4th problem set for Computer Simulations, 2024

Hand-in: see schedules for your group, linked in the TeachCenter.
Please upload/submit your program files in the TeachCenter *before* the conversation with the tutor.

Traveling Salesman

Write a Simulated Annealing program for solving the Traveling Salesman Problem.

- Generate the distribution of N towns in Matlab by

  ```
  N=25;   seednumber=122;
  rand('seed',seednumber);
  xy=rand(N,2);
  ```

  If you do not use Matlab, you can also obtain these xy-coordinates from the file *city-positions.txt* provided in the TeachCenter.

- Note that the „random" number sequence is determined deterministically by the seednumber. After generating the towns, you should therefore set the random number seed differently each time for your actual simulated annealing runs, for example by using the clock time.

- The "energy" is determined by the total length of the round trip:

$$E(\{\mathbf{x}\}) = \sum_{i=1}^{N} |\mathbf{x}_{i+1} - \mathbf{x}_i| \quad \text{mit} \quad \mathbf{x}_{N+1} = \mathbf{x}_1.$$

  The starting point is arbitrary.

- Use the Boltzman distribution

$$p_E(\{\mathbf{x}\}|T) = \frac{1}{Z} \exp\left\{-\frac{1}{T} E(\{\mathbf{x}\})\right\}$$

- Propose a new configuration K' by the following rule: Select two towns S_a and S_b at random. Exchange their position in the path:

K**:** $\quad S_1, \ldots, S_{a-1}, \underbrace{S_a, S_{a+1}, \ldots, S_{b-1}, S_b}_{old}, S_{b+1}, \ldots, S_N$

K'**:** $\quad S_1, \ldots, S_{a-1}, \underbrace{S_b, S_{b-1}, \ldots, S_{a+1}, S_a}_{new}, S_{b+1}, \ldots, S_N$

The order of towns in between S_a and S_b is reversed by this procedure. The corresponding energy difference is easy to calculate. For the case $a < b$ and $(a, b) \neq (1, N)$, it is

$$E(K') - E(K) = \overline{S_{a-1}S_b} + \overline{S_aS_{b+1}} - \overline{S_{a-1}S_a} - \overline{S_bS_{b+1}}$$

- Acceptance probability: $\min\left(1, \frac{p_E(K'|T)}{p_E(K|T)}\right)$
- Number of steps per temperature: rule of thumb: $n_{steps} \simeq N^2$
- Cooling strategy: For this example, it is sufficient to choose the k-th temperature as

  $$T_k \;=\; T_{\text{start}}\, k^{-q} \quad \text{with} \;\; q \;\simeq\; 1$$

  (This can result in very fast cooling, and may not be suitable for other applications of simulated annealing.)
- Use $T_{\text{start}} = 1$.
- For each temperature separately, determine the best energy which has been reached, and the following averages
  - average energy: $\overline{E} = \frac{1}{L}\sum_j E_j$
  - variance: $(\Delta E)^2 \simeq \overline{E^2} - \overline{E}^2$

  Plot these three values (*not* the running averages) versus inverse temperature β (or versus the index k) Also plot the current path several times during the simulation if possible.
- Important: find and use a suitable convergence criterion.
- Plot the best path. How long is it ?
- You cannot be certain to have reached the global optimum. Experiment with different choices of parameters, namely with
  - different values of q and n_{steps},
  - different sequences of random numbers (*after* generating the location of towns)
  - and also try to further improve on the result of a run, by using it as the starting configuration of an additional optimization, which should then be started at a higher temperature (≤ 1, your choice).

(continued on following page)

Bonus (2 points): Jackknife

This calculation is separate from the optimization above. It treats a very simple example of the Jackknife procedure.

Choose some temperature T for which you have observed sizable fluctuations in the energy above. Run your traveling salesman program for many steps n_{steps} (see below) at that fixed temperature. Measure and record the energy at each step. Subdivide your measurements into $N_B = 30$ blocks and calculate:

1. the variance $(\Delta E)^2$ of the energy and the uncertainty of that variance by using the Jackknife procedure.
2. For comparison, calculate the variance separately in each block (not Jackknife) and estimate its uncertainty from the 30 values thus obtained.

Perform this comparison separately for $n_{steps} = 100, 1000, 10000, 100000$ (if feasible; maybe more). Use the same starting configuration (path) each time. Do not plot the path, to save time.

Is the Jackknife analysis indeed better? That is, for a given number of data, does it provide statistically correct estimates for the variance with a smaller uncertainty than the second method, as one would expect ?

Note that the advantages of Jackknife become more pronounced in more complicated analyses, like ratios of averages, fits of fit results, etc.