ON THE ERDŐS DISTANCE PROBLEM

T. AGAMA

In loving memory of Doctor Margaret Lesley McIntyre

Abstract. In this paper, using the method of compression, we recover the lower bound for the Erdős unit distance problem and provide an alternative proof to the distinct distance conjecture. In particular, in \( \mathbb{R}^k \) for all \( k \geq 2 \), we have

\[
\# \left\{ ||\vec{x}_j - \vec{x}_t|| : ||\vec{x}_j - \vec{x}_t|| = 1, 1 \leq t, j \leq n, \vec{x}_t, \vec{x}_j \in \mathbb{R}^k \right\} \geq C\frac{\sqrt{k}}{2}n^{1+o(1)}
\]

for some \( C > 0 \). We also show that

\[
\# \left\{ d_j : d_j = ||\vec{x}_s - \vec{y}_t||, d_j \neq d_i, 1 \leq s, t \leq n \right\} \geq D\frac{\sqrt{k}}{2}n^{\frac{5}{2}-o(1)}
\]

for some \( D > 0 \). These lower bounds generalizes the lower bounds of the Erdős unit distance and the distinct distance problem to higher dimensions.

1. Introduction

The Erdős distinct distance conjecture is the assertion that

**Conjecture 1.1.** The number of distinct distances that can be formed from \( n \) points in the plane should at least be \( n^{1-o(1)} \).

Progress on this conjecture has developed overtime. Let us denote \( g(n) \) the counting function for such construction. Then the first lower bound of the form

\[
g(n) \gg n^\frac{2}{3}
\]

was given in \( \mathbb{R} \), which improves on an earlier version of Erdős. This was eventually improved to

\[
g(n) \gg \frac{n^{\frac{4}{3}}}{\log n}
\]

in \( \mathbb{R} \) and

\[
g(n) \gg n^{\frac{5}{8}}
\]

in \( \mathbb{R}^k \). The best currently known lower bound can be found in \( \mathbb{R}^k \), which essentially solves the problem. In this paper by using the method of compression and its accompanied estimates, we provide an alternative solution to the conjecture in the following result:

**Theorem 1.1.**

\[
\# \left\{ d_j : d_j = ||\vec{x}_s - \vec{y}_t||, d_j \neq d_i, 1 \leq s, t \leq n, \vec{x}, \vec{y} \in \mathbb{R}^k \right\} \geq D\frac{\sqrt{k}}{2}n^{\frac{5}{2}-o(1)}
\]
for some $D > 0$.

Using this method, we provide a lower bound for the Erdős unit distance problem, that takes into consideration the dimension of the space in which the points reside in the form:

**Theorem 1.2.** Let $I = \{ ||\vec{x}_j - \vec{x}_t|| : ||\vec{x}_j - \vec{x}_t|| = 1, 1 \leq t, j \leq n, \vec{x}_j, \vec{x}_t \in \mathbb{R}^k, \}$, then we have

$$\#I \geq C \sqrt{\frac{k}{2}} n^{1+o(1)}$$

for some $C > 0$.

### 1.1. Notations and conventions.

Throughout this paper, we will assume that $n$ is sufficiently large for any number of $n$ points in the euclidean plane. We write $f(s) \gg g(s)$ if there exists a constant $c > 0$ such that $f(s) \geq c|g(s)|$ for all $s$ sufficiently large. If the constant depends of some variable, say $t$, then we denote the inequality by $f(s) \gg_t g(s)$. We write $f(s) = o(g(s))$ if the limits holds

$$\lim_{s \to \infty} \frac{f(s)}{g(s)} = 0.$$  

In particular, $f(s) = o(1)$ implies that $f(s) \to 0$ as $s \to \infty$.

### 2. Compression

In this section we launch the notion of compression of points in space. We study the mass of compression and its accompanied estimates. These estimates turn out to be useful for estimating the gap of compression, which we will launch in the sequel.

**Definition 2.1.** By the compression of scale $0 < m \leq 1$ on $\mathbb{R}^n$, we mean the map $V_m : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$V_m[(x_1, x_2, \ldots, x_n)] = \left( \frac{m}{x_1}, \frac{m}{x_2}, \ldots, \frac{m}{x_n} \right)$$

for $n \geq 2$ and with $x_i \neq 0$ for all $i = 1, \ldots, n$.

**Remark 2.2.** The notion of compression is in some way the process of rescaling points in $\mathbb{R}^n$ for $n \geq 2$. Thus it is important to notice that a compression roughly speaking pushes points very close to the origin away from the origin by certain scale and similarly draws points away from the origin close to the origin. Intuitively, compression induces some kind of motion on points in the Euclidean space.

**Proposition 2.1.** A compression of scale $0 < m \leq 1$ with $V_m : \mathbb{R}^n \to \mathbb{R}^n$ is a bijective map. In particular the compression $V_m : \mathbb{R}^n \to \mathbb{R}^n$ is a bijective map of order 2.

**Proof.** Suppose $V_m[(x_1, x_2, \ldots, x_n)] = V_m[(y_1, y_2, \ldots, y_n)]$, then it follows that

$$\left( \frac{m}{x_1}, \frac{m}{x_2}, \ldots, \frac{m}{x_n} \right) = \left( \frac{m}{y_1}, \frac{m}{y_2}, \ldots, \frac{m}{y_n} \right).$$

It follows that $x_i = y_i$ for each $i = 1, 2, \ldots, n$. Surjectivity follows by definition of the map. Thus the map is bijective. The latter claim follows by noting that $V_m^2[\vec{x}] = \vec{x}$. \qed
2.1. **The mass of compression estimates.** In this section we study the mass of a compression in a given scale. We use the upper and lower estimates of the mass of compression to establish corresponding estimates for the gap of compression. These estimates will form an essential tool for establishing the main result of this paper.

**Definition 2.3.** By the mass of a compression of scale $0 < m \leq 1$ we mean the map $M : \mathbb{R}^n \rightarrow \mathbb{R}$ such that

$$M(V_m[(x_1, x_2, \ldots, x_n)]) = \sum_{i=1}^{n} \frac{m}{x_i}.$$ 

**Lemma 2.4.** We have

$$\sum_{n \leq x} \frac{1}{n} = \log x + \gamma + O \left( \frac{1}{x} \right)$$

where $\gamma = 0.5772 \cdots$.

**Remark 2.5.** Next we prove upper and lower bounding the mass of the compression of scale $0 < m \leq 1$.

**Proposition 2.2 (The mass of compression estimates).** Let $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ with $x_i \neq x_j$ for $1 \leq i, j \leq n$ with $i \neq j$ with $x_i \neq 0$ for all $1 \leq i \leq n$, then the estimates hold

$$m \log \left( 1 - \frac{n-1}{\sup(x_j)} \right) - 1 \ll M(V_m[(x_1, x_2, \ldots, x_n)]) \ll m \log \left( 1 + \frac{n-1}{\inf(x_j)} \right)$$

for $n \geq 2$.

**Proof.** Let $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ for $n \geq 2$ and $x_i \neq x_j$ ($i \neq j$) with $x_i \neq 0$ for all $1 \leq i \leq n$. Then it follows that

$$M(V_m[(x_1, x_2, \ldots, x_n)]) = m \sum_{j=1}^{n} \frac{1}{x_j} \leq m \sum_{k=0}^{n-1} \frac{1}{\inf(x_j) + k}$$

and the upper estimate follows by the estimate for this sum by appealing to Lemma 2.4. The lower estimate also follows by noting the lower bound

$$M(V_m[(x_1, x_2, \ldots, x_n)]) = m \sum_{j=1}^{n} \frac{1}{x_j} \geq m \sum_{k=0}^{n-1} \frac{1}{\sup(x_j) - k}.$$

\[\square\]

It is important to notice that the condition $x_i \neq x_j$ for $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ is not only a quantifier but a requirement; otherwise, the statement for the mass of compression will be flawed completely. To wit, suppose we take $x_1 = x_2 = \cdots = x_n$, then it will follow that $\inf(x_j) = \sup(x_j)$, in which case the mass of compression of scale $m$ satisfies

$$m \sum_{k=0}^{n-1} \frac{1}{\inf(x_j) - k} \leq M(V_m[(x_1, x_2, \ldots, x_n)]) \leq m \sum_{k=0}^{n-1} \frac{1}{\inf(x_j) + k}.$$
and it is easy to notice that this inequality is absurd. By extension one could also try to equalize the sub-sequence on the bases of assigning the supremum and the infimum and obtain an estimate but that would also contradict the mass of compression inequality after a slight reassignment of the sub-sequence. Thus it is important for the estimate to make any good sense to ensure that any tuple \((x_1, x_2, \ldots, x_n) \in \mathbb{R}^n\) must satisfy \(x_i \neq x_j\) for all \(1 \leq i, j \leq n\). Hence in this paper this condition will be highly extolled. In situations where it is not mentioned, it will be assumed that the tuple \((x_1, x_2, \ldots, x_n) \in \mathbb{R}^n\) is such that \(x_i \neq x_j\) for \(1 \leq i, j \leq n\).

2.2. Compression gap estimates. In this section we recall the notion of the gap of compression and its various estimates. We prove upper and lower bounding the gap of a point under compression of any scale.

**Definition 2.6.** Let \((x_1, x_2, \ldots, x_n) \in \mathbb{R}^n\) with \(x_i \neq 0\) for all \(i = 1, 2, \ldots, n\). Then by the gap of compression of scale \(m\) for the compression \(V_m\), denoted \(G \circ V_m[(x_1, x_2, \ldots, x_n)]\), we mean the quantity

\[
G \circ V_m[(x_1, x_2, \ldots, x_n)] = \left\| \left( x_1 - \frac{m}{x_1}, x_2 - \frac{m}{x_2}, \ldots, x_n - \frac{m}{x_n} \right) \right\|
\]

**Proposition 2.3.** Let \((x_1, x_2, \ldots, x_n) \in \mathbb{R}^n\) for \(n \geq 2\) with \(x_j \neq 0\) for \(j = 1, \ldots, n\), then we have

\[
G \circ V_m[(x_1, x_2, \ldots, x_n)]^2 = M \circ \nu_1 \left[ \left( \frac{1}{x_1^2}, \ldots, \frac{1}{x_n^2} \right) \right] + m^2 M \circ \nu_1[(x_1^2, \ldots, x_n^2)] - 2mn.
\]

In particular, if each \(x_i > 1\) for \(1 \leq i \leq n\), we have the estimate

\[
G \circ V_m[(x_1, x_2, \ldots, x_n)]^2 = M \circ \nu_1 \left[ \left( \frac{1}{x_1^2}, \ldots, \frac{1}{x_n^2} \right) \right] - 2mn + O \left( m^2 M \circ \nu_1[(x_1^2, \ldots, x_n^2)] \right)
\]

with \(m := m(n) = o(1)\) as \(n \to \infty\).

Proposition 2.3 offers us an extremely useful identity. It allows us to pass from the gap of compression on points to the relative distance to the origin. It tells us that points under compression with a large gap must be far away from the origin than points with a relatively smaller gap under compression. That is to say, the inequality

\[
G \circ \nu_m[\bar{x}] < G \circ \nu_m[\bar{y}]
\]

with \(m := m(n) = o(1)\) as \(n \to \infty\) if and only if \(||\bar{x}|| < ||\bar{y}||\) for \(\bar{x}, \bar{y} \in \mathbb{R}^n\) with each \(x_i, y_i \geq 1\) for all \(1 \leq i \leq n\). This important transference principle will be mostly put to use in obtaining our results.

**Corollary 2.1.** Let \((x_1, x_2, \ldots, x_n) \in \mathbb{R}^n\) for \(n \geq 2\) with \(x_j \neq x_i\) for \(j \neq i\) and \(x_i, x_j \geq 1\) for each \(1 \leq i, j \leq n\). If \(m := m(n) = o(1)\) as \(n \to \infty\), then we have

\[
G \circ \nu_m[(x_1, x_2, \ldots, x_n)]^2 \geq n \text{Inf}(x_j^2) - 2mn + O \left( m^2 M \circ \nu_1[(x_1^2, \ldots, x_n^2)] \right)
\]

and

\[
G \circ \nu_m[(x_1, x_2, \ldots, x_n)]^2 \leq \text{nSup}(x_j^2) - 2mn + O \left( m^2 M \circ \nu_1[(x_1^2, \ldots, x_n^2)] \right)
\]

**Lemma 2.7** (Compression gap estimates). Let \((x_1, x_2, \ldots, x_n) \in \mathbb{R}^n\) for \(n \geq 2\) with \(x_j \neq x_i\) for \(j \neq i\) and \(x_i, x_j \geq 1\) for each \(1 \leq i, j \leq n\). If \(m := m(n) = o(1)\) as \(n \to \infty\), then we have

\[
G \circ \nu_m[(x_1, x_2, \ldots, x_n)]^2 \ll \text{nSup}(x_j^2) + m^2 \log \left( 1 + \frac{n-1}{\text{Inf}(x_j^2)} \right) - 2mn
\]
and
\( \mathcal{G} \circ \mathcal{V}_m[(x_1, x_2, \ldots, x_n)]^2 \gg n \text{Inf}(x_j^2) + m^2 \log \left( 1 - \frac{n - 1}{\text{sup}(x_j^2)} \right)^{-1} - 2mn. \)

**Proof.** The estimates follows by leveraging the estimates in Proposition 2.2 and noting that
\[ n \text{Inf}(x_j^2) \ll \mathcal{M} \circ \mathcal{V}_1 \left[ \left( \frac{1}{x_1^2}, \ldots, \frac{1}{x_n^2} \right) \right] \ll n \text{sup}(x_j^2). \]

\[ \square \]

**Remark 2.8.** It is important to note that the inequality in Corollary 2.1 implies the inequalities in Lemma 2.7. At any given moment, we will decide which of the versions of these inequalities to use. Indeed the inequalities in Corollary 2.1 are mostly applicable to various problems that the one in Lemma 2.7.

In this paper, when we say points are concentrated around the origin; In particular, if a set of \( n \) points in \( \mathbb{R}^k \) are concentrated around the origin, we mean
\[ \inf_{1 \leq j \leq n} x_j = \sup_{1 \leq j \leq n} x_j = n^{o(1)} \]
for \( 1 \leq j \leq n. \)

### 3. Application to the Erdős unit distance and distinct distance conjecture

In this section we leverage the estimate of the gap of compression to study the problem of determining the number of unit distances that can be formed from \( n \) points. We state our main theorem that takes into consideration the dimension of the space in which the points reside.

**Theorem 3.1.** Let \( \mathcal{I} = \left\{ \|x_j - x_t\| : \|x_j - x_t\| = 1, 1 \leq t, j \leq n, x_j, x_t \in \mathbb{R}^k \right\} \), then we have
\[ \# \mathcal{I} \geq C \frac{\sqrt{k}}{2} n^{1 + o(1)} \]
for some \( C > 0. \)

**Proof.** First, we set \( m := m(k) = o(1) \) as \( k \to \infty \) and carefully choose \( n \) points \( \bar{x}_j \) for \( 1 \leq j \leq n \) in \( \mathbb{R}^k \) such that \( \left\lceil \frac{n}{2} \right\rceil \) of these points also has their image points under compression. In particular, we choose \( n \) points such that \( \left\lceil \frac{n}{2} \right\rceil \) of those points \( \bar{x}_j \) satisfies \( \inf_{i=1}^k(x_j)_i = \sup_{i=1}^k(x_j)_i = n^{o(1)} \) and for each of these points we also include their image points under compression \( \mathcal{V}_m(\bar{x}_j) \). This ensures that \( \|\bar{x}_j - \)
\( V_m[\vec{x}_j] = 1 \) for all \( n \) sufficiently large. Consequently, we have
\[
\#I = \# \left\{ \left| x_j^* - x_t^* \right| : \left| x_j^* - x_t^* \right| = 1, \ 1 \leq t, j \leq n, \ x_j^*, x_t^* \in \mathbb{R}^k \right\}
\]
\[
\geq \# \left\{ \left| x_j^* - \bar{x}_t^* \right| : \left| x_j^* - \bar{x}_t^* \right| = 1, \ 1 \leq t, j \leq n, \ x_j^*, \bar{x}_t^* \in \mathbb{R}^k, \min\{\text{Inf}(x_j, )\}_{1 \leq j \leq \frac{n}{2}} = n^{o(1)} \right\}
\]
\[
\geq \# \left\{ \left| x_j^* - \bar{x}_t^* \right| : x_j^* \in \mathbb{R}^k, \left| x_j^* - \bar{x}_t^* \right| = 1, \ 1 \leq j \leq \frac{n}{2}, \ V_1[\bar{x}_j] = \bar{x}_t, \min\{\text{Inf}(x_j, )\}_{1 \leq j \leq \frac{n}{2}} = n^{o(1)} \right\}.
\]
It is important to note that the quantity above cannot be zero, since the condition
\[
\min\{\text{Inf}(x_j, )\}_{1 \leq j \leq \frac{n}{2}} = n^{o(1)}
\]
is required for \( \left| x_j^* - \bar{x}_t^* \right| = 1 \) with \( V_1[\bar{x}_j] = \bar{x}_t \) and the configuration exists by construction. This follows from the requirement of the construction that the chosen set of points in \( \mathbb{R}^k \) has \( \frac{n}{2} \) of the points concentrated around the origin, which mean
\[
\inf(x_j, )_{i=1} = \sup(x_j, )_{i=1} = n^{o(1)}
\]
for \( 1 \leq j \leq \frac{n}{2} \). The right-hand side is basically the sum
\[
\sum_{1 \leq j \leq \frac{n}{2}} G \circ V_m[\vec{x}_j] = \sum_{1 \leq j \leq \frac{n}{2}} G \circ V_m[\vec{x}_j] = n^{o(1)}
\]
By taking \( m := m(k) = o(1) \) as \( k \rightarrow \infty \), in particular, if we choose \( m = O\left(\frac{1}{\log k}\right) \) then we have the lower bound for the right hand side
\[
\sum_{1 \leq j \leq \frac{n}{2}} G \circ V_m[\vec{x}_j] \geq \sum_{1 \leq j \leq \frac{n}{2}} C \text{Inf}(x_j, )_{1 \leq s \leq k} \sqrt{k}
\]
\[
\geq C \frac{\sqrt{k}}{2} \min\{\text{Inf}(x_j, )\}_{1 \leq j \leq \frac{n}{2}} = n^{o(1)}
\]
\[
= C \frac{\sqrt{k}}{2} n^{1+o(1)}
\]
for some \( C > 0 \), by an application of Lemma 2.7 which establishes the claimed lower bound for the construction. \( \square \)

It is important to point out that the lower estimate for the construction provided in Theorem 2.11 was achieved by counting not all possible unit distances but only unit distance that correspond to compression gaps of unit length. We state the second theorem as an application, which gives lower bound for the number of distinct distances that can be formed from \( n \) points in the Euclidean space \( \mathbb{R}^k \) for all \( k \geq 2 \).
Theorem 3.2. We have
\[ \# \{ d_j : d_j = ||x_s' - y_t'||, d_j \neq d_i, 1 \leq s, t \leq n, x, y \in \mathbb{R}^k \} \geq D \frac{\sqrt{k}}{2} n^{\frac{2}{3} - o(1)} \]
for some $D > 0$.

Proof. First, we set $m := m(k) = o(1)$ as $k \rightarrow \infty$ and carefully choose $n$ points $x_j$ for $1 \leq j \leq n$ in $\mathbb{R}^k$ such that $\left\lceil \frac{n}{m} \right\rceil$ of these points also have their image points under compression. That is, for each $x_j$ we also include $V_m[x_j]$. Next for $\left\lceil \frac{n}{m} \right\rceil$ of those points we make the assignment $\sup(x_{j_i}) = n^{1 - \frac{2}{3} + \epsilon}$ for any small $\epsilon := \epsilon(i) > 0$ and $\inf(x_{j_i}) \geq 1$. This ensures that
\[ \max_{1 \leq j \leq n} G \circ V_m[x_j] = n^{1 - \frac{2}{3} + \epsilon} \]
for any small $\epsilon := \epsilon(i) > 0$. Now, we let $\{ d_j : d_j = ||x_s' - y_t'||, d_j \neq d_i, 1 \leq s, t \leq n, x, y \in \mathbb{R}^k \} = R$, then we notice that
\[
\# R \geq \# \left\{ d_j : d_j = ||x_s' - y_t'||, d_j \neq d_i, 1 \leq s, t \leq n, x, y \in \mathbb{R}^k, \sup(d_j) = n^{1 - \frac{2}{3} + o(1)} \right\}
\geq \# \left\{ d_j : d_j = G \circ V_m[x_j'], d_j \neq d_i, 1 \leq j \leq \frac{n}{2}, \sup(d_j) = n^{1 - \frac{2}{3} + o(1)}, x_j' \in \mathbb{R}^k, x_j \ (1 \leq s \leq k) \right\}
\geq 1, \ V[x_j'] = x_k'
\]
\[
= \sum_{d_j = G \circ V_m[x_j'], 1 \leq j \leq \frac{n}{2}, \sup(d_j) = n^{1 - \frac{2}{3} + o(1)}, x_j' \in \mathbb{R}^k, d_j \neq d_i, 1 \neq j} \frac{G \circ V_m[x_j']}{d_j}
\geq D \sqrt{k} \sum_{1 \leq j \leq \frac{n}{2}, \sup(d_j) = n^{1 - \frac{2}{3} + o(1)}, x_j' \in \mathbb{R}^k, d_j \neq d_i, 1 \neq j} \frac{\inf(x_{j_s})}{d_j} 1 \leq s \leq k
\geq D \sqrt{k} \sum_{1 \leq j \leq \frac{n}{2}, \sup(d_j) = n^{1 - \frac{2}{3} + o(1)}, d_j \neq d_i, 1 \neq j} \frac{1}{d_j}
\geq D \sqrt{k} \sum_{1 \leq j \leq \frac{n}{2}, \sup(d_j) = n^{1 - \frac{2}{3} + o(1)}, d_j \neq d_i, 1 \neq j} \frac{\varphi}{\sup(d_j)} 1 \leq j \leq \frac{n}{2}
\geq D \frac{\sqrt{k}}{2} n^{\frac{2}{3} - o(1)}
\]
and the claimed lower bound follows by Lemma 2.7 for this construction.

It needs to be said that the result in Theorem 3.2 can be viewed as providing an alternate solution to the Erdős distinct distance problem, that takes into consideration the dimension of the space in which the points reside. The lower bound of this type, it has to be said, exists in the literature (See [1]). But the method employed is completely different from the one we have used here. Theorem 3.1 and Theorem 3.2 can be considered as a generalization of the solution to both versions of the Erdős distance problem to any euclidean space of dimension \( k \geq 2 \). In particular we have the following theorems as consequences of the main results of this paper.

**Theorem 3.3.** The number of distinct distances that can be formed from \( n \) points in any euclidean space \( \mathbb{R}^{2n} \) for \( n \geq 2 \) is at least

\[
\geq D \frac{\sqrt{2}}{2} n^{\frac{1}{4} + \frac{1}{2} - o(1)}
\]

for some \( D > 0 \).

**Theorem 3.4.** The number of distinct distances that can be formed from \( n \) points in a euclidean space of dimension \( n^2 \) for \( n \geq 2 \) is at least

\[
\geq D \frac{n^{\frac{1}{2} + 1 - o(1)}}{2}
\]

for some \( D > 0 \).

\[ \square \]

**References**

1. Guth, Larry and Katz, Nets Hawk, On the Erdős distinct distances problem in the plane, Annals of Mathematics, JSTOR, 2015, pp 155–190.
2. Chung, Fan RK and Szemerédi, Endre and Trotter, William T. The number of different distances determined by a set of points in the Euclidean plane, Discrete & Computational Geometry, vol. 7:1, Springer, 1992, pp 1–11.
3. Moser, Leo On the different distances determined by \( n \) points, The American Mathematical Monthly, vol. 59:2, Taylor & Francis, 1952, pp 85–91.
4. Solymosi, József and Tóth, Cs D Distinct distances in the plane, Discrete & Computational Geometry, vol. 25:4, Springer, 2001, pp 629–634.

Department of Mathematics, African Institute for Mathematical Science, Ghana

Email address: theophilus@aims.edu.gh/emperordagama@yahoo.com