Assignment 1 DESIGN.pdf

Description of Program:

The program will print the Collatz sequence starting from some positive integer n. We create a sequence of integers S starting with some positive integer n. That is, S1 = n. The next term of the sequence is based on the previous term of the sequence.

$$S_{k+1} = \begin{cases} 1+3S_k & \text{if } S_k \text{ is odd} \\ \frac{1}{2}S_k & \text{if } S_k \text{ is even} \end{cases}$$

Purpose:

We should write a script called plot.sh to run the given C program (collaz.c) and create three figures by using the gnuplot.

Files to be included in directory "asgn1":

1. plot.sh: This bash script should produce the Collatz sequence plots used in your report. This script should produce plots similar to Figures 2, 3, and 4. You can create more plots than these if you choose to.

2. collatz.c: This file is provided and contains the implementation of the Collatz sequence program. Has been given in the resources repository.

3. Makefile: This file is provided and directs the compilation process of the Collatz sequence program. Also has been given in the resources repository.

4. README.md: This must use proper Markdown syntax. It must describe how to use your script and Makefile. should also list and explain any command-line options that your program accepts.

5. DESIGN.pdf: This document must be a proper PDF. Describe my design and design process for your program with enough detail. Describe how your program works with supporting pseudocode.

6. WRITEUP.pdf: This document must be a proper PDF. This writeup must include the plots that you produced using your bash script, as well as discussion on which UNIX commands you used to produce each plot and why you chose to use them.

Pseudocode / Structure:

Use the output from the given c file to get the length, Maximum Value of the Sequence and frequency of the sequence length.

Remove -f /tmp/length.dat
Remove -f /tmp/maxvalue.dat
Remove -f /tmp/lengthdots.dat
Remove -f /tmp/frequency.dat
remove the .dat files Just in case I created a file with the same name earlier since I will make them later.

For n in range(2, 10000):
    Put the value of n into /tmp/length.dat
    Also the value of n into /tmp/maxvalue.dat
    Give the positive integer n to colltz and count the length of the output sequence and put it into /tmp/length.dat
    Give the positive integer n to colltz and sort the output sequence to get the maximum value in the sequence and put it into /tmp/maxvalue.dat
    Give the positive integer n to colltz and count the length of the output sequence and put it into /tmp/lengthdots.dat too.
For Loop done

Outside the loop I sort and count how many times the sequence length appears.

This is the heredoc that is sent to gnuplot to create PDF files
figure2.pdf
gnuplot <<END
    set output as pdf files
    set output name is"figure2.pdf"
    set title "Collatz Sequence Lengths"
    set x-label "n"
    set y-label "length"
    set zeroaxis
    Use the points in the "/tmp/length.dat" file to draw a dot plot
END

figure3.pdf
gnuplot <<END
    set output as pdf files
    set output name "figure3.pdf"
    set title "Maximum Collatz Sequence Value"
    set x-label "n"
    set y-label "value"
    set y-range [0:100000]
    set zeroaxis
        Use the points in the "/tmp/maxvalue.dat" file to draw a dot plot

END

```
# figure4.pdf
# still working on it
gnuplot <<END
    set output as pdf files
    set output name "figure4.pdf"
    set title "Collatz Sequence Length Histogram"
    set x-label "frequency"
    set y-label "length"
    set zeroaxis
    use the points in the "/tmp/frequency.dat" file and change the x and
y to draw a box plot

END
```

Notes:
1. How to run collatz ?
./collatz -n 2
2
1

2. What do we mean by length in the Fugure2 ?
For example:
./collatz -n 2
2
1

The length of the sequence is 2

3. What do we mean by Maximun value in The Figure3 ?
For example:
./collatz -n 2
2
1

The Maximun value of the sequence is 2

Error Handing:
1. When I was trying to get the points of figure2, I couldn't find a way
to   separate the x-coordinate from the y-coordinate. What I got is looks
like:
11
22

38
But I actually need is like:
1 1
2 2
3 8
I figure it out by just put a \ after the n. Like: $n\

2. When I want to run plot.sh, I got an error: -bash: ./plot.sh: Permission
denied. I post an ed post to ask for it
https://edstem.org/us/courses/16730/discussion/986520
and then I found the solution in assignment PDF.

3. When I use uniq -c, I accidentally put this command in the loop and
made it run multiple times
which prevented me from getting the correct .dat file.
I look at the man page for uniq and search the way to use set to solve
that problem.

4. Gnuplot keeps giving this error when I plot ".dat" file with histograms
title:
gnuplot> plot "/tmp/frequency.dat" using 2:1 with histograms title ""
line 0: Too many columns in using specification
ed post : https://edstem.org/us/courses/16730/discussion/987660
I solved it by just using the boxes instead of histogram:

plot "/tmp/frequency.dat" using 2:1 with boxes title

and plot the data file with solid filled boxes with a small vertical space
separating them:

set boxwidth 0.9 relative
set style fill solid 1.0

Credit:
1. I learned a lot about how to make a bash file, how does c program works,
how to run a for loop in bash, how to use gnuplot and so on. These are
things I've never been exposed to before. It took me a long time to read
asgn1.pdf because it took me a long time to understand and practice what
I didn't know.

2. I attended Eugene's Lab section on Friday 1/7/22, and he explained
collatz.c and Makefile which provided in resources and he taught us how
to run collatz.c with a Makefile. He also taught us how to use wc -l command
to get the length of a sequence, how to remove the newline created by echo

and how to get the maximum value from a sequence using head or tail.