Improved lower bound towards Chen-Chvátal conjecture

Congkai Huang[1]

[1]School of Mathematical Sciences, Peking University, Beijing, China

August 16, 2023

Abstract

We prove that in every metric space where no line contains all the points, there are at least $\Omega(n^{2/3})$ lines. This improves the previous $\Omega(\sqrt{n})$ lower bound on the number of lines in general metric space, and also improves the previous $\Omega(n^{4/7})$ lower bound on the number of lines in metric spaces generated by connected graphs.

1 Introduction

A classic theorem in plane geometry states that every noncollinear set of n points in the Euclidean space determine at least n lines. This is a special case of a combinatorial theorem of De Bruijn and Erdős [**?**] in 1948. In 2006, Chen and Chvátal suggested that the theorem might be generalized to arbitrary metric spaces. In a metric space (V, ρ), for every pair of distinct points $a, b \in V$, the *line* $\overline{ab}$ is defined to be

$$\begin{aligned}\overline{ab} = \{x : &\rho(x,b) = \rho(x,a) + \rho(a,b) \text{ or} \\ &\rho(a,b) = \rho(a,x) + \rho(x,b) \text{ or } \rho(a,x) = \rho(a,b) + \rho(b,x)\}.\end{aligned}$$

If there is a line containing all the points, i.e. $\overline{ab} = V$, then V is called a *universal line*. With this definition of the lines, Chen and Chvátal conjectured (see [**?**])

Conjecture 1. *In every finite metric space (V, ρ), either there is a universal line, or else there are at least $|V|$ distinct lines.*

It was proved in [**?**] that every finite metric space without a universal line contains $\Omega(|V|^{1/2})$ lines. In this article we improve the lower bound to $\Omega(|V|^{2/3})$:

Theorem 1. *In every finite metric space (V, ρ) without a universal line, there are at least $\Omega(|V|^{2/3})$ lines.*

Every connected graph $G = (V, E)$ generates a metric space (V, ρ) in the natural way — for each pair of vertices u and v, $\rho(u, v)$ is defined to be the length of the shortest path from u to v, i.e., the minimum number of edges one needs to travel from u to v. In [**?**] it was also proved that every finite metric space generated by a connected graph either contains a universal line, or else has $\Omega(|V|^{4/7})$ lines. Our work also improves the bound in this special case to $\Omega(|V|^{2/3})$.

There are special cases of Conjecture **??** where progresses are made in the past years. Kantor [**?**] proved that in the plane with L_1 metric a non-collinear set of n points induces at least $\lceil n/2 \rceil$ lines, improving an earlier lower bound of $n/37$ by Kantor and Patkós [**?**]. For metric spaces with a constant number of distinct distances, Aboulker, Chen, Huzhang, Kapadia, and Supko in [**?**] gave a $\Omega(n)$ lower bound, they also proved that every metric space on n points with distances in $\{0, 1, 2, 3\}$ has $\Omega(n^{4/3})$ distinct lines. Many other interesting results and stories related to the conjecture can be found in Chvátal's survey [**?**].

In Section **??** we give some notations used throughout the paper, and give a characterization of pairs generating the same line. In Section **??** we study the structure of the relations for the pairs generating the same line. When the number of lines is small, there must be a line with many different generating pairs. The key idea allows us to improve the lower bound is a careful study of the "interlocked" (we defined them as a green relation) generating pairs. In Section **??** we study the structure of each component connected by the green relations. This study allow us to find many lines if the green component is large. Finally in Section **??** we combine all the pieces together and prove Theorem **??**.

2 Notations and preliminaries

For distinct points a_0, a_1, ..., a_k,

$$[a_0a_1...a_k] \text{ means } \rho(a_0, a_k) = \sum_{i=0}^{k-1} \rho(a_i, a_{i+1})$$

With this notation, for a metric space, we call three distinct points a, b, c *collinear* if $[acb]$ or $[cab]$ or $[abc]$, and the line $\overline{ab}$ is the set of points consisting of a, b, and any c that is collinear with a and b.

Definition 1. *When $k \geq 1$ and $[a_0a_1...a_k]$ happens, we call $(a_0, a_1, \ldots, a_k)$ a* collinear sequence*, and call the set $\{a_0, a_1, \ldots, a_k\}$ a collinear set.*

For a collinear triple $\{a, b, c\}$, when $[acb]$ we say c is between *a and b, when $[cab]$ we say c is on the a-*side *of $\{a, b\}$, and when $[abc]$ we say c is on the b-*side *of $\{a, b\}$. In the latter two cases we say c is* outside *$\{a, b\}$.*

We also use the standard notations about sequences.

Definition 2. *For a sequence $\pi = (x_1, x_2, \ldots, x_k)$, its* reverse *is $\pi^R = (x_k, x_{k-1}, \ldots, x_1)$.*

For sequences $\pi = (x_1, x_2, \ldots, x_k)$ and $\sigma = (y_1, y_2, \ldots, y_s)$, their concatenation *is $\pi \circ \sigma = (x_1, x_2, \ldots, x_k, y_1, y_2, \ldots, y_s)$.*

The following facts are obvious. We list them and will use them frequently, sometimes implicitly.

Fact 1. *If (V, ρ) is a metric space and a, b, c, d, a_i for $i = 0, ..., k$ and b_j for $j = 1, ..., s$ are distinct points of V, then*

(a) $[abc] \Leftrightarrow [cba]$;

(b) $[abc]$ and $[acb]$ cannot both hold;

(c) $[abc]$ *and* $[acd]$ *implies* $[abcd]$*;*

(d) more generally, $[a_0 \dots a_k]$ *and* $[a_i b_1 \dots b_s a_{i+1}]$ *imply* $[a_0 \dots a_i b_1 \dots b_s a_{i+1} \dots a_k]$*;*

(d) $[a_0 a_1 ... a_k]$ *implies* $\rho(a_s, a_t) = \sum_{i=s}^{t-1} \rho(a_i, a_{i+1})$ *for every pair* s *and* t *such that* $s < t$*;*

(e) $[a_0 a_1 ... a_k]$ *implies* $[a_i a_j a_s]$ *for every pair* i *and* j *such that* $0 \leq i < j < s \leq k$.

The following is more general than Fact **??** (a) and (b).

Fact 2. *The elements of every collinear set can form exactly two collinear sequences and they reverse each other.*

Proof. redTOPROVE 0 □

Fact 3. *If* u, v, w, s *are four distinct points of* V *satisfying* $[usv]$*,* $[vsw]$*, and* $[wsu]$*, then* $\{u, v, w\}$ *is not collinear.*

Proof. redTOPROVE 1 □

Definition 3. *Let* (V, ρ) *be a metric space and let* L *be a line of* (V, ρ)*, define the set of its* generating pairs

$$K = K(L) := \left\{ \{a, b\} \in \binom{V}{2} : \overline{ab} = L \right\}. \tag{1}$$

For the rest of this work, we use ab to denote the binary set $\{a, b\}$.

First we discuss the possible relations between two pairs generating the same line L. [1] In order to formally do this, we introduce a little notation.

Definition 4. *for a pair of points* $e = \{a, b\}$*, we denote* $\rho(e) = \rho(a, b)$*.*

Fact 4. *For any* $e_1, e_2 \in K(L)$ *where* $\rho(e_1) \geq \rho(e_2)$*, exactly one of the following happens.*

(o) ordered relation*:*

(o.1) $e_1 = e_2$*;*

[1] Some of the classifications here are similar to those in Section 6 of [**?**].

(o.2) $|e_1 \cap e_2| = 1$*, they can be written as* $e_1 = \{a, b\}$*,* $e_2 = \{a, c\}$*, and* $[acb]$*;*

(o.3) $|e_1 \cap e_2| = 0$*, they can be written as* $e_1 = \{a, b\}$*,* $e_2 = \{c, d\}$*, and* $[acdb]$*.*

(b) blue relation*:*

(b.1) $|e_1 \cap e_2| = 1$*, they can be written as* $e_1 = \{a, b\}$*,* $e_2 = \{a, c\}$*, and* $[bac]$*;*

(b.2) $|e_1 \cap e_2| = 0$*, they can be written as* $e_1 = \{a, b\}$*,* $e_2 = \{c, d\}$*, and* $[abcd]$*.*

(g) green relation*:* $|e_1 \cap e_2| = 0$*, they can be written as* $e_1 = \{a, b\}$*,* $e_2 = \{c, d\}$*, and* $[acbd]$*.*

(r) red relation*:* $|e_1 \cap e_2| = 0$*, they can be written as* $e_1 = \{a, b\}$*,* $e_2 = \{c, d\}$*, and there are positive reals* x *and* y *such that* $\rho(a, b) = \rho(c, d) = x$*,* $\rho(a, c) = \rho(b, d) = y$*, and* $\rho(a, d) = \rho(b, c) = x + y$*.*

(p) purple relation*:* $|e_1 \cap e_2| = 0$*, they can be written as* $e_1 = \{a, b\}$*,* $e_2 = \{c, d\}$*, and there are positive reals* x *and* y *such that* $\rho(a, c) = \rho(b, d) = x$*,* $\rho(a, d) = \rho(b, c) = y$*, and* $\rho(a, b) = \rho(c, d) = x + y$*.*

These relations are depicted in Figure **??**. (It is helpful to see the corresponding cases in Figure **??** when reading the following proof; it is also helpful to use the figure when reading the rest of the paper.)

`generators.eps`

Figure 1: relations for $e_1, e_2 \in K(L)$.

Proof. redTOPROVE 2 □

Next, we define a graph on the pairs of points. Throughout the paper we use *points* for the elements of the metric space, and *vertices* for the vertices of the graph, so each vertex is a pair of points.

Definition 5. *Define a relation $\mathcal{P}_L = (K(L), \preccurlyeq)$ as $e_2 \preccurlyeq e_1$ if and only if e_1 and e_2 satisfy one of the ordered relations (o) in Fact* **??***.*

Define a coloured graph $\mathcal{G}_L$ on $K(L)$, e_1 and e_2 has an edge with colour blue (respectively, green, red, purple) whenever they have the blue (respectively, green, red, purple) relation as in Fact **??***, and we denote this by $e_1 \sim_b e_2$ (respectively, $e_1 \sim_g e_2$, $e_1 \sim_r e_2$, $e_1 \sim_p e_2$).*

By Fact **??**, it is easy to check that $\mathcal{P}_L$ is a partially ordered set (poset). Now we give a partition of $K(L)$.

Definition 6. *For $e \in K(L)$, define $\ell(e)$ be the length of the longest chain in the poset $\mathcal{P}_L$ with e as its maximum element.*

For a positive integer k, define

$$\mathcal{P}_L^{(k)} = \ell^{-1}(k) = \{e \in K(L) : \ell(e) = k\}.$$

and $\mathcal{G}_L^{(k)}$ be the (coloured) induced subgraph of $\mathcal{G}_L$ on $\mathcal{P}_L^{(k)}$.

Let $h(L)$ be the largest integer k for which $\mathcal{P}_L^{(k)} \neq \emptyset$. This is the height *of $\mathcal{P}_L$.*

It is a well known fact in partially ordered sets that

Fact 5. *For each positive integer k, $\mathcal{P}_L^{(k)}$ is an antichain.*

Proof. redTOPROVE 3 $\square$

Consequently, since Fact **??** tells us that any two elements in $K(L)$ either are in the ordered relation or they form one of the coloured edges, we have

Fact 6. *For every integer k with $1 \leq k \leq h(L)$, the graph $\mathcal{G}_L^{(k)}$ is a complete graph (with coloured edges).*

We also note that it is clear from Fact **??** and Fact **??** that

Fact 7. *For every $ab \in \mathcal{P}_L^{(k)}$, there are at least $k-1$ inner points collinear with a and b, i.e., there are distinct points $x_1, x_2, \ldots, x_{k-1}$ in $V \setminus \{a, b\}$ such that*

$$[ax_1x_2 \ldots x_{k-1}b].$$

3 The structure of $K(L)$

Definition 7. *We call an element $e \in K(L)$ purple if $e \sim_p f$ for some $f \in K(L)$, and denote $U(L)$ the set of all the purple elements; call an element $e \in K(L)$ red if $e \sim_r f$ for some $f \in K(L)$, and denote $D(L)$ the set of all the red elements.*

Fact 8. *For a red element $e = \{a, b\} \in D(L)$, no point (other than a and b) in V is between a and b; furthermore, e is a minimal element in the poset $\mathcal{P}_L$.*

Proof. redTOPROVE 4 □

Fact 9. *For a purple element $e = \{a, b\} \in U(L)$, every point in L (other than a and b) is between a and b, i.e., for every $v \in L$, we have $[avb]$.*

Proof. redTOPROVE 5 □

Fact 10. *When the set of purple elements $U(L) \neq \emptyset$, we have*

(a) $U(L)$ is the set of all maximal elements in the poset $\mathcal{P}_L$;

(b) For any $e \neq f \in U(L)$, $e \sim_p f$;

(c) For any $e \in U(L)$ and $f \in K(L) \setminus U(L)$, $f \preccurlyeq e$.

(d) $U(L) = \mathcal{P}_L^{(h(L))}$, the last level of the poset $\mathcal{P}_L$.

Proof. redTOPROVE 6 □

Now we turn to the study $\mathcal{P}_L^{(k)}$ for $2 \leq k \leq h(L) - 1$. By Facts **??** and **??**, there are no red nor purple elements in such levels; Fact **??** implies that $\mathcal{G}_L^{(k)}$ is a complete graph with blue and green edges.

Definition 8. *For a line L and an index k with $2 \leq k \leq h(L)-1$, $\mathcal{R}_L^{(k)}$ is the green sub-graph of $\mathcal{G}_L^{(k)}$, it has the vertex set $\mathcal{P}_L^{(k)}$ and has all the green edges of $\mathcal{G}_L^{(k)}$. Denote $Q_L^{(k)}$ the set of isolated vertices in $\mathcal{R}_L^{(k)}$, denote $c_L(k)$ the number of connected components of size at least 2 in $\mathcal{R}_L^{(k)}$, we call them the* green components in level k, *denote $P_L^{(k,i)}$ $(i = 1, 2, \ldots, c_L(k))$ the vertex set for each green component.*

For every subset $U \subseteq \mathcal{P}_L^{(k)}$, denote $V(U) \subseteq V$ the union of elements (each is a pair of points in V) of U, i.e., all the endpoints of generating pairs of L in U.

Fact 11. *For every $2 \leq k \leq h(L)-1$, $\left|Q_L^{(k)}\right| \leq |V|/(k-1)$.*

Proof. redTOPROVE 7 □

4 The structure of a green component

In this section, we fix a line L and an index k with $2 \leq k \leq h(L)-1$, denote $\mathcal{U} = \mathcal{P}_L^{(k)}$, and denote the green subgraph $\mathcal{R}_L^{(k)}$ by $\mathcal{R}$.

Definition 9. *For a subset $\mathcal{W} \subseteq \mathcal{U}$, we call a permutation π of the set of its endpoints $V(\mathcal{W})$ a* collinear ordering *for $\mathcal{W}$ if π is a collinear sequence. For each pair $ab \in \mathcal{W}$, where a comes before b in π, we call a the* opening point *of ab, and b the corresponding* closing point. *Since any two pairs $ab, cd \in \mathcal{W}$ has either $ab \sim_b cd$ or $ab \sim_g cd$, all the opening points are distinct; we call the sequence of opening points, sorted by their position in π from the earliest to the latest, the* opening sequence *of π.*

Let a be an opening point and b be its corresponding closing point, and v be a point in $L = \overline{ab}$. We say v is on the left side *of a in π if $[vab]$, otherwise (when $[avb]$ or $[abv]$) v is on the* right side *of a. We say v is on the* left side *of b in π if $[vab]$ or $[avb]$, otherwise (when $[abv]$) v is on the* right side *of b.*

We will prove the existence of a collinear ordering for every connected subgraph of $\mathcal{R}$. Before this, we first discuss some properties of such an ordering if one exists, as we will need them in the inductive proof.

Lemma **??** analyzes the opening and closing points on a collinear ordering, and the relation of a single point $v \in L$ to the opening-closing pairs. In turn, Lemma **??** gives the shape of the line L with respect to a collinear ordering. (See Figure **??**.)

greencpt.eps

Figure 2: The shaded area contains all points of L. In a collinear ordering, $a_1, a_2, \ldots, a_5$ are the opening points, $b_1, b_2, \ldots, b_5$ are corresponding closing points. $V = \overline{a_i b_i}$ for all $i = 1, 2, 3, 4, 5$.

Lemma 1. *Let H be a connected subgraph of $\mathcal{R}$ with order at least 2, π a collinear ordering for the vertex set of H, with $(a_1, a_2, \ldots, a_t)$ as its opening sequence; let b_i be the corresponding closing point to a_i $i = 1, 2, \ldots, t$. Then*

(a) The b_i's are distinct and their order in π, from left to the right, is $b_1, b_2, \ldots, b_t$;

(b) $a_i b_i \sim_g a_{i+1} b_{i+1}$ for every $1 \leq i < t$ and their order on π is $a_i, a_{i+1}, b_i, b_{i+1}$, and these four points are distinct;

(c) For every point $v \in L$, either v appears on π, or π has a partition $\pi = \sigma \circ \tau$ such that v is on the right side of all points in σ and on the left side of all points in τ. (Here σ and τ can be the empty sequence.)

Proof. redTOPROVE 8 $\square$

The following lemma is well-known in graph theory. [2]

Lemma 2. *Every connected graph G with order at least 2 has a vertex that is not a cut vertex, i.e., $G - v$ is still connected.*

[2] This is pointed to us by Vašek Chvátal — If the graph K_0 with no vertices is considered connected, then the unique vertex of K_1 is considered not to be a cut point, and the lower bound on the order of G may be dropped in the lemma. Arguments for and against declaring K_0 connected are presented on pages 42-43 of [**?**].

Proof. redTOPROVE 9 □

Lemma 3. *For every connected subgraph H of $\mathcal{R}$, a collinear ordering π for the vertex set of H, and every point $v \in L$ that does not appear on π, v can be inserted into π in a unique way to form a collinear sequence.*

Proof. redTOPROVE 10 □

Definition 10. *For every connected subgraph H of $\mathcal{R}$, a collinear ordering π for the vertex set of H, and every point $v \in L$ that does not appear on π, we say v is* outside H *if $(v) \circ \pi$ or $\pi \circ (v)$ is a collinear sequence, otherwise we say v is* inside H.

Note that the term *outside* is consistent with the terminology introduced in Definition **??**: if we view the pair $\{a, b\}$ as a one-vertex connected subgraph H for the line $\overline{ab}$, then v is outside this H if and only if it is outside $\{a, b\}$ according to Definition **??**.

Lemma 4. *The vertex set of every connected subgraph of $\mathcal{R}$ has a collinear ordering.*

Proof. redTOPROVE 11 □

Definition 11. *By Lemma **??** and Fact **??**, the vertex set of every green component $\mathcal{C}$ in level k has exactly two collinear orderings reverse each other. We call one of them the* standard collinear ordering *for $\mathcal{C}$.*

5 Proof of the main theorem

Our first lemma is a main ingredient in [**?**], a more rudimental form of this idea was first presented in [**?**], we prove it here for completeness.

Lemma 5. *If t distinct points $v_1, v_2, \ldots, v_t$ satisfy $[v_1 v_2 \ldots v_t]$ in a metric space without a universal line, then there are at least t distinct lines.*

Proof. redTOPROVE 12 □

Definition 12. *Let L be a line in a metric space (V, ρ) without a universal line, let k be index with $2 \leq k \leq h(L) - 1$, let $\mathcal{C}$ be a green component in level k of $\mathcal{P}_L$, and let $(a_1, a_2, \ldots, a_t)$ be the opening sequence of $\mathcal{C}$'s standard collinear ordering. For every i such that $1 \leq i < t$, $\overline{a_i a_{i+1}} \neq V$, so we can pick (any) $u_i \in V$ such that $\{u_i, a_i, a_{i+1}\}$ is not collinear. Set $L_i = \overline{u_i a_{i+1}}$ so $a_i \notin L_i$.*

We call u_i the i-th special point and L_i the i-th special line with respect to the green component $\mathcal{C}$.

Note that the u_i's may well be repeated for different index i. However, now we are going to show that the L_i's are distinct.

Lemma 6. *For a line L in a metric space (V, ρ) without a universal line, a level index k with $2 \leq k \leq h(L) - 1$, and any green component $\mathcal{C}$ in level k of $\mathcal{P}_L$, all the special lines with respect to $\mathcal{C}$ are distinct.*

Proof. redTOPROVE 13 □

Lemma 7. *For a line L in a metric space (V, ρ) without a universal line, a level index k with $2 \leq k \leq h(L) - 1$, any two special lines with respect to two different components in level k of $\mathcal{P}_L$ are distinct.*

Proof. redTOPROVE 14 □

Lemma 8. *For a line L in a metric space (V, ρ) without a universal line, a level index k with $2 \leq k \leq h(L) - 1$, and a green component $\mathcal{C}$ in level k, a special line with respect to $\mathcal{C}$ does not contain any point $v \in L$ but outside $\mathcal{C}$.*

Proof. redTOPROVE 15 □

Lemma 9. *For a line L a metric space (V, ρ) without a universal line, a level index k with $2 \leq k \leq h(L) - 1$, all the special lines in level k are distinct.*

Proof. redTOPROVE 16 □

Now we prove the main theorem.

Proof. redTOPROVE 17 □

6 Discussion – the structure of two green components

The following structure of two green components can be proved by an induction. After simplification we do not need it in the proof of the main theorem, and we omit the proof.

Fact 12. *For every index k such that $2 \leq k \leq h(L) - 1$, let π and σ be the standard ordering of two different green components in level k, there is a collinear sequence $\pi_1 \circ \sigma_1$, where π_1 is either π or π^R, and σ_1 is either σ or σ^R.*

Acknowledgement

The author heard this problem from the sessions in the Semicircle math club. We would like to thank Xiaomin Chen, Qiao Sun, Zhenyuan Sun, Chunji Wang and Qicheng Xu for the happy discussions on this work in the club. We thank Xiaomin Chen and Vašek Chvátal for their great help in writing this paper. We also thank two anonymous referees for their nice and helpful comments and suggestions.

References

[1] P. Aboulker, X. Chen, G. Huzhang, R. Kapadia and C. Supko, Lines, betweenness and metric spaces, *Discrete & Computational Geometry* **56** (2016), 427 – 448.

[2] X. Chen and V. Chvátal, Problems related to a de Bruijn - Erdős theorem, *Discrete Applied Mathematics* **156** (2008), 2101 – 2108.

[3] V. Chvátal, A de Bruijn-Erdős theorem in graphs?, in *Graph theory—favorite conjectures and open problems. 2.* Problem Books in Mathematics, Springer (2018), 149 – 176.

[4] N. G. De Bruijn and P. Erdős, On a combinatorial problem, *Indagationes Mathematicae* **10** (1948), 421 – 423.

[5] F. Harary and R. C. Reed, Is the null-graph a pointless concept?, *Graphs and Combinatorics: Proceedings of the Capital Conference on Graph Theory and Combinatorics at the George Washington University June 18–22, 1973* Springer Berlin Heidelberg (2006), 37 – 44.

[6] I. Kantor, Lines in the Plane with the L_1 Metric, *Discrete & Computational Geometry* **70** (2022), 960 – 974.

[7] I. Kantor and B. Patkós, Towards a de Bruijn-Erdős theorem in the L_1-metric, *Discrete & Computational Geometry* **49** (2013), 659 – 670.