

Assignment 1 Write Up

Nicholas Chu

January 12, 2022

1 Code Overview

Below is the code used in the shell script for plot.sh:

```
#!/bin/bash
#Creating all needed temporary .dat files for plotting.
>/tmp/seqlen.dat
>/tmp/seqmax.dat
>/tmp/seqlenhis.dat

for (( n=2; n < 10001; n++))
do

#Step 1: Putting Collatz sequence of n into the temp .dat file collatz.dat.
make clean && make collatz
./collatz -n $n > /tmp/collatz.dat

#Step 2: Getting length of Collatz sequence and along with n, adding it to
#seqlen.dat.
printf "%d " $n >> /tmp/seqlen.dat
wc -l < /tmp/collatz.dat >> /tmp/seqlen.dat

#Step 3: Getting the max value of the Collatz sequence and along with n,
#adding it to seqmax.dat.
printf "%d " $n >> /tmp/seqmax.dat
sort -n -r /tmp/collatz.dat | head -n 1 >> /tmp/seqmax.dat

#Step 4: Getting the frequency of lengths of the produced collatz sequences and along with n,
wc -l < /tmp/collatz.dat >> /tmp/seqlenhis.dat
done

sort -n /tmp/seqlenhis.dat > /tmp/seqlenhis1.dat
uniq -c /tmp/seqlenhis1.dat | sed 's/^ */' > /tmp/seqlenhis2.dat

gnuplot <<END
```

```

set terminal pdf
set output "CollatzSequenceLengths.pdf"
set title "Collatz Sequence Lengths"
set xlabel "n"
set ylabel "length"
plot "/tmp/seqlen.dat" with dots title ""
END

gnuplot <<END
set terminal pdf
set output "MaximumCollatzSequence.pdf"
set title "Maximum Collatz Sequence Value"
set xlabel "n"
set ylabel "value"
set yrange [0:100000]
plot "/tmp/seqmax.dat" with dots title ""
END

gnuplot <<END
set terminal pdf
set style data histogram
set style histogram cluster gap 1
set style fill solid
set output "CollatzSequenceLengthHistogram.pdf"
set title "Collatz Sequence Length Histogram"
set xlabel "length"
set xrange [0:225]
set xtics (0,25,50,75,100,125,150,175,200,225)
set ylabel "frequency"
set yrange [0:200]
set ytics (0,25,50,75,100,125,150,175,200)
plot "/tmp/seqlenhis2.dat" using 1 title ""
END

```

The beginning writes some of the temp files needed for storing the data that is being used to plot. This creates the files if they do not exist yet, or clears them in case they had anything in them, like data from a previous run of plot.sh.

Next is the for loop, which gathers the data from the Collatz sequences produced by the numbers (n) in the range of 2 through 10000 and saves the appropriate data to their corresponding data files. I have separated the contents of the for loop into steps for clarity and to better explain how it works.

- Step 1 of the for loop makes and cleans collatz.c, calls collatz.c with the given n, and then saves the Collatz sequence produced from n into a data file called collatz.dat.
- Step 2 uses the command "printf" to save the current n into the first column of seqlen.dat, and the "wc -l" command takes the length of the collatz sequence stored in the data file

collatz.dat and stores it into the second column. The file, seqlen.dat, will be used to plot the Collatz Sequence Lengths plot.

- Step 3 uses "printf" to save the n, then in the second command first uses "sort -n -r" to first numerically sort the Collatz Sequence of n and reverse it, since it always sorts from least to greatest. Then, still in the same line, using "head -n" it gets the first element, which should now be the largest element in the sequence, and appends it to the data file seqmax.dat. The file seqmax.dat will be used to plot the Maximum Collatz Sequence Value plot.
- Step 4 is just appending only the length of the sequence n into seqlenhis.dat which will be later used to save data into seqlenhis2.dat in order to plot the length-frequency histogram.

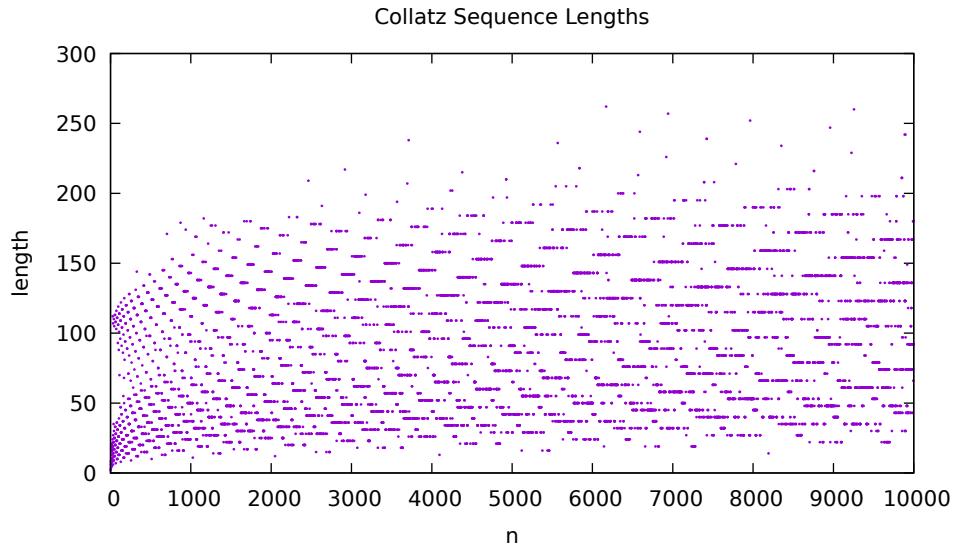
Once out of the for loop, "sort -n" numerically sorts the contents of seqlenhis.dat and puts it into seqlenhis1.dat. Then, using "uniq -c" on seqlenhis1.dat it finds the unique number of lengths and their number of occurrences and saves them into seqlenhis2.dat which is used to plot the length-frequency histogram.

The rest of the code below formats gnuplot to take in the data from the data files, stylize the plots, and output them to pdf files.

2 Results

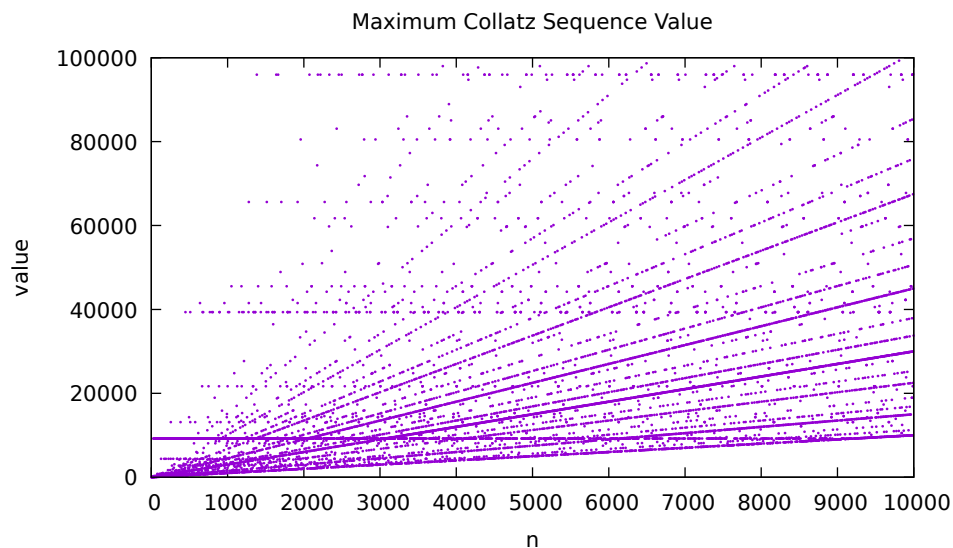
The following subsections show the results of running plot.sh and the plots created by it, along with a brief examination of what these plots imply about the Collatz sequence.

2.1 Sequence Lengths



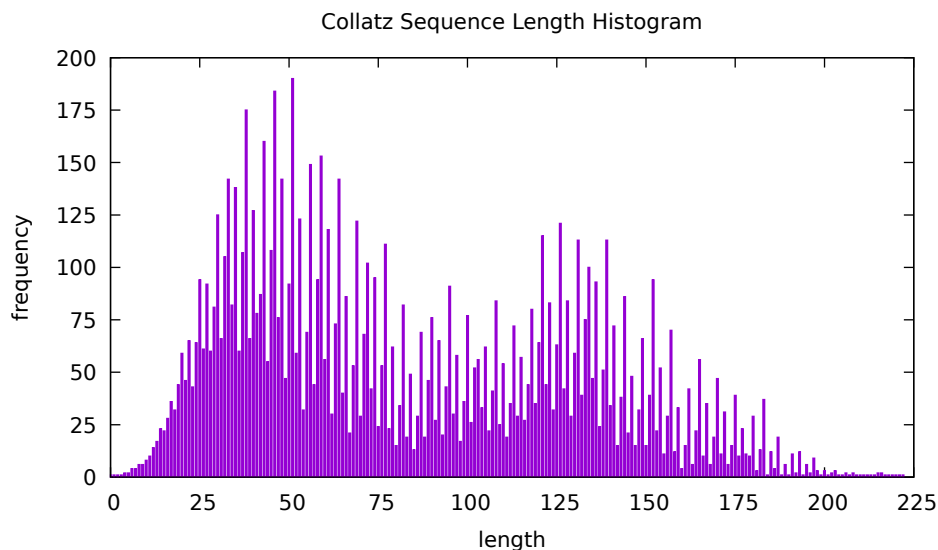
This plot produced by plot.sh represents the lengths of the Collatz sequences produced from 2 to 10,000 by collatz.c. From first looking over the graph, the growth of the sequence lengths appear to be logarithmic, so we can assume that while n increases, we can expect a logarithmic growth in the lengths of the sequences being produced as well.

2.2 Maximum Sequence Value



This plot produced by plot.sh represents the maximum values inside each of the Collatz sequences produced from 2 to 10,000 by collatz.c. From looking over the graph, the growth of the maximum values in the sequences appears to be linearly increasing, so we can assume that while the n increases we can expect to also see a linear increase in the maximum values of the sequences.

2.3 Maximum Sequence Value



This plot shows the frequency of the Collatz sequence lengths produced by the numbers ranging from 2 to 10,000. From this we can see that the most frequent sequence lengths are around 25 and 75. The highest frequency of sequence lengths peaking near length size 50, and after that there

is a general decrease in frequency as the sequence lengths get greater. There is another rise in frequency of sequence lengths starting around the length of 100, which begins to fall again around length 135.

3 Conclusion

Through this assignment I got more practice using UNIX and was able to better familiarize myself with many of the basic commands and how to use them. Along with this I found out what a shell script is and how to create one. When creating the shell script I also learned and practiced how to use data to create plots using gnuplot.