

A note on maximally repeated sub-patterns of a point set

Véronique Cortier, Xavier Goaoc, Mira Lee, Na Hyeon-Suk

► **To cite this version:**

Véronique Cortier, Xavier Goaoc, Mira Lee, Na Hyeon-Suk. A note on maximally repeated sub-patterns of a point set. *Discrete Mathematics*, Elsevier, 2006, 306 (16), pp.1965-1968. <hal-00097239>

HAL Id: hal-00097239

<https://hal.archives-ouvertes.fr/hal-00097239>

Submitted on 21 Sep 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A note on maximally repeated sub-patterns of a point set

Véronique Cortier^a and Xavier Goaoc^{b 1} and Mira Lee^{c 1} and Hyeon-Suk Na^{d 2}

^aLORIA - CNRS, 615 rue du Jardin Botanique, B.P. 101, 54602 Villers-les-Nancy cedex, France. Email: cortier@loria.fr

^bLORIA - INRIA Lorraine, 615 rue du Jardin Botanique, B.P. 101, 54602 Villers-les-Nancy cedex, France. Email: goaoc@loria.fr

^cDivision of Computer Science, Korea Advanced Institute of Science and Technology (KAIST), 373-1, Guseong-dong, Yuseong-gu, Daejeon, 305-701, Republic of Korea. Email: mira@kaist.ac.kr

^dSchool of Computing, Soongsil University, 1-1, Sangdo-dong, Dongjak-gu, Seoul, 156-743, Republic of Korea. Email: hyeonsuk@gmail.com. Corresponding author, (phone) +82 2 828 7170 (fax) +82 2 822 3622

Abstract

We answer a question raised by P. Brass on the number of maximally repeated sub-patterns in a set of n points in \mathbb{R}^d . We show that this number, which was conjectured to be polynomial, is in fact $\Theta(2^{n/2})$ in the worst case, regardless of the dimension d .

Key words: Discrete geometry, point sets, repeated configurations.

1 Introduction

Let \mathcal{S} be a set of n points in \mathbb{R}^d . A *sub-pattern*, i.e. a subset, of \mathcal{S} is repeated if it can be translated to another subset of \mathcal{S} . A sub-pattern $P \subseteq \mathcal{S}$ is *maximally repeated* if for any subset Q such that $P \subsetneq Q \subseteq \mathcal{S}$ there exists a translation that maps P to a subset of \mathcal{S} without mapping Q to a subset of \mathcal{S} . In other words, a pattern is maximally repeated if it cannot be extended without losing

¹ This work was supported by the French-Korean STAR project 11844QJ.

² This work was supported by the Korea Research Foundation Grant funded by the Korean Government(R04-2004-000-10004-0).

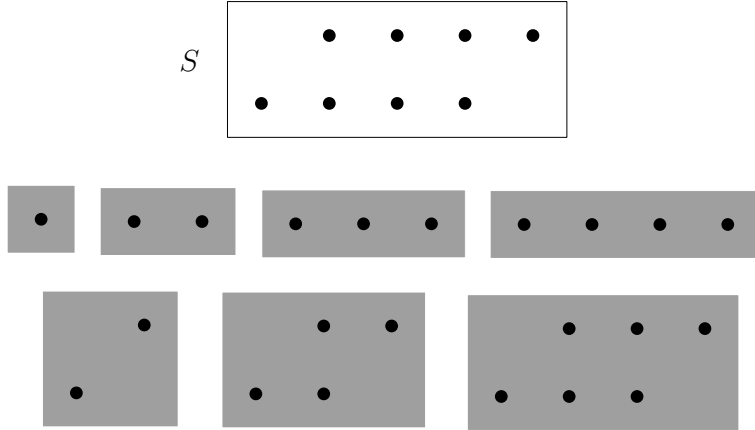


Fig. 1. A pattern S and all its maximally repeated sub-patterns.

at least one of its occurrences. Fig. 1 shows a pattern $S \subseteq \mathbb{R}^2$ and all its maximally repeated sub-patterns.

Maximally repeated sub-patterns (MRSP for short) originated from the field of pattern matching to solve the following problem: given two point sets X and Y , can Y be translated to a subset of X ? P. Brass [1, Theorem3] gave an algorithm that answers such queries in time $O(|Y| \log |X|)$ whose preprocessing time depends on the number of distinct MRSP of X , where two MRSP are *distinct* if they are not equal up to a translation. A natural question is thus to give a theoretical bound on this number of MRSP in order to provide an upper bound on the time requirement of that algorithm. This number was conjectured [1] [2, p.267] to be $O(n^d)$ where d is the dimension in which the point set is embedded.

In this note we prove that the number of MRSP of a set of n points in \mathbb{R}^d is actually $\Theta(2^{n/2})$ in the worst case and thus finding sub-patterns via this approach leads to exponential worst-case running time. More precisely, we show the following theorem:

Theorem 1 *A set of n points has at most $16 \cdot 2^{\lceil n/2 \rceil}$ distinct MRSP and for arbitrary large n there exist sets S of n points with $2^{\lfloor n/2 \rfloor - 1}$ distinct MRSP.*

Our proof is based on combinatorial rather than geometrical properties of the point set, which explains that the bound is independent of the dimension d in which the points are considered.

2 The Proof of Theorem 1

Let us first introduce some terminology. Given a set of points $P \subseteq \mathbb{R}^d$ and a translation $t \in \mathbb{R}^d$, $P + t := \{x + t \mid x \in P\}$ is the set of translated points of

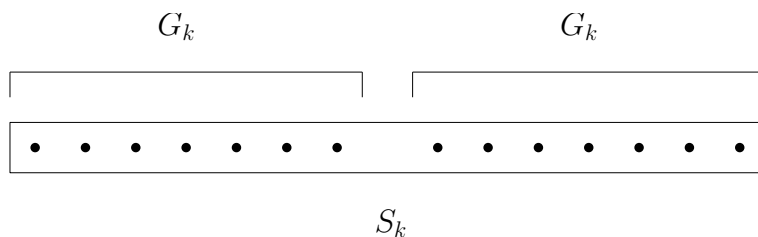


Fig. 2. A set S_k of $2k$ points with at least 2^{k-1} distinct MRSP.

P by t . A subset $P \subseteq \mathcal{S}$ is a repeated sub-pattern if there exists a translation $t \neq \mathbf{0}$ such that $P + t \subseteq \mathcal{S}$. P is a *maximally repeated sub-pattern* (MRSP) if, in addition, for any subset Q such that $P \subsetneq Q \subseteq \mathcal{S}$ there exists a translation t such that $P + t \subseteq \mathcal{S}$ and $Q + t \not\subseteq \mathcal{S}$. Two MRSP are *distinct* if they are not equal up to a translation.

In the sequel, we present a set of n points in \mathbb{R} having at least $2^{\lfloor n/2 \rfloor - 1}$ distinct MRSP (Section 2.1) and then prove that any set of n points in \mathbb{R}^d can have at most $16 \cdot 2^{\lfloor n/2 \rfloor}$ distinct MRSP (Section 2.2).

2.1 Lower bound

We build our example on a 1-dimensional grid which can, of course, be considered as embedded in \mathbb{R}^d for any $d \geq 1$. Let k be an integer, G_k denote the set of integers $\{1, \dots, k\}$ and \mathcal{S}_k be $G_k \cup (G_k + (k + 1))$, that is two copies of G_k separated by a gap of one point at $k + 1$ (see Figure 2).

Let P be a subset of the first copy of G_k , $Q \subseteq \mathcal{S}_k$ be a proper super-set of P and $p^* \in Q \setminus P$. If $p^* \geq k + 2$ then $P + (k + 1) \subseteq \mathcal{S}_k$ and $Q + (k + 1) \not\subseteq \mathcal{S}_k$. If $p^* \leq k$ then $P + (k + 1 - p^*) \subseteq \mathcal{S}_k$ and $Q + (k + 1 - p^*) \not\subseteq \mathcal{S}_k$. This proves that P is a MRSP. Two subsets of G_k containing 1 cannot be equal up to a non-trivial translation. Thus, all subsets of G_k containing 1 are distinct MRSP and \mathcal{S}_k admits at least 2^{k-1} MRSP. This proves the first statement of Theorem 1.

2.2 Upper bound

Recall that $(x_1, \dots, x_d) <_L (y_1, \dots, y_d)$ in the *lexicographic order* on vectors of \mathbb{R}^d if $x_1 < y_1$ or for some $r = 1, \dots, d - 1$:

$$x_1 = y_1, \dots, x_r = y_r \text{ and } x_{r+1} < y_{r+1}.$$

Let $\mathcal{S} = \{a_1, \dots, a_n\} \subseteq \mathbb{R}^d$ be a set of n points and $\mathcal{T} \subseteq \mathbb{R}^d$ the set of translations defined by $\mathcal{T} := \mathcal{S} - \mathcal{S} = \{x - y \mid (x, y) \in \mathcal{S}^2\}$.

Let \mathcal{A} denote the set of MRSP P such that no translation $t <_L \mathbf{0}$ satisfies $P + t \subseteq \mathcal{S}$. The set \mathcal{A} contains exactly one representative of each equivalence class of MRSP under translation, namely the one with the smallest point. To bound $|\mathcal{A}|$, we first partition this set in the following families:

$$\mathcal{A}_{ij} = \{P \in \mathcal{A} \mid \{a_i, a_j\} \subseteq P \subseteq \{a_i, \dots, a_j\}\}.$$

Informally, \mathcal{A}_{ij} is the set of MRSP spanning the range $\{a_i, \dots, a_j\}$. Since $\mathcal{A}_{11} = \{a_1\}$ and \mathcal{A}_{ii} is empty for $i \geq 2$, we have

$$|\mathcal{A}| = 1 + \sum_{1 \leq i < j \leq n} |\mathcal{A}_{ij}|. \quad (1)$$

There is an injection between the MRSP of \mathcal{A}_{ij} and the subsets of $\{a_{i+1}, \dots, a_{j-1}\}$. Hence,

$$|\mathcal{A}_{ij}| \leq 2^{j-i-1} \quad (2)$$

which will be enough to bound the number of MRSP spanning a “small” range. To bound the number of MRSP spanning a “wide” range, we describe them by the set of translations they allow. Let ϕ denote the function:

$$\phi : \begin{cases} 2^{\mathcal{S}} \rightarrow 2^{\mathcal{T}} \\ P \mapsto \{t \in \mathcal{T} \mid P + t \subseteq \mathcal{S}\} \end{cases}$$

If two elements of \mathcal{A} , P_1 and P_2 , have the same image by ϕ then:

$$\phi(P_1 \cup P_2) = \phi(P_1) = \phi(P_2).$$

By definition of MRSP, this implies that $P_1 \cup P_2 = P_1 = P_2$. Thus, ϕ is an injection from \mathcal{A} to the subsets of \mathcal{T} . For $1 \leq i < j \leq n$, let

$$\mathcal{T}_{ij} = \{t \in \mathcal{T} \mid t \geq_L \mathbf{0} \text{ and } \{a_i, a_j\} + t \subseteq \mathcal{S}\}$$

be the set of all non-negative translations compatible with a_i and a_j . MRSP in \mathcal{A}_{ij} only allow translations in \mathcal{T}_{ij} , so ϕ is an injection from \mathcal{A}_{ij} to the subsets of \mathcal{T}_{ij} and it follows that $|\mathcal{A}_{ij}| \leq 2^{|\mathcal{T}_{ij}|}$. Any $t \in \mathcal{T}_{ij} \setminus \{\mathbf{0}\}$ can be identified by the element $a_y = a_j + t$. Thus, the size of \mathcal{T}_{ij} is bounded by the number of such indexes y , which is at most $n - j$. Finally, we obtain that

$$|\mathcal{A}_{ij}| \leq 2^{n-j}. \quad (3)$$

Combining Equations (2), (3) and (1) we obtain:

$$|\mathcal{A}| \leq 1 + \sum_{1 \leq i < j \leq n} 2^{\min(n-j, j-i-1)}.$$

Splitting the sum at $j = \lceil \frac{n+i}{2} \rceil + 1$, we have

$$|\mathcal{A}| \leq 1 + 2 \sum_{i=1}^n \sum_{j=i+1}^{\lceil \frac{n+i}{2} \rceil + 1} 2^{j-i-1} \leq 1 + 2 \sum_{i=1}^n 2^{\lceil \frac{n-i}{2} \rceil + 1} \leq 1 + 8 \sum_{\ell=1}^{\lceil \frac{n}{2} \rceil} 2^{\ell}$$

and finally $|\mathcal{A}| \leq 16 \cdot 2^{\lceil n/2 \rceil}$, which proves the second statement of Theorem 1.

References

- [1] P. Brass. Combinatorial geometry problems in pattern recognition. *Discrete and Computational Geometry*, 28:495–510, 2002.
- [2] P. Brass, W. Moser, and J. Pach. *Research Problems in Discrete Geometry*. Springer-Verlag, 2005.