ITERATIVE METHODS IN EXPERIMENTAL MATHEMATICS

BY EMILY KUKURA

An essay submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Master of Science

Graduate Program in Mathematics

Written under the direction of

Dr. Doron Zeilberger

and approved by

____

____

____

New Brunswick, New Jersey

April, 2017

Acknowledgements

I would like to thank my advisor, Dr. Doron Zeilberger, who gave so much of his time to guiding me through this project, both for his time and wisdom, as well as for his encouragement throughout the process. I would also like to thank Dr. Vladimir Retakh and Dr. Shubhangi Saraf for taking the time to be on my committee, and to thank numerous other professors I have been taught by both here at Rutgers and at the University of North Carolina—Chapel Hill, for helping to grow my mathematical curiosity. Lastly, I would like to thank Nathan Vaughn for providing helpful comments on an earlier version of this paper.

Table of Contents

Chapter 1

Introduction

Often when studying a given iterative procedure which can be applied to objects (i.e. sets, integers, etc.) falling into a variety of size classes, we may want to learn about how these procedures behave for an arbitrary size class. One way to gain information about this behavior, especially in the case that *a priori* information seems hard to obtain initially, is to study it experimentally by running various test cases. In order to get large amounts of experimental data we need to use computer software—in particular a computer algebra system like Maple can be very helpful. In this paper I will use experimental methods to study two very different iterative processes, the first related to the so-called "6174 Phenomenon", and the second to a version of Garsia and Milne's Involution Principle.

Chapter 2

The 6174 Phenomenon

Let b be an arbitrary base and l a positive integer. Then for any number η such that $0 \le \eta \le b^l - 1$ there is a unique expansion of η in base b, $\eta = \sum_{i=0}^{l-1} \eta_i \cdot b^i$, where $0 \le \eta_i \le b - 1$. We call this the base-b, length-l expansion with coefficients η_i. In particular, note that *only* the integers between 0 and $b^l - 1$ have base-b, length-l expansions, so that the set of all integers expressible in this way is finite (with cardinality b^l.) Let $S_{l,b}(\eta)$ be the multiset whose elements are the η_i's. In particular, $S_{l,b}$ is always a multiset of cardinality l, *even if the number η has leading 0's in its expansion*, for example, since $81 = 0 \cdot 5^3 + 3 \cdot 5^2 + 1 \cdot 5 + 1$, $S_{4,5}(81) = \{0, 1, 1, 3\}$, rather than $\{1, 3, 3\}$. We can then perform the iterative procedure described below, which is sometimes called the Kaprekar routine for reasons mentioned shortly (see 2.2).

2.1 Kaprekar Routine $K_{l,b}$ for length l, base b:

Arrange the elements of $S_{l,b}(\eta)$ in non-decreasing order $\eta_{k_0} \le \eta_{k_1} \le ... \le \eta_{k_{l-2}} \le \eta_{k_{l-1}}$.
Let $M_{l,b}(\eta) = \sum_{i=0}^{l-1} \eta_{k_i} \cdot b^i$ and $m_{l,b}(\eta) = \sum_{i=0}^{l-1} \eta_{k_i} \cdot b^{l-1-i}$, i.e. $M_{l,b}(\eta)$ is the maximum number that can be expressed as a base-b, length-l expansion whose coefficients are exactly the elements of the multiset $S_{l,b}(\eta)$, and $m_{l,b}(\eta)$ is the minimum such number.

Then we say that the 'Kaprekar function' $K_{l,b}(\eta) = M_{l,b}(\eta) - m_{l,b}(\eta)$ gives the successor to η. Note in particular that since $m_{l,b}(\eta) \le M_{l,b}(\eta)$, we have that $0 \le K_{l,b}(\eta) \le M_{l,b}(\eta)$, and so the integer $K_{l,b}(\eta)$ is expressible by a base-b length-l expansion. If $K_{l,b}(\eta) = \eta$ we stop, and we call such a point a *fixed point* of the Kaprekar routine for base b and length l. Otherwise, we can apply the same Kaprekar function to the successor $K_{l,b}(\eta)$ to find the second successor to η in this routine, denoted by $K_{l,b}^{(2)}(\eta) = K_{l,b}(K_{l,b}(\eta))$. If $K_{l,b}^{(2)}(\eta) \in \{\eta, K_{l,b}(\eta)\}$, then we stop. Otherwise we continue. In general at the j-th step we can consider the set P_{j-1} of all previously encountered numbers, i.e. letting $K_{l,b}^{(0)}(\eta) = \eta$, we have that $P_{j-1} = \{K_{l,b}^{(i)}(\eta) \ : \ 0 \le i \le j-1\}$. Then if $K_{l,b}^{(j)}(\eta) \in P_{j-1}$, we stop, and otherwise we continue. Since for all j, $K_{l,b}^{(j)}(\eta)$ has a base-b, length-l expansion, and only b^l integers have such an expansion, there is a $t \le b^l$ such that $K_{l,b}^{(t)}(\eta) \in P_{t-1}$. In particular, there is an $s < t$ for which $K_{l,b}^{(t)}(\eta) = K_{l,b}^{(s)}(\eta)$. At this point, we terminate the routine, and we call the list $\left[K_{l,b}^{(s)}(\eta) \ , \ K_{l,b}^{(s+1)}(\eta) \ , \ ... \ , \ K_{l,b}^{(t-1)}(\eta)\right]$ the *limiting orbit* of η under the Kaprekar routine, denoted $O_{l,b}(\eta)$. The term *orbit* is used here because, if instead of terminating once we reach the desired integer t, we continued applying the Kaprekar function indefinitely, it would continuously cycle through the elements of the orbit $O_{l,b}(\eta)$.

2.2 History

D. R. Kaprekar pointed out in 1949 (see [4]) that in base 10 the 4-digit number 6174 is a fixed point under the function $K_{4,10}$, since $K_{4,10}(6174) = 7641 - 1467 = 6174$. What is more, he noted that in fact *every* base 10 length 4 integer which has at least two distinct digits (e.g. excluding the '*degenerate*' integers $nnnn$ for n from 0 to 9) eventually terminates to the number 6174 under the Kaprekar routine. In other words, when $b = 10$ and $l = 4$, for every such integer η there exists a positive integer j such that $K^{(j)}_{4,10}(\eta) = 6174$, and so since 6174 is a fixed point, $O_{4,10}(\eta) = [6174]$.

It is also the case in base 10 that every non-degenerate length-3 number eventually terminates to the number 495 under the Kaprekar routine, i.e. $O_{3,10}(\eta) = [495]$ for every non-degenerate 3-digit η. Since for base 10 we see that both $l = 3$ and $l = 4$ have the property that all non-degenerate integers terminate in a single specific integer, one may ask if this property always holds for any length l in base 10. However, far from being the case, for the lengths $2, 5, 6, 7,$ and 8 it is not only false that the non-degenerate integers fall into a single length 1 orbit, but for some of these lengths (specifically, $l = 5, 6, 8$) there is not even one specific orbit that they all fall into. For example, the set of all possible orbits ranging over all non-degenerate η for length 5 in base 10 is: $\{[53955, 59994], [61974, 82962, 75933, 63954], [62964, 71973, 83952, 74943]\}$, so there is a length 2 orbit and two distinct length 4 orbits that a number can fall into.

In fact for a general base b and length l it is very rare that there only be one possible length 1 terminating orbit. Thus the natural question arises: Is there any way to characterize what the set of terminating orbits looks like for a general b and l? For some specific lengths several mathematicians and computer scientists have discovered patterns for bases of a certain form, for example in [1] Eldridge and Sagong characterize what the set of limiting orbits is in length 3, showing that all even bases follow a certain pattern and all odd bases follow a different pattern. The specific theorem is given below, see 2.4.1.

One should note that in some formulations of the Kaprekar routine leading zeros *are* thrown away, however a little thought shows that this modification does not affect the overall set of limiting orbits, all it does is make it possible for more numbers which are originally non-degenerate in a given base to become degenerate at some point in the routine, and thus lead to an orbit of 0. For example, under Kaprekar's original formulation and the one used in this paper, where leading zeros are *not* thrown away, for length 3, base 10 we have that $100 \to 100 - 001 = 099 \to 990 - 099 = 891 \to 981 - 189 = 792 \to 972 - 279 = 693 \to 963 - 369 = 594 \to 954 - 459 = \underline{495} \to 954 - 459 = \underline{495}$, while in the modified version we simply have $100 \to 100 - 001 = 99 \to 99 - 99 = \underline{0}$. In general when leading zeros are kept, for bases $b \geq 3$, if η is non-degenerate to begin with, it will never become degenerate, a fact which could be seen using some of the simple algebraic proof techniques shown below, for example a similar method to those used in the proofs of Propositions 1 and 2 below for general lengths would work.

2.3 Further Notation

Note that when the base and length are understood, the base-b, length-l expansion of η can be expressed by the list $\eta'(l,b) = [\eta_{l-1}, \eta_{l-2}, ..., \eta_2, \eta_1, \eta_0]$, where $0 \leq \eta_i \leq b-1$ for $0 \leq i \leq l-1$. For ease of notation from this point forth we shall always refer to this list form of the number η simply by η', with the dependence on b and l implied. Similarly, we will often simply refer to the Kaprekar function K (instead of $K_{l,b}$) and the operators, M and m (instead of $M_{l,b}$ and $m_{l,b}$, respectively). Note using this terminology $M(\eta)'$ is the list of length l with the same entries as η' sorted in non-increasing order, and $m(\eta)'$ is the analogous list with entries sorted in non-decreasing order. Using this list notation we can also talk of K, M, and m acting directly on lists of integers, e.g. for $b = 10$ and $l = 4$, $K([6,1,7,4]) = [6,1,7,4]$, $M([6,1,7,4]) = [7,6,4,1]$, and $m([6,1,7,4]) = [1,4,6,7]$. In particular, for any integer η we have that $K(\eta)' = K(\eta')$, $M(\eta)' = M(\eta')$ and $m(\eta)' = m(\eta')$.

Additionally, if we have an integer which is represented in base b by the expansion $s = [\eta_{l-1}, \eta_{l-2}, ..., \eta_2, \eta_1, \eta_0]$, then the notation s^* denotes the underlying integer, e.g. in base 10, $[6,7,1,4]^* = 6714$. So in particular for any non-negative integer η, $(\eta')^* = \eta$ and for any list of non-negative integers s, $(s^*)' = s$. In addition, when the base b is understood we shall often simply refer to "the number" $[\eta_{l-1}, \eta_{l-2}, ..., \eta_2, \eta_1, \eta_0]$, but this slight abuse of language should cause no confusion.

Lastly, for a given length l and base b, let $Orb(l,b)$ denote the set of all possible limiting orbits for length l and base b, i.e. $Orb(l,b) = \{O_{l,b}(\eta) : 0 \leq \eta \leq b^l - 1\}$. Let $N(l,b)$ denote the number of limiting orbits (so $N(l,b) = |Orb(l,b)|$). Finally, if we enumerate the set $Orb(l,b)$ as a list (of length $N(l,b)$) $[o_1, o_2, ...o_{N(l,b)}]$, we can let $L(l,b)$ be the corresponding list of orbit lengths, sorted in increasing order. So for example we have $Orb(4,4) = \{ \ [\ [3,0,2,1]\ ], [\ [1,3,3,2],\ [2,0,2,2]\ ]\}$, and thus $L(4,4) = [1,2]$.

2.4 Previous Results

Note that all theorems, propositions and conjectures stated in the remainder of this chapter assume that the starting number η is non-degenerate, i.e. not all of the coefficients η_i in the base-b, length-l expansion of η are the same.

2.4.1 Length 3

The following theorems were proved in full, in addition to further results on the Kaprekar routine for length 3, by Eldridge and Sagong in [1]. I will give my own proof below in order to give the reader an idea of some of the basic mechanisms that may be needed to prove the (much more complicated) conjectures which follow later on.

Theorem 1 (Even Base)**.** *If $b = 2n$ is an even positive integer s.t. $b \geq 2$, then $N(3, b) = 1$ and $L(3, b) = [1]$. In particular, the single limiting orbit is given by:*

$$[\ [n-1, 2n-1, n]\ ]$$

Theorem 2 (Odd Base)**.** *If $b = 2n+1$ is an odd positive integer s.t. $b \geq 3$, then $N(3, b) = 1$ and $L(3, b) = [2]$. In particular, the single limiting orbit is given by:*

$$[\ [n-1, 2n, n+1], [n, 2n, n]\ ]$$

First we prove the following easy algebraic result:

Proposition 1. *Suppose $\eta = [\eta_0, \eta_1, \eta_2]$, where η is non-degenerate, and let $n_2 \geq n_1 \geq n_0$ be a listing of the entries of η in non-increasing order. Let $\delta = \delta(\eta)$ be the difference $\delta = n_2 - n_0$, (so in particular since η is non-degenerate, $\delta \geq 1$). Then for any base b, the successor $K(\eta)$ to η under one iteration of the Kaprekar function is:*

$$K(\eta) = [\delta - 1, b - 1, b - \delta]$$

<u>Proof:</u>
One iteration of the Kaprekar function gives us:

$$\begin{array}{rrllll} & M(\eta^*) = & n_2 \cdot b^2 & + \ n_1 \cdot b & + \ n_0 \\ - & m(\eta^*) \ = & n_0 \cdot b^2 & + \ n_1 \cdot b & + \ n_2 \\ \hline & K(\eta^*) \ = & \delta \cdot b^2 & + & - \ \delta \\ & = & (\delta - 1) \cdot b^2 + & (b-1) \cdot b + & b - \delta \end{array}$$

Thus $K(\eta) = [\delta - 1, b - 1, b - \delta]$, as desired. ■

Now, to the proofs of Theorems 1 and 2.

Proof:

Suppose $\eta = [\eta_0, \eta_1, \eta_2]$ is non-degenerate. Then by Proposition 1, letting $\alpha_0 = \delta(\eta)$ we have that $K^{(1)}(\eta) = K(\eta) = [\alpha_0 - 1, b - 1, b - \alpha_0]$. In particular note $\alpha_0 \neq 0$ since η is non-degenerate.

Assuming we have not already terminated, let α_1 be the difference $\delta(K^{(1)}(\eta))$. Then $b - 1 \geq \max\{\alpha_0 - 1, b - \alpha_0\}$ (because $b - 1$ is the maximum possible coefficient in base-b), and thus since $(\alpha_0 - 1) + (b - \alpha_0) = b - 1$, by definition of δ we have:

$$\alpha_1 = \delta(K^{(1)}(\eta)) = (b - 1) - \min\{\alpha_0 - 1, b - \alpha_0\} = \max\{\alpha_0 - 1, b - \alpha_0\}.$$

In general at each step in the process until we terminate we can define $\alpha_i = \delta(K^{(i)}(\eta))$. Assume we have reached the j-th step in the process without terminating. Then we have by the definition of α_{j-1} and by Proposition 1 that:

$$K^{(j)}(\eta) = K(K^{(j-1)}(\eta)) = [\alpha_{j-1} - 1, b - 1, b - \alpha_{j-1}]$$

and so since $(\alpha_{j-1} - 1) + (b - \alpha_{j-1}) = b - 1$:

$$\alpha_j = \delta(K^{(j)}(\eta)) = (b - 1) - \min\{\alpha_{j-1} - 1, b - \alpha_{j-1}\} = \max\{\alpha_{j-1} - 1, b - \alpha_{j-1}\}$$

Thus for all j throughout the procedure we have:

$$\alpha_j = \delta(K^{(j)}(\eta)) \quad \text{(by definition)},$$

$$K^{(j)}(\eta) = K(K^{(j-1)}(\eta)) = [\alpha_{j-1} - 1, b - 1, b - \alpha_{j-1}] \quad \text{(by Proposition 1)}, \qquad \text{and}$$

$$\alpha_j = \max\{\alpha_{j-1} - 1, b - \alpha_{j-1}\}.$$

Let $n = \lfloor b/2 \rfloor$, so that if b is even, $b = 2n$ and if b is odd, $b = 2n+1$. Now note that:

($\star$) *If before terminating we reached an i s.t. $\alpha_i = n$, then $O_{l,b}(\eta)$ is as specified in the theorems, depending on whether b is even or odd:*

If $b = 2n$ is even, then at step $i+1$ we have that:

$$K^{(i+1)}(\eta) = [\alpha_i - 1, 2n-1, 2n-\alpha_i] = [n-1, 2n-1, n]$$

But then $\alpha_{i+1} = \max\{\alpha_i - 1, 2n-\alpha_i\} = \max\{n-1, n\} = n$ and so also $K^{(i+2)}(\eta) = [n-1, 2n-1, n] = K^{(i+1)}(\eta)$, and so in particular since η can only fall into one orbit, this orbit must be $O_{l,b}(\eta) = [\ [n-1, 2n-1, n]\ ]$.

If $b = 2n+1$ is odd, then:

$$\alpha_i = n \;\Rightarrow\; K^{(i+1)}(\eta) = [\alpha_i - 1, 2n, 2n+1-\alpha_i] = [n-1, 2n, n+1]$$

$$\Rightarrow\; \alpha_{i+1} = \max\{\alpha_i - 1, 2n+1-\alpha_i\} = \max\{n-1, n+1\} = n+1 \;\Rightarrow\; K^{(i+2)}(\eta) = [n, 2n, n]$$

$$\Rightarrow\; \alpha_{i+2} = \max\{\alpha_{i+1} - 1, 2n+1-\alpha_{i+1}\} = \max\{n, n\} = n \;\Rightarrow\; K^{(i+3)}(\eta) = [n-1, 2n, n+1].$$

Thus $K^{(i+3)}(\eta) = K^{(i+1)}(\eta)$ and so again since η can only fall into one orbit, we must have $O_{l,b}(\eta) = [\ [n-1, 2n, n+1], [n, 2n, n]\ ]$.

Thus if $\alpha_0 = n$ then in either case we get the desired limiting orbit, and so we may assume $\alpha_0 \neq n$. Next, I claim we either have that $\alpha_1 = \alpha_0 - 1$, or else $\alpha_2 = \alpha_1 - 1$. This simple claim is proved immediately below the end of this proof (see Claim 1).

Thus there is some smallest $t \in \{1, 2\}$ such that $\alpha_t = \alpha_{t-1} - 1$, and further it can easily be seen that since $\alpha_0 \neq n$, we have not yet terminated at t (in particular, for $t = 1$ this is proved as part of of Claim 1 below, and the $t = 2$ case can be shown in a similar manner). Then without loss of generality $t = 1$, otherwise we can shift all the indices below by 1 and the result still follows. Consider the 2nd step. Since $\alpha_1 = \alpha_0 - 1$, we have that:

$$K^{(2)}(\eta) = K(K^{(1)}(\eta)) = [(\alpha_0 - 1) - 1, b-1, b-(\alpha_0 - 1)] = [\alpha_0 - 2, b-1, b-\alpha_0+1],$$

so we see that since $K^{(1)}(\eta) = [\alpha_0 - 1, b-1, b-\alpha_0]$, applying K to $K^{(1)}(\eta)$ decreased the first entry by one and increased the 3rd entry by one, while the middle entry stayed constant. Thus so long as at each step we have $\alpha_j = \alpha_{j-1} - 1$, we see by inductively applying the same process that:

$$K^{(j+1)}(\eta) = [\alpha_{j-1} - 2, b-1, b-\alpha_{j-1}+1] = [\alpha_0 - 1 - j, b-1, b-\alpha_0+j],$$

so that the first entry is j smaller than $\alpha_0 - 1$ and the 3rd entry is j larger than $b - \alpha_0$, while the middle entry stays constant at $b-1$. However, clearly since each entry must be non-negative this process cannot continue indefinitely (in particular since $\alpha_0 - 1 \leq b - 2$, it cannot continue for more than $b-1$ steps), and so there must be some first $\hat{\jmath}$ after t, so $2 \leq \hat{\jmath} \leq b$, such that $\alpha_{\hat{\jmath}} \neq \alpha_{\hat{\jmath}-1} - 1$. In particular, since $\alpha_{\hat{\jmath}} = \max\{\alpha_{\hat{\jmath}-1} - 1, b - \alpha_{\hat{\jmath}-1}\}$, we must have $\alpha_{\hat{\jmath}-1} - 1 < b - \alpha_{\hat{\jmath}-1}$. What happens when we reach this $\hat{\jmath}$? Well, at this point we have both that $\alpha_{\hat{\jmath}-1} = \alpha_{\hat{\jmath}-2} - 1$ and that $\alpha_{\hat{\jmath}-1} - 1 < b - \alpha_{\hat{\jmath}-1}$. Now:

$$\alpha_{\hat{\jmath}-1} = \alpha_{\hat{\jmath}-2} - 1 \Rightarrow \alpha_{\hat{\jmath}-2} - 1 \geq b - \alpha_{\hat{\jmath}-2} \Rightarrow \alpha_{\hat{\jmath}-2} \geq \frac{(b+1)}{2} \Rightarrow \alpha_{\hat{\jmath}-1} = \alpha_{\hat{\jmath}-2} - 1 \geq \frac{(b-1)}{2}$$

and

$$\alpha_{\hat{\jmath}-1} - 1 < b - \alpha_{\hat{\jmath}-1} \Rightarrow \alpha_{\hat{\jmath}-1} - 1 \leq b - \alpha_{\hat{\jmath}-1} - 1 \Rightarrow \alpha_{\hat{\jmath}-1} \leq \frac{b}{2}.$$

Thus altogether, $\frac{(b-1)}{2} \leq \alpha_{\hat{\jmath}-1} \leq \frac{b}{2}$, and so $\alpha_{\hat{\jmath}-1} = \lfloor b/2 \rfloor = n$, and so we're done by ($\star$) on the previous page. ■

Claim 1. *Using the notation of the above proof of Theorems 1 and 2, and assuming that $\alpha_0 \neq n$, when applying the Kaprekar function to $\eta = [\eta_0, \eta_1, \eta_2]$, we must have either that $\alpha_1 = \alpha_0 - 1$, or else $\alpha_2 = \alpha_1 - 1$.*

Proof:

If $\alpha_1 \neq \alpha_0 - 1$, then $\alpha_0 - 1 < b - \alpha_0$, and so first of all we couldn't have terminated the Kaprekar routine yet, for that could only have happened if $[\eta_0, \eta_1, \eta_2] = \eta = K^{(1)}(\eta) = [\alpha_0 - 1, b-1, b-\alpha_0]$. In particular since $\alpha_0 - 1 < b - \alpha_0 \leq b-1$, $\eta_0 < \eta_2 \leq \eta_1$, and so $\alpha_0 = \delta(\eta) = \eta_1 - \eta_0 = b - 1 - \eta_0$. So in particular:

$$\eta_0 = \alpha_0 - 1 \quad \text{and} \quad \alpha_0 = b - 1 - \eta_0 \quad \Rightarrow \quad 2\eta_0 = b - 2$$

Thus in particular this would forece b to be even, so $b = 2n$. But then $\eta_0 = \frac{2n-2}{2} = n - 1$, which implies that $\alpha_0 = b - 1 - \eta_0 = 2n - 1 - (n-1) = n$, contrary to the assumption that $\alpha_0 \neq n$. Thus we haven't terminated after the first step.

Now, since $\alpha_1 = \max\{\alpha_0 - 1, b - \alpha_0\}$, we have both that $\alpha_1 = b - \alpha_0$ and $\alpha_0 - 1 < b - \alpha_0$, so together we have that $\alpha_0 - 1 < \alpha_1$, so

$$\alpha_0 \leq \alpha_1, \quad \alpha_0 + \alpha_1 = b, \quad \text{and by assumption} \quad \alpha_0 \neq n = \lfloor b/2 \rfloor$$

and together these force $\alpha_1 > \lfloor b/2 \rfloor$, i.e. $\alpha_1 \geq n+1$. But once we know this, we see that:

$$\alpha_1 - 1 \geq n = 2n+1-(n+1) \geq 2n+1-\alpha_1 = 2\lfloor b/2 \rfloor + 1 - \alpha_1 \geq b - \alpha_1.$$

and so we thus have that $\alpha_2 = \max\{\alpha_1 - 1, b - \alpha_1\} = \alpha_1 - 1$. ■

2.4.2 Length 4

The following theorem for length-4 numbers in bases of the form $b = 5 \cdot 2^n$ is proved (using somewhat different terminology) in [3], see Theorem 4.1 in that source. The fact that essentially that entire work is devoted to proving this theorem goes to show how complex the problem of classifying the limiting orbits under the Kaprekar routine really is, even for small lengths.

Theorem 3 (Base $5 \cdot 2^n$)**.** * *If $b \geq 10$ is a positive integer of the form $b = 5 \cdot 2^n$, then:*

i) If n is even, $N(4,b) = 2$. Specifically, if $n \pmod 4 = 0$ then $L(4,b) = [1, n+1]$, and if $n \pmod 4 = 2$ then $L(4,b) = [1, 2(n+1)]$. In particular, in either case the length 1 orbit is given by $[\ [3{\cdot}2^n, 2^n{-}1, 2^{n+2}{-}1, 2^{n+1}]\ ]$. For $n \pmod 4 = 0$, the length $n+1$ orbit is generated by $[2^n-1, 5\cdot 2^n - 1, 5 \cdot 2^n - 1, 2^{n+2}]$; for $n \pmod 4 = 2$, the length $2(n+1)$ orbit is generated by $[2^{n+1}-1, 5{\cdot}2^n-1, 5{\cdot}2^n-1, 3{\cdot}2^n]$.

ii) If n is odd, $N(4,b) = 1$ and $L(4,b) = [1]$. In particular, the singular orbit is given by

$$[\ [3 \cdot 2^n, 2^n - 1, 2^{n+2} - 1, 2^{n+1}]\ ]$$

2.5 Conjectures

2.5.1 Finding Conjectures using Maple

Once we go beyond small lengths and bases, it quickly becomes very difficult to prove specific characterizations of the set of limiting orbits. In particular, most of the patterns that can be found for a specified length l occur for bases of a certain form b. For length 3 we showed that the simple base forms b even or b odd suffice to characterize all limiting orbits. However, in general we often need to break down the bases b into classes based on their prime factorizations, just as for length 4 Theorem 2.4.2 gives a pattern for bases with the prime factorization $b = 5 \cdot 2^n$. Since such base classes grow exponentially, it can be very hard to find these patterns by hand. For example, just to compare the results from the first 6 bases in the above class $b = 5 \cdot 2^n$, i.e. to look at these bases for $0 \leq n \leq 5$, we need to look at numbers in bases as high as $b = 5 \cdot 2^5 = 160$. Thus the use of computer programs is essential in the task of locating

*For the longer orbit in the n even cases, Hasse & Prichett chose a different generator for the orbit than the one given here and confirming, or at least providing empirical evidence in support of, such patterns. Thus in order to aid in the study of the behavior of numbers under the Kaprekar routine, I created the Maple package `6174phenom.txt` (see [5] for full text file, as well as input and output files), whose main features I will now overview.

The central method in this file is `listPhenom(n,b)`, which performs the length-l, base-b-Kaprekar routine on the number `n` given in its length-l, base-b list form, and outputs a list `[m, s, O, L]` where `m` is the number of iterations needed until the first element of the limiting orbit is reached, `O` is that limiting orbit (so in my previous notation, $\texttt{O} = O_{l,b}(\texttt{n})$), $\texttt{s} = |\texttt{O}|$, and `L` is a list giving the entire sequence of integers that the routine iterates through until it reaches the terminating orbit. Note in particular that the length l does not have to be input—it is given implicitly by the length of the list `n`. For example, we have `listPhenom([9,7,2], 10) =`

```
[4, 1, [[4, 9, 5]], [[9, 7, 2], [6, 9, 3], [5, 9, 4], [4, 9, 5], [4, 9, 5]]]
```

In order to look for patterns across several lengths or bases, we need a procedure which will return all of the limiting orbits of the Kaprekar routine for a specified base b and length l, i.e. which will give $Orb(l, b)$. This task is completed by the `allOrbits(l,b,f)` procedure. The inputs `l` and `b` are the length and base, respectively, and `f` is a format string, specifically either "list" or "string" which tells the procedure whether or not to output the orbits in as a list of strings which represent integers (for example the number 495 in base 10 would be represented by the string "495"), or as a list of lists in the format we have described in this paper. The string format is more readable for bases smaller than 10, but for higher bases the list format is much more user and computer-friendly. For example, since as mentioned earlier $Orb(4, 4) = \{ \ [\ [3, 0, 2, 1] \ ], [\ [1, 3, 3, 2], \ [2, 0, 2, 2] \ ]\}$, and thus $L(4, 4) = [1, 2]$, we have that: `allOrbits(4, 4, "list") =` `{[[3, 0, 2, 1]], [[1, 3, 3, 2], [2, 0, 2, 2]]}`.

The `allOrbits` procedure was used as a tool to come up with many of the conjectures in this paper, particularly the odd base length-4 conjectures (see Conjecture 5 below). However, again for patterns that rely on the bases being in some family whose elements grow very rapidly, like in $b = 5 \cdot 2^n$, this procedure quickly becomes impractical. For, as the bases increase the procedure quickly takes a very long time, e.g. running `allOrbits(4, 51, "list")` took several hours to compute. This was even after the procedure had been sped up by accounting for the follwing fact: when looking for all limiting orbits, instead of looking at all $b^l - b$ non-degenerate length-l, base-b numbers, one can look without loss of generality at just the set of numbers whose corresponding lists representations have entry multi-sets are distinct from each other. For example, one need not consider both the numbers $[2, 4, 3, 1]$ and $[3, 1, 2, 4]$. This is a result of the fact that the Kaprekar routine at each iteration sorts the lists representing the numbers into increasing and decreasing form, so that e.g. in base 5 the numbers $[2,4,3,1]$ and $[3,1,2,4]$ will both end up in the same orbit (in particular, $O_{4,5}([2,4,3,1]) = O_{4,5}([3,1,2,4]) = [\ [3,0,3,2]\ ]$). The number of integers with distinct length-l, base-b lists is exactly $\binom{b+l-1}{l}$, since for example there is a bijective correspondence between such lists and the monomials of total degree l in the b variables $v_0, v_2, ...v_{b-1}$, where v_j corresponds to the integer j, which is given by :

$$v_0^{c_1} \cdots v_{b-1}^{c_{b-1}} \text{ s.t. } \sum_{i=0}^{b-1} c_i = l \quad \Leftrightarrow \quad \text{length-}l \text{ list containing } c_j \text{ of the integer } j, \text{ for } 0 \leq j \leq b-1.$$

and we know there are $\binom{b+l-1}{l}$ such distinct monomials. Thus again since we only want non-degenerate numbers, we can reduce from looking at $b^l - b$ numbers to looking at $\binom{b+l-1}{l} - b$ numbers. For example in the case $b = 10$ and $l = 4$ we have reduced from a set of size $10^4 - 10 = 9990$ to one of size $\binom{13}{4} = 705$. Still, since the binomial coefficients grow exponentially for a fixed l, the computation time still increases very rapidly. So, the `allOrbits` procedure quickly becomes impractical as a tool for pattern-finding.

In order to get around this difficulty and still get a large set of data from which patterns can be found, the file `6714phenom.txt` also includes procedures that can generate and analyze large sets of random non-degenerate numbers in base-b and length-l. The procedure `randomEG(l, b)` returns the result of one run of `listPhenom(n,b)`, on a number `n` chosen uniformly at random from the set of non-degenerate base-b length-l numbers. The procedure `randomAnalysis(l,b,t)` then uses the `randomEG` procedure to find all orbits encountered across `t` randomly generated length-`l`, base-`b` numbers. For example, using the `allOrbits` procedure the complete set $Orb(4, 51)$ was found, as mentioned above, in several hours. By comparison, upon running 20 trials of `randomAnalysis(4,51,2000)`, each time the procedure returned the correct set of (44) distinct orbits, in an average time of 3.835 seconds. In particular, in all of these trials checking only $2,000$ numbers, instead of $\binom{54}{4} = 316,251$, sufficed to produce the entire set of orbits.

For each conjecture made below, there is a corresponding "checking" procedure which is run on a given set of orbits and checks to see if, for the specified input base and the length implicitly determined by the input orbits, the orbit set agrees with conjecture for that base and length. For example the checking procedure for the odd-base length-4 conjectures stated below is `checkOddBaseLength4Conjectures(o, b)`) and returns true if the set of orbits `o` agrees with this conjecture.

Now, because of the dramatic increase in speed of the `randomAnalysis` procedure over the `allOrbits` procedure, this random procedure can be used in tandem with the "checking" procedure for a given conjecture to give strong experimental evidence for if the conjecture holds or not. Naturally, a `true` return by the relevent checking procedure on the result of `randomAnalysis(l,b,t)` gives evidence in support of the truth of the conjecture for that l, b which is proportional to how specific the conjecture is. For example, if we have two conjectures for length l numbers, $c1$ and $c2$, and $c1$ only sets forth claims about the number of limiting orbits for a given base and length l, while $c2$ also determines exactly what these orbits will be, then a `true` return from the checking procedure for $c1$ would give less evidence than one for $c2$ in support of their respective conjectures. Also, it is important to note that all orbits produced by the `randomAnalysis` procedure must be a subset of the actual correct set of orbits, and so depending on the specific conjecture under study, this fact could lend even more strength to any experimental evidence found by `randomAnalysis` in support of the conjecture.

The conjectures given below all apply to a specific length l, and for that length, to a set of bases of a specified form (e.g. even bases, or bases which are powers of 2). If someone using the `6174phenom.txt` file wants to determine, for a specific set of lengths `lS` and set of bases `bS`, if the relevant conjectures hold for those bases and lengths, they could use the `getNPCconjecturesPaper(lS, bS)` procedure. The length set `lS` input must be a subset of those lengths with available conjectures, theorems, or propositions—specifically the set $\{3, 4, 6, 7\}$. The output of this procedure call will be a "paper" which has a section for each length in the input set. For each length the procedure checks to see if there are any available conjectures which apply to any of the input bases, and if there are it prints the conjecture itself, followed by a description of the results when the conjecture is applied to each relevant base. In particular it states whether or not the conjecture holds for these bases. In determining this, it utilizes the *verification* procedure corresponding to the conjecture, which itself uses the `allOrbits` procedure in tandem with the relevant checking procedure to determine if the conjecture is true for that base. For example, the odd-base length-4 conjecture, which had "checking" procedure `checkOddBaseLength4Conjectures`, has verification procedure `verifyOddBaseLength4Conjectures`.

Another question one might have regarding the Kaprekar routine is how many iterations it usually takes a number to reach a limiting orbit. Recall from above that this number of iterations is the first output of the procedure `listPhenom(n,b)`. In order to study this, the procedure `getIterationPGF(l,b,x)` outputs a probability generating function $\sum_m c_m \cdot x^m$, where c_m is the fraction of non-degenerate base-`b`, length-`l` integers which took m iterations to reach their limiting orbit.

To study what this distribution looks like, we use the procedure `moms(f,x,i)`, which takes as arguments a probability generating function `f` (in particular, we use `f = getIterationPGF(l,b,x)`), the symbol `x`, and an integer `i` between 1 and 6, and returns the mean μ if $i = 1$, the standard deviation σ if $i = 2$, and the i-th standardized moment $\frac{\mu_k}{\sigma^k}$ for $3 \leq i \leq 6$. Since the `getIterationPGF(l,b,x)` procedure tests over every possible base-`b`, length-`l` integer, we again have the problem that for bases and lengths that are not very small, the procedure takes a long time to run. Thus we also have the procedure `getSampleIterationPGF(l,b,x,N)` which uses the `randomEG` procedure to test over `N` random base-`b`, length-`l` integers.

Now, we move on to the conjectures found using `6174phenom.txt`.

2.5.2 Length 4

Most of the conjectures found using this code were about length-4 numbers. As a result, before we get to these conjectures, we give the following preliminary proposition which details how one iteration of the Kaprekar function behaves for length-4 numbers.

Proposition 2. *Suppose $\eta = [\eta_0, \eta_1, \eta_2, \eta_3]$, where not all entries of η are the same, and let $n_3 \geq n_2 \geq n_1 \geq n_0$ be a listing of the entries of η in non-increasing order. Let $\delta_0 = \delta_0(\eta)$ be the outer difference $\delta_0 = n_3 - n_0$ and let $\delta_1 = \delta_1(\eta)$ be the inner difference $\delta_1 = n_2 - n_1$ (so in particular since not all numbers are identical, $\delta_0 \geq 1$). Then for any base b, the successor $K(\eta)$to η under one iteration of the Kaprekar routine is:*

$$K(\eta) = \begin{Bmatrix} [\delta_0, \delta_1 - 1, b - \delta_1 - 1, b - \delta_0] & \text{if } \delta_1 \geq 1 \\ [\delta_0 - 1, b - 1, b - 1, b - \delta_0] & \text{if } \delta_1 = 0 \end{Bmatrix}$$

<u>Proof:</u>
Recall $M(\eta)$ is the length 4 list with the same entries as η but sorted in non-decreasing order and $m(\eta)$ is the corresponding list sorted in non-increasing order. Thus:

$$M(\eta^*) = n_3 \cdot b^3 + n_2 \cdot b^2 + n_1 \cdot b + n_0 \quad \text{and} \quad m(\eta^*) = n_0 \cdot b^3 + n_1 \cdot b^2 + n_2 \cdot b + n_3$$

and so:

$$\textbf{(1)} \qquad M(\eta^*) - m(\eta^*) = \delta_0 \cdot b^3 + \delta_1 \cdot b^2 - \delta_1 \cdot b - \delta_0 = \delta_0 \cdot b^3 + (\delta_1 - 1) \cdot b^2 + (b - \delta_1 - 1) \cdot b + (b - \delta_0)$$

Then since $0 \leq \delta_1 \leq b - 1$ and $1 \leq \delta_0 \leq b - 1$:

<u>Case 1:</u> If $\delta_1 \geq 1$ all of the coefficients of b^k for $0 \leq k \leq 3$ in the rightmost expansion of (1) are between 0 and $b - 1$, and so $K(\eta) = [\delta_0, \delta_1 - 1, b - \delta_1 - 1, b - \delta_0]$

Case 2: If $\delta_1 = 0$, we have that:

$$M(\eta^*) - m(\eta^*) = \delta_0 \cdot b^3 - \delta_0 = \delta_0 \cdot b^3 - b^2 + (b-1)\cdot b + (b-\delta_0) = (\delta_0 - 1)\cdot b^3 + (b-1)b^2 + (b-1)\cdot b + (b-\delta_0)$$

and so since $\delta_0 \geq 1$, $K(\eta) = [\delta_0 - 1, b-1, b-1, b-\delta_0]$ ■

Conjecture 1 (Base 2^n). *If $b \geq 4$ is a positive integer of the form $b = 2^n$, then:*

i) If n is even, $N(4,b) = n$. Specifically, writing $n = 2k$ we have that

$$L(4,b) = [k, k+1, 2k, ..., 2k, 2(k+1), ...2(k+1)].$$

where the number of $2k$'s and $2(k+1)$'s is $k-1$.

ii) If n is odd, $N(4,b) = n-1$. Specifically, writing $n = 2k+1$ we have that

$$L(4,b) = [k, k+1, 2k, ..., 2k, 2(k+1), ..., 2(k+1)$$

where the number of $2k$'s and $2(k+1)$'s is k.

Conjecture 2 (Base $3 \cdot 2^n$). *If $b \geq 6$ is a positive integer of the form $b = 3 \cdot 2^n$, then:*

i) If n is even, $N(4,b) = 2$. Specifically, writing $n = 2k$ we have that $L(4,b) = [2k+1, 2(2k+1)]$. In particular, the orbit of length $2k+1$ is generated by $[2^n - 1, 3\cdot 2^n - 1, 3\cdot 2^n - 1, 2^{n+1}]$ and the orbit of length $2(2k+1)$ is generated by $[2^n, 2^{n-1} - 1, 5\cdot 2^{n-1} - 1, 2^{n+1}]$.

ii) If n is odd, $N(4,b) = 1$. Specifically, writing $n = 2k+1$ we have that $L(4,b) = [6(k+1)]$. In particular, the singular orbit is generated by $[2^n, 2^{n-1} - 1, 5\cdot 2^{n-1} - 1, 2^{n+1}]$.

Conjecture 3 (Base $7 \cdot 2^n$). *If $b \geq 14$ is a positive integer of the form $b = 7 \cdot 2^n$, then:*

i) If $n \pmod 3 = 0$, $N(4,b) = 2$. Specifically, writing $n = 3k$ we have that $L(4,b) = [3, 3k+1]$. In particular, orbit of length 3 is generated by $[3\cdot 2^n, 2^n - 1, 3\cdot 2^{n+1} - 1, 2^{n+2}]$ and the orbit of length $3k+1$ is generated by $[2^n - 1, 7\cdot 2^n - 1, 7\cdot 2^n - 1, 3\cdot 2^{n+1}]$.

ii) If $n \pmod 3 \neq 0$, $N(4,b) = 1$ and $L(4,b) = [3]$. In particular, the singular orbit is generated by:

$$[3\cdot 2^n, 2^n - 1, 3\cdot 2^{n+1} - 1, 2^{n+2}].$$

In particular an easy application of Proposition 2 gives us the following component of Conjecture 3:

Proposition 3. *If $l = 4$ and $b = 7 \cdot 2^n$ then $[3 \cdot 2^n, 2^n - 1, 3 \cdot 2^{n+1} - 1, 2^{n+2}]$ generates a length 3 orbit.*

Proof:
Recall from Proposition 1 that $K_b(\eta) = [\delta_0(\eta), \delta_1(\eta) - 1, b - \delta_1(\eta) - 1, b - \delta_0(\eta)]$ if $\delta_1 \geq 1$, and write:

$$\eta = [3 \cdot 2^n, 2^n - 1, 3 \cdot 2^{n+1} - 1, 2^{n+2}] = [3 \cdot 2^n, 2^n - 1, 6 \cdot 2^n - 1, 4 \cdot 2^n]$$

Then since $6 \cdot 2^n - 1 \geq 4 \cdot 2^n \geq 3 \cdot 2^n \geq 2^n - 1$, we have that $\delta_0(\eta) = 5 \cdot 2^n$ and $\delta_1(\eta) = 2^n \geq 1$. So:

$$\eta_1 = K_b(\eta) = [5 \cdot 2^n, 2^n - 1, 7 \cdot 2^n - 2^n - 1, 7 \cdot 2^n - 5 \cdot 2^n] = [5 \cdot 2^n, 2^n - 1, 6 \cdot 2^n - 1, 2 \cdot 2^n]$$

Then $\delta_0(\eta_1) = 5 \cdot 2^n$ and $\delta_1(\eta_1) = 3 \cdot 2^n \geq 1$, so that $\eta_2 = K_b(\eta_1) = [5 \cdot 2^n, 3 \cdot 2^n - 1, 4 \cdot 2^n - 1, 2 \cdot 2^n]$. Finally, then $\delta_0(\eta_2) = 3 \cdot 2^n$ and $\delta_1(\eta_2) = 2^n \geq 1$, so that $\eta_3 = K_b(\eta_2) = [3 \cdot 2^n, 2^n - 1, 6 \cdot 2^n - 1, 4 \cdot 2^n] = \eta$, and thus η generates an orbit of length 3. ■

Conjecture 4 (Base 3^n)**.** *If $b \geq 9$ is a positive integer of the form $b = 3^n$, then $N(4, b) = \sum_{i=1}^{n-1} 3^i - 1$. In particular $L(4, b) = [3, 3, ..., 3^k, ..., 3^k, ..., 3^{n-1}, ..., 3^{n-1}]$, where k ranges from 1 to $n - 1$ and the integer 3^k appears $3^k - 1$ times in $L(4, b)$*

Conjecture 5 (Odd Base)**.** *If $b \geq 5$ is an odd positive integer s.t. $b = 2n + 1$, then:*

i) Parity Conjecture*: If $n \pmod 4 \in \{0, 1\}$, then $N(4, b)$ is even, while if $n \pmod 4 \in \{2, 3\}$, then $N(4, b)$ is odd.*

ii) Orbit Lengths Division Conjecture*: Let $\mu = \max\{l : l \in L(4, b)\}$, i.e. μ is the maximum length of a limiting orbit. Then each integer in $L(4, b)$ divides μ.*

iii) Elements Conjecture: *Let U be the set of all 4-digit numbers which occur in some limiting orbit for base b and length 4, then:*

$$U = \bigcup_{k=1}^{n-1} \bigcup_{i=0}^{k-1} \{\ [\ 2k + 1, 2i, 2(n - i) - 1, 2(n - k)\ ]\ \}$$

In particular[†]*, $|U| = \binom{n}{2}$.*

2.5.3 Length 6

Proposition 4 (Even Base)**.** *If b is an even integer, so $b = 2n$ for $n \geq 1$, then the number:*

$$\eta = [n, n - 1, 2n - 1, 2n - 1, n - 1, n]$$

[†] In the `6174phenom.txt` file, the procedure `correspondence(n,k,m)` gives the bijective counterpart in the set U of the 2-subset `{k,m}` of `[n]`.

has $[\eta] \in Orb(6,b)$, *and so* η *is a fixed point under the Kaprekar function, i.e.* $K(\eta) = \eta$.

<u>Proof:</u>
We have that $2n-1 \geq 2n-1 \geq n \geq n \geq n-1 \geq n-1$ and so:

$$\begin{array}{rrcccccc} & M(\eta^*) = & (2n-1)\cdot b^5 & + (2n-1)\cdot b^4 & + n\cdot b^3 & + n\cdot b^2 & + (n-1)\cdot b & + n-1 \\ - & m(\eta^*) = & (n-1)\cdot b^5 & + (n-1)\cdot b^4 & + n\cdot b^3 & + n\cdot b^2 & + (2n-1)\cdot b & + 2n-1 \\ \hline & K_b(\eta^*) = & n\cdot b^5 & + n\cdot b^4 & + & + & - n\cdot b & - n \\ & = & n\cdot b^5 & + (n-1)\cdot b^4 & + (b-1)\cdot b^3 & + (b-1)\cdot b^2 & + (b-n-1)\cdot b & + b-n \end{array}$$

Thus $K(\eta) = [n, n-1, b-1, b-1, b-n-1, b-n] = [n, n-1, 2n-1, 2n-1, n-1, n] = \eta$. ■

2.5.4 Length 7

Conjecture 6 (Base 0 (mod 4))**.** *If* b *is an integer of the form* $b = 4n$ *for* $n \geq 6$*, then* $N(7,b) = 2$ *and* $L(7,b) = [1, 19]$ *In particular, the length 1 orbit is given by:*

$$[\,[3n, 2n, n-1, 4n-1, 3n-1, 2n-1, n]\,]$$

and the length 19 orbit is generated by:

$$[3n-1, 2n-2, n-2, 4n-1, 3n, 2n+1, n+1]$$

Conjecture 7 (Base 2 (mod 4))**.** *If* $b = 4n+2$ *is an integer of the form* $b = 4n+2$ *for* $n \geq 4$*, then* $N(7,b) = 1$ *and* $L(7,b) = [4]$*. In particular, the single limiting orbit is given by:*

$$[\,[3n+1, 2n, n-2, 4n+1, 3n+2, 2n+1, n+1], [3n+3, 2n+1, n, 4n+1, 3n, 2n, n-1],$$

$$[3n+2, 2n+3, n-1, 4n+1, 3n+1, 2n-2, n], [3n+2, 2n+2, n+2, 4n+1, 3n-2, 2n-1, n]\,].$$

Chapter 3
The Involution Principle

3.1 Background

The Involution principle, in its initial form, was presented by Garsia and Milne (see [2]) in their 1981 article *A Rogers–Ramanujan Bijection*, where it was used, as the title suggests, for a combinatorial proof of the Rogers–Ramanujan identities. This principle is reviewed in a modified form in Zeilberger's *Enumerative and Algebraic Combinatorics* (see [6]) as part of an overview of bijective methods, and it is this version that will be presented and analyzed here, and is now formulated as follows:

Let A and B be finite sets of the same size, and $A' \subset A, B' \subset B$ be proper subsets of these sets, also with the same size, i.e. $|A| = |B| = n$ and $|A'| = |B'| = k$ for some positive integers $k < n$. Suppose we have the natural bijections $f : A \to B$ and $g : A' \to B'$. Then we know *a priori* that the complementary subsets $\overline{A'} = A \setminus A'$ and $\overline{B'} = B \setminus B'$ also have the same size, specifically $|\overline{A'}| = |\overline{B'}| = n - k$. Thus there must be a bijection between these sets as well, but how to construct it? The following version of the Involution principle answers this question.

Theorem 4. *An explicit bijection $h : \overline{A'} \to \overline{B'}$ is given by applying the following iterative process, for each $a \in \overline{A'}$:*

Let $a^{(1)} = a$, and let $b^{(1)} = f(a^{(1)}) = f(a)$, so in particular $b^{(1)} \in B$. Then if $b^{(1)} \in \overline{B'}$ we terminate the process and set $h(a) = b^{(1)}$. Otherwise, $b^{(1)} \in B'$, and so we can use the bijection $g : A' \to B'$ to set $a^{(2)} = g^{-1}(b^{(1)}) \in A'$ and then set $b^{(2)} = f(a^{(2)}) = f \circ g^{-1} \circ f(a) \in B$. Again, if $b^{(2)} \in \overline{B'}$ we set $h(a) = b^{(2)}$. Otherwise $b^{(2)} \in B'$, so that it is in the domain of g^{-1} and we can set $a^{(3)} = g^{-1}(b^{(2)})$. In general, if we have not terminated before the i-th iteration of the process then we reach $a^{(i)} \in A'$. At this point we set:

$$b^{(i)} = f(a^{(i)}) \in B \quad \text{((‡) note further that } b^{(i)} = (f \circ g^{-1})^{(i-1)} \circ f(a)\text{)}$$

Then, if $b^{(i)} \in \overline{B'}$, we terminate and let $h(a) = b^{(i)}$. Otherwise, $b^{(i)} \in B'$, so it is in the domain *of g^{-1} and we can begin the $(i+1)$-st iteration by defining:*

$$a^{(i+1)} = g^{-1}(b^{(i)}) \in A'$$

Eventually for some $j \leq n$ we must have $b^{(j)} \in \overline{B'}$, at which point we set $h(a) = b^{(j)}$.

Proof:
We will first show that for each $a \in \overline{A'}$ the procedure indeed terminates in at most n iterations. Then, we will show that the resulting map is in fact a bijection.

Proof the process terminates in at most n iterations:

This follows from the fact that, assuming we have not terminated before the $i-th$ step, $a^{(i)} \neq a^{(j)}$ and $b^{(i)} \neq b^{(j)}$ for any $j \leq i-1$, that is, each $a^{(i)}$ is a distinct element from all those that come before it in the sequence $a^{(1)}, a^{(2)}, ...a^{(i-1)}$, and the same holds for $b^{(i)}$. Thus since the sets A and B are finite, we must eventually reach an element $b^{(j)} \in \overline{B'}$ (in fact, this must happen within the first $k+1 \leq n$ iterations). The fact that $a^{(i)}$ and $b^{(i)}$ are distinct from their predecessors can be seen as follows:

First note $(*)$ that for each i, once we have shown that $a^{(i)} \neq a^{(j)}$ for any $j < i$, it follows that $b^{(i)} \neq b^{(j)}$ for any $j < i$, since for each $j < i$ we have by construction that $b^{(i)} = f(a^{(i)})$ and $b^{(j)} = f(a^{(j)})$, so since f is bijective $a^{(i)} \neq a^{(j)} \Rightarrow b^{(i)} \neq b^{(j)}$. The remainder of the proof follows by induction:

For the base case $i = 2$, we have $a^{(2)} \in A'$ while $a^{(1)} = a \in \overline{A'}$ by assumption. For general i, we assume the property holds for all $j \leq i-1$. Then we have that:

$$a^{(i)} = g^{-1}(b^{(i-1)}) \in A',$$

$$a^{(1)} = a \in \overline{A'}, \quad \text{and}$$

$$a^{(j)} = g^{-1}(b^{(j-1)}) \in A' \quad \text{for any } j \quad \text{s.t. } 2 \leq j \leq i-1$$

Thus $a^{(i)} \neq a^{(1)}$ since they are members of complementary sets, while $a^{(i)} \neq a^{(j)}$ for j s.t. $2 \leq j \leq i-1$ since by the inductive hypothesis and the note $(*)$ above we have that $b^{(i-1)} \neq b^{(j-1)}$, and so since g is a bijection $a^{(i)} = g^{-1}(b^{(i-1)}) \neq g^{-1}(b^{(j-1)}) = a^{(j)}$. ■.

Proof h is a bijection:

First note that since h is a map between two finite sets, it suffices to show that h is one-to-one. Now, assume for contradiction that h is not one-to-one, so there exist an $a, \hat{a} \in \overline{A'}$ s.t. $a \neq \hat{a}$ and $h(a) = h(\hat{a})$. Now, by construction there exist non-negative integers s, t (without loss of generality $s \geq t$) such that:

$$h(a) = (f \circ g^{-1})^{(s)} \circ f(a) \quad \text{and} \quad h(\hat{a}) = (f \circ g^{-1})^{(t)} \circ f(\hat{a})$$

Thus since $h(a) = h(\hat{a})$ and f, g are bijections, we have that

$$f \circ (g^{-1} \circ f)^{(s-t)}(a) = (f \circ g^{-1})^{(s-t)} \circ f(a) = f(\hat{a})$$

and so letting $m = s - t \geq 0$ we have that since f is a bijection, $(g^{-1} \circ f)^{(m)}(a) = \hat{a}$. Since $a \neq \hat{a}$ we have that $m \geq 1$. But then the composite function $(g^{-1} \circ f)^{(m)}$ is a map from the domain of f to the codomain of g^{-1}, i.e. $(g^{-1} \circ f)^{(m)} : A \to A'$. Thus $(g^{-1} \circ f)^{(m)}(a) \in A'$ while $\hat{a} \in \overline{A'}$, and so it is impossible that $(g^{-1} \circ f)^{(m)}(a) = \hat{a}$, and this contradiction means our assumption that h is not one-to-one must be false. ■

3.2 Using Maple to Analyze the Involution Principle

For small n and a given bijection $g : A' \to B'$ the corresponding bijection h can easily be constructed by hand. However, as n and k grow this becomes impractical very quickly, and certainly any large-scale experimental study of the behavior of this iterative process for large n requires the aid of a computer. Thus, in order to analyze the behavior of the above described procedure, the file `InvolutionPrinciple.txt` was created.

3.2.1 Notation and Basic Maple Procedures

First we set forth the notation used in the central procedure of `InvolutionPrinciple.txt`, which is `compBijection(n, indA', indB', τ, i)`. This procedure fulfills the role of the bijection $h : \overline{A'} \to \overline{B'}$ above. Now, the input `n` is just the same as $n = |A| = |B|$. In order to understand the remaining inputs, we need the following:

Let $A = \{a_1, a_2, ..., a_n\}$ be some enumeration of the set A, and then enumerate B using the bijection f by setting $b_i = f(a_i)$ for $1 \leq i \leq n$. Then $A' = \{a_i \ : \ i \in I_{A'}\}$, and $B' = \{b_i \ : \ i \in I_{B'}\}$ for some index sets $I_{A'} \subset [n]$ and $I_{B'} \subset [n]$ of size $k = |A'| = |B'|$. The inputs `indA'`, `indB'`, and $\boldsymbol{\tau}$ specify the bijection $g : A' \to B'$ (in a way set forth shortly).

Finally, the input `i` is the index of the element of the complementary set $\overline{A'}$ (so $\mathtt{i} \notin I_{A'}$) that we wish to find the bijective counterpart of, and the output of the procedure is this counterpart, i.e. $h(a_i) =$ `compBijection(n, indA', indB', τ, i)`.

Now, the input `indA'` is a list which specifies a certain ordering of the index set $I_{A'}$, and similarly with `indB'` and $I_{B'}$ (so in particular these inputs are both length-k lists of positive integers). For example, suppose $n = 6$, so we have some enumeration $A = \{a_1, a_2, a_3, a_4, a_5, a_6\}$ of A, and B enumerated by $b_i = f(a_i)$. Suppose $A' = \{a_1, a_4, a_5\}$ and $B' = \{b_3, b_4, b_6\}$, so that $k = 3$, $I_{A'} = \{1, 4, 5\}$ and $I_{B'} = \{3, 4, 6\}$. Then we could for choose to represent A' by the list `indA' = [ 4,1,5 ]` and B' by `indB' = [ 6,3,4 ]`. The input $\boldsymbol{\tau}$ is then a permutation on k elements which specifies the bijection $g : A' \to B'$ in terms of the chosen ordered lists `indA'` and `indB'`. This permutation is represented by a list of length k, where `τ[i]= j` if and only if g sends the element of A' whose index is the i-th element of the list `indA'` to the element of B' whose index is the j-th element of the list `indB'` (i.e. $g(a_{\mathtt{indA'[i]}}) = b_{\mathtt{indB'[j]}}$). In particular, continuing with our example above, the permutation `τ = [3, 1, 2]` specifies that since the 1-st element of `indA'` is 4, and $\boldsymbol{\tau}[1] = 3$, a_4 is mapped under g to the element of B' whose index is the 3-rd element of `indB'`, 4, i.e. $g(a_4) = b_4$. Similarly, this permutation tells us that $g(a_1) = g(a_{\mathtt{indA'[2]}}) = b_{\mathtt{indB'[\boldsymbol{\tau}[2]]}}) = b_{\mathtt{indB'[1]}} = b_6$ and $g(a_5) = g(a_{\mathtt{indA'[3]}}) = b_{\mathtt{indB'[\boldsymbol{\tau}[3]]}} = b_{\mathtt{indB'[2]}} = b_3$.

Now, lets apply the procedure described in Theorem 4 to this example to find what the bijection $h : \overline{A'} \to \overline{B'}$ should look like. In this case we have $\overline{A'} = \{2, 3, 6\}$ and $\overline{B'} = \{1, 2, 5\}$, and we worked out above that g maps $a_1 \to b_6$, $a_4 \to b_4$, and $a_5 \to b_3$. Thus applying the iterative process of Theorem 4 to the elements in $\overline{A'}$, we have:

$$a_2 \to f(a_2) = b_2 \in \overline{B'}, \quad \text{so} \quad h(a_2) = b_2.$$

$$a_3 \to f(a_3) = b_3 \notin \overline{B'} \to g^{-1}(b_3) = a_5 \to f(a_5) = b_5 \in \overline{B'} \quad \text{so} \quad h(a_3) = b_5.$$

$$a_6 \to f(a_6) = b_6 \notin \overline{B'} \to g^{-1}(b_6) = a_1 \to f(a_1) = b_1 \in \overline{B'} \quad \text{so} \quad h(a_6) = b_1.$$

Thus we have:

```
compBijection(6, [4, 1, 5], [6, 3, 4], [3, 1, 2], 2) = b2

compBijection(6, [4, 1, 5], [6, 3, 4], [3, 1, 2], 3) = b5

compBijection(6, [4, 1, 5], [6, 3, 4], [3, 1, 2], 6) = b1
```

Often we may want to look at the full bijection h at once instead of just seeing the bijective counterpart of one element of $\overline{A'}$. This is achieved in the maple program by the procedure `fullCompBijection(n, indA', indB', τ)`, where all of the inputs are the same as described for `compBijection`, but now the output gives the full bijection h by outputting the list of ordered pairs $[a_i, b_j] = [a_i, h(a_i)]$ for all elements $a_i \in \overline{A'}$. For example we have:

`fullCompBijection(6, [4, 1, 5], [6, 3, 4], [3, 1, 2])` $=$ `[[`$\mathtt{a}_2, \mathtt{b}_2$`], [`$\mathtt{a}_3, \mathtt{b}_5$`], [`$\mathtt{a}_6, \mathtt{b}_1$`]]`

3.2.2 Complexity Analysis

One way to study the behavior of the iterative process used to define the bijection h on the complementary sets is to look at the number of iterations of the process it takes for the bijective counterpart of an element $a \in \overline{A'}$ to be found. We know that $h(a) = (f \circ g^{-1})^{(m_a - 1)} \circ f(a)$ for some positive integer m_a, and the number of iterations it took for the counterpart of a to be found is exactly this m_a. To study this value m_a we use the procedure `compBijectionComplexity(n, indA', indB', τ,i)`, whose output is a list of two elements. The first gives $h(a_i)$, just as in `compBijection`. The second gives the integer m_a such that $h(a) = (f \circ g^{-1})^{(m_a - 1)} \circ f(a)$. For instance, we saw in the example above that it took 2 iterations to find $h(a_3)$. Thus:

`compBijectionComplexity(6, [4, 1, 5], [6, 3, 4], [3, 1, 2], 3)` $=$ `[`$\mathtt{b}_5, 2$`]`

We will denote the second output of this procedure, i.e. the number of iterations m_a for $h(a)$ to by found, by $c(a)$. To study the complexity of the creation of the entirety of a specific h, we can look at the values:

$$C(h) = \sum_{a \in \overline{A'}} c(a) \quad \text{and} \quad \overline{C(h)} = \frac{C(h)}{|\overline{A'}|}$$

In other words, $C(h)$ is the total number of iterations of the procedure needed to find the bijective counterparts for all of the elements in $\overline{A'}$, and $\overline{C(h)}$ is the average (over the set $\overline{A'}$) number of iterations needed to find $h(a)$ for a single $a \in \overline{A'}$. These values are output by the `fullCompBijectionComplexity` procedure. Specifically, we have:

$[C(h), \overline{C(h)}] =$ `fullCompBijectionComplexity(n, indA', indB', τ)`

For example, for the h constructed above we have:

$$C(h) = c(a_2) + c(a_3) + c(a_6) = 1 + 2 + 2 = 5 \quad \text{and} \quad \overline{C(h)} = \frac{5}{3} \quad \text{and so}$$

`fullCompBijectionComplexity(6, [4, 1, 5], [6, 3, 4], [3, 1, 2])` $= \left[\, 5, \frac{5}{3} \,\right]$

In order to study these complexities we need to look at what values they will take for arbitrary cases, e.g. for any fixed $n = |A| = |B|$ and $k = |A'| = |B'|$, and an arbitrary bijection

$g : A' \to B'$, what are the 'typical' values for $C(h)$ and $\overline{C(h)}$ for the resulting h constructed using the given form of the Involution principle?

In order to study this experimentally we use the procedure `generateEG(n,k)` which generates a "random example". In other words, given the inputs `n` for the size of $|A| = |B|$ and `k` for the size of $|A'| = |B'|$, `generateEG(n,k)`:

1. Chooses the subsets A', B' at random from all `k`-subsets of $|A|, |B|$
2. Given A' and B', chooses a random bijection $g : A' \to B'$
3. Constructs the resulting complementary bijection h
4. Determines the associated complexities $C(h)$ and $\overline{C(h)}$

As before, we enumerate A and B such that $f(a_i) = b_i$ for i from 1 to `n`, and specify the subsets A' and B' by their corresponding index sets $I_{A'}$ and $I_{B'}$. Thus task 1 can be completed by choosing at random two `k`-subsets of `[n]` for $I_{A'}$ and $I_{B'}$.

Next note that while the procedure `fullCompBijection` allows for the specification of the bijection $g : A' \to B'$ by the triple of inputs (`indA'`, `indB'`, $\boldsymbol{\tau}$), without loss of generality we can always take $\boldsymbol{\tau}$ to be the identity permutation, $\boldsymbol{\tau}$ = `[1,2,...,k]`, since then, given any desired bijection g between the `[k]` element sets, we can specify it completely for any ordering of `indA'` by ordering `indB'` such that $b_{\texttt{indB'}[i]} = g(a_{\texttt{indA'}[i]})$. For example the bijection g from the previous example, which sent $a_1 \to b_6, a_4 \to b_4$, and $a_5 \to b_3$, had been specified by the triple (`[4, 1, 5], [6, 3, 4], [3, 1, 2]`), but could also be specified (in fact, more intuitively) by the triple (`[1, 4, 5], [6, 4, 3], [1, 2, 3]`). Thus task 2 can be completed by choosing `indA'` and `indB'` at random from all length-`k` lists which contain each of the integers in `[k]` exactly once, and simply using the identity permutation for $\boldsymbol{\tau}$.

Once tasks 1 and 2 are complete, tasks 3 and 4 can be completed using the procedures `fullCompBijection` and `fullCompBijectionComplexity`, respectively. Then, using these randomly generated examples, we can use the procedure `runTest(n, k, N)` to study the 'typical' complexity when $|A| = |B| =$ `n` and $|A'| = |B'| =$ `k`. This procedure inputs `N`, the number of trials, in addition to `n` and `k`. For each trial, it uses `generateEG` to randomly generate A', B' and a bijection $g : A' \to B'$, and to get the corresponding bijection $h : \overline{A'} \to \overline{B'}$ and the associated total complexity $C(h)$. It then returns the average value of $C(h)$ over all `N` trials. Thus if `N` is very large, we should have:

$$\texttt{runTest(n, k, N)} \to \mathbb{E}[C(h)],$$

where $\mathbb{E}$ denotes the expected value, and we are considering $C(h)$ here as a "random variable" dependent on n and k.

So, what *is* the expected value of $C(h)$?

Well, for an arbitrary n, k and, given these, a random bijection g between two 3-subsets $A' \subset A$ and $B' \subset B$ isomorphic to $[k]$, consider the probability that, for any randomly selected $a \in \overline{A'}$, the number of iterations for $h(a)$ to be found is exactly m, denoted $\Pr[c(a) = m]$. Since we have n total elements in the codomain of $f : A \to B$ and k elements in B', we have $n - k$ elements in $\overline{B'}$. $h(a)$ is found on the first try if and only if $f(a) \in \overline{B'}$. Thus since the subset B' of B was randomly chosen, the probability $h(a)$ is found on the first iteration is exactly $\frac{n-k}{n}$, i.e. $\Pr[c(a) = 1] = \frac{n-k}{n}$.

Now, recall that in general (see the note (‡) in Theorem 4), if we terminate at exactly the m-th step for $m \geq 2$, then we had:

$$b^{(i)} = (f \circ g^{-1})^{(i-1)} \circ f(a) \in B' \qquad \text{for } i < m \qquad \text{and}$$

$$b^{(m)} = (f \circ g^{-1})^{(m-1)} \circ f(a) \in \overline{B'}$$

Thus we have that:

$$\textbf{(1)} \qquad \Pr[c(a) = m] = \Pr[(b^{(1)} \in B') \wedge (b^{(2)} \in B') \wedge ...(b^{(m-1)} \in B') \wedge (b^{(m)} \in \overline{B'})]$$

$$= \Pr[b^{(1)} \in B'] \cdot \Pr[b^{(2)} \in B' \ \mid \ b^{(1)} \in B'] \cdot ... \cdot \Pr[b^{(m)} \in \overline{B'} \ \mid \ (b^{(i)} \in B' \ \text{ for } \ 1 \leq i \leq m-1)]$$

Now, for any $i \leq m$ we reached the i-th step without having terminated because for all $j \leq i-1$, $b^{(j)} \in B'$. Further, we had that $b^s \neq b^t$ for any $s < t \leq i-1$, i.e. Thus at the i-th step we had already found $i-1$ elements in B', and we know that $b^{(i)}$ cannot be any of these. Thus $b^{(i)}$ must be one of the $n-(i-1)$ elements remaining in B, $k-(i-1)$ of which are in B', and $n-k$ of which are in $\overline{B'}$. Again, since the subsets A', B' and the bijection g between them were all chosen randomly, $b^{(i)}$ is equally likely to be any of the remaining elements in B. Thus we see that:

$$\textbf{(2)} \qquad \begin{cases} \Pr[b^{(1)} \in B'] = \frac{k}{n} & \\ \Pr[b^{(i)} \in B' \ \mid \ (b^{(j)} \in B' \ \text{ for } \ 1 \leq j \leq i-1)] = \frac{k-(i-1)}{n-(i-1)} & \text{for } 2 \leq i \leq m-1 \\ \Pr[b^{(m)} \in \overline{B'} \ \mid \ (b^{(i)} \in B' \ \text{ for } \ 1 \leq i \leq m-1)] = \frac{n-k}{n-(m-1)} & \end{cases}$$

And so finally from (1) and (2) we have

$$\textbf{(3)} \qquad \Pr[c(a)=m] = \left(\prod_{i=1}^{m-1} \frac{k-(i-1)}{n-(i-1)}\right) \cdot \frac{n-k}{n-(m-1)} = (n-k)\cdot\left(\prod_{j=0}^{m-1} \frac{1}{n-j}\right)\cdot\left(\prod_{j=0}^{m-2}(k-j)\right)$$

$$= (n-k)\cdot\frac{(m-1)!\cdot\binom{k}{m-1}}{m!\binom{n}{m}} = (n-k)\cdot\frac{\binom{k}{m-1}}{m\binom{n}{m}}$$

Recall that since B' has only k elements, we must have that $c(a) \leq k+1$. At this point we can now determine the expected value of $c(a)$ for an arbitrary $a \in \overline{A'}$ using (3):

$$\textbf{(4)} \qquad \mathbb{E}[c(a)] = \sum_{m=1}^{k+1} \Pr[c(a)=m]\cdot m = \sum_{m=1}^{k+1}(n-k)\cdot\frac{\binom{k}{m-1}}{m\binom{n}{m}}\cdot m = (n-k)\sum_{m=1}^{k+1}\frac{\binom{k}{m-1}}{\binom{n}{m}}$$

Finally, since $\overline{A'}$ has $n-k$ elements and $C(h) = \sum_{a\in\overline{A'}} c(a)$, by linearity of expectation and (4) we have that:

$$\mathbb{E}[C(h)] = \sum_{a\in\overline{A'}} \mathbb{E}[c(a)] = (n-k)\cdot(n-k)\sum_{m=1}^{k+1}\frac{\binom{k}{m-1}}{\binom{n}{m}} = (n-k)^2\sum_{m=1}^{k+1}\frac{\binom{k}{m-1}}{\binom{n}{m}}$$

and so we have derived an expression for $\mathbb{E}[C(h)]$, and can test it against our expectation that `runTest(n, k, N)` $\to \mathbb{E}[C(h)]$. For instance, in the example we've been using in this paper, we had `n` $= 6$ and `k` $= 3$. Now, one trial of `runTest` gives us:

`runTest(6,3, 10000)` $= 5.2473$

while by comparison:

$$\mathbb{E}[C(h)] = (n-k)^2\sum_{m=1}^{k+1}\frac{\binom{k}{m-1}}{\binom{n}{m}}\Bigg|_{\substack{n=6\\k=3}} = 3^2\sum_{m=1}^{4}\frac{\binom{3}{m-1}}{\binom{6}{m}} = 9\left(\frac{\binom{3}{0}}{\binom{6}{1}}+\frac{\binom{3}{1}}{\binom{6}{2}}+\frac{\binom{3}{2}}{\binom{6}{3}}+\frac{\binom{3}{3}}{\binom{6}{4}}\right) = 9\left(\frac{1}{6}+\frac{3}{15}+\frac{3}{20}+\frac{1}{15}\right) = 5.25$$

Similarly if we instead had $k=4$ we get for one trial of `runTest` that:

`runTest(6,4, 10000)` $= 4.6627$

while by comparison:

$$\mathbb{E}[C(h)] = 2^2\sum_{m=1}^{5}\frac{\binom{4}{m-1}}{\binom{6}{m}} = 4\left(\frac{\binom{4}{0}}{\binom{6}{1}}+\frac{\binom{4}{1}}{\binom{6}{2}}+\frac{\binom{4}{2}}{\binom{6}{3}}+\frac{\binom{4}{3}}{\binom{6}{4}}+\frac{\binom{4}{4}}{\binom{6}{5}}\right) = 4\left(\frac{1}{6}+\frac{4}{15}+\frac{6}{20}+\frac{4}{15}+\frac{1}{6}\right) = \frac{14}{3} \approx 4.6667$$

and so we see in both of these cases we get something very close to what we expect, and we could easily do these comparisons for arbitrary values of n and k.

In order to experimentally study the behavior of the complexity for large n and variable $k < n$, we can use `runTest` to generate plots showing how the complexity $C(h)$ varies for a given n as the proportion of elements in the complementary set $\overline{A'}$ increases (i.e. as k decreases). In the below, we normalized the complexity to $\frac{C(h)}{n} = \frac{C(h)}{|A|}$ so that we could accurately compare the behavior for variable n:

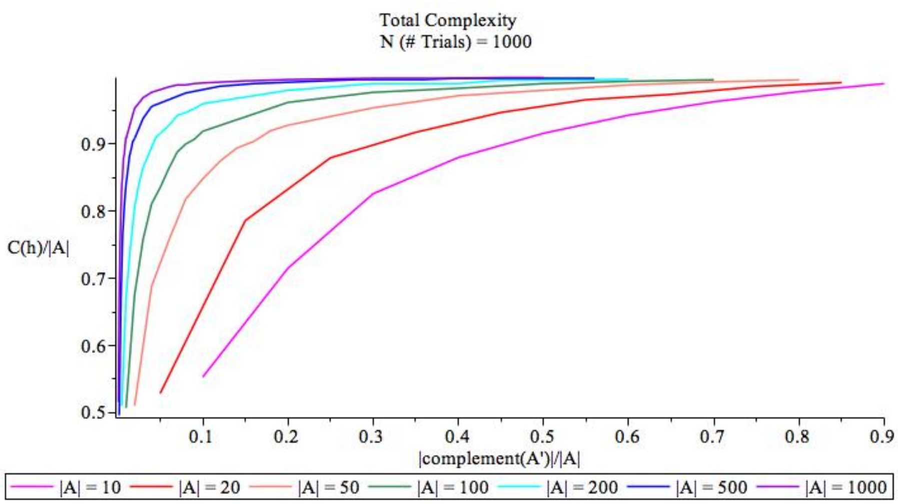

Figure 3.1: Complexity Analysis for variable n between 10 and 1000

Thus we see that the larger $n = |A|$ is, the more rapidly the complexity increases to its full potential ($C(h) = |A|$), i.e. when n is large, we only need a very small proportion of elements to be in $\overline{A'}$ in order to achieve close to the maximum possible complexity for the generation of the bijection $h : \overline{A'} \to \overline{B'}$. For instance, when $|A| = 1000$ and $|\overline{A'}|$ is only 10 (so $|A'| = k = 990$) over $\texttt{N} = 100$ trials, one instance of `runTest` gives us:

$$\texttt{runTest(1000,990,100)} = 924.04, \quad \text{out of a possible 1000.}$$

Lastly, just as in `6174phenom.txt` we had a procedure to create a probability generating function and a procedure to find the mean, standard deviation, and 3-rd to 6-th standardized moments over several random trials, we have analogous procedures in `InvolutionPrinciple.txt` Specifically, `runTestPGF(n,k,N,x)` outputs a probability generating function $\sum_m c_m \cdot x^m$, where c_m is the fraction of `N` independent trials of `generateEG(n,k)` which result in a total complexity $C(h) = m$. Again we can study this distribution using the procedure `moms(f,x,i)`, which takes as arguments a probability generating function `f` (e.g. the output of `runTestPGF`), the symbol `x`, and an integer `i` between 1 and 6, and returns the mean μ if $i = 1$, the standard deviation σ if $i = 2$, and the i-th standardized moment $\frac{\mu_k}{\sigma^k}$ for $3 \leq i \leq 6$.

References

[1] Eldridge, Klaus E., and Sagong, Seok. *"The Determination of Kaprekar Convergence and Loop Convergence of All Three-Digit Numbers.* The American Mathematical Monthly, vol. 95, no. 2, 1988, pp. 105112., www.jstor.org/stable/2323062

[2] Garsia, A.M., and Milne, S.C. , *"A Rogers-Ramanujan bijection"*, Journal of Combinatorial Theory, Series A, vol. 31, Issue 3, 1981, pp. 289-339, ISSN 0097-3165, http://dx.doi.org/10.1016/0097-3165(81)90062-5

[3] Hasse, H., and Prichett, G.D. *"The Determination of All Four-Digit Kaprekar Constants."* Journal Für die reine und angewandte Mathematik, vol. 299/300, 1978, pp. 113-124.

[4] Kaprekar, D.R., *"Another Solitaire Game"*, Scripta Mathematica, vol. 15, 1949, pp. 244-245.

[5] Kukura, Emily L. *"Iterative Methods in Experimental Mathematics."* Web. http://www.math.rutgers.edu/~elk69/masters/mastersHome.html

[6] Zeilberger, Doron. *"Enumerative and Algebraic Combinatorics"*, 2004; in: *Princeton Companion to Mathematics* (Timothy Gowers, ed.), Princeton University Press, 2008, pp. 550-561. http://www.math.rutgers.edu/~zeilberg/mamarim/mamarimhtml/enu.html