An Analysis of Simulated Annealing applied to the Traveling Salesman Problem

Mehmet Bahadir Gülhan (14125250)
m.b.guelhan@gmail.com
Nicolas Schrama (11881437)
nicolas.schrama@student.uva.nl

December 24, 2021

Abstract

We investigate the effect of initial temperature, Markov chain length, and cooling schemes on convergence to the global optimum in an application of the simulated annealing method to the travelling salesman problem. Our simulation results indicate (i) the geometrical schemes outperform arithmetic-geometric and linear schemes, (ii) that a low initial temperature yields faster convergence, and (iii) that in the range of Markov chain lengths we experiment with, the longer the Markov chain, the better the solution we find. Based on these results, we attempt to convergence to the true solution of the problem instance. Our solution is less than twenty percent longer than the global optimum route. Furthermore, convergence is rapid. Hence, we conclude that simulated annealing is an effective method to solve instances of travelling salesman problems.
Key words: Simulated annealing, travelling salesman problem, Markov chains, Markov Chain Monte Carlo, Simulation

1 Introduction

Optimizing functions with many optima in a small interval poses challenges to standard optimization techniques, such as Newtonian methods, and yield only a local optimum.

Simulated annealing can solve the problem misbehaving functions pose. The main idea of Simulated Annealing is using Markov-Chain Monte Carlo sampling methods to generate a candidate solution - also called a state - and accept or reject this candidate based on an acceptance criterion. This criterion is based on the Boltzmann distribution. With each iteration of the Markov chain, the temperature parameter, which determines the shape of the Boltzmann distributions, is lowered according to a cooling scheme. When the temperature decreases, the strictness of the acceptance criterion increases and the probability of accepting poorer solutions for variation's sake decreases. Hence, the distribution centers more around the optimal value as the temperature decreases.

In this paper, we investigate the simulated annealing method. In particular, we analyze the effect of initial temperature, Markov chain length, and cooling schemes on convergence to the global optimum.

To this end, we apply the simulated annealing method to the traveling salesman problem (TSP). First, we determine which initial temperature, Markov chain length, and cooling scheme yield the best tour. Second, we attempt to find the shortest route for the TSP instance using the best configuration of initial temperature, Markov chain length and cooling scheme in our simulation results, and compare the results of the best configuration to the true shortest route.

In section two, we provide a theoretical overview of simulated annealing and the TSP. In section three, we elaborate on the experimental setup. Section four discusses the simulation results. Finally, section five concludes.

2 Theoretical Background

In this theoretical overview, we elaborate on the simulated annealing (SA) procedure and its mechanisms and on the traveling salesman problem.

2.1 Simulated annealing

SA [1] combinatorial optimization problems. In these problems, values of a function with many degrees of freedom and many local optima are optimized [2]. Inspired by crystal structure forming under cooling, SA makes it possible to prevent getting stuck at local optima while converging to a global optimum. To this end, the algorithm samples random points and evaluates their fitness. Improvements compared to the previous state are always accepted. Crucially, deteriorations are accepted as well in the early stages of the sampling process for the sake of variation. This allows the algorithm to identify high hills in the search space. As the sampling process goes on, acceptance of deteriorations becomes increasingly strict. Therefore, only improvements are accepted in the final stages of the sampling process. During this stage of sampling, the algorithm is climbing a hill in the search space.

To create this mechanism, SA uses the Boltzmann distribution $p_i \approx e^{-\frac{h(X)}{T}}$ [3]. Here, $p(X)$ is the probability density function, h the objective function, and T a control parameter known as temperature. For high T i.e., as T converges to infinity, $p(X)$ converges to uniform density. As T converges to zero, $p(X)$ converges to the Dirac-delta around the optimum. As a result, the probability of points other than the optimum being sampled decreases as T decreases. This works under the condition that the algorithm has found the hill containing the global optimum before T reaches zero. In practice, the SA algorithm consists of the following six steps [3]. First, select a symmetric proposal density $q(X_{i+1}|X_i)$. Second, sample the next state $x_{i+1} = q(X|x_i)$. Third, sample a random draw U from the U(0,1) distribution. Fourth, compute $\alpha(x_i) = \min(e^{-\frac{h(x_{i+1})-h(x_i)}{T}}, 1)$. Fifth, if $\alpha(x_i) \geq U$, the next state is the sampled state x_{i+1}. If $\alpha(x_i) < U$, the next state equals the current state x_i. Sixth, decrease T.

To obtain convergence to the global optimum, the length of the Markov chain, the initial temperature, and the cooling schedule of the temperature are essential. The length of the Markov chain must be high enough such that the algorithm has enough time to find the highest hill and that it may climb it when the temperature becomes low. On the other hand, the length cannot be too large, as the algorithm may accept a worse hill while searching the search space such that it climbs a hill with a local optimum.

Furthermore, the initial temperature determines the acceptance of deteriorations in the early stages of sampling. If T_0 is high, acceptance is high, while it is lower if T_0 is low.

Finally, the cooling schedule determines the rate at which the temperature decreases during the sampling processes. Multiple variations exist. We consider four schemes here: linear, logarithmic, geometrical, and arithmetic-geometric. First, in a linear scheme [4], the temperature decreases linearly from T_0 to zero. Linear schemes do not guarantee convergence to the global optimum but do converge relatively quickly to strong optima, i.e., those close to the global optima.

Second, [5] defines a logarithmic scheme as follows:

$$T_t = \frac{c}{\log(1+t)} \tag{1}$$

Here, c is a problem-dependent positive constant. Although this cooling schedule asymptotically converges towards the global optimum theoretically, it converges slowly in practice. Third, geometrical schemes [4] behave as follows:

$$T_t = T_0 \alpha^t \tag{2}$$

Here, is a constant between zero and one. The closer is to one, the slower the cooling. Geometrical schemes cool faster than logarithmic schemes [2] [6].

Fourth, arithmetic-geometric schemes [4] are defined as:

$$T_{t+1} = \alpha T_t + b \tag{3}$$

This scheme simplifies to an arithmetic scheme when $\alpha = 1$ and to a geometric scheme. Three behaviors may emerge in this scheme. When $\alpha > 1$, the progression diverges to $+\inf$. When $|\alpha| < 1$, the progression converges to $\frac{b}{1-a}$. When $\alpha \leq 1$, the progression diverges.

2.2 Traveling salesman problem

The traveling salesman (TSP) problem takes its name from a salesman's wish to travel between n cities in the shortest amount of time possible, making a round trip while visiting each city exactly once [7]. Hence, it is a combinatorial optimization problem.

To solve the TSP, a few different approaches have been taken. First, exhaustive methods that check all permutations and select the cheapest one always find the global optimum. However, the number of permutations is n!, making this approach impractical even for twenty cities. Other exact approaches include [8], using the cutting-plane method.

Furthermore, heuristic algorithms are used to deliver an approximated solution in reasonable computational time. Among these is the 2-opt – also known as pairwise exchange – method [9]. This entails deleting two non-adjacent edges and reconnecting the four affected cities such that a new circuit is created. It achieves good results on real world Euclidean instances with respect to running time and accuracy.

3 Experimental setup

In this section, we discuss the three experiments we conduct, the parameter values used in them, and the dataset.

In the three experiments, we investigate the effect of initial temperature, Markov chain length, and various cooling schedules on the convergence to the global optimum route. Each experiment is based on fifty simulations. The configurations used in each experiment are summarized in Table 1.

Experiment	Initial temperature	Markov chain length (in thousands)	Cooling schedule
Experiment 1	1, 10, 100, 1000	5100	Geometrical 1
Experiment 2	100	1, 5, 10, 25, 50, 100, 150	Geometrical 1
Experiment 3	100	100	Geometrical 1, Geometrical 2, geometrical-arithmetic, linear

Table 1: Configurations of the experiments

The geometrical 1 cooling schedule uses $\alpha_1 = 1 - 10/mc_{length}$, which equates to 0.9999 when the Markov chain length is 100 000. The geometrical 2 cooling schedule uses $\alpha_2 = 1 - 5/mc_{length}$, which equates to 0.99995 when the Markov chain length is 100 000. The arithmetic-geometric cooling schedule uses α_2 and $b = (1 - \alpha_1) * 3 * 10^{\log(mc_{length})-5}$, which equates to 0.0003 when the Markov chain length is 100 000. Figure 1 provides a visual comparison of these methods.

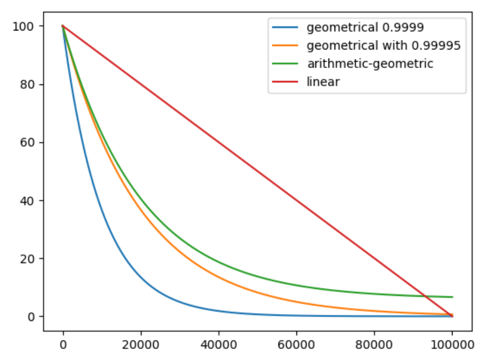


Figure 1: The cooling of the temperature using four schemes.

Figure 1 shows that the temperature decreases fastest in the geometrical 1 cooling scheme, then in the geometrical 2 cooling scheme. In the arithmetic-geometric cooling scheme, the temperature decreases slower and converges to 6 – at least for initial temperature 100 and Markov chain length 100 000. Temperature decreases linearly in the linear cooling schedule. The parameter selection of α_1, α_2, and b is such that the cooling schemes are proportional to those presented in figure 1 for all other values of Markov chain length and initial temperature.

The dataset we apply the methods to is the a280 Ludwig drilling problem, which contains 280 cities. The length of the optimal route is 2579.

4 Results & Discussion

In this section, we use the simulated annealing algorithm to optimize the route in the a280 problem using the thirteen parameter configurations presented in Table 1. We do not elaborate on all simulation results. The results of the geometrical 2 and arithmetic-geometric cooling schemes experiments, the results for the initial temperatures $T_0 = 10$ and $T_0 = 100$, and the results for the Markov chain lengths of 1000, 5000, 25000 and 100000 can be found in the appendix.

Second, using the optimal configuration of initial temperature, Markov chain length, and cooling scheme, we attempt to converge to the globally optimal route.

4.1 Cooling Schemes

Figure 2 presents the average route lengths of the four cooling schemes.

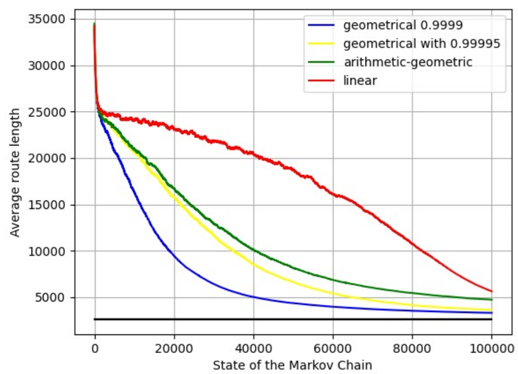
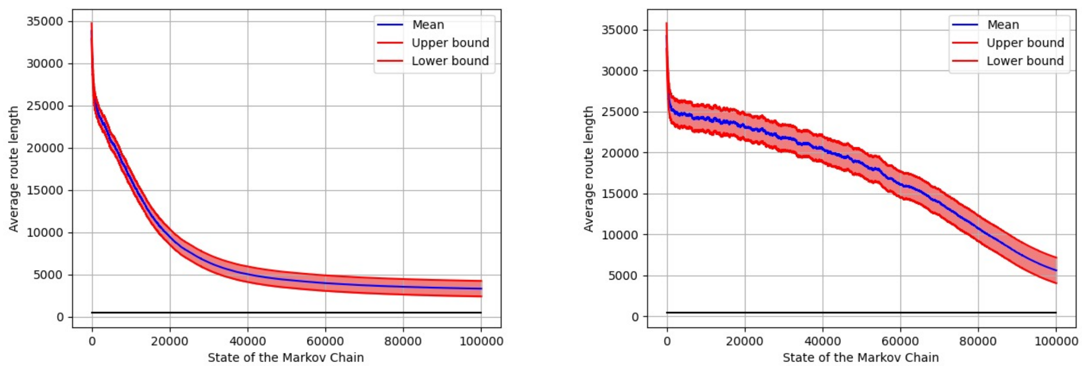


Figure 2: Average route length per cooling scheme over 50 simulations. Confidence intervals are omitted for readability.

Figure 2 shows that the geometric cooling schemes perform better than the linear scheme. They approach the minimum value faster, especially the geometrical 1 cooling scheme. The faster decrease of the temperature under the geometrical 1 scheme leads to lower acceptance rates for worse solutions. Hence, there is more time for climbing hills in exchange for less time for variation. The simulation results suggest this is a favorable trade-off.

Figures 3a and 3b show that the confidence bounds for the geometrical 1 lie closer to the average than for the linear cooling-scheme. Because the temperature goes to zero the slowest in the linear case, the condition of accepting a new route is less restrictive. Therefore, less optimal routes find their way into the Markov chain. Because it is less restrictive, a broader range of states have the possibility to meet the acceptance criterion, explaining the larger confidence bounds seen in 3b.



(a) (b)

Figure 3: The average of the route lengths of the Markov chains of (a) the geometrical 1 approach and (b) the linear approach. The average is taken over 25 simulations.

Thus, the geometrical 1 cooling scheme provides the best approach in reducing the temperature and finding the minimum. We use this cooling scheme for finding our optimal solution for the TSP in 4.4.

4.2 Initial Temperatures

Figure 4 presents the average route lengths for four different initial temperatures.

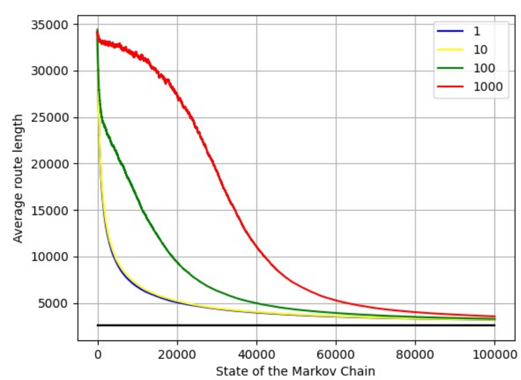
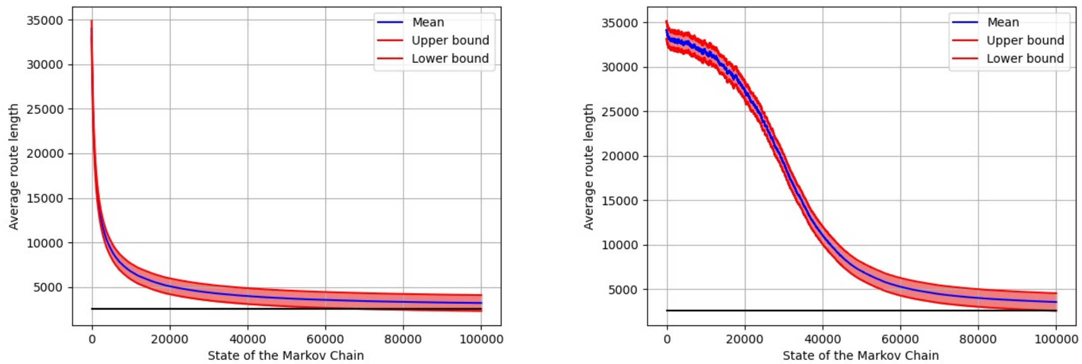


Figure 4: Average route length for four initial temperatures over 50 simulations.

For $T_0 = 1$ and $T_1 = 10$, there is not much difference in the convergence rate. They also both give the smoothest functions. The convergence of $T_0 = 1000$, on the other hand, is not as smooth for the first few thousand states of the the Markov chain. Interestingly, $T_0 = 1$ is not too small such that the distribution of the acceptance condition cannot approach the Dirac-delta distribution of the minimum point. The bigger the T_0, the more the acceptance condition approaches the uniform distribution. This is why fluctuations in the mean route length appear at the initial states, when the temperatures have not been lowered much yet, since the acceptance is less restrictive which means less optimal routes get accepted.

Furthermore, Figure 5 contains the convergence of the initial temperatures 1 and 1000 including confidence intervals. These figures show that variation between runs is limited, similar to when we varied Markov chain length.



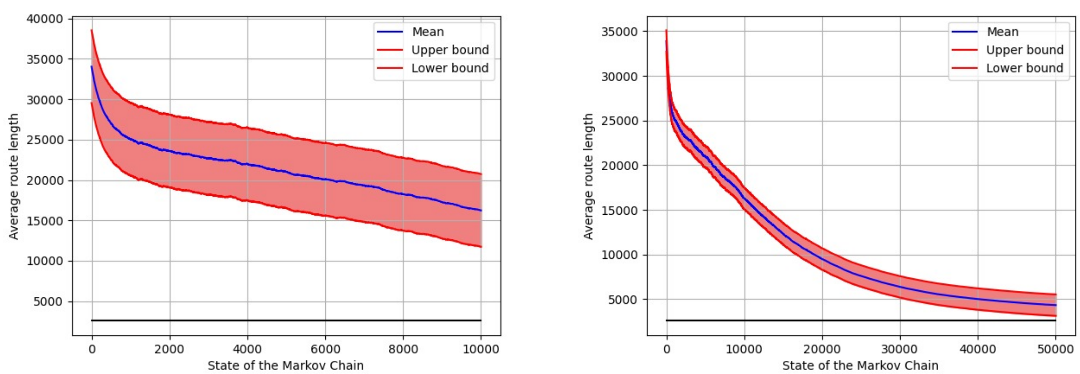
(a) (b)

Figure 5: Average route length for (a) the initial temperature of 1 and (b) the initial temperature of 1000. Averages are taken over 50 simulations.

Since the convergence rate of $T_0 = 1$ slightly outperforming that of $T_0 = 10$, we use $T_0 = 1$ to calculate the optimal solution of the TSP.

4.3 Markov Chain lengths

In Figures 6a, 6b and 7 the average route lengths for Markov chain lengths 10000, 50000 and 150000 are shown.



(a) (b)

Figure 6: Average route lengths when (a) the Markov chain length is 10000 and (b) the Markov chain length is 50000. The average is taken over 50 simulations.

It is apparent from Figure 7 that the larger the Markov chain length, the more the route converges to the optimum route. Therefore, we opt for choosing the highest Markov chain length that we experimented with, which is 150000, to find our optimal solution.

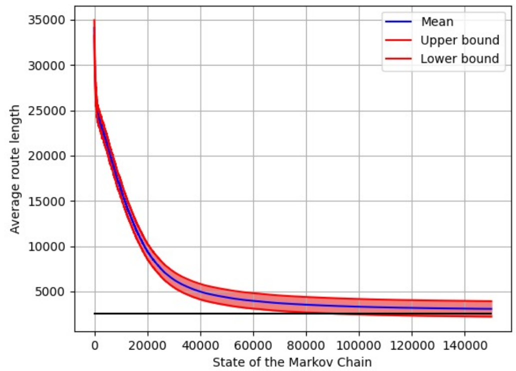


Figure 7: Average route length when the Markov chain length is 150000. The average is taken over 50 simulations.

4.4 Optimal solution

Figure 8 presents the convergence of our best solution to the TSP using simulated annealing. It uses an initial temperature of 1, a Markov chain length of 150000, and the geometrical 1 cooling scheme.

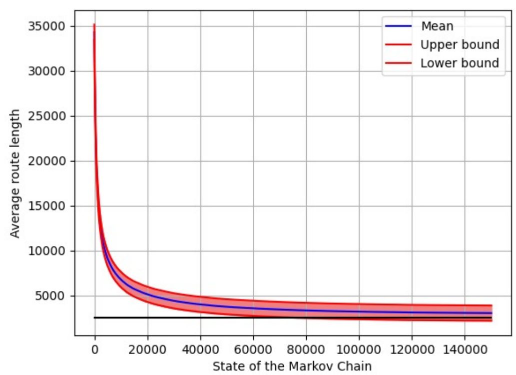
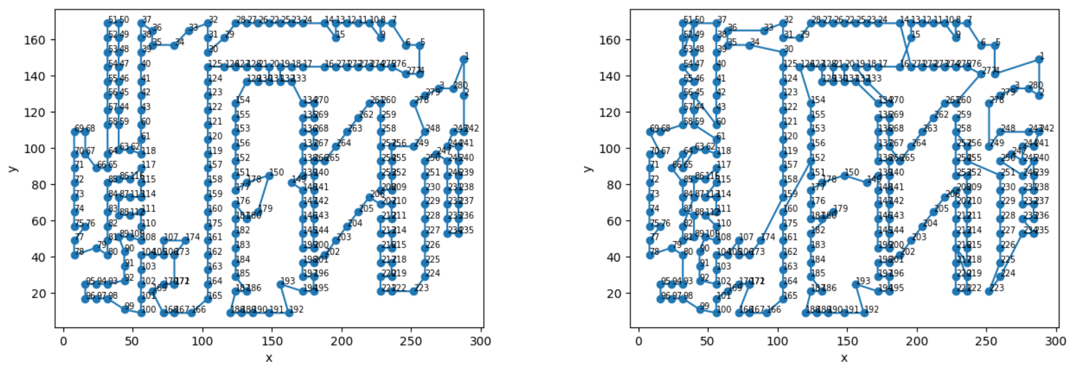


Figure 8: Average route length of the our best solution, where we applied the geometrical 1 cooling scheme, an initial temperature of 1 and a Markov chain length of 150000. The average is taken over 50 simulations.

Figure 8 shows that our solution is fairly close but not identical to the global optimum. Whereas our solution converges to approximately 2900, the global optimum is 2579. Figure 9 shows the differences between the tours.



(a) (b)

Figure 9: Route according to (a) the globally optimal route (b) our simulated annealing approach.

At a glance, the differences between the optimal tour and our solution appear small. Indeed the difference in length is a few hundred meters. Hence, solving the Traveling Salesman Problem using SA yields results that come close to the benchmark minimum route length. Convergence is rapid: most of the gain is reaped within 20000 iterations.

5 Conclusion

In this paper, we investigate the effect of initial temperature, Markov chain length, and cooling schemes on convergence to the global optimum in the simulated annealing method.

To this end, we apply it to the a280 Ludwig drilling instance of the traveling salesman problem. First, we determine which initial temperature, Markov chain length, and cooling scheme yield the best tour. Second, we attempt to find the shortest route for the TSP instance using the best configuration of initial temperature, Markov chain length, and cooling scheme in our simulation results, and compare the results of the best configuration to the true shortest route.

First, our simulation results indicate the geometrical schemes outperform arithmetic-geometric and linear schemes. Second, they indicate that a low initial temperature yields faster convergence. Third, in the range of Markov chain lengths we experiment with, the longer the Markov chain, the better the solution we find.

Based on these results, we attempt to convergence to the true solution of the problem instance. Our solution is less than twenty percent longer than the global optimum route. Furthermore, convergence is rapid. Hence, we conclude that simulated annealing is an effective method to solve instances of travelling salesman problems.

Further research could investigate other instances of travelling salesman problems to verify whether our conclusions with respect to initial temperature, cooling scheme, and Markov chain length are robust across instances. Alternatively, it could examine whether the returns to increasing the Markov chain length become negative for larger lengths.

References

[1] Kirkpatrick, S.; Gelatt Jr, C. D.; Vecchi, M. P. (1983). "Optimization by Simulated Annealing". *Science*. 220 (4598): 671–680.

[2] Nourani, Y. & Andresen, B. (1998). A comparison of simulated annealing cooling strategies. *Journal of Physics A: Mathematical and General*, Vol. 31, No. 41, pp. 8373–8385.

[3] Arora, J. S. (1989). *Introduction to optimum design*. New York: McGraw-Hill.

[4] Mahdi, Walid & Medjahed, Seyyid Ahmed & Ouali, Mohammed. (2017). Performance Analysis of Simulated Annealing Cooling Schedules in the Context of Dense Image Matching. Computación y Sistemas. 21. 493-501. 10.13053/CyS-21-3-2553.

[5] Geman, S. & Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 6, No. 6, pp. 721–741.

[6] Ortner, M., Descombes, X., & Zerubia., J. (2007). Adaptive simulated annealing cooling schedule for object detection in images. *Research Report*, INRIA, Vol. 6336.

[7] Merrill M. Flood, (1956) The Traveling-Salesman Problem. Operations Research 4(1):61-75.

[8] Dantzig, George Bernard, D. R. Fulkerson, and Selmer Martin Johnson, Solution of a Large-Scale Traveling-Salesman Problem. Santa Monica, CA: RAND Corporation, 1954.

[9] Englert, M., Röglin, H., & Vöcking, B. (2007, January). Worst case and probabilistic analysis of the 2-Opt algorithm for the TSP. In SODA (pp. 1295-1304).

6 Appendix

In the the appendix we will show the results not discussed in detail in the Results section.

6.1 Cooling Schemes

In this section of the appendix we will show the results not discussed in detail in the Results section.

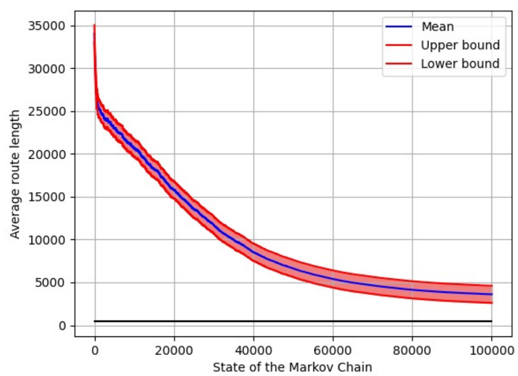


Figure 10: The average of the route lengths of the Markov chains for the geometrical 2 approach. The average is taken from 50 simulations.

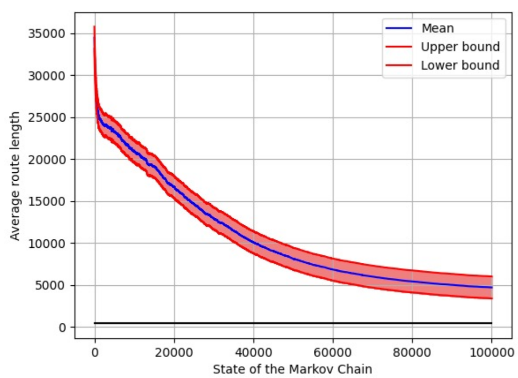


Figure 11: The average of the route lengths of the Markov chains for the arithmetic-geometric approach. The average is taken from 50 simulations.

6.2 Initial Temperatures

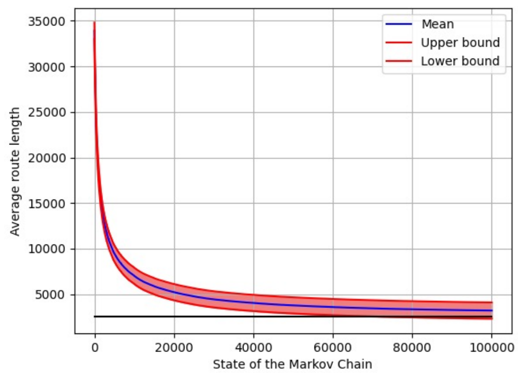


Figure 12: The average of the route lengths of the Markov chains for the initial temperature of 10. The average is taken over 50 simulations.

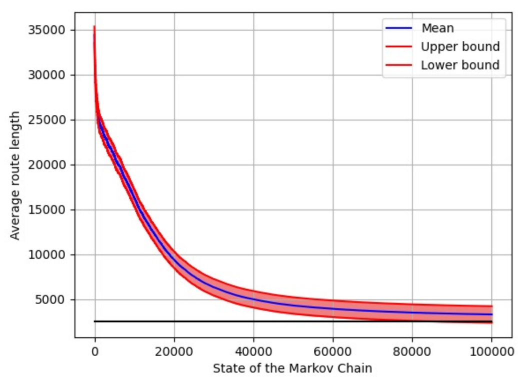


Figure 13: The average of the route lengths of the Markov chains for the initial temperature of 100. The average is taken over 50 simulations.

6.3 Markov Chain lengths

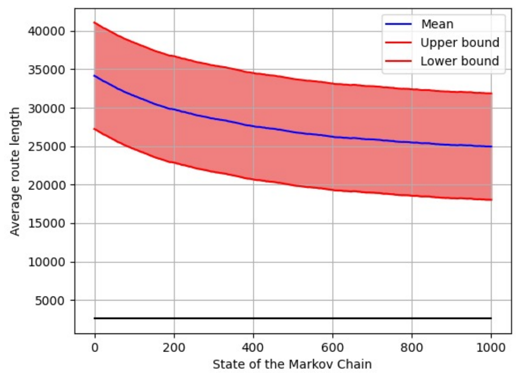


Figure 14: The average route length when the Markov chain length is 1000. The average is taken over 50 simulations.

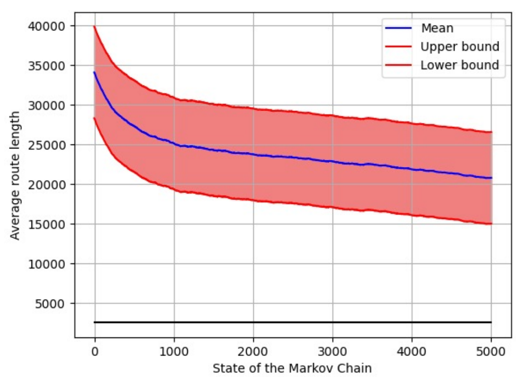


Figure 15: The average route length when the Markov chain length is 5000. The average is taken over 50 simulations.

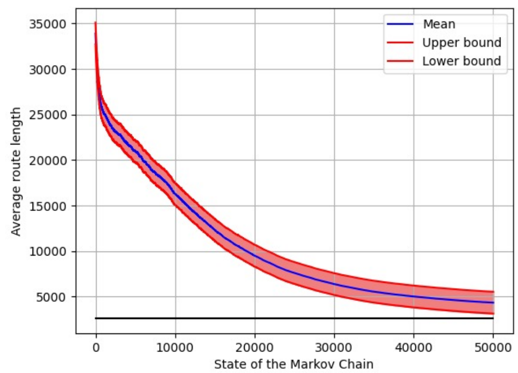


Figure 16: The average route length when the Markov chain length is 25000. The average is taken over 50 simulations.

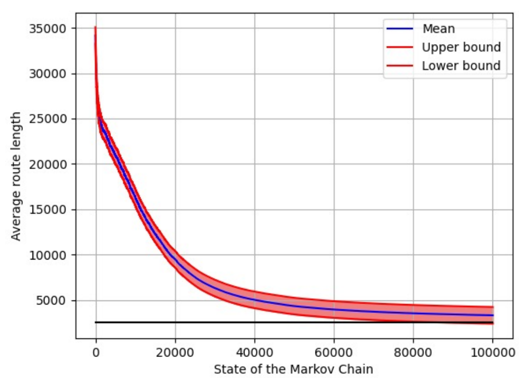


Figure 17: The average route length when the Markov chain length is 100000. The average is taken over 50 simulations.