Graph Distance and Euclidean Distance on the Grid

J. Pach*, New York, USA
R. Pollack**, New York, USA
J. Spencer, New York, USA

Abstract

Given a connected graph $G = (V, E)$, $V = \mathbf{Z}^2$, on the lattice points of the plane, let $d_G(p,q)$ and $d(p,q)$ denote the graph distance and the Euclidean distance between p and q respectively. In this note we prove that for every $\epsilon > 0$ there is a graph $G = G_\epsilon$ and a constant $d = d_\epsilon$ such that

$$|d_G(p,q) - d(p,q)| < \epsilon d(p,q)$$

for every pair $p, q \in V$ with $d(p,q) \geq d$. It remains open whether or not there is a graph G and a suitable constant K which satisfies

$$|d_G(p,q) - d(p,q)| < K$$

for all $p, q \in \mathbf{Z}^2$.

1 Defining the Graph

Let G_0 denote the graph on $\mathbf{Z}^2$ which can be obtained by joining two integer points with an edge if and only if their Euclidean distance is 1. The idea for the construction stems from the observation that, although $\sup \frac{d_G(p,q)}{d(p,q)} = \sqrt{2}$, if the straight line pq is almost parallel to the $x-$axis or to the $y-$axis, then the graph distance in G_0 and the Euclidean distance are very close to each other. Therefore we are going to define a large number of square lattices $\Lambda_i \subseteq \mathbf{Z}^2$ with various side lengths so that the directions of their axes are fairly uniformly distributed in $[0, 2\pi)$. For every i, we shall construct a graph G_i similar to G_0, whose main nodes are the elements of Λ_i, and we shall

*Is also at the Mathematical Institute of the Hungarian Academy of Sciences and is supported in part by NSF grant CCR-8901484 and Hungarian SF grant OTKA-1814
**Supported in part by NSF grant CCR-8901484 and NSA grant MDA904-89-H-2030

R. Bodendiek, R. Henn (Eds.)
Topics in Combinatorics and Graph Theory

glue these graphs together. The most important property of G_i will be that $d_{G_i}(p,q)$ and $d(p,q)$ do not differ too much, whenever $p, q \in \Lambda_i$ and the line determined by p and q is almost parallel to one of the axes of Λ_i.

To be more precise, we need some notation. Let $\epsilon > 0$ be fixed, and choose a set of vectors (a_i, b_i), $1 \le i \le N_\epsilon$, with the properties

(i) a_i, b_i are positive integers;

(ii) $\{\frac{b_i}{a_i} : a \le i \le N_\epsilon\}$ is well distributed in $[-1,+1]$ in the sense that for every; $x \in [-1,+1]$ there exists an i with $|x - \frac{b_i}{a_i}| < \frac{\epsilon}{100}$;

(iii) the length S_1 of (a_1, b_1) is at least K_ϵ, and $S_i = \sqrt{a_i^2 + b_i^2} > 5S_{i-1}$ for $2 \le i \le N_\epsilon$ (where $K_\epsilon > 50$ is a constant which will be fixed later).

Let $\Lambda_i \subseteq \mathbf{Z}^2$ denote the square lattice generated by (a_i, b_i), i.e.,

$$\Lambda_i = \{m(a_i, b_i) + n(-b_i, a_i) : m, n \in \mathbf{Z}^2\}, 1 \le i \le N_\epsilon$$

For simplicity, let

$$(m,n)_i = m(a_i, b_i) + n(-b_i, a_i) = (ma_i - nb_i, mb_i + na_i),$$

and let $C_i(m,n)$, the cell of $(m,n)_i$ be defined as the convex hull of $\{(m,n)_i, (m+1,n)_i, (m,n+1)_i, (m+1,n+1)_i\}$. Further, let $(m,n)_i^*$ denote a point $(m',n')_{i-1}$ which is closest to $(m,n)_i$ and whose cell $C_{i-1}(m',n') \subset C_i(m,n), (i \ge 2)$.

Assume recursively that for every $j < i$ and for every pair of integers (m,n) we have already defined three paths $Pr_j(m,n), Pu_j(m,n)$ and $P_j(m,n)$ which satisfy

1. $Pr_j(m,n)$ connects $(m,n)_j$ to $(m+1,n)_j$,
   $Pu_j(m,n)$ connects $(m,n)_j$ to $(m,n+1)_j$,
   $P_j(m,n)$ connects $(m,n)_j$ to $(m,n)_j^*$;
2. $Pr_j(m,n)$ and $Pu_j(m,n)$ have length $\lceil S_j \rceil$, while the length of $P_j(m,n)$ is $\lceil d((m,n)_j, (m,n)_j^*) \rceil$;
3. All of these paths are internally disjoint from one another and from all previously defined paths, i.e., they can only meet at their endpoints;
4. All internal vertices of $Pr_j(m,n)$, $Pu_j(m,n)$ and $P_j(m,n)$ are in the interior of the cell $C_j(m,n)$.

Observe that at this point of the construction the number of nonisolated vertices in $C_i(m,n)$ is at most

$$\sum_{j<i}\left(\frac{S_i+2\sqrt{2}S_j}{S_j}\right)^2 4\lceil S_j\rceil < \sum_{j<i}\left(\frac{2S_i}{S_j}\right)^2 5S_j = 20S_i^2\sum_{j<i}\frac{1}{S_j} < \frac{S_i^2}{2},$$

provided that $S_1 \geq 50$. Hence, there are plenty of isolated vertices left in $C_i(m,n)$ that can be used to form the paths $Pr_i(m,n)$, $Pu_i(m,n)$ and $P_i(m,n)$ preserving the above properties.

Suppose that we have carried out the construction for all $i \leq N_\epsilon = N$, and consider the graph consisting of all paths already defined. For each point $p \in \mathbf{Z}^2$ which remained isolated, we add an edge to the nearest point of Λ_i.

2 Estimating the Distances

Since the length of every path which begins and ends at a vertex of some Λ_i is at least as large as the Euclidean distance between its endpoints, and every internal vertex belongs to some Λ_i,

$$d_G(p,q) > d(p,q) - 2\sqrt{2}S_N.$$

Bounding $d_G(p,q)$ from above is just a bit more complicated.

Lemma 1 *For every point $(m,n)_N \in \Lambda_N$, the graph G contains a path $P(m,n)$ of length at most S_N, all of whose vertices are in the cell $C_N(m,n)$ and which visits at least one vertex p_j of each Λ_j, $1 \leq j \leq N$.*

Proof: Consider the path, $P(m,n)$, obtained by concatenating $P_N(m,n)$, $P_{N-1}(m',n'), P_{N-2}(m'',n''),\ldots,P_2(m^{(N-2)},n^{(N-2)})$, as defined in (1)–(4) above. Its length is clearly at most $\sum_{i<N}\lceil 2S_i\rceil < S_N$, and it visits exactly one vertex of each Λ_i, $1 \leq i \leq N$. □

Lemma 2 *For any two points $x = (m_x,n_x)_i$ and $y = (m_y,n_y)_i$ of Λ_i*

$$d_G(x,y) < (1+\delta)(1+\frac{1}{S_i})d(x,y)$$

where $\delta = \min(\frac{n_x-n_y}{m_x-m_y}, \frac{m_x-m_y}{n_x-n_y}) \leq 1$.

Proof: Using only paths Pr_i and Pu_i, we obtain that

$$\begin{aligned} d_G(x,y) &\le (|m_x - m_y| + |n_x - n_y|)\lceil S_i \rceil \\ &\le (1+\delta)\sqrt{(m_x - m_y)^2 + (n_x - n_y)^2} S_i (1 + \frac{1}{S_i}) \\ &= (1+\delta)(1+\frac{1}{S_i}) d(x,y), \end{aligned}$$

as desired.□

Lemma 3 *Given any integer point $p \in \mathbf{Z}^2$ and $1 \le i \le N$, there is a point $p_i \in \Lambda_i$ such that $d_G(p, p_i) < 10 S_N$ and $d(p, p_i) \le \sqrt{2} S_N$.*

Proof: Let $p \in C_N(m,n)$. Clearly, there is a $j \le N$ such that p has graph distance at most $2\lceil S_j \rceil$ from some $q_j \in \Lambda_j \cap C_N(m,n)$. Let p_j (p_i) be the vertex of Λ_j (Λ_i) which is on $P(m,n)$, and whose existence is guaranteed by Lemma 1. By Lemma 2 there is a path in G from q_j to p_j whose length is at most $3d(p_j, q_j) \le 3\sqrt{2} S_N < 5 S_N$. Thus

$$\begin{aligned} d_G(p, p_i) &\le d_G(p, q_j) + d_G(q_j, p_j) + d_G(p_j, p_i) \\ &\le 2\lceil S_j \rceil + 5 S_N + S_N \\ &< 10 S_N. \end{aligned}$$

Since $p, p_i \in C_N(m,n)$, $d(p, p_i) \le \sqrt{2} S_N$ readily follows. □

To establish the result stated in the abstract it is sufficient to prove the following

Theorem 1 *There exists a constant d_ϵ such that*

$$d_G(p,q) < (1+\epsilon) d(p,q)$$

for every pair $p, q \in \mathbf{Z}^2$ whose Euclidean distance exceeds d_ϵ.

Proof: Given $p, q \in \mathbf{Z}^2$, choose i so that $\tan\theta$, the slope of the line determined by p and q, differs from $\tan\theta_i = \frac{b_i}{a_i}$ by less than $\frac{\epsilon}{100}$. Choose $p_i = (m_1, n_1)_i$ and $q_i = (m_2, n_2)_i$ as in Lemma 3. Let α denote the angle between the lines pq and $p_i q_i$. Then $\sin\alpha \le \frac{\sqrt{2} S_N}{d(p,q)/2}$, whence

$$\tan\alpha \le \frac{\sqrt{2} S_N}{\sqrt{\frac{d^2(p,q)}{4} - 2S_N^2}} = \sqrt{\frac{8 S_N^2}{d^2(p,q) - 8 S_N^2}}.$$

Obviously,

$$\begin{aligned}|\frac{n_2 - n_1}{m_2 - m_1}| &< \tan(|\theta - \theta_i| + \alpha) \\ &< \tan(\frac{\epsilon}{100} + \alpha) \\ &< \frac{\epsilon}{10},\end{aligned}$$

provided that $d(p,q)$ is so large that $\tan\alpha < \frac{\epsilon}{100}$. Thus, by Lemma 2,

$$d_G(p_i, q_i) < (1 + \frac{\epsilon}{10})(1 + \frac{1}{S_i})d(p_i, q_i).$$

On the other hand, Lemma 3 implies that

$$\begin{aligned}d_G(p,q) &\leq d_G(p,p_i) + d_G(p_i,q_i) + d_G(q_i,q) \\ &\leq 20S_N + (1 + \frac{\epsilon}{10})(1 + \frac{1}{S_i})(d(p,q) + 2\sqrt{2}S_N) \\ &\leq (1+\epsilon)d(p,q),\end{aligned}$$

if S_1 and $d(p,q)$ are sufficiently large. □

Our thanks are due to Paul Erdös, who communicated the problem discussed in this note to us. The question originated in computer science. If the Euclidean metrics could be approximated by graph metrics, then we could run efficient graph algorithms for solving geometric problems on-screen.