Lab 11 Report

PHY407 Week 10 Assignment

Landon Wang, Yishi Liu

December 6, 2022

Contribution: Landon Wang writes the entirety of question 2 and completes this document, Yinshi Liu writes the entirety of question 1.

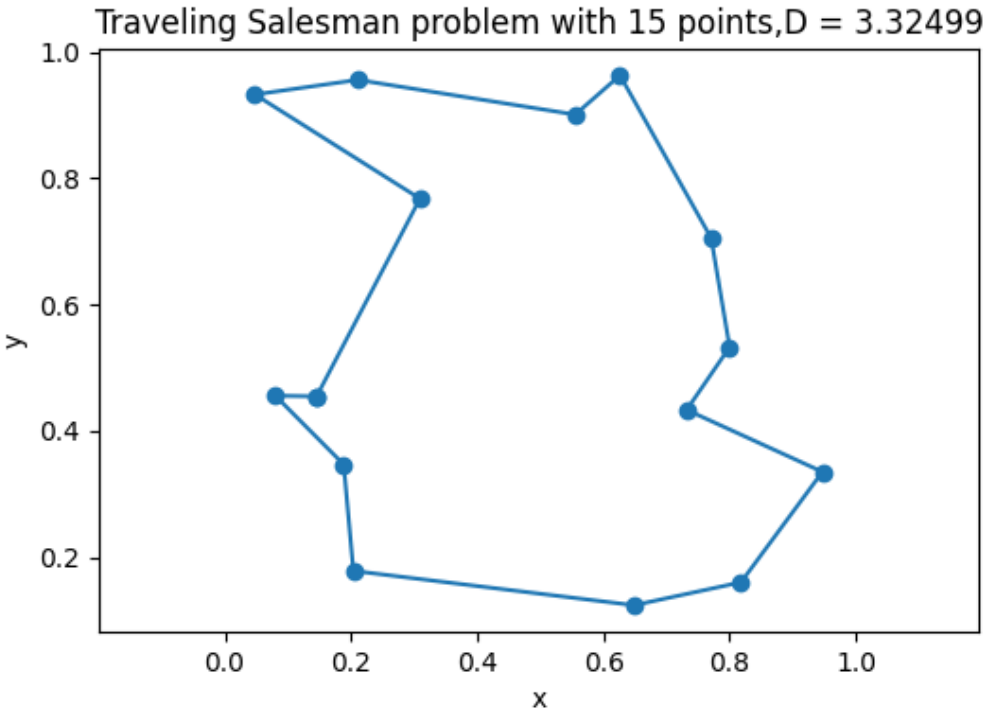
Question 1

PART A

Code Output:

Parameters: Seed = 100, N = 15, Initial Temperature = 10, tau = 10000

Sample Distribution of points and paths



Note: Since there is a large amount of figures in this question, We condensed the feature of each figure into the chart below. The images are attached in a separate appendix.

Table 1.1 Total distance D for different seeds and step sizes

Seed #/Step size	2tau	tau	1/2tau
100	3.24616	3.32499	3.44764
101	3.44764	3.63841	3.63841
102	3.29355	3.29355	3.24616
103	3.24616	3.24616	3.29355
104	3.24616	3.44764	3.24616

Depending on the different seeds for each run, the total distance D varies slightly. We noticed that while there are some minor differences in path among smaller 'clusters' of points, the general shape of the paths remains similar. This means that our algorithm is doing its job by ruling out the less optimal routes. In some cases, the path has signs of imperfections (crisscrossing paths, etc.) which have some minor impacts on D.

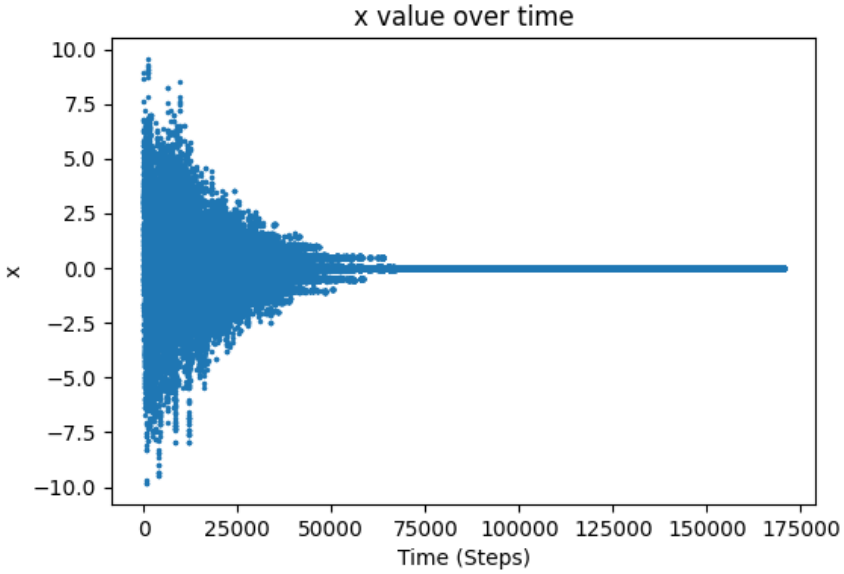
As the chart has shown, an increase in step size would make the temperature of the system cool slower, which increased the chance of the system settling in the global minima rather than the local one. This is indeed the case for our simulation. Out of the 5 trials we have run, as the step size doubles, all runs have ended in a lower or equal energy state than the normal runs.

Similarly, a decrease in step size would allow the system temperature cool faster, which shortens the simulation time and lessens the computation load. On the other hand, a shorter step size allows less time for the system to settle to the global minima, which may cause the system to be 'frozen' in a local minimum without the energy to escape. This is what happened when the step size is halved. Out of the 5 runs, 4 of them stayed at a higher/equal energy level, with one run reaching the lowest energy level. We believe that this run (seed # = 104) has reached the global minima by chance alone and got 'frozen in' as the temperature drops rapidly

PART B

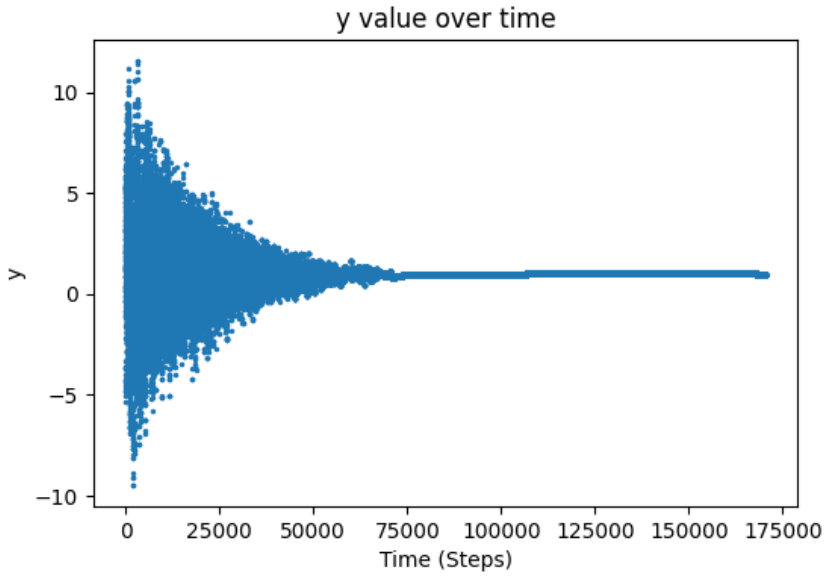
Code Output

Figure 1.1: The value of x over time in SA algorithm



As the system gradually cools, the possible space for x-value decreases exponentially as time evolves. Eventually, the system settles in on one several possible 'states', which are all local minima. This is evidenced by the parallel horizontal lines on the figure. As time evolves further, the x-value settles near x = 1 and remained in this location as temperature cools.

Figure 1.2: The value of y over time in SA algorithm



In this case, the y-value converges towards final state faster. This is caused by a lack of other potential local minima competing with the global minimum.

In conclusion, the final output of x and y is:

```
final x value: -0.0002546534175133018
final y value: 1.0001228920944358
```

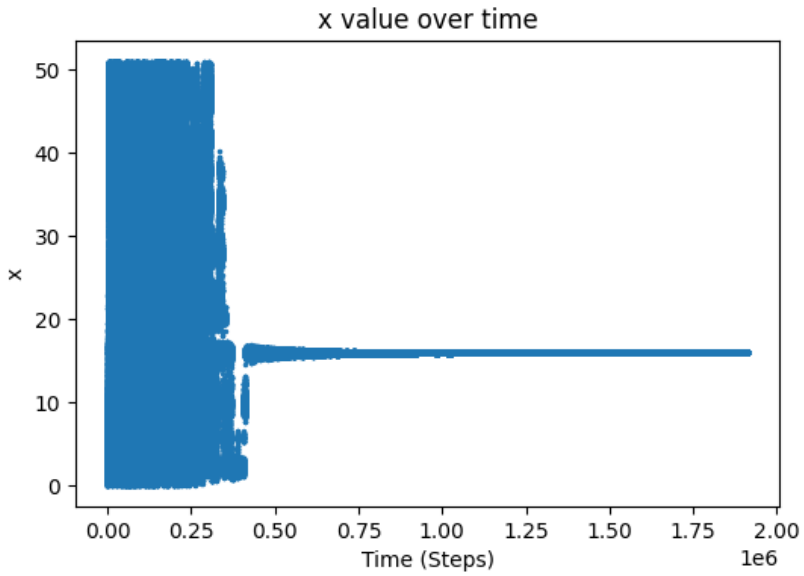
Which is around the theoretical value of (0, 1).

PART C

In this case, determining the global minimum is more difficult. After running multiple trials, we observed that there is an about equal chance of the system settling in competing minima
(42, 1), (2, 1) as well as the correct value (16, 1). In order to make our simulation settle at the correct value, we attempted to increase the time constant tau, as well as to raise the initial temperature (or drop the final temperature) in order to allow the simulation to run longer and move itself out of the local minima. However, the effect is limited. We suppose that is because the values of competing minima has a similar value as the global one.

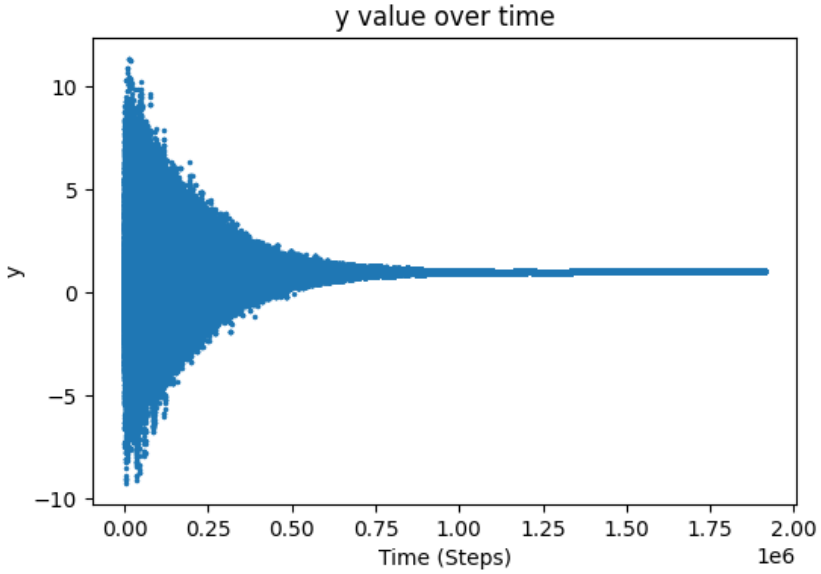
Code Output

Figure 1.3



Note that the shape of the x values does not decrease like Figure 1.1. Instead, x goes through multiple possible states before settling in the correct minimum by chance.

Figure 1.4



In comparison, there are no other competing minima for the y values. Therefore, its behaviour is very similar to that of Figure 1.2.

Finally, the output of the code is:

```
final x value: 15.961447290429982
final y value: 0.9976473493357647
```

which is close to the correct value of (16, 1).

Question 2

PART A

For this part, we run the code for three sets of different initial condition

Case 1: $N = 30, T = 1.5, \epsilon = -5, n = 100000$

Energy averaged over last half of simulations is: -42.95

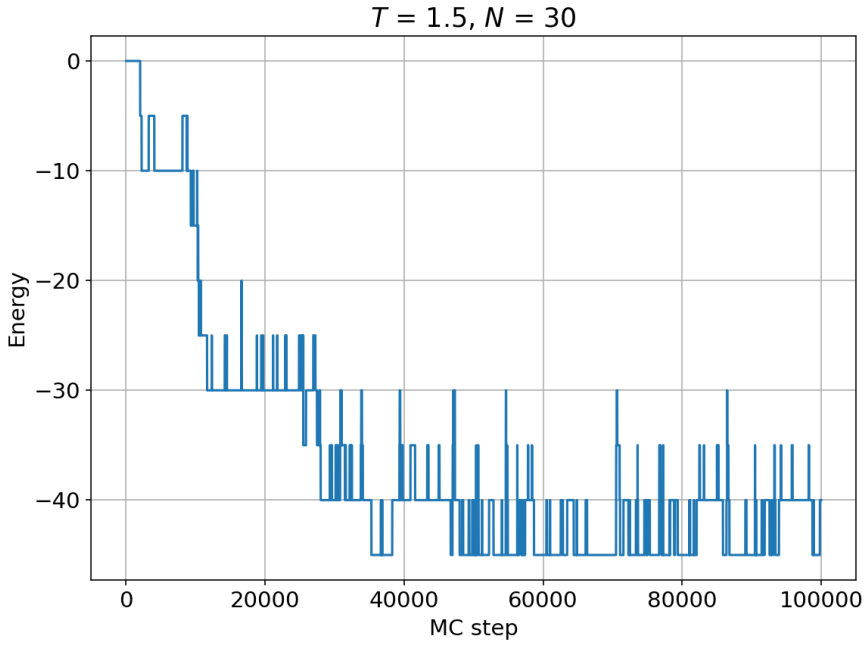


Figure 2.1 Energy plot for a run of initial condition case 1

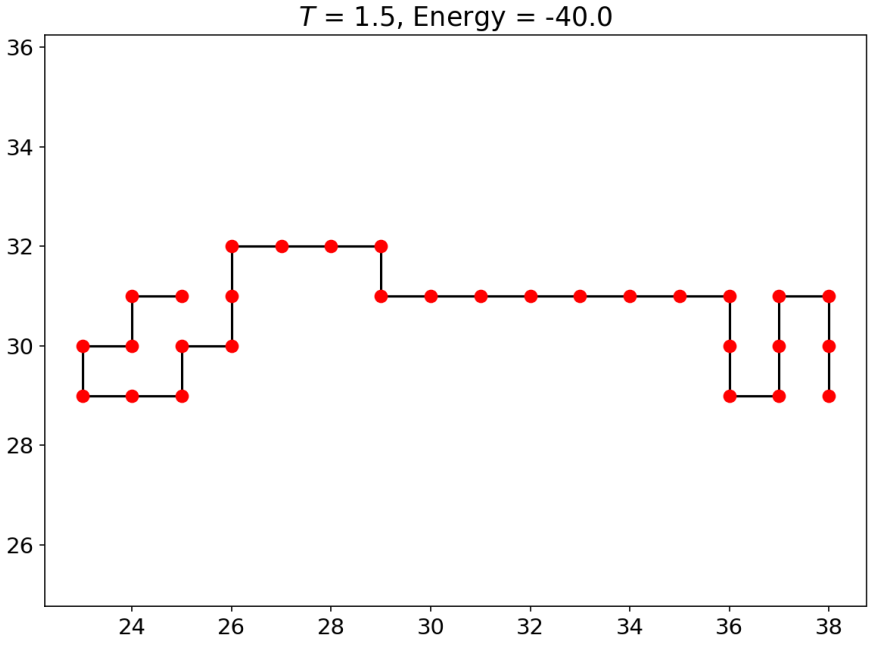


Figure 2.2 Final protein configuration of a run of initial condition case 1

Case 2: $N = 30, T = 0.5, \epsilon = -5, n = 100000$

Energy averaged over last half of simulations is: -10.00

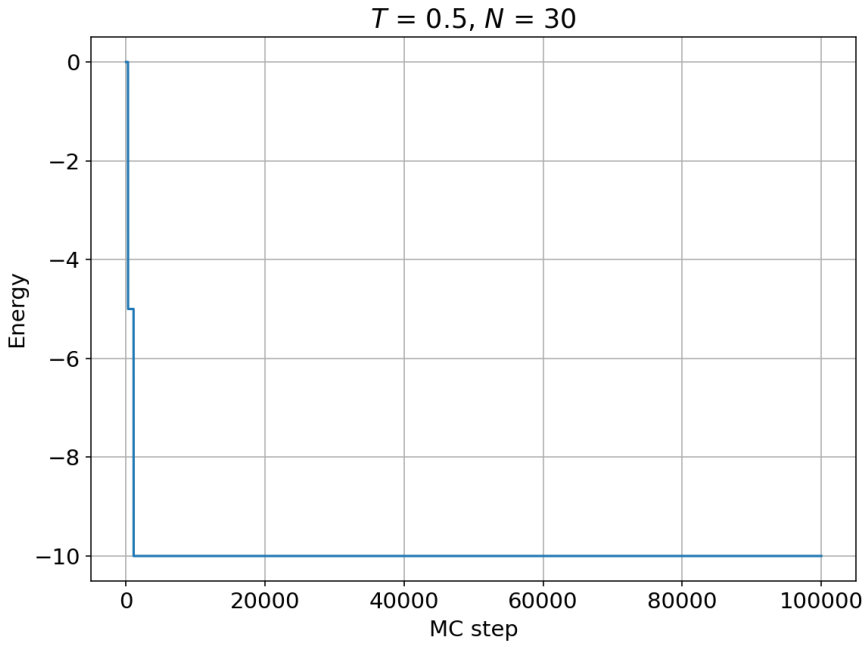


Figure 2.3 Energy plot for a run of initial condition case 2

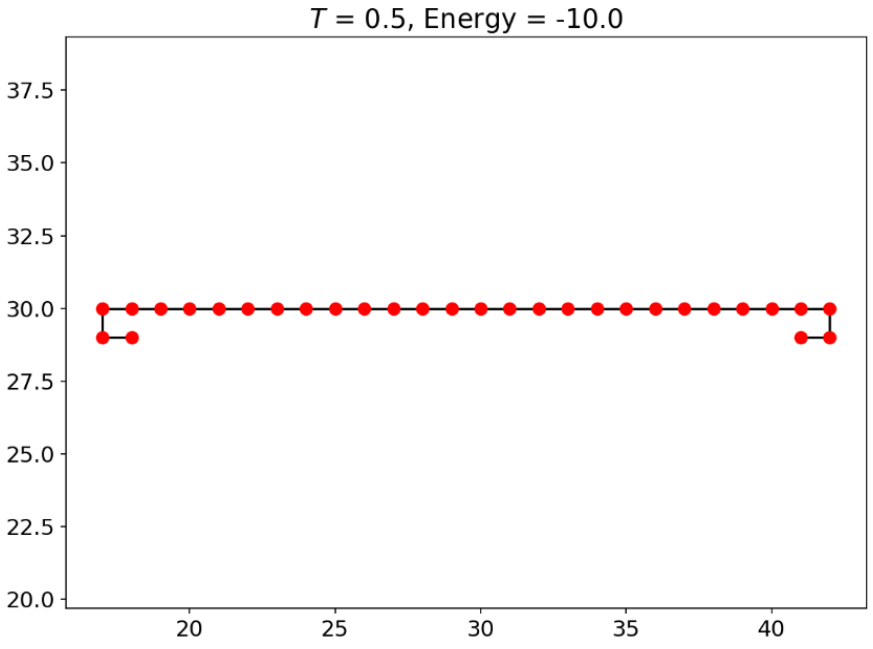


Figure 2.4 Final protein configuration of a run of initial condition case 2

Case 3: $N = 30, T = 5, \epsilon = -5, n = 100000$

Energy averaged over last half of simulations is: -67.44

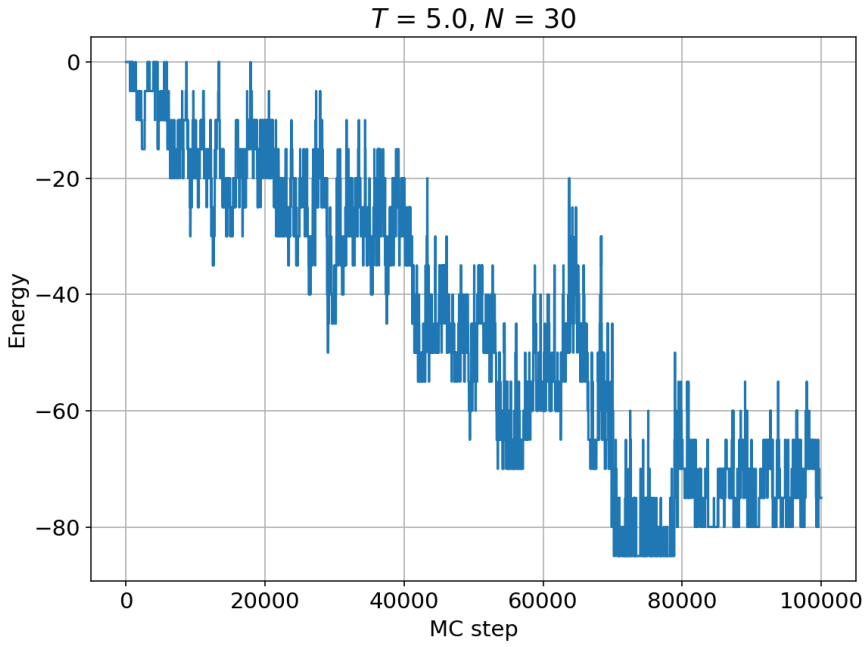


Figure 2.5 Energy plot for a run of initial condition case 3

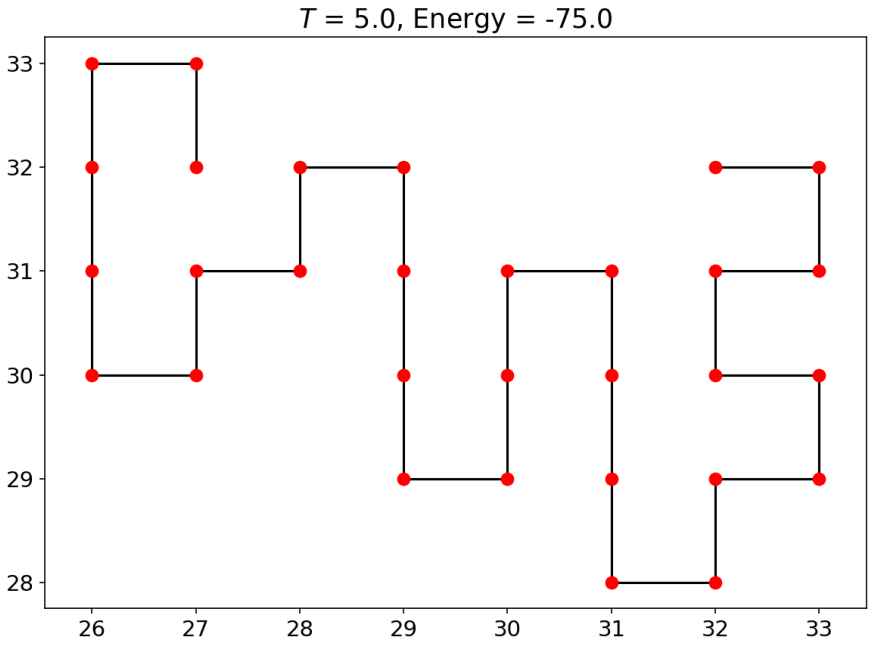


Figure 2.6 Final protein configuration of a run of initial condition case 3

From the most basic thermodynamic perspective, the higher the temperature T, the higher the thermal energy $E = k_B T$, which leads to more probable change of configuration of a long protein chain. Therefore, it should not be unexpected that we have a most simple, energetic, straight like protein configuration with $T = 0.5$, and the most complicated, least energetic tangled protein configuration with $T = 5$.

From the figures 2.1, 2.3, 2.5, we observed a clear trend that with low temperature, the fluctuation in the energy shuttle to a stable equilibrium very fast at probably less than 100 Markov steps as shown in figure 2.3.

For a higher temperature, as showed in figure 2.1, the energy steadily decreases over the first 30000 steps, than flatulate at around $E = -40$ for the rest of simulation.

For the even higher temperature condition as figure 2.5 showed, the decrease in energy extends beyond 75000 steps than strangely increased slightly and maintained at around $E = -75$.

From the algorithm perspective, the reason behind this behaviour is due to the step 4 of the computational background section of the lab manual regarding to this simulation. Here, the algorithm calls that if the change of configuration will be made if the new state has lower energy. Otherwise, a random decision will be made following the Boltzmann factor.

The Boltzmann factor is $\exp -\Delta E/T$ which is increases when T increases (especially in the small numbers of T.) Therefore, given the same ΔE for the change in state, the higher the T, the more likely that a change in configuration will happen.

Based on the analysis, understand that why would the behaviour shown in the figures are making sense algorithmically.

PART B

Run the same code for more simulations, we have the figure 2.7 – 2.10 showing the result with 10 times more steps.

In the figure 2.7 with $T = 1.5$, we do notice the lowered average energy which indicates about 2 more interactions within the protein structure. This is expected because we have given it more time to simulate, the random contribution to the fluctuation of the energy will has less meaningful impact to the result, which is the law of large numbers in statistics.

Notice that for this simulation, the protein energy stay at around $E = -30$ when the first 100000 steps finished which is higher than the simulation in figure 2.1, but the system did reduce the energy further and ends up at around step 350000 which it maintained at around $E = -50$.

In the other case, figure 2.9 shows the similar result as figure 2.3, that is, very stable configuration for a very long time. Something different here is that a random increase energy at about step 300000 and suddenly reduced the energy to an even lower state at $E = -15$ which is the last successful configuration change. This event is likely due to a successful change in configuration at protein #5 or #6 then followed by another successful configuration change to the other one of the two, resulting the configuration we see at the end.

In the few attempts on producing figure 2.3, we notice the similar behaviour, that a random occasion of continues successive configuration change would lead to a lower energy state given the low probability of any successful changes.

In summery, the extension of simulation provides more opportunities for low probability events to happen, which results a lower energy state.

Case 1: $N = 30, T = 1.5, \epsilon = -5, n = 1000000$

Energy averaged over last half of simulations is: -53.64

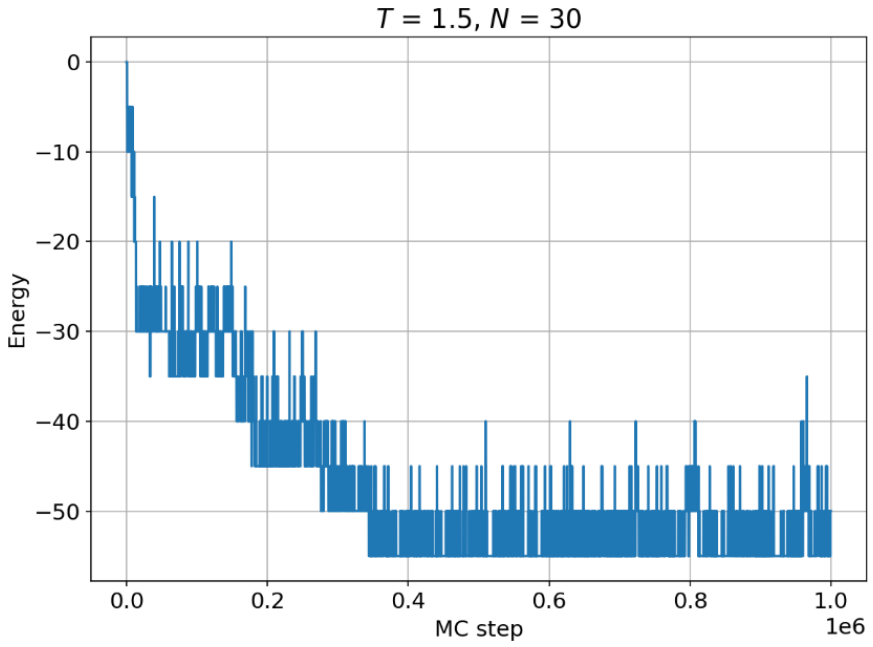


Figure 2.7 Energy plot for a run of initial condition case 1

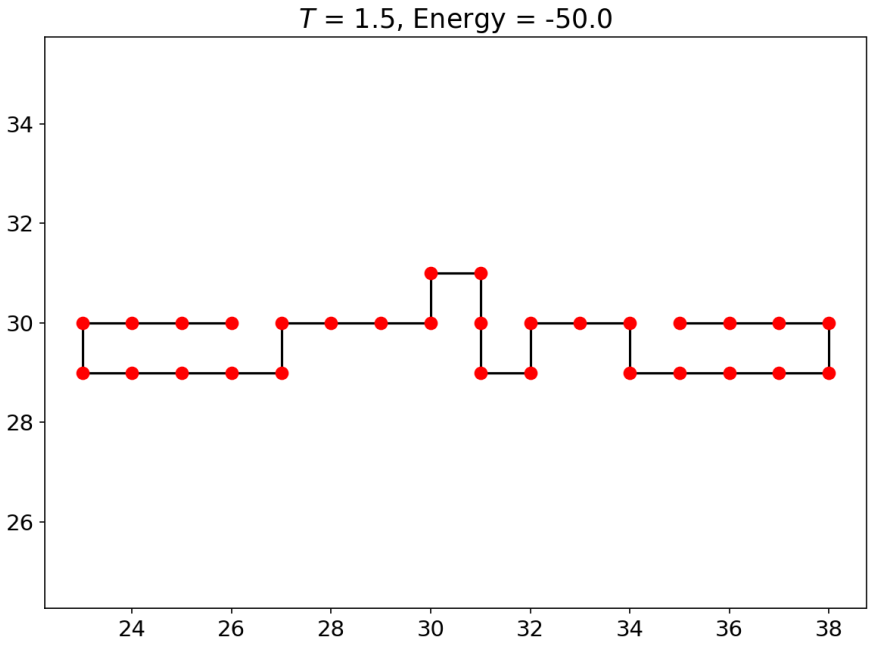


Figure 2.8 Final protein configuration of a run of initial condition case 1

Case 2: $N = 30, T = 0.5, \epsilon = -5, n = 1000000$

Energy averaged over last half of simulations is: -15.00

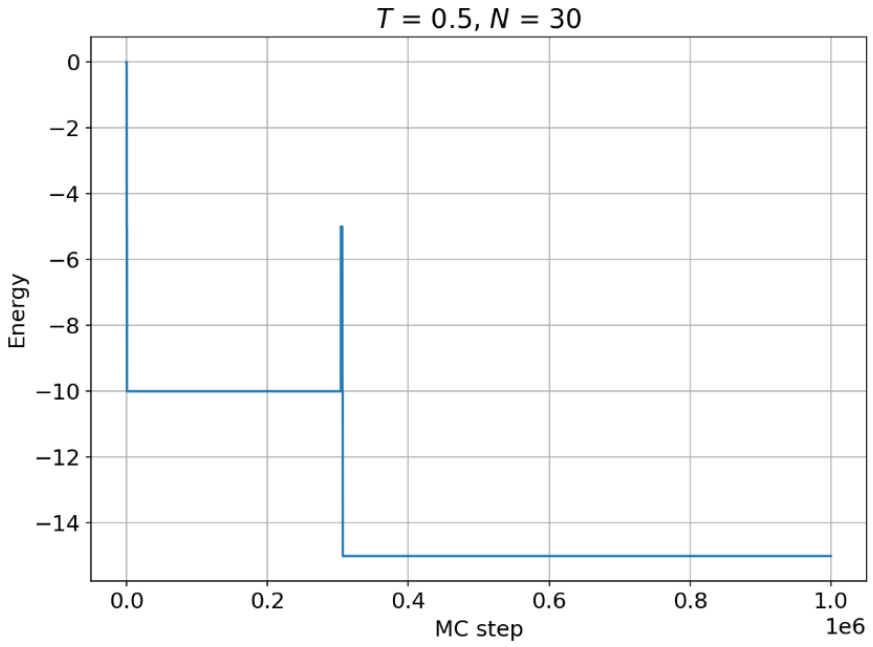


Figure 2.9 Energy plot for a run of initial condition case 2

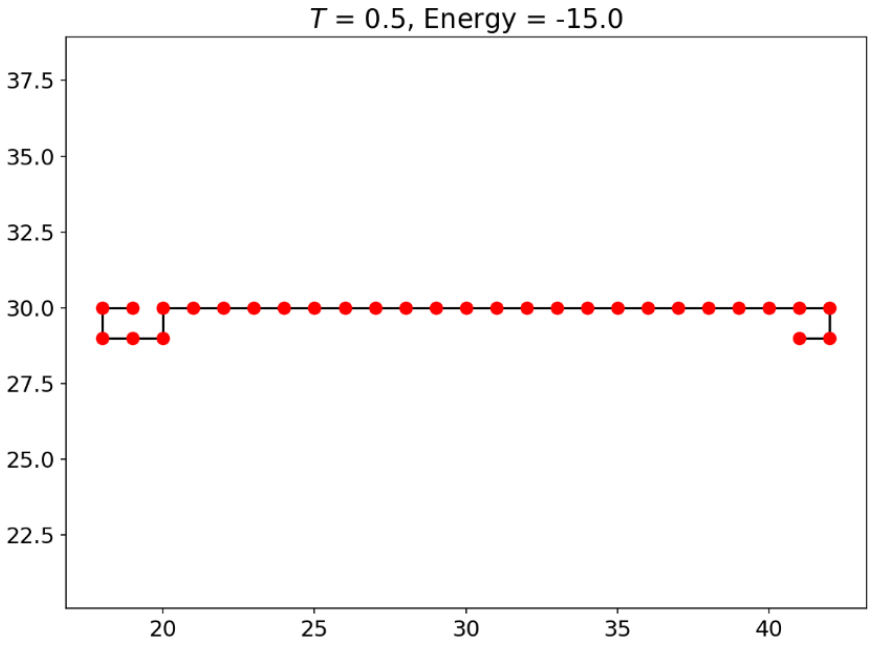


Figure 2.10 Final protein configuration of a run of initial condition case 2

PART D

Using stepping down temperature, we have the result

Energy averaged over last half of simulations is: -64.77

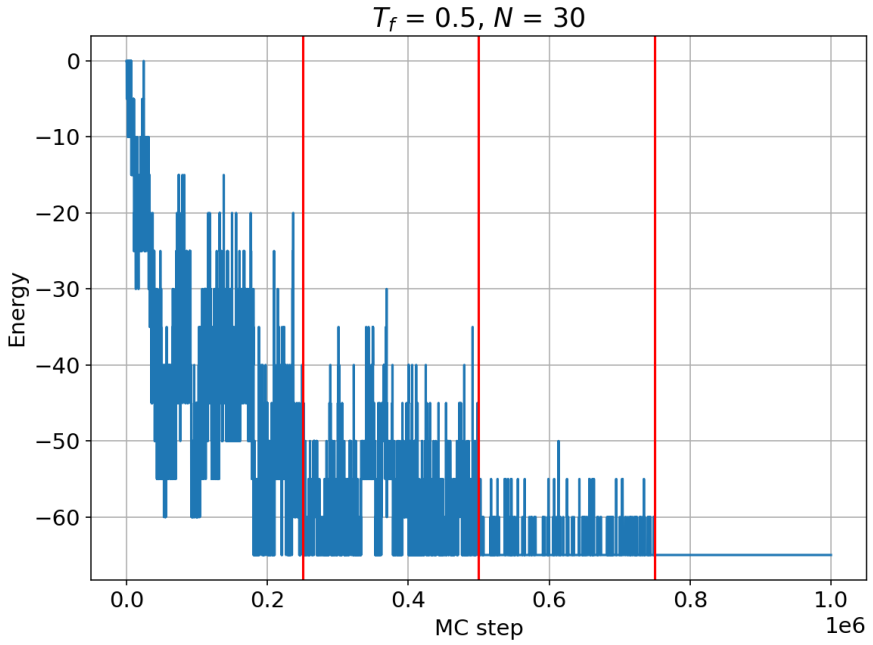


Figure 2.11 Energy plot for a run of stepping down temperature with $T_f = 0.5$

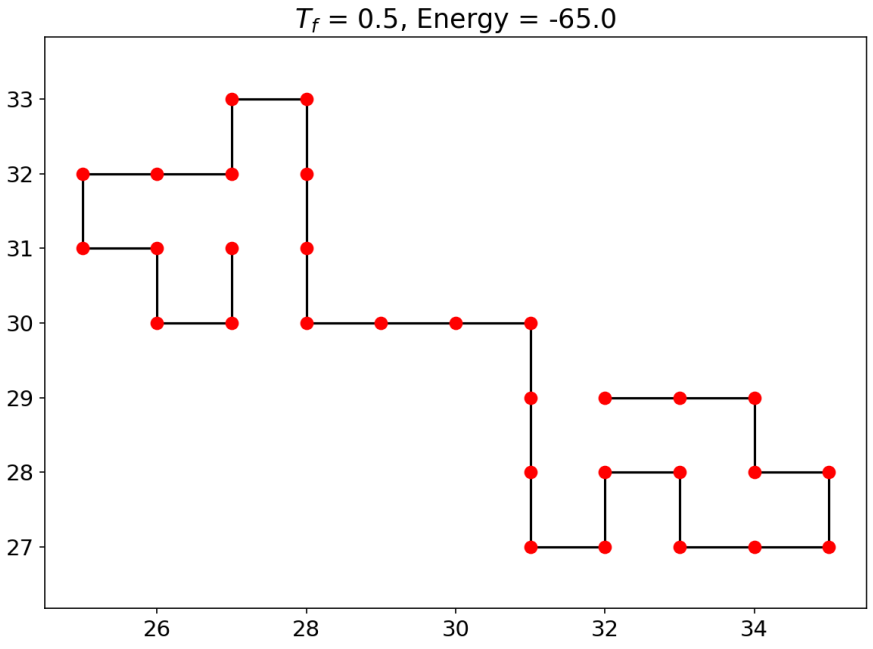


Figure 2.12 Final protein configuration of a run of stepping down temperature with $T_f = 0.5$

In the stage of making the plot, we decided to add a vertical line at each change in temperature to explicitly show the impact of changing temperature.

We observed from the figure 2.11 that reducing of energy did improve a lot compared with figure 2.9. We learned from the previous part that the first few thousands of steps will reducing the energy very rapidly.

When the temperature drops, we can clearly observe that according the first red vertical line, the energy fluctuation is much contained, and the lowest energy state ever achieved during this simulation has occurred very often.

Similar situation happens at the second vertical line, which the already contained fluctuation on energy reduces again. It seems that only one occasion that three energy states above the lowest has achieved during those steps.

After the last red vertical line, there are no longer any change in the energy state of the protein in this simulation, which agree with the result from figure 2.9 that very low probability of changing states in this temperature. And during the 250000 steps of the lowest temp, we have not hit any changes whatsoever.

Notice that figure 2.12 has a very similar configuration as the figure 2.6 because of the initial high temperature leads to rapid energy reduction of the protein. It should be expected that during the last section with lowest temperature, none if any potential changes to the configuration is successful, because hardly any changes of configuration would reduce the energy state anymore. This is the reason why evenith some high temperature system like we have in figure 2.6 will ends up to a similar energy state like this example.

PART E

We run the simulation for 20 different temperature each with 500000 steps, the result is plotted in figure 2.13 and data in table 2.1.

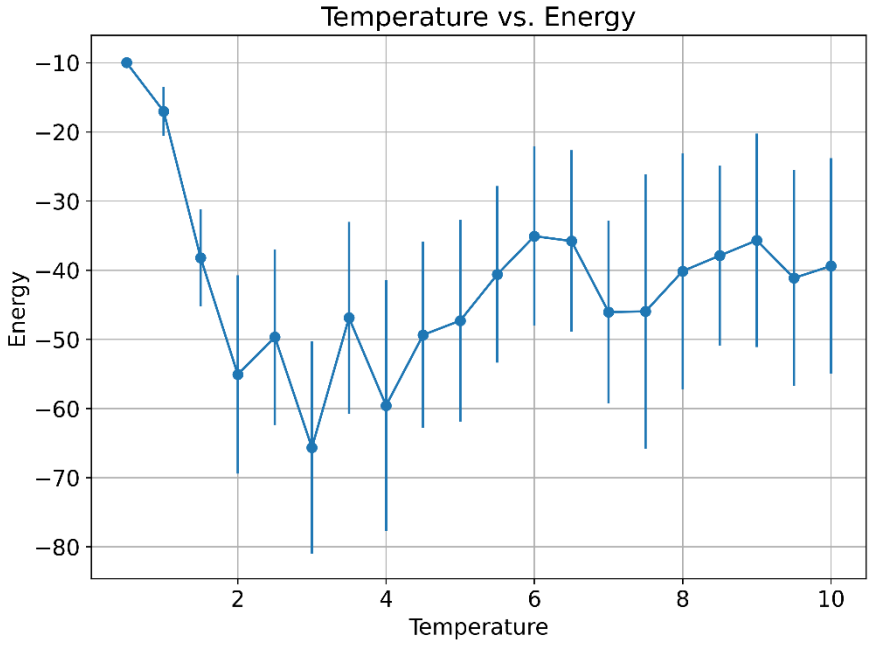


Figure 2.13 Temperature vs Energy plot

Notice that for $T = 0.5$, the error bar is too small to be seen, but the data in the table indicate how small the fluctuation in energy is.

For the temperature between 0.5 to 2, we have a clear trend that higher the temperature, the lower the energy states.

Between $T = 2$ and $T = 6$, there seems to be a upward trend in the data, which is very likely to be the bias of the simulation. Because we only simulate 1 protein with 500000 steps in each temperature, there is a very high chance that configuration of the protein is becomes a dead end for any further configuration changes. Especially, with a relatively low temperature, the chance of few consecutive configuration changes is unlikely. This could explain why we have the plot shape as is.

Temperature range between 6 to 10 is somewhat consistent, it is totally possible the result we have is heavily biased. With the computer we have, it will take ages to run few thousands protein simulations to have a better idea what is going on. With what we have, we might hit some bottleneck issue with the short protein chain. With a longer protein chain, we might further decrease the total energy of the protein.

There seems to be no phase transitions in the simulation result we have. But only 1 sample can't tell the full story. We expect that it's almost impossible (unless extremely lucky) to find any phase changes using 1 protein simulation data. A faster computer with a better algorithm might work, but for this lab, we did not find any.

Table 2.1 Simulation data

T	Mean Energy	Standard Deviation	T	Mean Energy	Standard Deviation
10	-39.3913	15.5976	5	-47.2882	14.603
9.5	-41.1265	15.6115	4.5	-49.349	13.4664
9	-35.6775	15.4406	4	-59.5611	18.1131
8.5	-37.886	13.032	3.5	-46.8729	13.8537
8	-40.128	17.0681	3	-65.655	15.3772
7.5	-45.9573	19.8189	2.5	-49.6552	12.7078
7	-46.0332	13.2351	2	-55.0753	14.334
6.5	-35.7762	13.1294	1.5	-38.2143	7.03093
6	-35.0701	12.9566	1	-17.0159	3.53488
5.5	-40.6087	12.7705	0.5	-9.98357	0.390231

Appendix

For Question 1 Part A

Figure A.1, seed = 100, normal time constant

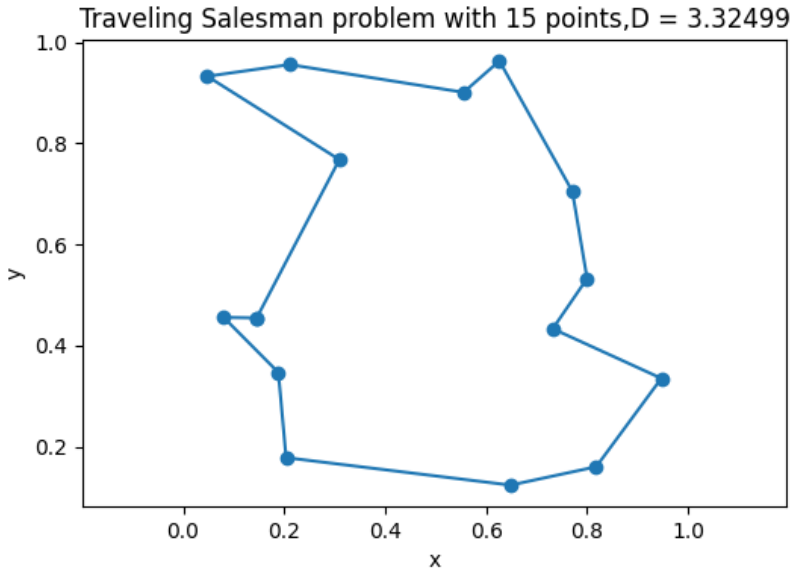


Figure A.2, seed = 101

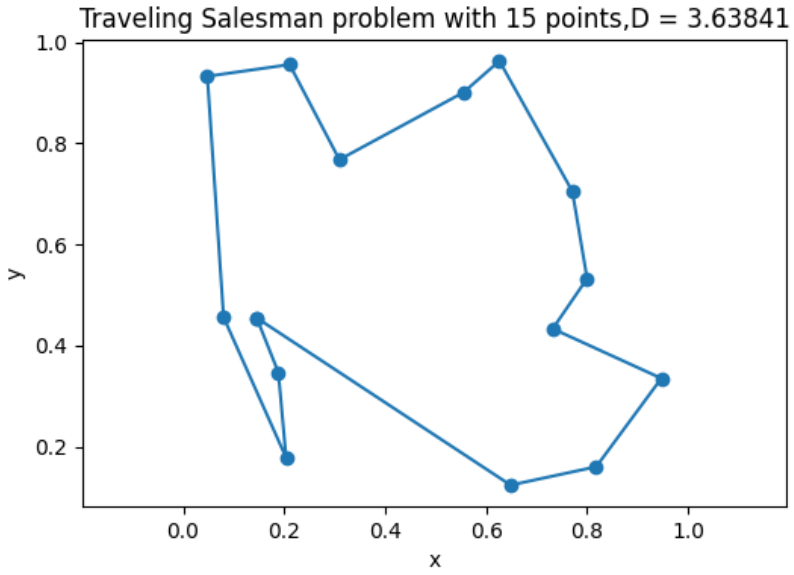


Figure A.3, seed = 102

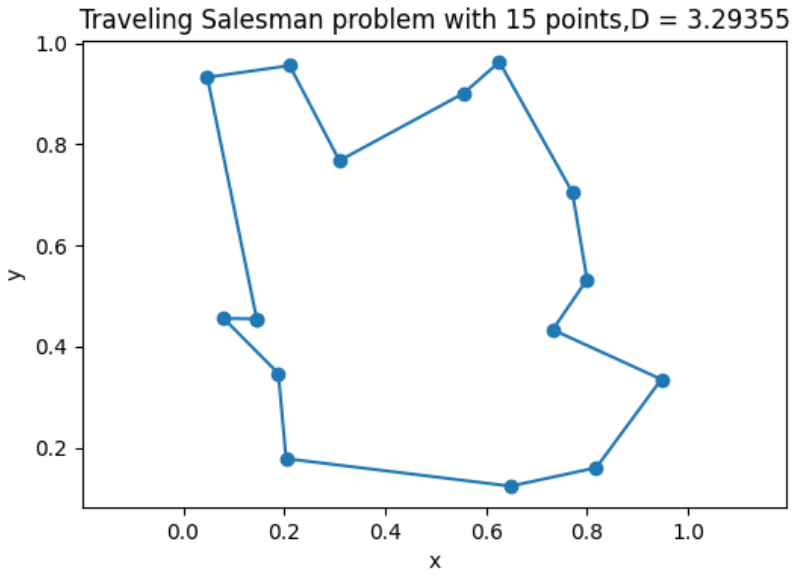


Figure A.4, seed = 103, (this is the optimum path)

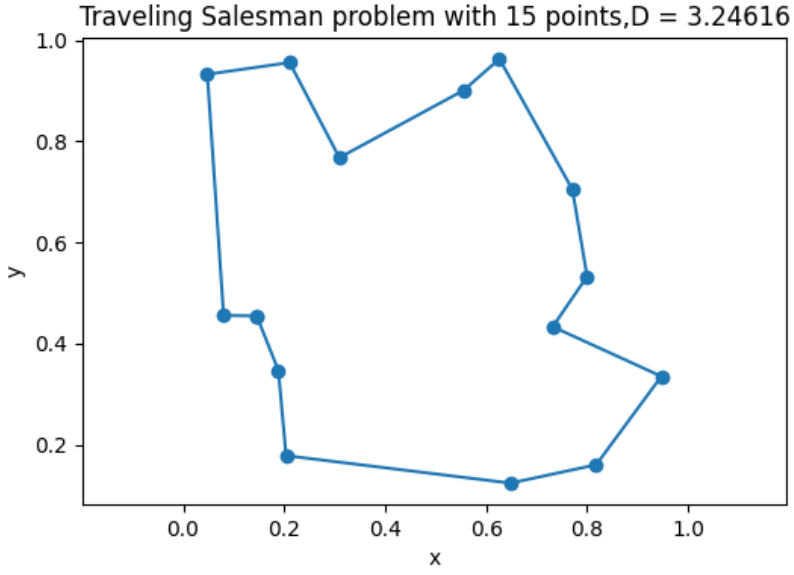


Figure A.5, seed = 104, the crisscrossing pattern makes the path poorly optimized.

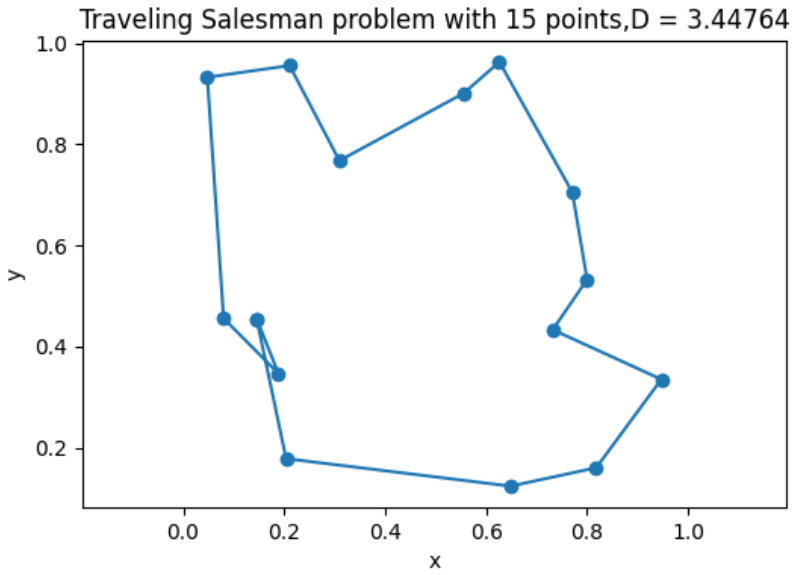


Figure A.6, seed = 100, time constant is halved. This is a poorer path compared to Figure 1.

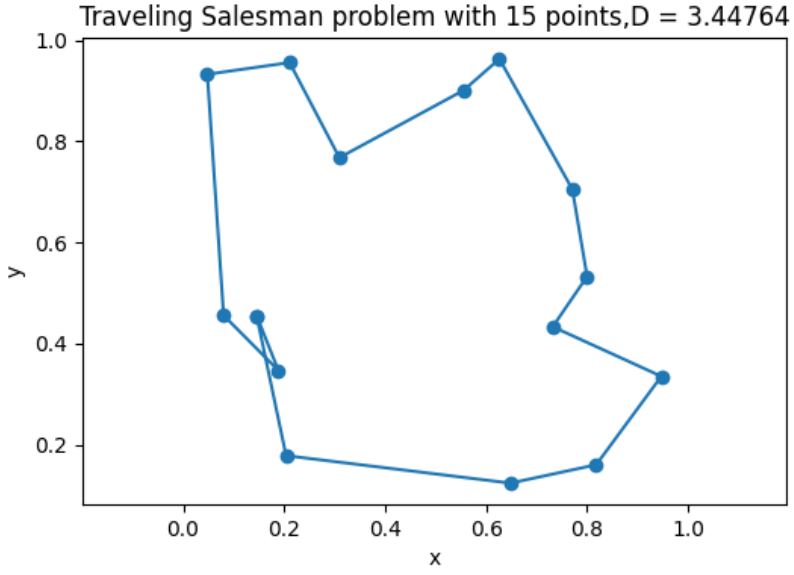


Figure A.7, seed = 101, again, distance becomes longer.

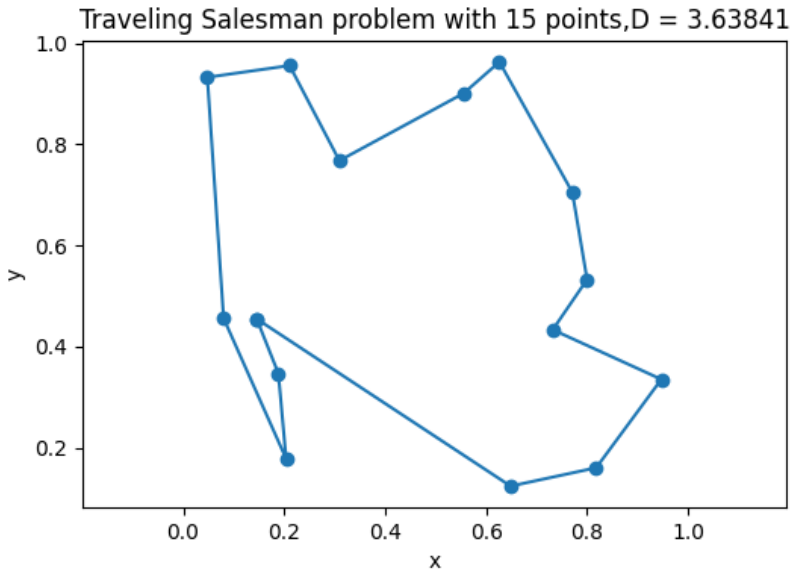


Figure A.8, seed = 102

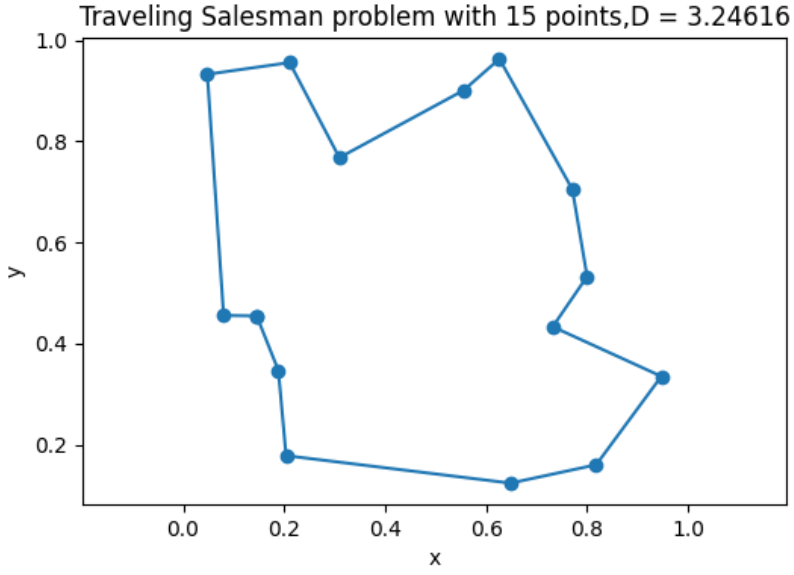


Figure A.9, seed = 103

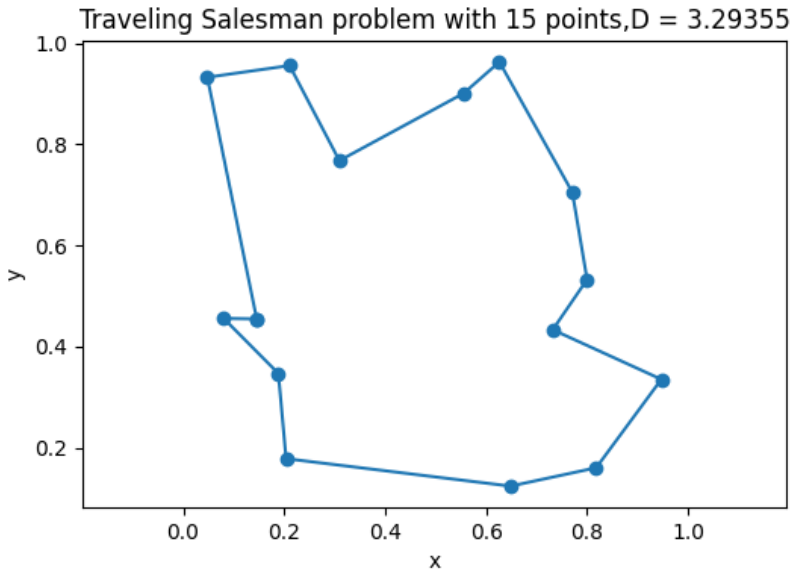


Figure A.10, seed = 104

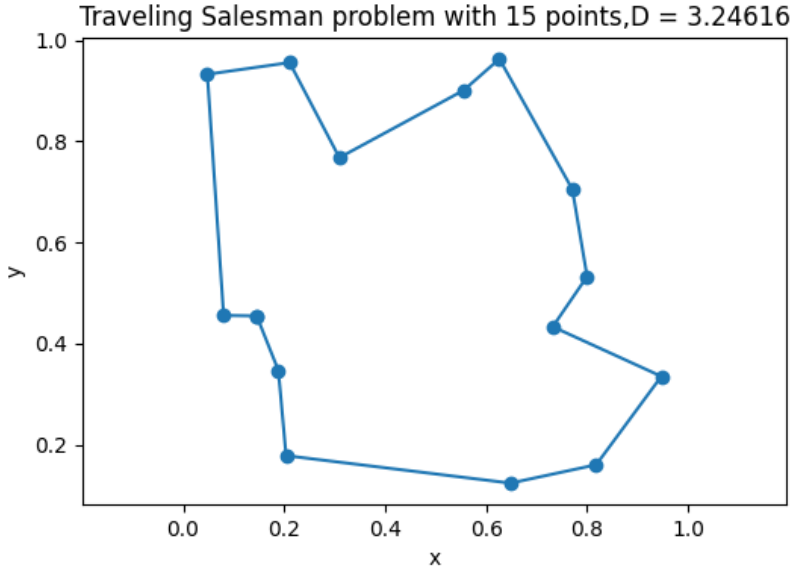


Figure A.11, seed = 100, the time constant is doubled. A longer run time optimized the path.

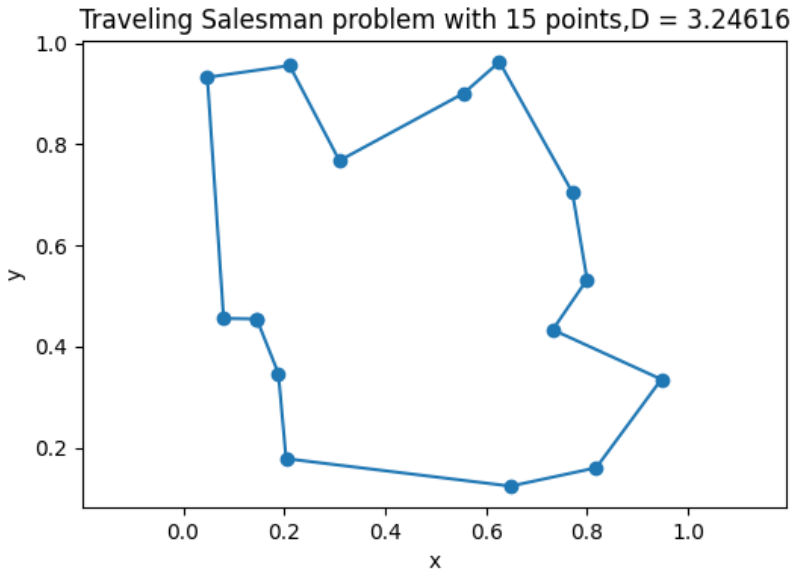


Figure A.12, seed = 101, the solution is still not perfect, but better than Figure 2

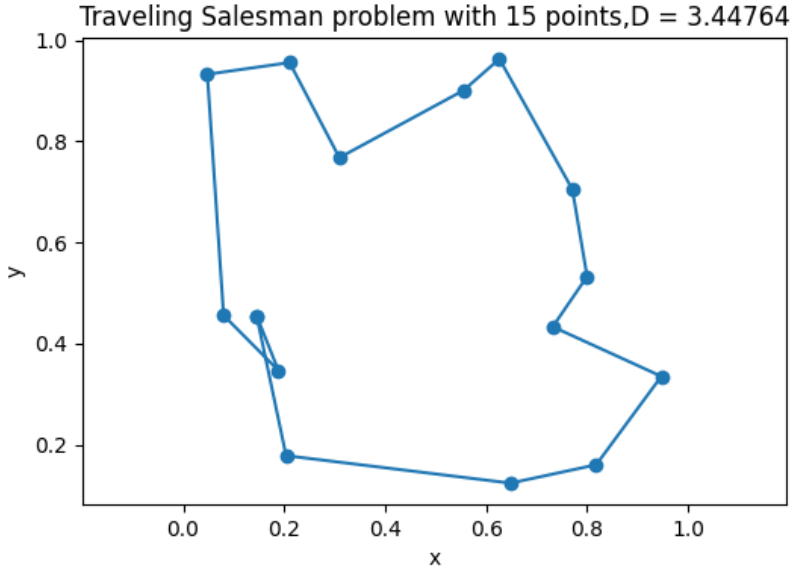


Figure A.13, seed = 102

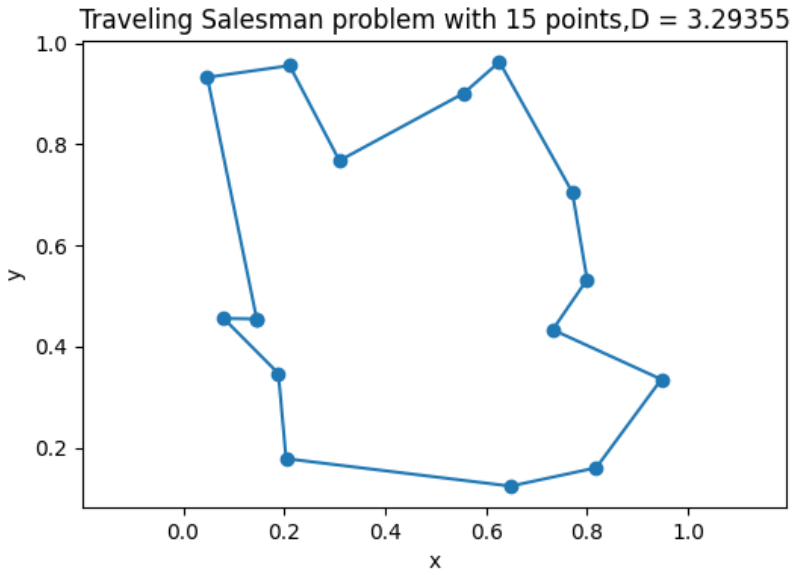


Figure A.14, seed = 103

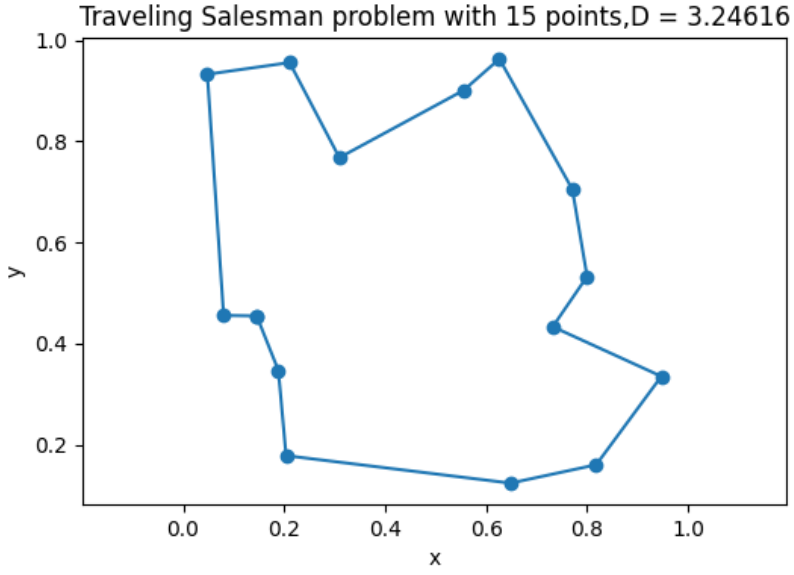


Figure A.15, seed = 104