Untitled Number Problem

Vikrum Nijjar (`vikrum@slique.net`)

September 30, 2003

Abstract

I recently took UC Berkeley's introductory Computer Science course. It is taught using a dialect of LISP called Scheme from the book Structure and Interpretation of Computer Programs (Abelson and Sussman). One of the problems in the book refers to a certain algorithm[1] which generates the Cartesian product of $\mathbb{Z}^+$ and $\mathbb{Z}^+$ (i.e. in LISP, an infinite stream of pairs using the stream of positive integers as input) in a particular order; it asks for how the pairs are placed, in general, and the positioning of a few pairs, in particular. The solution I found behaves strangely under a specific condition and my attempts to learn more about the function haven't been very fruitful. I'd appreciate any input on where I could find more information about: the type of function itself and others in its class; how to analyze the function (particularly the "strange behavior" it exhibits); and how to effectively visualize the input data.

1 Background

The problem statement asks for "any general comments about the order in which the pairs are placed into the stream" and about the positioning of (1 100), (99 100), and (100 100) – which turn out to be computationally unfeasible to find by means of brute force. The first 250 pairs are found in Table 1.

2 Observations

From this we can make some immediate observations about the properties of the stream and the rules, dictated by the algorithm, which govern the placement of pairs. (We let p be the position of pair $(m\ n)$, some function π such $\pi(m,n) = p$, and assign the first pair, (1 1), to be position $p = 1$.)

1. For all pairs, $m \geq 1$.
2. For all pairs, $n \geq 1$.
3. For all pairs, $p \geq 1$.
4. For any particular pair, $m \leq n$.
5. For all n, where $n \neq 2$ and $m = 1$, p is odd.
6. For all pairs $p \geq 2$: $\pi(m,m) > \pi(m-1,m)$.

n. (Others?)

After a cursory analysis of the beginning of the stream, I thought it too noisy to find anything which would describe it in general. I next brute forced the positions of the following pairs:

Pair	(1 100)	(2 100)	(3 100)	(4 100)	(5 100)
Pos	198	393	779	1543	3055

3 Position Formula

From this I quickly found a recursive relationship between a pair's position and the position of the one which proceeds it. Based on this I was able to find a non-recursive relationship shortly after. Let this function, f, which relates p to $(m\ n)$ be given as

$$p = f(m,n) = 2^m n - [((2m-1)(2^{m-1}) + 1] \qquad (1)$$

Checking the position found by f against a small set of arbitrary pairs found the relationship to hold.

3.1 Inconsistency

With the electronic submission deadline for the problem approaching, I churned out the positions of the pairs to be found using Equation (1); we find pair (99 100) at position

$$f(99,100) = 950737950171172051122527404031$$

and pair (100 100) at position

$$f(100,100) = 633825300114114700748351602687$$

I immediately noticed that the position of (100 100) as returned by f broke Rule 6 of the stream: the position of (100 100) must be strictly greater than (99 100).

4 Composite Position Formula

Upon f I placed the restriction that $m \neq n$ and that there must consequently exist some other function g such that

[1] Exercise 3.66 from SICP (http://mitpress.mit.edu/sicp/full-text/book/book-Z-H-24.html#%_sec_3.5.2)

$\star$	1	2	3	4	5	6	7	8	9	0
00	(1 1)	(1 2)	(2 2)	(1 3)	(2 3)	(1 4)	(3 3)	(1 5)	(2 4)	(1 6)
01	(3 4)	(1 7)	(2 5)	(1 8)	(4 4)	(1 9)	(2 6)	(1 10)	(3 5)	(1 11)
02	(2 7)	(1 12)	(4 5)	(1 13)	(2 8)	(1 14)	(3 6)	(1 15)	(2 9)	(1 16)
03	(5 5)	(1 17)	(2 10)	(1 18)	(3 7)	(1 19)	(2 11)	(1 20)	(4 6)	(1 21)
04	(2 12)	(1 22)	(3 8)	(1 23)	(2 13)	(1 24)	(5 6)	(1 25)	(2 14)	(1 26)
05	(3 9)	(1 27)	(2 15)	(1 28)	(4 7)	(1 29)	(2 16)	(1 30)	(3 10)	(1 31)
06	(2 17)	(1 32)	(6 6)	(1 33)	(2 18)	(1 34)	(3 11)	(1 35)	(2 19)	(1 36)
07	(4 8)	(1 37)	(2 20)	(1 38)	(3 12)	(1 39)	(2 21)	(1 40)	(5 7)	(1 41)
08	(2 22)	(1 42)	(3 13)	(1 43)	(2 23)	(1 44)	(4 9)	(1 45)	(2 24)	(1 46)
09	(3 14)	(1 47)	(2 25)	(1 48)	(6 7)	(1 49)	(2 26)	(1 50)	(3 15)	(1 51)
10	(2 27)	(1 52)	(4 10)	(1 53)	(2 28)	(1 54)	(3 16)	(1 55)	(2 29)	(1 56)
11	(5 8)	(1 57)	(2 30)	(1 58)	(3 17)	(1 59)	(2 31)	(1 60)	(4 11)	(1 61)
12	(2 32)	(1 62)	(3 18)	(1 63)	(2 33)	(1 64)	(7 7)	(1 65)	(2 34)	(1 66)
13	(3 19)	(1 67)	(2 35)	(1 68)	(4 12)	(1 69)	(2 36)	(1 70)	(3 20)	(1 71)
14	(2 37)	(1 72)	(5 9)	(1 73)	(2 38)	(1 74)	(3 21)	(1 75)	(2 39)	(1 76)
15	(4 13)	(1 77)	(2 40)	(1 78)	(3 22)	(1 79)	(2 41)	(1 80)	(6 8)	(1 81)
16	(2 42)	(1 82)	(3 23)	(1 83)	(2 43)	(1 84)	(4 14)	(1 85)	(2 44)	(1 86)
17	(3 24)	(1 87)	(2 45)	(1 88)	(5 10)	(1 89)	(2 46)	(1 90)	(3 25)	(1 91)
18	(2 47)	(1 92)	(4 15)	(1 93)	(2 48)	(1 94)	(3 26)	(1 95)	(2 49)	(1 96)
19	(7 8)	(1 97)	(2 50)	(1 98)	(3 27)	(1 99)	(2 51)	(1 100)	(4 16)	(1 101)
20	(2 52)	(1 102)	(3 28)	(1 103)	(2 53)	(1 104)	(5 11)	(1 105)	(2 54)	(1 106)
21	(3 29)	(1 107)	(2 55)	(1 108)	(4 17)	(1 109)	(2 56)	(1 110)	(3 30)	(1 111)
22	(2 57)	(1 112)	(6 9)	(1 113)	(2 58)	(1 114)	(3 31)	(1 115)	(2 59)	(1 116)
23	(4 18)	(1 117)	(2 60)	(1 118)	(3 32)	(1 119)	(2 61)	(1 120)	(5 12)	(1 121)
24	(2 62)	(1 122)	(3 33)	(1 123)	(2 63)	(1 124)	(4 19)	(1 125)	(2 64)	(1 126)

Table 1: First 250 pairs in sequence

$p = g(m, m)$ holds true. Curious about the anomaly being returned by f, I dug around some more. It turns out that the position value $q = f(m, m)$ does have some significance: it can be found as a component of a pair nearby! It can be found within the n component of the pair $(1\ (q + 2))$ — which is located right after $(m\ m)$. We then let the function g be defined as

$$p = g(m) = f(1, f(m, m) + 2) - 1 \tag{2}$$

In other words, a composition of the f can be used to find the true position of all pairs:

$$\pi(m, n) = \begin{cases} f(m, n) & \text{if } m \neq n, \\ f(1, f(m, n) + 2) - 1 & \text{if } m = n. \end{cases}$$

I am very interested in learning about other functions which exhibit this behavior, or more succinctly, functions which explicitly return the "wrong" result in certain cases but which can be used to find the true answer. Looking again at how the algorithm places pairs, I found that the position of $(m\ m)$ could also be found using another function h defined as

$$p = h(m) = 2^m - 1 \tag{3}$$

It turns out that the composition of f as defined by g reduces to h.

I've run into a brick wall in trying to learn more about π; why does this sudden dip in position values occur? How does the function mix position values of pairs with the component values of *other* pairs? How would one go about effectively visualizing this function? I've generated visualizations of the function π of little value since:

1. π is not continuous
2. The inputs m and n face restrictions as governed by the rules of the stream.
3. When left unbounded, π grows extremely fast — easily debilitating commonly available plotting applications.

How would one go about visualizing the properties of the algorithm based on the data alone? What branches of math does the visualization of this type of data and function fall? Any insight would be greatly appreciated.

★