Master de Physique
Faculté de Physique & Ingénierie

Faculté de physique et ingénierie
Université de Strasbourg

Optimisation : Simulated annealing, the travelling salesman problem

C Boily, F Thalmann, R Schotter

Season 2023-24

Abstract

Optimisation consists in finding the optimal (or if it is too hard, a very good) configuration of a complex problem. Goodness takes here the form of a mathematical cost function. Beside standard techniques of function minimisation (*e.g.* dichotomy, gradient descent, conjugate gradient descent...) there are a number of approaches that sometimes outperform the previous ones especially in the presence of many competing local minima. One of them is the **simulated annealing**. The term annealing comes from the field of material science where it refers to heating/cooling cycles in order to heal defects and improve material properties such as crystallinity. Simulated annealing is strongly linked to statistical physics and Monte-Carlo canonical simulations.
The travelling salesman is a discrete mathematics - graph theory optimisation problem. Given a set of cities (graph vertices) and paths (graph edges connecting pairs of vertices), one looks for the Hamiltonian path (a path that passes once and only once by each city) that minimises the total length (or total travelling time). The hardness of the problem comes from the fact that graph edges can have arbitrary distances attached to them and that the number of configurations (the number of hamiltonian paths) grows very fast with the number of cities (as the number of permutations $N!$ with N the number of cities).
The aim of the project is to deploy a simulated annealing approach to solving the travelling salesman problem in a few funny and realistic situations.

1 Simulated annealing, Monte-Carlo and Statistical Physics

Discretisation of a configuration space

Let us consider a set of configurations $\{\mathcal{C}_\alpha\}$ with index $1 \leq \alpha \leq \mathcal{N}$, $\mathcal{N}$ large. Each configuration has an energy E_α. The canonical distribution is the unique probability distribution p_α such that the ratio for any pair (α, γ) of configurations obeys

$$\frac{p_\alpha}{p_\gamma} = \exp(\beta(E_\gamma - E_\alpha)) \tag{1}$$

β being an "inverse temperature" arbitrary parameter. The numbers p_α are positive and adds to one: $\sum_\alpha p_\alpha = 1$.

Importance sampling and Metropolis algorithm

Importance sampling consists in generating a "trajectory" visiting the configurations $\mathcal{C}_\alpha$ with unequal probabilities, the chances of visiting $\mathcal{C}_\alpha$ being proportional to $\exp(-\beta E_\alpha)$ (or alternatively p_α). In this way, configurations are visited (sampled) according to their "importance" $\exp(-\beta E_\alpha)$. Low energy configurations can be visited many times while high energy configurations are hardly explored Metropolis et al. (1953); Krauth (2006). Taking i as a discrete time index, $\alpha[i]$ as the configuration at time step i, the Metropolis algorithm is the following.
At each step i, pick-up at random a configuration $\mathcal{C}_\gamma \neq \mathcal{C}_{\alpha[i]}$. Consider $\alpha[i] \to \gamma$ as a move.

- if $E_\gamma \leq E_{\alpha[i]}$ accept the move, *i.e.* set $\alpha[i+1] = \gamma$

- if $E_\gamma > E_{\alpha[i]}$, draw a uniform random number $0 < r < 1$. Accept the move if $r < \exp(-\beta(E_{\alpha[i]} - E_\gamma))$, *i.e.* set $\alpha[i+1] = \gamma$. Reject the move if $r \geq \exp(-\beta(E_{\alpha[i]} - E_\gamma))$, *i.e.* set $\alpha[i+1] = \alpha[i]$.

A T steps trajectory $\alpha[i]$ is a sample of configurations. In many cases of interest one can show that the time average of any observable $\mathcal{O}(\mathcal{C}_\alpha)$

$$\overline{\mathcal{O}} = \lim_{T\to\infty} \frac{1}{T} \sum_{i=1}^{T} \mathcal{O}(\mathcal{C}_{\alpha[i]}) \tag{2}$$

converges towards the canonical average

$$\langle \mathcal{O} \rangle = \sum_{\alpha=1}^{\mathcal{N}} \mathcal{O}(\mathcal{C}_\alpha) p_\alpha \tag{3}$$

Note the absence of weighting in the first expression, the canonical distribution being realized a priori. The Metropolis algorithm lies at the basis of the Monte-Carlo sampling of canonical spaces.

Quenching

In the limit of the inverse parameter β going to infinity ∞, and in the case of a non degenerated minimum, the canonical probability is shown to concentrate into the optimal configuration $\alpha = A$, *i.e.* the one of minimal energy. In other words $\forall \alpha; E_\alpha > E_A$ and $p_\alpha \to \delta_{\alpha A}$.

The idea of simulated annealing consists in using the Metropolis algorithm, starting from high temperature (low β) and decreasing temperature progressively (increasing β) until it finds the optimal state. With a lucky time temperature evolution, it is possible to sample the global configuration space, escape from local minima and end up eventually in the global minimum of the problem. In the absence of local minimum the algorithm will always find the global minimum (one may have to reduce the step size as the same time as temperature decreases). Ideally the acceptance/rejection rate of the moves should be not too high (too close to 1) nor to low (too close to 0). There is room for many heuristic strategies.

2 The travelling salesman case

General idea

Let us consider N cities, indexed by j. Every pair of cities j, k is assigned a distance $d_{j,k} = d_{k,j}$ or equivalently a time of travel. The problem can be abstracted to a graph with weighted edges. It can also be mapped onto a real symmetric matrix $\mathbf{D} = d_{j,k}$. Setting $d_{j,k}$ to a very large value amounts to disconnecting the given pair of cities.

A configuration $\mathcal{C}_\alpha$ of the problem is simply a permutation $J[j]$ of the N cities that minimises the total cost function

- $$C_c(J) = \sum_{j=1}^{N} d_{J[j],J[j+1]} \tag{4}$$

  with $J[N+1] \equiv J[1]$, corresponding to a closed path, the salesman coming back to the starting place. Check that the starting place does not matter and that without loss of generality one can decide that $J[1] = 1$.

- $$C_o(J) = \sum_{j=1}^{N-1} d_{J[j],J[j+1]} \tag{5}$$

  corresponding to an open path. The starting place then obviously matters.

To generate configurations (*i.e.* permutations) it is convenient to adopt the mathematical notation (j, k) corresponding to a transposition $j \to k$, $k \to j$. For instance the permutation of $\{1, 2, 3, 4, 5, 6\}$ subject to the composition of 3 permutations (1,4), (1,3), (5,6) is $\{4, 2, 3, 1, 6, 5\}$ and $J[1] = 4, J[2] = 2, J[3] = 3, J[4] = 1, J[5] = 6, J[6] = 5$.

Every permutation can be decomposed into a product of transpositions. Applying transpositions with randomly chosen j and k will sample uniformly the set $\mathfrak{S}_N$ of all possible permutations. Transpositions of randomly chosen pairs are natural candidates for the "moves" needed in the Metropolis algorithm.

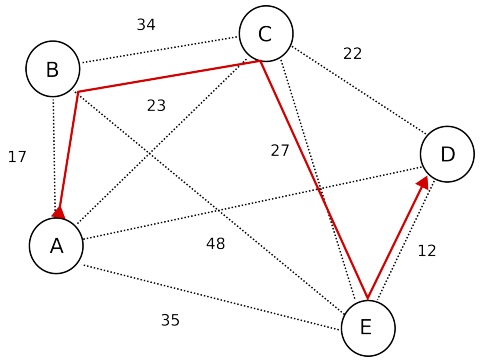

Figure 1: Example of complete graph of cities and a particular hamiltonian path

3 Simple trial case

Consider the following set of 6 French cities: { Strasbourg, Nancy, Paris, Mulhouse, Dijon, Besançon}. Use an application such as Google Map for finding out the travel times or distances between any pair of cities (15 pairs total).

Find an appropriate numerical representation for the distance matrix **D**. Find as well a convenient representation for any arbitrary configuration/ permutation $J[j]$. Of course the costs functions $C_o(J)$, $C_c(J)$ must be defined.

The number of cities is sufficiently small for an exhaustive enumeration of all configurations (720 permutations, the famous greedy algorithm). You will find out the best closed path, the best open path starting from Strasbourg, the best open path (starting point to determine). If curious, you can compare train and car trips (the result may depend on conditions, days... just make your example explicit).

Try now to implement a Metropolis sampling of the configurations. Use random number generators provided by packages such as `random` or `numpy.random`. There is some freedom as far as associating an "energy" E_J to the cost function $C(J)$. The exponentiation $\exp(\beta\Delta E))$ should not lead to ridiculously small or large numbers, so consider first something like $E_J = \ln(C(J))$.

An important issue is whether the cost function needs or not to be reevaluated completely at each step. If you choose to go for an optimized evaluation of energy differences such as the one appearing in the Metropolis move, make sure it is consistent with the true cost function.

You are now ready to perform a long sampling $J_i[j]$ with i the time index and j the city index. Increase β progressively and keep track of the best solution $J_{\text{opt}}[j]$ obtained after a time T

$$C(J_{\text{opt}}) = \min_{i \le T} C(J_i) \tag{6}$$

Compare the result of your simulated annealing with your exhaustive greedy algorithm. A plot of $C(J_{\text{opt}})$ *vs* T is usually instructive. Use the Unix `time` command to estimate the efficiency of your codes.

4 Earth Travel

Simulated annealing may not seem competitive if too few cities are considered. This changes as N increases. We suggest now the following development. Below is a list of 18 cities:
{ Paris, Madrid, Athens, Helsinki, Beyrouth, New Delhi, Bangkok, Beijing, Tokyo, Seoul, Sidney, Buenos Aires, Brasilia, Caracas, Mexico City, Chicago, Quebec, Reykjavik }.

Find out (Google Maps, Wikipedia) the coordinates (longitude ϕ, latitude λ) of the cities. Compute the mutual distances using the Gaussian spherical geometry (grand circle). Inspired by spherical coordinates each city j corresponds a unit vector $\mathbf{u}_j = (\cos(\lambda)\cos(\phi), \cos(\lambda)\sin(\phi), \sin(\lambda))$. Then define

$$d_{j,k} = R_\oplus \arccos(\mathbf{u}_j \cdot \mathbf{u}_k) \tag{7}$$

with $R_\oplus$ the Earth radius, assumed to be spherical.

Fill in the distance matrix $\mathbf{D}$.
Weight up using the Stirling approximation whether the greedy algorithm is worth a trial. Then adapt the previous heuristic simulated annealing approach.

You can change the random number seed and see if your result is sensitive to it. How do different good solutions differ from each other ?

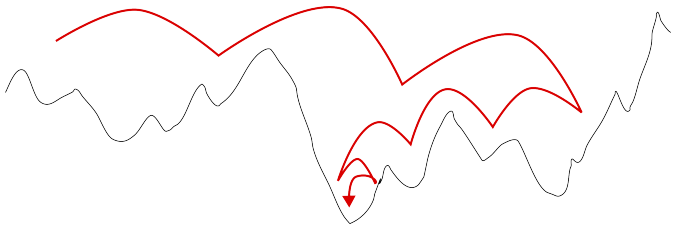

Figure 2: Illustration of a simulated annealing search

References

Krauth W., 2006, Statistical Mechanics, Algorithms and Computations. Oxford Master Series in Statistical, Computational and Theoretical Physics, Oxford University Press, New York

Metropolis N., Rosenbluth A. W., Rosenbluth M. N., Teller A. H., Teller E., 1953, The Journal of Chemical Physics, 21, 1087