Lecture 11: Simulated Annealing

- Annealing
- Metropolis algorithm
- Simulated Annealing
- Statistical properties
- Markov theory
- Cooling schemes
- Telecrypto again
How are fine crystals obtained?

The idea for obtaining fine crystals in a solid with low energy:

- First melt the solid by increasing the temperature.
- Then, slowly cool it so it crystallizes into a perfect lattice.

Converse: annealing - quenching.
The Metropolis algorithm


Suppose a state of the solid with energy $E_i$ and subsequent energy state $E_j$.

If the energy difference $E_j - E_i$ is less than or equal to zero, then state $j$ is accepted as current state.

Otherwise, let the probability of transforming be proportional to

$$
\exp\left(\frac{E_i - E_j}{k_B T}\right)
$$

T - Temperature, $k_B$ - Boltzmann constant.
If the lowering of the temperature is sufficiently slow, the solid reaches thermal equilibrium at each temperature.

At thermal equilibrium the probability distribution is given by

\[ P_T(X = i) = \frac{1}{Z(T)} \exp\left(\frac{-E_i}{k_B T}\right) \]

\[ Z(T) = \sum_j \exp\left(\frac{-E_j}{k_B T}\right) \]

The Boltzmann distribution
Simulated annealing

We want to minimize $f(x)$. Let the domain be denoted $S$ according to Simulated Annealing and Boltzmann Machines.

Equivalences between a physical many-particle system and optimization:

- Solutions - states of the physical system
- Objective function - energy
- Control parameter - temperature

Suppose a feasible solution $x_i$ is given. Should one change to solution $x_j$? Acceptance criterion:

$$P_c = \begin{cases} 
1 & \text{if } f(x_j) \leq f(x_i) \\
\exp\left(\frac{f(x_i) - f(x_j)}{c}\right) & \text{if } f(x_j) > f(x_i)
\end{cases}$$

A transition consists of two steps.

1. Generate a new solution $x_j$
2. Decide whether the new solution should be accepted or not
Algorithm

1. Init \((x_i, c, L)\),

2. Perform \(L\) transitions of the type:
   (a) Generate a new solution \(x_j\)
   (b) If \(x_j\) is accepted (according to probability \(P_c\)), set \(x_i = x_j\).

3. Update \(c\) and \(L\)

4. if the stop criterion is not fulfilled, goto 2.
Conjecture: After (sufficiently) many steps at a fixed value of $c$, the probability of being in $x_i$

$$P_c(X = x_i) = \frac{1}{N_0(c)} \exp\left(\frac{-f(x_i)}{c}\right)$$ \hspace{1cm} (1)$$

where

$$N_0(c) = \sum_{x \in S} \exp\left(\frac{-f(x)}{c}\right)$$

is a normalization constant.

What happens when $c \to 0$?

**Theorem 1.**

$$\lim_{c \to 0} P_c(x_i) = \begin{cases} 
\frac{1}{N_{opt}} & \text{if } x_i \text{ is optimal} \\
0 & \text{otherwise}
\end{cases}$$

where $N_{opt}$ is the number of optimal points.

Proof.
Example:
Suppose we have $S = \{x_1, \ldots, x_7\}$ with values \{3, 1, 4, 7, 6, 2, 1\} of the objective function. Then the Boltzmann and the empirical estimated distribution at $c = 10, 1, 0.1$ according the below plots.
Suppose we reach thermal equilibrium at temperature $c$.

The **expected cost** of the objective function is

$$E_c(f) = E[f(x_i)] = \sum_{x \in S} f(x)P_c(x)$$

The **expected square cost** is

$$E_c(f^2) = E[f^2(x_i)] = \sum_{x \in S} f^2(x)P_c(x)$$

The **variance of the cost** is

$$\sigma_c^2(f) = V[f(x_i)] = \sum_{x \in S} (f(x) - E_c(f))^2P_c(x)$$

The **entropy** of the objective function is

$$S_c = -\sum_{x \in S} P_c(x) \ln(P_c(x))$$
At each level of the control parameter $c$, one can estimate:

- The acceptance ratio $\chi(c)$ - the ratio of the number of accepted transitions and the number of proposed transitions.

- The average of the objective function

$$\bar{f}(c) \approx E_c(f)$$

$$\bar{f}(c) = \frac{1}{L} \sum_{i=1}^{L} f(x_i)$$

The standard deviation $\bar{\sigma}(c) \approx \sigma_c$

$$\bar{\sigma}(c) = \left( \frac{1}{L} \sum_{i=1}^{L} (f(x_i) - \bar{f}(c))^2 \right)^{1/2}$$

Note the typical behaviours of these three functions.
A quantitative analysis

Introduce the solution density

\[ \omega(f)df = \frac{1}{|S|} \left| \{ x_i \in S \mid f \leq f(x_i) < f + df \} \right| \]

The corresponding solution density at thermal equilibrium is

\[ \Omega(f, c)df = \frac{\omega(f) \exp(\frac{f_{opt} - f}{c}) df}{\int \omega(f') \exp(\frac{f_{opt} - f'}{c}) df'} \]

It is possible to express \( E_c(f) \) and \( \sigma_c^2 \) using \( \Omega(f, c) \)

\[ E_c(f) = \int f' \Omega(f', c) df' \]

\[ \sigma_c^2 = \int ( (f' - E_c(f))^2 ) \Omega(f', c) df' \]

If one could estimate or guess \( \omega(f) \), then one could compute how both \( E_c(f) \) and \( \sigma_c^2 \) would look like.
Example of solution density.
A crude guess of $\omega(f)$.

Postulate: A typical optimization problem has a solution density that can be approximated with a normal distribution in a region $R_1$ close to $E_\infty(f)$ and an exponential distribution in a region $R_2$ close to $f_{opt}$.

Given these approximations, then one can compute $E_c$ and $\sigma_c$ for large values of $c$ (region $R_1$) and small (region $R_2$).
Markov theory

Simulated annealing can be regarded as a markov chain

Let $p(k)$ be vector where element $i$ denote the probability of being in solution $x_i$ after $k$ steps. Then, the following matrix equation holds

$$p_k = A(c)p_{k-1}$$

where $A(c)$ is a transition matrix dependent on $c$.

With markov theory, one can show that for each fixed value of $c$, then $p(k)$ converges to a stationary distribution (1).

One can also estimate convergence speed and get information of how many steps are necessary to be close to the stationary distribution.
What does the convergence analysis tell us?

Consider the traveling salesman problem with $n$ cities and let two tours be neighbours if one tour can be obtained from the other by replacing two edges.

Then one can show that the stationary solution with a tolerance $\epsilon = 1/n$ is reached after $k = O(n^{2n-1})$ steps.

Thus, exhaustive search is faster!
Cooling schedules

Rigorous application of simulated annealing is not practically possible.

Find an ad hoc cooling schedule which is practical and keeps the distribution close to thermal equilibrium.

The cooling schedule includes:

- Choice of start temperature \( c_0 \). Sufficiently large to accept all transitions.

- Choice of lowering temperatures \( c_k \). Often \( c_{k+1} = \alpha c_k \), with \( 0.8 \leq \alpha \leq 0.99 \).

- Choice of stop criterion. For example, when the same solution is stuck under several \( k \).

- Choice of transition jumps \( L_k \).

If one makes too large steps in \( c \), one has to compensate with larger \( L \) to reach thermal equilibrium.

Cooling schedule with polynomial time complexity
Telekryptot igen!

Rutin för poängsättning och lösningsgenerering.

jogga - genererar en slumpmässig granne
poangsatt - målfunktionen

% Initiering
eval(initt);
xi=eval(slump);
fi=eval([poangsatt,’(xi)’]);
c=10; L=5000;
fortsatt=1;
while c>0.001,
  %Uppdatera c  
c=c*0.98

for i=1:L;
  % Generera xj.
  xj=eval(['jogga,',(xi)']);
  fj=eval(['poangsatt,',(xj)']);
  if rand<exp( (fi-fj)/c ),
      % Accepterad övergång.
      xi=xj;
      fi=fj;
      chi(ci)=chi(ci)+1;
  end;
end;
end;
$\chi, E_c(f), \sigma_c$
Kod: . ågvsehinäröökydlaumtpqbczfwxj Poang: 6.662
Text: asyr tio sko iust tara mata est oika
grad äia vaktai ave slråtsa us amargam iedan
svån ek sadat ar ör iu om aiena se låre
kroka kuihama er maskolå are kaiåna van
lank genå ond heig akpt odui sei arm gast
mästeiskal skail ku urand tar ien rea so
koi tiamla tie sed ån gus svada rake otta
atre anai kun eiåk rots krå saia skar not
kiav ioman ona må eka dam ananas iög ra
hog rag ove stava ös laiås demoreïång