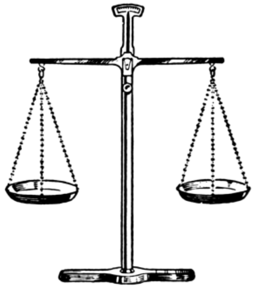


Simulated Annealing

Simulated annealing

- **Unconstrained optimization method**

$$\min_{\boldsymbol{x}} f(\boldsymbol{x})$$

- **Simulated annealing**
  - Start from an initial point
  - Repeatedly consider various new solution points
  - Accept or reject some of these solution candidates
  - Converge to the optimal solution

Simulated annealing

- **Unconstrained optimization method**

$$\min_{\boldsymbol{x}} f(\boldsymbol{x})$$

- **Simulated annealing**
  - It was introduced by Metropolis in 1953
  - It is based on “similarities” and “analogies with the way that alloys manage to find a nearly global minimum energy level when they are cooled slowly.

Simulated annealing

- Simulated Annealing is a **stochastic optimization method**
  - It uses randomness strategically to explore the solution space
  - Randomness can help escape local minima
  - This increases the chance of searching near the global optimum

Local optimization vs simulated annealing

Local optimization

Start from an initial point

Repeatedly consider various new solution points

Reduce cost function at each iteration

Converge to optimal solution

Local optimization vs simulated annealing

Local optimization

Start from an initial point

Repeatedly consider various new solution points

Reduce cost function at each iteration

Converge to optimal solution

Simulated annealing

Start from an initial point

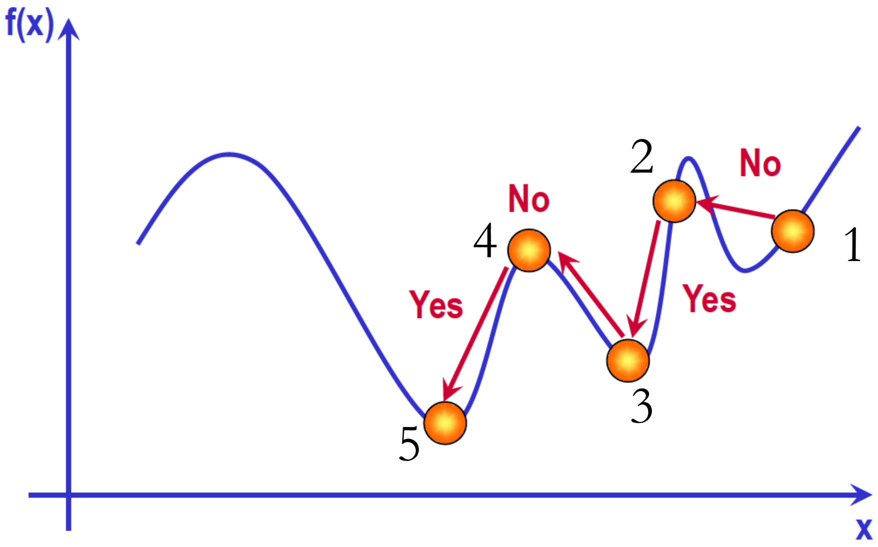
Repeatedly consider various new solution points

Accept/reject new solution using probability at each iteration

Converge to optimal solution

Local optimization vs simulated annealing

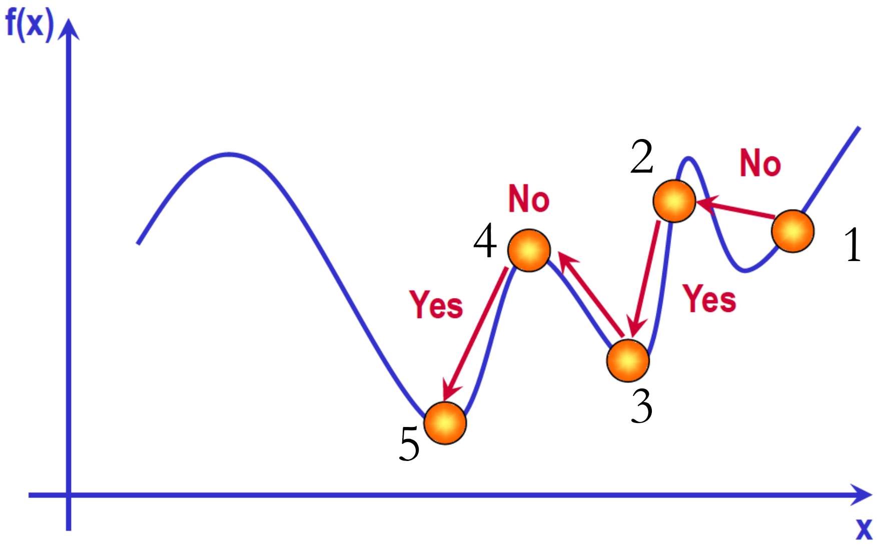
Local optimization



Local optimization attempts to reduce cost function at each iteration

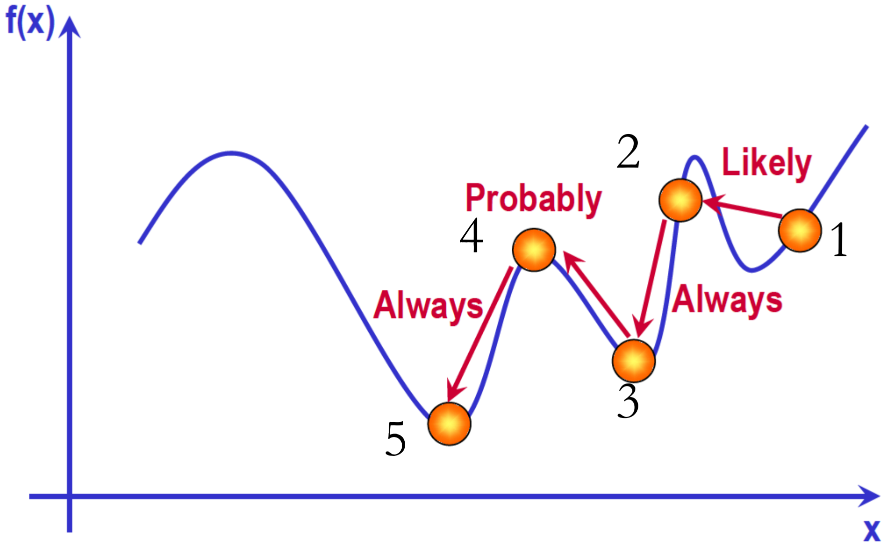
Local optimization vs simulated annealing

Local optimization



Local optimization attempts to reduce cost function at each iteration

Simulated annealing



Simulated annealing accept/reject new solution candidate based on probability

Simulated annealing

- Intelligently controls the degree of randomness added to stochastic search methods
- Initially, the randomness added to function evaluations is large
- The “temperature” is then slowly lowered according to a predetermined “annealing schedule”

Simulated annealing

Step 1: start from an initial point $X = X_0$ & $K = 0$

Step 2: evaluate cost function $F = f(X_K)$

Step 3: randomly move from X_K to a new solution X_{K+1}

Step 4: if $f(X_{K+1}) < F$, then

Accept new solution

$X = X_{K+1}$ & $F = f(X_{K+1})$

End if

Step 5: if $f(X_{K+1}) \geq F$, then

Accept new solution with certain probability

$X = X_{K+1}$ & $F = f(X_{K+1})$ iff $rand() < \varepsilon$

End if

Step 6: $K = K + 1$ & go to Step 2

Simulated annealing

Step 1: start from an initial point $X = X_0$ & $K = 0$

Step 2: evaluate cost function $F = f(X_K)$

Step 3: randomly move from X_K to a new solution X_{K+1}

Step 4: if $f(X_{K+1}) < F$, then

Accept new solution

$X = X_{K+1}$ & $F = f(X_{K+1})$

End if

Step 5: if $f(X_{K+1}) \geq F$, then

Accept new solution with certain probability

$X = X_{K+1}$ & $F = f(X_{K+1})$ iff $rand() < \varepsilon$

End if

Step 6: $K = K + 1$ & go to Step 2

Simulated annealing

Step 5: if $f(X_{K+1}) \geq F$, then

Accept new solution with certain probability

$X = X_{K+1}$ & $F = f(X_{K+1})$ iff rand() < ε

End if

- Option 1 : constant probability ε (e.g. ε=0.1)
- Option 2 : Dynamically varying probability, i.e., decreasing over time

Simulated annealing

Step 5: if $f(X_{K+1}) \geq F$, then

Accept new solution with certain probability

$X = X_{K+1}$ & $F = f(X_{K+1})$ iff rand() $< \varepsilon$

End if

- Metropolis criterion probability of acceptance

(Use Boltzmann distribution to determine the probability)

$$\varepsilon = exp\left[-\frac{f(X_{k+1})-F}{T_{k+1}}\right]$$

T_{k+1} is a "temperature" parameter that gradually decreases

E.g., $T_{k+1} = \alpha \cdot T_k$ where $\alpha < 1$

Simulated annealing

Step 5: if $f(X_{K+1}) \geq F$, then

Accept new solution with certain probability

$X = X_{K+1}$ & $F = f(X_{K+1})$ iff rand() <= $exp\left[-\frac{f(X_{k+1})-F}{T_{k+1}}\right]$

End if

High temperature

Attempt to accept all new solutions even if $[f(X_{k+1}) - F]$ is large

Low temperature

Only accept the new solutions where $[f(X_{k+1}) - F]$ is small

Simulated annealing

Temperature Annealing schedules

- Logarithmic annealing schedule

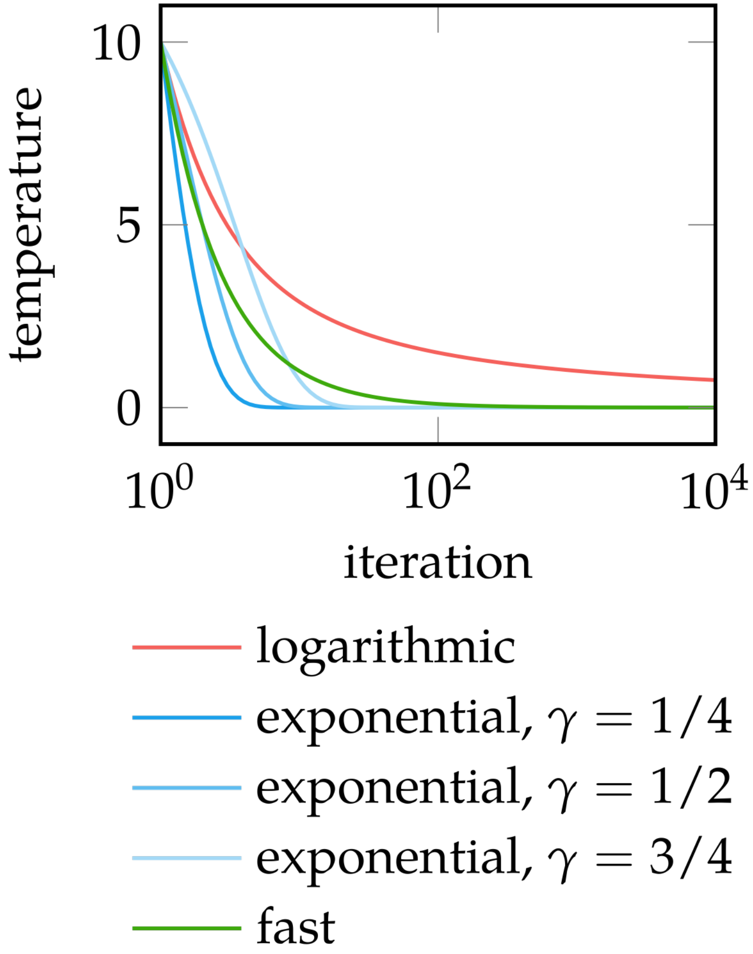
$$t^{(k)} = t^{(1)} \frac{\ln(2)}{\ln(k+1)}$$

- Exponential annealing schedule

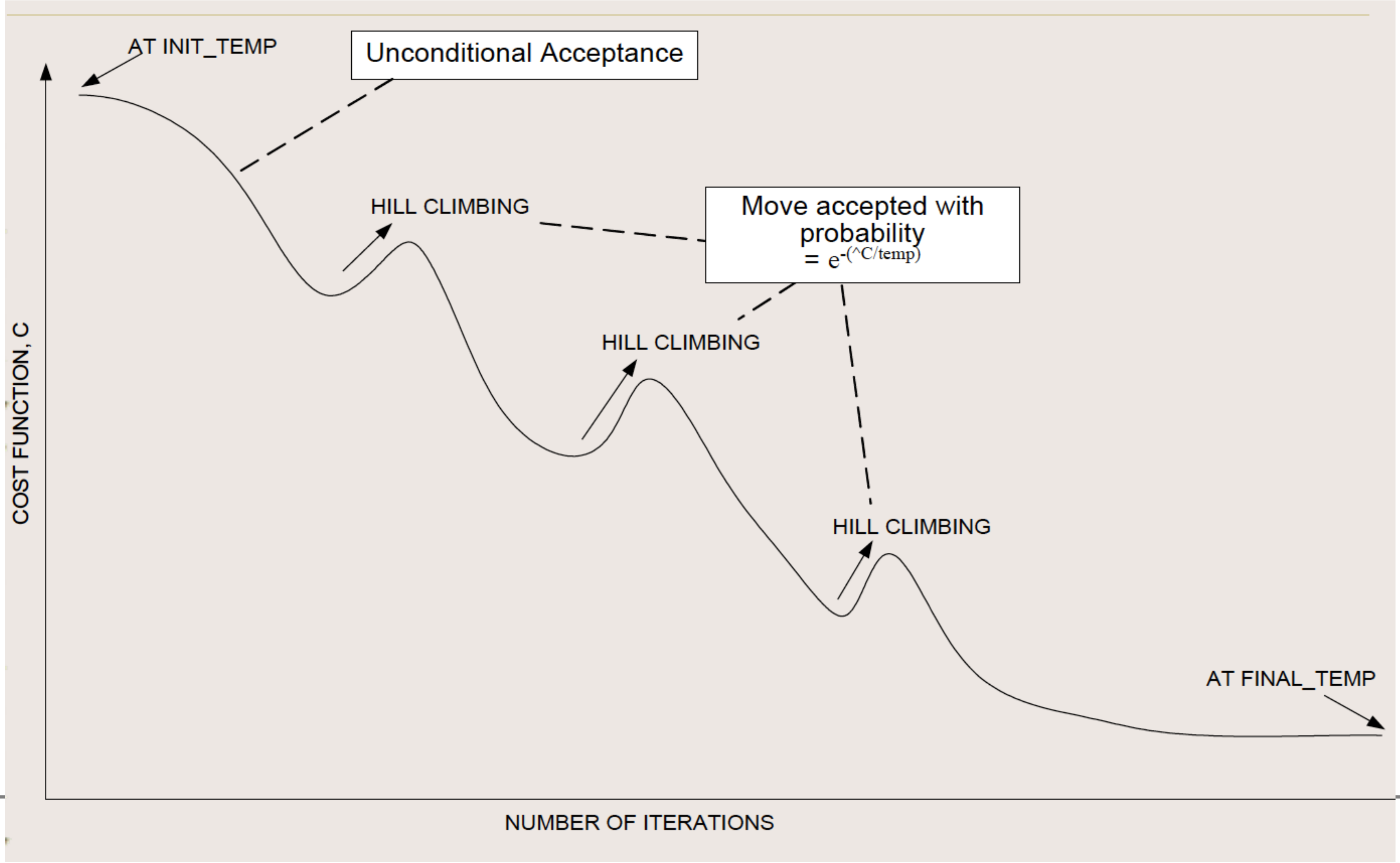
$$t^{(k+1)} = \gamma t^{(k)}$$

- Fast annealing

$$t^{(k)} = \frac{t^{(1)}}{k}$$



Simulated annealing



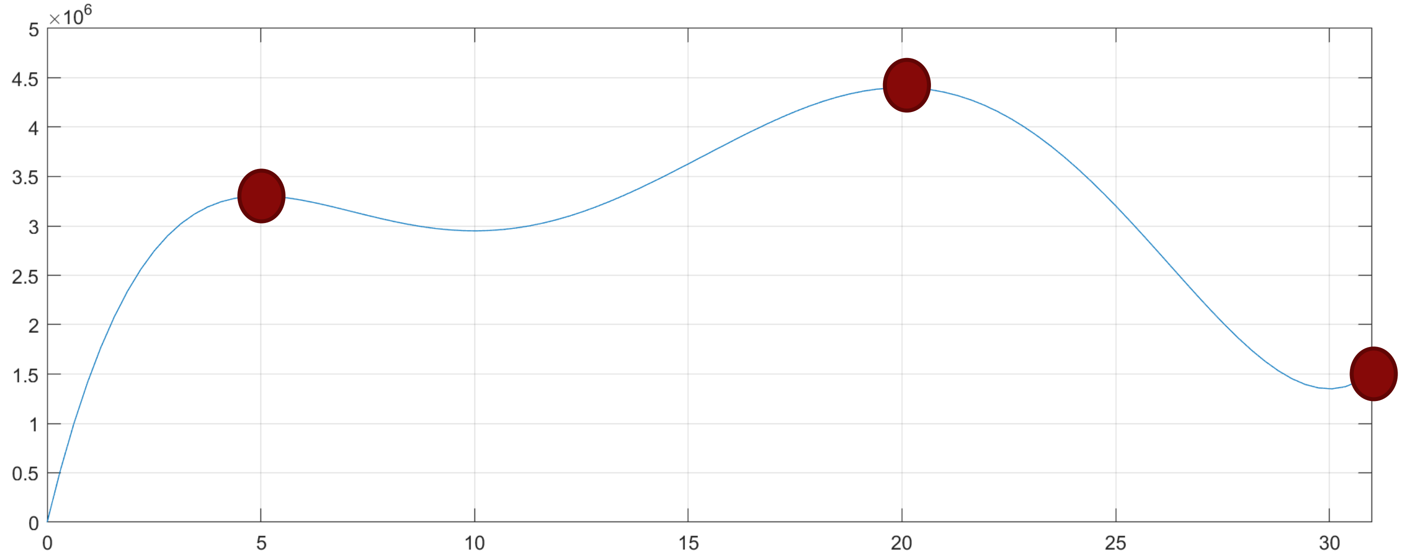
Simulated Annealing for non linear optimization

- Suppose we want to solve this problem

$$\max f(x) = 12x^5 - 975x^4 + 28000x^3 - 345000x^3 + 1800000x$$

Subject to $\quad 0 \leq x \leq 31$

- Graphically we have:



- We can see that there are tre local maxima in $x = 5,20,31$ but only $x = 20$ is the global maximum

Simulated Annealing for non linear optimization

- **Starting point**: we can choose the initial point randomly, however it is always better to start from a good initial solution. In this case, since no information is available we can choose to start from $x = 15.5$

- **Neighborhood Structure** : all feasible solutions can be considered as candidate solutions. In this case we prefer feasible solutions that are relatively close to the current solution.

- **Selection of a candidate solution**: the new candidate solution can be randomly sampled from a normal distribution with mean $\mu = 0$ and variance $\sigma = \frac{31-0}{6}$ (the denominator equal to 6 increases the probability of choosing a feasible solution), i.e.

$$x_{k+1} = x_k + N(0, \sigma)$$

If x_{k+1} is unfeasible we sample again until we find a feasible one.

- **Temperature schema**: we decrease the temperature every n=5 iterations

$$T_1 = 0.2\, f(x_0)$$

$$T_2 = 0.5T_1$$

$$T_3 = 0.5T_2$$

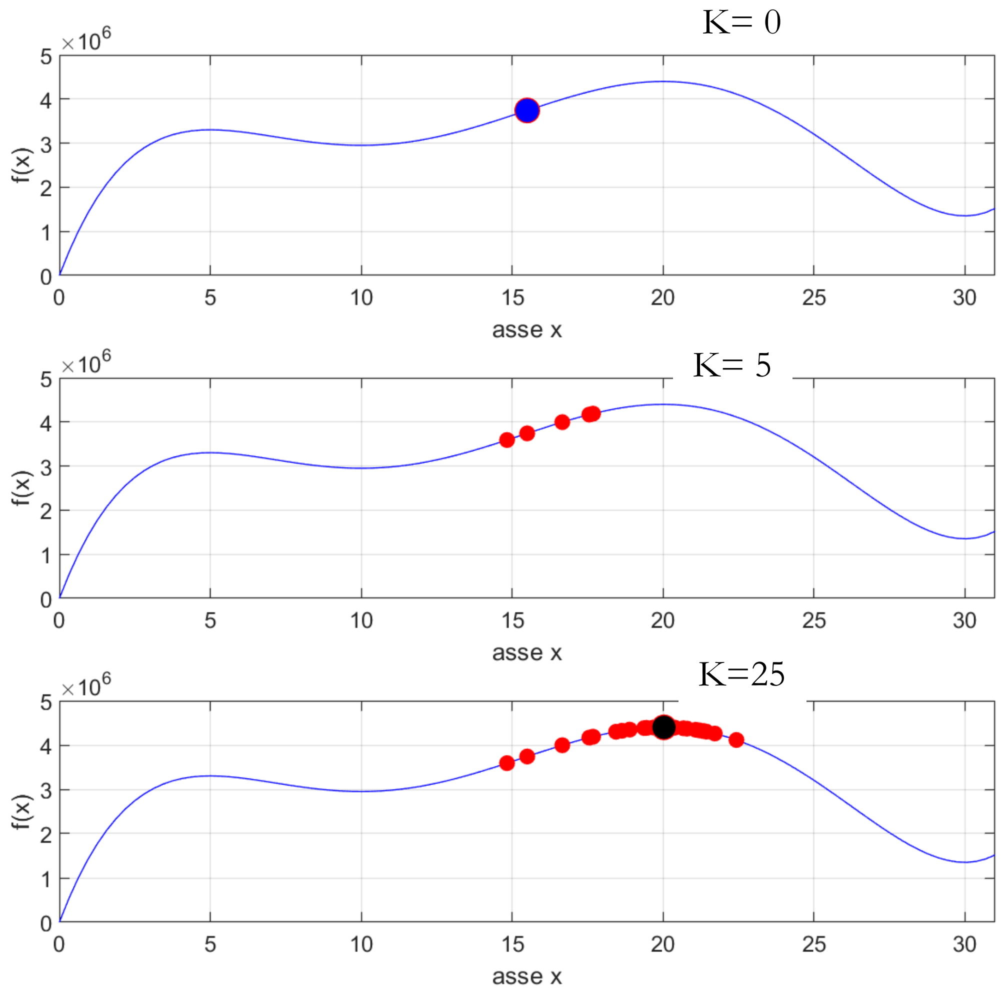
$$T_4 = 0.5T_3$$

$$T_5 = 0.5T_4$$

Simulated Annealing for non linear optimization

Iterazione	T	Soluzione	f(x)
0		$x = 15.5$	3 741 121.000
1	748224	$x = 17.557$	4 167 533.956
2	748224	$x = 14.832$	3 590 466.203
3	748224	$x = 17.681$	4 188 641.364
4	748224	$x = 16.662$	3 995 966.078
5	748224	$x = 18.444$	4 299 788.258
6	374112	$x = 19.445$	4 386 985.033
7	374112	$x = 21.437$	4 302 36.329
8	374112	$x = 18.642$	4 322 687.873
…	…	…	…
20	46764	$x = 20.680$	4 378 591.085
21	46764	$x = 20.031$	4 399 955 913
22	46764	$x = 20.184$	4 398 462.299
…	…	…	…
25	46764	$x = 19.377$	4 383 048.039

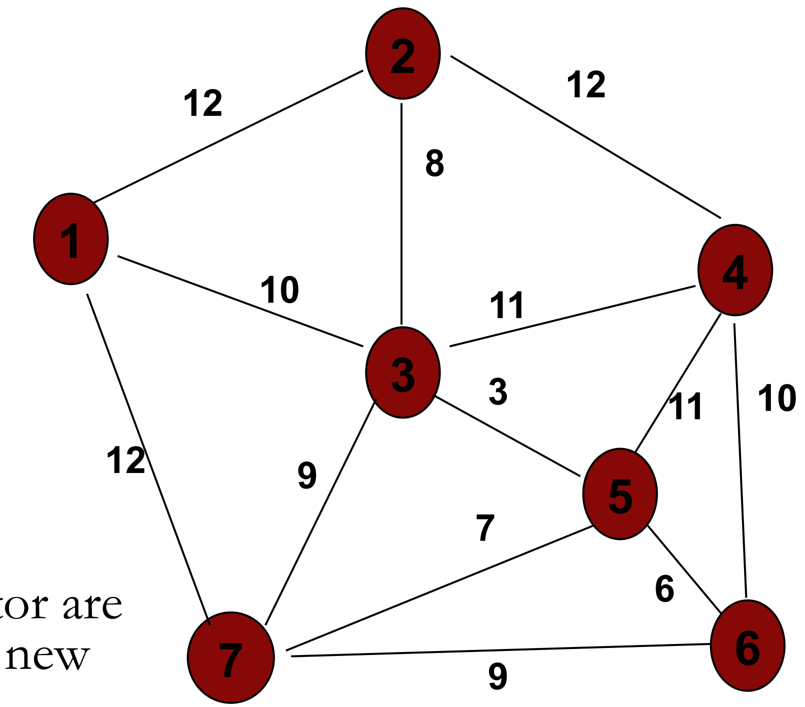
Maximum!



- We can notice that in the first iterations f(x) changes considerablt since the Temperature is high
- Subsequently solutions tend to concentrate in the neighborhood of the optimal solution

Simulated Annealing for TSP

- How can a solution be represented?
  - A vector of integers, i.e., (1,4,2,3,6,5,1)
- How is the initial solution chosen?
  - Every feasible solution is eligible as initial solution
- How is the neighborhood chosen ?
  - A neighbour solution is obtained by swapping 2 integers
  - (1,4,2,3,6,5,1) → (1,4,3,2,6,5,1)
- Mechanism of random selection
  - The initial and final points of a subsequence of the solution vector are randomly selected and the subsequent is swapped (as long as the new candidate solution is feasible)
- Cooling schedule (temperature schedule)
  - $T_1 = 0.2 \cdot f(x_0)$
  - $T_n = 0.5 \cdot T_{n-1}$. every 5 iterations



Settaggio iniziale dei parametri

- Initial solution
  - x_0=(1,2,3,4,5,6,7,1)
  - $f(x_0)$ =69
  - T_1=13.8
- r=0.2779 -> start from 3

0<= r <0.2	da 2
0.2<= r <0.4	da 3
0.4<= r <0.6	da 4
0.6<= r <0.8	da 5
0.8<= r <=1	da 6

- r=0,161 -> end in 4

0<= r <1/3	Fino a 4
1/3<= r <2/3	Fino a 5
2/3<= r <1	Fino a 6

x_0= (1,2,3,4,5,6,7,1)
x_1=(1,2,4,3,5,6,7,1)
Feasible? yes

$f(x_1)=65< f(x_0)= 69$

Settaggio iniziale dei parametri

- Initial solution
  - x_0=(1,2,3,4,5,6,7,1)
  - $f(x_0)$ =69
  - T_1=13.8
- r=0.2779 -> start from 3

0<= r <0.2	da 2
0.2<= r <0.4	da 3
0.4<= r <0.6	da 4
0.6<= r <0.8	da 5
0.8<= r <=1	da 6

- r=0,161 -> end in 4

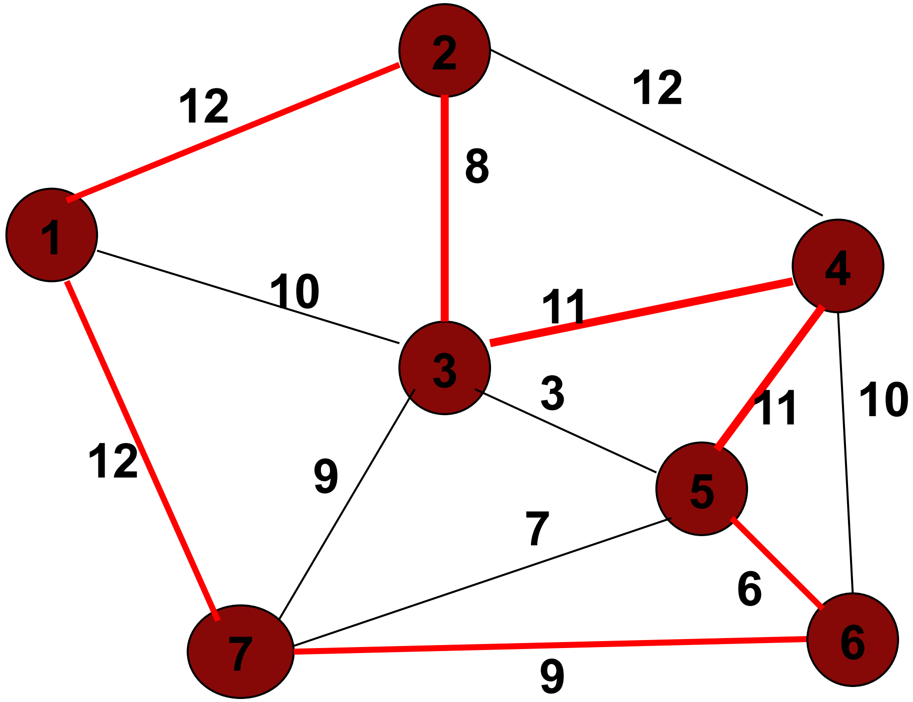
0<= r <0.25	Fino a 4
0.25<= r <0.5	Fino a 5
0.5<= r <0.75	Fino a 6
0.75<= r <=1	Fino a 7

x_0= (1,2,3,4,5,6,7,1)
x_1=(1,2,4,3,5,6,7,1)

Feasible? yes

$f(x_1)=65 < f(x_0)= 69$

Simulated Annealing per il TSP

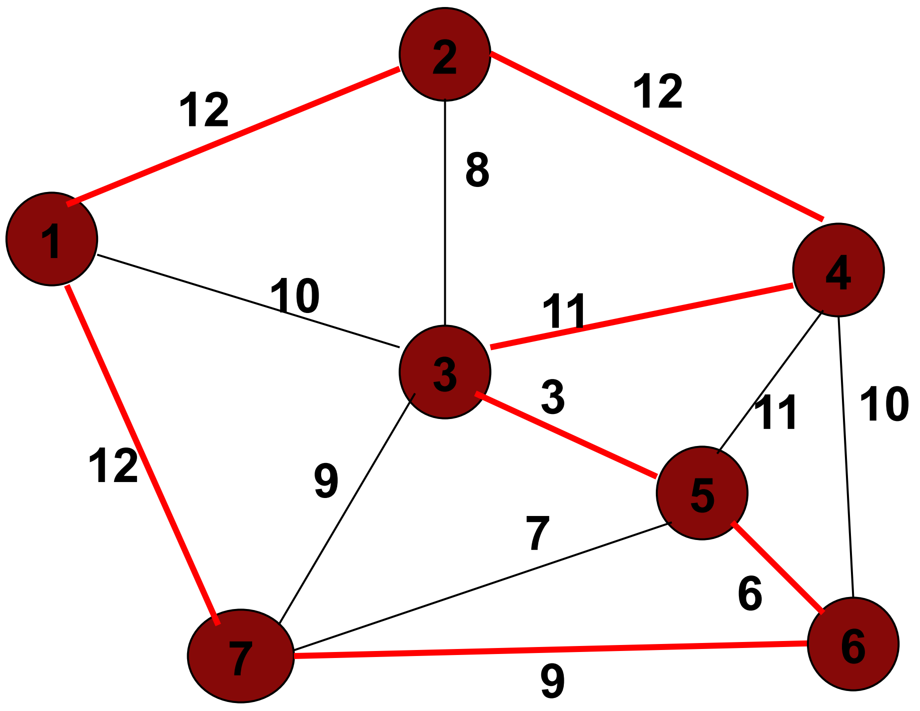


Initial solution: 1-2-3-4-5-6-7-1

Distance: 69

swap 3-4

Simulated Annealing per il TSP



Iteration 0:

Initial solutionle x_0 : 1-2-3-4-5-6-7-1

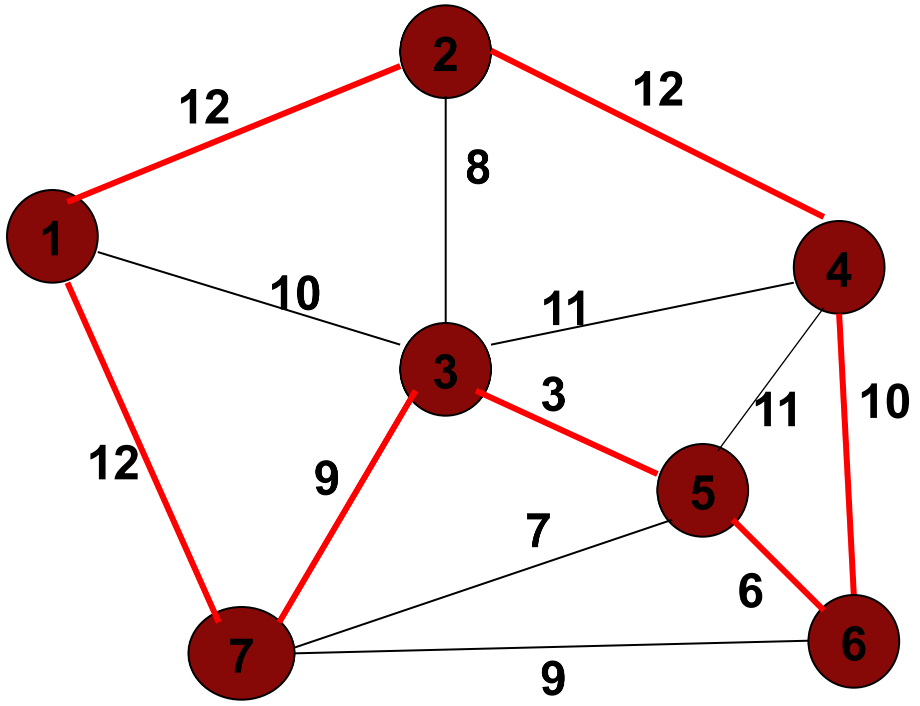
Distance: 69

Iteration 1: swap 3-4

New solution x_1 : 1-2-4-3-5-6-7-1

Distance: 65

Simulated Annealing per il TSP



Iteration 0:

Initial solution x_0: 1-2-3-4-5-6-7-1

Distance: 69

Iteration 1: swap 3-4

New solution x_1 : 1-2-4-3-5-6-7-1

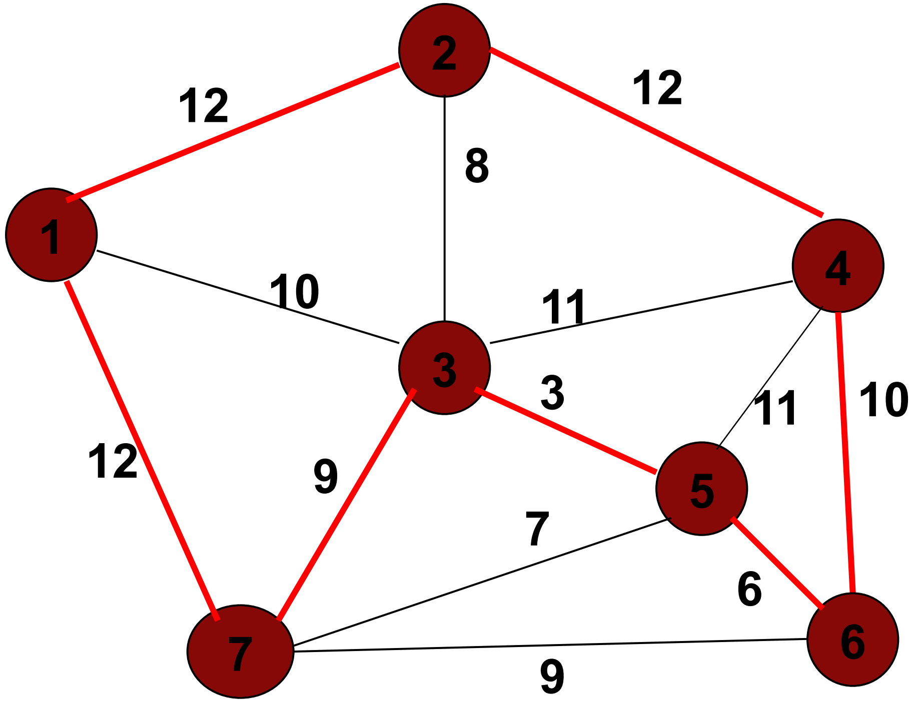
Distance: 65

Iteration 2: swap 3-5-6

New solution x_2 : 1-2-4-6-5-3-7-1

Distance: 64

Simulated Annealing per il TSP



Iteration 2: swap 3-5-6

New solution x_2 : 1-2-4-6-5-3-7-1

Distance: 64

Iteration 3: Swap 3-7

New solution x_3 : 1-2-4-6-5-3-7-1

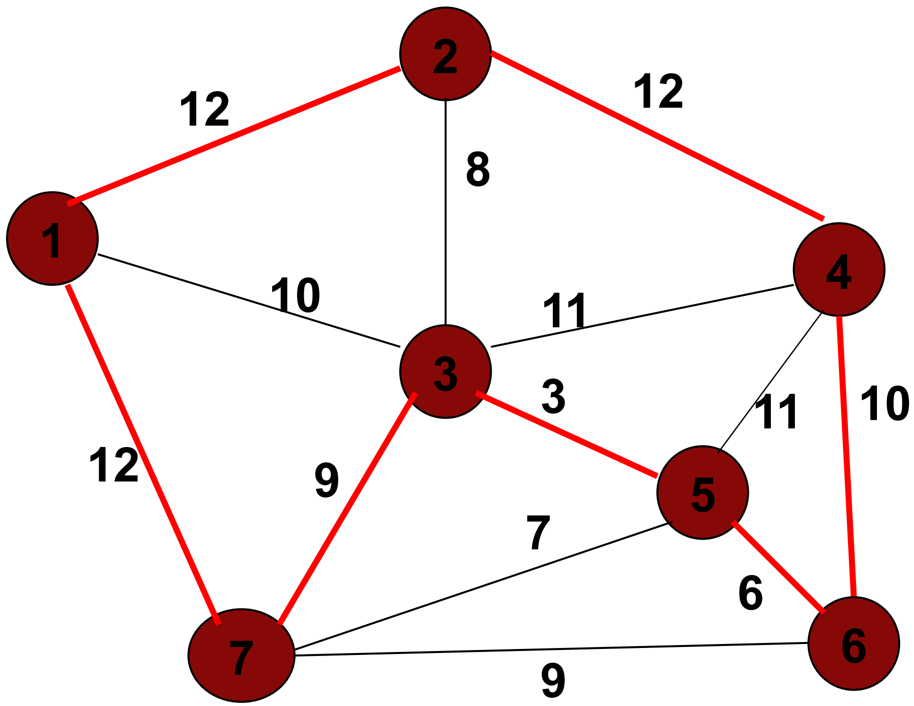
Shell I accept x_3 ?

$f(x_2)=64<f(x_3)=66$ $\quad$ $T = 13.83$

$\text{Prob(acceptance)}=e^{(f(x3)-f(x2))}/T=e^{-2/13.8}=0.865$

x_3 will be accepeted with probability 0.865 !!!

Simulated Annealing per il TSP



Iteration 2: swap 3-5-6

New solution x_2 : 1-2-4-6-5-3-7-1

Distance: 64

Iteration 3: Swap 3-7

New solution x_3 : 1-2-4-6-5-3-7-1

Shell I accept x_3 ?

$f(x_2)=64<f(x_3)=66$ T = 13.83

$Prob(acceptance)=e^{(f(x3)-f(x2))}/T=e^{-2/13.8}=0.865$

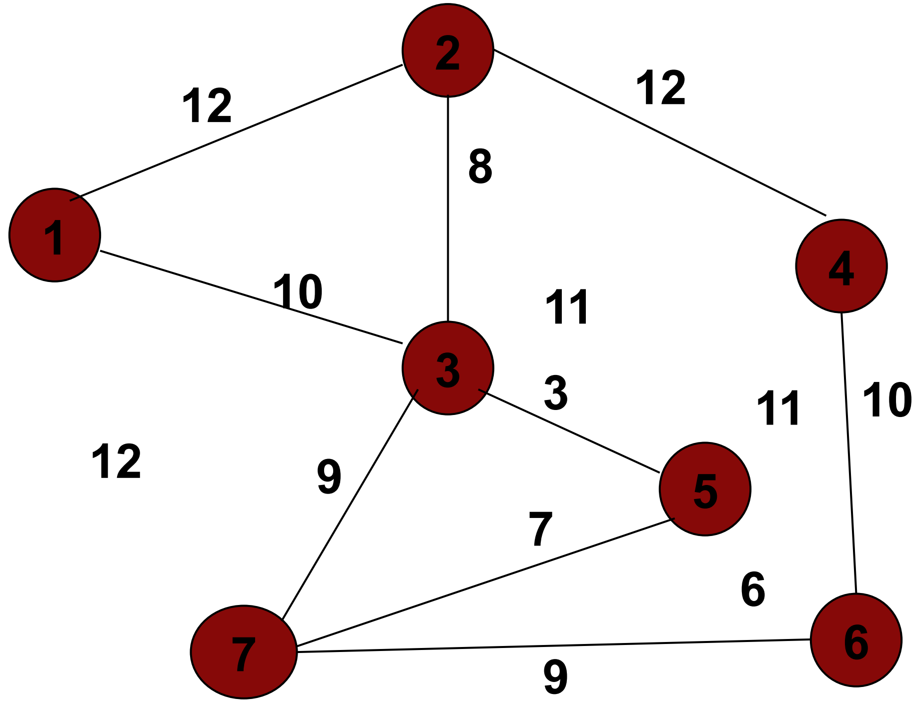
x_3 will be accepeted with probability 0.865 !!!

Simulated Annealing: results

Iterazione	T	Soluzione	f(x)
0		1-2-3-4-5-6-7-1	69
1	13.8	1-3-2-4-5-6-7-1	68
2	13.8	1-2-3-4-5-6-7-1	69
3	13.8	1-3-2-4-5-6-7-1	68
4	13.8	1-3-2-4-6-5-7-1	65
5	13.8	1-2-3-4-6-5-7-1	66
6	6.9	1-2-3-4-5-6-7-1	69
7	6.9	1-3-2-4-5-6-7-1	68
8	6.9	1-2-3-4-5-6-7-1	69
…	…	…	…
14	3.43	1-3-5-7-6-4-2-1	63
15	3.43	1-3-7-5-6-4-2-1	66
16	1.725	1-3-5-7-6-4-2-1	63
…	…	…	…
25	0.8625	1-3-7-5-6-4-2-1	66

Min

Min



This is a new simulation of the SA starting from the same initial solution 1-2-3-4-5-6-7-1

Note that we generate the optimal soltion at iterations 14 and 16

Simulated annealing final remarks

- A number of decisions must be taken when applying SA
  - Solution Representation and generation
  - Initial temperature T ?
  - Temperature schedule ?
  - How many iterations with the same T value ?
  - Stop criteria ?
- **Practical issues**
  - The initial temperature must be such that about 50% of worsening solutions are initially accepted
  - The cooling schedule should be slow for example 10%
  - The final temperature should be such that no worsening solutions are accepted, i.e. $T \approx 0$
- **Note:** Can be teoretically proved that SA asymptotically converges to the global optimum. In practice however its convergence speed is highliy influenced by the **cooling schedule**

Simulated annealing

Simulated annealing does not guarantee global optimum

However, it tries to avoid a large number of local minima

Therefore, it often yields a better solution than local optimization

Simulated annealing is not deterministic

Whether accept or reject a new solution is random

You can get different answers from multiple runs

Simulated annealing is more expensive than local optimization

It is the price you must pay to achieve a better optimal solution