

# Assignment 1 Design Document

Nicholas Chu

January 10, 2022

## Purpose

Need to create a bash script that uses gnuplot to plot the Collatz sequence which is given by collatz.c. The script will need to produce plots that represent the relationship like the Collatz sequence lengths(n and length of sequence), maximum value (n and maximum value in the sequence), and the sequence lengths (length of sequences and frequency of those lengths).

## Files

The files that will need to be included in the directory asgn1:

- plot.sh: Bash script that plots the Collatz sequences produced by collatz.c.
- collatz.c: Provided file that is used to produce Collatz sequences based on an input n.
- Makefile: Provided file that directs the compilation of the Collatz sequence program.
- README.md: Description on how to run the script and Makefile.
- DESIGN.pdf: This current document which describes the design process of the program.
- WRITEUP.pdf: Will include the plots created by plot.sh and a discussion on which UNIX commands were used to produce each.

## Pseudocode

The script, plot.sh, needs to take the Collatz sequences produced from 2 to 10000 by collatz.c and put them into data files, then organizes these data files to produce the appropriate plots by using gnuplot.

Planning to use 5 data files for this process: collatz.dat will be used to store the current sequence produced by collatz.c when taking the input n, seqlen.dat will be used to store n and the length of its sequence for Collatz Sequence Lengths, seqmax.dat will store n and the maximum value of its sequence for Maximum Collatz Sequence Value, seqlenhis1.dat which stores the frequency of sequence lengths, and finally seqlenhis2.dat which uses the lengths from seqlenhis1.dat to make data file that can be used to plot the Collatz Sequence Length Histogram.

For each n from 2 to 10000:

Put the Collatz sequence for current n into a data file collatz.dat.

Get the length of the sequence by using wc (wc -l), then storing n and the result of wc into seqlen.dat.

Get the max value of the sequence through first numerically sorting (sort -nr) collatz.dat and getting maximum value by using head to get the top value, then storing n and the maximum value into seqmax.dat.

Add length of the sequence of n to seqlenhis1.dat by using wc.

Use uniq on seqlenhis1.dat and to find lengths and number of repeated lengths, which is then saved to seqlenhis2.dat. Then set up gnuplot to plot all three plots using seqlen.dat, seqmax.dat, and seqlenhis2.dat.

Diagram below is what the .dat files will hopefully look like when running the script. (I plan on using 5 .dat files in this script, but this might change if I am able to cut that number down or end up needing more.)

for n in the range 2 to 10000:

1. collatz.c with input n writes to Collatz sequence generated to collatz.dat.

collatz.dat:

n

.

.

1

Would completely change every iteration based on n.

2. seqlen.dat:

n length (wc -l collatz.dat)

.

.

.

Would add to existing seqlen.dat

3. seqmax.dat:

n max (sort -nr collatz.dat, head collatz.dat)

.

.

.

Would add to existing seqmax.dat

4. seqlenhist.dat:

length frequency of said length in seqlenhist.dat

.

.

.

Would add to existing seqlenhist.dat

(After for loop.)

seqlenhist2.dat:

length frequency of said length in seqlenhist1.dat

.

.

.