IE 607 Heuristic Optimization

Simulated Annealing

Origins and Inspiration from Natural Systems

- **Metropolis et al.** (Los Alamos National Lab), "Equation of State Calculations by Fast Computing Machines," *Journal of Chemical Physics*, 1953.

- In annealing, a material is heated to high energy where there are frequent state changes (disordered). It is then gradually (*slowly*) cooled to a low energy state where state changes are rare (approximately *thermodynamic equilibrium*, "frozen state").

  For example, large crystals can be grown by very slow cooling, but if fast cooling or quenching is employed the crystal will contain a number of imperfections (glass).

- Metropolis's paper describes a Monte Carlo simulation of interacting molecules.
- Each state change is described by:
  - If E decreases, P = 1
  - If E increases, P = $e^{-(\frac{\Delta E}{kT})}$

    where T = temperature (Kelvin), k = Boltzmann constant, k > 0, $\Delta E =$ energy change

To summarize behavior:

For large T, P → $\frac{1}{e^0} = 1$

For small T, P → $\frac{1}{e^\infty} = 0$

SA for Optimization

- **Kirkpatrick, Gelatt and Vecchi**, "Optimization by Simulated Annealing," *Science*, 1983.
- Used the Metropolis algorithm to optimization problems – spin glasses analogy, computer placement and wiring, and traveling salesman, both difficult and classic combinatorial problems.

Spin Glasses Analogy

- Spin glasses display a reaction called *frustration*, in which opposing forces cause magnetic fields to line up in different directions, resulting in an energy function with *several low energy states*

  → in an optimization problem, there are likely to be *several local minima* which have cost function values close to the optimum

Thermodynamic Simulation vs Combinatorial Optimization

Simulation	Optimization
System states	Feasible solutions
Energy	Cost (Objective)
Change of state	Neighboring solution (Move)
Temperature	Control parameter
Frozen state	Final solution

Canonical Procedure of SA

- Notation:

T_0	starting temperature
T_F	final temperature ($T_0 > T_F$, $T_0, T_F \geq 0$)
T_t	temperature at state *t*
m	number of states
n(t)	number of moves at state *t*
	(total number of moves = n * m)
s	move operator

Canonical Procedure of SA

$\mathbf{x}_0$	initial solution
$\mathbf{x}_i$	solution i
$\mathbf{x}_{\mathrm{F}}$	final solution
$\mathrm{f}(\mathbf{x}_i)$	objective function value of $\mathbf{x}_i$
a	cooling parameter

- **Procedure** (Minimization)

Select $\mathbf{x}_0$, T_0, m, n, $\boldsymbol{\alpha}$

Set $\mathbf{x}_1 = \mathbf{x}_0$, $T_1 = T_0$, $\mathbf{x}_F = \mathbf{x}_0$

for (t = 1,…,m) {

for (i = 1,…n) {

$\mathbf{x}_{\text{TEMP}} = \boldsymbol{\sigma}(\mathbf{x}_i)$

if $f(\mathbf{x}_{\text{TEMP}}) = f(\mathbf{x}_i)$, $\mathbf{x}_{i+1} = \mathbf{x}_{\text{TEMP}}$

else

if $U(0,1) \le e^{-\frac{f(\mathrm{x}_{TEMP}) - f(\mathrm{x}_i)}{T_t}}$, $\mathbf{x}_{i+1} = \mathbf{x}_{\text{TEMP}}$

else $\mathbf{x}_{i+1} = \mathbf{x}_i$

if $f(\mathbf{x}_{i+1}) = f(\mathbf{x}_F)$, $\mathbf{x}_F = \mathbf{x}_{i+1}$ }

$T_{t+1} = \boldsymbol{\alpha}\ T_t$ }

return $\mathbf{x}_F$

Combinatorial Example: TSP

Continuous Example:
2 Dimensional Multimodal Function

SA: Theory

Assumption: can reach any solution from any other solution

1. **Single T –** *Homogeneous Markov Chain*

   - constant T

   - global optimization does not depend on initial solution

   - basically works on law of large numbers

2. Multiple T

- *sequence of homogeneous Markov Chains* (one at each T) (by Aarts and van Laarhoven)

→ number of moves is at least quadratic with search space at T

→ running time for SA with such guarantees of optimality will be *Exponential*!!

- a *single non-homogeneous Markov Chain* (by Hajek)

→ if $T(k) = \frac{c}{\log(1+k)}$, the SA converges if c = depth of largest local minimum where k = number of moves

2. Multiple T (cont.)

- $T_{t+1} = aT_t \quad a < 1$ usually successful between 0.8 & 0.99

- $T_{t+1} = \dfrac{T_t}{1 + bT_t}$ Lundy & Mees

- $T_{t+1} = \dfrac{T_t}{1 + \dfrac{T_t \ln(1 + \Delta E)}{3s_T}}$ Aarts & Korst

s_T : standard deviation at temp T

Variations of SA

- **Initial starting point**
  - construct "good" solution (pre-processing)
    → search must be commenced at a lower temperature
    → save a substantial amount of time (when compared to totally random)
  - Multi-start

Variations of SA (cont.)

- **Move**

  - change neighborhood during search (usually contract)

    → e.g. in TSP, 2-opt neighborhoods restrict two points “close” to one another to be swapped

  - non-random move

Variations of SA (cont.)

- **Annealing Schedule**

  - constant T: such a temperature must be high enough to climb out of local optimum, but cool enough to ensure that these local optima are visited → problem-specific & instance-specific

  - 1 move per T

Variations of SA (cont.)

- **Annealing Schedule (cont.)**

  - change schedule during search (include possible reheating)

  → # of moves accepted

  → # of moves attempted

  → entropy, specific heat, variance of solutions at T

Reheating

- **Kirkpatrick et al.**

  Use it when the process got stock in a local optimum. Can be triggered by the rate of change in the cost function or the ratio of acceptances.

Reheating (cont.)

- **Dowsland**

Cooling $$T_{t+1} = \frac{T_t}{1 + bT_t}$$

(every time a move is accepted)

Reheating $$T_{t+1} = \frac{T_t}{1 - gT_t}$$

(every time a move is rejected)

$b = kg$ → need to go through k heating iterations to balance one cooling

Entropy

A measure of the disorder of the system

$$S(T) = -\sum_{i \in S} P_i(T) \ln P_i(T)$$

i: state, P_i: probability, *T*: temperature

S(T) monotonically decreases as temperature decreases.

Specific Heat

The rate of change of mean energy

$$\frac{dS(T)}{dT} = \frac{C(T)}{T} \Rightarrow S(T_1) - S(T_0) = \int_{T_0}^{T_1} \frac{C(T')dT}{T}$$

A large value of C (peak) indicates “freezing” has begun so very slow cooling is needed.

Variations of SA (cont.)

- **Acceptance Probability**

$$P = e^{-\frac{\Delta E}{T}}$$

$$P = 1 - \frac{\Delta E}{T}$$

By Johnson et al.

$$P = \frac{\Delta E}{T}$$

By Brandimarte et al., need to decide whether to accept moves for the change of energy = 0

Variations of SA (cont.)

- **Stopping Criteria**
  - total moves attempted
  - no improvement over n attempts
  - no accepted moves over m attempts
  - minimum temperature

Variations of SA (cont.)

- **Finish with Local Search**

  \- post-processing

  → apply a descent algorithm to the final annealing solution to ensure that at least a local minimum has been found

Variations of SA (cont.)

- **Parallel Implementations**

  - multiple processors proceed with different random number streams at given T→the best result from all processors is chosen for the new temperature

  - if one processor finds a neighbor to accept → convey to all other processors and search start from there

Applications of SA

- Graph partitioning & graph coloring
- Traveling salesman problem
- Quadratic assignment problem
- VLSI and computer design
- Scheduling
- Image processing
- Layout design
- A lot more…

Some Websites

- http://www.svengato.com/salesman.html
- http://www.ingber.com/#ASA
- http://www.taygeta.com/annealing/simanneal.html