A QUANTITATIVE VERSION OF THE ERDŐS-ANNING THEOREM

T. AGAMA

Abstract. Let $\mathcal{R}\subset \mathbb{R}^n$ be an infinite set of collinear points and $\mathcal{S}\subset \mathcal{R}$ be an arbitrary and finite set with $\mathcal{S}\subset \mathbb{N}^n$. Then the number of points with mutual integer distances on the shortest line containing points in $\mathcal{S}$ satisfies the lower bound

$$\gg_n |\mathcal{S}|\sqrt{n}\mathrm{min}_{\vec{x}\in \mathcal{S}}\mathrm{Inf}(x_j)_{j=1}^{n}\sum \limits_{\substack{k\leq \mathrm{max}_{\vec{x}\in \mathcal{S}}\mathcal{G}\circ \mathbb{V}_1[\vec{x}]\\k\in \mathbb{N}\\k>1}}\frac{1}{k},$$

where $\mathcal{G}\circ \mathbb{V}_1[\vec{x}]$ is the compression gap of the compression induced on $\vec{x}$. This proves that there are infinitely many collinear points with mutual integer distances on any line in $\mathbb{R}^n$ and generalizes the well-known Erdős-Anning Theorem in the plane $\mathbb{R}^2$.

1. **Introduction**

The well-known Erdős-Anning Theorem is the assertion that infinite number of points in the plane $\mathbb{R}^2$ have can mutual integer distances only if all the points lie on the straight line. The theorem was first proved by Paul Erdős and Norman H. Anning [1]. In this paper we obtain a quantitative lower bound for the number of points with mutual integer distances in any finite subset of an infinite set of points on the same line in the space $\mathbb{R}^n$. In particular for any finite subset $\mathcal{S}\subset \mathbb{N}^2$ of an infinite set of points on the same line in the plane $\mathbb{R}^2$ the number of points on the shortest line with mutual integer distances containing points in $\mathcal{S}$ must satisfy the lower bound

$$\gg_2 |\mathcal{S}|\sqrt{2}\mathrm{min}_{\vec{x}\in \mathcal{S}}\mathrm{Inf}(x_j)_{j=1}^{2}\sum \limits_{\substack{k\leq \mathrm{max}_{\vec{x}\in \mathcal{S}}\mathcal{G}\circ \mathbb{V}_1[\vec{x}]\\k\in \mathbb{N}\\k>1}}\frac{1}{k}$$

where $\mathcal{G}\circ \mathbb{V}_1[\vec{x}]$ is the compression gap of the compression induced on $\vec{x}\in \mathbb{N}^2$. As it is being hinted at the notion of compression developed and the tools developed therein (see [2]) plays an instrumental role. By applying the notion of the mass of compression, the compression gap and associated estimates with the notion of the lines induced by compression on points in space, we can get a handle on a lower bound for any such points on the line. The immediate consequence of this is the assertion that there are infinitely many points with mutual distances on a line in a plane. So our result in a way supplies an estimate to the Erdős-Anning Theorem in the plane and more generally in the space $\mathbb{R}^n$ for $n\geq 2$.

Date: July 29, 2021.

2000 *Mathematics Subject Classification.* Primary 54C40, 14E20; Secondary 46E25, 20C20.

Key words and phrases. points; collinear.

The notations we have adopted in this paper are quite cumbersome so we will feel the need to clarify them at appropriate places in the paper were it is used.

In the sequel we write $f(n) \gg g(n)$ to mean there exists some constant $c > 0$ such the inequality $f(n) \geq cg(n)$ holds for all sufficiently large values of n. In situations where the constant depends on some variable say s then we will write $f(n) \gg_s g(n)$. Similarly we write $f(n) \ll g(n)$ if there exists some constant $c > 0$ such that we write $f(n) \leq cg(n)$ for all sufficiently large values of n.

2. **Preliminary results**

Definition 2.1. By the compression of scale $m > 0$ ($m \in \mathbb{R}$) fixed on $\mathbb{R}^n$ we mean the map $\mathbb{V} : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ such that

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

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

Remark 2.2. The notion of compression is in some way the process of re scaling 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.

Proposition 2.1. *A compression of scale $m > 0$ with $\mathbb{V}_m : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ is a bijective map.*

Proof. Suppose $\mathbb{V}_m[(x_1, x_2, \ldots, x_n)] = \mathbb{V}_m[(y_1, y_2, \ldots, y_n)]$, 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. $\square$

2.1. **The mass of compression.**

In this section we recall the notion of the mass of compression on points in space and study the associated statistics.

Definition 2.3. By the mass of a compression of scale $m > 0$ ($m \in \mathbb{R}$) fixed, we mean the map $\mathcal{M} : \mathbb{R}^n \longrightarrow \mathbb{R}$ such that

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

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 follows that $\mathrm{Inf}(x_j) = \mathrm{Sup}(x_j)$, in which case the mass of compression of scale m satisfies

$$m\sum_{k=0}^{n-1} \frac{1}{\mathrm{Inf}(x_j) - k} \leq \mathcal{M}(\mathbb{V}_m[(x_1, x_2, \ldots, x_n)]) \leq m\sum_{k=0}^{n-1} \frac{1}{\mathrm{Inf}(x_j) + k}$$

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 $x_i \leq x_j$ for $1 \leq i, j \leq n$.

Lemma 2.4. *The estimate holds*

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

where $\gamma = 0.5772 \cdots$.

Remark 2.5. Next we prove upper and lower bounding the mass of the compression of scale $m \geq 1$.

Proposition 2.2. *Let* $(x_1, x_2, \ldots, x_n) \in \mathbb{N}^n$, *then the estimates holds*

$$m \log\left(1 - \frac{n-1}{\sup(x_j)}\right)^{-1} \ll \mathcal{M}(\mathbb{V}_m[(x_1, x_2, \ldots, x_n)]) \ll m \log\left(1 + \frac{n-1}{\mathrm{Inf}(x_j)}\right)$$

for $n \geq 2$.

Proof. Let $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ for $n \geq 2$ with $x_j \geq 1$. Then it follows that

$$\begin{aligned}\mathcal{M}(\mathbb{V}_m[(x_1, x_2, \ldots, x_n)]) &= m \sum_{j=1}^{n} \frac{1}{x_j} \\ &\leq m \sum_{k=0}^{n-1} \frac{1}{\mathrm{Inf}(x_j) + k}\end{aligned}$$

and the upper estimate follows by the estimate for this sum. The lower estimate also follows by noting the lower bound

$$\begin{aligned}\mathcal{M}(\mathbb{V}_m[(x_1, x_2, \ldots, x_n)]) &= m \sum_{j=1}^{n} \frac{1}{x_j} \\ &\geq m \sum_{k=0}^{n-1} \frac{1}{\sup(x_j) - k}.\end{aligned}$$

□

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 $\mathbb{V}_m$, denoted $\mathcal{G} \circ \mathbb{V}_m[(x_1, x_2, \ldots, x_n)]$, we mean the expression

$$\mathcal{G} \circ \mathbb{V}_m[(x_1, x_2, \ldots, x_n)] = \left|\left|\left(x_1 - \frac{m}{x_1}, x_2 - \frac{m}{x_2}, \ldots, x_n - \frac{m}{x_n}\right)\right|\right|$$

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*

$$\mathcal{G} \circ \mathbb{V}_m[(x_1, x_2, \ldots, x_n)]^2 = \mathcal{M} \circ \mathbb{V}_1\left[\left(\frac{1}{x_1^2}, \ldots, \frac{1}{x_n^2}\right)\right] + m^2 \mathcal{M} \circ \mathbb{V}_1[(x_1^2, \ldots, x_n^2)] - 2mn.$$

In particular, we have the estimate

$$\mathcal{G}\circ\mathbb{V}_m[(x_1,x_2,\ldots,x_n)]^2=\mathcal{M}\circ\mathbb{V}_1\left[\left(\frac{1}{x_1^2},\ldots,\frac{1}{x_n^2}\right)\right]-2mn+O\bigg(m^2\mathcal{M}\circ\mathbb{V}_1[(x_1^2,\ldots,x_n^2)]\bigg)$$

for $\vec{x}\in\mathbb{N}^n$*, where* $m^2\mathcal{M}\circ\mathbb{V}_1[(x_1^2,\ldots,x_n^2)]$ *is the error term in this case.*

Lemma 2.7 (Compression estimate)**.** *Let* $(x_1,x_2,\ldots,x_n)\in\mathbb{N}^n$ *for* $n\geq 2$*, then we have*

$$\mathcal{G}\circ\mathbb{V}_m[(x_1,x_2,\ldots,x_n)]^2\ll n\mathrm{sup}(x_j^2)+m^2\log\left(1+\frac{n-1}{\mathrm{Inf}(x_j)^2}\right)-2mn$$

and

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

3. **Compression lines**

In this section we study the notion of lines induced under compression of a given scale and the associated geometry. We first launch the following language.

Definition 3.1. Let $\vec{x}=(x_1,x_2,\ldots,x_n)\in\mathbb{R}^n$ with $x_1\neq 0$ for $1\leq i\leq n$. Then by the line $L_{\vec{x},\mathbb{V}_m[\vec{x}]}$ produced under compression $\mathbb{V}_m:\mathbb{R}^n\longrightarrow\mathbb{R}^n$ we mean the line joining the points $\vec{x}$ and $\mathbb{V}_m[\vec{x}]$ given by

$$\vec{r}=\vec{x}+\lambda(\vec{x}-\mathbb{V}_m[\vec{x}])$$

where $\lambda\in\mathbb{R}$.

Remark 3.2. In striving for the simplest possible notation and to save enough work space, we will choose instead to write the line produced under compression $\mathbb{V}_m:\mathbb{R}^n\longrightarrow\mathbb{R}^n$ by $L_{\mathbb{V}_m[\vec{x}]}$. Next we show that the lines produced under compression of two distinct points not on the same line of compression cannot intersect at the corresponding points and their images under compression.

Lemma 3.3. *Let* $\vec{a}=(a_1,a_2,\ldots,a_n)\in\mathbb{R}^n$ *with* $\vec{a}\neq\vec{x}$ *and* $a_i,x_j\neq 0$ *for* $1\leq i,j\leq n$*. If the point* $\vec{a}$ *lies on the corresponding line* $L_{\mathbb{V}_m[\vec{x}]}$*, then* $\mathbb{V}_m[\vec{a}]$ *also lies on the same line.*

Proof. Pick arbitrarily a point $\vec{a}$ on the line $L_{\mathbb{V}_m[\vec{x}]}$ produced under compression for any $\vec{x}\in\mathbb{R}^n$. Suppose on the contrary that $\mathbb{V}_m[\vec{a}]$ cannot live on the same line as $\vec{a}$. Then $\mathbb{V}_m[\vec{a}]$ must be away from the line $L_{\mathbb{V}_m[\vec{x}]}$. Produce the compression line $L_{\mathbb{V}_m[\vec{a}]}$ by joining the point $\vec{a}$ to the point $\mathbb{V}_m[\vec{a}]$ by a straight line. Then It follows from Proposition 2.3

$$\mathcal{G}\circ\mathbb{V}_m[\vec{x}]>\mathcal{G}\circ\mathbb{V}_m[\vec{a}].$$

Again pick a point $\vec{c}$ on the line $L_{\mathbb{V}_m[\vec{a}]}$, then under the assumption it follows that the point $\mathbb{V}_m[\vec{c}]$ must be away from the line. Produce the compression line $L_{\mathbb{V}_m[\vec{c}]}$ by joining the points $\vec{c}$ to $\mathbb{V}_m[\vec{c}]$. Then by Proposition 2.3 we obtain the following decreasing sequence of lengths of distinct lines

$$\mathcal{G}\circ\mathbb{V}_m[\vec{x}]>\mathcal{G}\circ\mathbb{V}_m[\vec{a}]>\mathcal{G}\circ\mathbb{V}_m[\vec{c}].$$

By repeating this argument, we obtain an infinite descending sequence of lengths of distinct lines

$$\mathcal{G} \circ \mathbb{V}_m[\vec{x}] > \mathcal{G} \circ \mathbb{V}_m[\vec{a_1}] > \cdots > \mathcal{G} \circ \mathbb{V}_m[\vec{a_n}] > \cdots .$$

This proves the Lemma. □

Proposition 3.1. *Let $\vec{a} = (a_1, a_2, \ldots, a_n) \in \mathbb{R}^n$ with $\vec{a} \neq \vec{x}$ and $a_i, x_j \neq 0$ for $1 \leq i, j \leq n$. Also let $\mathcal{G} \circ \mathbb{V}_m[\vec{x}], \mathcal{G} \circ \mathbb{V}_m[\vec{a}] \in \mathbb{N}$. If the point $\vec{a}$ lies on the corresponding line $L_{\mathbb{V}_m[\vec{x}]}$ then the mutual distances between the points $\vec{x}, \vec{a}, \mathbb{V}_m[\vec{x}], \mathbb{V}_m[\vec{a}]$ are also integers.*

Proof. Under the main assumption with $\mathcal{G} \circ \mathbb{V}_m[\vec{x}], \mathcal{G} \circ \mathbb{V}_m[\vec{a}] \in \mathbb{N}$ then appealing to Lemma 3.3, we have the inequality

$$\mathcal{G} \circ \mathbb{V}_m[\vec{x}] > \mathcal{G} \circ \mathbb{V}_m[\vec{a}]$$

and the line $L_{\mathbb{V}_m[\vec{a}]}$ is only a segment of the line $L_{\mathbb{V}_m[\vec{x}]}$ by virtue of the estimate in Proposition 2.3 so that $||\vec{x} - \vec{a}||, ||\mathbb{V}_m[\vec{x}] - \mathbb{V}_m[\vec{a}], ||\vec{x} - \mathbb{V}_m[\vec{a}], ||\vec{a} - \mathbb{V}_m[\vec{x}]|| \in \mathbb{N}$. This completes the proof of the proposition. □

4. **Main result**

In this section we prove the main result of this paper.

Theorem 4.1. *Let $\mathcal{R} \subset \mathbb{R}^n$ be an infinite set of collinear points and $\mathcal{S} \subset \mathcal{R}$ be an arbitrary and finite set with $\mathcal{S} \subset \mathbb{N}^n$. Then the number of points with mutual integer distances on the shortest line containing points in $\mathcal{S}$ satisfies the lower bound*

$$\gg_n |\mathcal{S}|\sqrt{n}\min_{\vec{x} \in \mathcal{S}} \mathrm{Inf}(x_j)_{j=1}^n \sum_{\substack{k \leq \max_{\vec{x} \in \mathcal{S}} \mathcal{G} \circ \mathbb{V}_1[\vec{x}] \\ k \in \mathbb{N} \\ k > 1}} \frac{1}{k}.$$

Proof. Let us pick arbitrarily the lattice point $\vec{x} \in \mathbb{N}^n$ and apply the compression $\mathbb{V}_1[\vec{x}]$. Next construct the line induced by compression $L_{\mathbb{V}_1[\vec{x}]}$ and cover all the points on the line by the set $\mathcal{R}$. Let us choose $\mathcal{S} \subset \mathcal{R}$ be the set of all lattice points on the line of compression induced. Now, the number of points with mutual integer distances on the line of compression $L_{\mathbb{V}_1[\vec{x}]}$ can be lower bounded by virtue of Lemma 3.3 by the sum

$$\begin{aligned}
\sum_{\substack{\vec{x} \in \mathcal{S} \\ \mathcal{G} \circ \mathbb{V}_1[\vec{x}] = k \\ k \in \mathbb{N} \\ k > 1 \\ k \leq \max_{\vec{x} \in \mathcal{S}} \mathcal{G} \circ \mathbb{V}_1[\vec{x}]}} 1 &= \sum_{\substack{\vec{x} \in \mathcal{S} \\ k \leq \max_{\vec{x} \in \mathcal{S}} \mathcal{G} \circ \mathbb{V}_1[\vec{x}] \\ k \in \mathbb{N} \\ k > 1}} \frac{\mathcal{G} \circ \mathbb{V}_1[\vec{x}]}{k} \\
&\gg_n \sqrt{n} \sum_{\substack{\vec{x} \in \mathcal{S} \\ \vec{x} = (x_1, x_2, \ldots, x_n) \\ k \leq \max_{\vec{x} \in \mathcal{S}} \mathcal{G} \circ \mathbb{V}_1[\vec{x}] \\ k \in \mathbb{N} \\ k > 1}} \frac{\mathrm{Inf}(x_j)_{j=1}^n}{k} \\
&\geq \sqrt{n}\min_{\vec{x} \in \mathcal{S}} \mathrm{Inf}(x_j)_{j=1}^n |\mathcal{S}| \sum_{\substack{k \leq \max_{\vec{x} \in \mathcal{S}} \mathcal{G} \circ \mathbb{V}_1[\vec{x}] \\ k \in \mathbb{N} \\ k > 1}} \frac{1}{k}
\end{aligned}$$

thereby establishing the lower bound. □

Corollary 4.1. There are infinitely many collinear points with mutual integer distances on any line in $\mathbb{R}^n$ for all $n \geq 2$.

[1].

References

1. Anning, Norman H and Erdös, Paul and others *Integral distances*, Bulletin of the American Mathematical Society, vol. 51:8, American Mathematical Society, 1945, 598–600.
2. T. Agama *The Compression method and applications*, arXiv preprint arXiv:1912.08075, 2019.

Department of Mathematics, African Institute for Mathematical science, Ghana
E-mail address: theophilus@aims.edu.gh/emperordagama@yahoo.com

[1] .