

COMPUTATIONAL SOLUTIONS TO SOME CHALLENGING NUMBER
THEORY AND COMBINATORIAL PROBLEMS

by

Shanzhen Gao

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Computing and Information Systems

Charlotte

2017

Approved by:

Dr. Keh-Hsun Chen

Dr. Zbigniew Ras

Dr. Yu Wang

Dr. Aidong Lu

Dr. Yong Zhang

ABSTRACT

SHANZHEN GAO. Computational Solutions to Some Challenging Number Theory and Combinatorial Problems. (Under the direction of DR. KEH-HSUN CHEN)

We use computational assist approach to tackle some challenging and interesting problems in number theory and combinatorics, such as Markoff-Hurwitz equations, integer matrix enumeration, integer sequences, and self-avoiding walks. We present the background, what people did in the past, what we have obtained.

In the more than 100 years since Markoff-Hurwitz Equations, they play a decisive role, have turned up in an astounding variety of different settings, from number theory to combinatorics, from classical groups and geometry to the world of graphs and words, from discrete mathematics to scientific computation. We will first introduce other people's work in this area. Then we present algorithms for searching and generating solutions to the equation $x_1^2 + x_2^2 + \dots + x_n^2 = kx_1x_2\dots x_n$. Solutions are reported for $n = 2, 3, \dots, 9$. Properties of solutions are discussed. We will prove that the solutions do not exist when $n=4$ and $k=2$ or 3 ; $n=5$ and $k=2$ or 3 . Conjectures based on computational results are discussed.

The enumeration of integer-matrices has been the subject of considerable study and it is unlikely that a simple formula exists. The number in question can be related in various ways to the representation theory of the symmetric group or of the complex general linear group, but this does not make their computation any easier. We will discuss the following five problems: (1) the number of $m \times n$ matrices over $\{0, 1\}$ with each row summing to s and each column summing to t ; (2) the number of

$(0, 1)$ - matrices with restriction; (3) the number of nonnegative integer matrices of size $m \times n$ with each row sum equal to s and each column sum equal to t ; (4) the number of $(-1, 0, 1)$ - matrices of size $n \times n$; (5) the number of nonnegative matrices with restriction. We will present many conjectures based on our computation.

For self-avoiding walks, we will present: A self-avoiding walk (SAW) is a sequence of moves on a lattice not visiting the same point more than once. A SAW on the square lattice is prudent if it never takes a step towards a vertex it has already visited. Prudent walks differ from most sub-classes of SAWs that have been counted so far in that they can wind around their starting point. Some problems and some sequences arising from prudent walks are also discussed.

ACKNOWLEDGMENTS

I would like to express my most sincere gratitude to my research advisor Dr. Ken Chen for his guidance and scientific inspiration throughout my Ph.D. studies at The University of North Carolina at Charlotte. Supported by his knowledge, wisdom, encouragement, and sometimes remarkable patience, I have grown into a computer scientist. I am indebted to my supervisory committee members, Dr. Zbigniew Ras, Dr. Yu Wang, Dr. Aidong Lu and Dr. Yong Zhang for their advice, time and support throughout my years at UNCC. I thank my parents for their love and encouragement. Last but not least, I wish to express my deepest thanks to my wife Dr. Jing Ye and my two lovely children, Jia and May, for their love and unconditional support, understanding and sacrifice during my graduate studies.

TABLE OF CONTENTS

LIST OF FIGURES	viii
LIST OF TABLES	ix
CHAPTER 1: Markoff-Hurwitz Equations	1
1.1. Introduction	1
1.2. Other People's Work	2
1.3. Some Well Known Theorems	6
1.4. General Properties of Solutions	7
1.5. Search for Solutions	9
1.6. Generating Solutions	15
1.7. Results for General n and k	18
1.8. $n=3$	22
1.9. $n=4$	26
1.10. $n=5$	32
1.11. $n=6$	39
1.12. $n=7$	42
1.13. Sequences	44
1.14. Conjectures	46
CHAPTER 2: Enumeration of Integer Matrices	48
2.1. Integer Sequence	48
2.2. Zero-One Matrices with Fixed Row Sum and Column Sum	52
2.2.1. Algorithm	59

	vii
2.3. (0,1)-Matrices with Restriction	62
2.3.1. Algorithm	68
2.4. Nonnegative Integer Matrices	74
2.4.1. Algorithm	76
2.5. (-1,0,1)-Matrices	82
2.6. Nonnegative Matrices with Restriction	84
2.7. Open Problems	84
2.8. Applications	85
CHAPTER 3: Computation of Walks and Paths	86
3.1. Introduction	86
3.2. Prudent Self-Avoiding Walks: Definitions and Examples	88
3.2.1. Properties of a PSAW	89
3.3. Some Sequences Arising from One-sided PSAWs	92
3.4. Some Sequences Arising from Two-sided PSAWs	95
REFERENCES	101

LIST OF FIGURES

FIGURE 1: Tree for $n=3$ and $k=3$	5
FIGURE 2: Tree for $n=4$ and $k=4$	6
FIGURE 3: PSAW	89
FIGURE 4: PSAW2	90
FIGURE 5: Not PSAW	90
FIGURE 6: Two-sided PSAW	91

LIST OF TABLES

TABLE 1: $n = 3, 4, 5, 6$	13
TABLE 2: $n = 7, 8, 9$	14
TABLE 3: Solution Generating	17

CHAPTER 1: MARKOFF-HURWITZ EQUATIONS

1.1 Introduction

The Diophantine equation

$$x_1^2 + x_2^2 + \dots + x_n^2 = kx_1x_2\dots x_n \quad (1)$$

with k a nonzero integer and $n \geq 3$ is known as a Hurwitz or Markoff-Hurwitz equation or generalized Markoff equation. Such equations were first studied by Hurwitz [26] who thought of them as generalizations of the Markoff equation.

The Markoff equation

$$x^2 + y^2 + z^2 = 3xyz$$

was first studied by Markoff in 1879 [31], [24]. He made it famous when he noted the connection between its integral solutions, classes of quadratic forms, and Diophantine approximation. Using a descent argument, he showed all the integral solutions (except $(0,0,0)$) can be generated by the fundamental solution $(1,1,1)$ and a group of automorphisms. Its set of integer solutions is infinite and nontrivial, yet is easy to describe [7], [26], [19], [4]. In [19] and [20], equation (1) was tackled by using computational assist approach. They discussed general properties of solutions and presented an efficient systematic solution space search algorithm, and reported the search finding. They gave the theorems on non-existence of solutions on

some n & k combinations suggested from search results for x_i values up to 1,000,000. They also present an extremely fast algorithm for generating the solutions, which match exactly the result from the systematic searching/checking. Some conjectures were proposed based on their finding and observation. They also proved solution does not exist when $n = 3$ and $k = 2$.

1.2 Other People's Work

The solution triples (x, y, z) to equation (1) with $x, y, z > 0$, are called Markoff (or Markov) triples, and the numbers that appear in such a triple are called Markoff (or Markov) numbers. The first 12 Markoff triples are: $(1, 1, 1)$, $(1, 1, 2)$, $(1, 2, 5)$, $(1, 5, 13)$, $(2, 5, 29)$, $(1, 13, 34)$, $(1, 34, 89)$, $(2, 29, 169)$, $(5, 13, 194)$, $(1, 89, 233)$, $(5, 29, 433)$, $(89, 233, 610)$.

The particular interest of the Markoff equation lies in the fact that it is a quadratic equation in each of x , y and z , and hence new solutions can be obtained by a simple process from any given one. In [31] Markoff demonstrated that every Markoff triple can be obtained from $(1, 1, 1)$ by repeatedly generating new neighbors.

Conjecture 1. *Uniqueness conjecture: Every Markoff number appears exactly once as the maximum in a Markov triple.*

The unicity conjecture states that for a given Markov number c , there is exactly one normalized solution having c as its largest element: proofs of this conjecture have been claimed but none seems to be correct.[24], [35]

Frobenius [19] conjectured that for any positive integer x , there exists at most one pair of integers (y, z) with $y \geq z \geq 0$, such that (x, y, z) is solution to the Markoff

equation.

Many people have studied this conjecture. This conjecture turns up in an amazing number of different variants, from numbers and matrices to geometry and matchings of graphs. The conjecture has become widely known when Kassels mentioned it in [27]. It has been proved only for some rather special subsets of the Markoff numbers. The following result for Markoff numbers which are prime powers or 2 times prime powers was first proved independently and partly by A. Baragar [26] (for primes and 2 times primes), Button [14] (for primes but can be easily extended to prime powers) and Schmutz [32] (for prime powers but the proof works also for 2 times prime powers) using either algebraic number theory or hyperbolic geometry. A Markoff number is unique if it is a prime power or 2 times a prime power. In [43] Zhang claimed that if c is an even Markoff number then $c \equiv 2 \pmod{32}$. A Markoff number c is unique if one of $3c + 2$ and $3c - 2$ is a prime power, 4 times a prime power, or 8 times a prime power in [5].

A. Baragar described the Markoff equations and their orbits of integer solutions in [7]. He showed that the number of orbits of integer solutions is finite, and he described a sequence of equations for which this number goes to infinity. This is described in more detail in [5]. He also described the asymptotic growth of Markoff numbers in each orbit, and sketched a proof which requires an assumption which he later removed in [4], and improved in [6].

In [1], Aigner showed that the triples $(1, 1, 1)$ and $(2, 1, 1)$ are the only Markov triples with repeated numbers.

D. Zagier [42] investigated the asymptotic growth for the number of solutions to

the Markov equation ($k=n=3$) below a given bound and Baragar [4] investigated the cases $n \geq 4$. Several other researchers also studied the asymptotic growth or the ratio of the numbers from the generalized equation (1).

In [23] Enrique and Jose studied the solutions of the Rosenberg–Markoff equation (a generalization of the well-known Markoff equation) $ax+by+cz=dx yz$. They specifically focus on looking for solutions in arithmetic progression that lie in the ring of integers of a number field. With the help of previous work by Alvanos and Poulakis, they give a complete decision algorithm, which allows them to prove finiteness results concerning these particular solutions.

In [3] Ioulia studied Generalizations of the Markoff–Hurwitz equations over residue class rings. A number of novel features of Markoff numbers were found from the graph-theoretical standpoint in [25]. Namely, for a given Markoff number there exist a pair of graphs, caterpillar and linearly growing polyomino, whose topological index and perfect matching number are, respectively, equal to that number. Efficient step wise algorithms and recursion formulas are found for enumerating these two characteristic quantities of these special graphs, which have either mirror or rotational symmetry.

The Markoff equation is quadratic in each variable, so given a solution (x, y, z) , we can find solutions $(3yz - x, y, z)$, $(x, 3xz - y, y)$, $(x, y, 3xy - z)$. Using this map, permutations of the variables, and the fundamental solution $(1, 1, 1)$, we can construct a Markoff tree of positive ordered solutions. We let the three coordinates in a solution be in decreasing order. The following (Figure 1) is such a tree.

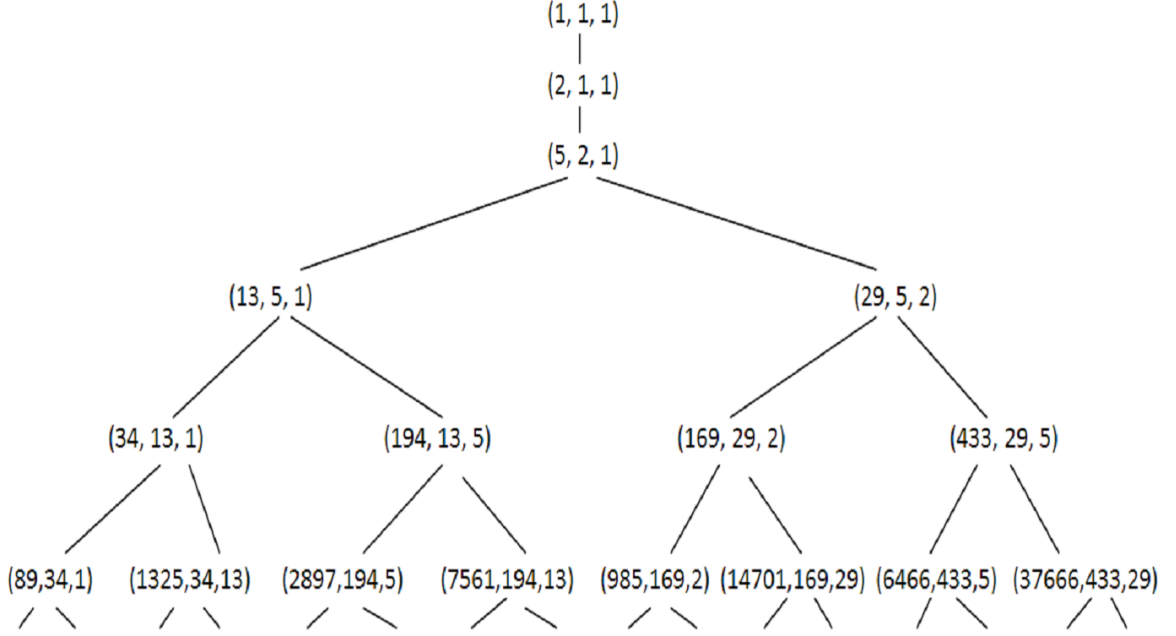


Figure 1: Tree for $n=3$ and $k=3$

We also construct a tree (Figure 2) for

$$x^2 + y^2 + z^2 + w^2 = 4xyzw.$$

Hurwitz generalized Markoff tree and showed that for equation (1), with $n \geq 3$ and $k \leq n$, any existed positive integer solution can be listed in one of a finite number of trees formed from a well-defined set of root solutions. Arthur [4] characterized all pairs of n and k with $k \geq 2(n-1)^{1/2}$ for which the Hurwitz equation has positive integer solutions. And the possibility that there may be more than one tree.

There are many approaches to equation (1), we will concentrate on the existence and the structures of solutions, and algorithms.

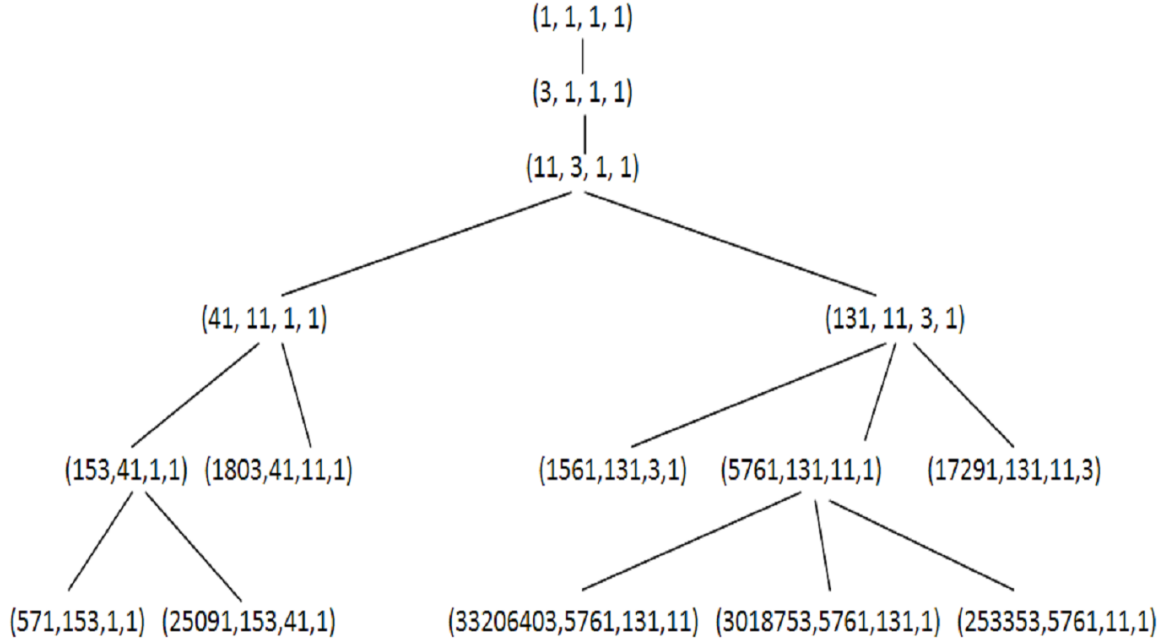


Figure 2: Tree for $n=4$ and $k=4$

1.3 Some Well Known Theorems

As a foundation, we recall the following important related theorems. We will show new proofs for some of those theorems in section 1.7 "Results for General n and k ".

Theorem 2. *When $n = 2$, $x_1^2 + x_2^2 = kx_1x_2$ has solution iff $k = 2$.*

For any positive integer c , $x_1 = x_2 = c$ is a solution.

Theorem 3. *If $n = k$, $x_1 = x_2 = \dots = x_n = 1$ is a solution to equation (1).*

Theorem 4. *If x_1, x_2, \dots, x_n is a solution to equation (1), then $x_1, x_2, \dots, x_{i-1}, kx_1x_2\dots x_{i-1}x_{i+1} \dots x_n - x_i, x_{i+1}, \dots, x_n$ is also a solution (for each $1 \leq i \leq n$).*

Theorem 5. *Equation (1) does not have solutions when $k > n$ for $n \geq 2$. [26], [18], [20].*

For the rest of the paper, we shall focus on the cases $n \geq 3$ and $1 \leq k \leq n$ for positive integer solutions x_1, x_2, \dots, x_n .

1.4 General Properties of Solutions

In this section, we shall define our notations and terminologies and discuss some general properties of solutions to equation (1).

We call a solution x_1, x_2, \dots, x_n of equation (1) an ordered solution if $x_1 \geq x_2 \geq \dots \geq x_n$.

Let $x_i' = kx_1x_2\dots x_{i-1}x_{i+1}\dots x_n - x_i$. By Theorem 4, we know if $x_1, x_2, \dots, x_i, \dots, x_n$ form a solution of equation (1), then $x_1, x_2, \dots, x_i', \dots, x_n$ also form a solution. We say the new solution comes from the original solution by applying Theorem 4 on index i .

Lemma 6. *Let $X: x_1, x_2, \dots, x_i, \dots, x_n$ be an ordered solution and $i \geq 2$. Then*

$$x_i' = kx_1x_2\dots x_{i-1}x_{i+1}\dots x_n - x_i > x_1.$$

Proof. $kx_1x_2\dots x_{i-1}x_{i+1}\dots x_n - x_i$.

$$\begin{aligned} &= (x_1^2 + x_2^2 + \dots + x_n^2) / x_i - x_i \\ &= x_1^2 / x_i + (x_2^2 + \dots + x_n^2) / x_i - x_i > x_1^2 / x_i \\ &\geq x_1 \end{aligned}$$

□

Definition 7. *Let $X: x_1, x_2, \dots, x_i, \dots, x_n$ and $Y: y_1, y_2, \dots, y_i, \dots, y_n$ be two ordered solutions of (1). We define $X < Y$ (X comes before Y in lexical order) iff there exists i such that $1 \leq i \leq n$ and $x_j = y_j$ for $1 \leq j < i$ and $x_i < y_i$. We say X is the minimum solution if $X < Y$ for any other solution Y of (1).*

Theorem 8. *If $X: x_1, x_2, \dots, x_i, \dots, x_n$ is an ordered solution of (1), then $X_i': x_i', x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_n$, where x_i' is as defined in Lemma 6, is also an ordered solution and $X < X_i'$ for $2 \leq i \leq n$. In notation, $X \mid_{-i} X_i'$.*

Proof. Directly follows Theorem 4 and Lemma 6. □

Note that for an ordered solution x_1, x_2, \dots, x_n , x_1' may be $<$, $>$, or $= x_1$. For examples, for $n = 7$ & $k = 5$: $3, 1, 1, 1, 1, 1, 1$ is a solution, applying Theorem 4 to index 1, we get a new solution $2, 1, 1, 1, 1, 1, 1$ which is smaller. From $2, 1, 1, 1, 1, 1, 1$, we get $3, 1, 1, 1, 1, 1, 1$, which is larger. For $n = 5$ & $k = 4$: $2, 1, 1, 1, 1$ is a solution, applying Theorem 3 on index 1, we get $2, 1, 1, 1, 1$ itself. So if we apply the rule in Theorem 4 on index 1 to an ordered solution $X: x_1, x_2, \dots, x_n$, the resulting solution $X_1': x_1', x_2, \dots, x_n$ may not be an ordered solution.

Let insert x_1' in proper order into x_2, \dots, x_n to form an ordered solution X_1'' . Then we shall use the notation $X \mid_{-1} X_1''$.

Definition 9. *Let X, Y be ordered solutions of (1).*

We define $X \mid - Y$ iff $X \mid_{-i} Y$ for some $i = 1, 2, \dots, n$,

and $X \mid^ Y$ iff there exists X_1, X_2, \dots, X_m such that $X = X_1$, $Y = X_m$, and $X_i \mid - X_{i+1}$ for $i = 1, 2, \dots, m-1$.*

Note that x_i and $x_i' = kx_1x_2\dots x_{i-1}x_{i+1}\dots x_n - x_i$ are the two solutions of the quadratic equation

$$y^2 - (kx_1x_2\dots x_{i-1}x_{i+1}\dots x_n)y + (x_1^2 + x_2^2 + \dots + x_{i-1}^2 + x_{i+1}^2 + \dots + x_n^2) = 0.$$

When $i > 1$, $x_i < x_i'$. When $i = 1$, x_1 may $>$, $<$, or $= x_1'$.

Theorem 10. *If equation (1) has one solution then it has infinitely many solutions.*

Proof. Let X be an ordered solution of (1). Then

$$X \mid_n X_1 \mid_n X_2 \mid_n \dots$$

$$X < X_1 < X_2 < \dots /$$

There are infinitely many solutions. □

Definition 11. *Let X be an ordered solution of equation (1) and $X \mid_{-1} Y$. If $X \leq Y$, then we call X a fundamental solution of equation (1).*

It is still an open question whether equation (1) has a fundamental solution other than the minimum solution for some n & k .

Let X be an ordered solution of (1) and $X \mid_{-i} Y_i$ for $i = 1, 2, \dots, n$. If X is not a fundamental solution then $Y_1 < X$ and $Y_i > X$ for $i = 2, 3, \dots, n$. Furthermore, $Y_i \mid_{-1} X$ for $i = 2, 3, \dots, n$ and $Y_1 \mid X$.

Definition 12. *If $X \mid Y$ where $X \neq Y$ are ordered solutions of (1), then we say X and Y are adjacent ordered solutions.*

The following theorem is obvious.

Theorem 13. *Let $X = x_1, x_2, \dots, x_n$ be an ordered solution of (1). X has $|\{x_1, x_2, \dots, x_n\}|$ adjacent solutions if X is not a fundamental solution, otherwise it may have one less adjacent solutions than $|\{x_1, x_2, \dots, x_n\}|$ (in case $X \mid_{-1} X$).*

1.5 Search for Solutions

We shall present an algorithm for checking all potential ordered solutions in lexical order with each component $x_i \leq \text{Limit}$ where Limit is an input constant.

The algorithm will skip any range where the nonexistence of solutions can be inferred by the following proposition.

Proposition 1. If $kx_1x_2...x_i > x_1^2 + x_2^2 + ... + x_i^2 + (n-i)$ with $i < n$ then $kx_1x_2...x_n > x_1^2 + x_2^2 + ... + x_n^2$ for any $x_{i+1}, x_{i+2}, \dots, x_n$ such that $x_i \geq x_{i+1} \geq x_{i+2} \geq \dots \geq x_n > 0$.

Proof. Case 1. $x_{i+1} = 1$

Then $x_{i+2} = x_{i+3} = \dots = x_n = 1$.

$$kx_1x_2...x_ix_{i+1}...x_n = kx_1x_2...x_i$$

$$> x_1^2 + x_2^2 + ... + x_i^2 + (n-i) = x_1^2 + x_2^2 + ... + x_n^2$$

Case 2. $x_{i+1} \geq 2$

$$kx_1x_2...x_ix_{i+1} \geq kx_1x_2...x_i * 2$$

$$> x_1^2 + x_2^2 + ... + x_i^2 + (n-i) + x_1^2 + x_2^2 + ... + x_i^2 + (n-i)$$

$$= x_1^2 + x_2^2 + ... + x_i^2 + x_1^2 + (x_2^2 + ... + x_i^2 + (n-i) + (n-i))$$

$$> x_1^2 + x_2^2 + ... + x_i^2 + x_{i+1}^2 + (n-(i+1))$$

We can consider one extra x_j at a time until $j = n$ or $x_j = 1$ (Case 1). Hence the product will always be bigger than the sum of squares under the condition. \square

Our systematic search/checking algorithm with cut-offs based on Proposition 1 is coded below in a C-like pseudo code:

Algorithm 1 Systematic Search for Solutions

// Check all possible ordered solutions with each

// component $1 \leq x[i] \leq \text{Limit}$

// $s[i] = x[1]*x[1] + x[2]*x[2] + \dots + x[i]*x[i]$

// $p[i] = k*x[1]*x[2]*\dots*x[i]$

```

s[0]=0;

p[0]=k;

for (j = i; j < n; j++)

x[j]=1;

x[n] = 0;

x[0]=Limit;

while (x[n] < Limit) {

for (j = n; j >= 1; j- -) {

if (x[j] < x[j]-1) {

    x[j]++;

    for (j1=j+1; j1 <= n; j1++)

x[j1]=1;

    break;

}

}

for (i = 1; i <= n; i++) {

s[i] = s[i-1];

p[i] = p[i-1];

s[i] += x[i]*x[i];

p[i] *= x[i];

if (p[i] > s[i]+n-i) {

//cut off based on the Proposition 1

j = i-1;

```

```

for (i1 = j; i1 >= 1; i1- -) {
    if (x[i1] < x[i1-1]) {
        x[i1]++;
        for (j1 = i1+1; j1 < n; j1++)
            x[j1]=1;
        x[n]=0;
        break;
    }
}
goto L;
}
}

if (s[n] == p[n]) {
    fprintf(out, "\n");
    for (j = 1; j <= n; j++)
        fprintf(out, " %d, ", x[j]);
}

L: continue;
}

```

The average time complexity of Algorithm 1 is about $O(Limit^2)$ for given n , since the cut-offs usually occur at early stages when $i = 2$ or 3 .

Results

We implemented the Algorithm 1 and ran on a half dozen PCs of Intel Core i7-

2600K Quad-Core Processor 3.4 Ghz or equivalent for $n = 3, 4, \dots, 9$ and $k = 1, 2, \dots, n$. The Limit was set to 1,000,000. Each (n, k) case used a single thread. A PC ran up to 8 cases simultaneously. The timing data in hours have been converted to that of Intel Core i7-2600K 3.4 Ghz for comparison. Table 1 summarizes the results for $n = 3, 4, 5, 6$ and Table 2 for $n = 7, 8, 9$.

Table 1: $n = 3, 4, 5, 6$

	Limit 1,000,000	N = 3	N = 4	N = 5	N = 6
k=1	#Ord. Solutions	35	23	177	0
	Min.Solutions	3,3,3	2,2,2,2	4,3,3,1,1	No
	Proc. Time in hrs.	13.5	58.8	138.3	208.53
k=2	#Ord. Solutions	0	0	0	0
	Min.Solutions	No	No	No	No
	Proc. Time in hrs.	8.2	28.8	64.3	104
k=3	#Ord. Solutions	40	0	0	76
	Min.Solutions	1,1,1	No	No	2,2,1,1,1,1
	Proc. Time in hrs.	4.3	16	37.6	59.43
k=4	#Ord. Solutions		24	30	0
	Min.Solutions		1,1,1,1	2,1,1,1,1	No
	Proc. Time in hrs.		11.4	27.87	41.5
k=5	#Ord. Solutions			17	0
	Min.Solutions			1,1,1,1,1	No
	Proc. Time in hrs.			21.7	32
k=6	#Ord. Solutions				15
	Min.Solutions				1,1,1,1,1,1
	Proc. Time in hrs.				26.2

Table 2: $n = 7, 8, 9$

	Limit 1,000,000	$N = 7$	$N = 8$	$N = 9$
k=1	#Ord. Solutions	124	61	0
	Min.Solution	3,2,2,2,1,1,1	4,2,2,2,1,1,1,1	No
	Proc. Time in hrs.	279.95	356.11	376.87
k=2	#Ord. Solutions	64	0	0
	Min.Solution	2,2,2,1,1,1,1	No	No
	Proc. Time in hrs.			
k=3	#Ord. Solutions	65	0	0
	Min.Solution	3,2,1,1,1,1,1	No	No
	Proc. Time in hrs.	72.8	90.6	96.6
k=4	#Ord. Solutions	0	0	0
	Min.Solution	No	No	No
	Proc. Time in hrs.	51.1	63.5	69.6
k=5	#Ord. Solutions	42	0	0
	Min.Solution	2,1,1,1,1,1,1	No	No
	Proc. Time in hrs.	41.45	48.2	52.8
k=6	#Ord. Solutions	0	0	33
	Min.Solution	No	No	2,1,1,1,1,1,1,1
	Proc. Time in hrs.	31.3	38.4	40.67
k=7	#Ord. Solutions	12	0	0
	Min.Solution	1,1,1,1,1,1,1	No	No
	Proc. Time in hrs.	25.7	31.7	34.6
k=8	#Ord. Solutions		10	0
	Min.Solution		1,1,1,1,1,1,1,1	No
	Proc. Time in hrs.		26.9	29.4
k=9	#Ord. Solutions			10
	Min.Solution			1,1,1,1,1,1,1,1,1
	Proc. Time in hrs.			25.4

There are no solutions with all $x_i \leq 1,000,000$ for (n, k) cases: $(3, 2)$, $(4, 2)$, $(4, 3)$, $(5, 2)$, $(5, 3)$, $(6, 1)$, $(6, 2)$, $(6, 4)$, $(6, 5)$, $(7, 4)$, $(7, 6)$, $(8, 2)$, $(8, 3)$, $(8, 4)$, $(8, 5)$, $(8, 6)$, $(8, 7)$, $(9, 1)$, $(9, 2)$, $(9, 3)$, $(9, 4)$, $(9, 5)$, $(9, 7)$, $(9, 8)$.

There is a strong possibility that no solutions at all on these cases. We have found mathematical proofs for non-existence of solutions for cases $(3, 2)$, $(4, 2)$, $(4, 3)$, $(5, 2)$, $(5, 3)$ and are working on proofs for the rest cases. You will see some of the proofs in section 1.8, 1.9 and 1.10.

The minimum solutions for $3 \leq n \leq 9$ and $1 \leq k \leq n$ other than the cases listed above contain no components > 4 .

For a fix n , the processing time decreases as k increases since larger k induces more cut-offs (Proposition 1). For a fix k , the processing time increases as n increases since the equation (1) gets more complex on larger n .

1.6 Generating Solutions

In this section, we shall present a fast algorithm for generating all ordered solutions up to a given value limit on the solution components.

Algorithm 2 Generating Solutions

We use Algorithm 1 to find the minimum solution M (within the limit). If the minimum solution is found, proceeds as follows:

Keep a lexically ordered linked list L of all ordered-solutions found so far. Initially L consists of the minimum solution M alone. Traverse through L until the end. Let $X: x_1, x_2, \dots, x_n$ be initialized to the 2nd ordered solution in the linked list L .

while $(X \neq \text{null})$ {


```

for (i = 2; i <= n; i++) {
    if ( $x_i \neq x_{i-1}$ ) {
        generate a new solution Y using the
        rule in Theorem 4 on index i.
        if  $Y > X$  and all yj with in limit
        then insert Y to L in lexical order
        //if  $Y \leq X$ , Y is already in L
    }
}
X = the next ordered solution in L.
}

```

The time complexty of Algorithm 2 is $O(s^2)$ where s is the number of ordered solutions within the limit when case (n, k) has solutions, the same as Algorithm 1 when there is no solutions. Algorithm 2 is several order of magnitude faster than Algorithm1 when solutions exist. When we used 1,000,000 as the value limit, this algorithm took just a small fraction of a second to solve equation (1) (if there is a solution) when Algorithm 1 had taken days.

The solution generating algorithm produced exactly the same sets of solutions in $\{1, 2, \dots, 1,000,000\}$ for $3 \leq n \leq 9$ and $k \leq n$. We can conclude that there is at most one fundamental solution with components $\leq 1,000,000$ of equation (1) for each (n, k) where $3 \leq n \leq 9$ and $k \leq n$. Of course this fundamental solution is the minimum solution.

We implemented and ran Algorithm 2 to value limit 10^{18} on cases of tables 1 & 2

having solutions. We get the results as in Table 3 below.

Limit: 1,000,000,000,000,000

#OS: Number of ordered solutions

PTS: Processing time in seconds

Table 3: Solution Generating

		N = 3	N = 4	N = 5	N = 6	N = 7	N = 8	N = 9
k=1	#OS	311	253	3010	0	2560	1195	
	PTS	0.010	0.002	0.076		0.055	0.019	
k=2	#OS	0	0	0	0	998	0	0
	PTS					0.01		
k=3	#OS	328	0	0	1092	927	0	0
	PTS	0.03			0.007	0.012		
k=4	#OS		263	378	0	0	0	0
	PTS	0.002	0.003					
k=5	#OS			172	0	479	0	0
	PTS			0.002		0.005		
k=6	#OS				139	0	0	366
	PTS				0.002			0.002
k=7	#OS					106	0	0
	PTS					0.002		
k=8	#OS						92	0
	PTS						0.001	
k=9	#OS							79
	PTS							0.001

1.7 Results for General n and k

We will discuss and prove some theorems in this section.

We will only consider $n \geq 3$ in the following.

Theorem 14. *If x_1, x_2, \dots, x_n is a solution to equation (1), then $x_1, x_2, \dots, x_{i-1}, (\frac{k}{x_i} \prod_{j=1}^n x_j) - x_i, x_{i+1}, \dots, x_n$ is also a solution ($1 \leq i \leq n$).*

Proof. If x_1, x_2, \dots, x_n is a solution to equation (1), then

$$x_1^2 + x_2^2 + \dots + x_n^2 = kx_1x_2\dots x_n.$$

We want to show that

$$\begin{aligned} & x_1^2 + x_2^2 + \dots + x_{i-1}^2 + \left(\frac{k}{x_i} \prod_{j=1}^n x_j - x_i\right)^2 + x_{i+1}^2 + \dots + x_n^2 \\ &= \frac{kx_1x_2\dots x_n}{x_i} \left(\frac{kx_1x_2\dots x_n}{x_i} - x_i\right). \end{aligned}$$

Now we consider

$$\begin{aligned} & x_1^2 + x_2^2 + \dots + x_{i-1}^2 + \left(\frac{k}{x_i} \prod_{j=1}^n x_j - x_i\right)^2 + x_{i+1}^2 + \dots + x_n^2 \\ &= x_1^2 + x_2^2 + \dots + x_{i-1}^2 + k \frac{kx_1^2x_2^2\dots x_n^2}{x_i^2} - 2kx_1x_2\dots x_n + x_i^2 + x_{i+1}^2 + \dots + x_n^2 \\ &= x_1^2 + x_2^2 + \dots + x_n^2 - 2kx_1x_2\dots x_n + k \frac{kx_1^2x_2^2\dots x_n^2}{x_i^2} \\ &= k \frac{kx_1^2x_2^2\dots x_n^2}{x_i^2} - kx_1x_2\dots x_n \\ &= \frac{kx_1x_2\dots x_n}{x_i} \left(\frac{kx_1x_2\dots x_n}{x_i} - x_i\right). \end{aligned}$$

Therefore, $x_1, x_2, \dots, x_{i-1}, (\frac{k}{x_i} \prod_{j=1}^n x_j) - x_i, x_{i+1}, \dots, x_n$ is also a solution to equation (1)

□

Theorem 15. *Equation (1) does not have solution when $k > n$ for $n \geq 2$.*

Proof. We give an outline of a constructive proof of this theorem in the following.

(1) Claim: $(n - 1) + u^2 \leq nu$ when $n > 3$ and $1 \leq u \leq n - 1$.

To prove this claim, we let $f(u) = u^2 + (n - 1) - nu$. Then $\frac{d}{du}(u^2 + (n - 1) - nu) = 2u - n$.

We need to consider:

(1) $u = 1$ or $u = n - 1$.

(2) $1 < u < n/2$.

(3) $n/2 \leq u < n - 1$.

Case 1: $x_1 = x_2 = \dots = x_n$

Solution does not exist.

Case 2: $x_1 = x_2 = \dots = x_{n-1} \neq x_n$

Solution does not exist.

Case 3: $x_1 \leq x_2 \leq \dots \leq x_n$ (at least one " $<$ ")

Then $kx_1x_2\dots x_{n-1} - x_n < x_n$.

Case 3-1: $2x_n = kx_1x_2\dots x_{n-1} + \sqrt{k^2x_1^2x_2^2\dots x_{n-1}^2 - 4(x_1^2 + x_2^2 + \dots + x_{n-1}^2)}$

Case 3-2: $2x_n = kx_1x_2\dots x_{n-1} - \sqrt{k^2x_1^2x_2^2\dots x_{n-1}^2 - 4(x_1^2 + x_2^2 + \dots + x_{n-1}^2)}$

Therefore, the sum of the n components of the solution $x_1, x_2, \dots, x_{n-1}, kx_1x_2\dots x_{n-1} - x_n$ is smaller than that of the solution $x_1, x_2, \dots, x_{n-1}, x_n$.

Transforming $x_1, x_2, \dots, x_{n-1}, x_n$ to $kx_1x_2\dots x_{n-1} - x_n$ finite times, resulting two kinds of solution:

$x_1 = x_2 = \dots = x_n$ or $x_1 = x_2 = \dots = x_{n-1} \neq x_n$. None of them is a solution. \square

Theorem 16. *If x_1, x_2, \dots, x_n , ($x_1 \geq x_2 \geq \dots \geq x_n$) is a solution to equation (1), then there are infinite solutions of the form $a, b, x_3, x_4, \dots, x_n$ with distinct leading terms (i.e., the first component).*

Proof. Let a, b, x_3, \dots, x_n , ($a \geq b \geq \dots \geq x_n$) be a solution to equation (1), then

$$a^2 + b^2 + x_3^2 + \dots + x_n^2 = kabx_3 \dots x_n.$$

$$\begin{aligned} & (kax_3 \dots x_n - b)^2 + a^2 + x_3^2 + \dots + x_n^2 - k(kax_3 \dots x_n - b)ax_3 \dots x_n \\ &= (kax_3 \dots x_n)^2 - 2kabx_3 \dots x_n + b^2 + a^2 + x_3^2 + \dots + x_n^2 - (kax_3 \dots x_n)^2 + kabx_3 \dots x_n \\ &= -kabx_3 \dots x_n + a^2 + b^2 + x_3^2 + \dots + x_n^2 \\ &= 0 \end{aligned}$$

And $(kax_3 \dots x_n - b) > a$.

Thus, $(kax_3 \dots x_n - b), b, x_3, \dots, x_n$ is another solution to equation (1).

Therefore, the theorem holds. □

We defined minimum solution in section 1.3. We also can define minimum solution in the following way.

Definition 17. *A minimum solution is a solution with x_1 is minimum among all solutions, where solution is in the form $x_1 \geq x_2 \geq \dots \geq x_n$.*

Definition 18. *Given two positive integers n and k , the Diophantine sequence of (1) is the list of x'_1 's in nondecreasing order over all ordered solutions, ($x_1 \geq x_2 \geq \dots \geq x_n$). In short, we say a sequence for n and k .*

Conjecture 19. *If x_1, x_2, \dots, x_n and y_1, y_2, \dots, y_n are two non-identical solutions to equation (1) with component numbers in non-increasing order, then $x_1 \neq y_1$.*

Theorem 20. *There is a solution with all components has a common divisor c ($c > 2$) for equation (1) if and only if $n = 3$ and $k = 1$.*

Proof. When $n = 3$ and $k = 1$, $x_1 = x_2 = x_3 = 3$ is a solution to $x_1^2 + x_2^2 + x_3^2 = x_1 x_2 x_3$.

Let $x_i = cy_i$ for $i = 1, 2, \dots, n$, be a solution to equation (1).

$$y_1^2 + y_2^2 + \dots + y_n^2 = c^{n-2} k y_1 y_2 \dots y_n.$$

$$\text{Let } f(n) = c^{n-2} k - n.$$

$$\frac{d}{dn}(c^{n-2} k - n) = c^{n-2} (\ln c) k - 1 > 0 \text{ for } n > 2.$$

Thus $f(n)$ is an increasing function when $n > 2$.

When $n \geq 3$, $\min(f(n)) = 0$ iff $c = 3$, $n = 3$ and $k = 1$.

Thus $f(n) > 0$ when $n \geq 4$ and $k \geq 1$.

$$y_1^2 + y_2^2 + \dots + y_n^2 = k 2^{n-2} y_1 y_2 \dots y_n \text{ has no solution when } n \geq 4.$$

The theorem holds. □

Theorem 21. *Let $x_i = 2y_i + 1$ ($i = 1, 2, \dots, n$). If x_1, x_2, \dots, x_n is a solution to equation (1), then either n and k are both even, or n and k are both odd.*

Theorem 22. *Let $x_1 = x_2 = \dots = x_n = c$. If x_1, x_2, \dots, x_n is a solution to equation (1), then $n = k$ and $c = 1$, or $n = 3$ and $k = 1$, or $n = 4$ and $k = 1$.*

Proof. Let $x_i = c$ ($i = 1, 2, \dots, n$) be a solution to equation (1). Then

$$nc^2 = kc^n, n = kc^{n-2}$$

If $c = 1$, then $n = k$.

If $c = 2$, then $n = 4$ and $k = 1$. (from a previous theorem)

If $c = 3$, then $\frac{n}{k} = 3^{n-2}$.

For $c \geq 3$, we define $f(n) = kc^{n-2} - n$.

$$\frac{d}{dn}(kc^{n-2} - n) = kc^{n-2} \ln c - 1$$

$$kc^{n-2} \ln c - 1 > 0.$$

$f(n) = kc^{n-2} - n$ is an increasing function.

If $f(3) = kc^{3-2} - 3 = 0$, then $k = 1$ and $c = 3$, since $c \geq 3$.

$$f(n) > 0 \text{ for } n > 3.$$

Therefore there is no solution of the form $x_i = c$ ($i = 1, 2, \dots, n$) when $n > 3$ and $c \geq 3$.

□

1.8 n=3

Theorem 23. *Solution does not exist when $n = 3$ and $k = 2$.*

Proof. Suppose that x_1, x_2, x_3 is a solution. Then $x_1^2 + x_2^2 + x_3^2 = 2x_1x_2x_3$.

x_1, x_2, x_3 can not be all odd.

Case 1: x_1, x_2, x_3 : one even, two odd

Say $x_1 = 2a + 1, x_2 = 2b + 1, x_3 = 2c$.

$$(2a + 1)^2 + (2b + 1)^2 + (2c)^2 = 2(2a + 1)(2b + 1)2c$$

$$4a^2 + 4a + 4b^2 + 4b + 4c^2 + 2 = 4(2a + 1)(2b + 1)c$$

left side = $2 \pmod{4}$, right side = $0 \pmod{4}$.

Solution does not exist for this case.

Case 2: x_1, x_2, x_3 : two even, one odd

Say $x_1 = 2a, x_2 = 2b, x_3 = 2c + 1$.

$$\text{Left side} = (2a)^2 + (2b)^2 + (2c + 1)^2 = 1 \pmod{4}$$

$$\text{Right side} = 0 \pmod{4}$$

Solution does not exist for this case

Case 3: x_1, x_2, x_3 : all even

Say $x_1 = 2a, x_2 = 2b, x_3 = 2c$.

$$(2a)^2 + (2b)^2 + (2c)^2 = 2(2a)(2b)(2c)$$

$$a^2 + b^2 + c^2 = 4abc$$

Solution does not exist for this case.

Therefore, solution does not exist when $n = 3$ and $k = 2$. □

Theorem 24. *If x_1, x_2, x_3 is a solution to the equation (1) for $n = 3$ and $k = 3$ if and only if $3x_1, 3x_2, 3x_3$ is a solution to the equation (1) for $n = 3$ and $k = 1$.*

Proof. Let x_1, x_2, x_3 be a solution for $n = 3$ and $k = 3$, then $x_1^2 + x_2^2 + x_3^2 = 3x_1x_2x_3$.

$$\text{Then } 9x_1^2 + 9x_2^2 + 9x_3^2 = 9 \times 3x_1x_2x_3.$$

$$(3x_1)^2 + (3x_2)^2 + (3x_3)^2 = (3x_1)(3x_2)(3x_3).$$

Therefore, $3x_1, 3x_2, 3x_3$ is a solution for the case $n = 3$ and $k = 1$.

Let $3x_1, 3x_2, 3x_3$ is a solution for the case $n = 3$ and $k = 1$.

$$\text{Then } (3x_1)^2 + (3x_2)^2 + (3x_3)^2 = (3x_1)(3x_2)(3x_3).$$

$$\text{Then } x_1^2 + x_2^2 + x_3^2 = 3x_1x_2x_3.$$

Therefore, x_1, x_2, x_3 is a solution for the case $n = 3$ and $k = 3$. □

Theorem 25. *If $x_1^2 + x_2^2 + x_3^2 = x_1x_2x_3$, then $x_i \equiv 0 \pmod{3}$ for $i = 1, 2, 3$.*

We can easily prove this theorem by using the previous theorem. We give another proof in the following.

Proof. We want to show that x_1, x_2, x_3 is a solution, then $x_i \equiv 0 \pmod{3}$ for $i = 1, 2, 3$.

If $x_1 = 3a + x$, $x_2 = 3b + y$, $x_3 = 3c + z$ with $0 \leq x, y, z \leq 2$ is a solution, then

$$(3a + x)^2 + (3b + y)^2 + (3c + z)^2 = (3a + x)(3b + y)(3c + z).$$

$$\begin{aligned} & 9a^2 + 6ax + 9b^2 + 6by + 9c^2 + 6cz + x^2 + y^2 + z^2 \\ &= 27abc + 9abz + 9yac + 3yaz + 9xbc + 3xbz + 3yxc + xyz \end{aligned}$$

$$\text{left side} = x^2 + y^2 + z^2 \pmod{3}$$

$$\text{right side} = xyz \pmod{3}$$

Case 1: all of $x, y, z \neq 0$.

Case 1-1: all of $x, y, z = 1$, then left side $= 0 \pmod{3}$, right side $= 1 \pmod{3}$.

Case 1-2: all of $x, y, z = 2$, then left side $= 0 \pmod{3}$, right side $= 2 \pmod{3}$.

Case 1-3: $x = 1, y = 1, z = 2$, then left side $= 0 \pmod{3}$, right side $= 2 \pmod{3}$.

Case 1-4: $x = 1, y = 2, z = 2$, then left side $= 0 \pmod{3}$, right side $= 1 \pmod{3}$.

Case 2: only $x = 0$

Case 2-1: $y = 1, z = 1$

$$\text{left side} = 2 \pmod{3}$$

$$\text{right side} = 0 \pmod{3}$$

Case 2-2: $y = 1, z = 2$

$$\text{left side} = 2 \pmod{3}$$

$$\text{right side} = 0 \pmod{3}$$

Case 2-3: $y = 2, z = 2$

$$\text{left side} = 2 \pmod{3}$$

$$\text{right side} = 0 \pmod{3}$$

Case 3: $x = 0, y = 0, z \neq 0$

Case 3-1: $z = 1$

left side= $1 \bmod 3$

right side= $0 \bmod 3$

Case 3-2: $z = 2$

left side= $1 \bmod 3$

right side = $0 \bmod 3$

Therefore $x_1 = 0 \bmod 3, x_2 = 0 \bmod 3$ and $x_3 = 0 \bmod 3$.

□

K.Guy and R.Nowakowski asked:”For what pairs of integers a, b does ab exactly divide $a^2 + b^2 + 1$?” [29]. We answer this question in the following theorem.

Theorem 26. *If $a^2 + b^2 + 1 = kab$ has solutions, then $k = 3$. And there are infinite solutions.*

Proof. We already proved that solution does not exist when $n = 3$ and $k = 2$ in the equation (1).

When $k = 3$, $a = b = 1$ is a solution.

If $k = 1$, then $a^2 + b^2 + 1 = ab$. It is not true since $a^2 + b^2 + 1 > ab$.

Therefore, solution dose not exist when $k = 1$ or 2 . Solutions do exist when $k = 3$.

Let a and b with $a \geq b \geq 1$ be a solution to $a^2 + b^2 + 1 = 3ab$.

We will show that $3a - b, a$ is another solution for $a^2 + b^2 + 1 = kab$, i.e., $(3a - b)^2 + a^2 + 1 = 3(3a - b)a$, i.e., $(3a - b)^2 + a^2 + 1 = 3(3a - b)a$.

$$(3a - b)^2 + a^2 + 1 - 3(3a - b)a = a^2 + b^2 + 1 - 3ab = 0.$$

It is clear that $3a - b > a$ and $a > b$.

Therefore, there are infinite solutions.

□

Theorem 27. *The leading terms of any solution and its adjacent generated solutions of $a^2 + b^2 + c^2 = abc$ are distinct.*

Proof. If a, a, b is a solution, then $2a^2 + b^2 = a^2b$.

Then $b^2 \bmod a^2 = 0, b \bmod a = 0$

Let $b = aq$, then $2 + q^2 = aq$.

$2 \bmod q = 0$, then $q = 2$ or 1 .

If $q = 1$, then $b = a$. The unique solution is $3, 3, 3$.

If $q = 2$, then $b = 2a$. Then $a^2 + a^2 + (2a)^2 = 2a^2a$. $a = 3$. The unique solution is $6, 3, 3$.

Assume $a > b > c \geq 3$ satisfying $a^2 + b^2 + c^2 = abc$.

We can generate three solutions from a solution a, b, c :

$ab - c, a, b$, $(ab - c > a)$

$ac - b, a, c$, $(ac - b > a, ab - c > ac - b)$

$bc - a, b, c$, $(bc - a < ab - c, bc - a < ac - b)$

Therefore, the leading terms of the four solutions are distinct. □

1.9 n=4

Theorem 28. *Solution to equation (1) does not exist when $n = 4$ and $k = 2$.*

Proof. Suppose x_1, x_2, x_3, x_4 is a solution when $n = 4$ and $k = 2$, then $x_1^2 + x_2^2 + x_3^2 + x_4^2 = 2x_1x_2x_3x_4$.

Case 1: x_1, x_2, x_3, x_4 are all odd.

Let $x_1 = 2y_1 + 1, x_2 = 2y_2 + 1, x_3 = 2y_3 + 1, x_4 = 2y_4 + 1$.

$$(2y_1 + 1)^2 + (2y_2 + 1)^2 + (2y_3 + 1)^2 + (2y_4 + 1)^2 = 2(2y_1 + 1)(2y_2 + 1)(2y_3 + 1)(2y_4 + 1)$$

$$\text{Left side} = (2y_1 + 1)^2 + (2y_2 + 1)^2 + (2y_3 + 1)^2 + (2y_4 + 1)^2 = 0 \pmod{4}$$

$$\text{Right side} = 2(2y_1 + 1)(2y_2 + 1)(2y_3 + 1)(2y_4 + 1) = 2 \pmod{4}$$

Case 2: x_1, x_2, x_3, x_4 are all even.

$$\text{Let } x_1 = 2y_1, x_2 = 2y_2, x_3 = 2y_3, x_4 = 2y_4.$$

$$(2y_1)^2 + (2y_2)^2 + (2y_3)^2 + (2y_4)^2 = 2(2y_1)(2y_2)(2y_3)(2y_4)$$

$$y_1^2 + y_2^2 + y_3^2 + y_4^2 = 8y_1y_2y_3y_4$$

Solution dose not exist.

Case 3: x_1, x_2, x_3, x_4 : two even and two odd.

$$\text{Let } x_1 = 2y_1, x_2 = 2y_2, x_3 = 2y_3 + 1, x_4 = 2y_4 + 1.$$

$$(2y_1)^2 + (2y_2)^2 + (2y_3 + 1)^2 + (2y_4 + 1)^2 = 2(2y_1)(2y_2)(2y_3 + 1)(2y_4 + 1)$$

$$\text{Left side} = 2 \pmod{4}$$

$$\text{Right side} = 0 \pmod{4}$$

Case 4: x_1, x_2, x_3, x_4 : three even and one odd.

$$\text{Let } x_1 = 2y_1, x_2 = 2y_2, x_3 = 2y_3, x_4 = 2y_4 + 1.$$

$$(2y_1)^2 + (2y_2)^2 + (2y_3)^2 + (2y_4 + 1)^2 = 2(2y_1)(2y_2)(2y_3)(2y_4 + 1)$$

$$\text{Left side} = 1 \pmod{4}$$

$$\text{Right side} = 0 \pmod{4}$$

Case 5: x_1, x_2, x_3, x_4 : one even and three odd.

$$\text{Let } x_1 = 2y_1, x_2 = 2y_2 + 1, x_3 = 2y_3 + 1, x_4 = 2y_4 + 1.$$

$$(2y_1)^2 + (2y_2 + 1)^2 + (2y_3 + 1)^2 + (2y_4 + 1)^2 = 2(2y_1)(2y_2 + 1)(2y_3 + 1)(2y_4 + 1)$$

$$\text{Left} = 3 \pmod{4}$$

$$\text{Right} = 0 \pmod{4}$$

Therefore, there is no solution when $n = 4$ and $k = 2$.

□

Theorem 29. *There is a solution with all components even for equation (1) if and only if $n = 4$ and $k = 1$.*

Proof. When $n = 4$ and $k = 1$, $x_1 = x_2 = x_3 = x_4 = 2$ is a solution of

$$x_1^2 + x_2^2 + x_3^2 + x_4^2 = x_1 x_2 x_3 x_4.$$

Case 1: $n = 3$

We will see in a following theorem that solution does not exist when $n = 3$ and $k = 2$.

Assume $x_1 = 2y_1, x_2 = 2y_2, x_3 = 2y_3$ is a solution of $x_1^2 + x_2^2 + x_3^2 = kx_1 x_2 x_3$. Then $y_1^2 + y_2^2 + y_3^2 = 2ky_1 y_2 y_3$. There is no solution for $k > 1$. For $k = 1$, $y_1^2 + y_2^2 + y_3^2 = 2y_1 y_2 y_3$ does not have solution.

Case 2: $n \geq 4$

Let $x_i = 2y_i$ for $i = 1, 2, \dots, n$, be a solution to equation (1). Then

$$4(y_1^2 + y_2^2 + \dots + y_n^2) = 2^n k y_1 y_2 \dots y_n$$

$$y_1^2 + y_2^2 + \dots + y_n^2 = k 2^{n-2} y_1 y_2 \dots y_n$$

We define a function $f(n) = k 2^{n-2} - n$.

We can see $f(3) = 2k - 3, f(4) = 4k - 4, f(5) = 8k - 5$.

We want to show that $f(n) > 0$ for $n > 4$.

It is clear that $\frac{d}{dn}(k 2^{n-2} - n) = \frac{1}{4} k 2^n \ln 2 - 1 > 0$ for $n > 2$.

Thus $f(n)$ is an increasing function when $n > 2$.

When $n \geq 4$, $\min(f(n)) = 0$ iff $n = 4$ and $k = 1$.

$f(n) > 0$ when $n \geq 4$ and $k \geq 1$ except $n = 4$ and $k = 1$.

Therefore,

$y_1^2 + y_2^2 + \dots + y_n^2 = k2^{n-2}y_1y_2\dots y_n$ has no solution when $n > 4$.

Thus the theorem holds. \square

Theorem 30. *Solution does not exist when $n = 4$ and $k = 3$.*

Proof. Suppose x_1, x_2, x_3, x_4 is a solution when $n = 4$ and $k = 3$, then $x_1^2 + x_2^2 + x_3^2 + x_4^2 = 3x_1x_2x_3x_4$.

Case 1: x_1, x_2, x_3, x_4 : all odd.

Let $x_1 = 2y_1 + 1, x_2 = 2y_2 + 1, x_3 = 2y_3 + 1, x_4 = 2y_4 + 1$.

$$(2y_1 + 1)^2 + (2y_2 + 1)^2 + (2y_3 + 1)^2 + (2y_4 + 1)^2 = 3(2y_1 + 1)(2y_2 + 1)(2y_3 + 1)(2y_4 + 1)$$

Left side $\equiv 0 \pmod{2}$

Right side $\equiv 1 \pmod{2}$

Case 2: x_1, x_2, x_3, x_4 : all even.

Let $x_1 = 2y_1, x_2 = 2y_2, x_3 = 2y_3, x_4 = 2y_4$.

$$(2y_1)^2 + (2y_2)^2 + (2y_3)^2 + (2y_4)^2 = 3(2y_1)(2y_2)(2y_3)(2y_4)$$

$$y_1^2 + y_2^2 + y_3^2 + y_4^2 = 12y_1y_2y_3y_4$$

Therefore, solution does not exist.

Case 3: x_1, x_2, x_3, x_4 : two even and two odd

Let $x_1 = 2y_1, x_2 = 2y_2, x_3 = 2y_3 + 1, x_4 = 2y_4 + 1$.

$$(2y_1)^2 + (2y_2)^2 + (2y_3 + 1)^2 + (2y_4 + 1)^2 = 3(2y_1)(2y_2)(2y_3 + 1)(2y_4 + 1)$$

Left side $\equiv 2 \pmod{4}$

Right side $\equiv 0 \pmod{4}$

Case 4: x_1, x_2, x_3, x_4 : three even and one odd.

Let $x_1 = 2y_1, x_2 = 2y_2, x_3 = 2y_3, x_4 = 2y_4 + 1$.

$$(2y_1)^2 + (2y_2)^2 + (2y_3)^2 + (2y_4 + 1)^2 = 3(2y_1)(2y_2)(2y_3)(2y_4 + 1)$$

Left side $\equiv 1 \pmod{4}$

Right side $\equiv 0 \pmod{4}$

Case 5: x_1, x_2, x_3, x_4 : one even and three odd.

Let $x_1 = 2y_1, x_2 = 2y_2 + 1, x_3 = 2y_3 + 1, x_4 = 2y_4 + 1$.

$$(2y_1)^2 + (2y_2 + 1)^2 + (2y_3 + 1)^2 + (2y_4 + 1)^2 = 3(2y_1)(2y_2 + 1)(2y_3 + 1)(2y_4 + 1)$$

Left side $\equiv 1 \pmod{2}$

Right side $\equiv 0 \pmod{2}$

Now we can conclude: there is no solution when $n = 4$ and $k = 3$. \square

Theorem 31. x_1, x_2, x_3, x_4 is a solution of equation (1) when $n = 4$ and $k = 4$ if and only if $2x_1, 2x_2, 2x_3, 2x_4$ is a solution of equation (1) when $n = 4$ and $k = 1$.

Proof. Let x_1, x_2, x_3, x_4 be a solution when $n = 4$ and $k = 4$, then $x_1^2 + x_2^2 + x_3^2 + x_4^2 = 4x_1x_2x_3x_4$.

$$\text{Then } 4x_1^2 + 4x_2^2 + 4x_3^2 + 4x_4^2 = 16x_1x_2x_3x_4.$$

$$(2x_1)^2 + (2x_2)^2 + (2x_3)^2 + (2x_4)^2 = (2x_1)(2x_2)(2x_3)(2x_4).$$

Therefore, $2x_1, 2x_2, 2x_3, 2x_4$ is a solution when $n = 4$ and $k = 1$.

Now let $2x_1, 2x_2, 2x_3, 2x_4$ be a solution when $n = 4$ and $k = 1$. Then $(2x_1)^2 + (2x_2)^2 + (2x_3)^2 + (2x_4)^2 = (2x_1)(2x_2)(2x_3)(2x_4)$.

$$\text{Then } x_1^2 + x_2^2 + x_3^2 + x_4^2 = 4x_1x_2x_3x_4.$$

Therefore, x_1, x_2, x_3, x_4 is a solution when $n = 4$ and $k = 4$. \square

Theorem 32. *If $x_1^2 + x_2^2 + x_3^2 + 1 = kx_1x_2x_3$ has integer solution, then $k = 4$.*

Proof. We give a straightforward proof.

It is easy to see that $k < 5$.

Next we want to show that x_1, x_2, x_3 are all odd numbers.

Consider $x_1^2 + x_2^2 + x_3^2 + 1 = kx_1x_2x_3$.

Case 1: x_1, x_2, x_3 : all even

Left side= 1 mod 4.

Right side= 0 mod 4.

Case 2: x_1, x_2, x_3 : two even, one odd

Left side= 2 mod 4.

Right side= 0 mod 4.

Case 3: x_1, x_2, x_3 : one even, two odd

Left side= 1 mod 2.

Right side= 0 mod 2.

Since the solution exists, x_1, x_2, x_3 are all odd numbers.

Then $x_1^2 + x_2^2 + x_3^2 + 1 = kx_1x_2x_3 = 0 \text{ mod } 4$.

Therefore, $k = 4$. □

Theorem 33. *$x_1^2 + x_2^2 + 1 + 1 = 4x_1x_2$ has infinite solutions.*

Proof. $x_1 = 1, x_2 = 1$ is solution of $x_1^2 + x_2^2 + 1 + 1 = 4x_1x_2$.

Let $x_1 = a, x_2 = b$ with $a \geq b$ is a solution, then $a^2 + b^2 + 1 + 1 = 4ab$.

Consider $x_1 = 4a - b, x_2 = a$,

$$(4a - b)^2 + a^2 + 1 + 1 - 4(4a - b)a = a^2 + b^2 + 1 + 1 - 4ab = 0$$

Also, $x_1 = 4a - b > 0$, $x_2 = a > b$.

Therefore, $x_1 = 4a - b$, $x_2 = a$ is another solution of $x_1^2 + x_2^2 + 1 + 1 = 4x_1x_2$.

Therefore, there are infinite solutions. \square

$$1.10 \quad n=5$$

Lemma 34. *If $x \geq a > 0$, $y \geq b > 0$, x, y, a and b are integers, then $xy \geq bx + ay - ab$.*

Proof. $xy - (bx + ay - ab)$

$$= x(y - b) - a(y - b)$$

$$= (x - a)(y - b) \geq 0. \quad \square$$

Lemma 35. *If x_1, x_2, \dots, x_n is a solution of $x_1^2 + x_2^2 + \dots + x_n^2 = kx_1x_2\dots x_n$ with $x_1 \leq x_2 \leq \dots \leq x_n$ (at least one " $<$ ") then*

$$x_n > \frac{k}{n}x_1x_2\dots x_{n-1}.$$

Proof. From $x_1 \leq x_2 \leq \dots \leq x_n$ (at least one " $<$ ") and $x_1^2 + x_2^2 + \dots + x_n^2 = kx_1x_2\dots x_n$, we get

$$nx_n^2 > kx_1x_2\dots x_n$$

$$x_n > \frac{k}{n}x_1x_2\dots x_{n-1}. \quad \square$$

Theorem 36. *Solution to equation (1) does not exist when $n = 5$ and $k = 2$.*

Proof. Suppose x_1, x_2, x_3, x_4, x_5 is a solution when $n = 5$ and $k = 2$, then $x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 = 2x_1x_2x_3x_4x_5$.

Then the number of odd numbers in x_1, x_2, x_3, x_4, x_5 must be even.

Case 1: x_1, x_2, x_3, x_4, x_5 are all even.

Let $x_1 = 2y_1, x_2 = 2y_2, x_3 = 2y_3, x_4 = 2y_4, x_5 = 2y_5$.

$$(2y_1)^2 + (2y_2)^2 + (2y_3)^2 + (2y_4)^2 + (2y_5)^2 = 2(2y_1)(2y_2)(2y_3)(2y_4)(2y_5)$$

$$y_1^2 + y_2^2 + y_3^2 + y_4^2 + y_5^2 = 16y_1y_2y_3y_4y_5$$

Solution does not exist.

Case 2: x_1, x_2, x_3, x_4, x_5 : four even and one odd

No solution.

Case 3: x_1, x_2, x_3, x_4, x_5 : three even and two odd.

Let $x_1 = 2y_1, x_2 = 2y_2, x_3 = 2y_3, x_4 = 2y_4 + 1, x_5 = 2y_5 + 1$.

$$(2y_1)^2 + (2y_2)^2 + (2y_3)^2 + (2y_4 + 1)^2 + (2y_5 + 1)^2 = 2(2y_1)(2y_2)(2y_3)(2y_4 + 1)(2y_5 + 1)$$

Left side $\equiv 2 \pmod{4}$

Right side $\equiv 0 \pmod{4}$

Case 4: x_1, x_2, x_3, x_4, x_5 : two even and three odd

No solution.

Case 5: x_1, x_2, x_3, x_4, x_5 : one even and four odd.

Case 5-1: $x_1 = x_2 = x_3 = x_4 \neq x_5$

Then $4x_1^2 + x_5^2 = 2x_1^4x_5$.

Let $x_5 = 2y$.

$$x_1^2 + y^2 = x_1^4y.$$

Then $y^2 \equiv 0 \pmod{x_1^2}, y \equiv 0 \pmod{x_1}$.

We let $y = x_1q$ ($q > 1$).

Then $x_1^2 + x_1^2q^2 = x_1^5q$.

$$1 + q^2 = x_1^3q$$

$$x_1^3 = q + \frac{1}{q}$$

Therefore, solution does not exist in this case.

Case 5-2: $x_1 = x_2 = x_3 = 1$

Then $3 + x_4^2 + x_5^2 = 2x_4x_5$.

$$(x_4 - x_5)^2 = -3.$$

Therefore, solution does not exist in this case.

Case 5-3: $x_1 \leq x_2 \leq x_3 \leq x_4 \leq x_5$ (at least one " $<$ ")

It is easy to see that $x_1 = 1, x_2 = 1, x_3 = 2, x_4 = 2, x_5$ is not a solution.

Let $x_1 \geq 1, x_2 \geq 1, x_3 \geq 2, x_4 \geq 3$.

From a previous lemma: $2x_5 > \frac{4}{5}x_1x_2x_3x_4$.

From $x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 = 2x_1x_2x_3x_4x_5$, we get

$$2x_5 = 2x_1x_2x_3x_4 \pm \sqrt{4x_1^2x_2^2x_3^2x_4^2 - 4(x_1^2 + x_2^2 + x_3^2 + x_4^2)}$$

If $2x_5 = 2x_1x_2x_3x_4 - \sqrt{4x_1^2x_2^2x_3^2x_4^2 - 4(x_1^2 + x_2^2 + x_3^2 + x_4^2)}$, then

$$2x_1x_2x_3x_4 - \sqrt{4x_1^2x_2^2x_3^2x_4^2 - 4(x_1^2 + x_2^2 + x_3^2 + x_4^2)} > \frac{4}{5}x_1x_2x_3x_4,$$

$$\frac{6}{5}x_1x_2x_3x_4 > \sqrt{4x_1^2x_2^2x_3^2x_4^2 - 4(x_1^2 + x_2^2 + x_3^2 + x_4^2)}$$

$$\frac{36}{25}x_1^2x_2^2x_3^2x_4^2 > 4x_1^2x_2^2x_3^2x_4^2 - 4(x_1^2 + x_2^2 + x_3^2 + x_4^2)$$

$$25(x_1^2 + x_2^2 + x_3^2 + x_4^2) > 16x_1^2x_2^2x_3^2x_4^2$$

Remember: $x_1 \geq 1, x_2 \geq 1, x_3 \geq 2, x_4 \geq 3$.

Then $x_1^2x_3^2 \geq 4, x_2^2x_4^2 \geq 9$.

$$x_1^2x_2^2x_3^2x_4^2 \geq 9x_1^2x_3^2 + 4x_2^2x_4^2 - 36 \text{ (using a previous lemma)}$$

$$\text{Then } x_1^2x_2^2x_3^2x_4^2 \geq 36x_1^2 + 9x_3^2 + 36x_2^2 + 4x_4^2 - 36 \times 3$$

$$= 2(x_1^2 + x_2^2 + x_3^2 + x_4^2) + 34x_1^2 + 3x_2^2 + 7x_3^2 + 2x_4^2 - 36 \times 3$$

$$\geq 2(x_1^2 + x_2^2 + x_3^2 + x_4^2) + 34 + 34 + 28 + 18 - 36 \times 3$$

$$> 2(x_1^2 + x_2^2 + x_3^2 + x_4^2)$$

$$\text{Then } 16x_1^2x_2^2x_3^2x_4^2 > 32(x_1^2 + x_2^2 + x_3^2 + x_4^2) > 25(x_1^2 + x_2^2 + x_3^2 + x_4^2)$$

Contradiction!

$$\text{If } 2x_5 = 2x_1x_2x_3x_4 + \sqrt{4x_1^2x_2^2x_3^2x_4^2 - 4(x_1^2 + x_2^2 + x_3^2 + x_4^2)}$$

$$2x_5 > 2x_1x_2x_3x_4$$

$$2x_1x_2x_3x_4 - x_5 < x_5$$

Remember $x_1 \leq x_2 \leq x_3 \leq x_4 \leq x_5$

If x_1, x_2, x_3, x_4, x_5 is a solution, then

$$2x_1x_2x_3x_4 - x_5 < x_5$$

Then "the sum of $x_1, \dots, x_4 - x_5$ " is smaller than "the sum of x_1, \dots, x_5 ".

We use this transformation finite times, we get

$$x_1 = \dots = x_5, \text{ or } x_1 = \dots = x_4 \neq x_5.$$

However, none of them is a solution.

Therefore, solution does not exist for the equation when $n = 5$ and $k = 2$. \square

Theorem 37. *Solution does not exist when $n = 5$ and $k = 3$.*

Proof. Suppose x_1, x_2, x_3, x_4, x_5 is a solution when $n = 5$ and $k = 3$, then $x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 = 3x_1x_2x_3x_4x_5$.

Case 1: x_1, x_2, x_3, x_4, x_5 are all even.

$$\text{Let } x_1 = 2y_1, x_2 = 2y_2, x_3 = 2y_3, x_4 = 2y_4, x_5 = 2y_5.$$

$$(2y_1)^2 + (2y_2)^2 + (2y_3)^2 + (2y_4)^2 + (2y_5)^2 = 3(2y_1)(2y_2)(2y_3)(2y_4)(2y_5)$$

$$y_1^2 + y_2^2 + y_3^2 + y_4^2 + y_5^2 = 24y_1y_2y_3y_4y_5$$

Solution does not exist.

Case 2: x_1, x_2, x_3, x_4, x_5 : four even and one odd.

Let $x_1 = 2y_1, x_2 = 2y_2, x_3 = 2y_3, x_4 = 2y_4, x_5 = 2y_5 + 1$.

$$(2y_1)^2 + (2y_2)^2 + (2y_3)^2 + (2y_4)^2 + (2y_5 + 1)^2 = 3(2y_1)(2y_2)(2y_3)(2y_4)(2y_5 + 1)$$

Left side $\equiv 1 \pmod{2}$

Right side $\equiv 0 \pmod{2}$

Case 3: x_1, x_2, x_3, x_4, x_5 : three even and two odd.

Let $x_1 = 2y_1, x_2 = 2y_2, x_3 = 2y_3, x_4 = 2y_4 + 1, x_5 = 2y_5 + 1$.

$$(2y_1)^2 + (2y_2)^2 + (2y_3)^2 + (2y_4 + 1)^2 + (2y_5 + 1)^2 = 3(2y_1)(2y_2)(2y_3)(2y_4 + 1)(2y_5 + 1)$$

Left side $\equiv 2 \pmod{4}$

Right side $\equiv 0 \pmod{4}$

Case 4: x_1, x_2, x_3, x_4, x_5 : two even and three odd.

Let $x_1 = 2y_1, x_2 = 2y_2, x_3 = 2y_3 + 1, x_4 = 2y_4 + 1, x_5 = 2y_5 + 1$.

$$(2y_1)^2 + (2y_2)^2 + (2y_3 + 1)^2 + (2y_4 + 1)^2 + (2y_5 + 1)^2 = 3(2y_1)(2y_2)(2y_3 + 1)(2y_4 + 1)(2y_5 + 1)$$

Left side $\equiv 1 \pmod{2}$

Right side $\equiv 0 \pmod{2}$

Case 5:

From the above, we know that the possible solution should be: x_1, x_2, x_3, x_4, x_5 : one even and four odd; or five odd.

Case 5-1: $x_1 = x_2 = x_3 = x_4 = x_5$

No solution.

Case 5-2: $x_1 = x_2 = x_3 = x_4 \neq x_5$

$$\text{Then } 4x_1^2 + x_5^2 = 3x_1^4 x_5$$

$$x_5^2 \pmod{x_1^2} = 0, x_5 \pmod{x_1} = 0.$$

Let $x_5 = x_1q$.

$$4x_1^2 + x_1^2q^2 = 3x_1^5q$$

$$4 + q^2 = 3x_1^3q$$

$$3x_1^3 = q + \frac{4}{q}, (q \neq 1)$$

If $q = 2$, then $3x_1^3 = 4$. (No solution.)

If $q = 4$, $3x_1^3 = 5$. (No solution.)

Thus solution does not exist for this case.

Case 5-3: $x_1 \leq x_2 \leq x_3 \leq x_4 \leq x_5$ with $x_1 \geq 1, x_2 \geq 1, x_3 \geq 1, x_4 \geq 3$

Case 5-3-1 $x_1 = x_2 = x_3 = 1 \neq x_4$

Then $3 + x_4^2 + x_5^2 = 3x_4x_5$.

$$x_4^2 + x_5^2 = 3x_4x_5 - 3$$

$$x_4^2 = 0 \text{ or } 1; x_5^2 = 0 \text{ or } 1; 3x_4x_5 - 3 = 0 \pmod{3}$$

Then $x_4^2 = 0; x_5^2 = 0 \pmod{3}$

Let $x_4^2 = 3k_1, x_5^2 = 3k_2$.

$$9k_1^2 + 9k_2^2 = 27k_1k_2 - 3$$

$$3k_1^2 + 3k_2^2 = 9k_1k_2 - 1$$

$$3k_1^2 + 3k_2^2 \neq 9k_1k_2 - 1 \pmod{3}$$

Therefore, $x_1 = x_2 = x_3 = 1 \neq x_4, x_5$ is not a solution.

Case 5-3-2 $x_1 \geq 1, x_2 \geq 1, x_3 \geq 3, x_4 \geq 3$

$$\text{Then } 2x_5 = 3x_1x_2x_3x_4 \pm \sqrt{9x_1^2x_2^2x_3^2x_4^2 - 4(x_1^2 + x_2^2 + x_3^2 + x_4^2)}$$

$2x_5 > \frac{6}{5}x_1x_2x_3x_4$ from a previous lemma.

Case 5-3-2-1

If $2x_5 = 3x_1x_2x_3x_4 - \sqrt{9x_1^2x_2^2x_3^2x_4^2 - 4(x_1^2 + x_2^2 + x_3^2 + x_4^2)}$, then

$$2x_5 = 3x_1x_2x_3x_4 - \sqrt{9x_1^2x_2^2x_3^2x_4^2 - 4(x_1^2 + x_2^2 + x_3^2 + x_4^2)}$$

$$3x_1x_2x_3x_4 - \sqrt{9x_1^2x_2^2x_3^2x_4^2 - 4(x_1^2 + x_2^2 + x_3^2 + x_4^2)} > \frac{6}{5}x_1x_2x_3x_4$$

$$\frac{9}{5}x_1x_2x_3x_4 > \sqrt{9x_1^2x_2^2x_3^2x_4^2 - 4(x_1^2 + x_2^2 + x_3^2 + x_4^2)}$$

$$\frac{81}{25}x_1^2x_2^2x_3^2x_4^2 > 9x_1^2x_2^2x_3^2x_4^2 - 4(x_1^2 + x_2^2 + x_3^2 + x_4^2)$$

$$81x_1^2x_2^2x_3^2x_4^2 > 25 \times 9x_1^2x_2^2x_3^2x_4^2 - 25 \times 4(x_1^2 + x_2^2 + x_3^2 + x_4^2)$$

$$100(x_1^2 + x_2^2 + x_3^2 + x_4^2) > 144x_1^2x_2^2x_3^2x_4^2$$

$$100(x_1^2 + x_2^2 + x_3^2 + x_4^2) > 144x_1^2x_2^2x_3^2x_4^2 \geq 144(x_1^2 + x_2^2 + x_3^2 + x_4^2)$$

Contradiction.

Case 5-3-2-2

If $2x_5 = 3x_1x_2x_3x_4 + 3\sqrt{9x_1^2x_2^2x_3^2x_4^2 - 4(x_1^2 + x_2^2 + x_3^2 + x_4^2)}$, then

Remember: $x_1 \geq 1, x_2 \geq 1, x_3 \geq 3, x_4 \geq 3$.

Then $x_1^2x_3^2 \geq 9, x_2^2x_4^2 \geq 9$

$$x_1^2x_2^2x_3^2x_4^2 \geq 9x_1^2x_3^2 + 9x_2^2x_4^2 - 81 \text{ (using a previous lemma)}$$

$$\text{Then } x_1^2x_2^2x_3^2x_4^2 \geq 81x_1^2 + 9x_3^2 + 81x_2^2 + 9x_4^2 - 81 \times 3$$

$$= 4(x_1^2 + x_2^2 + x_3^2 + x_4^2) + 77x_1^2 + 77x_2^2 + 5x_3^2 + 5x_4^2 - 36 \times 3$$

$$\geq 4(x_1^2 + x_2^2 + x_3^2 + x_4^2) + 77 + 77 + 5 + 5 - 36 \times 3$$

$$> 4(x_1^2 + x_2^2 + x_3^2 + x_4^2)$$

$$2x_5 = 3x_1x_2x_3x_4 + 3\sqrt{9x_1^2x_2^2x_3^2x_4^2 - 4(x_1^2 + x_2^2 + x_3^2 + x_4^2)} > \frac{6}{5}x_1x_2x_3x_4$$

$$(2x_5 - 3x_1x_2x_3x_4) = 3\sqrt{9x_1^2x_2^2x_3^2x_4^2 - 4(x_1^2 + x_2^2 + x_3^2 + x_4^2)}$$

$$4x_5^2 - 12x_5x_1x_2x_3x_4 + 9x_1^2x_2^2x_3^2x_4^2 = 81x_1^2x_2^2x_3^2x_4^2 - 36(x_1^2 + x_2^2 + x_3^2 + x_4^2)$$

$$4x_5^2 + 36(x_1^2 + x_2^2 + x_3^2 + x_4^2) = 72x_1^2x_2^2x_3^2x_4^2 + 12x_5x_1x_2x_3x_4$$

$$x_5^2 + 9(x_1^2 + x_2^2 + x_3^2 + x_4^2) = 18x_1^2x_2^2x_3^2x_4^2 + 3x_5x_1x_2x_3x_4 = 18x_1^2x_2^2x_3^2x_4^2 + x_1^2 + x_2^2 +$$

$$x_3^2 + x_4^2 + x_5^2$$

$$8(x_1^2 + x_2^2 + x_3^2 + x_4^2) = 18x_1^2x_2^2x_3^2x_4^2 > 72(x_1^2 + x_2^2 + x_3^2 + x_4^2)$$

Contradiction!

Therefore, solution does not exist in this case.

Therefore, there is no solution when $n = 5$ and $k = 3$. □

Theorem 38. *When $n = 5$ and $k = 1$, equation (1) has no solution of the form $x_1, x_2, 1, 1, 1$. There are infinite solutions of the form $x_1, x_2, 3, 1, 1$.*

Proof. Assume $a, b, 1, 1, 1$ is a solution with $a \geq b$.

$$\text{Then } a^2 + b^2 + 1 + 1 + 1 = ab.$$

$$\text{Then } a = \frac{1}{2}b + \frac{1}{2}\sqrt{-3b^2 - 12}.$$

Therefore, there is no solution of the form $x_1, x_2, 1, 1, 1$.

Let $a, b, 3, 1, 1$ be a solution with $a \geq b \geq 3$. Then

$$a^2 + b^2 + 3^2 + 1 + 1 - 3ab = 0$$

$$\text{Then } (3a - b)^2 + a^2 + 3^2 + 1 + 1 - 3(3a - b)a = a^2 - 3ab + b^2 + 11 = 0$$

Thus, $3a - b, a, 3, 1, 1$ is another solution with $3a - b > a$.

Therefore, there are infinite solutions of the form $x_1, x_2, 3, 1, 1$. □

1.11 n=6

Theorem 39. *If equation (1) has a solution when $n = 6$ and $k = 1$, then the 6 solution components must be 2 even and 4 odd.*

Proof. Suppose that $x_1, x_2, x_3, x_4, x_5, x_6$ is a solution.

Case 1: $x_1, x_2, x_3, x_4, x_5, x_6$: all even: No solution

Case 2: $x_1, x_2, x_3, x_4, x_5, x_6$: five even and one odd

Let $x_1 = 2y_1, x_2 = 2y_2, x_3 = 2y_3, x_4 = 2y_4, x_5 = 2y_5, x_6 = 2y_6 + 1$.

$$(2y_1)^2 + (2y_2)^2 + (2y_3)^2 + (2y_4)^2 + (2y_5)^2 + (2y_6 + 1)^2 = (2y_1)(2y_2)(2y_3)(2y_4)(2y_5)(2y_6 +$$

1)

Left side $\equiv 1 \pmod{4}$

Right side $\equiv 0 \pmod{4}$

Case 3: $x_1, x_2, x_3, x_4, x_5, x_6$: four even and two odd.

Let $x_1 = 2y_1, x_2 = 2y_2, x_3 = 2y_3, x_4 = 2y_4, x_5 = 2y_5 + 1, x_6 = 2y_6 + 1$.

$$(2y_1)^2 + (2y_2)^2 + (2y_3)^2 + (2y_4)^2 + (2y_5 + 1)^2 + (2y_6 + 1)^2 = (2y_1)(2y_2)(2y_3)(2y_4)(2y_5 +$$

1)(2y_6 + 1)

Left side $\equiv 2 \pmod{4}$

Right side $\equiv 0 \pmod{4}$

Case 4: $x_1, x_2, x_3, x_4, x_5, x_6$: three even and three odd.

Let $x_1 = 2y_1, x_2 = 2y_2, x_3 = 2y_3, x_4 = 2y_4 + 1, x_5 = 2y_5 + 1, x_6 = 2y_6 + 1$.

$$(2y_1)^2 + (2y_2)^2 + (2y_3)^2 + (2y_4 + 1)^2 + (2y_5 + 1)^2 + (2y_6 + 1)^2 = (2y_1)(2y_2)(2y_3)(2y_4 +$$

1)(2y_5 + 1)(2y_6 + 1)

Left side $\equiv 3 \pmod{4}$

Right side $\equiv 0 \pmod{4}$

Case 5: $x_1, x_2, x_3, x_4, x_5, x_6$: one even and five odd.

Let $x_1 = 2y_1, x_2 = 2y_2 + 1, x_3 = 2y_3 + 1, x_4 = 2y_4 + 1, x_5 = 2y_5 + 1, x_6 = 2y_6 + 1$.

$$(2y_1)^2 + (2y_2 + 1)^2 + (2y_3 + 1)^2 + (2y_4 + 1)^2 + (2y_5 + 1)^2 + (2y_6 + 1)^2 = (2y_1)(2y_2 +$$

1)(2y_3 + 1)(2y_4 + 1)(2y_5 + 1)(2y_6 + 1)

Left side $\equiv 1 \pmod{2}$

Right side $\equiv 0 \pmod{2}$

Case 6: $x_1, x_2, x_3, x_4, x_5, x_6$: all odd.

Let $x_1 = 2y_1 + 1, x_2 = 2y_2 + 1, x_3 = 2y_3 + 1, x_4 = 2y_4 + 1, x_5 = 2y_5 + 1, x_6 = 2y_6 + 1$.

Left side $\equiv 0 \pmod{2}$

Right side $\equiv 1 \pmod{2}$

Therefore, solution can not be any of the above six cases. \square

Theorem 40. $x_1 = x_2 = x_3 = x_4, x_5, x_6$ is not a solution to equation (1) when $n = 6$ and $k = 1$.

Proof. If $x_1 = x_2 = x_3 = x_4, x_5, x_6$ is a solutions, then

$$4x_1^2 + x_5^2 + x_6^2 = x_1^4 x_5 x_6$$

$$x_1^2 = \frac{4 \pm \sqrt{16 + 4x_5 x_6 (x_5^2 + x_6^2)}}{8x_5 x_6}$$

$$x_1^2 = \frac{2 + \sqrt{4 + x_5 x_6 (x_5^2 + x_6^2)}}{x_5 x_6} > x_5 x_6$$

$$x_5^2 + x_6^2 = x_1^2 (x_1^2 x_5 x_6 - 4) \geq x_5 x_6 (x_5 x_6 x_5 x_6 - 4) > x_5^2 + x_6^2$$

Contradiction! \square

Theorem 41. $x_1 = x_2 = x_3 = 1 \neq x_4, x_5 \leq x_6$ is not a solution to equation (1) when $n = 6$ and $k = 1$.

Proof. Suppose that $x_1 = x_2 = x_3 = 1 \neq x_4, x_5 \leq x_6$ is a solution.

If $x_1 = x_2 = x_3 = 1, x_4 = 2, x_5, x_6$ is a solution, then

$$3 + 4 + x_5^2 + x_6^2 = 2x_5 x_6, \text{ then } x_6 = x_5 \pm i\sqrt{7}. \text{ Contradiction.}$$

Without loss of generality, assume $10 \leq x_4, 10 \leq x_5 \leq x_6$

$$3 + x_4^2 + x_5^2 + x_6^2 = x_4 x_5 x_6, \text{ solution } x_6 = \frac{1}{2}x_4 x_5 \pm \frac{1}{2}\sqrt{x_4^2 x_5^2 - 12 - 4x_4^2 - 4x_5^2}$$

$$\text{If } x_6 = \frac{1}{2}x_4 x_5 - \frac{1}{2}\sqrt{x_4^2 x_5^2 - 12 - 4x_4^2 - 4x_5^2}$$

$$\frac{1}{2}x_4 x_5 - \frac{1}{2}\sqrt{x_4^2 x_5^2 - 12 - 4x_4^2 - 4x_5^2} \geq x_5$$

$$x_4x_5 - 2x_5 \geq \sqrt{x_4^2x_5^2 - 12 - 4x_4^2 - 4x_5^2}$$

$$x_4^2x_5^2 - 4x_4x_5^2 + 4x_5^2 \geq x_4^2x_5^2 - 12 - 4x_4^2 - 4x_5^2$$

$$12 + 4x_4^2 + 8x_5^2 \geq 4x_4x_5^2$$

$$3 + x_4^2 + 2x_5^2 \geq x_4x_5^2 \geq 16x_4 + 3x_5^2 - 48 > 3 + x_4^2 + 2x_5^2$$

$$3 + x_4^2 + 2x_5^2 \geq x_4x_5^2 \geq 100x_4 + 10x_5^2 - 1000 > 3 + x_4^2 + 2x_5^2$$

Contradiction. □

$$1.12 \quad n=7$$

Theorem 42. *If equation (1) has a solution for $n = 7$ and $k = 4$, then the seven solution components must be four odd and three even.*

Theorem 43. *If there is a solution for $n = 7$ and $k = 6$, then the seven solution components must be four odd and three even.*

It is easy to prove the above two theorems although there are many cases to be considered.

Theorem 44. *Given two positive integers n and k with $n \geq 2$, $k \leq n$, equation (1) has either 0 or infinite number of positive integer solutions.*

Proof.

$$x_1^2 + x_2^2 = kx_1x_2 \tag{2}$$

$k = 2$, (2) has infinite number of solutions.

$k = 1$, (2) has no solution.

$n \geq 3$, by an earlier theorem, $x_1 \geq x_2 \geq \dots \geq x_n$ is a solution, then

$$(kx_1x_2\dots x_{n-1} - x_n), x_1, x_2, \dots, x_{n-1} \text{ is also a solution.} \quad (3)$$

Claim:

$$kx_1x_2\dots x_{n-1} - x_n > x_1 \quad (4)$$

Case 1: $x_1 = x_2 = \dots = x_n = 1$

Then $k = n \geq 3$.

Thus $kx_1x_2\dots x_{n-1} - x_n = k - 1 \geq 2 \geq x_1$

Case 2: $x_1 = x_2 = \dots = x_n = c > 1$

$$nc^2 = kc^n, n = kc^{n-2}$$

Case 2-1: $n = 3$

Then $c = 3$ and $k = 1$. (4) holds.

Case 2-2: $n = 2$

Then $k = 1$ and $c = 2$. (4) holds.

Case 2-3: $n \geq 5$

Then $n/k = c^{n-2}$. No $c > 1$ possible.

Case 3: $x_1 > x_n$

Case 3-1 $x_2 \geq 2$

$$kx_1x_2\dots x_{n-1} - x_n \geq 2x_1 - x_n > x_1$$

Case 3-2 $k \geq 2$

$$kx_1x_2\dots x_{n-1} - x_n \geq 2x_1 - x_n > x_1$$

Case 3-3 $x_2 = 1$ and $k = 1$

$$x_1^2 + (n - 1) = x_1$$

$$x_1 = \frac{1}{2} \pm \frac{1}{2} \sqrt{5 - 4n}$$

Equation (1) has no solution.

So (4) holds.

Repeately using (3), we can generate unlimited number of solutions. \square

1.13 Sequences

Given n & k , the leading term of the ordered solutions of equation (1) in non-decreasing order form an integer sequence if the solutions exist. So far we have not found same leading term from two different ordered solutions for a given (n, k) . Some of the sequences below can be found at [36] .

Sequence for $n=3$ and $k=3$:

1,2,5,13,29,34,89,169,194,233,433,610,985,1325,1597,2897,4181,...

The number of distinct prime divisors of any of the first 93 terms is less than 6. The second, third, 4th, 5th, 6th, 10th, 11th, 15th, 16th, 18th, 20th, 24th, 25th, 27th, 30th, 36th, 38th, 45th, 48th, 49th, 69th, 79th, 81th, 86th, 91th terms are primes.

This sequence is called Markoff numbers.

It seems that the prime Markoff numbers have density zero among all Markoff numbers. It might be also true that infinitely many Markoff numbers are prime. [9]

Sequence for $n=3$ and $k=1$:

3,6,15,39,87,102,267,507,582,699,1299,1830,2955,3975,4791,8691,...

This sequence is the previous sequence multiplied by three.

Sequence for $n=4$ and $k=1$:

2,6,22,82,262,306,1142,3122,3606,4262,11522,15906,34582, 37202, 50182, ...

Sequence for $n=4$ and $k=4$:

1,3,11,41,131,153,571,1561,1803,2131,5761,7953,17291, 18601, 25091, 29681, ...

Sequence for $n=5$ and $k=1$:

4,5,9,12,23,31,33,35,44,57,60,81,107,123,157,179,204, 212, 273, 293, 311, ...

Sequence for $n=5$ and $k=4$:

2,7,26,55,97,362,433,727,1351,1538,3079, 3409, 5042, 10087, 18817, 20330, ...

Sequence for $n=5$ and $k=5$:

1,4,19,91,379,436,2089,7561,8644,10009, 36001,47956,144019, 150841, 198379...

Sequence for $n=6$ and $k=3$:

2,4,10,11,23,26,64,68,119,131,134,178,274,373,466,551, 779, 781, 1220, 1418, ...

Sequence for $n=6$ and $k=6$:

1,5,29,169,869,985,5741,26041,29405,33461, 151201, 195025, 756029, 780361, ...

Sequence for $n=7$ and $k=1$:

3,5,10,18,23,37,39,58,67,119,138,178,181,250,274,307,338,359,515,551,738, ...

Sequence for $n=7$ and $k=2$:

2,6,15,22,47,82,118,239,262,306,370,527,929,1126,1142,2255,2913,3122,...

Sequence for $n=7$ and $k=3$:

3,7,17,18,47,62,99,123,151,305,322,377,551,577,843,1299,1342,2207,2537, ...

Sequence for $n=7$ and $k=5$:

2,3,9,14,43,67,89,206,209,321,881,987,1538,1934,3121,4003,4689, 4729, ...

Sequence for $n=7$ and $k=7$:

1,6,41,281,1721,1926,13201,72241,80646,90481,493921,620166,2963561,...

Sequence for $n=8$ and $k=1$:

4,14,31,52,110,194,223,244,494,724,866,991,1454,1921,3076,3554,6158,...

Sequence for $n=8$ and $k=8$:

1,7,55,433,3079,3409,26839,172369,190519,211303,1354753,1663585,9483319,...

Sequence for $n=9$ and $k=6$:

2,4,11,23,64,131,134,373,551,781,1561,2174,4223,4552,8644,12671,13201,17291,...

Sequence for $n=9$ and $k=9$:

1,8,71,631,5111,5608,49841,367921,403208,442961,3265921,3936808,26127431,...

1.14 Conjectures

We believe the following conjectures are most likely to be true.

Conjecture 45. *Given (n, k) , any ordered solution S of equation (1) can be generated from the minimum ordered solution M using the component replacement method in Theorem 4 in a finite number of steps. i. e. $M \mid^* S$ for any ordered solution S .*

That is, the minimum solution is the only fundamental solution. Since any \mid -step is reversible, $S \mid^* M \mid^* S'$. Any solution can be generated from any other solution in a finite number of the component replacement steps.

Conjecture 46. *If x_1, x_2, \dots, x_n and x_1', x_2', \dots, x_n' are two different ordered solutions of (1) then $x_1 \neq x_1'$. All our experiments suggest each ordered-solution has a unique x_1 for a given (n, k) .*

Conjecture 47. *For $(n, k) = (6, 1), (6, 2), (6, 4), (6, 5), (7, 4), (7, 6), (8, 2), (8, 3), \dots, (8, 7), (9, 1), \dots, (9, 5), (9, 7), (9, 8)$, equation (1) has no solutions.*

We didn't see any solutions with each variable value $\leq 1,000,000$ for the cases in this conjecture. It highly unlikely the minimum solution would contain a component $> 1,000,000$, we have not seen a minimum solution with leading term > 4 .

Conjecture 48. *The minimum solution of (1) consists of at least a 1 for $n \geq 5$.*

This conjecture is true for all the minimum solutions that we have found for $5 \leq n \leq 20$.

Conjecture 49. *The leading term in the minimum ordered solution of (1) is not greater than n for $n \geq 3$.*

The maximum leading term among all the minimum solutions of (1) that we have obtained is 5 for $3 \leq n \leq 20$.

We challenge the readers to prove or disprove our conjectures.

CHAPTER 2: ENUMERATION OF INTEGER MATRICES

2.1 Integer Sequence

We will talk about integer sequences since we have obtained many sequences in this dissertation. An integer sequence is a sequence (i.e., an ordered list) of integers. An integer sequence may be specified explicitly by giving a formula for its n th term, or implicitly by giving a relationship between its terms. For example, the sequence $0, 1, 1, 2, 3, 5, 8, 13, \dots$ (the Fibonacci sequence) is formed by starting with 0 and 1 and then adding any two consecutive terms to obtain the next one: an implicit description. The sequence $0, 3, 8, 15, \dots$ is formed according to the formula $n^2 - 1$ for the n th term: an explicit definition. Alternatively, an integer sequence may be defined by a property which members of the sequence possess and other integers do not possess [36]. An integer sequence is a computable sequence, if there exists an algorithm which given n , calculates a_n , for all $n > 0$. An integer sequence is a definable sequence, if there exists some statement $P(x)$ which is true for that integer sequence x and false for all other integer sequences. The set of computable integer sequences and definable integer sequences are both countable, with the computable sequences a proper subset of the definable sequences (in other words, some sequences are definable but not computable). The set of all integer sequences is uncountable (with cardinality equal to that of the continuum); thus, almost all integer sequences

are incomputable and cannot be defined.

Why does one integer follow another? What is the pattern? What rule or formula dictates the position of each integer? Most people think deeply about sequences only when confronted by one on a test, but for mathematicians, computer scientists, and others, sequences are part and parcel of their work. Today sequences are especially important in number theory, combinatorics, and discrete mathematics, but sequences have been known and wondered about even before the time of Pythagoras, who discovered an infinite sequence of integer trip (a, b, c) such that $a^2 + b^2 = c^2$. In medieval times, bell ringers relied on sequences to cycle through all possible combinations of bells. But no one in the intervening millennia had thought to compile sequences into a collection that could be referenced by others. Neil Sloane started collecting integer sequences as a graduate student in 1965 to support his work in combinatorics. The database was at first stored on punch cards. He published selections from the database in book form twice: [38] containing 2372 sequences in lexicographic order and assigned numbers from 1 to 2372. [34] containing 5488 sequences. These books were well received and, especially after the second publication, mathematicians supplied Sloane with a steady flow of new sequences. The collection became unmanageable in book form, and when the database had reached 16,000 entries Sloane decided to go online—first as an e-mail service (August 1994), and soon after as a web site (1996). As a spin-off from the database work, Sloane founded the Journal of Integer Sequences in 1998. The database continues to grow at a rate of some 10,000 entries a year. Sloane has personally managed 'his' sequences for almost 40 years, but starting in 2002, a board of associate editors and volunteers has helped maintain the database.

The On-Line Encyclopedia of Integer Sequences (OEIS), also cited simply as Sloane's, is an online database of integer sequences, created and maintained by N. J. A. Sloane, a researcher at AT&T Labs. OEIS records information on integer sequences of interest to both professional mathematicians and amateurs, and is widely cited. As of 25 September 2015 it contains over 260,000 sequences, making it the largest database of its kind. And 15,000 new entries are added each year. Each entry contains the leading terms of the sequence, keywords, mathematical motivations, literature links, and more, including the option to generate a graph or play a musical representation of the sequence. The database is searchable by keyword and by subsequence. [37], [2].

Sequences can come from anywhere. Computational fields not surprisingly generate a lot of sequences. Computer science, to a large extent based on discrete math, also makes use of sequences (e.g. number of steps to sort n things). While it makes sense that sequences appear in mathematics, they are all around. The Fibonacci sequence in particular appears in nature: the growth of branches, pinecone rows, sandollar, and the number petals in many flowers all relate to the Fibonacci sequence. The sequence appears in art and literature too. Sloane originally started the sequence collection as an aid to research so that anyone coming upon a sequence in their calculations could immediately get additional terms and maybe a formula. This use of the OEIS is more important than ever today, since many computer-related tasks can be stated in terms of a sequence: minimizing the number of steps needed to count a set of items, ranking a list of unsorted numbers from lowest to highest, even characterizing the behavior of a program or algorithm. As more applications today depend on ideas

and concepts taken from pure mathematics—cryptography, the use of graphs to study social networks, the ranking of search engine listings—sequences increasingly play a more direct role in solving real-world problems.

Sequence data is pervasive in our lives, and understanding sequence data is of grand importance. Much research has been conducted on sequence data mining in the last dozen years. Hundreds if not thousands of research papers have been published in forums of various disciplines, such as data mining, database systems, information retrieval, biology and bioinformatics, industrial engineering, etc. The area of sequence data mining has developed rapidly, producing a diversified array of concepts, techniques and algorithmic tools. [16]

There are many research topics on integer sequence. For example: (1) How to find a good formula for a sequence with a bad formula or no formula at all? Sometimes it is not very hard to find the first a few terms of a sequence by hand calculation. It might be very tough to find a formula. (2) How to find a good algorithm to compute more terms for a sequence if you could not get a formula? People have been working on some sequences for more than one hundred years. However, they still could not get the first one hundred terms, or even not the first thirty terms. (3) The applications and data structure of some sequences. (4) Find some new sequences.

In the following, we will concentrate on integer matrix enumeration and some conjectures which agree to the results obtained from our computation.

2.2 Zero-One Matrices with Fixed Row Sum and Column Sum

“Let $f(n)$ be the number of $n \times n$ matrices M of zeros and ones such that every row and column of M has exactly three ones, $f(0) = 1$, $f(1) = f(2) = 0$, $f(3) = 1$. The most explicit formula known at present for $f(n)$ is

$$f(n) = 6^{-n} \sum \frac{(-1)^\beta n!^2 (\beta + 3\gamma)! 2^\alpha 3^\beta}{\alpha! \beta! \gamma! 2^6 6^\gamma} \quad ((ii))$$

where the sum is over all $(n+2)(n+1)/2$ solutions to $\alpha + \beta + \gamma = n$ in nonnegative integers. This formula gives very little insight into the behavior of $f(n)$, but it does allow one to compute $f(n)$ faster than if only the combinatorial definition of $f(n)$ were used. Hence with some reluctance we accept (ii) as a “determination” of $f(n)$. Of course if someone were later to prove $f(n) = (n-1)(n-2)/2$ (rather unlikely), then our enthusiasm for (ii) would be considerably diminished.” [39], [21]

The enumeration of integer-matrices has been the subject of considerable study. It has been the subject of considerable study, and it is unlikely that a simple formula exists. The number in question can be related in various ways to the representation theory of the symmetric group or of the complex general linear group, but this does not make their computation any easier. We got and will prove a more efficient formula than (ii). Let $f(m, n, s, t)$ be the number of $(0, 1)$ - matrices of size $m \times n$ such that each row has exactly s ones and each column has exactly t ones ($sm = nt$). The determination of $f(m, n, s, t)$ is an unsolved problem, except for very small s, t .

In some row, let $x_{i_1} x_{i_2} \cdots x_{i_k}$ denote the $i_1 - th$ column, the $i_2 - th$ column, \cdots , the $i_k - th$ column entries are 1 in some row and other entries are all 0, where

$$i_1, i_2, \dots, i_k \in \{1, 2, \dots, n\}.$$

Example: Let $m = n = 4, s = t = 3$, then $x_1x_2x_3|x_1x_2x_4|x_1x_3x_4|x_2x_3x_4$ denotes the following matrix:

$$\begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{pmatrix}$$

Obviously, $f(m, n, s, t)$ equals the coefficient of $x_1^t x_2^t \cdots x_n^t$ in the symmetric polynomial

$$\left(\sum_{i_1 < i_2 < \cdots < i_s} x_{i_1} x_{i_2} \cdots x_{i_s} \right)^m$$

where $i_1, i_2, \dots, i_s \in \{1, 2, \dots, n\}$, and the sum is over all the possible of s - combinations from $\{1, 2, \dots, n\}$ with $i_1 < i_2 < \cdots < i_s$. It is easy to get,

$$f(m, n, s, t) = f(n, m, t, s), \quad (sm = tn)$$

$$f(m, n, s, t) = f(n, m, n - s, m - t), \quad (sm = tn)$$

$$f(m, n, 1, t) = \frac{m!}{(t!)^n} \quad (m = tn)$$

$$f(m, n, s, 1) = \frac{n!}{(s!)^m} \quad (sm = n)$$

We let:

$$\alpha_1 = x_1 + x_2 + \cdots + x_n;$$

$$\alpha_2 = x_1^2 + x_2^2 + \cdots + x_n^2;$$

.....

$$\alpha_k = x_1^k + x_2^k + \cdots + x_n^k;$$

.....

$$\sigma_1 = \sum_{i=0}^n x_i = x_1 + x_2 + \cdots + x_n = \alpha_1;$$

$$\sigma_2 = \sum_{i < j} x_i x_j \quad \text{where} \quad i, j \in \{1, 2, \cdots, n\};$$

$$\sigma_3 = \sum_{i < j < k} x_i x_j x_k \quad \text{where} \quad i, j, k \in \{1, 2, \cdots, n\};$$

.....

$$\sigma_k = \sum_{i_1 < i_2 < \cdots < i_k} x_{i_1} x_{i_2} \cdots x_{i_k} \quad \text{where} \quad i_1, i_2, \cdots, i_k \in \{1, 2, \cdots, n\};$$

Theorem 50. $f(n, n, 2, 2) = \frac{n!}{2^n} \sum_{r_0=0}^n \binom{n}{r_0} \frac{(-1)^{n-r_0} (2r_0)!}{2^{r_0} r_0!}$

$$= \frac{1}{4^n} \left((2n)! + \sum_{k=1}^n (-2)^k \binom{n}{k}^2 k! (2(n-k))! \right)$$

$$= 4^{-n} \sum_{i=0}^n \frac{(-2)^i (n!)^2 (2n-2i)!}{i! ((n-i)!)^2}$$

Proof. From $\sigma_2^n = \left(\frac{1}{2} (\alpha_1^2 - \alpha_2) \right)^n = 2^{-n} \sum_{i=0}^n \binom{n}{i} (-1)^i \alpha_2^i \alpha_1^{(2n-2i)}$

we know that the coefficient of $x_1^2 x_2^2 \cdots x_n^2$ is

$$\begin{aligned} & 2^{-n} \sum_{i=0}^n \binom{n}{i} (-1)^i \binom{n}{i} i! \underbrace{\binom{2n-2i}{2 \quad 2 \quad \cdots \quad 2}}_{n-i} \\ &= 2^{-n} \sum_{i=0}^n \frac{(-1)^i (n!)^2 (2n-2i)!}{i! ((n-i)!)^2 2^{n-i}} \\ &= 4^{-n} \sum_{i=0}^n \frac{(-2)^i (n!)^2 (2n-2i)!}{i! ((n-i)!)^2}. \end{aligned}$$

□

Example: $f(2, 2, 2, 2) = 1$, $f(3, 3, 2, 2) = 6$

Some numbers obtained are listed in the following.

1

6

90

2040

67950

3110940

187530840

14398171200

1371785398200

158815387962000

21959547410077200

3574340599104475200

676508133623135814000

147320988741542099484000

Theorem 51. $f(m, n, 2, 3) = 2^{-m} \sum_{i=0}^n \frac{(-1)^i m! n! (2m-2i)!}{i! (m-i)! (n-i)! 6^{n-i}}$

Proof. From $\sigma_2^m = \left(\frac{1}{2}(\alpha_1^2 - \alpha_2)\right)^m = 2^{-m} \sum_{i=0}^m \binom{m}{i} (-1)^i \alpha_2^i \alpha_1^{(2m-2i)}$

we know that the coefficient of $x_1^3 x_2^3 \cdots x_n^3$ is

$$2^{-m} \sum_{i=0}^n \binom{m}{i} (-1)^i \binom{n}{i} i! \binom{2m-2i}{\underbrace{1 \ 1 \ \cdots \ 1}_i \underbrace{3 \ 3 \ \cdots \ 3}_{n-i}}$$

$$= 2^{-m} \sum_{i=0}^n \frac{(-1)^i m! n! (2m-2i)!}{i! (m-i)! (n-i)! 6^{n-i}}.$$

□

Some numbers obtained are listed in the following.

1860

90291600

31082452632000

46764764308702440000

229747284991066934931840000

3031982831164890119435183865600000

93453554057243260025029337978773248000000

Theorem 52. $f(n, n, 3, 3) = 6^{-n} \sum_{\alpha=0}^n \sum_{\beta=0}^{n-\alpha} \frac{(-1)^\beta 2^\alpha 3^\beta (n!)^2 (3n-3\alpha-2\beta)!}{\alpha! \beta! (n-\alpha-\beta)! 2^6 (n-\alpha-\beta)^2}$

Proof. From

$$\begin{aligned} \sigma_3^n &= \left(\frac{1}{6} (2\alpha_3 - 3\alpha_1\alpha_2 + \alpha_1^3) \right)^n \\ &= 6^{-n} \sum_{\alpha+\beta+\gamma=n} \binom{n}{\alpha \quad \beta \quad \gamma} 2^\alpha \alpha_3^\alpha (-3)^\beta \alpha_1^\beta \alpha_2^\beta \alpha_1^{3\gamma} \\ &= 6^{-n} \sum_{\alpha+\beta+\gamma=n} \frac{2^\alpha (-3)^\beta n! \alpha_3^\alpha \alpha_2^\beta \alpha_1^{\beta+3\gamma}}{\alpha! \beta! \gamma!} \end{aligned}$$

we know that the coefficient of $x_1^3 x_2^3 \cdots x_n^3$ is

$$\begin{aligned} &6^{-n} \sum_{\alpha+\beta+\gamma=n} \frac{2^\alpha (-3)^\beta n!}{\alpha! \beta! \gamma!} \binom{n}{\alpha} \alpha! \binom{n-\alpha}{\beta} \beta! \times \\ &\quad \binom{\beta+3\gamma}{\underbrace{1 \quad 1 \quad \cdots \quad 1}_\beta \quad \underbrace{3 \quad 3 \quad \cdots \quad 3}_\gamma} \\ &= 6^{-n} \sum_{\alpha+\beta+\gamma=n} \frac{(-1)^\beta 2^\alpha 3^\beta (n!)^2 (\beta+3\gamma)!}{\alpha! \beta! (\gamma!)^2 6^\gamma} \\ &= 6^{-n} \sum_{\alpha=0}^n \sum_{\beta=0}^{n-\alpha} \frac{(-1)^\beta 2^\alpha 3^\beta (n!)^2 (3n-3\alpha-2\beta)!}{\alpha! \beta! (n-\alpha-\beta)! 2^6 (n-\alpha-\beta)^2} \end{aligned}$$

□

Some numbers obtained are listed in the following.

0

0

1

24

2040

297200

68938800

24046189440

12025780892160

8302816499443200

7673688777463632000

9254768770160124288000

14255616537578735986867200

27537152449960680597739468800

65662040698002721810659005184000

190637228506535883540302038364160000

665825560532772251175492202972938240000

276780648054221571651550187279222472704000

13564406360915457771720399143711430952267776000

77705104689340239554388061645133412621507133440000

Conjecture 53.

$$f(m, n, 4, 2) = \frac{n!}{2^n} \sum_{r_0=0}^m \sum_{r_1=0}^{m-r_0} \frac{m!}{r_0!r_1!(m-r_0-r_1)!} \frac{(-1)^{2(m-r_0)-r_1}}{(n-2m+2r_0+r_1)!} \frac{(4r_0+2r_1)!}{24^{r_0}2^{(m-r_0)}}$$

Some numbers obtained are listed in the following.

1

90

44730

56586600

154700988750

807998767676100

7373018003758407000

109829050417159537464000

2532230252503738514963235000

Conjecture 54.

$$f(m, n, 5, 2) = \frac{n!}{2^n} \sum_{r_0=0}^m \sum_{r_1=0}^{m-r_0} \frac{m!}{r_0!r_1!(m-r_0-r_1)!} \frac{(-1)^{r_1+2(m-r_0-r_1)}(4r_0+2r_1+m)!}{(n+r_1-2m+2r_0)!120^{r_0}6^{r_1}2^{(m-r_0-r_1)}}$$

Some numbers obtained are listed in the following.

1

756756

25989269017140

9647422924194982967040

24935177268489106332174087326700

Conjecture 55. $f(m, n, 6, 2) = \frac{n!}{2^n} \sum_{r_0=0}^m \sum_{r_1=0}^{m-r_0} \sum_{r_2=0}^{m-r_0-r_1}$

$$\frac{m!}{r_0!r_1!r_2!(m-r_0-r_1-r_2)!} \frac{(-1)^{3m-3r_0-2r_1-r_2}}{(n+2r_1+r_2-3m+3r_0)!}$$

Some numbers obtained are listed in the following.

1

1680

32496156

2604964362000

666569141498660400

452489654840076972518400

2.2.1 Algorithm

The algorithm used to verify the equations presented counts all the possible matrices, but does not construct them. It is best described with an example. Suppose we wanted to compute $f(12, 9, 3, 4)$. We first create a state vector of length 9, filled with 4s:

$$\#(4 \ 4 \ 4 \ 4 \ 4 \ 4 \ 4 \ 4 \ 4)$$

Each state vector can be thought of as a container to inform us how many ones need to go into each column. The '#' symbol reminds us that we must count the number of possibilities that we can put the indicated number of ones into each column. We assign where the ones will go in the first row. Clearly, 3 ones need to go in the first row somewhere, and there are $(9 \text{ take } 3) = 84$ possibilities for this placement. Hence, we simply assign them to go in the leftmost positions. Then, our state vector drops to $\#(3 \ 3 \ 3 \ 4 \ 4 \ 4 \ 4 \ 4 \ 4)$ noting that however many possibilities there are to fill in the remaining 11 rows, we multiply this by $(9 \text{ take } 3)$. Thus, we have

$$\#(4 \ 4 \ 4 \ 4 \ 4 \ 4 \ 4 \ 4 \ 4) = 84 * \#(3 \ 3 \ 3 \ 4 \ 4 \ 4 \ 4 \ 4 \ 4).$$

Eventually, we would like to drop the state vector to $\#(0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0)$ after

(exactly) all 12 rows have been assigned, reflecting a properly filled-in matrix. Now, for the second row, there are again 3 ones to place. Some of them can go in columns where ones are above, and some of them can go in columns where ones haven't been placed yet. The possibilities are as follows: 3/0, 2/1, 1/2, and 0/3, where x/y denotes putting x ones in the "left part" (where ones have been placed before) and y ones in the "right part" (where ones haven't been placed yet). We calculate each in turn.

For 3/0, there is only $(3 \text{ take } 3) = 1$ way to place all 3 ones in the left part, and $(6 \text{ take } 0) = 1$ way to place 0 ones in the right part. Hence, in this case we drop our state vector to $\#(2 \ 2 \ 2 \ 4 \ 4 \ 4 \ 4 \ 4 \ 4)$, since 2 ones will need to be placed in the leftmost three columns during subsequent row assignments, and we note that we'll multiply the ways to fill in a matrix this way by $(3 \text{ take } 3) * (6 \text{ take } 0) = 1 * 1$.

We also consider 2/1. There are $(3 \text{ take } 2) = 3$ ways to place 2 ones in the left part, and $(6 \text{ take } 1) = 1$ way to place a one in the right part. Now, as before, we will elect to place these ones in the leftmost area of each part.

Since 2 ones will be placed in the leftmost area of the left part, and 1 one will be placed in the leftmost area of the right part, our state vector in this case drops to

$$\#(2 \ 2 \ 3 \ 3 \ 4 \ 4 \ 4 \ 4 \ 4).$$

We also consider 1/2. There are $(3 \text{ take } 1) = 3$ ways to place 1 one in the left part, and $(6 \text{ take } 2) = 15$ ways to place a one in the right part. Hence, our state vector in this case drops to

$$\#(2 \ 3 \ 3 \ 3 \ 3 \ 4 \ 4 \ 4 \ 4).$$

We also consider 0/3. There is $(3 \text{ take } 0) = 1$ way to place 0 ones in the left part, and $(6 \text{ take } 3) = 20$ ways to place a one in the right part. Hence, our state vector in

this case drops to

$$\#(3\ 3\ 3\ 3\ 3\ 3\ 4\ 4\ 4).$$

Thus, in total, we have

$$\begin{aligned} \#(3\ 3\ 3\ 4\ 4\ 4\ 4\ 4\ 4) &= (1 * 1 * \#(2\ 2\ 2\ 4\ 4\ 4\ 4\ 4\ 4)) + (3 * 1 * \#(2\ 2\ 3\ 3\ 4\ 4\ 4\ 4\ 4)) \\ &+ (3 * 15 * \#(2\ 3\ 3\ 3\ 3\ 4\ 4\ 4\ 4)) + (1 * 20 * \#(3\ 3\ 3\ 3\ 3\ 3\ 4\ 4\ 4)). \end{aligned}$$

We would then proceed to work on each sub-state vector in turn. One final example: to compute $\#(2\ 2\ 3\ 3\ 4\ 4\ 4\ 4\ 4)$, we see that we have three parts: the left part (consisting of two columns) where 2 ones have already been placed, the middle part (consisting of two columns) where 1 one has already been placed, and the right part (consisting of five columns) where 0 ones have been placed. To assign our third row, we (again) need to place 3 ones, so we consider all the possibilities.

We see that 3/0/0 is not possible since there are only two columns in the left part. Similarly, 0/3/0 is not possible. We then compute the remaining possibilities: 2/1/0, 2/0/1, 1/2/0, 1/1/1, 1/0/2, and 0/0/3, and continue on.

After 11 of the 12 row assignments, we will either get state vectors like $\#(0\ 0\ 0\ 0\ 0\ 0\ 1\ 1\ 1)$ in which case we can terminate with a 1, or vectors like

$$\#(0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 1) \text{ or } \#(0\ 0\ 0\ 0\ 0\ 0\ 0\ 1\ 2)$$

in which case we can terminate with a 0, since it is impossible to fill in 3 ones in the last row in the prescribed manners.

This is the backbone of the algorithm. We remark that it is very possible to take different paths to get the same state vector later on, so we only compute its count once, storing it for later use if it shows up again. In its current implementation, the

calculation engine is completely separated from the storage object, so improvements to reading/writing from/to the storage object can be explored independently. We've found that in Scheme, a tree with ten branches at each node seems to optimize reading and writing, once the state vector is hashed (uniquely) into a whole number. Other node widths are certainly possible.

2.3 (0,1)-Matrices with Restriction

Let $f_s(n)$ be the number of $(0,1)$ - matrices of size $n \times n$ such that each row has exactly s 1's and each column has exactly s 1's and with the restriction that no 1 stands on the main diagonal. In this section we give rather involved closed formulas for $f_1(n)$ and $f_2(n)$, and a conjecture for $f_3(n)$, and present three instructive reformulations of the problem, and description of a algorithm.

Reformulation One:

There are $s \times n$ balls with s balls labelled $A_i, i = 1, 2, \dots, n$. Distribute these $s \times n$ balls into n distinct boxes numbered $1, 2, \dots, n$, such that each box contains s different balls, and the i -th box i does not contain $A_i, i = 1, 2, \dots, n$. How many distributions are there?

Reformulation Two:

There are $s \times n$ letters, each letter A_i appears exactly s times($i = 1, 2, \dots, n$). Let these $s \times n$ letters be arranged in a row according to: (from the left to the right on the row, we define the first location, the second location, \dots , the $s \times n$ -th location)

There is only one letter in each location.

There are no two equal letters A_i ($i = 1, 2, \dots, n$) in any two of the following s

locations: the $(sk + 1)$ -th location, $(sk + 2)$ -th location, \dots , the $(sk + s)$ -th location ($k = 0, 1, 2, \dots, s - 1$).

If $A_{i_1}, A_{i_2}, \dots, A_{i_s}$ are in the following s locations respectively: the $(sk + 1)$ -th location, $(sk + 2)$ -th location, \dots , the $(sk + s)$ -th location ($k = 0, 1, 2, \dots, s - 1$), then $i_1 < i_2 < \dots < i_s$.

A_i ($i = 1, 2, \dots, n$) is not in any of the $(sk + 1)$ -th location, $(sk + 2)$ -th location, \dots , the $(sk + s)$ -th location ($k = 0, 1, 2, \dots, s - 1$).

How many arrangements are there?

Reformulation Three:

$f_s(n)$ is equal to the number of labeled s -regular bipartite simple graphs on $2n$ vertices with the vertex set $V = V_1 \cup V_2$, $V_1 \cap V_2 = \emptyset$, $V_1 = \{u_1, u_2, \dots, u_n\}$, $V_2 = \{v_1, v_2, \dots, v_n\}$, and no edge between u_i and v_i , $i = 1, 2, \dots, n$.

Obviously, $f_s(n) = f_{n-s-1}(n)$, ($n > 3$).

In some row, let $x_{i_1}x_{i_2}\dots x_{i_k}$ denote the i_1 -th column, the i_2 -th column, \dots , the i_k -th column entries are 1 in some row and other entries are all 0, where $i_1, i_2, \dots, i_k \in \{1, 2, \dots, n\}$.

Example: Let $m = n = 4, s = t = 3$, then $x_1x_2x_3|x_1x_2x_4|x_1x_3x_4|x_2x_3x_4$ denotes the matrix as follows:

$$\begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{pmatrix}$$

In the following, we let:

$$\alpha_1 = x_1 + x_2 + \cdots + x_n;$$

$$\alpha_2 = x_1^2 + x_2^2 + \cdots + x_n^2;$$

$$\alpha_3 = x_1^3 + x_2^3 + \cdots + x_n^3;$$

$$\sigma_1 = \sum_{i=1}^n x_i = x_1 + x_2 + \cdots + x_n = \alpha_1;$$

$$\sigma_2 = \sum_{i < j} x_i x_j \quad \text{where } i, j \in \{1, 2, \dots, n\};$$

$$\sigma_3 = \sum_{i < j < k} x_i x_j x_k \quad \text{where } i, j, k \in \{1, 2, \dots, n\}$$

$[x]$ denotes the greatest integer such that $[x] \leq x$

Theorem 56. $f_1(n) = \sum_{k=0}^n (-1)^k \frac{n!}{k!}$.

Proof. Since $f_1(n)$ equals the coefficient of $x_1 x_2 \cdots x_n$ in the symmetric polynomial

$$\prod_{i=1}^n (\sigma_1 - x_i)$$

$$\prod_{i=1}^n (\sigma_1 - x_i) = \sum_{k=0}^n \sum_{\{i_1, i_2, \dots, i_k\} \in \{1, 2, \dots, n\}} (-1)^k \sigma_1^{n-k} x_{i_1} x_{i_2} \cdots x_{i_k}$$

Thus the coefficient of $x_1x_2\cdots x_n$ in $\prod_{i=1}^n(\sigma_1 - x_i)$ is

$$\sum_{k=0}^n (-1)^k \binom{n}{k} (n-k)! = \sum_{k=0}^n (-1)^k \frac{n!}{k!}$$

□

Some numbers obtained are listed in the following.

0

1

2

9

44

265

1854

14833

133496

1334961

14684570

176214841

2290792932

32071101049

481066515734

This is also the number of de-arrangements of length n .

Theorem 57. $f_2(n) = \sum_{k=0}^n \sum_{s=0}^k \sum_{j=0}^{n-k}$

$$\frac{(-1)^{k+j-s} n!(n-k)!(2n-k-2j-s)!}{s!(k-s)!((n-k-j)!)^2 j! 2^{2n-2k-j}}.$$

Some numbers obtained are listed in the following.

0

0

1

9

216

7570

357435

22040361

1721632024

166261966956

19459238879565

2714812050902545

445202898702992496

84798391618743138414

18567039007438379656471

This also the number of labeled 2-regular digraphs with n nodes.

Conjecture 58.

$$f_3(n) = \sum_{k=0}^n \sum_{t=0}^k \sum_{s=0}^t \sum_{p=0}^{n-k} \sum_{q=0}^{n-k-p} \sum_{r=0}^{k-t} \sum_{w=0}^{k-t} \frac{(-1)^{k+t+q+r-s} n! (n-k)!}{3^{2n-2k-p-2q-r+w} 2^{2n-2t-2p-q-r} p! q!} \frac{(k-t)! (q+r)!}{(n-k-p-q)! w! (k-t-w)!} \frac{(3n-k-3p-2q-t-2r-s)!}{r! (k-t-r)! s! (t-s)! (q+r-w)! (n-k-p-q-r+w)!}$$

Some numbers obtained are listed in the following.

1

44

7570

1975560

749649145

399035751464

28902113634903

277435664056527360

345023964977303838105

545099236551025860229460

107559520380415169555562244

This also the number of labeled 3-regular digraphs with n nodes.

2.3.1 Algorithm

Enclosed is a walkthrough for the Lefty algorithm which computes the number of $n \times n$ 0-1 matrices with t ones in each row and column, but none on the main diagonal.

The algorithm used to verify the equations presented counts all the possible matrices, but does not construct them.

It is called "Lefty", it is reasonably simple, and is best described with an example.

Suppose we wanted to compute the number of 6×6 0-1 matrices with 2 ones in each row and column, but no ones on the main diagonal. We first create a state vector of length 6, filled with 2s:

$\#(2\ 2\ 2\ 2\ 2\ 2)$

This state vector symbolizes the number of ones we must yet place in each column. We accompany it with an integer which we call the "puck", which is initialized to 1. This puck will increase by one each time we perform a ones placement in a row of the matrix (a "round"), and we will think of the puck as "covering up" the column that we won't be able to place ones in for that round.

Since we are starting with the first row (and hence the first round), we place two ones in any column, but since the puck is 1, we cannot place ones in the first column. This corresponds to the forced zero that we must place in the first column, since the 1,1 entry is part of the matrix's main diagonal.

The algorithm will iterate over all possible choices, but to show each round, we shall make a choice, say the 2nd and 6th columns. We then drop the state vector by subtracting 1 from the 2nd and 6th values, and advance the puck:

$\#(2\ 1\ 2\ 2\ 2\ 1); 2$

For the second round, the puck is 2, so we cannot place a one in that column. We choose to place ones in the 4th and 6th columns instead and advance the puck:

$\#(2\ 1\ 2\ 1\ 2\ 0); 3$

Now at this point, we can place two ones anywhere but the 3rd and 6th columns. At this stage the algorithm treats the possibilities differently: We can place some ones before the puck (in the column indexes less than the puck value), and/or some ones after the puck (in the column indexes greater than the puck value). Before the puck, we can place a one where there is a 1, or where there is a 2; after the puck, we can place a one in the 4th or 5th columns. Suppose we place ones in the 4th and 5th columns. We drop the state vector and advance the puck once more:

$\#(2\ 1\ 2\ 0\ 1\ 0); 4$

For the 4th round, we once again notice we can place some ones before the puck, and/or some ones after.

Before the puck, we can place:

- (a) two ones in columns of value 2 (1 choice)
- (b) one one in the column of value 2 (2 choices)
- (c) one one in the column of value 1 (1 choice)
- (d) one one in a column of value 2 and one one in a column of value 1 (2 choices).

After we choose one of the options (a)-(d), we must multiply the listed number of choices by one for each way to place any remaining ones to the right of the puck.

So, for option (a), there is only one way to place the ones.

For option (b), there are two possible ways for each possible placement of the

remaining one to the right of the puck. Since there is only one nonzero value remaining to the right of the puck, there are two ways total.

For option (c), there is one possible way for each possible placement of the remaining one to the right of the puck. Again, since there is only one nonzero value remaining, there is one way total.

For option (d), there are two possible ways to place the ones.

We choose option (a). We drop the state vector and advance the puck:

$\#(1\ 1\ 1\ 0\ 1\ 0); 5$

Since the puck is "covering" the 1 in the 5th column, we can only place ones before the puck. There are $(3 \text{ take } 2)$ ways to place two ones in the three columns of value 1, so we multiply 3 by the number of ways to get remaining possibilities. After choosing the 1st and 3rd columns (though it doesn't matter since we're left of the puck; any two of the three will do), we drop the state vector and advance the puck one final time:

$\#(0\ 1\ 0\ 0\ 1\ 0); 6$

There is only one way to place the ones in this situation, so we terminate with a count of 1. But we must take into account all the multiplications along the way:

$1*1*1*1*3*1 = 3$. So, this string of rounds counts the following three matrices:

0 1 0 0 0 1	0 1 0 0 0 1	0 1 0 0 0 1
0 0 0 1 0 1	0 0 0 1 0 1	0 0 0 1 0 1
0 0 0 1 1 0	0 0 0 1 1 0	0 0 0 1 1 0
1 0 1 0 0 0	1 0 1 0 0 0	1 0 1 0 0 0
1 1 0 0 0 0	1 0 1 0 0 0	0 1 1 0 0 0 <- only variation

0 0 1 0 1 0 0 1 0 0 1 0 1 0 0 0 1 0

Another way of thinking of the varying row is to start with the first matrix, focus on the lower-left 2x3 submatrix, and note how many ways there were to permute the columns of that submatrix. Since there are only 3 such ways, we get 3 matrices.

We cannot optimize by permuting submatrices that contain an entry of the main diagonal, since that is a 'fixed' position that must contain a zero.

We note that, in the actual implementation, after each round, the state vector values to the left of the puck are sorted (but the values to the right of the puck maintain their exact positions) to make counting possibilities easier. Hence, we would have in the third and fourth rounds, respectively,

#(1 2 2 1 2 0); 3

#(1 2 2 0 1 0); 4

In a larger example (13x13 matrix with 3 ones in each row/column), we might come across the following state:

#(0 1 1 1 2 2 3 3 0 1 0 0 1); 9

To place three ones in this case, the algorithm would branch depending on how many ones it wishes to place to the right of the puck, make that choice, and then multiply by the possibilities for placing the remaining ones to the left of the puck. Hence,

Case 1: Right of the puck gets 3 ones.

Not possible since there are only two nonzero columns there.

Case 2: Right of the puck gets 2 ones.

Only one way to do this, but there are three different ways to place the third one

to the left of the puck:

(a) under a column with a 1 value (3 ways), with resultant state $\#(0\ 0\ 1\ 1\ 2\ 2\ 3\ 3\ 0\ 0\ 0\ 0\ 0)$; 10

(b) under a column with a 2 value (2 ways), with resultant state $\#(0\ 1\ 1\ 1\ 1\ 2\ 3\ 3\ 0\ 0\ 0\ 0\ 0)$; 10

(c) under a column with a 3 value (2 ways), with resultant state $\#(0\ 1\ 1\ 1\ 2\ 2\ 2\ 3\ 0\ 0\ 0\ 0\ 0)$; 10.

Case 3: Right of the puck gets 1 one.

There are two ways to do this, so we have to branch depending on if it's going in the 10th column or 13th column.

Subcase 1: 10th column.

To place the other two ones to the left of the puck, we have choices:

(d) both ones under a 1-value ((3 take 2) ways),
with resultant state $\#(0\ 0\ 0\ 1\ 2\ 2\ 3\ 3\ 0\ 0\ 0\ 0\ 1)$; 10

(e) one one under 1-value, one under 2-value ((3 take 1)*(2 take 1) ways),
with resultant state $\#(0\ 0\ 1\ 1\ 1\ 2\ 3\ 3\ 0\ 0\ 0\ 0\ 1)$; 10

(f) one one under 1-value, one under 3-value ((3 take 1)*(2 take 1) ways),
with resultant state $\#(0\ 0\ 1\ 1\ 2\ 2\ 2\ 3\ 0\ 0\ 0\ 0\ 1)$; 10

(g) both ones under 2-value ((2 take 2) ways),
with resultant state $\#(0\ 1\ 1\ 1\ 1\ 1\ 3\ 3\ 0\ 0\ 0\ 0\ 1)$; 10

(h) one one under 2-value, one under 3-value ((2 take 1)*(2 take 1) ways),
with resultant state $\#(0\ 1\ 1\ 1\ 1\ 2\ 2\ 3\ 0\ 0\ 0\ 0\ 1)$; 10

(i) both ones under 3-value ((2 take 2) ways),

with resultant state $\#(0\ 1\ 1\ 1\ 2\ 2\ 2\ 2\ 0\ 0\ 0\ 1)$; 10.

Subcase 2: 13th column.

The options (j)-(o) are the same as (d)-(i) in the above subcase, but the resultant states have $\#(\dots\ 0\ 1\ 0\ 0\ 0)$ at the end instead.

Case 4: Right of the puck gets 0 ones.

So all three ones go to the left of the puck. We have choices:

(p) all ones under 1-value ((3 take 3) ways),

with resultant state $\#(0\ 0\ 0\ 0\ 2\ 2\ 3\ 3\ 0\ 1\ 0\ 0\ 1)$; 10

(q) two ones under 1-value, one under 2-value ((3 take 2)*(2 take 1) ways),

with resultant state $\#(0\ 0\ 0\ 1\ 1\ 2\ 3\ 3\ 0\ 1\ 0\ 0\ 1)$; 10

(r) two ones under 1-value, one under 3-value ((3 take 2)*(2 take 1) ways),

with resultant state $\#(0\ 0\ 0\ 1\ 2\ 2\ 2\ 3\ 0\ 1\ 0\ 0\ 1)$; 10

(s) two ones under 2-value, one under 3-value ((2 take 2)*(2 take 1) ways),

with resultant state $\#(0\ 1\ 1\ 1\ 1\ 1\ 2\ 3\ 0\ 1\ 0\ 0\ 1)$; 10

(t) one one under 2-value, two under 3-value ((2 take 1)*(2 take 2) ways),

with resultant state $\#(0\ 1\ 1\ 1\ 1\ 2\ 2\ 2\ 0\ 1\ 0\ 0\ 1)$; 10

In all options (a)-(t), the state would be resorted: since the puck moved from the 9th column to the 10th column, it will reveal a 0 in the 9th column, which will then get moved to the front of the state vector.

In general, Lefty will iterate over all possible choices (optimizing for permutations below the main diagonal by multiplying by the indicated cofactors), add up the values, and produce the result. To provide a further speedup, a storage object is used to store each state vector for which a count has been acquired, so that if that state vector is

seen again, the count can be produced from memory instead of recalculated. This speedup is necessary, and without it the algorithm will take too long.

2.4 Nonnegative Integer Matrices

The enumeration of nonnegative integer matrices has been the subject of considerable study. The determination of $t(m, n, s, t)$ is an unsolved problem and it is unlikely that a simple formula exists except for very small s, t . Equivalently, $t(m, n, s, t)$ counts 2-way contingency tables of order $m \times n$ such that the row marginal sums are all s and the column marginal sums are all t . Another equivalent description is that $t(m, n, s, t)$ is the number of semiregular labelled bipartite multigraphs with m vertices of degree s and n vertices of degree t . The matrices counted by $t(m, n, s, t)$ arise frequently in many areas of sciences, for example enumeration of permutations with respect to descents and statistics.

Conjecture 59. $t(n, n, 2, 2) = 4^{-n} \sum_{i=0}^n \frac{2^i (n!)^2 (2n-2i)!}{i! ((n-i)!)^2}$.

Some numbers obtained are listed in the following.

1

3

21

282

6210

202410

9135630

545007960

41514583320

3930730108200

452785322266200

62347376347779600

10112899541133589200

1908371363842760216400

414517594539154672566000

Conjecture 60. $t(n, m, 3, 2) = 2^{-m} \sum_{i=0}^n \frac{m!n!(2m-2i)!}{i!(m-i)!(n-i)!6^{n-i}}.$

Some numbers obtained are listed in the following.

7

16260

747558000

250071339672000

369820640830881240000

1796185853884657144990080000

23511842995969107700302647865600000

720289186703359375552628986978410240000000

Conjecture 61. $t(n, n, 3, 3) = 6^{-n} \sum_{\alpha=0}^n \sum_{\beta=0}^{n-\alpha} \frac{2^\alpha 3^\beta (n!)^2 (\beta+3(n-\alpha-\beta))!}{\alpha! \beta! (n-\alpha-\beta)!^2 6^{(n-\alpha-\beta)}}.$

Some numbers obtained are listed in the following.

1

4

55

2008

153040

20933840

4662857360

1579060246400

772200774683520

523853880779443200

477360556805016931200

569060910292172349004800

868071731152923490921728000

1663043727673392444887284377600

3937477620391471128913917360384000

2.4.1 Algorithm

The algorithm used to verify the equations presented counts all possible matrices, but does not construct them. It is a bit involved, so it is best described with an example.

Suppose we wanted to compute the number of 4×6 matrices over nonnegative integers with row sum 12 and column 8. We first create a list of all nonincreasing partitions of 12: 12, 11 1, 10 2, 10 1 1, 9 3, etc., and store this in memory. We make sure that each partition stored is not of length greater than the number of columns of the matrix. We then create a state vector of length 6 filled with 8s:

$\#(8\ 8\ 8\ 8\ 8\ 8)$

This state vector symbolizes the sum of integers we must place in each column, and each time the state changes, it is sorted in nondecreasing order.

An additional vector, called the cap vector, is created when we deal with a new state. It records the length of the contiguous blocks of numbers found in the state. Here, it is

$\#(6)$.

Next, we iterate over each of the (valid) partitions of 12 that we could possibly use for the choice of the first row of the matrix. Here, our first partition is 8 4. We then create a partition block (pb) vector, which is exactly a “cap vector” of the partition instead of the state. Here, it is

$\#(1\ 1)$.

Finally, we create all the assignment vectors that are valid for this partition and this cap vector. An assignment vector dictates where the indicated element of the partition will be placed in the row. Assignment vectors always have the same length as the partition we are planning to use. The entries of the assignment vector refer to the (zero-based) indices of the cap vector. Since the cap vector in this case only has one index (namely, 0) and both 8 and 4 can be elements in the matrix row, we assign 8 and 4 to the 0th index:

$\#(0\ 0)$

In other words, both the 8 and the 4 will appear in block 0 of the state. Now, there are $\binom{6}{1}\binom{5}{1}$ ways of placing the 8 and 4, so we note that when we drop the state vector. We pretend that the first row of the matrix will be (8 4 0 0 0 0), and so, dropping

the state vector, the remaining three rows must sum to

$$\#(0\ 4\ 8\ 8\ 8\ 8)$$

and we record that the number of ways of obtaining a matrix of state $\#(8\ 8\ 8\ 8\ 8\ 8)$ is 30 times the number of ways we can obtain a matrix of state $\#(0\ 4\ 8\ 8\ 8\ 8)$.

Of course, we must add to our count the other ways to assign the 8 and 4. Since there are no other ways, no more assignment vectors can be constructed. We then add to our count the ways in which we can use the partition 8 3 1 (with all applicable assignment vectors), and then 8 2 2 (with all applicable assignment vectors), and so forth.

To get a better feel for how the assignment vectors are created, let's say that, in the middle of our counting, we achieve the state

$$\#(1\ 1\ 4\ 6\ 6\ 6)$$

with two rows left to fill. Our cap vector is then

$$\#(2\ 1\ 3)$$

and suppose we are considering the partition 4 4 3 1. Its pb is $\#(2\ 1\ 1)$. Since the cap vector has length 3, the indices for it are 0, 1, and 2, so the entries of each assignment vector can be comprised only of 0, 1, and/or 2.

To create the first assignment vector, we note that the first element of the partition, 4, cannot be placed in block 0 of the state (the block of two 1s), since $4 > 1$. A single 4 can be placed in block 1 of the state (the block consisting of the single 4), so the first 4 in the partition can be assigned to block 1:

$$\#(1\ ?\ ?\ ?)$$

But block 1 is only length 1 (as noted by the cap vector's entry of 1 at index 1), so

no more 4s can go in that block. The second 4 in the partition can also be placed in block 2 of the state (the block of three 6s), since $<$. Thus, our assignment vector changes to

$$\#(1\ 2\ ?\ ?).$$

Next in the partition, we have a 3, which is also greater than 1, so it too cannot go into block 0. Block 1 has already been taken by the 4. Hence the only remaining place for it is in block 2:

$$\#(1\ 2\ 2\ ?)$$

Finally, the last element of the partition is a 1, which can go anywhere in the state.

We begin by assigning it to block 0, giving the resulting assignment vector as

$$\#(1\ 2\ 2\ 0).$$

How many ways could these assignments be carried out? The first 4 has only one way. The second 4 and the 3 are both in block 2, but they are different numbers, so they can be inserted in $\binom{3}{1}\binom{2}{1}$ ways. Finally, the 1 has $\binom{2}{1}$ ways to be inserted into block 0. Hence we multiply to get 12 ways for this assignment vector, and dropping the state, we get $\#(0\ 1\ 0\ 2\ 3\ 6)$. Sorting it, it becomes $\#(0\ 0\ 1\ 2\ 3\ 6)$, which we will process after we deal with the remaining assignment vectors possible for 4 4 3 1.

To get the next assignment vector, we note that we can keep everything the same, but the 1 in the partition can be put in block 2. This gives

$$\#(1\ 2\ 2\ 2)$$

and to compute the number of ways, we have $\binom{3}{1}\binom{2}{1}\binom{1}{1} = 6$.

To get the next assignment vector, we note we've exhausted all possibilities for $\#(1\ 2\ ?\ ?)$, so we then find the 'next' way to assign the two 4s in the partition. The only

remaining option is to put them both in block 2, so we start with

$$\#(2 \ 2 \ ? \ ?).$$

Now, the 3 can go in block 1 and the 1 can go in block 0, giving

$$\#(2 \ 2 \ 1 \ 0)$$

and total number of ways $\binom{3}{1}\binom{1}{1}\binom{2}{1} = 6$.

Now, we think of a “block” of the assignment vector as the entries that correspond to an equal number in the partition; here, the first two entries correspond to the partition entry 4, so they form a block. The pb tells us the length of each block of the assignment vector. For example, recall that here, pb is $\#(2 \ 1 \ 1)$, so each assignment vector corresponding to this partition has three blocks, the first of which has length two, and the remaining two have length one. We construct assignment vectors that are nondecreasing in each block, though we can have a decrease when we move to a new block from an old one. The remaining three assignment vectors and the number of ways to make the assignment are then

$$\#(2 \ 2 \ 1 \ 2) \text{ with ways } \binom{3}{2}\binom{1}{1}\binom{1}{1} = 3$$

$$\#(2 \ 2 \ 2 \ 0) \text{ with ways } \binom{3}{2}\binom{1}{1}\binom{2}{1} = 6$$

$$\#(2 \ 2 \ 2 \ 1) \text{ with ways } \binom{3}{2}\binom{1}{1}\binom{1}{1} = 3.$$

Let’s consider a larger example. Suppose the state was

$$\#(0 \ 1 \ 1 \ 1 \ 1 \ 2 \ 2 \ 2 \ 3 \ 3 \ 3 \ 3 \ 4 \ 5 \ 5)$$

with row sum 18. This state will produce a cap vector of $\#(4 \ 3 \ 4 \ 1 \ 2)$ (since zeroes in the state are ignored). Let’s suppose we were considering the partition

$$3 \ 3 \ 3 \ 2 \ 2 \ 2 \ 1 \ 1 \ 1,$$

which gives a pb of $\#(3\ 3\ 3)$. There are 433 total assignment vectors for this partition.

The first one we could construct is

$\#(2\ 2\ 2\ 1\ 1\ 1\ 0\ 0\ 0)$ with ways $\binom{4}{3}\binom{3}{3}\binom{4}{3} = 16$,

an intermediate one we could construct is

$\#(2\ 3\ 4\ 1\ 1\ 2\ 0\ 1\ 2)$

with ways $\binom{4}{1}\binom{1}{1}\binom{2}{1}$ (for placing the three 3s) $\times \binom{3}{2}\binom{3}{1}$ (for placing the three 2s) $\times \binom{4}{1}\binom{1}{1}\binom{2}{1}$ (for placing the three 1s) totalling 576,

and the last one we could construct is

$\#(3\ 4\ 4\ 2\ 2\ 2\ 1\ 1\ 2)$

with ways $\binom{1}{1}\binom{2}{2}$ (for placing the three 3s) $\times \binom{4}{3}$ (for placing the three 2s) $\times \binom{3}{2}\binom{3}{1}$ (for placing the three 1s) totalling 12.

Notice that each block of each assignment vector has its entries in nondecreasing order, but often there is a decrease when we move from block to block. Since the state vectors are nondecreasing, this is to be expected.

In general, for each state vector that is achieved, this algorithm will iterate over all assignment vectors for each valid partition, multiplying cofactors and adding the results. When fitting the last row, though, the calculation is surprisingly easy: continuing the example we had above, if we examine the state $\#(0\ 0\ 1\ 2\ 3\ 6)$, we see that there is only one possible partition of 12 that fits it (namely $6\ 3\ 2\ 1$) and there is only one way to fit it in. Hence, there is only one way to achieve this state. The situation is the same for every state with one row left to be filled.

For further speedup, a fast storage object must be used, so that if a given state is seen again, we can recall from memory how many partially-filled matrices can

produce it. This speedup is necessary, for without it, the algorithm will take too long. Other approaches and improvements are certainly possible, such as storing all possible assignment vectors for each partition and later recalling the applicable ones.

2.5 $(-1,0,1)$ -Matrices

Let $r(m, n, s, t)$ be the number of $(-1, 0, 1)$ -matrices of size $m \times n$ with each row sum equal to s and each column sum equal to t ($sm = nt$). $r(m, n, s, t)$ is related to the following two problems:

Can you obtain a formula for the number of connected labeled 2-regular pseudodigraphs? ("connected" means there is a path between any pair of vertices disregard the directions of edges, i.e. only one component).

Can you obtain a formula for the number of non-isomorphic connected labeled 2-regular pseudodigraphs?

This interest stems from some research in theoretical physics on the fractional quantum Hall effect. The graphs essentially arise in a Wick expansion of some physical quantity, and in physical terms the CONNECTED labeled 2-regular pseudodigraphs are Feynman diagrams. The graphs must be 2-regular because each vertex represents a 2-body interaction. For the non-isomorphic graphs, they are really physically different.

To compute the number of connected labeled 2-regular pseudodigraphs (multiple arcs and loops allowed) of order n . You start with the numbers of all labeled 2-regular pseudodigraphs and apply the standard recurrence relation to produce the numbers of those which are connected (weakly connected, to be more precise). That recurrence

can be found in. Then you can get a sequence: 0,1,2,14,201,4704,160890,... .

The enumeration of $(0, 1)$ –matrices is a special case of enumeration of $(-1, 0, 1)$ –matrices.

Some Easy Questions:

Question 1.

The number of $(-1, 0, 1)$ –matrices with matrix sum n ?

Matrix sum=the sum of all the entries of the matrix.

Answer : $+\infty$

Question 3.

The number of $(0, -1, 1)$ –matrices with m entries and matrix sum n ?

Answer:

$$\sum_{i=n}^{\lfloor (m+n)/2 \rfloor} \frac{d(m)m!}{i!(i-n)!(m+n-2i)!}.$$

Question 3.

The number of $r \times r$ $(-1, 0, 1)$ –matrices with matrix sum n ?

Answer:

$$\sum_{i=n}^{\lfloor (r^2+n)/2 \rfloor} \frac{(r^2)!}{i!(i-n)!(r^2+n-2i)!}.$$

Question 4. The number of $a \times r$ $(-1, 0, 1)$ –matrices with each row sum n ?

Answer:

$$\left(\sum_{i=n}^{\lfloor (r+n)/2 \rfloor} \frac{r!}{i!(i-n)!(r+n-2i)!} \right)^a$$

Let $f_2(n)$ be the number of $(-1, 0, 1)$ –matrices with each row and each column with exactly one "1" and one "−1". It is clearly that $f_2(n)$ is even.

$$f_2(n) = n(n-1)^2 f_2(n-2) + n(n-1) f_2(n-1)$$

$$f_2(1) = 0, f_2(2) = 2.$$

2.6 Nonnegative Matrices with Restriction

The number $(h(m, n, s, t))$ of nonnegative matrices of size $m \times n$ with row sum s and column t and the entries on the line from a_{11} to a_{kk} ($k = \min\{m, n\}$) are all 0 ($sm = nt$).

It is easy to get,

$$h(m, n, s, t) = h(n, m, t, s) \quad (sm = tn)$$

$$h(n, sn, s, 1) = g(sn, n, 1, s) = \sum_{k=0}^n (-1)^k \frac{n!(sn-k)!}{k!(n-k)!(s!)^{n-k}(s-1)!^k} \quad (s > 1)$$

Conjecture 62. $h(m, n, s, 2) =$

$$\begin{aligned} & \sum_{k=0}^m \sum_{r_0^1 + \dots + r_p^1 = k} \sum_{r_0^2 + \dots + r_q^2 = m-k} \\ & \frac{(-1)^k m! (n-k)!}{r_0^1! r_1^1! \dots r_p^1! r_0^2! r_1^2! \dots r_p^2! \prod_{i=0}^p ((s-1-2i)! i! 2^i)^{r_i^1}} \\ & \frac{(ms-k-2 \sum_{i=1}^p i r_i^1 - 2 \sum_{j=1}^q j r_j^2)!}{\prod_{j=0}^q ((s-2j)! j! 2^j)^{r_j^2} (n-k - \sum_{i=1}^p i r_i^1 + \sum_{j=1}^q j r_j^2)!} \\ & \frac{1}{2^{(n-k - \sum_{i=1}^p i r_i^1 + \sum_{j=1}^q j r_j^2)}} \\ & (2n = sm, p = \lfloor \frac{s-1}{2} \rfloor \text{ and } q = \lfloor \frac{s}{2} \rfloor). \end{aligned}$$

2.7 Open Problems

Problem 63. *How many $n \times n$ matrices in F_q (q is a prime) exist up to similarity?*

Problem 64. *Up to similarity, compute the number of $n \times n$ matrices with entries in F_2 with each column and each row with exactly one "1".*

That's the number of partitions of n with ordering.

Problem 65. *How many $n \times n$ trace zero matrices in F_q (q is a prime) exist up to similarity?*

2.8 Applications

Concepts in integer matrix and graph theory have been applied to the development of (a) a computerized method for determining structural identity (isomorphism) between kinematic chains, (b) a method for the automatic sketching of the graph of a mechanism defined by its incidence matrix, and (c) the systematic enumeration of general, single-loop constrained spatial mechanisms. These developments, it is believed, demonstrate the feasibility of computer-aided techniques in the initial stages of the design of mechanical systems. [17]

(0,1)-matrices with fixed row and column sum vectors; namely, determining its rank and in case the matrices are square its eigenvalues. It turns out that the trace of the structure matrix has some interesting properties. The rank of the structure matrix has the values 1, 2, or 3; this yields a classification of econometric models. [33]

There are more applications of integer matrices in [12], [28], [13], [15], [8].

CHAPTER 3: COMPUTATION OF WALKS AND PATHS

3.1 Introduction

A well-known long standing problem in combinatorics and statistical mechanics is to find the generating function for self-avoiding walks (SAW) on a two-dimensional lattice, enumerated by perimeter. A SAW is a sequence of moves on a square lattice which does not visit the same point more than once. In spite of this simple definition, many of the most basic questions about this model are difficult to resolve in a mathematically rigorous fashion. In particular, we do not know much about how far an n step self-avoiding walk typically travels from its starting point, or even how many such walks there are. These and other important questions about the self-avoiding walk remain unsolved in the rigorous mathematical sense, although the physics and chemistry communities have reached consensus on the answers by a variety of non-rigorous methods, including computer simulations. But there has been progress among mathematicians as well, much of it in the last decade, and the primary goal of this book is to give an account of the current state of the art as far as rigorous results are concerned. A second goal of this book is to discuss some of the applications of the self-avoiding walk in physics and chemistry, and to describe some of the non-rigorous methods used in those fields. The model originated in chemistry several decades ago as a model for long-chain polymer molecules. Since then

it has become an important model in statistical physics, as it exhibits critical behavior analogous to that occurring in the Ising model and related systems such as percolation [30]. It has been considered by more than one hundred researchers in the past one hundred years, including George Polya, Tony Guttmann, Laszlo Lovasz, Donald Knuth, Richard Stanley, Doron Zeilberger, Mireille Bousquet-Mélou, Thomas Prellberg, Neal Madras, Gordon Slade, Agnes Dittel, E.J. Janse van Rensburg, Harry Kesten, Stuart G. Whittington, Lincoln Chayes, Iwan Jensen, Arthur T. Benjamin, and others. More than three hundred papers and a few volumes of books were published in this area. A SAW is interesting for simulations because its properties cannot be calculated analytically. Calculating the number of self-avoiding walks is a common computational problem [40], [22]. In the past few decades, many mathematicians have studied the following two problems:

Problem 1

What is the number of SAWs from $(0, 0)$ to $(n - 1, n - 1)$ in an $n \times n$ grid, taking steps from $\{\uparrow, \downarrow, \leftarrow, \rightarrow\}$?

Donald Knuth claimed that the number is between 1.3×10^{24} and 1.6×10^{24} for $n = 11$ and he did not believe that he would ever in his lifetime know the exact answer to this problem in 1975. However, after a few years, Richard Schroepel pointed out that the exact value is $1, 568, 758, 030, 464, 750, 013, 214, 100 = 2^2 3^2 5^2 31 \times 115\,422\,379 \times 487\,148\,912\,401$ [11]. It is still an unsolved problem for $n > 25$.

Problem 2

What is the number $f(n)$ of n -step SAWs, on the square lattice, taking steps from $\{\uparrow, \downarrow, \leftarrow, \rightarrow\}$?

The number $f(n)$ is known for $n \leq 71$ [41].

A recently proposed model called prudent self-avoiding walks (PSAW) was first introduced to the mathematics community in an unpublished manuscript of Pr  a, who called them *exterior walks*. A prudent walk is a connected path on square lattice such that, at each step, the extension of that step along its current trajectory will never intersect any previously occupied vertex. Such walks are clearly self-avoiding [10]. We will talk about some sequences arising from PSAWs in the following.

3.2 Prudent Self-Avoiding Walks: Definitions and Examples

A PSAW is a proper subset of SAWs on the square lattice. The walk starts at $(0, 0)$, and the empty walk is a PSAW. A PSAW grows by adding a step to the end point of a PSAW such that the extension of this step - by any distance - never intersects the walk. Hence the name prudent. The walk is so careful to be self-avoiding that it refuses to take a single step in any direction where it can see - no matter how far away - an occupied vertex. The following walk is a PSAW.

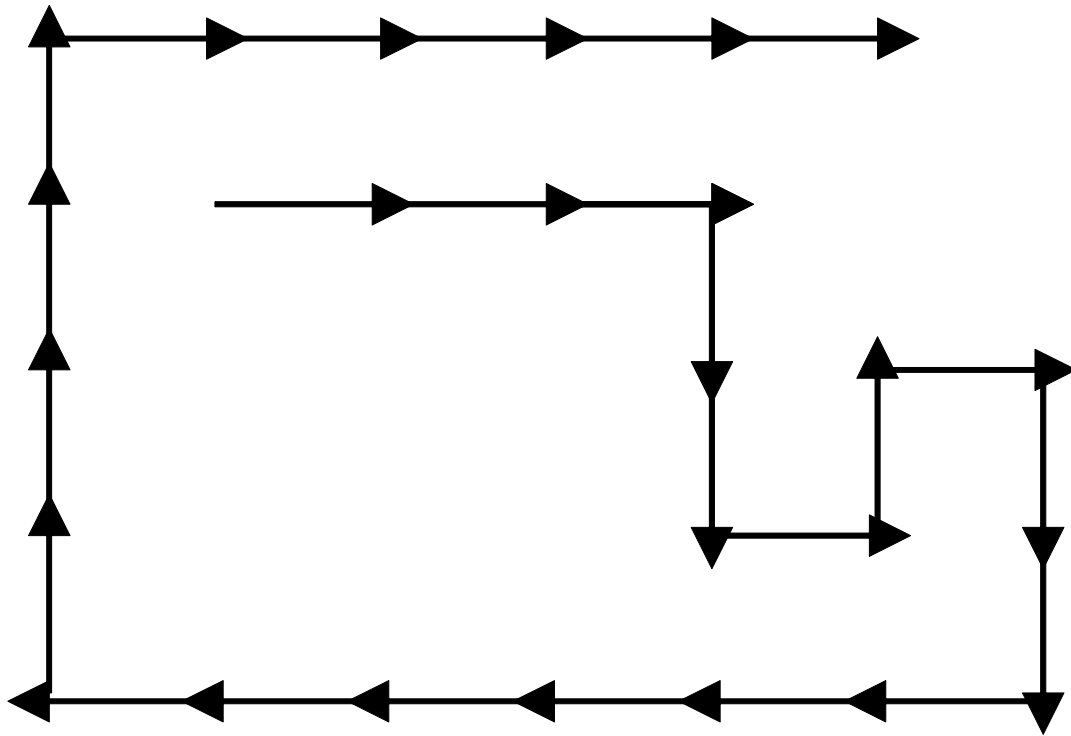


Figure 3: PSAW

3.2.1 Properties of a PSAW

Unlike SAW, PSAW are usually not reversible. There is such an example in the following figure.

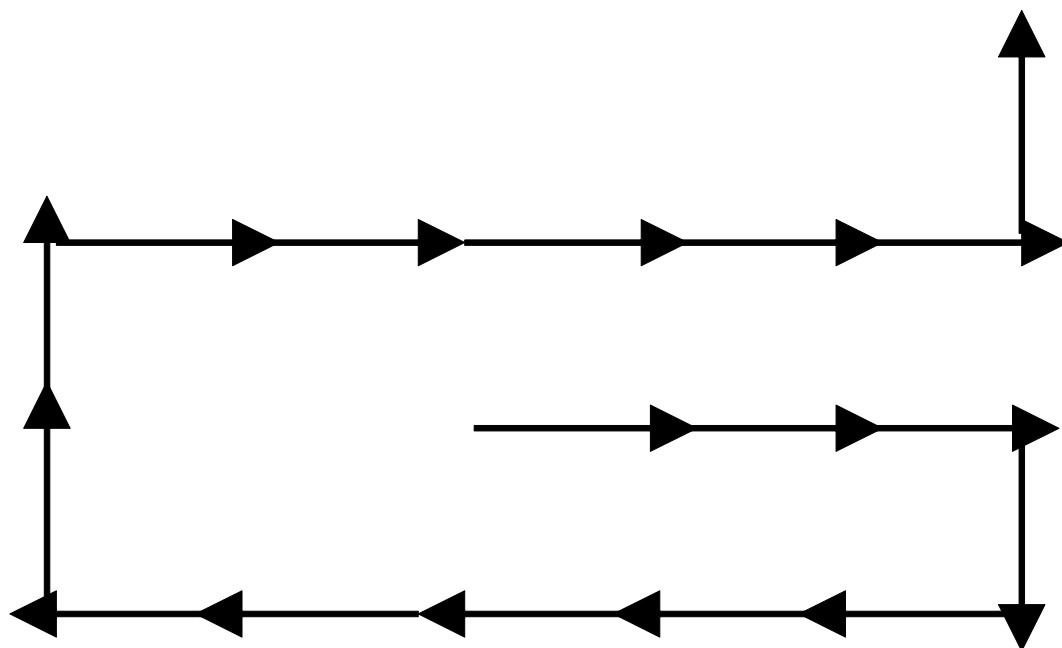


Figure 4: PSAW2

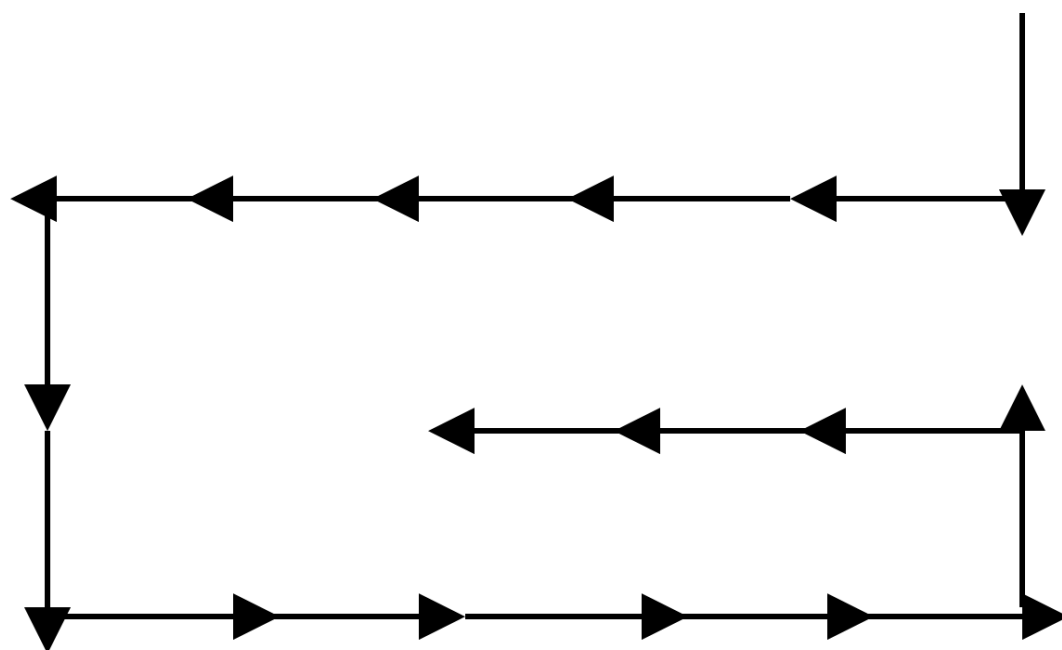


Figure 5: Not PSAW

Each PSAW possesses a minimum bounding rectangle, which we call *box*. Less obviously, the endpoint of a prudent walk is always a point on the boundary of the box. Each new step either inflates the box or walks (prudently) along the border. After an inflating step, there are 3 possibilities for a walk to go on. Otherwise, only 2.

In a *one-sided* PSAW, the endpoint lies always on the top side of the box. The walk is *partially directed*.

A prudent walk is *two-sided* if its endpoint lies always on the top side, or on the right side of the box. The walk in the following figure is a two-sided PSAW.

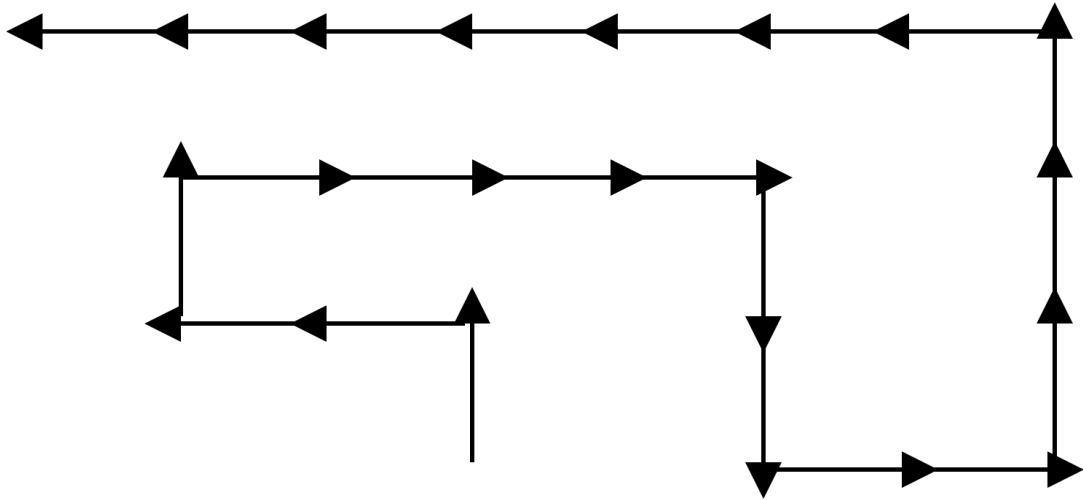


Figure 6: Two-sided PSAW

3.3 Some Sequences Arising from One-sided PSAWs

Sequence 1

What is the number (say $f(n)$) of one-sided n -step prudent walks, taking steps from $\{\uparrow, \leftarrow, \rightarrow\}$?

The generating function equals

$$\begin{aligned} \sum_{n \geq 0} f(n) t^n &= \frac{1+t}{1-2t-t^2} \\ &= 1 + 3t + 7t^2 + 17t^3 + 41t^4 + 99t^5 + \dots \end{aligned}$$

Also,

$$\begin{aligned} f(n) &= 2f(n-1) + f(n-2) \\ &= \frac{(1-\sqrt{2})^n + (1+\sqrt{2})^n}{2} \\ &= \begin{bmatrix} 1 & 0 \end{bmatrix} \left(\sum_{k=0}^{n+1} \binom{n+1}{k} \begin{bmatrix} 0 & 1 \\ 2 & 0 \end{bmatrix}^k \right) \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \end{aligned}$$

We obtain sequence A001333 of the On-Line Encyclopedia of Integer Sequences.

Sequence 2

The number of one-sided n -step prudent walks, starting from $(0,0)$ and ending on y -axis, taking steps from $\{\uparrow, \leftarrow, \rightarrow\}$ is

$$1 + \sum_{k=1}^{\lfloor (n-1)/2 \rfloor} \sum_{i=1}^{\min\{n-2k, k\}} \binom{n-2k+1}{i} \binom{k-1}{k-i} \binom{n-k-i}{k}.$$

We obtain sequence A136029.

Sequence 3

Consider the number of one-sided prudent walks starting from $(0, 0)$ to (x, y) , taking steps from $\{\uparrow, \leftarrow, \rightarrow\}$. The number of such walks with $k + x$ right \rightarrow steps, k left \leftarrow steps and y up \uparrow steps, is

$$\sum_{i=1}^{\min\{y, k+x\}} \binom{y+1}{i} \binom{k+x-1}{k+x-i} \binom{y+k-i}{k}.$$

If $k = 2$ and $x = y = n$, we obtain sequence A119578.

Sequence 4

The number of one-sided n -step prudent walks, from $(0, 0)$ to (x, y) , ($n - x - y$ is even) taking steps from $\{\uparrow, \leftarrow, \rightarrow\}$ is

$$\sum_{i=0}^{\min\{y, \frac{n+x-y}{2}\}} \binom{y+1}{i} \binom{\frac{n+x-y}{2}-1}{\frac{n+x-y}{2}-i} \binom{\frac{n-x+y}{2}-i}{\frac{n-x-y}{2}}.$$

If $x = y = 3$, we obtain sequence A163761.

Sequence 5

What is the number of the one-sided n -step prudent walks, avoiding k or more consecutive east steps, $\rightarrow^{\geq k}$?

The generating function equals

$$\frac{1 + t - t^k}{1 - 2t - t^2 + t^{k+1}}$$

If $k = 2$, we obtain sequence A006356, counting the number of paths for a ray of light that enters two layers of glass and then is reflected exactly n times before leaving the layers of glass.

If $k = 3$, we obtain sequence A033303 (see also page 244 in [39]).

Sequence 6

The number of one-sided n -step prudent walks, taking steps from $\{\uparrow, \leftarrow, \rightarrow, \nearrow\}$ equals

$$\frac{5 + \sqrt{17}}{2\sqrt{17}} \left(\frac{3 + \sqrt{17}}{2} \right)^n - \frac{5 - \sqrt{17}}{2\sqrt{17}} \left(\frac{3 - \sqrt{17}}{2} \right)^n.$$

We obtain sequence A055099.

Sequence 7

What is the number of one-sided n -step prudent walks, taking steps from $\{\rightarrow, \leftarrow, \uparrow, \nearrow, \searrow\}$?

The generating function is

$$\frac{1 + t}{1 - 4t - 3t^2}.$$

We obtain sequence A126473.

Sequence 8

What is the number of one-sided n -step prudent walks in the first quadrant, starting from $(0, 0)$ and ending on the y -axis, taking steps from $\{\uparrow, \leftarrow, \rightarrow\}$?

The generating function is

$$\frac{1}{2t^3}((1 + t)(1 - t)^2 - \sqrt{(1 - t^4)(1 - 2t - t^2)}).$$

Sequence 9

What is the number of one-sided n -step prudent walks *exactly* avoiding \leftarrow^k , taking steps from $\{\uparrow, \leftarrow, \rightarrow\}$?

The generating function equals

$$\frac{1 + t - t^k + t^{k+1}}{1 - 2t - t^2 + t^{k+1} - t^{k+2}}.$$

If $k = 1$, we obtain sequence A078061.

Sequence 10

What is the number of one-sided n -step prudent walks exactly avoiding $\leftarrow^{=k}$ and $\uparrow^{=k}$ (both at the same time)?

The generating function is

$$\frac{1 + t - 2t^k + 2t^{k+1}}{1 - 2t - t^2 + 2t^{k+1} - 2t^{k+2}}.$$

For $k = 1$,

$$f(n) = (2^{n+2} - (-1)^{\lfloor n/2 \rfloor} + 2(-1)^{\lfloor (n+1)/2 \rfloor}) / 5,$$

also,

$$f(n) = 2f(n-1) - f(n-2) + 2f(n-3)$$

$$\text{with } f(1) = 1, f(2) = 3, f(3) = 7.$$

This is sequence A007909.

3.4 Some Sequences Arising from Two-sided PSAWs

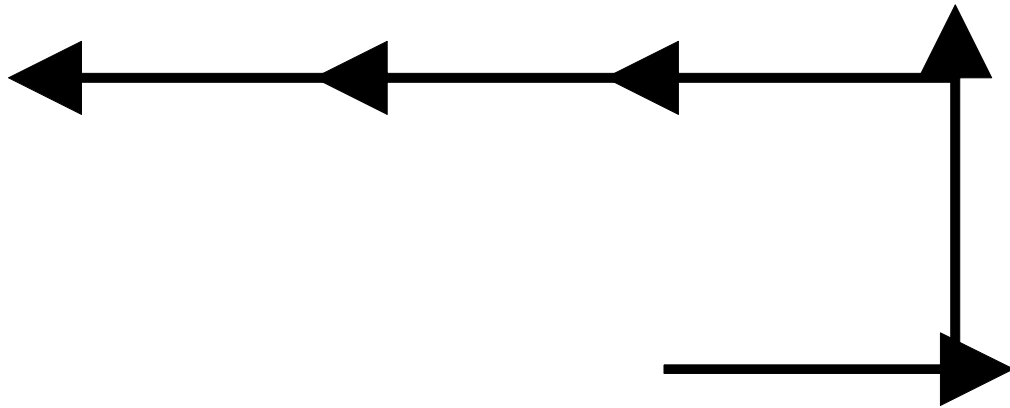
What is the number of two-sided, n -step prudent walks ending on the top side of their box avoiding both patterns $\leftarrow^{\geq 2}$, $\downarrow^{\geq 2}$ (both at the same time), taking steps from $\{\uparrow, \downarrow, \leftarrow, \rightarrow\}$?

Theorem 66. *The generating function (say $T(t, u)$) of the above two-sided prudent walks ending on the top side of their box satisfies*

$$\left(1 - t^2u - \frac{tu}{u-t}\right) T(t, u) = 1 + tu + T(t, t)t \frac{u-2t}{u-t}, \quad (5)$$

where u counts the distance between the endpoint and the north-east (NE) corner of the box.

For instance, in the following figure, a walk takes 5 steps, and the distance between the endpoint and the north-east corner is 3. So we can use $t^5 u^3$ to count this walk.



Outline of the proof of the theorem:

Case 1: Neither the top nor the right side has ever moved; the walk is only a west step. This case contributes 1 to the generating function.

Case 2: The last inflating step goes east. This implies that the endpoint of the walk was on the right side of the box before that step. After that east step, the walk has made a sequence of north steps to reach the top side of the box. Observe that, by symmetry, the series $T(t, u)$ also counts walks ending on the right side of the box by the length and the distance between the endpoint and the north-east corner. These two observations give the generating function for this class as $T(t, t)$.

Case 3: The last inflating step goes north. After this step, there is either a west step or a bounded sequence of East steps. This gives the generation function for this

class as

$$\left(t^2u + \frac{tu}{u-t}\right) T(t, u) - \frac{t^2}{u-t} T(t, t)$$

Putting the three cases together, we get the generating function (5) for $T(t, u)$.

Solve this generating function for $T(t, u)$ using the Kernel Method:

From

$$\left(1 - t^2u - \frac{tu}{u-t}\right) T(t, u) = 1 + tu + T(t, t) \left(t - \frac{t^2}{u-t}\right),$$

we can get

$$\begin{aligned} & (1 - tu)(u - tu - t - t^2u^2 + t^3u) T(t, u) \\ &= (u - t)(1 - tu)(1 + tu) - T(t, t)(1 - tu)t(2t - u) \end{aligned}$$

Set $(1 - tu)(u - tu - t - t^2u^2 + t^3u) = 0$, then there is only one *power series* solution for u

$$u = \frac{1}{2t^2} \left(1 - t + t^3 - \sqrt{(1 - t - t^3)^2 - 4t^4}\right).$$

Let U be this solution,

$$U = U(t) = \frac{1}{2t^2} \left(1 - t + t^3 - \sqrt{(1 - t - t^3)^2 - 4t^4}\right). \quad (6)$$

Set

$$(1 + tu)(u - t)(1 - tu) + T(t, t)(1 - tu)t(u - 2t) = 0,$$

and replace u by U :

$$T(t, t) = (1 + tU) \frac{t - U}{t(U - 2t)}. \quad (7)$$

From

$$\begin{aligned} & (1 - tu) (u - t - tu - t^2 u^2 + t^3 u) T(t, u) \\ &= (u - t)(1 - tu)(1 + tu) - T(t, t) (1 - tu) t (2t - u) \end{aligned}$$

get

$$T(t, u) = \frac{(t - u)(1 - tu)(1 + tu) + T(t, t) (1 - tu) t (2t - u)}{(1 - tu) (u - t - tu - t^2 u^2 + t^3 u)}.$$

Replace $T(t, t)$ by (7). Now

$$\begin{aligned} T(t, u) &= \frac{(1 + tu)(u - t)}{u - t - tu - t^2 u^2 + t^3 u} \\ &\quad - \frac{(1 + tU) (U - t) (1 - tu) (u - 2t)}{(U - 2t) (1 - tu) (u - t - tu - t^2 u^2 + t^3 u)} \end{aligned}$$

where $U(t)$ has been defined in (6).

Sequence 11

Notice that $T(t, 1)$ is the generating function of the number of two-sided n -step prudent walks ending on the top side of their box avoiding both patterns $\leftarrow^{\geq 2}, \downarrow^{\geq 2}$, taking steps from $\{\uparrow, \downarrow, \leftarrow, \rightarrow\}$, thus $T(t, 1) =$

$$\begin{aligned} & \frac{(1 - 2t) (1 - t) \sqrt{(1 - t - t^3)^2 - 4t^4}}{2t (1 - 2t - t^2 + t^3) (1 - 2t - 2t^3)} \\ & - \frac{(1 + t) (1 - 7t + 14t^2 - 11t^3 + 10t^4 - 4t^5)}{2t (1 - 2t - t^2 + t^3) (1 - 2t - 2t^3)} \\ &= 1 + 3t + 6t^2 + 15t^3 + 35t^4 + 83t^5 + \dots \end{aligned}$$

Sequence 12

Note that $T(t, 0)$ is the generating function of the number of two-sided n -step prudent walks ending at the north-east corner of their box avoiding both patterns

$\leftarrow^{\geq 2}, \downarrow^{\geq 2}$, taking steps from $\{\uparrow, \downarrow, \leftarrow, \rightarrow\}$, so $T(t, 0) =$

$$\begin{aligned} & \frac{(1-t) \sqrt{(1-t-t^3)^2 - 4t^4} - 1 + 3t - t^2 + t^3 + t^4}{(1-2t-2t^3)t} \\ &= 1 + 2t + 4t^2 + 10t^3 + 24t^4 + 56t^5 + \dots \end{aligned}$$

Sequence 13

Furthermore, $2T(t, 1) - T(t, 0)$ is the generating function of the number of two-sided n -step prudent walks ending on the top side or right side of their box avoiding both patterns $\leftarrow^{\geq 2}, \downarrow^{\geq 2}$, taking steps from $\{\uparrow, \downarrow, \leftarrow, \rightarrow\}$, thus $2T(t, 1) - T(t, 0) =$

$$\begin{aligned} & \frac{t(1-t)^2 \sqrt{(1-t-t^3)^2 - 4t^4}}{(1-2t-t^2+t^3)(1-2t-2t^3)} \\ &+ \frac{1-t-2t^2-2t^3-2t^4+4t^5-t^6}{(1-2t-t^2+t^3)(1-2t-2t^3)} \\ &= 1 + 4t + 8t^2 + 20t^3 + 46t^4 + 110t^5 \\ &+ 260t^6 + 616t^7 + 1456t^8 + 3442t^9 + \dots \end{aligned}$$

Open Problem 1

What is the number of two-sided n -step prudent walks, ending on the top side of their box, avoiding both $\leftarrow^{\geq k}$, and $\downarrow^{\geq k}$ ($k > 2$) taking steps from $\{\uparrow, \downarrow, \leftarrow, \rightarrow\}$?

The generating function satisfies:

$$\begin{aligned} & \left(1 - t^2 u \frac{1 - t^k u^k}{1 - tu} - \frac{tu}{u - t}\right) T(t, u) \\ &= 1 + tu \frac{1 - t^k u^k}{1 - tu} + \frac{u - 2t}{u - t} t T(t, t), \end{aligned}$$

where u counts the distance between the endpoint and the north-east corner of the box. For $k = 3$,

$$\begin{aligned} & \frac{u - t - t^2 u^2 + t^3 u - t^3 u^3 + t^4 u^2 - t^4 u^4 + t^5 u^3 - tu}{u - t} T(t, u) \\ &= 1 + tu + t^2 u^2 + t^3 u^3 + \frac{u - 2t}{u - t} t T(t, t) \end{aligned}$$

i.e.,

$$\begin{aligned} & (-t + (1 + t^3 - t)u + (t^4 - t^2)u^2 \\ & + (t^5 - t^3)u^3 - t^4 u^4) T(t, u) \\ &= (1 + tu + t^2 u^2 + t^3 u^3)(u - t) + t(u - 2t) T(t, t). \end{aligned}$$

Set $-t + (1 + t^3 - t)u + (t^4 - t^2)u^2 + (t^5 - t^3)u^3 - t^4 u^4 = 0$, and solve for u , as a power series of t . We obtained the first one hundred terms for u , beginning with

$$u = t + t^2 + t^3 + t^4 + 2t^5 + 4t^6 + 8t^7 + 16t^8 + 33t^9 + 69t^{10} + \dots$$

Using this u , we can get many examples for the sequence.

Open Problem 2

What is the number of two-sided n -step prudent walks, ending on the top side of their box, exactly avoiding both $\leftarrow^2, \downarrow^2$, taking steps from $\{\uparrow, \downarrow, \leftarrow, \rightarrow\}$?

The generating function is

$$\begin{aligned} & \left(1 - \frac{t^2 u}{1 - tu} - \frac{tu}{u - t} + u^2 t^3\right) T(t, u) \\ &= \frac{1}{1 - tu} - u^2 t^2 + \frac{u - 2t}{u - t} t T(t, t). \end{aligned}$$

REFERENCES

- [1] M. Aigner. *Markov's theorem and 100 years of the uniqueness conjecture*. Springer, 2015.
- [2] D. Applegate, O. Pol, and N. Sloane. The toothpick sequence and other sequences from cellular automata, congress. numerant., 206 (2010), 157–191. *arXiv preprint arXiv:1004.3036*.
- [3] I. N. Baoulina. Generalizations of the markoff–hurwitz equations over residue class rings. *International Journal of Number Theory*, 11(03):909–930, 2015.
- [4] A. Baragar. Integral solutions of markov-hurwitz equations. *Journal of Number Theory*, 49(1):27–44, 1994.
- [5] A. Baragar. On the unicity conjecture for markoff numbers. *Canadian Mathematical Bulletin*, 39(1):3–9, 1996.
- [6] A. Baragar. The exponent for the markoff–hurwitz equations. *pacific journal of mathematics*, 182(1):1–21, 1998.
- [7] A. Baragar. The markoff-hurwitz equations over number fields. *JOURNAL OF MATHEMATICS*, 35(3), 2005.
- [8] N. Bebiano. *Applied and Computational Matrix Analysis: MAT-TRIAD, Coimbra, Portugal, September 2015 Selected, Revised Contributions*, volume 192. Springer, 2017.

- [9] J. Bourgain, A. Gamburd, and P. Sarnak. Markoff triples and strong approximation. *Comptes Rendus Mathématique*, 354(2):131–135, 2016.
- [10] M. Bousquet-Mélou. Families of prudent self-avoiding walks. *Journal of Combinatorial Theory, Series A*, 117(3):313–344, 2010.
- [11] M. Bousquet-Mélou, A. Guttmann, and I. Jensen. Self-avoiding walks crossing a square. *Journal of Physics A: Mathematical and General*, 38(42):9159, 2005.
- [12] R. A. Brualdi. Matrices of zeros and ones with fixed row and column sum vectors. *Linear algebra and its applications*, 33:159–231, 1980.
- [13] R. A. Brualdi. *Combinatorial matrix classes*, volume 13. Cambridge University Press, 2006.
- [14] J. Button. The uniqueness of the prime markoff numbers. *Journal of the London Mathematical Society*, 58(1):9–17, 1998.
- [15] G. Dahl. Network flows and combinatorial matrix theory. *Retrieved from University of Oslo: <http://heim.ifi.uio.no/~geird/CombMatrix.pdf>*, 2010.
- [16] G. Dong and J. Pei. *Sequence data mining*, volume 33. Springer Science & Business Media, 2007.
- [17] F. Freudenstein and L. Dobjansky. Some applications of graph theory to the structural analysis of mechanisms. *ASME Transaction Journal of Engineering for Industry*, 89(1):153–158, 1967.

- [18] S. Gao, C. Caliskan, and S. Rong. Some properties of n-dimensional generalized markoff equation. *CONGRESSUS NUMERANTIUM*, 177:217, 2005.
- [19] S. Gao and K.-H. Chen. Tackling markoff-hurwitz equations. In *Computational Science and Computational Intelligence (CSCI), 2014 International Conference on*, volume 1, pages 341–346. IEEE, 2014.
- [20] S. Gao and K. H. Chen. Computation in markoff-hurwitz equations. In *2016 International Conference on Computational Science and Computational Intelligence (CSCI)*, pages 1292–1297, Dec 2016.
- [21] S. Gao and K. H. Chen. Computation in the enumeration of integer matrices. In *2016 International Conference on Computational Science and Computational Intelligence (CSCI)*, pages 1315–1320, Dec 2016.
- [22] S. Gao and K. H. Chen. Unsolved problems in computational science iii: A special self-avoiding walk. In *2016 International Conference on Computational Science and Computational Intelligence (CSCI)*, pages 1371–1374, Dec 2016.
- [23] E. González-Jiménez and J. M. Tornero. Markoff-rosenberger triples in arithmetic progression. *Journal of Symbolic Computation*, 53:53–63, 2013.
- [24] R. Guy. Unsolved problems in number theory problem books in mathematics subseries: Unsolved problems in intuitive mathematics, vol. 1, 2004.
- [25] H. Hosoya. Graph-theoretical properties of markoff numbers. topological indices of symmetrical brocomb graphs and perfect matching numbers of symmetrical stepomino graphs. 2011.

- [26] A. Hurwitz. Über eine aufgabe der unbestimmten analysis. *Arch. Math. Phys*, 3(11):185–196, 1907.
- [27] J. Kassels. An introduction to diophantine approximations. *Cambridge Tracts in Math. Phys*, (45), 1957.
- [28] H. L. Keng. Integer matrices and their applications. In *Introduction to Number Theory*, pages 365–404. Springer, 1982.
- [29] K.Guy and R.Nowakowski. An old four-squares chestnut. *The American Mathematical Monthly*, 100(4):589, 1993.
- [30] N. Madras and G. Slade. *The self-avoiding walk*. Springer Science & Business Media, 2013.
- [31] A. Markoff. Sur les formes quadratiques binaires indéfinies. *Mathematische Annalen*, 17(3):379–399, 1880.
- [32] P. Schmutz. Systoles of arithmetic surfaces and the markoff spectrum. *Mathematische Annalen*, 305(1):191–203, 1996.
- [33] G. Sierksma and E. Sterken. The structure matrix of $(0, 1)$ -matrices: Its rank, trace, and eigenvalues. an application to econometric models. *Linear Algebra and Its Applications*, 83:151–166, 1986.
- [34] N. Sloan and S. Plouffe. The encyclopedia of integer sequences, 1995.
- [35] N. Sloan, S. Plouffe, and many more. <http://oeis.org/wiki/>, 2017.

- [36] N. J. Sloane et al. The on-line encyclopedia of integer sequences.
- [37] N. J. A. Sloane. My favorite integer sequences. In *Sequences and their Applications*, pages 103–130. Springer, 1999.
- [38] N. J. A. Sloane. *A handbook of integer sequences*. Academic Press, 2014.
- [39] R. P. Stanley. Enumerative combinatorics. vol. 1, vol. 49 of cambridge studies in advanced mathematics, 1997.
- [40] I. Tishby, O. Biham, and E. Katzav. The distribution of path lengths of self avoiding walks on erdős–rényi networks. *Journal of Physics A: Mathematical and Theoretical*, 49(28):285002, 2016.
- [41] E. J. van Rensburg. Monte carlo methods for the self-avoiding walk. *Journal of Physics A: Mathematical and Theoretical*, 42(32):323001, 2009.
- [42] D. Zagier. On the number of markoff numbers below a given bound. *Mathematics of Computation*, 39(160):709–723, 1982.
- [43] Y. Zhang. Congruence and uniqueness of certain markoff numbers. *arXiv preprint math/0612620*, 2006.