Rose-Hulman Undergraduate Mathematics Journal

Volume 24
Issue 2

Article 6

k-Distinct Lattice Paths

Eric J. Yager
Valparaiso University, eric.yager@valpo.edu

Marcus Engstrom
Valparaiso University, 2018marcus2022@gmail.com

k-Distinct Lattice Paths

Cover Page Footnote

Special thanks to Rick Gillman for sponsoring and guiding our research, and to Nick Rosasco for providing computing resources.

ℝose-Hulman Undergraduate Mathematics Journal
Volume 24, Issue 2, 2023

k-Distinct Lattice Paths

By *Marcus Engstrom* and *Eric Yager*

Abstract. Lattice paths can be used to model scheduling and routing problems, and, therefore, identifying maximum sets of k-distinct paths contributes to optimizing solutions to these problems. We extend the work previously done by Gillman et. al. to determine the order of a maximum set of k-distinct lattice paths. In particular, we disprove a conjecture by Gillman that a greedy algorithm gives this maximum order and also refine an upper bound given by Brewer et. al. We illustrate that brute force is an inefficient method to determine the maximum order, as it has time complexity $\mathrm{O}(n^k)$.

1 Introduction

Imagine driving home from work. You're tired of taking the same route home every day, so you want to see some new scenery (without taking extra time to drive around). How many different ways are there to drive home, so you see at most four of the same landmarks each day? How about five? Can you go a whole week where you never see more than the same three landmarks on the way home? We can model a simplified version of this situation by examining paths on an $m \times n$ lattice.

Imagine that each edge has a unique landmark, and we wish to count paths that share at most some number, $k-1$, of edges. Generally, on an $m \times n$ lattice, there are a total of $m+n$ edges in each path, and two paths are k-distinct if they share fewer than k edges. Otherwise, they are k-equivalent. A set of k-distinct paths, $\mathbf{P}$, contains paths $p_1, p_2, \ldots, p_{\mathrm{P}}$ for a total of P paths. These paths may be ordered alphabetically. We want to determine the maximum number, $\mathrm{P}(m,n,k)$, of k-distinct paths on the $m \times n$ lattice. We know that the set of all possible paths, $\mathbf{C}$, has order $\mathrm{C} = \binom{m+n}{n}$ [2].

On the 4×2 lattice shown in Figure 1, we can move along the edges to travel from the lower left corner to the upper right corner as shown. Then the blue line is the path EENNEE, the green path is NEENEE, and the red path is NEEEEN. Now suppose that $k = 3$. The blue path is 3-equivalent to the green path, which is 3-equivalent to the red path, but the blue and red paths are not 3-equivalent to each other.

It is this lack of transitivity that makes the problem interesting.

Mathematics Subject Classification. 05E15
Keywords. counting paths, efficient algorithms

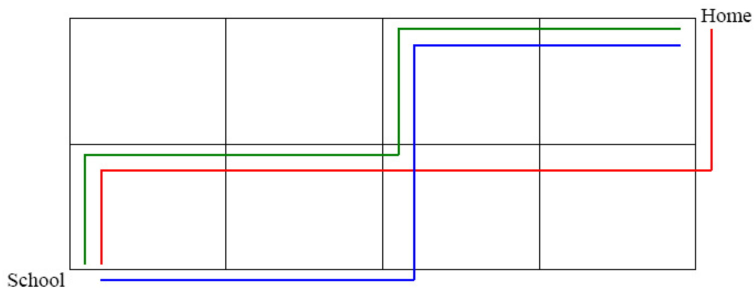

Figure 1: Lattice path from school to home. There are three different possible paths shown.

Earlier work by Gillman [5] obtained the following results for some extreme values of m, n, and k.

Theorem 1.1. *For the appropriate values of m, n, and k, the following are true.*

1. *For all $m, n > 0$, $\mathrm{P}(m,n,1) = 2$.*
2. *For all $m, n \geq 3$ or $m = n = 2$, $\mathrm{P}(m,n,2) = 4$.*
3. *If $m \geq 3$ and $n = 2$ or $m = 3$ and $n = 1$, then $\mathrm{P}(m,n,2) = 3$.*
4. *For all m, n, $\mathrm{P}(m,n,m+n) = \mathrm{C}$.*
5. *For all m, n, $\mathrm{P}(m,n,m+n-1) = \mathrm{C}$.*
6. *If $m \geq 3$, then for any $k = 2, 3, \ldots, m$, $\mathrm{P}(m,1,k) = 1 + \lfloor m/(m-k) \rfloor$.*

Gillman conjectured that the greedy algorithm finds the maximum set of k-distinct paths. We show this conjecture false in Section 2. He also obtained the following result which we will improve on in Section 3 by relaxing the conditions of the hypothesis.

Theorem 1.2. *If $n \leq k \leq \frac{m-1}{n+1}$, then $\mathrm{P}(m,n,k) = n+1$.*

As an aside, finding a maximal set of k-distinct paths is equivalent to finding the independence number on a particular family of graphs. These graphs have the set of all paths in the $m \times n$ lattice as their vertices, with edges connecting k-equivalent paths. They were explored in some detail by Brewer, et. al. [1] and Gillman [6] .

2 Computational Results

Given an $m \times n$ lattice, the greedy algorithm generates a set, $\mathbf{G}$, of k-distinct paths through the following steps:

1. Generate an alphabetical list of all possible paths.

2. Add the first path to **G**.

3. While path is not the last path,

   - Select the next alphabetical path.

   - If the path is k-distinct from all paths in **G**, add the path to **G**.

4. After checking all of the paths, return **G** and its order, G.

The greedy algorithm executes in $O(n^2)$ time[3]. The greedy algorithm first constructs a list of all paths on the given lattice in alphabetical order. Then, it iterates through the list and greedily adds paths to a set of k-distinct paths. Thus, the first path is not compared to any other paths, the second path is compared to at most one path, the third path is compared to at most two paths, and so on. Because there are $\sum_{i=0}^{n} i \frac{n(n+1)}{2}$ comparisons, the time complexity is $O(n^2)$.

To test the results from the greedy algorithm, we constructed all possible path sets of order $G+1$ and determined whether each was a set of k-distinct paths. If none were, the the greedy algorithm had yielded a maximal result. If one or more were sets of k-distinct paths, we repeated this process for path sets of order $G+2$, and so forth, until no k-distinct path sets were found. The code we used to run these calculations is available on Github.com [7].

In contrast to the greedy algorithm, this algorithm is very computationally complex. Generating all combinations of a set of a given size requires $O(n^k)$ time according to Canonne [4].

Using this algorithm, we constructed k-distinct sets for small values of m, n, and k. For $m = 4$, $n = 3$, and $k = 3$, the greedy algorithm produces the set

`EEENNN, EENENNE, ENENEEN, ENNENEE, NEEEENN, NNEEENE`

and the brute force algorithm shows that this is one of three sets of k-distinct paths of order six. When tested against sets of order seven, we found the following maximal sets of 3-distinct paths. (Each row is a different set of seven, 3-distinct, paths.)

```
EEEENNN, EENENNE, ENENNEE, ENNEEEN, NEEEENN, NENNEEE, NNEEENE
EEEENNN, EENENNE, ENENNEE, ENNEEEN, NEEENEN, NENNEEE, NNEEENE
EEEENNN, EENNNEE, ENEENNE, ENNEEEN, NEEEENN, NENNEEE, NNEEENE
EEENENN, EENENNE, ENENNEE, ENNEEEN, NEEENEN, NENNEEE, NNEEENE
EEENNEN, EENNENE, ENEEENN, ENNENEE, NEEENNE, NENEEEN, NNENEEE
EEENNEN, EENNENE, ENEEENN, ENNENEE, NEEENNE, NENEEEN, NNNEEEE
EEENNEN, EENNENE, ENEEENN, ENNENEE, NEEENNE, NNEEEEN, NNNEEEE
```

```
EEENNEN, EENNENE, ENEEENN, ENNNEEE, NEEENNE, NENEEEN, NNEENEE
EEENNEN, EENNNEE, ENEEENN, ENNEENE, NEEENNE, NNEEEEN, NNNEEEE
EEENNNE, EENEENN, ENENNEE, ENNEEEN, NEEENEN, NENNEEE, NNEEENE
```

There were no sets of 3-distinct paths of order eight, so we not only found that the greedy algorithm does not always produce a maximal set of k-distinct paths, but also found the maximum order in this case. However, the following table, constructed using the greedy algorithm, suggests that it may not be wrong often, or by much; the brute force algorithm was used to completed testing through the $m = 5$ row before it began using months of computing time.

$m\backslash k$	1	2	3	4	5	6	7	8	9	10	11	12	13
3	2	4	8	10	20								
4	2	4	~~6~~ 7	13	19	35							
5	2	4	6	~~9~~ 10	20	28	56						
6	2	4	5	8	13	30	44	84					
7	2	4	5	7	11	18	42	60	120				
8	2	4	5	6	10	16	25	57	85	165			
9	2	4	4	5	8	12	20	33	76	110	220		
10	2	4	4	4	6	10	16	24	41	98	146	286	
11	2	4	4	4	6	8	13	18	31	50	124	182	364

Table 1: Table of $\mathrm{P}(m, 3, k)$ values generated by the Greedy Algorithm. Terms with a slash through them are terms where the greedy algorithm does not give an optimal value, as found by the brute force algorithm.

3 Theoretical Results

We begin this section with a theorem which establishes conditions for when $n+1$ is a lower bound on $\mathrm{P}(m, n, k)$ and then establish when this bound is obtained, improving on Theorem 1.2 above.

Theorem 3.1. *If $k \geq n$, then $\mathrm{P}(m, n, k) \geq n+1$.*

Proof. Assuming without loss of generality that $m \geq n$, if $m = 1$, then $n = 1$ and we have two paths, so $2 \geq 1+1$ and the conclusion holds. Consider an $m \times n$ lattice, where $m \geq 1$. Let $\mathbf{P}$ be the set of paths where $p_l = \mathrm{N}^l \mathrm{E}^m \mathrm{N}^{n-l}$ for $0 \leq l \leq n$, so that the order of $\mathbf{P}$ is $\mathrm{P} = n+1$. Now consider two distinct elements $p_i, p_j \in \mathrm{P}$ such that $p_i = \mathrm{N}^{l_i} \mathrm{E}^m \mathrm{N}^{n-l_i}$ and $p_j = \mathrm{N}^{l_j} \mathrm{E}^m \mathrm{N}^{n-l_j}$, and $l_i < l_j$. Then p_i and p_j share $l_i + (n - l_j) = n + (l_i - l_j) < n$ edges, since their east edges are distinct. We then have, for $k \geq n$, all of the paths in $\mathbf{P}$ distinct and as a consequence $\mathrm{P}(m, n, n) \geq n+1$ for all $m \geq 1$. $\square$

We now study when $n+1$ is an upper bound on $\mathrm{P}(m,n,k)$ and show that this occurs for large m. This result depends on the following lemma about a very specific type of sequence.

Given an integer $n \geq 0$, let $\mathbf{S} = \{s_1, s_2, \ldots, s_{\mathrm{S}}\}$ be a set of non-decreasing integer sequences of length x on the interval $[0,n]$, in which any two sequences share at most $k-1$ elements. That is, for any two sequences s_i and s_j, $s_{iz} = s_{jz}$ at most $k-1$ times.

Example Let $n=1, x=5, k=3$. We find the set

$$\mathbf{S} = \{s_1, s_2\} = \{\{0,0,0,1,1\},\{1,1,1,1,1\}\},$$

since s_1, s_2 are non-decreasing integer sequences of length five on the interval [0, 1] in which the two sequences share two elements (elements 4 and 5 of both sequences are 1).

Lemma 3.2. *Let the sequences in* $\mathbf{S}$ *have length* $x = n(k-1)+k+c$ *for integers* $c \geq 0$ *and* $k \geq 1$. *Then the order of* $\mathbf{S}$ *is* $\mathrm{S} \leq n+1$.

Proof. Assume the sequences in $\mathbf{S}$ have length $n(k-1)+k+c$ for some fixed integers $c \geq 0, k \geq 1$. We proceed by induction on $n \geq 0$.

If $n=0$, we have sequences of length $k+c$ with a single element 0. There is only one such sequence, and $1 \leq n+1$.

Now let the sequences in $\mathbf{S_N}$ have length $x_{\mathrm{N}} = \mathrm{N}(k-1)+k+c$ on the interval $[0,\mathrm{N}]$ and assume that $\mathrm{S_N} \leq \mathrm{N}+1$. Now consider the sequences in $\mathbf{S_{N+1}}$ with length $x_{\mathrm{N}+1} = (\mathrm{N}+1)(k-1)+k+c$ on the interval $[0,\mathrm{N}+1]$. We will show that $\mathrm{S_{N+1}} \leq (\mathrm{N}+1)+1 = \mathrm{N}+2$.

Without loss of generality, we may assume sequences in $\mathbf{S_{N+1}}$ are sorted alphabetically (or more precisely, lexicographically) and label the first sequence s_1. Since $\mathrm{N}+1 \geq 1$ and $k \geq 1$, a maximal set must have at least two sequences, as $\{0,\ldots,0\}$ and $\{1,\ldots,1\}$ share no elements. Thus we consider sets of order $\mathrm{S_{N+1}} \geq 2$.

Assume that the k-th element of the second sequence of the set, s_2, is 0. Since the sequences are non-decreasing, the first k elements of s_2 must also be 0. Since the sequences are ordered alphabetically, the first k elements of s_1 must also be 0, leading to a contradiction of the fact that the two sequences only share $k-1$ elements. Hence the k-th element of s_2 is at least 1. It follows from the alphabetical ordering that the k-th element of each of $s_2, \ldots, s_{\mathrm{S}}$ is least 1.

Remove the first $k-1$ elements of the sequences $s_2, \ldots, s_{\mathrm{S}}$, and call this new set of truncated sequences $\mathbf{S_A}$. Then $\mathrm{S_A} = \mathrm{S} - \mathrm{N} + 1 - 1$. Each sequence in this set has length

$$[(\mathrm{N}+1)(k-1)+k+c]-(k-1) = \mathrm{N}(k-1)+k+c = x_{\mathrm{N}}$$

elements. Further, since each sequence is comprised of integers on the interval $[1,\mathrm{N}+1]$, we can subtract 1 from every element. This gives an equivalent set of sequences, comprised of integers on the interval $[0,\mathrm{N}]$. Let us consider a maximal set of sequences of this form, $\mathbf{S'_A}$. By the inductive assumption, $\mathrm{S'_A} \leq \mathrm{N}+1$, so $\mathrm{S_A} \leq \mathrm{S'_A} \leq \mathrm{N}+1$.

If two sequences share (at least) k elements in $\mathbf{S_A}$, then they must share (at least) k elements in $\mathbf{S_{N+1}}$, since the sequences in $\mathbf{S_A}$ are constructed as the latter part of sequences in $\mathbf{S_{N+1}}$. Then $\mathrm{S_A} \le \mathrm{S}'_{\mathrm{A}} \le \mathrm{N}+1$ implies $\mathrm{S_{N+1}} - 1 \le \mathrm{N}+1$, so $\mathrm{S_{N+1}} \le \mathrm{N}+2$. □

Example. Continuing the previous example, let $n = 1, k = 3, c = 0$, so that $x = 1(3-1) + 3 + 0 = 5$. Then **S** is a set of non-decreasing integer sequences of length 5 on the interval [0, 1]. By Lemma 1, there are at most $n+1 = 2$ sequences in **S**. We can manually check all possible sets of three sequences from among the six sequences available (00000, 00001, 00011, 00111, 01111, 11111) and verify there is no valid set of three sequences. We can verify that $\{\{0,0,0,0,0\},\{1,1,1,1,1\}\}$ is a valid set, so the maximal order of **S** is exactly 2.

Theorem 3.3. *If* $m \ge n(k-1) + k$, *then* $\mathrm{P}(m,n,k) \le n+1$.

Proof. If $k = 1$, then $\mathrm{P}(m,n,k) = 2 \le n+1$. Now we can assume $k \ge 2$. If two paths share at least k east moves, then they are k-equivalent. So consider the maximal set of paths that share fewer than k east edges, $\mathbf{P_H}$. Then any set of at least $\mathrm{P_H}+1$ paths will have two paths that share at least k east moves and thus share at least k total moves. Thus $\mathrm{P}(m,n,k) \le \mathrm{P_H}$. With this in mind, let us label the $m \times n$ lattice from $(0,0)$ to (m,n). Let $m = n(k-1) + c$ for some non-negative integer c. Then we have a total of $n(k-1)+c$ east moves, and north coordinates $0,1,2,\dots n$. Let **L** be a set of sequences of $n(k-1)+k+c$ non-decreasing integers from 0 to $n = n+0$, where the integers are the north coordinates and the elements are the east moves (from left to right). Then the maximal number of sequences where no two sequences share k or more elements, L, is $\mathrm{L} = \mathrm{P_H}$. By Lemma 3.2, $\mathrm{P_H} \le n+1$.

Thus $\mathrm{P}(m,n,k) \le \mathrm{P_H} \le n+1$ whenever $m \ge n(k-1)+k$. □

Example. Consider $\mathrm{P}(5,1,3)$. This is a continuation of the previous example. Since $m \ge n(k-1)+k$, by Theorem 4 $\mathrm{P}(5,1,3) \le 2$. In fact we have paths NEEEEE and EEEEEN which share 0 edges, so $\mathrm{P}(m,n,k) = n+1$ in this example.

Now we have sufficient conditions to determine when $\mathrm{P}(m,n,k) = n+1$.

Corollary 3.4. *If* $k \ge n$ *and* $m \ge n(k-1)+k$, *then* $\mathrm{P}(m,n,k) = n+1$.

Proof. Follows directly from Theorem 3.1 and Theorem 3.3. □

This result is an improvement on Theorem 1.2. Rewriting the second condition for Theorem 1.2, we have

$$m \ge k(n+1) + 1 = kn + k + 1 + n - n = n(k-1) + k + (n+1),$$

while Corollary 3.4 only requires

$$m \ge n(k-1) + k,$$

which reduces this bound on m by $n+1$.

If we let $k = n$ in Theorem 3.3, we have $m = n^2$ The following theorem begins to describe what happens for $m < n^2$.

Theorem 3.5. *If* $1 < m < n^2$, *then* $\mathrm{P}(m,n,n) > n+1$.

Proof. Let $\mathbf{P} = \{\mathrm{N}^l\mathrm{E}^m\mathrm{N}^{n-l} \mid 0 \le l \le n\}$. From the proof of Theorem 3.1, we know $\mathbf{P}$ is a set of $n+1$ k-distinct paths. The north moves are either before any east moves or after all east moves. This also means all east moves are at the same vertical position. Take any path $p_i = \mathrm{N}^i\mathrm{E}^m\mathrm{N}^{n-i} \in \mathbf{P}$ and consider the following cases.

Case 1: Assume $m \le n$ and consider the path $p_{n+2} = \mathrm{E}^{m-1}\mathrm{N}^n\mathrm{E}$. Clearly p_{n+2} and p_i share no north edges and at most $m-1 \le n-1 < n$ east edges. Thus p_{n+2} is k-distinct from all paths $p_i \in \mathbf{P}$ and the conclusion holds.

Case 2: Now assume $m > n$ and $\frac{m}{n-1} \in \mathbb{Z}$. Consider the path

$$p_{n+2} = (\mathrm{E}^{n-1}\mathrm{N})^{\frac{m}{n-1}-1}\mathrm{N}^{n-(\frac{m}{n-1}-1)}\mathrm{E}^{n-1}.$$

We can verify this is a valid path since it has m east moves and n north moves, and $\frac{m}{n-1} - 1 \le \frac{n^2-1}{n-1} - 1 = \frac{(n+1)(n-1)}{n-1} - 1 = n+1-1 = n \le n$ as $m < n^2 \Rightarrow m \le n^2 - 1$. Since $m > n$ we must have that $\frac{m}{n-1} > 1$. Since $\frac{m}{n-1} \in \mathbb{Z}$, we have that $\frac{m}{n-1} - 1 \ge 2 - 1 = 1$, so we start with an east move. Since the last move is an east move, p_{n+2} and p_i share no north edges. Since p_{n+2} and p_i share at most $n-1 < n$ east edges, p_{n+2} is k-distinct from all paths in $\mathbf{P}$ and the conclusion holds.

Case 3 Now we can assume $m > n$ and $\frac{m}{n-1} \notin \mathbb{Z}$. Consider the path

$$p_{n+2} = (\mathrm{E}^{n-1}\mathrm{N})^{\lfloor\frac{m}{n-1}\rfloor}\mathrm{N}^{n-\lfloor\frac{m}{n-1}\rfloor}\mathrm{E}^{m-(n-1)\lfloor\frac{m}{n-1}\rfloor}.$$

We can verify that p_{n+2} has m east moves and n north moves. Since $m > n$, p_{n+2} starts with an east move. Since $\frac{m}{n-1} \notin \mathbb{Z}$, $\lfloor\frac{m}{n-1}\rfloor < m$. This means p_{n+2} ends with an east move and thus shares no north moves with p_i. Since $m - (n-1)\lfloor\frac{m}{n-1}\rfloor < m - (n-1)(\frac{m}{n-1} + 1) = m - (m - (n-1)) = n-1 < n$, we know p_{n+2} and p_i share at most $n-1 < n$ east edges. Thus p_{n+2} is k-distinct from all paths in $\mathbf{P}$ and the conclusion holds.

This exhausts all cases. Thus $\mathrm{P}(m,n,n) > n+1$. □

4 Open Questions

The brute force algorithm described in Section 2 may be improved somewhat by more efficient programming, but it will always require substantial computational time. It may be that an alternate approach, such as some form of an evolutionary algorithm, may be able to find solutions quicker than using direct computation.

While it is obvious that $\mathrm{P}(m,n,k)$ is a non-decreasing function in k, it may also be true that $\mathrm{P}(m,n,k)$ is a non-increasing function in m, at least for non-extreme values of m.

Finally, it is possible to find more particular results such as the following conjecture: If $m = n(k-1) + k - 1$ and $k \geq n$, then $\mathrm{P}(m, n, k) > n + 1$.

References

[1] Brewer, M., et al. "Graphs of Essentially Equivalent Lattice Paths." *Geombinatorics*, vol. 13, no. 1, 2003, pp. 5-9.

[2] Bona, M. (editor) Handbook of Enumerative Combinatorics CRC Press, Boca Raton, 2015.

[3] Brown, R. Advanced Mathematics: Precalculus with Discrete Mathematics and Data Analysis. Houghton-Mifflin Co., Boston, 1994.

[4] Canonne, C. "Approximation of combination $\binom{n}{k} = \Theta\left(n^k\right)$?" *Stack Exchange*, May 3, 2015, math.stackexchange.com/questions/1265519/approximation-of-combination-n-choose-k-theta-left-nk-right/4134185

[5] Gillman, Rick. "Enumerating and Constructing Essentially Equivalent Lattice Paths." *Geombinatorics*, vol. 11, no. 2, 2001, pp. 37-42.

[6] Gillman, Rick, et al. "On the Edge Set of Graphs and Lattice Paths." *International Journal of Mathematics and Mathematical Sciences*, vol. 61, no. 1, 2004, pp. 3291-3299.

[7] Yager, E. Lattice Path Research Code [Computer software]. github.com/ejyager00/lattice_paths.

Marcus Engstrom
Valparaiso University
`2018marcus2022@gmail.com`

Eric Yager
Valparaiso University
`ejyager00@protonmail.com`