International Journal of Mathematics Trends and Technology
ISSN: 2231-5373/ https://doi.org/10.14445/22315373/IJMTT-V71I7P103

Volume 71 Issue 7, 28-32, July 2025
© 2025 Seventh Sense Research Group®

Original Article

A Boundary Property of Dataset Π^C

Yan Liu[1], Xingbo Wang[1,2*]

[1]*Guangzhou College of Applied Science and Technology, Guangzhou City, China.*
[2]*Foshan University, Foshan City, China.*

[2]*Corresponding Author : xbwang@fosu.edu.cn*

Received: 20 May 2025 Revised: 27 June 2025 Accepted: 12 July 2025 Published: 27 July 2025

Abstract *- The dataset ΠC has been shown to exhibit the characteristic of locally accumulating multiples of the divisors of a given composite integer, and it is employed to assist in identifying divisors of unfactorized integers. While several properties of this dataset have been described in the existing literature, its boundary characteristics remain unexplored. This paper demonstrates that the first and last rows of the dataset ΠC contain an identical number of multiples of the divisors of the composite integer under consideration, and this number corresponds to the maximum count across all rows. The identified property enhances the overall understanding of the dataset structure and may facilitate the design and optimization of practical search algorithms.*

Keywords *- Folding transformation, Geometry, Point symmetry, Randomized algorithm.*

1. Introduction

The construction of a high-quality dataset is undoubtedly beneficial for scientific computation. This viewpoint has long been supported by researchers in the field of scientific computing, as demonstrated by several survey papers [1]–[7]. Generally, dataset construction encompasses two essential components: the formal representation of the dataset and the characterization of its properties, both of which contribute to ensuring clarity and precision in its application. Two recent studies [8] and [9] have introduced datasets, referred to as Π^C, designed to identify divisors of odd composite integers. These datasets are structured in matrix form, consisting of rows and columns. By defining the term "host" as an integer that shares a common divisor with a given odd composite integer N, it has been shown that both Π and Π^C exhibit structural features conducive to the local accumulation of hosts. Consequently, these datasets can be searched using random search algorithms, such as those described in [10].

A search process achieves higher efficiency when the target locations are approximately known, for instance, being located near the center or along the boundaries. Paper [8] explicitly illustrates the distribution pattern of the hosts on the dataset; however, paper [9] does not provide a clear account of the corresponding distribution on the dataset Π^C. This lack of clarity may lead to inefficiencies in the search process, particularly due to the existence of large gaps between hosts, as theoretically proven in [11]. Therefore, this paper aims to present a more detailed analysis of the host distribution within the dataset Π^C. We demonstrate that the number of hosts located in the first and last rows is no less than that in the other rows. This property is further validated through computational experiments conducted using the Maple software.

The paper is organized as follows: Section 1 contextualizes the research landscape. Section 2 presents preliminaries. Section 3 gives the main results. Section 4 concludes with implications. Finally, an appendix section lists the Maple programs to test the results.

2. Preliminaries

Given a positive integer g and an odd composite integer $N = pq$ with divisors p and q satisfying $2 < p < q$, by letting $p_b = \lfloor\sqrt{N}\rfloor + 1$and $p_u = N - 1 - \lfloor\sqrt{N}\rfloor$, the dataset Π^C is an integer set whose elements are calculated by

$$\omega_{X,Y} = p_u + X - \left\lfloor\frac{Y-1}{g}\right\rfloor - 1 \tag{1}$$

Where integer variables X and Y are restricted by

$$\begin{cases} X \in [1, N - p_u] \\ Y \in [g(N - p_u) + 1, g(p_u - p_b - 1) + 1] \end{cases} \quad (2)$$

Denote $\xi = \lfloor \sqrt{N}/p \rfloor$, article [9], ^{C} has the following properties:

(i) Each column contains ξ or $\xi + 1$hosts p.
(ii) Once a host appears, it occurs consecutively for g times in the same row.
(iii) There must be a certain number of hosts in a row.
(iv) If $\omega_{X,Y} = h$it is a host of p or q, then $\omega_{X+1,Y+g} = h$ the hosts are distributed along a direction parallel to the line$y = gx$.
(v) The hosts are symmetric with respect to the center of Π^C.

Using a small circle 'o o o o ' to denote a host and a dot '.' to denote a non-host, Fig. 1 demonstrates the distribution of the hosts and non-hosts generated by N=21 and g = 3; Fig. 2 is the case for N=35 and g = 2.

o . . . . . . o o o . . . o o o o o o .

. o o o . . . . . . o o o . . . o o o o

o . . . o o o . . . . . . o o o . . . o

o o o o . . . o o o . . . . . . o o o .

. o o o o o o . . . o o o . . . . . . o

Fig. 1 Hosts and non-hosts in Π^C for N = 21 and g = 3

. . . . . o o o o . . . . o o o o . . . . . . o o . . . . o o .

o . . . . . . o o o o . . . . o o o o . . . . . . o o . . . . o

. o o . . . . . . o o o o . . . . o o o o . . . . . . o o . . .

. . . o o . . . . . . o o o o . . . . o o o o . . . . . . o o .

o . . . . o o . . . . . . o o o o . . . . o o o o . . . . . . o

. o o . . . . o o . . . . . . o o o o . . . . o o o o . . . . .

Fig. 2 Hosts and non-hosts in Π^C for N = 35 and g = 2

3. Main Results

This section proves Proposition 1 and demonstrates the results with Maple software.

Proposition 1. The number of hosts in the first row is equal to that in the last row, and is not less than that in the other rows.

Proof. Denote ^{C} to be a rectangle labeled by $ABCD$, as shown in Fig. 3. Let A be the origin and imagine two line segments AE and DF that are parallel to the line$y = gx$. Then $AEDF$ forms a parallelogram, and the two triangles, ACE and FDB, are congruent. Referring to [12], the hosts are distributed on ridges parallel to AE, indicating that the portion of each row covered by the parallelogram $AEDF$ contains the same number of hosts as the corresponding portion of the first row. Denote n to be this number. If the two triangles cover no hosts, n is valid for all the rows. Hence, the proposition holds in this case. In the case that there are hosts covered by the two triangles, each triangle covers the same number of hosts due to the central symmetrical property of the hosts. Take those hosts covered by the triangle ACE as an example. Because each ridge reaches the last row while no one reaches the first row, the number of hosts in the last row covered by the triangle ACE is not less than that in the other rows. Again, by the central symmetrical property, the number of hosts in the first row covered by the triangle FDB is not less than that in the other rows. Consequently, the proposition holds in all cases.

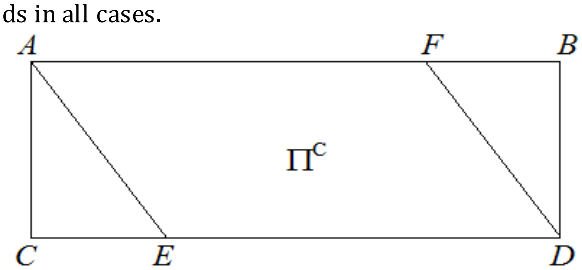

Fig. 3 Segments AE and DF are parallel to the line$y = gx$

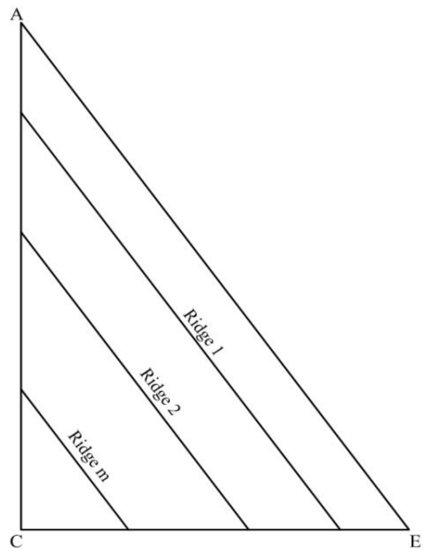

Fig. 4 Ridges in triangle *ACE*

With the Maple program, which is listed in the Appendix section, Proposition 1 is easily tested. Figures 5, 6, and 7 present several cases of the results from running the program. In these figures, the number at the end of a row indicates the number of the hosts in that row. It can be seen that the first row in each figure has the same number of hosts as the last row, and that number reaches the maximum one.

. . . . . . .oooooo. . . . . .oooooo. . . . . .ooo. ooo.18

o. . . . . .oooooo. . . . . .oooooo. . . . . .ooo. o17

.ooo. . . . . .oooooo. . . . . .oooooo. . . . . .ooo. . . .18

. . . .ooo. . . . . .oooooo. . . . . .oooooo. ooo.18

o. . . . . .ooo. . . . . .oooooo. . . . . .oooooo. o17

.ooo. . . . . .ooo. . . . . .oooooo. . . . . .oooooo. 18

Fig. 5 Number of hosts in the rows for *N* = 35 and *g* = 3.

. . . . .oooooooo. . . . . .oooooooo. . . . . .oooo. oooo.24

o. . . . . .oooooooo. . . . . .oooooooo. . . . . .oooo. o22

.oooo. . . . . .oooooooo. . . . . .oooooooo. . . . . .oooo. 24

. . . . .oooo. . . . . .oooooooo. . . . . .oooooooo. oooo.24

o. . . . . . . .oooo. . . . . .oooooooo. . . . . .oooooooo. o22

.oooo. . . . . . . .oooo. . . . . .oooooooo. . . . . .oooooooo. 24

Fig. 6 Number of hosts in the rows for *N* = 35 and *g* = 4.

. . . . .oooooooooo. . . . . .oooooooooo. . . . . .ooooo. ooooo.30

o. . . . . .oooooooooo. . . . . .oooooooooo. . . . . .ooooo. o27

.ooooo. . . . . .oooooooooo. . . . . .oooooooooo. . . . . .ooooo. 30

. . . . . .ooooo. . . . . .oooooooooo. . . . . .oooooooooo. ooooo.30

o. . . . . .ooooo. . . . . .oooooooooo. . . . . .oooooooooo. o27

.ooooo. . . . . .ooooo. . . . . .oooooooooo. . . . . .oooooooooo. 30

Fig. 7 Number of hosts in the rows for *N* = 35 and *g* = 5.

4. Conclusion

Through elementary geometry, this paper elucidates the boundary property of the dataset Π^C, showing that its first and last rows contain the maximum number of hosts. This property can help subsequent operations on Π^C. For example, searching the first row or the last row will have a higher chance of hitting a host. As a mathematical result, it might have other applications. This is our future exploration, and we hope to receive more attention.

Author Contributions

Author Yan Liu contributes programming and testing, and Author Xingbo Wang contributes theoretical modeling and mathematical reasoning.

Acknowledgment

The research work is supported by Characteristic Innovative Project from the Department of Education of Guangdong Province under the number. 2024KTSCX116 and Innovative Guiding Project from the Zhaoqing Bureau of Science and Technology under no. 241225090090875.

References

[1] Adriane Chapman et al., "Dataset Search: A Survey," *The VLDB Journal*, vol. 29, pp. 251-272, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[2] Youdi Gong et al., "A Survey on Dataset Quality in Machine Learning," *Information and Software Technology*, vol. 162, pp. 1-12, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[3] Minghao Liu et al., "Automatic Dataset Construction (ADC): Sample Collection, Data Curation, and Beyond," *arXiv:2408.11338v1*, pp. 1-20, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[4] Norman W. Paton, Jiaoyan Chen, and Zhenyu Wu, "Dataset Discovery and Exploration: A Survey," *ACM Computing Surveys*, vol. 56, no. 4, pp. 1-37, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[5] Xinyan Velocity Yu et al., "Beyond Counting Datasets: A Survey of Multilingual Dataset Construction and Necessary Resources," *arXiv:2211.15649v1*, pp. 1-19, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[6] Yuzhen Lu, and Sierra Young, "A Survey of Public Datasets for Computer Vision Tasks in Precision Agriculture," *Computers and Electronics in Agriculture*, vol. 178, pp. 1-13, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[7] Amandalynne Paullada et al., "Data and its (Dis)contents: A Survey of Dataset Development and Use in Machine Learning Research," *arXiv:2012.05345v1*, pp. 1-14, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[8] Hao Li, Zhiying Liu, and Xingbo Wang, "A Dataset and a Dynamic Distributed Parallel Network for Identifying Divisors of Odd Composite Integers," *Journal of Advances in Mathematics and Computer Science*, vol. 40, no. 4, pp. 104-129, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[9] Zhiying Liu, Hao Li, and Xingbo Wang, "A Novel Dataset and Algorithm for Identifying Divisors of an Odd Composite Integer Using Distributed Parallel Network," *Journal of Advances in Mathematics and Computer Science*, vol. 40, no. 5, pp. 105-120, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[10] Xingbo Wang, "Leveraging Lévy Flight for Efficient Divisor Identification in Odd Composite Integers," *Journal of Advances in Mathematics and Computer Science*, vol. 39, no. 11, pp. 116-139, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[11] Xingbo Wang, "Gaps between Integers Having a Common Divisor with an Odd Semiprime," *Asian Research Journal of Mathematics*, vol. 21, no. 2, pp. 67-88, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[12] Xingbo Wang, "Distribution of Divisors of an Integer in a Triangle Integer Sequence," *JP Journal of Algebra, Number Theory and Applications*, vol. 63, no. 2, pp. 185-208, 2024. [CrossRef] [Google Scholar] [Publisher Link]

Appendix Maple Programs

Here lists the Maple programs for testing Proposition 1.

============ShowHosts==============

```
ShowHosts := proc(N, g)
   local X, Y, W, pu, pb, dd, Ly, Ry, cd, ct;
   ct := 0;
   pb := floor(sqrt(N));
   pu := N - 1 - pb;

   Ly := g*(N - pu);
   Ry := g*(pu - pb - 1) + 1;
```

```
    for X from N - pu do
      ct := 0;
      for Y from Ly to Ry do
          dd := floor((Y - 1)/g);
          W := pu + X - 1 - dd;
          cd := gcd(N, W);
          if  cd >1 then printf("o"); ct++;
          else printf("."); end if;
        end do;
        printf(" %2d \n", ct);
    end do;
  end proc
===============End================
```