Strictly Convex Drawings of Planar Graphs

Imre Bárány and Günter Rote

Received: July 13, 2005
Revised: June 21, 2006

Communicated by Günter Ziegler

Abstract. Every three-connected planar graph with n vertices has a drawing on an $O(n^2) \times O(n^2)$ grid in which all faces are strictly convex polygons. These drawings are obtained by perturbing (not strictly) convex drawings on $O(n) \times O(n)$ grids. Tighter bounds are obtained when the faces have fewer sides. In the proof, we derive an explicit lower bound on the number of primitive vectors in a triangle.

2000 Mathematics Subject Classification: Primary 05C62; Secondary 52C05.
Keywords and Phrases: Graph drawing, planar graphs.

1 Introduction

A *strictly convex* drawing of a planar graph is a drawing with straight edges in which all faces, including the outer face, are strictly convex polygons, i. e., polygons whose interior angles are less than 180°.

Theorem 1. (i) *A three-connected planar graph with n vertices in which every face has at most k edges has a strictly convex drawing on an $O(nw) \times O(n^2k/w)$ grid of area $O(n^3k)$, for any choice of a parameter w in the range $1 \le w \le k$.*

(ii) *In particular, every three-connected planar graph with n vertices has a strictly convex drawing on an $O(n^2) \times O(n^2)$ grid, and on an $O(n) \times O(n^3)$ grid.*

(iii) *For $k \le 4$, an $O(n) \times O(n)$ grid suffices.*

The drawings can be constructed in linear time.

When referring to a $W \times H$ grid of width W and height H, the constant hidden in the O-notation is on the order of 100 for the width and on the order of 10000 for the height. This is far too much for applications where one wants to draw graphs on a computer screen, for example. For the case $w = 1$, the bound is tighter: the grid size is approximately $14n \times 30n^2k$. For part (iii) of the theorem, the grid size is at most $14n \times 14n$, and if the outer face is a triangle, it is $2n \times 2n$.
The main idea of the proof is to start with a (non-strictly) convex embedding, in which angles of 180° are allowed, and to perturb the vertices to obtain strict convexity. We will use an embedding with special properties that is provided by the so-called *Schnyder embeddings*, which are introduced in Section 2.

HISTORIC CONTEXT. The problem of drawing graphs with straight lines has a long history. It is related to realizing three-connected planar graphs as three-dimensional polyhedra. By a suitable projection on a plane, one obtains from a polyhedron a straight-line drawing, a so-called *Schlegel diagram*. The faces in such a drawing are automatically strictly convex. By a projective transformation, it can be arranged that the projection along a coordinate axis is possible, and hence a suitable realization as a grid polytope gives rise to a grid drawing of the graph. However, the problem of realizing a graph as a polytope is more restricted: not every drawing with strictly convex faces is the projection of a polytope. In fact, there is an exponential gap between the known grid size for strictly convex planar drawings and for polytopes in space.
The approaches for realizing a graph as a polytope or for drawing it in the plane come in several flavors. The classical methods of Steinitz (for polytopes) and Fáry and Wagner (for graphs) work incrementally, making local modifications to the graph and adapting the geometric structure accordingly. Tutte [15, 16] gave a "one-shot" approach for drawing graphs that sets up a system of equations. This method yields also a polytope via the Maxwell-Cremona correspondence, see [11]. All these methods give embeddings that can be drawn on an integer grid but require an exponential grid size (or even larger, if one is not careful).
The first methods for straight-line drawings of graphs on an $O(n) \times O(n)$ grid were proposed for triangulated graphs, independently by de Fraysseix, Pach and Pollack [7] and by Schnyder [13]. The method of de Fraysseix, Pach and Pollack [7] is incremental: it inserts vertices in a special order, and modifies a partial grid drawing to accommodate new vertices. In contrast, Schnyder's method is another "one-shot" method: it constructs some combinatorial structure in the graph, from which the coordinates of the embedding can be readily determined afterwards. Both methods work in linear time. $O(n) \times O(n)$ is still the best known asymptotic bound on the size of planar grid drawings.
If graphs are not triangulated, the first challenge is to get faces which are convex. (Without the convexity requirement one can just add edges until the graph becomes triangulated, draw the triangulated supergraph and remove the extra edges from the drawing.) Many algorithms are now known that construct

convex (but not necessarily strictly convex) drawings with $O(n) \times O(n)$ size, for example by Chrobak and Kant [5] (à la Fraysseix, Pach and Pollack); or Schnyder and Trotter [14] and Felsner [8], see also [4] (à la Schnyder). Our algorithm builds on the output of Felsner's algorithm, which is described in the next section. Luckily, this embedding has some special features, which our algorithm uses.

The idea of getting a strictly convex drawing by perturbing a convex drawing was pioneered by Chrobak, Goodrich and Tamassia [6]. They claimed to construct strictly convex embeddings on an $O(n^3) \times O(n^3)$ grid, without giving full details, however. This was improved to $O(n^{7/3}) \times O(n^{7/3})$ in [12]. In this paper we further improve the "fine perturbation" step of [12] to obtain a bound of $O(n^2) \times O(n^2)$ for grid drawings. Theorem 1 gives better bounds when the faces have few sides, and we allow grids of different aspect ratios (keeping the same total area).

In the course of the proof, we need explicit (not just asymptotic) lower bounds on the number of primitive vectors in certain triangles. A primitive vector is an integer vector which is not a multiple of another integer vector; hence, primitive vectors can be used to characterize the directions of polygon edges. The existence of many short primitive vectors is the key to constructing strictly convex polygons with many sides. These lower bounds are derived in Section 5, based on elementary techniques from the geometry of numbers.

2 Preliminaries: Schnyder Embeddings of Three-Connected Plane Graphs

Felsner [8] (see also [9, 4]) has extended the straight-line drawing algorithm of Schnyder, which works for triangulated planar graphs, to arbitrary three-connected graphs. It constructs a drawing with special properties, beyond just having convex faces. These properties will be crucial for the perturbation step.

Felsner's algorithm works roughly as follows. The edges of the graph are covered by three directed trees which are rooted at three selected vertices a, b, c on the boundary, forming a *Schnyder wood*. The three trees define for each vertex v three paths from v to the respective root, which partition the graph into three regions. Counting the faces in each region gives three numbers x, y, z which can be used as barycentric coordinates for the point v with respect to the points a, b, and c. Selecting abc as an equilateral triangle of side length $f - 1$ (the number of interior faces of the graph) yields vertices which lie on a hexagonal grid formed by equilateral triangles of side length 1, see Figure 1a. Since $f \leq 2n$ this yields a drawing on a grid of size $2n \times 2n$.

This straight-line embedding has the following important property (see [8, Lemma 4 and Figure 11], [4, Fact 5]):

> *The Three Wedges Property.* Every vertex except the corners a, b, c has exactly one incident edge in each of the three closed 60° wedges shown in Figure 2a.

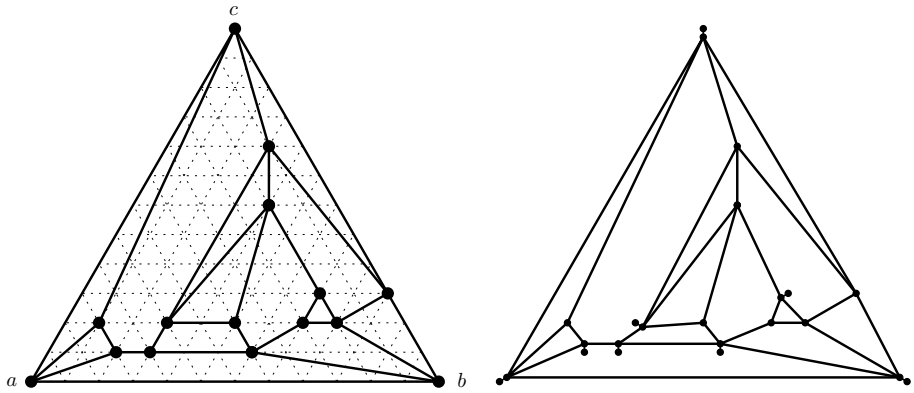


Figure 1: (a) A Schnyder embedding on a hexagonal grid and (b) on the refined grid after the initial (rough) perturbation

From this it follows immediately that there can be no angle larger than 180°, and hence all faces are convex. Moreover, it follows that the interior faces F have the *Enclosing Triangle Property*, see Figure 4a ([8, proof of Lemma 7], [4, Lemma 2]):

> *The Enclosing Triangle Property.* Consider the line $x = \text{const}$ through the point of F with maximum x-coordinate, and similarly for the other three coordinate directions. These three lines form a triangle T_F which encloses F. Then all vertices of F lie on the boundary of T_F, but F contains none of the vertices of T_F.

It follows that interior faces with $k \leq 4$ sides are already strictly convex. Throughout, we will call T_F the *enclosing triangle* of the face F.

The Schnyder wood and the coordinates of the points can be calculated in linear time. Recently, Bonichon, Felsner, and Mosbah [4], have improved the grid size to $(n-2) \times (n-2)$. However, the resulting drawing does not have the Three Wedges Property. An alternative algorithm for producing an embedding with a property similarly to the Enclosing Triangle Property is sketched in Chrobak, Goodrich and Tamassia [6]. It proceeds incrementally in the spirit of the algorithm of de Fraysseix, Pach and Pollack [7] and takes linear time. From the details given in [6] it is not clear whether the embedding has also the Three Wedges Property, which we need for our algorithm. The original algorithm of Chrobak and Kant [5] achieves a weak form of the Three Wedges Property, where F is permitted to contain vertices of T_F. Maybe, this algorithm can be modified to obtain the Three Wedges Property, at the expense of a constant-factor blow-up in the grid size.

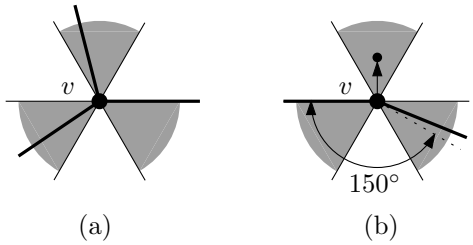


Figure 2: (a) Each closed shaded wedge contains exactly one edge incident to v. There may be additional edges in the interior of the white sectors. (b) A typical situation at a vertex which is perturbed.

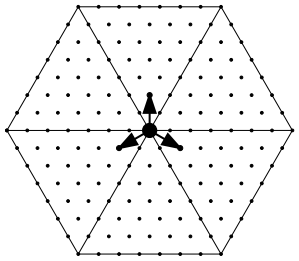


Figure 3: The three possible new positions for a single vertex in the rough perturbation. (Only the three boundary vertices a, b, c are pushed in directions opposite to these.)

3 Rough Perturbation

Before making all faces strictly convex, we perform an initial perturbation on a refined grid which is smaller by only a constant factor. This preparatory step will ensure that the subsequent "fine perturbation" can treat each face independently.

We overlay a triangular grid which is scaled by a factor of $1/7$, see Figures 3 and 5. A point may be moved to one of the three possible positions shown in Figure 3, by a distance of $\sqrt{3}/7$. The precise rules are as follows: A vertex v on an interior face F is moved if and only if the following two conditions hold.

(i) The interior angle of F at v is larger than 150° (including the possibility of a straight angle of 180°); and

(ii) v is incident to an edge of F which lies on the enclosing triangle T_F.

See Figure 2b for a typical case. Such a vertex is then pushed "out", perpendicular to the edge of T_F. We call the angle between the two edges incident

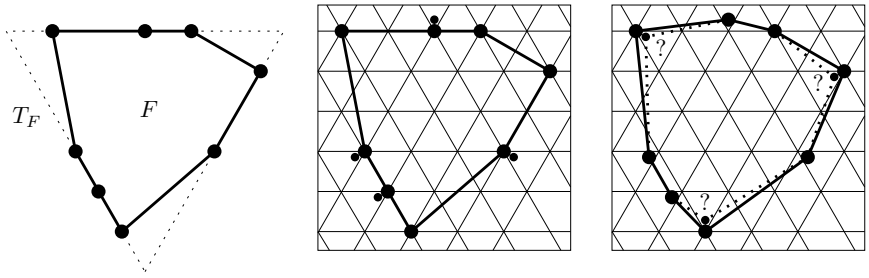


Figure 4: (a) A typical face F constructed by the convex embedding algorithm. (b) The new positions of the vertices of F which are pushed out are indicated. (c) The result of the rough perturbation. The perturbation of the vertices with question marks depends on the other faces incident to these vertices.

to F and v the *critical angle* of v. For a boundary vertex different from a, b, c, the exterior angle is the critical angle, but these vertices are not subject to the rough perturbation. The three corners a, b, and c are treated specially: they are pushed straight *into* the triangle by the rough perturbation, as illustrated in Figure 1. Examples can be seen in Figure 4b–c and Figure 5. The result of

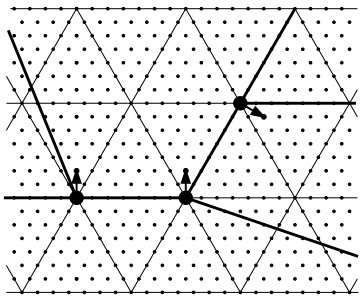
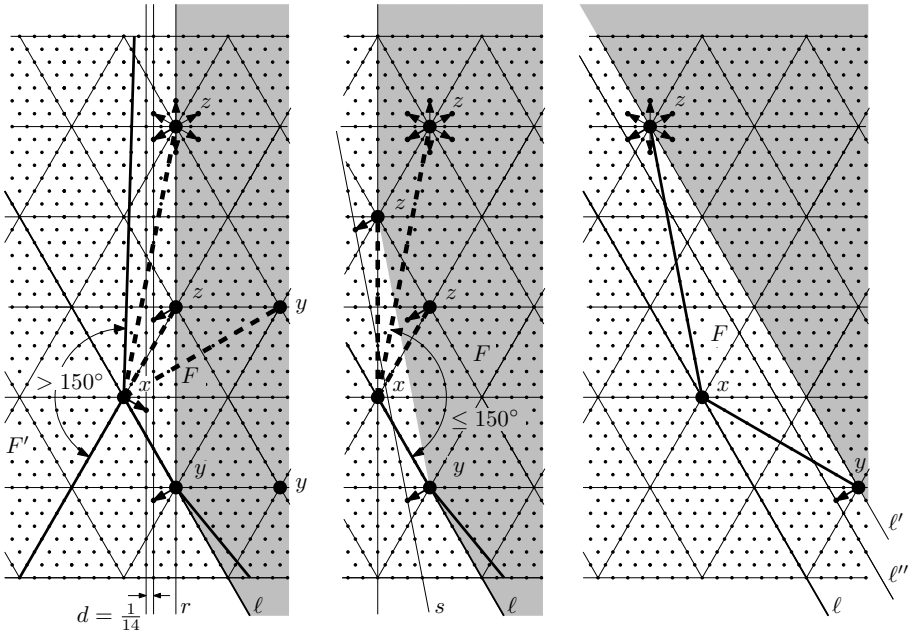


Figure 5: Example of the rough perturbation.

perturbing the example in Figure 1a is shown in Figure 1b.
There can be no conflict in applying the rules by regarding a vertex v as part of different faces: the bound of 150° on the angle, together with the Three Wedges Property ensures that there is at most one critical angle for every vertex (Figure 2b).
The result has the following properties:

LEMMA 1. *After the rough perturbation, all faces are still convex.*
Moreover, if each vertex is additionally perturbed within a disk of radius $1/30$*, the only concave angle that might arise at a vertex* v *is the critical angle of* v*.*



(a) Case I (b) Case IIa (c) Case IIb

Figure 6: The cases in the proof of Lemma 1. The figures show possible locations for the neighbors y and z of x.

Proof. It is evident that no critical angle can become bigger than 180°. For non-critical angles, this is also easy to see (cf. Figure 4c). (In fact, the second statement is a strengthening of this claim.)

We now prove this second statement of the lemma by considering different cases. The reader who is satisfied with the existence of *some* small enough perturbation bound $\varepsilon > 0$ may skip the rest of the proof. We continue to show that we can choose $\varepsilon = 1/30$.

Consider a non-critical angle yxz at a vertex x in a face F. We assume without loss of generality that x lies on the *lower left* edge ℓ of the enclosing triangle T_F.

Case I. The point x is incident to a critical angle of another face F', and thus x is pushed out of F'.

Without loss of generality, we can assume that x lies on the lower right edge of $T_{F'}$, and thus x is perturbed in the lower right direction, as in Figure 6a. (The other case, when x lies on the upper edge of $T_{F'}$ and is pushed vertically upward, is symmetric.) By the definition of critical angles, the angle in F' must be bigger than 150°. This excludes from F all points vertically above x or to the left of x. The upper neighbor z of x, which is a grid point, is therefore restricted to a closed halfplane right of a vertical line r at distance $1/2$ from x. The lower neighbor y must lie on or above the line ℓ that bounds

the enclosing triangle T_F. Thus, y and z are restricted to the shaded area in Figure 6a. Even if all three points are perturbed by the rough perturbation, they are still separated by a vertical strip of width $d = \frac{1}{2} - 2 \cdot \frac{3}{14} = \frac{1}{14}$. An additional perturbation of $\frac{1}{30} < \frac{1}{2 \cdot 14}$ cannot make the angle at x larger than $180°$.

Case II. The point x not perturbed by the initial perturbation.

Case IIa. The point x has a neighbor on ℓ.

We can assume w.l.o.g. that it is the lower neighbor y, see Figure 6b. The angle yxz must be at most $150°$ because otherwise x would be critical. It means that z cannot lie to the left of x, and thus y and z are restricted to the shaded area in Figure 6b. Even if they are perturbed, they remain above the line s, which is obtained by offsetting the edge of the shaded region that is closest to x. The distance from x to s is $1/7 \cdot \sqrt{3/7} \approx 0.0935 > \frac{2}{30}$. Thus, there is enough space to additionally perturb the points x, y and z without creating a concave angle. (Actually, the vertex x will not even be perturbed in the fine perturbation.)

Case IIb. The point x has no neighbors on ℓ, see Figure 6c.

This means that y and z lie on or beyond the next grid line ℓ' parallel to ℓ. The rough perturbation can move them closer to ℓ, but they remain beyond another parallel line ℓ'' whose distance from x is $5/7 \cdot \sqrt{3/4} \approx 0.618$. This leaves plenty of space for additional perturbations of x, y, and z. □

After the rough perturbation, we will subject every vertex v that is incident to a critical angle to an additional small perturbation of a distance at most $1/30$. The lemma ensures that, in order to achieve convexity at v without destroying convexity at another place, we only have to take care of *one* incident face when we decide the final perturbation of v. We can thus work on each face independently to make it strictly convex.

4 Fine Perturbation

We will now discuss how we go about achieving strict convexity of all faces. The rough perturbation helps us to reduce this task to the case of regularly spaced points on a line (Section 4.1). In Section 4.2, we will describe in detail how the perturbed strictly convex chain is constructed for this special case.

4.1 The Setting after the Rough Perturbation

After the rough perturbation, we are in the following situation. Consider a maximal chain $v_2, v_3, \ldots, v_{K-1}$ of successive critical angles on a face F. These angles must be made strictly convex by perturbing them inside their little disks. (The two extreme angles at v_2 and v_{K-1} might already be convex.) The vertices $v_2, v_3, \ldots, v_{K-1}$ lie originally on a common edge of the enclosing triangle T_F, We first discuss the case when the vertices lie on the upper edge ℓ of T_F, forming a horizontal chain, as in Figure 7a. (The extension to the other two cases is discussed in Section 4.3.) According to Lemma 1 we have to

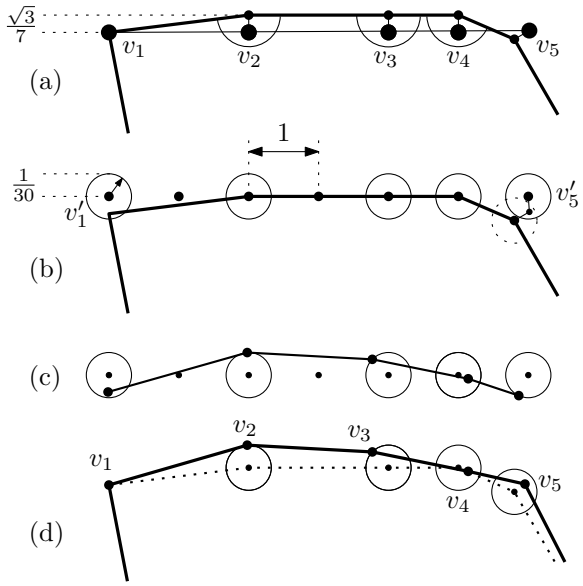


Figure 7: The setting of the fine perturbation process: (a) The initial situation after the rough perturbation. The angles in which it is necessary to ensure a convex angle are marked. (b) The circles in which the fine perturbation is performed. The size of the circles is exaggerated to make the perturbation more conspicuous. (c) A strictly convex polygon inside the circles. (d) The final result.

ensure that these critical angles are smaller than 180° after the perturbation. In Figure 7a, these are the vertices v_2, v_3, and v_4. Let us call these vertices *critical vertices*. In addition, we look at the two adjacent vertices v_1 and v_K on F. By the choice of a maximal chain, they are not critical for F. They may lie on the same line as the critical vertices, as the vertices v_1 and v_5 in Figure 7a, or they might lie below this line. To guide the perturbation of the points $v_2, \ldots, v_{K-1}$, we pretend that v_1 and v_K are part of the chain, and we create *surrogate positions* v'_1 and v'_K for these neighbors: First we move them from their original positions vertically upward to ℓ; if they don't land on a grid point, we move them outward by 1/2 unit. Since the angles at v_2 and v_{K-1} are bigger than 150°, we are sure that $v'_1, v_2, \ldots, v_{K-1}, v'_K$ lie on ℓ in this order. Finally, we subject v'_1 and v'_K to the same rough perturbation as the critical vertices between them, and move them vertically upward.

We place a disk of radius 1/30 around every perturbed point on this edge, including the two surrogate positions, see Figure 7b. In the next step, to be described in Section 4.2, we find a strictly convex chain which selects one vertex out of each little disk, as shown in Figure 7c.

This will make all angles at $v_2, \ldots, v_{K-1}$ strictly convex. Finally, we use these

perturbed positions for our critical vertices, but for v_1 and v_K, we ignore their perturbed surrogate positions, see Figure 7d. The true position of v_1 or v_K may be determined by a different face in which it forms a critical angle (as is the case for v_5 in the example), or it might just keep its original position (like v_1 in the example). We only have to check that the angle at the left-most and right-most critical vertex (v_2 and v_4 in this case) remains convex:

LEMMA 2. *Replacing the perturbed surrogate position v_1' and v_K' of the points v_1 and v_K by their true positions does not destroy convexity at their neighbors v_2 and v_{K-1} in F.*

Proof. We first show that the rough perturbation does not actually perturb v_1 and v_K to their surrogate positions v_1' or v_K'. It is conceivable that, say, v_1 lies on ℓ and is perturbed upwards because of its critical angle in a different face F', see Figure 8. However, this would contradict the Three Wedges Property for v_1 and F, creating two incident edges in a sector in which only a unique incident edge can exist.

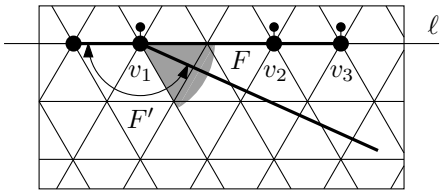
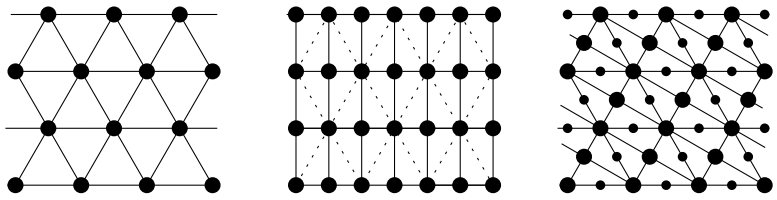


Figure 8: A neighbor of a critical vertex cannot be perturbed in the same direction.

Thus we conclude that v_1 and v_K lie below or on ℓ, and they are either perturbed not at all or in a direction below ℓ.
Vertices v_2 and v_4 in the example of Figure 7 represent the possible extreme cases that have to be considered. v_5 represents a vertex that is pushed downward in the rough perturbation, and then subjected to a fine perturbation anywhere in its little circle. For visual clarity, the circles in Figure 7 have been drawn with a much larger radius than 1/30. Since the circles are actually small enough, the angle at v_4 will be convex no matter where the point v_5 is placed in its own circle. (This position is determined when the critical face of v_5 is considered.) A similar statement holds at v_2, where the perturbed surrogate position of v_1 in Figure 7c is replaced by the *original* position of v_1; this will always turn the edge v_2v_1 counterclockwise and thus preserve convexity at v_2. The argument works also for a chain of vertices on an exterior edge of the enclosing triangle. In this case, $v_2, v_3, \ldots, v_{K-1}$ are perturbed around their original position on ℓ, whereas the neighbors v_1 and v_K are moved inside the triangle and below ℓ. Geometrically, the situation looks similar as for vertex v_1 in Figure 7, except that v_1 is not pushed down straight but at a $-30°$ angle. This movement is large enough to ensure convexity at v_2. □



(a) (b) (c)

Figure 9: The hexagonal grid (a) is contained in a rectangular grid (b). A hexagonal grid twice refined (c) contains rectangular grids in three different directions. One of these rectangular grids is highlighted by thicker points.

4.2 Convex Chains in the Grid

We have a number K of vertices $0 = a_1 < a_2 < \cdots < a_K \leq 2n-1$ on a horizontal line which form part of an array of $2n$ consecutive grid points. We want to perturb them into convex position. If the faces of the embedding have at most k sides, then $K \leq k$. It is more convenient to work with a rectangular grid. So we extend the hexagonal grid to a rectangular grid as shown in Figure 9. This grid will be refined sufficiently in order to allow a strictly convex chain to be drawn inside a sequence of circles. Figure 10 gives a schematic picture of the situation. (This drawing is not to scale.) It is more convenient to discuss the construction of an *upward* convex chain. Inside each disk (of radius 1/30) we fit a square of side length 1/50, which is subdivided into a subgrid of width w and height h. More precisely, we are looking for a sequence of points $p_i = (x_i, y_i)$ in these circles, whose coordinates measure the distance from the lower left corner of the first circle in units of little grid cells. Two successive circle centers at distance 1 in terms of the original grid have a distance of $S := 50w$ when measured in subgrid units. Thus we are looking for integer coordinates that satisfy $a_i \cdot S \leq x_i \leq a_i \cdot S + w$ and $0 \leq y_i \leq h$. Eventually, when the whole subgrid is scaled to the standard grid $\mathbb{Z} \times \mathbb{Z}$, x_i and y_i will become true distances again. The total size of the resulting integer grid will be $O(nw) \times O(nh)$.

The convex chain $p_1, p_2, \ldots, p_K$ has a descending part up to a point with minimum y-coordinate and an ascending part. We choose the two points with minimum y-coordinate to lie in the middle: We define $M := \lfloor K/2 \rfloor + 1$ and set $y_{M-1} = y_M = 0$. We will only describe the construction of the ascending chain from p_M to the right. The left half is constructed symmetrically.

The direction between two grid points is uniquely specified by a *primitive vector*, a vector whose components are relatively prime. We now take a sequence of primitive vectors $q_1, q_2, \ldots, q_{K-M}$, $q_i = (u_i, v_i)$ with $0 < u_i \leq w$ and $v_i > 0$, in order of increasing slope v_i/u_i. Then we choose the difference vectors Δp as appropriate multiples of these vectors, in the following way. We have already

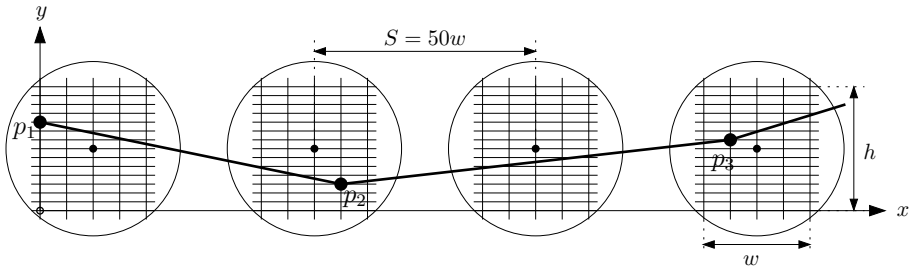


Figure 10: A convex chain formed by grid points in the circles. (Again, the radius of the circles is drawn much too large compared to their distance.)

defined $y_M := 0$, and we choose x_M arbitrarily within the permitted range of x-coordinates. Having defined p_{M+i-1}, we define

$$p_{M+i} := p_{M+i-1} + s \cdot q_i$$

by adding as many copies of q_i as are necessary to bring x_{M+i} into the desired box:

$$a_{M+i} \cdot S \le x_{M+i} \le a_{M+i} \cdot S + w$$

Since this box has width w, and $u_i \le w$, this is always possible.
We need $K - M \le K/2$ primitive vectors q_i (including the vector $(1, 0)$ from p_{M-1} to p_M.) The following theorem ensures that we can find these vectors in a triangle of sufficiently large area.

THEOREM 2. *The right triangle $T = (0, 0), (w, 0), (w, t)$, where $w \ge 1$, w integer, and $t \ge 2$, contains at least $wt/4$ primitive vectors.*

The general proof is given in Section 5. We can however easily give an explicit solution for the special case $t = 2$ (corresponding to the choice $w = k$ below, which leads to the most balanced grid dimensions): In this case, we can simply take the $1 + \lfloor w/2 \rfloor$ vectors $(w, 1)$, $(w - 1, 1)$, ..., $(\lceil w/2 \rceil, 1)$.
We use Theorem 2 as follows. We choose an arbitrary width $w \le k$ for the boxes. By Theorem 2, we can set $t := \max\{2, 2K/w\}$ to ensure that we find at least $K/2$ primitive vectors in the triangle T. The slope of these vectors is bounded by t/w. Let us estimate the necessary height h of the boxes. The last point p_K is connected to p_M by a chain of vectors with slope at most t/w. The distance of x-coordinates is at most the width of the whole grid on which the graph is embedded, i. e., at most $S \cdot 2n = O(wn)$; hence the difference in y-coordinates is at most $t/w \cdot O(wn) = O(tn) = O(kn/w)$. It follows that the height h of the boxes is $O(kn/w)$. The total height of the resulting grid is $O(hn) = O(kn^2/w)$.
This leads to part (i) of Theorem 1. Part (ii) is an easy corollary. As an extreme case, we can set $w = 1$ and perform only vertical perturbations. We get $h \le 2kn$ (without any additional constants depending on S).

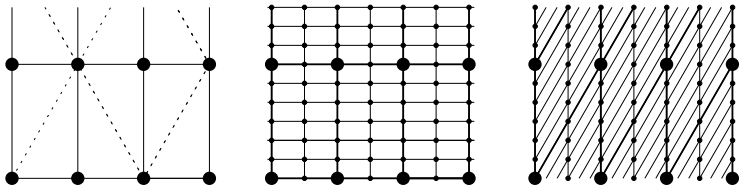


Figure 11: A rectangular grid (a), its 2×6 refinement (b), and a shearing (c) of the refined grid. Its grid-points coincide with the untransformed grid.

4.3 Perturbation of Vertices on Diagonal Lines

So far, we have treated only a sequence of vertices on a horizontal straight line. The same scheme can be applied to lines of the two other directions by applying the shearing transformation $\binom{x}{y} \mapsto \binom{x}{y+\sqrt{3}/2 \cdot x}$ or $\binom{x}{y} \mapsto \binom{x}{y-\sqrt{3}/2 \cdot x}$ which moves points only in vertical direction. If h is a multiple of w, the transformation will produce a grid like in Figure 11c which is contained in the original grid of Figure 11b. For the range of parameters which is interesting for the theorem ($w \leq k$), the height h of the subgrid is never smaller than the width w; thus, the choice of h as a multiple of w does not change the asymptotic analysis. One needs to reduce the size of the little square subgrid to ensure that the sheared square still fits inside the circle, and one has to adjust the quantity S accordingly. In addition, we have to select h and w as multiples of 14, to accommodate the grid of the rough perturbation and the refined rectangular grid of Figure 9b. All of this changes the analysis only by a constant factor.

For the case of a uniform stretching of both dimensions ($w = h$), one referee has pointed out a simpler alternative method. After a blow-up by a factor of two, the original triangular grid contains rectangular grids in all three grid directions, Figure 9c. Two further refinements by the factor 7 (for the rough perturbation) and then by the factor w are sufficient to accommodate the fine perturbation. On the exterior edges, the points must of course be perturbed to form an *outward* convex chain.

For part (iii) of the theorem we have already mentioned that interior faces with $k \leq 4$ sides are already strictly convex. If the outer face has 4 edges, it contains a single vertex on one of the sides of the outer triangle. The rough perturbation is thus sufficient to make the outer face strictly convex.

The whole procedure, as described above, is quite explicit and can be carried out with a linear number of arithmetic operations. We calculate the $O(k)$ primitive vectors q_i only once and store them in an array. Then, for every actual sequence of vertices on an edge, we can construct the perturbation very easily. The primitive vectors in the triangle $(0,0)$, $(w,0)$, (w,t) according to Theorem 2 can be selected from the $O(wt) = O(k)$ grid points in linear time with a sieve method.

	optimal		greedy	
$w = h$	n	$(w+1)/n$	n	$(w+1)/n$
0	2	0.5000	2	0.5000
1	4	0.5000	4	0.5000
2	6	0.5000	6	0.5000
4	10	0.5000	8	0.6250
6	14	0.5000	12	0.5833
8	16	0.5625	14	0.6429
10	20	0.5500	18	0.6111
12	22	0.5909	18	0.7222
20	32	0.6562	28	0.7500
40	58	0.7069	48	0.8542
100	122	0.8279	96	1.0521
200	212	0.9481	164	1.2256
400	366	1.0956	276	1.4529
1,000	758	1.3206	562	1.7811
2,000	1,292	1.5488	948	2.1108
4,000	2,206	1.8137	1,610	2.4851
10,000	4,468	2.2384	3,230	3.0963
20,000	7,592	2.6345	5,472	3.6552
40,000			9,250	4.3244
100,000			18,484	5.4101
200,000			31,192	6.4119
400,000			52,626	7.6008
1,000,000			105,012	9.5227
2,000,000			177,046	11.2965
4,000,000			299,494	13.3559

Table 1: The length of the longest strictly convex n-gon in a sequence of square cells of size $w \times w$, regularly spaced at distance $S = 50w$.

4.4 Numerical Experiments

We have presented a general systematic solution for finding a convex chain by selecting grid-points from a sequence of boxes. One can find the *optimal* (i.e., longest) convex chain in polynomial time by dynamic programming, as described in more detail below. Results of some experiments are shown in the first column of Table 1. We restrict ourselves to the standard situation of selecting an n-gon from n adjacent boxes ($K = n$) which are squares ($w = h$). For several different sizes w, we computed the largest n such that a strictly convex n-gon can be found in a sequence of cells of size $w \times w$. The factor $(w+1)/n$ determines the necessary grid size w in terms of n. (By the convention of Figure 10, a "$w \times w$" grid consists of $(w+1)^2$ vertices; thus we give the fraction $(w+1)/n$ instead of w/n.) Since the convex chain consists of a monotone

decreasing and a monotone increasing part, connected by a horizontal segment in the middle, the necessary height $w+1$ is at least $0.5\,n$. We see that this trivial lower bound is achieved for small values of n. The factor $(w+1)/n$ increases with n, but not very fast. (The rectangular $w \times h$ boxes constructed in the proof of Theorem 1 would have $w/n = 1$, but $h/n = 100$.)

The dynamic programming algorithm computes, for each point p in the $w \times h$ box, and for each possible previous point p' in the adjacent box to the left, the longest ascending and strictly convex chain (of length i) for which $p_{i-1} = p'$ and $p_i = p$. Knowing p' and p, it can be determined which points in the next box are candidate endpoints p_{i+1} of a chain of length $i+1$. One can argue that, among these points p_{i+1} that are reachable as a continuation of $p'p$, only the $w+1$ lowest points on each vertical line are candidates for endpoints p_{i+1} that form part of an optimal chain. Theoretically, the complexity of this algorithm is therefore $O(w^3h^2)$. It turns out that, with few exceptions, every point p has only one predecessor point p' that must be considered: all other predecessor points p_{i-1} have either a larger slope of the vector $p - p_{i-1}$ or they are reached by a shorter chain. Therefore, the algorithm runs in $O(w^2h) = O(wkn)$ time, in practice.

A simple greedy approach for selecting the points p_i one by one gives already a very good solution: we choose p_{i+1} from the possible grid points in the appropriate box in such a way that the segment $p_{i+1} - p_i$ has the slope as small as possible while still forming a convex angle at p_i. The results in the right column of Table 1 indicate that this algorithm is quite competitive with the optimum solution. The running time is $O(kw)$.

5 Grid Points in a Triangle

In this section we prove Theorem 2. We denote by $\mathbb{P} := \{\, (x,y) \mid \gcd(x,y) = 1 \,\}$ the set of primitive vectors in the plane.

It is known that the proportion of primitive vectors among the integer vectors in some large enough area is approximately $1/\zeta(2) = 6/\pi^2$ [10, Chapters 16–18]. Thus, a "large" triangle T should contain roughly $3/\pi^2 \cdot wt \approx 0.304wt$ primitive points. However, for very wide or very high triangles, the fraction of primitive vectors may be different. In fact, for $t = 2$, the bound $wt/4$ is tight except for an additive slack of at most 2.

We will use special methods for counting primitive vectors when T is "very high" (i. e., w is fixed and below some threshold and t is unbounded, Section 5.1), when T is "very wide" (t is fixed and w is unbounded, Section 5.2), and for the case when both t and w are large (Section 5.3). We use the help of the computer for the first two cases, but we use a general bound for the last case.

5.1 FIXED WIDTH, UNBOUNDED HEIGHT

For a fixed value of w, the function $f(t) := |T \cap \mathbb{P}|$ can be analyzed explicitly. It is periodically ascending:

$$f(t+w) = f(t) + C,$$

where $C = \sum_{i=1}^{w} \phi(i)$ is the number of primitive vectors in the triangle $(0,0)$, $(w,0)$, (w,w), excluding the point $(1,1)$. Euler's totient function $\phi(i)$ denotes the number of integers $1 \le j \le i$ that are relatively prime to i, or equivalently, the number of primitive vectors (i,j) on the vertical line segment from $(i,0)$ to $(i,i-1)$.

The reason for the periodic behavior is that the unimodular shearing transformation $(x,y) \mapsto (x,y+x)$ maps the triangle $(0,0)$, $(w,0)$, (w,t), to the triangle $(0,0)$, (w,w), $(w,w+t)$, which is equal to $(0,0)$, $(w,0)$, $(w,w+t)$ minus the triangle $(0,0)$, $(w,0)$, (w,w).

Therefore, it is sufficient to check that the "average slope" C/w of f is bigger than $w/4$, and to check

$$f(t) \ge tw/4 \tag{1}$$

for the initial interval $2 \le t \le 2+w$. This can be done by computer: We sort all primitive vectors (x,y) with $0 \le x \le w$ and $0 \le y/x \le (w+2)/w$ by their slope y/x. We gradually increase t from 2 to $w+2$. The critical values of t for which (1) must be checked explicitly are when a new primitive vector is just about to enter the triangle.

We ran a lengthy computer check to establish (1) for $w = 1, 2, \ldots, 250$ and for $2 \le t \le w+2$ (and hence for all t). In addition, we checked it for the range $w = 251, 252, \ldots, 800$ and for $2 \le t \le 250$.

5.2 LARGE WIDTH

In this section we prove Theorem 2 for small t and large w. T intersects each horizontal line $y = i$ in a segment of length $w - (w/t)i$. In any set of i consecutive grid points on this line, there are precisely $\phi(i)$ primitive vectors. We can subdivide the grid points on $y = i$ into $\lfloor (w-(w/t)i)/i \rfloor \ge w/i - w/t - 1$ groups of i consecutive points, leading to a total of at least $(w/i - w/t - 1)\phi(i)$ primitive vectors:

$$|T \cap \mathbb{P}| \ge 1 + \sum_{i=1}^{\lfloor t \rfloor} \Big(\frac{w}{i} - \frac{w}{t} - 1 \Big) \phi(i)$$

For a given value of $\lfloor t \rfloor$, one can evaluate the expression

$$|T \cap \mathbb{P}| \ge 1 + \sum_{i=1}^{\lfloor t \rfloor} \Big(\frac{w}{i} - \frac{w}{t} - 1 \Big) \phi(i) \ge 1 + \sum_{i=1}^{\lfloor t \rfloor} \Big(\frac{w}{i} - \frac{w}{\lfloor t \rfloor} - 1 \Big) \phi(i) \tag{2}$$

explicitly. The right-hand side of this bound is a linear function $g(w)$:

$$g(w) = \sum_{i=1}^{\lfloor t \rfloor} \Big(\frac{w}{i} - \frac{w}{\lfloor t \rfloor} - 1 \Big) \phi(i)$$

For example, for $\lfloor t \rfloor = 130$, we have $g(w) = w \cdot 39.514\ldots - 5153$. It follows that $g(w) > w \cdot 131/4 > wt/4$ for $w \geq 762$. Performing this calculation by computer for $\lfloor t \rfloor = 6, 7, \ldots, 130$ establishes Theorem 2 for $6 \leq t \leq 130$ and $w \geq 800$. The interval $4 \leq t < 6$ can be split into the ranges $4 \leq t < 4.5$, $4.5 \leq t < 5$, $5 \leq t < 5.5$, and $5.5 \leq t < 6$. For each range, we can use the above method with a tighter bound in (2) than $t \geq \lfloor t \rfloor$, and the estimate goes through in the same way.

So let us consider the remaining interval $2 \leq t \leq 4$: For $2 \leq t < 3$, we can evaluate $|T \cap \mathbb{P}|$ explicitly:

$$|T \cap \mathbb{P}| = 1 + (w + 1 - \lceil \tfrac{w}{t} \rceil) + (\lceil \tfrac{w}{2} \rceil - \lceil \tfrac{w}{t} - \tfrac{1}{2} \rceil), \tag{3}$$

counting the primitive vectors on the lines $y = 0$, $y = 1$, and $y = 2$, respectively. For $t \geq 3$, the right-hand side of (3) is still valid as a lower bound. We get

$$|T \cap \mathbb{P}| \geq 1 + \big(w + 1 - (\tfrac{w}{t} + 1) \big) + \big(\tfrac{w}{2} - (\tfrac{w}{t} - \tfrac{1}{2} + 1) \big) > w(\tfrac{3}{2} - \tfrac{2}{t})$$

The last expression is $\geq wt/4$ for $2 \leq t \leq 4$.

Thus we have proved the theorem for $2 \leq t \leq 130$ and $w \geq 800$.

5.3 Large triangles

Lemma 3. *Let T' be an axis-aligned right triangle of width a' and height b', whose right angle lies on a grid point. Then*

$$\operatorname{area} T' \leq |T' \cap \mathbb{Z}^2| \leq \operatorname{area} T' + \lfloor a' \rfloor + \lfloor b' \rfloor + 1$$

Proof. This is simple. Suppose the right angle is at the right bottom corner of T', see Figure 12a. Each lattice point in T' is the right bottom vertex of a unit square and these squares cover T'. To bound the area from below, we must subtract the squares which are not contained in T'. These squares form a monotone chain along the longest side of T', and their number is $\lfloor a' \rfloor + \lfloor b' \rfloor + 1$. □

Lemma 4. *Let T be the right triangle $(0,0)$, $(a,0)$, (a,b), with $a, b \geq 1$. Define T^* as $T \cap \{ (x,y) : y \geq 1 \}$. Then*

$$\frac{ab}{2} - a - b + \frac{a}{2b} \leq |T^* \cap \mathbb{Z}^2| \leq \frac{ab}{2} + b - \frac{a}{2b} \tag{4}$$

In particular,

$$\Big| |T^* \cap \mathbb{Z}^2| - \frac{ab}{2} \Big| \leq a + b$$

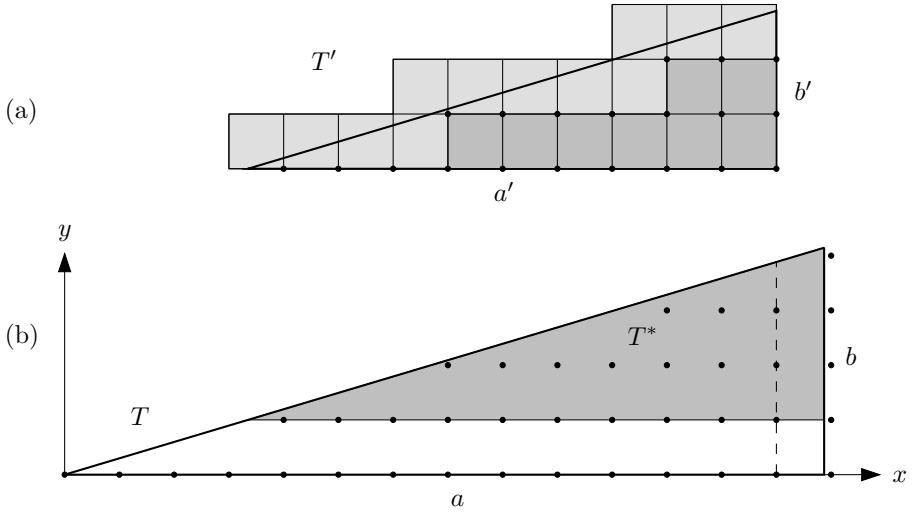


Figure 12: (a) The triangle T' in Lemma 3 and its covering by squares. (b) The triangles T and T^* (shaded) in Lemma 4.

Proof. See Figure 12b. The triangle T^* has length $a - a/b$, height $b - 1$ and area $\frac{1}{2}(b-1)(a-a/b) = ab/2 - a + a/(2b)$. Let T' denote the part of T^* that lies left of the line $x = \lfloor a \rfloor$. This triangle contains the same grid points as T^*. We assume first that T' is a nonempty triangle. The difference in areas lies in a rectangle strip of width < 1 and height $b - 1$:

$$\operatorname{area} T^* - (b-1) \le \operatorname{area} T' \le \operatorname{area} T^*$$

We can apply Lemma 3 to T' and obtain

$$|T^* \cap \mathbb{Z}^2| = |T' \cap \mathbb{Z}^2| \le \Big(\frac{ab}{2} - a + \frac{a}{2b}\Big) + \Big(a - \frac{a}{b}\Big) + (b-1) + 1,$$
$$|T^* \cap \mathbb{Z}^2| = |T' \cap \mathbb{Z}^2| \ge \operatorname{area} T' \ge \Big(\frac{ab}{2} - a + \frac{a}{2b}\Big) - (b-1),$$

from which the lemma follows.
The triangle T' may not exist, as in Figure 13. In this case, $T^* \cap \mathbb{Z}^2 = \emptyset$. Instead of arguing why the above derivation is valid also for this case, we establish the inequalities directly. Let $b' \ge b - b/a$ denote the vertical extent of T at $x = \lfloor a \rfloor$. Then the fact that T' is empty is equivalent to $b' < 1$.
Then, from $1 \ge b' \ge b - b/a$ we conclude that $ab < a + b$. It follows that the lower bound in (4) is at most 0:

$$\frac{ab}{2} + \frac{a}{2b} - a - b \le \frac{a+b}{2} + \frac{a}{2} - a - b \le 0$$

The claimed upper bound in (4) is always nonnegative, by the assumption $b \ge 1$. □

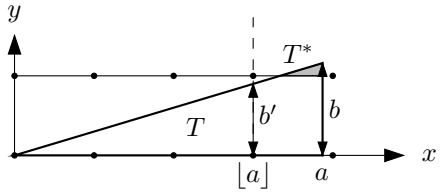


Figure 13: If T^* (shaded) contains no grid points, the triangle T' does not exist.

The number of primitive vectors can be estimated by an inclusion-exclusion formula, taking into account vectors which are multiples of single primes $2, 3, 5, 7, \ldots$, vectors which are jointly multiples of two primes, of three primes, and so on, see [10, Chapters 16–18]:

$$|T \cap \mathbb{P}| = 1 + |T^* \cap \mathbb{P}| = 1 + \sum_{i=1}^{S} \mu(i) \cdot \left|(\tfrac{1}{i} \cdot T^*) \cap \mathbb{Z}^2\right| = 1 + \sum_{i=1}^{S} \mu(i) \cdot \left|(\tfrac{1}{i} T)^* \cap \mathbb{Z}^2\right| \tag{5}$$

Here, $\mu(i)$ is the Möbius function: $\mu(i) = (-1)^k$ if i is the product of k distinct primes and $\mu(i) = 0$ otherwise. It is known that $\sum_{i=1}^{\infty} \frac{\mu(i)}{i^2} = 1/\zeta(2) = 6/\pi^2$, leading to the fact mentioned above that a fraction of approximately $6/\pi^2$ of the grid points in a large area are primitive vectors.
Our sum in (5) goes to $i = \infty$, but for $i > w$ or $i > t$, the set $(\frac{1}{i}T)^* \cap \mathbb{Z}^2$ is empty. Therefore, the formula is valid for $S := \min\{w, \lfloor t \rfloor\}$. We apply Lemma 4 and obtain

$$\begin{aligned} |T \cap \mathbb{P}| = 1 + \sum_{i=1}^{S} \mu(i) \cdot \left|(\tfrac{1}{i}T)^* \cap \mathbb{Z}^2\right| &\geq \frac{wt}{2} \sum_{i=1}^{S} \frac{\mu(i)}{i^2} - \sum_{i=1}^{S} \frac{w+t}{i} \\ &\geq \frac{wt}{2}\left(\frac{6}{\pi^2} - \frac{1}{S}\right) - H_S(w+t), \end{aligned}$$

where $H_S = 1+1/2+1/3+\cdots+1/S$ is the harmonic number. The last inequality comes from bounding the remainder $\sum_{i=S+1}^{\infty} \mu(i)/i^2 \leq \sum_{i=S+1}^{\infty} 1/i^2 < 1/S$ of the infinite series, whose value is $6/\pi^2$.
We distinguish the two cases for S: Case 1: $w \leq t$, and $S = w$. Then

$$|T \cap \mathbb{P}| \geq \frac{wt}{2}\left(\frac{6}{\pi^2} - \frac{1}{w}\right) - H_w(2t) = wt\left(\frac{3}{\pi^2} - \frac{1}{2w} - \frac{2H_w}{w}\right)$$

Case 2: $w \geq t$, and $S = \lfloor t \rfloor$.

$$\begin{aligned} |T \cap \mathbb{P}| &\geq \frac{wt}{2}\left(\frac{6}{\pi^2} - \frac{1}{\lfloor t \rfloor}\right) - H_{\lfloor t \rfloor}(w+t) \\ &\geq wt\left(\frac{3}{\pi^2} - \frac{1}{2(t-1)} - H_{\lfloor t \rfloor}\left(\frac{1}{t} + \frac{1}{w}\right)\right) \end{aligned} \tag{6}$$

Combining the two cases and setting $n := \min\{w, t\}$ gives

$$|T \cap \mathbb{P}| \geq wt \left(\frac{3}{\pi^2} - \frac{1}{2(n-1)} - \frac{2H_{\lfloor n \rfloor}}{n} \right)$$

Using the estimate $H_i \leq \gamma + \ln(i+1)$ with Euler's constant $\gamma \approx 0.57721$, it can be checked that this factor is bigger than $1/4$ for $n \geq 250$, thus proving the theorem for $w, t \geq 250$.
On the other hand, the factor in (6) is bigger than $1/4$ for $w \geq 800$ and $130 \leq t \leq 250$, proving the theorem also for this range.

WRAP-UP. The proof of Theorem 2 is now complete. On a high level, we distinguish three ranges for w: $1 \leq w \leq 250$, $251 \leq w \leq 800$, and $w \geq 800$.

- Range 1: For $1 \leq w \leq 250$, the theorem has been established in Section 5.1.
- Range 2: $251 \leq w \leq 800$. For $251 \leq w \leq 800$ and $1 \leq t \leq 250$, the theorem has been established in Section 5.1 as well. For $251 \leq w \leq 800$ and $t \geq 250$, it has been proved in Section 5.3.
- Range 3: Finally, for $w \geq 800$, there is a division into three cases: Section 5.2 takes care of the range $2 \leq t \leq 130$. Section 5.3 proves the bound separately for the ranges $130 \leq t < 250$ and $t \geq 250$. □

6 CONCLUSION

In practice, the algorithm behaves much better than indicated by the rough worst-case bounds that we have proved. We have not attempted to optimize the constants in the proof. For example, if we don't take a 7×7 subgrid but an 11×11 subgrid, and with a more specialized treatment of the outer face, the permissible amount of perturbation in Lemma 1 increases from $1/30$ to $1/9$, but it would make the pictures of the rough perturbation harder to draw.
Bonichon, Felsner, and Mosbah [4] have used a technique of eliminating edges from the drawing that can later be inserted in order to reduce the necessary grid size for (non-strictly) convex drawings. This technique can also be applied in our case: remove interior edges as long as the graph remains three-connected. These edges can be easily reinserted in the end, after all faces are strictly convex. (For non-strictly convex drawings in [4], the selection of removable edges and their reinsertion is actually a more complicated issue.) This technique might be useful in practice for reducing the grid size.

LOWER BOUNDS. The only known lower bound comes from the fact that a single convex n-gon on the integer grid needs $\Omega(n^3)$ area, see Bárány and Tokushige [3], or Acketa and Žunić [1, 2] for the easier case of a *square* grid. To achieve this area for an n-gon, one has to draw it in a quite round shape.

In contrast, the faces that are produced in our algorithm have a very restricted shape: when viewed from a distance, the look like the triangles, quadrilaterals, pentagons, or hexagons of the $n \times n$ grid drawing from which they were derived. To reduce the area requirement below $O(n^4)$ one has to come up with a new approach that also produces faces with a "rounder" shape.

Our bounds are however, optimal within the restricted class of algorithms that start with a Schnyder drawing or an arbitrary non-strictly convex drawing on an $O(n) \times O(n)$ grid and try to make it strictly convex by *local perturbations* only. Consider the case where $n-1$ vertices lie on the outer face, connected to a central vertex in the middle. The Schnyder drawing will place these vertices on the enclosing triangle, and at least $n/3$ vertices will lie on a common line. They have to be perturbed into convex position, as in Figures 7 or 10.

Let us focus on the standard situation when we want to perturb n equidistant vertices on a line, at distance 1 from each other. The $n-1$ edge vectors $p_{i+1}-p_i$ lie in a $2w \times 2h$ box; they must be non-parallel, and in particular, they must be *distinct*. If Δy is the average absolute vertical increment of these vectors, it follows that $\Delta y = \Omega(n/w)$, and the total necessary height h of the boxes is $\Omega(n(\Delta y)) = \Omega(n^2/w)$. Therefore, the total necessary area is $\Omega(hwn^2) = \Omega(n^4)$. The argument can be extended to the case when only $\Omega(n)$ selected grid vertices on a line of length $O(n)$ have to be perturbed. It can also be shown that our bounds in terms of k are optimal in this setting. The worst case occurs when there is a line of length n with $\Omega(k)$ consecutive grid points in the middle and two vertices at the extremes.

Extensions. The class of three-connected graphs is not the most general class of graphs which allow strictly convex embeddings. The simplest example of this is a single cycle. A planar graph, with a specified face cycle C as the outer boundary, has a strictly convex embedding if and only if it is three-connected to the boundary, i. e., if every interior vertex (not on C) has three vertex-disjoint paths to the boundary cycle. Equivalently, the graph becomes three-connected after adding a new vertex and connecting it to every vertex of C. These graphs cannot be treated directly by our approach, since the Schnyder embedding method of Felsner [8] does not apply. Partitioning the graph into three-connected components and putting them together at the end might work.

Acknowledgements. We thank the referees for helpful remarks which have lead to many clarifications in the presentation. Imre Bárány was partially supported by Hungarian National Science Foundation Grants No. T 037846 and T 046246.

References

[1] Dragan M. Acketa and Jovisa D. Žunić, *On the maximal number of edges of convex digital polygons included into a square grid*, počítače a umelá

inteligencia 1 (1982), no. 6, 549–558.

[2] ———, *On the maximal number of edges of convex digital polygons included into an $m \times m$-grid*, J. Comb. Theory, Ser. A 69 (1995), 358–368.

[3] Imre Bárány and Norihide Tokushige, *The minimum area of convex lattice n-gons*, Combinatorica 24 (2004), no. 2, 171–185.

[4] Nicolas Bonichon, Stefan Felsner, and Mohamed Mosbah, *Convex drawings of 3-connected plane graphs*, Graph Drawing: Proc. 12th International Symposium on Graph Drawing (GD 2004), September 29–October 2, 2004 (New York), János Pach, ed. Lecture Notes in Computer Science, vol. 3383, Springer-Verlag, 2005, pp. 60–70.

[5] M. Chrobak and G. Kant, *Convex grid drawings of 3-connected planar graphs*, Internat. J. Comput. Geom. Appl. 7 (1997), no. 3, 211–223.

[6] Marek Chrobak, Michael T. Goodrich, and Roberto Tamassia, *Convex drawings of graphs in two and three dimensions*, Proc. 12th Ann. Sympos. Comput. Geom., 1996, pp. 319–328.

[7] H. de Fraysseix, J. Pach, and R. Pollack, *How to draw a planar graph on a grid*, Combinatorica 10 (1990), no. 1, 41–51.

[8] Stefan Felsner, *Convex drawings of planar graphs and the order dimension of 3-polytopes*, Order 18 (2001), 19–37.

[9] ———, *Geodesic embeddings and planar graphs*, Order 20 (2003), 135–150.

[10] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, Oxford Science Publications, 1979.

[11] Jürgen Richter-Gebert, *Realization Spaces of Polytopes*, Lecture Notes in Mathematics, vol. 1643, Springer-Verlag, 1997.

[12] Günter Rote, *Strictly convex drawings of planar graphs*, Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), Vancouver, 2005, pp. 728–734.

[13] W. Schnyder, *Embedding planar graphs on the grid*, Proc. 1st ACM-SIAM Sympos. Discrete Algorithms, 1990, pp. 138–148.

[14] W. Schnyder and W. T. Trotter, *Convex embeddings of 3-connected plane graphs*, Abstracts of the AMS 13 (1992), no. 5, 502.

[15] W. T. Tutte, *Convex representations of graphs*, Proceedings London Mathematical Society 10 (1960), no. 38, 304–320.

[16] ———, *How to draw a graph*, Proceedings London Mathematical Society 13 (1963), no. 52, 743–768.

Imre Bárány
Rényi Institute
Hungarian Academy of Sciences
PoB 127
Budapest 1364, Hungary
barany@renyi.hu
and
Department of Mathematics
University College London
Gower Street
London WC1E 6BT, UK

Günter Rote
Freie Universität Berlin
Institut für Informatik
Takustraße 9
D-14195 Berlin, Germany
rote@inf.fu-berlin.de