ORIENTATION PRESERVING MAPS OF THE SQUARE GRID

Imre Bárány,[*] *Attila Pór,*[†] *and Pavel Valtr*[‡]

Abstract. For a finite set $A \subset \mathbb{R}^2$, a map $\varphi : A \to \mathbb{R}^2$ is *orientation preserving* if for every non-collinear triple $u, v, w \in A$ the orientation of the triangle u, v, w is the same as that of the triangle $\varphi(u), \varphi(v), \varphi(w)$. We prove that for every $n \in \mathbb{N}$ and for every $\varepsilon > 0$ there is $N = N(n, \varepsilon) \in \mathbb{N}$ such that the following holds. Assume that $\varphi : G(N) \to \mathbb{R}^2$ is an orientation preserving map where $G(N)$ is the grid $\{(i, j) \in \mathbb{Z}^2 : -N \le i, j \le N\}$. Then there is an affine transformation $\psi : \mathbb{R}^2 \to \mathbb{R}^2$ and $z_0 \in \mathbb{Z}^2$ such that $z_0 + G(n) \subset G(N)$ and $\|\psi \circ \varphi(z) - z\| < \varepsilon$ for every $z \in z_0 + G(n)$. This result was previously proved in a completely different way by Nešetřil and Valtr, without obtaining any bound on N. Our proof gives $N(n, \varepsilon) = O(n^4\varepsilon^{-2})$.

1 Introduction

This paper is about orientation preserving maps of the $n \times n$ grid. We denote by $G(N)$ the grid $\{(i, j) \in \mathbb{Z}^2 : -N \le i, j \le N\}$ and by $G^*(n)$ the grid $\{(i, j) \in \mathbb{Z}^2 : 0 \le i, j \le n - 1\}$. A map $\varphi : G(N) \to \mathbb{R}^2$ is *orientation preserving* if for every non-collinear triple $u, v, w \in G(N)$ the orientation of the triangle u, v, w is the same as that of the triangle $\varphi(u), \varphi(v), \varphi(w)$, or with a formula

$$\operatorname{sign}\det\begin{bmatrix} u & v & w \\ 1 & 1 & 1 \end{bmatrix} = \operatorname{sign}\det\begin{bmatrix} \varphi(u) & \varphi(v) & \varphi(w) \\ 1 & 1 & 1 \end{bmatrix}.$$

We are going to show that given an orientation preserving map $\varphi : G(N) \to \mathbb{R}^2$ there is an $n \times n$ subgrid of $G(N)$ whose image under φ is very close to an affine image of the $n \times n$ grid, provided that N is large enough (polynomial in n and $1/\varepsilon$). More precisely, we have the following result where the norm $\|\ \|$ is Euclidean.

Theorem 1. *For every $n \in \mathbb{N}$ and for every $\varepsilon > 0$ there is $N = N(n, \varepsilon)$ such that if $\varphi : G(N) \to \mathbb{R}^2$ is an orientation preserving map, then there is an affine transformation $\psi : \mathbb{R}^2 \to \mathbb{R}^2$ and $z_0 \in \mathbb{Z}^2$ such that $z_0 + G^*(n) \subset G(N)$ and for every $z \in z_0 + G^*(n)$.*

$$\|\psi \circ \varphi(z) - z\| < \varepsilon.$$

Here $N(n, \varepsilon) = O(n^4\varepsilon^{-2})$.

[*] *Alfréd Rényi Institute of Mathematics, Budapest Hungary, and Department of Mathematics University College London, UK,* barany.imre@renyi.hu

[†] *Department of Mathematics, Western Kentucky University, Bowling Green, KY, USA,* attila.por@wku.edu

[‡] *Department of Applied Mathematics, Charles University, Praha, Czech Republic,* valtr@kam.mff.cuni.cz

Theorem 1 without the explicit bound on $N(n,\varepsilon)$ was already proved by Nešetřil and Valtr [5, Lemma 10] as the key tool for proving several Ramsey-type results. The proof in the paper [5] relied on repeated compactness arguments, thus it could give no upper bound on N. Our bound $N(n,\varepsilon) = O(n^4\varepsilon^{-2})$ makes ground for giving explicit bounds for Ramsey-type results given in the paper [5]; see concluding remark (1) on page 105 of [5] where the lack of an explicit bound is discussed. From the (discrete and) computational geometry point of view, the most interesting consequences of our bound $N(n,\varepsilon) = O(n^4\varepsilon^{-2})$ in Theorem 1 might be those which are connected with the study of order types, as described in the next section.

Remark 1. The function $N(n,\varepsilon)$ in Theorem 1 satisfies the lower bound $N(n,\varepsilon) = \Omega(n^2\varepsilon^{-1})$. The example showing this is given in the last section.

2 Connections to order types and motivation

An *order type* of size n is an equivalence class of all n-point sets which can be mapped into each other by strongly order preserving maps, where a map $\varphi : A \to \mathbb{R}^2$ from a finite planar point set A to $\mathbb{R}^2$ is *strongly orientation preserving* if it is orientation preserving and, additionally, it maps collinear triples of A to collinear triples. If the sets of an order type are in general position then we say that the order type is *in general position*. Order types have been studied from various perspectives, for example, see the paper of Goodman and Pollack [1] for a classical result and the recent paper of Pilz and Welzl [6] for further references.

The *span* of a finite point set $A \subset \mathbb{R}^2$ is the ratio between the maximum distance in A and the minimum distance in A. Note that due to projective transformations the supremum of the spans of the sets of any fixed order type (of size at least three) is ∞. We define the *span* of an order type T as the infimum of the spans of the point sets in T. By famous results of Goodman, Pollack and Sturmfels [2] and of Kratochvíl and Matoušek [3], there are order types of size n with double exponential span.

Theorem 2. *For $n > 1$, let $f(n)$ be the smallest real number such that, for any order type T of size n in general position and for any $\delta > 0$, there exists a set A in T having the span smaller than $f(n) + \delta$. Then there are two positive constants c_1 and c_2 such that, for any integer $n > 3$,*

$$2^{2^{c_1 n}} \leq f(n) \leq 2^{2^{c_2 n}}.$$

Our Theorem 1 considers subsets of sets of some order type with a small span. In particular, an immediate consequence of Theorem 1 says that some order types have the property that any set of this order type contains a rather large subset whose affine image has a very small span (asymptotically as small as possible for the given size).

Theorem 3. *For any $N \geq 2$, there is an order type T_N of size N in general position such that any set A of T_N contains a subset B of size $n = \Omega(N^{1/3})$ which is an affine transform of a set having span $O(\sqrt{n})$.*

We remark that due to a simple packing argument the span of any set (or order type) of size $n \geq 2$ is at least $\Omega(\sqrt{n})$.

Another (almost immediate) consequence of Theorem 1 says that there are order types T of arbitrary size $n \geq 2$ in general position such that any set A of order type T contains a quite large subset of points which lie, one by one, in small neighborhoods of equidistantly distributed points along some line.

Theorem 4. *For any $N \geq 2$ and any $\varepsilon > 0$, there is an order type T_N of size N in general position such that any set A of T_N contains a subset B of size $n = \Omega(N^{1/4}\varepsilon^{1/2})$ such that for some line ℓ and for some n equally distributed points $p_1, \ldots, p_n$ on ℓ where the distance between p_i and p_{i+1} is exactly d for some fixed $d > 0$ and for each $i = 1, \ldots, n-1$, the following holds. There is exactly one point of B in the (εd)-neighborhood of p_i for each $i = 1, \ldots, n$.*

Since some of the ratios of distances among sufficiently many equidistantly distributed points on a line approximate (with any prescribed precision) l prescribed distance ratios, Theorem 4 immediately implies the following result of Nešetřil and Valtr [5, Theorem 6].

Theorem 5 (Nešetřil and Valtr [5])**.** *For any positive integer $l > 0$ and for any $l+1$ positive real numbers $\varepsilon, r_1, r_2, \ldots, r_l > 0$, there exists a (finite) order type T in general position such that any set of order type T determines $l+1$ distances $d_i, i = 0, 1, 2, \ldots, l$, such that $|\frac{d_i}{d_0} - r_i| < \varepsilon$ $(i = 1, 2, \ldots, l)$.*

3 Preparations and sketch of proof

We start with introducing basic notation and definitions. For distinct $u, v \in \mathbb{R}^2$, $L(u, v)$ denotes the line they span. The angle $\alpha(u, v)$ is defined as the angle the vector $v - u$ and the positive half of the x axis make. It is understood mod 2π.

Assume $\varphi_0 : G^*(n) \to \mathbb{R}^2$ is an orientation preserving map with $\varphi_0(0, 0) = (0, 0)$. Define $e, f \in \mathbb{R}^2$ via $\varphi_0(n-1, 0) = (n-1)e$ and $\varphi_0(0, n-1) = (n-1)f$. Suppose further that for all $u, v \in G^*(n)$ with $\alpha(u, v) \in \{0, \pi/4, \pi/2\}$

$$|\alpha(u, v) - \alpha(\varphi_0(u), \varphi_0(v))| < \gamma, \tag{1}$$

where $\gamma > 0$. For the proof of Theorem 1 we need the following lemma.

Lemma 1. *Assume $\gamma = O(n^{-2})$. Then, under the above conditions for every $(i, j) \in G^*(n)$ we have*

$$\|\varphi_0(i, j) - (ie + jf)\| < 10\gamma n^2(\|e\| + \|f\|).$$

The proof is given in Section 8.

An important notion is that of a *block* of a grid. The horizontal block H_i of $G(m)$ $(i = -m, \ldots, m)$ is the set of the lattice points on the segment $[(-m, i), (m, i)]$, its first and last points are $(-m,i),(m,i)$. The vertical block V_j for $(j=-m,\dots,m)$ is the set of lattice points on the segment $[(j,-m),(j,m)]$ with its first and last points defined analogously.

Similarly, the plus diagonal block D_i^+ of $G(m)$ is the set of lattice points of $G(m)$ on the line $x-y=i$ $(i=0,\pm1,\dots,\pm2m)$, and the minus diagonal block D_j^- is the set of lattice points of $G(m)$ on the line $x+y=i$ $(i=0,\pm1,\dots,\pm2m)$. Their first and last points are defined similarly. Two blocks are neighbourly if they lie on consecutive parallel lattice lines.

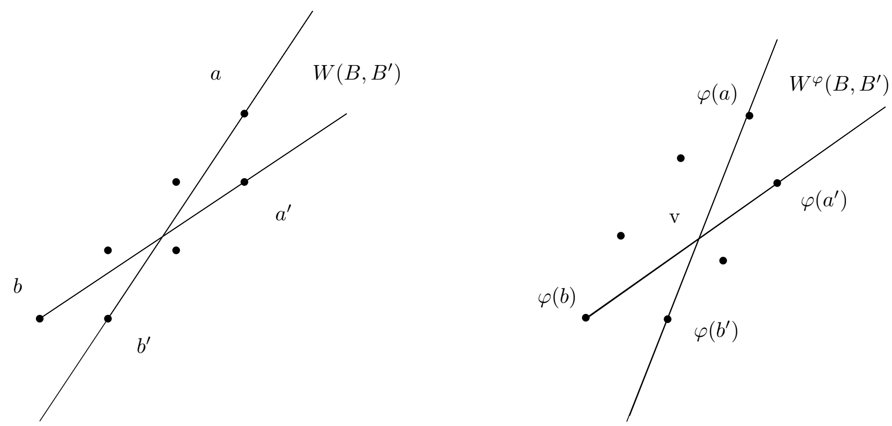

Figure 1: Neighbourly blocks and φ blocks separated

Given an orientation preserving map $\varphi : G(m)\to\mathbb{R}^2$ the image $\varphi(B)$ of a block B is called a φ block. We need separation properties (in the weak sense) of blocks and φ blocks. Let B and B' be two neighbourly blocks with distinct first and last point a,b and a',b', respectively. Here $b-a$ and $b'-a'$ are parallel and point in the same direction, see Figure 1. It is clear that both $L(a,b')$ and $L(a',b)$ separate B and B'. The orientation preserving properties of φ imply that the lines

$$L_1 = L(\varphi(a),\varphi(b')) \text{ and } L_2 = L(\varphi(a'),\varphi(b))$$

also separate $\varphi(B)$ and $\varphi(B')$, or, what is the same, conv $\varphi(B)$ and conv $\varphi(B')$. The lines L_1 and L_2 define a double cone (or wedge) $W^\varphi(B,B')$ with apex $v=L_1\cap L_2$ which is the double cone not containing $\varphi(B)$ and $\varphi(B')$. Similarly, let $W(B,B')$ be the double cone determined by $L(a,b')$ and $L(a',b)$, again the one not containing B and B'. The following simple facts are well known.

Fact 1. If u,w are in different components of the wedge $W^\varphi(B,B')$, then the line through v and parallel with $L(u,w)$ separates $\varphi(B)$ and $\varphi(B')$.

Fact 2. If $z_1,z_2\in G(m)$ are in different components of the wedge $W(B,B')$, then $\varphi(z_1)$ and $\varphi(z_2)$ lie in different components of the wedge $W^\varphi(B,B')$.

Remark 2. If $T:\mathbb{R}^2\to\mathbb{R}^2$ is an affine map (with positive determinant), then we can replace φ by $T\circ\varphi$ and this new map will be orientation preserving and so satisfy the conditions of Theorem 1. In the next section we will "normalize" φ by choosing a suitable T.

Proof sketch. Here is a quick sketch of the proof of Theorem 1. We set $N = 2m^2$ (m is large), $G(m) \subset G(N)$ of course. We check that the horizontal and vertical blocks of $G(m)$ are separated by certain horizontal and vertical lines, and also that its plus and minus diagonal blocks are separated by suitable lines of slope $+1$ and -1. This implies that the corresponding φ blocks are separated by parallel lines that can be defined using the map φ. We then apply an affine transformation so that the horizontal and vertical φ blocks are separated by horizontal and vertical lines, respectively. The diagonal separating lines remain parallel of course, and after suitable scaling the slope of the plus diagonal φ separating lines is very close to 1. Then we find a suitable $m \times m$ subgrid G_1 of $G(m)$ where the horizontal and vertical separating lines are distributed fairly equidistantly. These lines form a rectangular grid; see Figure 3. Each cell of this rectangular grid contains a unique point $w_{i,j} = \varphi(i,j)$ for each $(i,j) \in G_1$. We show next that if two such points w^1, w^2 belong to the same φ block, then their line $L(w^1, w^2)$ is almost horizontal, vertical, or diagonal depending on what kind of φ block w^1 and w^2 belong to. Finally we locate a small $n \times n$ subgrid of G_1 which satisfies the conditions of Lemma 1.

Preparations for the proof of Theorem 1 are given in Section 4, the proof itself is in Sections 5, 6, and 7. We remark that the proof does not use the full force of the orientation preserving property. It is only required for triples $u, v, w \in G(N)$ where one of the pairs u, v or v, w or w, u is horizontal or vertical or diagonal. Much stronger results can be proved using more triples. We hope to return to this question in a companion paper soon.

4 Normalizing φ

We set $N = 2m^2$. A simple computation shows that the lines $L_1 = L((0, N), (i, -m))$ and $L_2 = L((0, -N), (i, m))$ separate the neighbouring vertical blocks V_{i-1} and V_i of $G(m)$ for $i = 1, \dots, m$; see Figure 2. Then the vertical line through their intersection point also separates V_{i-1} and V_i. Note that the lines $L(\varphi(0, N), \varphi(i, -m))$ and $L(\varphi(0, -N), \varphi(i, m))$ are well-defined, let v be their point of intersection. Facts 1 and 2 show that the line passing through v and parallel with $L_0 = L(\varphi(0, N), \varphi(0, -N))$ separates $\varphi(V_{i-1})$ and $\varphi(V_i)$. All these lines are parallel with L_0. The same way we define separating lines for the $\varphi(V_{-i})$ and $\varphi(V_{-i+1})$ blocks. Consequently all vertical φ blocks are separated by parallel lines. It is important to point out that these parallel separating lines are uniquely determined by φ. The same method gives separating lines for the horizontal φ blocks of $G(m)$ that are all parallel with $L(\varphi(N, 0), \varphi(-N, 0))$.

The same argument works again for the diagonal blocks of $G(m)$. For instance one can check that the plus diagonal blocks D_i^+ and D_{i-1}^+ (for $i = 1, \dots, 2m$) are separated by the line through $(-N, -N)$ and the last point of D_i^+ and also by the line through (N, N) and the first point of D_i^+, we omit the straightforward computation.

This way we fix parallel separating lines for the horizontal, vertical, and plus and minus diagonal φ blocks of $G(m)$. We now use Remark 2 to modify φ so that the separating lines for the horizontal and vertical φ blocks of $G(m)$ are horizontal and vertical, respectively.

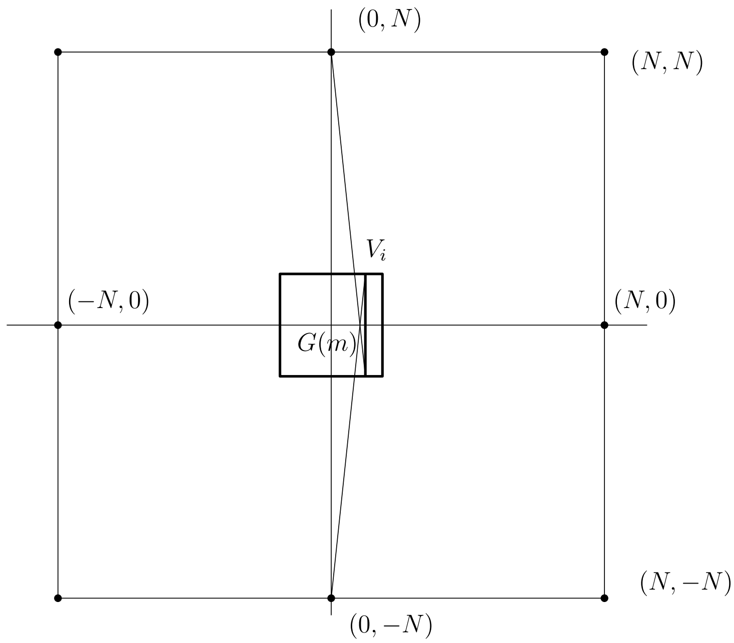

Figure 2: $G(N)$, $G(m)$ and a vertical block

They will be denoted by

$$L(b_j)^h = \{(x, y) : y = b_j\} \text{ and } L(a_i)^v = \{(x, y) : x = a_i\},$$

here $i, j \in [2m]$ and $a_1 < a_2 < \ldots < a_{2m}$ and $b_1 < b_2 < \ldots < b_{2m}$ where $[m]$ denotes the set $\{1, \ldots, m\}$. The upper indices h and v refer to horizontal and vertical. From this point onward we only work with points of the grid $G(m)$.

To have simple writing we keep the same notation for the modified φ. We note that there is still some freedom to define φ more precisely, a translation and scaling in horizontal and vertical directions are still allowed. That will come a little later.

Observe now that we have a grid-like structure (see Figure 3): the lines $L(a_i)^v$ and $L(b_j)^h$ determine $(2m-1)^2$ rectangular cells and each such cell contains the φ image of a unique point from $G(m)$. Precisely, the cell $C(i, j)$ is just the rectangle $[a_i, a_{i+1}] \times [b_j, b_{j+1}]$. It contains the point $w_{i,j} = \varphi(i, j)$, the image of a unique point in $G(m)$.

Suppose that m is large, $m > 10^5$ say. We also assume that m is a multiple of 4. Let $a_{p+4} - a_p$ be the minimal among the numbers $a_7 - a_3, a_8 - a_4, \ldots, a_{2m-2} - a_{2m-6}$ and let $b_{q+4} - b_q$ be the minimal among $b_7 - b_3, b_8 - b_4, \ldots, b_{2m-2} - b_{2m-6}$. Note that we have left out a “double frame” of the first and last two rows and columns. They will be needed later. Now either $p \le m$ or $p + 4 > m$. Similarly, either $q \le m$ or $q + 4 > m$. We can assume by symmetry that $p, q \le m$. We use now our freedom to fix φ by requiring that $a_p = b_q = 0$ and $a_{p+m} = b_{q+m} = m$. It follows then that $0 < a_{p+4}, b_{q+4} \le 4$. From this point onward we do not need the minus diagonal blocks, and a diagonal block will always mean a plus diagonal one. (Note that we would have kept the minus diagonal blocks in case $p \le m$ and $q + 4 > m$.)

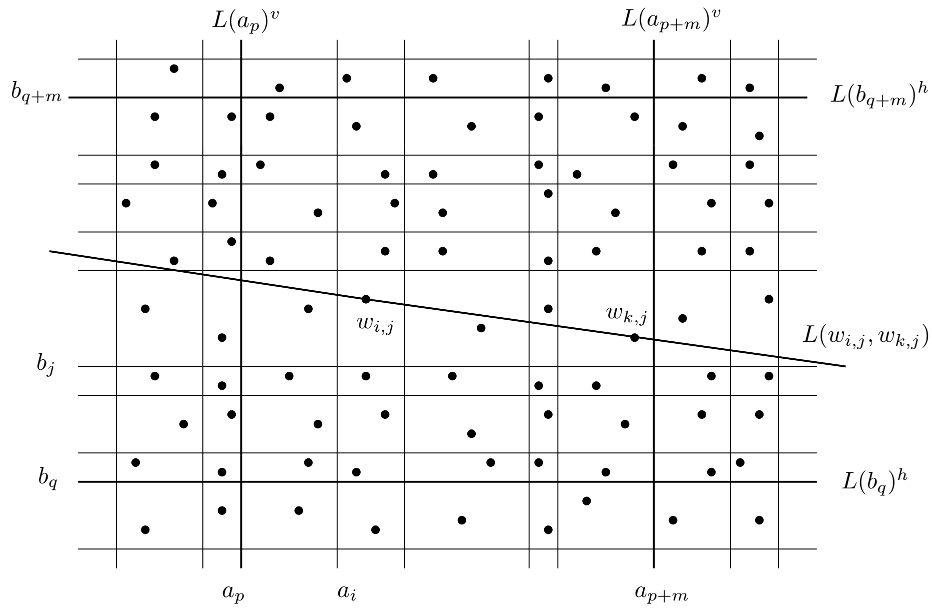

Figure 3: The grid-like structure and the line $L(w_{i,j}, w_{k,j})$

Remark 3. We fixed the map φ with the method just described and we keep the notation φ unchanged. We will recall at the end of Section 7 that it is of the form $T \circ \varphi$ with a well-defined affine transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$.

5 The rectangular grid and the subgrid G_1

With φ fixed this way, our first target is to show that the set of points $a_p, \ldots, a_{p+m}$ are distributed rather equidistantly on the interval $[0, m]$ on the x-axis, and the same for $b_q, \ldots, b_{q+m}$. Let R be the rectangle $[a_p, a_{p+m}] \times [b_q, b_{q+m}]$ and define G_1 as the $m \times m$ subgrid of $G(m)$ whose φ image lies in R. Horizontal, vertical, and diagonal blocks of G_1 are defined the same way as those of $G(m)$.

Assume B and B' are neighbouring diagonal blocks of G_1, L is their separating line, and $w_{i,j} \in B$ and $w_{i,j+1}, w_{i-1,j} \in B'$, see Figure 4. We will need the following key observation.

Claim 1. *The line L intersects both $C(i,j) \cup C(i-1,j)$ and $C(i,j) \cup C(i,j+1)$.*

The *proof* is simple: if L does not intersect the double cell $C(i,j) \cup C(i-1,j)$ (say), then it cannot separate the points $w_{i,j}$ and $w_{i-1,j}$. $\square$

Lemma 2. *If m is large enough, then $0 < a_{p+k+1} - a_{p+k} < 9$ and $0 < b_{q+k+1} - b_{q+k} < 9$ for all $k = 0, 1, \ldots, m-1$.*

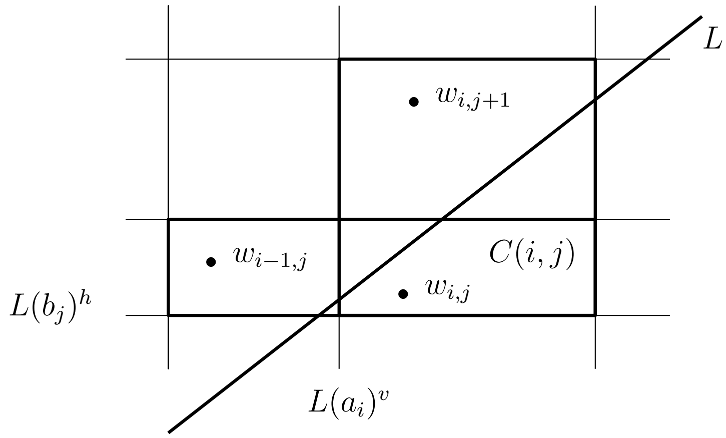

Figure 4: The line L and the cell $C(i, j)$.

Proof. We assume that $k > 4$ since the inequalities $a_{p+k+1} - a_{p+k} \leq 4$ and $b_{q+k+1} - b_{q+k} \leq 4$ are automatically satisfied for smaller k. Let B_i be the diagonal φ block of G_1 that contains the point $w_{p,q+i}$ for $i = 0, 1, 2, 3$ and let B_{-i} be the one containing $w_{p+i,q}$ for $i = 1, 2, 3$.

The seven diagonals $B_3, B_2, \ldots, B_{-3}$ are separated by six parallel lines L_3, L_2, L_1, L_{-1}, L_{-2}, L_{-3} in this order (L_0 is not defined). So for instance L_3 separates B_3 and B_2, see Figure 5. The key observation (Claim 1) implies that L_3 intersects $C(p, q+2) \cup C(p, q+3)$ and L_{-3} intersects $C(p+2, q) \cup C(p+3, q)$. Then the lines L_3 and L_{-3} intersect the rectangle $R_0 = [a_p, a_{p+4}] \times [b_q, b_{q+4}]$, so the distance between them is less than the diameter of R_0, which is at most $4\sqrt{2}$.

Define the rectangle $R_k = [a_{p+k}, a_{p+k+4}] \times [b_{q+k}, b_{q+k+4}]$ where $k = -2, -1, \ldots, m-3$. The above argument shows that the lines $L_3, \ldots, L_{-3}$ intersect the rectangle R_{p+m-4}. The line L_3 intersects both R_0 and R_{p+m-4} so its slope is a positive number. Consequently the angle β this line makes with the positive half of the x axis is strictly between 0 and $\pi/2$. Of course all diagonal separator lines have the same slope.

We claim that the cell $C(p+k+2, q+k+2) \subset R_k$ lies between the lines L_3 and L_{-3} for $k = -2, -1, \ldots, m-3$; see Figure 5. (This is where the double frame will be used.) Indeed, if it did not, then either the point (a_{p+k+2}, b_{q+k+3}) is above the line L_3, or the point (a_{p+k+3}, b_{q+k+2}) is below the line L_{-3}. In the former case L_3 does not intersect the union of the cells $C(p+k, q+k+3)$ and $C(p+k+1, q+k+3)$ contrary to the key observation. A similar argument works when the point (a_{p+k+3}, b_{q+k+2}) is below L_{-3}.

The line L_3 intersects $L(a_p)^v$ below the point (a_p, b_{q+4}), and intersects $L(a_{p+m})^v$ above the point (a_{p+m}, b_{q+m}), so its slope has to be at least $\frac{m-4}{m}$. Similar arguments show that the slope of the line L_{-3} cannot be larger than $\frac{m}{m-4}$. As both slopes are equal to $\tan\beta$ we have

$$\frac{m-4}{m} \leq \tan\beta \leq \frac{m}{m-4}. \tag{2}$$

So for m large, β is very close to $\pi/4$ and the strip between L_3 and L_{-3} (whose width is at most $4\sqrt{2}$) intersects both axes in a segment of length shorter than 9. This and

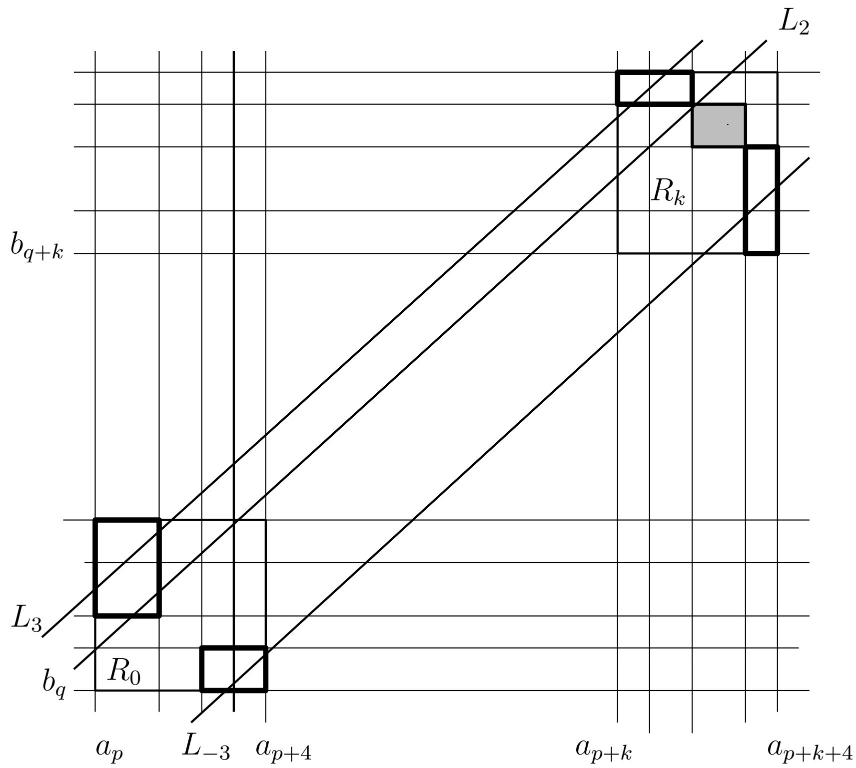

Figure 5: Only some of the lines L_i are shown, the cell $C(p+k+2, q+k+2)$ is shaded.

the fact that the cell $C(p+k+2, q+k+2)$ lies between the lines L_3 and L_{-3} finish the proof. □

The previous argument gives actually more. Namely, assume that B_0 is an arbitrary diagonal φ block of G_1 with neighbouring separating lines $L_3, L_2, \dots, L_{-3}$. Then every cell containing a point of B_0 lies between the lines L_3 and L_{-3}. But we will not use this fact.

We show next that if w^1 and w^2 belong to the same horizontal (or vertical) φ block, then their line $L(w^1, w^2)$ is almost horizontal (vertical). This is quite easy. Recall the notation $\alpha(w^1, w^2)$ for the angle of the line $L(w^1, w^2)$ with the positive x-axis.

Lemma 3. *Assume $p \le i < k \le p+m$ and $q+1 \le j \le q+m-2$. Then $|\tan \alpha(w_{i,j}, w_{k,j})| < \frac{27}{m}$. Similarly $p+1 \le i \le p+m-2$ and $q \le j < k \le q+m$ imply that $|\cot \alpha(w_{i,j}, w_{i,k})| < \frac{27}{m}$.*

Proof for the horizontal case. The line $L(w_{i,j}, w_{k,j})$ (see Figure 3) intersects the line $L(a_p)^v$ on the interval $[(a_p, b_{j-1}), (a_p, b_{j+2})]$, as otherwise the cell $C(p-1, j-1)$ or $C(p-1, j+1)$ from the double frame would be on the wrong side of $L(w_{i,j}, w_{k,j})$, contradicting the orientation preserving property of φ. Same way, the line $L(w_{i,j}, w_{k,j})$ intersects $L(a_{p+m})^v$ on the interval $[(a_{p+m}, b_{j-1}), (a_{p+m}, b_{j+2})]$. The length of both intervals is at most 27 by Lemma 2. Same proof applies in the vertical case. □

Lemma 2 shows that $a_{p'+k} - a_{p'} \le 9k$ when $p \le p' < p'+k \le p+m$ (and a similar bound for $b_{q'+k} - b_{q'}$). In fact a much stronger estimate holds.

Lemma 4. *Assume $k \le m/2$, $p \le p' < p'+k \le m$, and $q \le q' < q'+k \le q+m$. Define $a = a_{p'+k} - a_{p'}$ and $b = b_{q'+k} - b_{q'}$. Then $|a-b| < 46$. If k divides m, then $|a-k|$, $|b-k| < 46$.*

Proof. The main diagonal φ block in the rectangle $[a_{p'}, a_{p'+k}] \times [b_{q'}, b_{q'+k}]$, the one starting with $w_{p',q'}$, is separated from the next diagonal block above by the line L whose slope is $\tan\beta$. The key observation shows that this line intersects $C(p',q') \cup C(p',q'+1)$. Similarly, L intersects $C(p'+k-2, q'+k-1) \cup C(p'+k-1, q'+k-1)$. This implies, using Lemma 2, that

$$\frac{b-27}{a} < \tan\beta < \frac{b}{a-27},$$

which, combined with (2), gives

$$\left(1 - \frac{4}{m}\right)(a-27) < b < 27 + \left(1 + \frac{4}{m-4}\right)a.$$

The bounds here are of the form $a \pm 27$ plus (or minus) a small error term. By Lemma 2 $a = a_{p'+k} - a_{p'} \le 9k \le 9m/2$, and the error term is less than 27+19 (if $m > 80$, say) and $|a-b| < 46$ follows.

Here b can be any of the numbers $b(q') := b_{q'+k} - b_{q'}$ with $q \le q' \le m+q-k$. If $\min b(q') \le k \le \max b(q')$, then $|a-k| < 46$. Finally, if k divides m, then $\sum_0^{m/k-1} b(q+jk) = m$ so the average of these $b(q')$s is exactly k and $\min b(q') \le k \le \max b(q')$. So no a can differ from this average by more than 46. $\square$

We want to show the analogue of Lemma 3 for the diagonal φ blocks. For this purpose we have to consider a smaller rectangle R'. Recall that m is divisible by 4, set $p' = p + m/4, q' = q + m/4$, $m' = m/2$ and define $R' = [a_{p'}, a_{p'+m'}] \times [b_{q'}, b_{q'+m'}]$.

Lemma 5. *Assume $w^1, w^2 \in R'$ belong to the same diagonal φ block, B^*, of G_1. If $\alpha(w^1, w^2)$ differs from $\pi/4$ by δ, then $|\delta| < \frac{K}{m}$ where K is a constant, for instance $K = 150$ will do.*

Proof. Let $u_1 = w_{p',i}$ be the leftmost, and $u_2 = w_{p'+m'-1,i+m'-1}$ be the rightmost point of B^* in S, where S is the slab between the vertical lines $L(a_{p'})^v$ and $L(a_{p'+m'})^v$; see Figure 6. Due to the definition of R and R', both points $u_3 = w_{p',i-1}$ and $u_4 = w_{p'+m'-1,i+m'}$ are in R. For simplicity of notation, we further assume that B^* is the block B_0 (defined at the beginning of the proof of Lemma 5.2.). Then B_1 and B_{-1} are the blocks neigbouring the block $B^* = B_0$. Since φ is orientation preserving, $u_3 \in B_{-1}$ is below the line $L(w^1, w^2)$ and $u_4 \in B_1$ is above. It follows that the point $(a_{p'+1}, b_{i-1})$ (denoted by X on Figure 6) is below this line and the point $(a_{p'+m'-1}, b_{i+m'+1})$ (denoted by Z) is above. Analogously the point $w_{p'-1,i} \in B_1$ is above the line $L(w^1, w^2)$, and then so is $Y := (a_{p'-1}, b_{i+1})$. Similarly $w_{p'+m',i+m'-1} \in B_{-1}$ is below this line and then so is $V := (a_{p'+m'+1}, b_{i+m'-1})$; see Figure 6. Analogously $Z := (a_{p'+m'-1}, b_{i+m'+1})$ is above $L(w^1, w^2)$.

Consequently the slope of the line $L(w^1, w^2)$ is between the slopes of $L(Y, V)$ and $L(X, Z)$. By Lemma 4 (m is divisible by m') the numbers $a_{p'+m'} - a_{p'}$ and $b_{i+m'} - b_i$ differ from m' by less than 46. The slope of $L(Y, V)$ is

$$\frac{b_{i+m'-1} - b_{i+1}}{a_{p'+m'+1} - a_{p'-1}} > \frac{(m'-46) - 18}{(m'+46) + 18} = 1 - \frac{128}{m'+64},$$

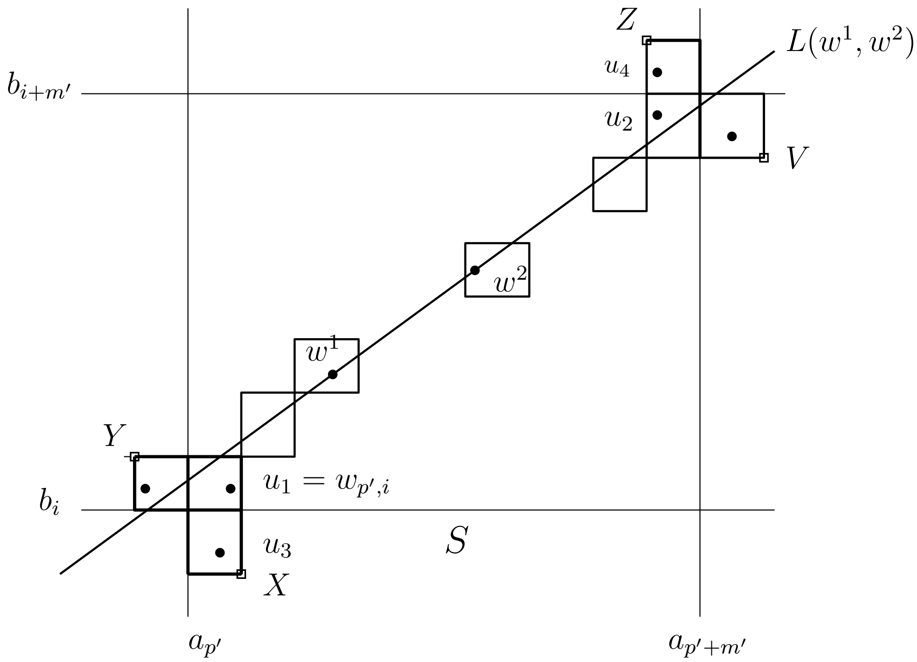

Figure 6: Some cells in B_0 and the line $L(w^1, w^2)$.

we also used that by Lemma 2 $a_{k+1} - a_k, b_{k+1} - b_k < 9$. Similarly the slope of $L(X, Z)$ equals

$$\frac{b_{i+m'+1} - b_{i-1}}{a_{p'+m'-1} - a_{p'+1}} < \frac{(m' + 46) + 18}{(m' - 46) - 18} = 1 + \frac{128}{m' - 64}.$$

A straightforward computation shows that slope of $L(w^1, w^2)$, which equals $\tan\left(\frac{\pi}{4} + \delta\right) \approx 1 + 2\delta$, differs from one by at most $256/(m - 128)$ (assuming that m is large enough) and $|\delta| < \frac{K}{m}$ follows with $K = 150$. $\square$

6 Finding an even smaller subgrid

We set $m = Cn^2\varepsilon^{-1}$ where $C > 0$ will be specified later. Set $p^* = p + m/2$ and $q^* = q + m/2$ and define the rectangle $R^* = [a_{p^*}, a_{p^*+n}] \times [b_{q^*}, b_{q^*+n}]$ and let G_2 consist of points $z \in G_1$ such that $\varphi(z) \in R^*$. Of course G_2 is an $n \times n$ subgrid of G_1, a translate of the set of grid points in $G^*(n)$: $G_2 = z_0 + G^*(n)$ for a suitable $z_0 \in \mathbb{Z}^2$. This is the subgrid that we are after as we shall see soon.

Note that $n < m' = m/2$, in fact much smaller. We assume (as we can) that n divides m. So R^* is a tiny rectangle in the middle of R whose sides have length between $n-46$ and $n+46$ because of Lemma 4. Then the diagonal φ blocks of G_1 that contain points from R^* are very close to the middle of R. It follows that if w^1, w^2 belong to a diagonal φ block B_0 (say) of G_2, then B_0 satisfies the conditions of Lemma 5. Lemma 3 applies to points in horizontal and vertical φ block of G_2.

Corollary 1. *If z^1, z^2 belong to the same horizontal, diagonal or vertical block of G_2, then $\alpha(\varphi(z^1), \varphi(z^2))$ deviates from $0, \pi/4, \pi/2$, respectively by at most $\gamma := K/m$.*

Lemma 6. *There is an affine map $\vartheta : \mathbb{R}^2 \to \mathbb{R}^2$ such that (with $C = 42K$) for every $z \in G_2$*

$$\|\varphi(z) - \vartheta(z)\| < \frac{\varepsilon}{2}.$$

Proof. Define $\varphi_0 : G^*(n) \to \mathbb{R}^2$ by

$$\varphi_0(i,j) = w_{p^*+i,q^*+j} - w_{p^*,q^*}$$

and set $e = \frac{1}{n-1}\varphi_0(n-1,0)$ and $f = \frac{1}{n-1}\varphi_0(0,n-1)$. Note that e and f are almost orthogonal because e is away from the horizontal direction by at most γ and f from the vertical one at most by the same amount. So the angle between e and f differs from $\pi/2$ by less than 2γ. Then $|e \cdot f| < 0.1$ if m is large enough (where $e \cdot f$ denotes the scalar product of e and f). The x component of $(n-1)e$ is close to n (by Lemma 4) and its y component is at most 9 (by Lemma 2) so $\|e\|$ is very close to one: $0.95 < \|e\| < 1.05$ if m is large enough. The same way we get that $0.95 < \|f\| < 1.05$, too. This shows that e and f form an almost orthonormal basis of $\mathbb{R}^2$, and $\|e\| + \|f\| < 2.1$.

The conditions of Lemma 1 are satisfied for φ_0 with $\gamma = K/m$. So its conclusion holds: for every $(i,j) \in G^*(n)$

$$\|\varphi_0(i,j) - (ie + jf)\| < 10\gamma n^2(\|e\| + \|f\|).$$

We define a linear map $L : \mathbb{R}^2 \to \mathbb{R}^2$ by setting $L(z) = L(x,y) = xe + yf$ and an affine map $\vartheta : \mathbb{R}^2 \to \mathbb{R}^2$ via $\vartheta(z) = L(z) + w_{p^*,q^*}$; L is the linear part of ϑ.

A given $z \in G_2$ can be written uniquely as $z_0 + (i,j)$ where $(i,j) \in G^*(n)$. For every $z = z_0 + (i,j) \in G_2$ we have

$$\begin{aligned} \|\varphi(z) - \vartheta(z)\| &= \|\varphi_0(i,j) - (ie + jf)\| \le 10\gamma n^2(\|e\| + \|f\|) \\ &< 21n^2 \frac{K\varepsilon}{Cn^2} < \frac{21K}{C}\varepsilon \le \frac{\varepsilon}{2} \end{aligned}$$

when choosing the constant $C \ge 42K$. □

7 Proof of Theorem 1

The proof is quite easy now. The linear part, L, of ϑ carries the vector $(1,0)$ and $(0,1)$ to e and f, respectively. Since vectors e and f form an almost orthonormal basis of $\mathbb{R}^2$, L is very close to the identity assuming that m and n is large enough. Then its inverse, L^{-1}, is also close to the identity implying that $\|L^{-1}(x)\| \le 2\|x\|$ for all $x \in \mathbb{R}^2$. In particular when $z \in G_2$ and $x := \varphi(z) - \vartheta(z)$ we have using Lemma 6

$$\|L^{-1}(\varphi(z) - \vartheta(z))\| < \varepsilon.$$

Observe now that

$$\vartheta^{-1}(\varphi(z)) - z = \vartheta^{-1}(\varphi(z)) - \vartheta^{-1}(\vartheta(z)) = L^{-1}\left(\varphi(z) - \vartheta(z)\right).$$

So in Theorem 1 the map ψ is ϑ^{-1}, or more precisely $\vartheta^{-1} \circ T$ where T is the affine map from Remark 2 $\square$

8 Proof of Lemma 1

We will need the following almost elementary fact. Consider the quadrilateral $Q = \text{conv}\,\{X, Y, Z, V\}$ as in Figure 7. Assume that

$$\begin{aligned} |\alpha(X,Y)|, |\alpha(V,Z)| < \gamma, |\alpha(X,Z) - \pi/4| &< \gamma \\ |\alpha(X,V) - \pi/2|, |\alpha(Y,Z) - \pi/2| &< \gamma. \end{aligned}$$

Define

$$M = \frac{1 + \tan 2\gamma}{(1 - \tan 2\gamma)\cos 2\gamma}.$$

Claim 2. *Under the above conditions*

$$M^{-1} < \frac{a}{b}, \frac{a'}{b}, \frac{b'}{b}, \frac{a'}{a}, \frac{b'}{a}, \frac{a'}{b'} < M.$$

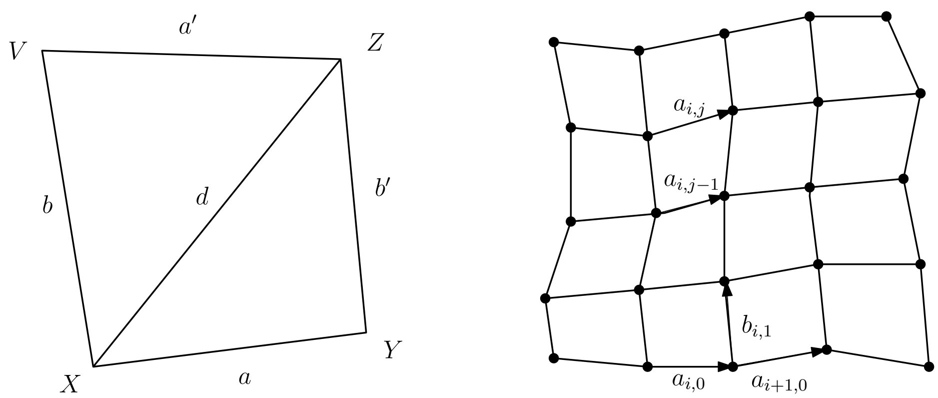

Figure 7: The quadrilateral Q and a piece of the φ grid.

The *proof* is simple: the sine theorem shows that, with the notation of Figure 7,

$$\frac{d}{\sqrt{2}}(\cos 2\gamma - \sin 2\gamma) < a, a', b, b' < \frac{d}{\sqrt{2}} \frac{\cos 2\gamma + \sin 2\gamma}{\cos 2\gamma}. \quad \square$$

Proof of Lemma 1. We are going to use the inequalities of the claim in the quadrilaterals with vertices $\varphi_0(i,j), \varphi_0(i-1,j), \varphi_0(i,j-1)$, and $\varphi_0(i-1,j-1)$. We define $a_{i,j} = \varphi_0(i,j) - \varphi_0(i-$

$1, j)$ for $i \in [n-1]$ and $j \in \{0, 1, \ldots, n-1\}$ and, analogously, $b_{i,j} = \varphi_0(i, j) - \varphi_0(i, j-1)$, for $i \in \{0, 1, \ldots, n-1\}$ and $j \in [n-1]$.

The claim shows that in the triangle with sides $a_{i+1,0}$ and $b_{i,1}$ (see Figure 7), and in the triangle with sides $b_{i,1}$ and $a_{i,0}$

$$M^{-1} < \frac{\|b_{i,1}\|}{\|a_{i+1,0}\|} < M \text{ and } M^{-1} < \frac{\|a_{i,0}\|}{\|b_{i,1}\|} < M.$$

Consequently

$$M^{-2} < \frac{\|a_{i,0}\|}{\|a_{i+1,0}\|} < M^2 \text{ and so } \max \|a_{i,0}\| \leq \min \|a_{i,0}\| M^{2(n-1)}. \tag{3}$$

The vectors e, f form a basis of $\mathbb{R}^2$, so a vector $a \in \mathbb{R}^2$ can be written uniquely as $a = a^x e + a^y f$. Our first target is to show that every $a^x_{i,0}$ is very close to one and every $a^y_{i,j}$ is very close to zero. The analogous statement for $b^y_{0,j}$ and $b^x_{i,j}$ would follow by symmetry.

Condition (1) implies that e and f are almost orthogonal, their angle differs from $\pi/2$ by less than 2γ. The same condition implies that in the triangle with vertices $0, a_{i,j}$, and $a^x_{i,j}e$ the angle at 0 is at most 2γ, the angle at $a_{i,j}$, and the one at $a^x_{i,j}e$ differs from $\pi/2$ by less than 2γ. Thus $a^x_{i,j} > 0$ follows. The sine theorem shows then that

$$\cos 2\gamma < \frac{\|a_{i,j}\|}{a^x_{i,j}\|e\|} < \frac{1}{\cos 2\gamma} \text{ and} \tag{4}$$

$$-\sin 2\gamma < \frac{a^y_{i,j}\|f\|}{a^x_{i,j}\|e\|} < \tan 2\gamma \text{ implying } |a^y_{i,j}| < a^x_{i,j}\frac{\|e\|}{\|f\|}\tan 2\gamma. \tag{5}$$

Another form of (4) is

$$\cos 2\gamma \max a^x_{i,j}\|e\| < \max \|a_{i,j}\| \text{ and } \min \|a_{i,j}\| < \min a^x_{i,j}\|e\|/\cos 2\gamma$$

where j is fixed and the maxima and minima are taken over all i. Putting these inequalities with $j = 0$ in (3) we see that

$$\frac{\max a^x_{i,0}}{\min a^x_{i,0}} \leq \frac{\max \|a_{i,0}\|}{\min \|a_{i,0}\| \cos^2 2\gamma} < \frac{M^{2(n-1)}}{\cos^2 2\gamma} =: 1 + \Delta.$$

The average of the $a^x_{i,0}$ for $i \in [n-1]$ is 1 because $\sum_1^{n-1} a^x_{i,0} = n-1$, so $\min a^x_{i,0} \leq 1 \leq \max a^x_{i,0}$ implying that $\max a^x_{i,0} < (1+\Delta)\min a^x_{i,0} \leq 1+\Delta$, and $\min a^x_{i,0} > \max a^x_{i,0}/(1+\Delta) \geq 1-\Delta$. Consequently

$$|a^x_{i,0} - 1| \leq \Delta \text{ for all } i \in [n-1].$$

We need to estimate Δ:

$$\Delta = \frac{M^{2(n-1)}}{\cos^2 2\gamma} - 1 = \left(\frac{1+\tan 2\gamma}{1-\tan 2\gamma}\right)^{2(n-1)} \left(\frac{1}{\cos^2 2\gamma}\right)^n - 1.$$

Here $1/\cos^2 2\gamma = 1/(1-\sin^2 2\gamma) < 1/(1-(2\gamma)^2) < 1+2(2\gamma)^2$ where the last inequality holds for small enough $\gamma > 0$ as one can check directly. Then, using the inequality $1+t < e^t$ we get

$$\left(\frac{1}{\cos^2 2\gamma}\right)^n < \left(1+2(2\gamma)^2\right)^n < \exp\{2n(2\gamma)^2\}.$$

The same way

$$\begin{aligned}\left(\frac{1+\tan 2\gamma}{1-\tan 2\gamma}\right)^{2(n-1)} &= \left(1+\frac{2\tan 2\gamma}{1-\tan 2\gamma}\right)^{2(n-1)} \\ &< \exp\left\{2(n-1)\frac{2\tan 2\gamma}{1-\tan 2\gamma}\right\}.\end{aligned}$$

Returning to Δ we have now

$$\Delta < \exp\left\{2(n-1)\frac{2\tan 2\gamma}{1-\tan 2\gamma} + 2n(2\gamma)^2\right\} - 1 < 10n\gamma.$$

The last inequality follows from $e^t \le 1+1.1t$ which is true if $t > 0$ is small enough and here $t = 2(n-1)\frac{2\tan 2\gamma}{1-\tan 2\gamma} + 2n(2\gamma)^2 \approx 8n\gamma$ is indeed small. Consequently

$$|a^x_{i,0} - 1| \le 10n\gamma \text{ and similarly } |b^y_{0,j} - 1| \le 10n\gamma. \tag{6}$$

The same way one can check that M^{n-1} is only slightly larger than one, namely, $1 < M^{n-1} < 1+5n\gamma$.

In the quadrilateral with sides $a_{i,j-1}$ and $a_{i,j}$ (see Figure 7) we have, using Claim 2 again, that $\frac{\|a_{i,j}\|}{\|a_{i,j-1}\|} < M$ and so $\|a_{i,j}\| < M^{n-1}\|a_{i,0}\|$. Applying (4) twice gives

$$\begin{aligned} a^x_{i,j} &\le \frac{\|a_{i,j}\|}{\cos 2\gamma\|e\|} < M^{n-1}\frac{\|a_{i,0}\|}{\cos 2\gamma\|e\|} < M^{n-1}\frac{a^x_{i,0}}{\cos^2 2\gamma} \\ &< (1+5n\gamma)(1+10n\gamma)(1+2(2\gamma)^2) < 1+16n\gamma.\end{aligned}$$

We could prove here that $a^x_{i,j} > 1-16n\gamma$ but this is not needed.

Claim 2 applies in the quadrilateral with vertices $0, (n-1)e, (n-1)f$, and $\varphi(n-1, n-1)$ and shows that $\frac{\|e\|}{\|f\|} < M$. Using this in (5) gives that

$$|a^y_{i,j}| \le a^x_{i,j} M \tan 2\gamma < (1+16n\gamma)\frac{1+\tan 2\gamma}{(1-\tan 2\gamma)\cos^2 2\gamma}\tan 2\gamma \approx \tan 2\gamma$$

which is approximately $\tan 2\gamma$ in the given range $\gamma = O(n^{-2})$. Since there $\tan 2\gamma < 3\gamma$ we have

$$|a^y_{i,j}| \le 4\gamma \text{ and similarly } |b^x_{i,j}| \le 4\gamma. \tag{7}$$

Next we estimate the difference $u = \varphi_0(i,j) - (ie+jf) = u^x e + u^y f$. We begin with u^x :

$$\begin{aligned} |u^x| &= |a^x_{1,0} + \ldots + a^x_{i,0} + b^x_{i,1} + \ldots b^x_{i,j} - i| \\ &\le |a^x_{1,0} - 1| + \ldots + |a^x_{i,0} - 1| + |b^x_{i,1}| + \ldots + |b^x_{i,j}| \\ &\le i10n\gamma + j4\gamma \le (n-1)(10n\gamma + 4\gamma) < 10\gamma n^2,\end{aligned}$$

where we used (6) and (7). Estimating $|u^y|$ is similar but starts with writing u^y as

$$b^y_{0,1} + \ldots + b^y_{0,j} + a^y_{j,1} + \ldots a^y_{j,i} - j.$$

Finally, we have

$$\begin{aligned} u^2 &= (u^x e + u^y f)^2 = (u^x)^2 e^2 + 2u^x u^y e \cdot f + (u^y)^2 f^2 \\ &< (10\gamma n^2)^2 (\|e\| + \|f\|)^2, \end{aligned}$$

implying that $\|\varphi_0(i,j) - (ie + jf)\| < 10\gamma n^2(\|e\| + \|f\|)$ indeed. □

9 The lower bound on $N(n, \varepsilon)$

The example showing the lower bound in Remark 1 is the projective map $\varphi(x, y) \to (x, y)/(N + 1 - y)$ which is orientation preserving on $G(N)$ and carries the horizontal line $y = j$ to the horizontal line $y = j/(N + 1 - j)$ for $j = 0, \pm 1, \ldots, \pm N$. Assume ψ is the affine map satisfying the conclusion of Theorem 1 with $z_0 = (a, b) \in \mathbb{Z}^2$, $z_0 + G^*(n) \subset G(N)$. Let z_0, z_1, z_2, z_3 be the vertices of the square $z_0 + G^*(n)$ with the subscripts chosen so that $z_1 - z_0 = z_3 - z_2 = (n - 1, 0)$ and $z_2 - z_0 = (0, n - 1)$.

The map ψ is affine so is of the form $\psi(x) = Lx + t$ where $L : \mathbb{R}^2 \to \mathbb{R}^2$ is a linear map and $t \in \mathbb{R}^2$ is a translation. The vectors $\varphi(z_1) - \varphi(z_0)$ and $\varphi(z_3) - \varphi(z_2)$ are parallel (actually both are horizontal). Thus

$$\begin{aligned} \frac{\|\psi \circ \varphi(z_1) - \psi \circ \varphi(z_0)\|}{\|\psi \circ \varphi(z_3) - \psi \circ \varphi(z_2)\|} &= \frac{\|L(\varphi(z_1) - \varphi(z_0))\|}{\|L(\varphi(z_3) - \varphi(z_2))\|} = \frac{\|\varphi(z_1) - \varphi(z_0)\|}{\|\varphi(z_3) - \varphi(z_2)\|} \\ = \frac{N + 1 - (b + n - 1)}{N + 1 - b} &= 1 - \frac{n - 1}{N + 1 - b}. \end{aligned}$$

Moreover, the vectors

$$\psi \circ \varphi(z_1) - \psi \circ \varphi(z_0) \text{ and } \psi \circ \varphi(z_3) - \psi \circ \varphi(z_2)$$

differ from $z_1 - z_0 = z_3 - z_2 = (n - 1, 0)$ by at most 2ε. Consequently the ratio of their lengths is at least $(n - 1 - 2\varepsilon)/(n - 1 + 2\varepsilon)$ which is equal to $1 - 4\varepsilon/(n - 1 + 2\varepsilon)$. So we have

$$\frac{n - 1}{N + 1 - b} \le \frac{4\varepsilon}{n - 1 + 2\varepsilon} \text{ or } N \ge b - 1 + \frac{(n - 1)(n - 1 + 2\varepsilon)}{4\varepsilon}.$$

Since $-N \le b \le N - (n - 1)$, we get

$$\begin{aligned} 2N &= N + N \ge N + b - 1 + \frac{(n - 1)(n - 1 + 2\varepsilon)}{4\varepsilon} \\ &\ge -1 + \frac{(n - 1)^2}{4\varepsilon} + \frac{n - 1}{2} > \frac{(n - 1)^2}{4\varepsilon} \end{aligned}$$

when $n \ge 4$. This implies $N(n, \varepsilon) = \Omega(n^2 \varepsilon^{-1})$, indeed.

Acknowledgments. The work of I. Bárány was partially supported by Hungarian National Research Grants No 131529, 132696, and 133819. The work by P. Valtr was supported by the grant no. 21-32817S of the Czech Science Foundation (GAČR). We thank three anonymous referees for carefully checking earlier versions of this paper and for pointing out several inaccuracies.

References

[1] Goodman, J. E. and Pollack, R. The complexity of point configurations. *Discrete Appl. Math.*, **31** (1991) 167–180.

[2] Goodman, J. E., Pollack, R. and B. Sturmfels. The intrinsic spread of a configuration in $\mathbb{R}^d$. *J. Amer. Math. Soc.*, **3** (1990) 639–651.

[3] Kratochvíl, J. and Matoušek, J. Intersection graphs of segments. *J. Combin. Theory Ser. B*, **62** (1994), 289–315.

[4] Nešetřil, J. and Valtr, P. A Ramsey-type theorem in the plane, *Combinatorics, Probablity and Computing*, **3** (1994), 127–135.

[5] Nešetřil, J. and Valtr, P. A Ramsey property of order types, *J. Comb. Theory, A*, **81** (1998), 88–107.

[6] Pilz, A. and Welzl, E. Order on order types. *Discrete Comput. Geom.*, **59** (2018), 886–922.