QingheGao
12589896
Maud Bremer
11347872

1 Abstract

Traveling salesman problem (TSP) is a classical problem in combinatorial optimization. A proposed method to solve TSP is simulated annealing. In order to improve the performance of simulated annealing, the following parameters for this model were discussed: initial temperature, end temperature, Markov chain length, α and the cooling method. Finally, exponential multiplicative cooling method, $T_i = 10, T_e = 10^{-5}, \alpha = 0.9995$, Markov chain length$= 70$ were chosen to get the final best distance (2793.2425).

2 Introduction

The traveling salesman problem (TSP) consists of a salesman having to visit n different cities. The problem tries to find the shortest route to visit all these cities and return again to the city the salesman started in. For small n, this can be easily solved by comparing all possible tours and evaluating the distances these tours form. The total amount of tours for n cities is: $(n-1)!/2$, where clockwise and anti-clockwise tours are taken as the same tour. However, with increasing n it is too expensive to compute all the possible different tours and its distances. So, it requires a different method to compute the best tour of all cities. In this report simulated annealing is used to solve the traveling salesman problem with 280 cities.
Simulated annealing (SA) is a method that cools a system from a high initial temperature to a lower end temperature. Then different tours are generated by switching two cities and calculating the new cost between them. After which the Metropolis algorithm is used to decide whether this tour is accepted or rejected, which relies on the current temperature of the system. At high temperature tours with a higher cost can also be explored, but as the the temperature decreases the system will converge into a minima. Without having to calculate all possible tours ($279!/2 \approx 3 \cdot 10^{562}$). SA can converge into a local or even global minimum.
Simulated annealing is used to compute the best tour for 280 cities. Temperature cooling can be performed in different manners. Here, three different cooling methods are compared: linear, exponential and quadratic multiplicative cooling. Then further optimal parameters are discussed. Initial and end temperatures of the cooling scheme, the rate of cooling down and the Markov chain length. And after the process of fine tuning parameters, a final improved result will be discussed in this report. These parameters were shown to improve the minimal distance of the tour for 280 cities.

3 Theory

3.1 Traveling Salesman Problem

The traveling salesman problem (TSP) is a classical problem in combinatorial optimization, where the optimal distance between n different cities is computed. Different cities in Germany are taken as an example, shown in Figure1. Here, the main task is to find the shortest tour. First, vector π represents a possible tour along all cities. The distance between all cities is stored in a matrix d. d_{ij} is the distance between city i and j. Thus the cost function C, which will compute the distance between two cities is given by:

$$C = \sum_{i=1}^{N} d_{i,\pi_i} \tag{1}$$

The number of tours is given by Equation 3.1, where clockwise and anti-clockwise tours are taken as the same tour. By evaluating the cost of each tour, the best tour can be found. For large n it can be expensive to compute the distance for every possible tour.

$$\text{Possible tours} = \frac{(n-1)!}{2} \tag{2}$$

Figure 1: Traveling Salesman Problem for 15 cities in Germany. Adapted from [1].

3.2 Markov chains and the Metropolis Algorithm

A Markov chain is a stochastic process in which the outcome of each state only depends on the outcome of the previous state. In this way, a Markov chain does not have any memory. The transition probability to go from state i to state j is given by Equation 3:

$$P_{ij}(k) = P[X_k = j | X_{k-1} = i] \tag{3}$$

where X_k denotes the state at the time k, i and j is one of states ($i, j = 1, 2, \cdots N$) and P_{ij} denotes the probability the state transfers i from j. In this way, a transition matrix can be formed:

$$a_j(k) = \sum_{N}^{i=1} P_{ij}(k) a_i(k-1) \tag{4}$$

where $a_j(k) = P[X_k = j]$. For homogeneous Markov chains this transition does not depend on k. Thus, $P_{ij}(k) = P_{ij}$.
Based on Markov chains, the Metropolis Algorithm has been used to obtain a sequence of random samples from a probability distribution indirectly. The steps for this algorithm are as followed:

1. Generate a Markov chain.
2. Calculate the cost of this new chain.
3. Accept this new step from state i to j with the transition probability Γ_{ij}.

$$\Gamma_{ij} = \begin{cases} 1 & \text{if } p(X_j) > p(X_i) \\ \frac{p(X_j)}{p(X_i)} & \text{if } p(X_j) < p(X_i) \end{cases} \tag{5}$$

4. Repeat from the beginning.

For the TSP problem, the Markov chains are generated using the Lin 2-opt transition. This transition uses the current sequence of cities and reverses two cities randomly to get the next sequence of cities and a new distance. Next, the acceptance will be evaluated. Then, the whole process is repeated till the end of the Markov chain length.

3.3 Simulated Annealing

Iterative improvement algorithms are a method to find the minimal cost for a given system with multiple possibilities. However, because this method calculates the gradient and follows or goes against the direction of this gradient, the system can get stuck in a local minimum.
In this report, simulated annealing has been used to solve the TSP. Annealing is a technique, in which a perfect crystal can be created by heating just below the melting temperature and cooling it down slowly. The Boltzmann distribution describes the probability that the crystal is in a state with energy *E*:

$$P(E_{in} = E) = \frac{1}{Z(T)} e^{-\frac{E}{kT}} \tag{6}$$

where T is the temperature, k is Boltzmann constant and $Z(t)$ is the partition function, which is usually highly dimensional and hard to calculate. An analogy of this annealing process has been used to solve the TSP. The possible tours in TSP act similarly to the possible states of a crystal; the cost of each tour resembles the energy (E) of each state of a crystal. Also the process of annealing to a minimal energy is just like finding the shortest tour between n different cities. Then Equation1 can be rewritten into:

$$P(configuration = i) = \frac{1}{Z(c)} e^{-\frac{C(i)}{c}} \tag{7}$$

The steps in the simulated annealing algorithm are:

1. Start with a random configuration.
2. Start at initial temperature T_i.
3. Use the Metropolis Algorithm to equilibrate the system and get the sample configuration from Boltzmann distribution.
4. Use a cooling method to decrease the temperature.
5. Start from step 3.

Thus, in simulated annealing the Boltzmann distribution has been used in the Metropolis Algorithms. Then the steps of Metropolis Algorithms:

1. Propose a new cost j from state i.
2. Calculated the difference of the cost δc.
3. Calculated the transition probability from state i to j.

$$A_{ij} = min[1, e^{-\frac{\delta c}{C}}] \tag{8}$$

4. if $A_{ij} = 1$, accept the new state j. if $A_{ij} = e^{-\frac{\delta c}{C}}$, generate the uniformly distributed random number R between 0 and 1. Accept the new state j when $A_{ij} < R$.

3.4 Method

Simulated annealing for the Traveling Salesman Problem was implemented using Python code from [2] and Python libraries: numpy, scipy, matplotlib. [3] [4] [5] The initial configuration of cities for simulated annealing was generated using a greedy algorithm, calculating the closest neighbour to each city and connecting these. First, the system was converged into a local minimum. To improve upon this, the convergence for different cooling schedules are compared. First, three different cooling methods are discussed.

3.4.1 Cooling methods

The cooling of the system can be performed in different methods. Here linear, exponential and quadratic cooling are discussed, how the temperature decreases for these different methods is shown in Figure 2. Linear cooling cools the temperature with a fixed step. This way, each step the temperature decreases with the same interval, α.

$$T_{i+1} = T_i - \alpha \tag{9}$$

The second method that is discussed is exponential cooling. In this method the cooling speed will first be fast, after which it will decrease. Meaning that in the first few steps the temperature steps are large and closer to the end temperature the steps will be very small.

$$T_k = T_0 \alpha^k \tag{10}$$

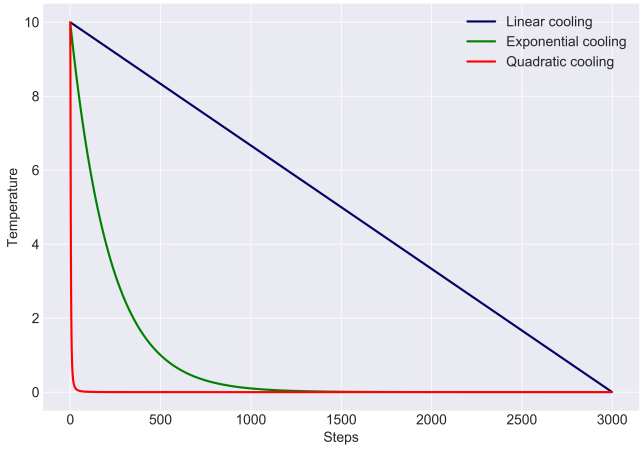


Figure 2: The temperature decreasing for linear cooling, exponentially and quadratic multiplicative cooling.

The last method discussed, is the quadratic multiplicative cooling. This is similar to exponential cooling, where the first steps the temperature will decrease more steeply than towards the end. But, with quadratic multiplicative cooling the first steps will decrease even more quickly.

$$T_k = \frac{T_0}{1 + \alpha k^2} \tag{11}$$

3.4.2 Different initial and final temperatures

Using exponential cooling, the convergence and the value of the minimum are studied for different initial and final temperatures, while keeping the temperature decreasing speed, α, constant. The initial temperatures are ranged from and the end temperatures are ranged from 0.1 to 50 and the end temperatures are ranged from 0.1 to $1e-8$.

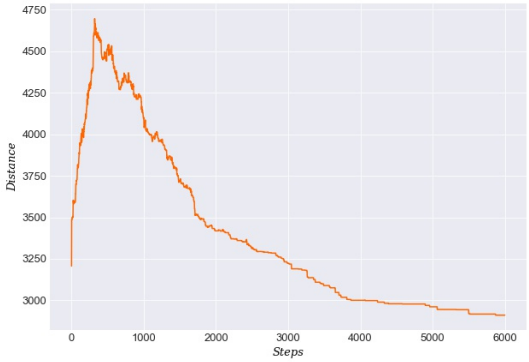
3.4.3 Effect of α and Markov Chain Length

The convergence and value of the minimum is also discussed for different α (ranging from 0.5 to 0.995 and for different Markov chain lengths (1to 70).
The total steps in one simulation depends on the Markov chain length and how quickly the temperature is decreasing, which is determined by α. This makes the total steps in one simulation given by: $k \cdot$ Markov Chain Length. k is calculated from Equation 3.4.1 and is given by : $\frac{log(T_k/T_0)}{log(\alpha)}$. The value of the best cost is compared for different $alpha$ while keeping the total steps the same, meaning that the Markov chain length changed according to the value of α to keep the total steps constant.

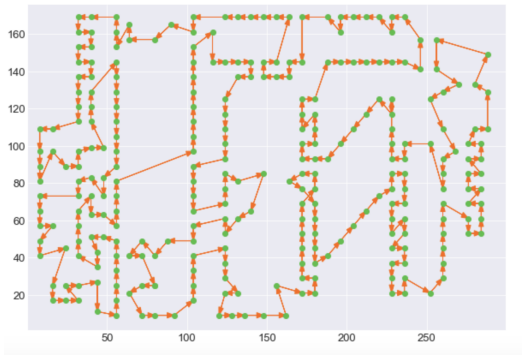
4 Results and Discussion

4.1 Simulated Annealing converges to a local minimum

Using Simulated annealing it was possible to converge to a local minimum, shown in Figure 3(a). First, the distance increases steeply, which is the result of the high temperature. At high temperature the transition probability for states with a higher cost will be higher. But with increasing steps, the temperature exponentially decreases and so does the distance of the tour, until it converges to a value of 2910.7668. This was not the global minimum. The global minimum for this set of 280 cities is: 2586.7696 (this was calculated using the optimal distance file).



(a) Convergence to a local minimum.



(b) Final tour found for 280 cities.

Figure 3: Finding a local minimum using simulated annealing for TSP with 280 cities. $\alpha = 0.999$, $T_i = 10$, $T_e = 0.1$, Markov chain length $= 50$.

4.2 Convergence differs for different cooling schedules

Figure 4 shows the convergence of three different cooling methods. Quadratic and exponential multiplicative cooling converge very similarly and to around the same value after 3000 steps. However, linear cooling acts differently. The distance first increases very steeply, after which it decreases and after 3000 steps it does not seem converged. This could result from larger steps at a small temperature. The other two methods had very small steps towards a small temperature and were converged. Additionally, the final value for linear cooling is far from the other two methods as the Figure shows. Between quadratic and exponential cooling the final value is similar, but exponential cooling had a peak of distance in the whole decreasing process, which would help to increase the possibility that the distance ”escapes” the local minimum. While quadratic first samples more space and then converges. For all methods the error margin decreased with increased steps. Between these three methods, exponential cooling method has been chosen for further experiments.

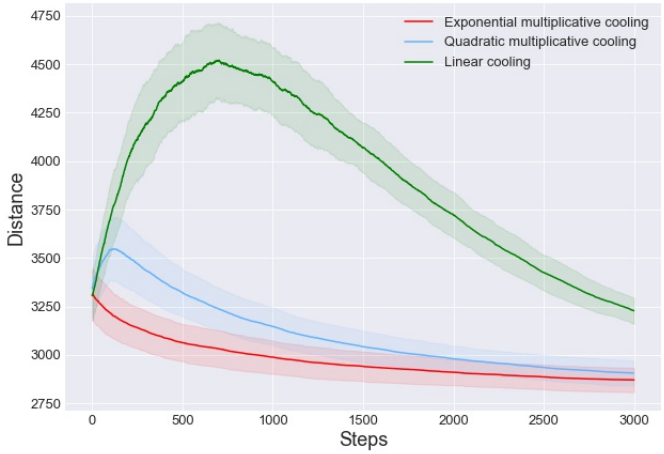
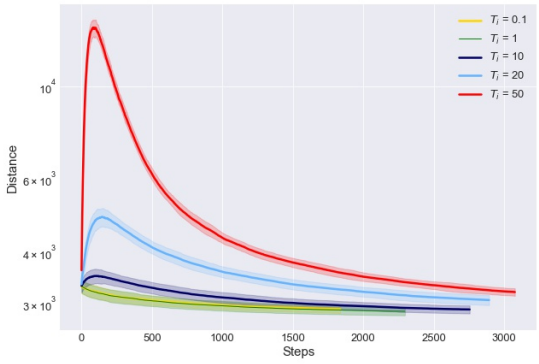


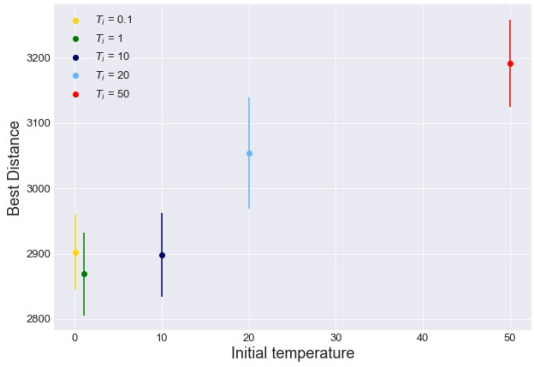
Figure 4: Convergence for three different methods for 3000 decreasing steps, averaged over 50 simulations.

4.3 Temperature influences convergence to minimum

Firstly, different initial temperatures have been explored, shown in Figure 5(a). When the $T_i = 0.1$, the distance always decreased with the decreasing of the temperature. That is because when the initial temperature is small, the probability of accepting a larger distance will be smaller. Reversely, when $T_i = 50$ a peak was present because of the probability. Figure 5(b) shows the best distance of each different initial temperature. $T_i = 1$ has the smallest best distance and $T_i = 50$ has the biggest best distance. Thus, $T_i = 10$ has been chosen, because distance can easily be stuck in a local minimum when either T_i is too big or small.



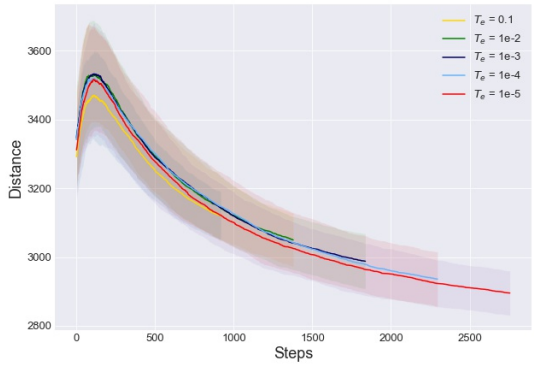
(a) Convergence for different initial temperatures.



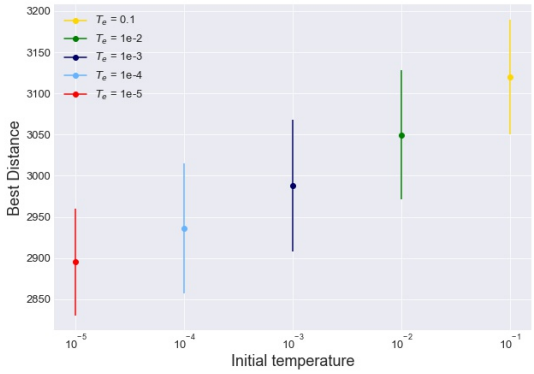
(b) Best cost for different initial temperatures.

Figure 5: Different initial temperatures for TSP with 280 cities. α = 0.995 , T_e= 10^{-5}, Markov chain = 50, Simulation times = 50.

Furthermore, different T_e have been explored, whilst keeping T_i and α constant. The smallest T_e with a value of 10^{-5} had the smallest best cost as Figure6(b) shows. While larger T_e had higher distances for the best found tour. This results from smaller T_e having more steps in the temperature decreasing towards the end. The curves all look similar, except for the number at which it converges. This is because of the exponential cooling. With this type of cooling the initial path will be very similar, because all have the same initial temperatures. However, the final part will be different since there are more steps taken towards the end with a lower temperature. Because of the exponential function these steps are very small. Overall, the lower the end temperature is, the lower the best distance will be.



(a) Convergence for different end temperatures.

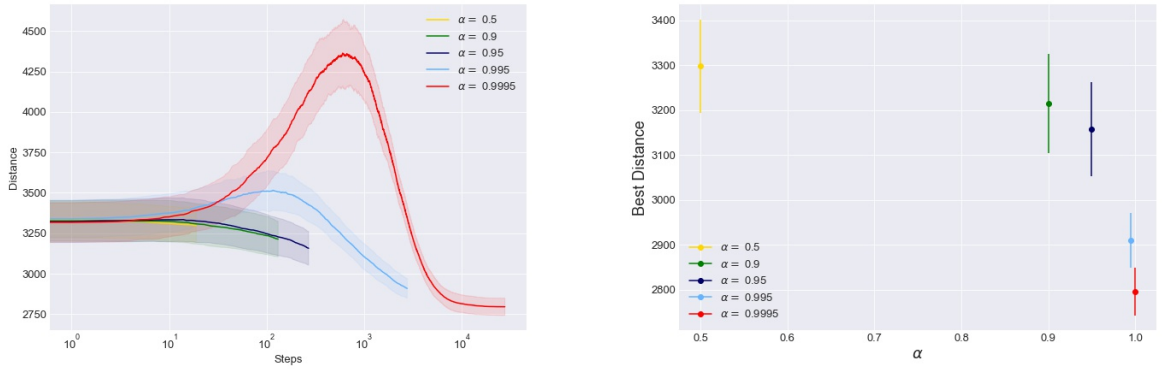


(b) Best distance for different end temperatures.

Figure 6: Different end temperatures for TSP with 280 cities. α = 0.995 , T_i= 10, Markov chain = 50, averaged over 50 simulations.

4.4 Convergence is influenced by α

The rate of decreasing the temperature is determined by α for exponential cooling. With a small α the temperature steps will be large and the system might get stuck into a minimum. Figure 7(a) shows the convergence for α ranging from 0.5 to 0.9995. The closer α got to 1, the higher the highest sampled distance was. This is a result of smaller steps when the system is at a high temperature. However, also with higher α the final cost got better. So, this increased sampling seems to be able to converge to a better solution. Figure 7(b) shows the final value for different values of α. The value of the best solution decreases with increasing α and so does the error margin. Thus, increasing α will result in a better solution for the TSP with a smaller standard deviation.



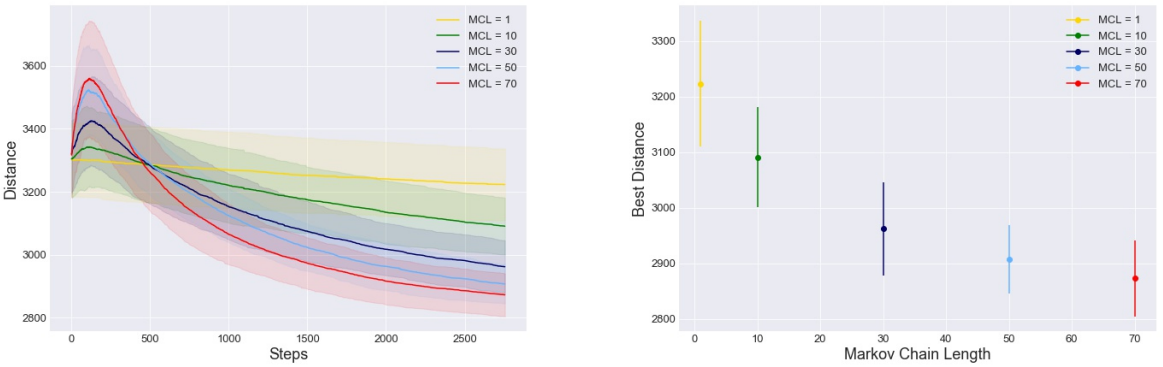
(a) Convergence for different α.

(b) Best cost for different values of α.

Figure 7: Convergence and best found distance for different values of α. Averaged over 50 simulations.

4.5 Effect of Markov Chain Length on convergence

The Markov chain length determines the number of iterations at the same temperature. Figure 8(a) shows the convergence for different Markov Chain lengths with same α. Increasing of the length, decreases the value at which the system converges, while the convergence seems to stay similar for different lengths. Also with higher Markov chain length, higher costs can be seen in the beginning of the simulations, which is due to more sampling at higher temperatures. This way more of the search space can be sampled. Figure 8(b) shows the last values for different Markov Chain lengths. Again, it shows that with increasing length the minimum will be lower. Also the standard deviation decreases slightly. It is expected that for large values of this length the convergence will not differ anymore, but this was not seen. This could be a result of not reaching these lengths, because it was computationally not possible. These results show that increasing Markov Chain lengths will result in better costs.



(a) Convergence for different Markov Chain lengths.

(b) Best found cost for different Markov Chain lengths.

Figure 8: Convergence and best found values for different Markov Chain lengths. Averaged over 50 simulations.

4.6 Effect of Markov chain length and α on best distance

For both increasing α and Markov chain length, the best found tour distance seems to be lower. Figure 9 shows which of these two has a higher weight to increasing the system results. While the total steps ($log(\frac{T_e}{T_i})/log(\alpha)$ · Markov chain length) stays constant, the values of α are changed and the Markov chain length according to that. For increasing α, and thus decreasing the Markov chain length, the best cost decreases from 3190 to 3110. This difference is small and indicates that both parameters are equally important in finding the best distance.

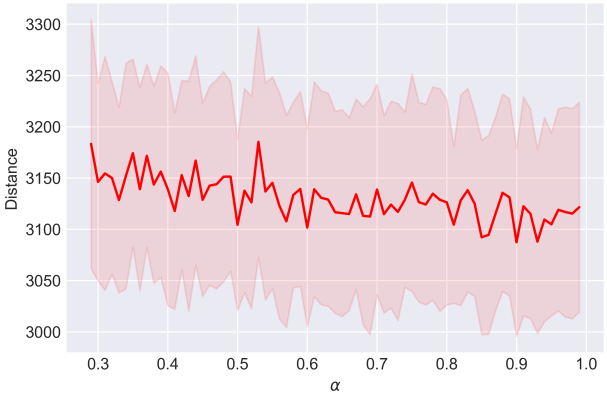
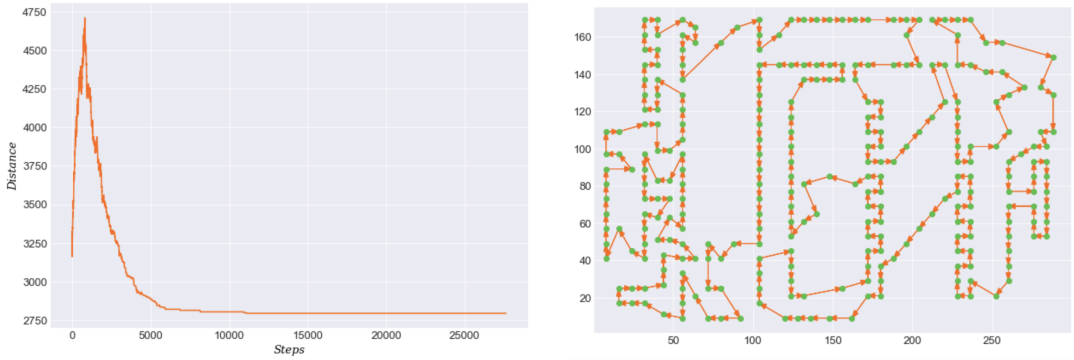


Figure 9: Best found different for values of α, while keeping the total steps equal to 20000. Averaged over 50 simulations, $T_i = 10$, $T_e = 1e - 8$.

4.7 Convergence to a minimum with improved parameters

After the process of tuning parameter of the simulated annealing, these are used to find a new minimum for TSP with 280 cities. The final result is shown in Figure10. The best distance of the final result converged to 2793.2425, which is an improvement of the previous simulation. However, due to the limited computing power it was difficult to get the global minimum: 2586.7696. Also, this was only one simulation instead of an average, but again due to limited computing power it is hard to do multiple simulations. But from this one simulation an improvement can be seen using the new parameters.



(a) Convergence to a local minimum.

(b) Final route for TSP with 280 cities.

Figure 10: Convergence for TSP with 280 cities with improved parameters. $\alpha = 0.9995$, $T_i = 10$, $T_e = 10^{-5}$, Markov chain length $= 70$, Best distance$= 2793.2425$.

5 Conclusion

A local minimum for a TSP problem with 280 cities was found using simulated annealing. The convergence and the value of the local minimum were compared for three methods: linear cooling, exponential multiplicative and quadratic multiplicative cooling. Exponential and quadratic cooling both converged quickly, but exponential cooling converged to a lower cost. This method was used to find other optimal parameters. The optimal initial temperature was found to be 10. The lower the end temperature the lower the convergence cost will be. The closer α comes to 1 the better the cost will be, also larger Markov chain length will result in better cost. The dependency between α and the chain length showed that both are equally important in finding the best distance. Using these improved parameters a better distance was found compared to the previous simulation. However, due to the limited computing power it was still not the same as the optimal distance. Further work can focus on simulations with larger α, larger Markov chain lengths or a better cooling method, such as decreasing slowly at the beginning and end temperatures and faster in the middle.

References

[1] Problem des handlungsreisenden. https://de.wikipedia.org/wiki/Problem_des_Handlungsreisenden. Accessed: 10-12-2019.

[2] Cihan Ceyhan. Simulated-annealing-tsp.

[3] Oliphant Travis E. *A guide to NumPy*. USA: Trelgol Publishing, 2006.

[4] S. Chris Colbert Stéfan van der Walt and Gaël Varoquaux. The numpy array: A structure for efficient numerical computation. *Computing in Science & Engineering*, 13:22–30, 2011.

[5] John D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9:90–95, 2007.