–222.
- [M] G. Meinardus, Zur additiven Zahlentheorie in mehreren Dimensionen, *Math. Ann.* **132** (1956), 333–346.
- [R] I. Z. Ruzsa, Private communication (1993).
- [S] Ya. G. Sinai, Probabilistic approach to analyse the statistics of convex polygonal curves, *Funktsional Anal. Appl.* **28** (1994), 41–48 (in Russian).
- [TS] P. Turán, M. Szalay, On some problems of the statistical theory of partitions with applications to characters of the symmetric group, *Acta Math. Hungar.* **29** (1977), 361–379.
- [V] A. M. Vershik, *Funktsional Anal. Appl.* **28** (1994), 16–25 (in Russian).
- [VK1] A. M. Vershik, S. V. Kerov, Asymptotic behavior of the Plancherel measure of the symmetric group and the limiting form of the Young tableaux, *Dokl. Akad. Nauk SSSR* **233** (1977), 1024–1027, (in Russian).
- [VK2] A. M. Vershik, S. V. Kerov, Asymptotic theory of the characters of the symmetric group, *Funktsional Anal. Appl.* **15** (1981), 246–255 (in Russian).
- [W] E. M. Wright, The number of partitions of a large bi-partite number, *Proc. London Math. Soc.* **7** (1957), 150–160.

Received December 27, 1993, and in revised form August 7, 1994.