Stats 318: Lecture # 13

Agenda: MCMC

- Global and combinatorial optimization
- Traveling salesman problem (TSP)
- Simulated annealing

General problem

- Finite set $\mathcal{X}$
- Objective $f : \mathcal{X} \longrightarrow \mathbb{R}$

Problem is to find $x^\star$ which minimizes or maximizes f

$$x^\star = \text{arg min/arg max } \{f(x) : x \in \mathcal{X}\}$$

Interested in situations where $\mathcal{X}$ is very large

Example: *hard-sphere model*

Find configuration with maximum number of occupied sites (max # 1's)

Traveling salesman problem (TSP)

A salesman must visit a set of $\{1, \ldots, d\}$ cities & return back home (one of these cities)
In which order should he visit the cities as to minimize the total distance travelled?
Suppose salesman leaves from city 1

- State space $\mathcal{X}$ is set of all feasible routes

$$x = (x(1), \ldots, x(d)) \quad x(1) = 1$$
$$x(2), \ldots, x(d) \text{ permutation of } \{2, \ldots, d\}$$

- $|\mathcal{X}| = (d-1)!$
  e.g. 100 cities $\Longrightarrow 10^{157}$ routes!
- Need to minimize $f(x) = \sum_i \text{dist}(x(i), x(i+1))$ s.t. $x(1) = 1$, $x(d+1) = 1$
- How are we going to do this?

Idea underlying simulated annealing (SA)

- View f as an energy functional
- Gibbs distribution : $\pi_T(x) = \frac{1}{Z_T} e^{-f(x)/T}$
- At low temperatures, π places most of its mass on states with low values of f

Lemma

(a)

$$\lim_{T \to \infty} \pi_T(x) = \frac{1}{|\mathcal{X}|}$$

(b)

$$\lim_{T \to 0} \pi_T(x) = \begin{cases} \frac{1}{|\mathsf{SOL}|} & \text{if } x \in \mathsf{SOL} \\ 0 & \text{otherwise} \end{cases}$$

SOL $\subset \mathcal{X}$ is set of global minima of f

Proof:

(a) For $x \neq x'$

$$\frac{\pi_T(x)}{\pi_T(x')} = e^{-\frac{1}{T}[f(x)-f(x')]} \underset{T\to\infty}{\longrightarrow} 1$$

(b) $x' \in$ SOL & $x \notin$ SOL

$$\frac{\pi_T(x)}{\pi_T(x')} \underset{T\to\infty}{\longrightarrow} 0$$

Remark

Competition between entropy & energy
Chain with equil. dist π
KL divergence $D(\pi^{(t)} \mid\mid \pi)$ decreases with time

$$D(\mu \mid\mid \pi) = \sum_x \mu(x) \log \frac{\mu(x)}{\pi(x)} = -S(\mu) + \frac{1}{T} \sum_x \mu(x) f(x) + \log Z_T$$

$$= \underbrace{-S(\mu)}_{\text{entropy}} + \frac{1}{T} \underbrace{\mathbb{E}_\mu f}_{\text{average energy}} + \log Z_T$$

Metropolis algorithm

$-S\left(\pi^{(t)}\right) + \frac{1}{T}\mathbb{E}_{\pi^{(t)}} f$ decreases with time (as t increases)

- Hot temp: entropy dominates & X_t will approx. behave like a uniform variable
- Low temp: energy term dominates & Metropolis alg essentially seeks to minimize f

Simulated annealing (SA)

Sequence of temperatures $T(t)$ with

$$T(t) \underset{T\to\infty}{\longrightarrow} 0$$

SA is similar to Metropolis algorithm with the proviso that the temperature is changing

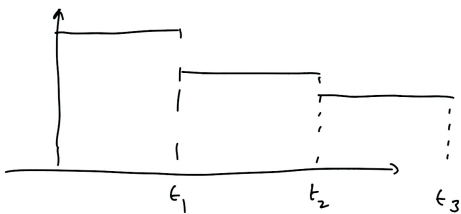
Algorithm

- Initialize X_0
- For $t = 1, 2, \ldots$
  - Propose a new state y
  - Accept new state wp $\exp\left(-\frac{1}{T(t)}[f(y) - f(X_t)]\right) \wedge 1$

Cooling schedules

- Sequence $T(t)$ is called the annealing/cooling schedule and is of critical importance
- If $T(t)$ decays too rapidly, $\{X_t\}$ can be trapped in a local minimum
- If $T(t)$ decays too sufficiently slowly, algorithm converges to a global minimizer wp 1 (more later)

- Possible cooling schedule:



$$T_1 > T_2 > T_3 > \dots$$

$$t_1 < t_2 < t_3 < \dots$$

Example: simulated annealing for TSP

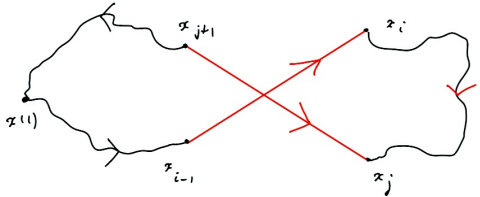
- Need a metropolis algorithm for π_T
- Need to define permutations we view as neighbors

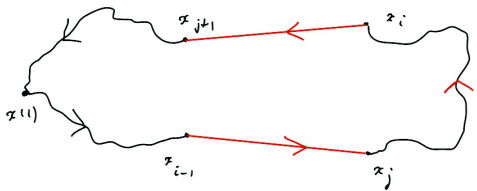
Two possibilities

(a) $x \sim x'$ if for a pair $2 \le i < j \le d$

$$x'(i) = x(j)$$
$$x'(j) = x(i) \qquad x = \{x(1), x(2), \ldots, x(i), \ldots, x(j), \ldots, x(d)\}$$

(b) $x \sim x'$ if for a pair $2 \le i < j \le d$, x' obtained by reversing segment $\{x(i), \ldots, x(j)\}$





In both cases:

$$Q(x, x') = \begin{cases} \frac{1}{\binom{d}{2}} & x' \sim x \\ 0 & \text{otherwise} \end{cases}$$

And $$P(x, x') = \begin{cases} Q(x, x')R(x, x') & \text{if } x' \sim x \\ 1 - \sum_{y : y \sim x} P(x, y) & \text{if } x' = x \\ 0 & \text{otherwise} \end{cases}$$

$R(x, x') = \exp\left(-\frac{1}{T}(f(x') - f(x))\right) \wedge 1$

Convergence of simulated annealing (SA)

Theorem

There is a constant C_f s.t. if

$$\lim_{t\to 0} T(t) = 0 \quad \& \quad \sum_{t \geq 1} exp\left(-\frac{C_f}{T(t)}\right) = \infty$$

Then SA will converge to a global minimizer of f

$$\lim_{t\to\infty} \mathbb{P}(X_t \in \mathsf{SOL}) = 1 \qquad \mathsf{SOL} : \{x : f(x) = f^\star\}.$$

Examples:

(i) $T(t) = \frac{C}{\log t}$

(ii) $T(t) = \frac{1}{k} \quad e^{(k-1)C} \leq t < e^{kC}$

If $C > C_f$
hypothesis of thm holds & SA converges

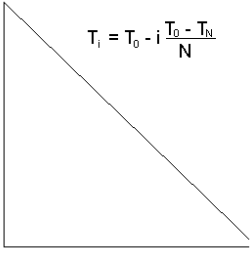
Problems

- Calculating C_f in general not possible
- Cooling schedule is too slow, will take forever

Example of cooling schedules (Brian T. Luke)

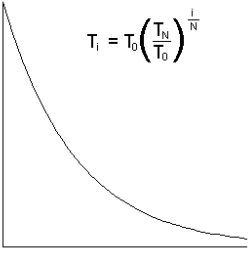
Cooling Schedule 0

$$T_i = T_0 - i\,\frac{T_0 - T_N}{N}$$



Cooling Schedule 1

$$T_i = T_0\left(\frac{T_N}{T_0}\right)^{\frac{i}{N}}$$

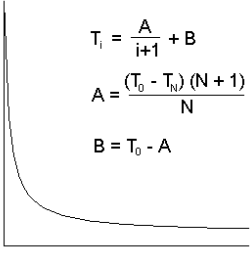


Cooling Schedule 2

$$T_i = \frac{A}{i+1} + B$$

$$A = \frac{(T_0 - T_N)\,(N+1)}{N}$$

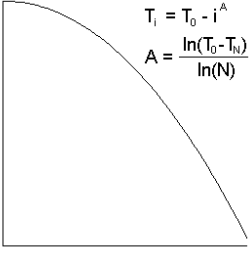
$$B = T_0 - A$$



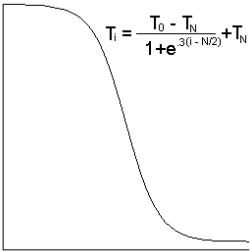
Cooling Schedule 3

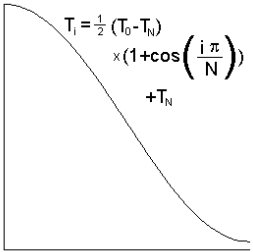
$$T_i = T_0 - i^A$$

$$A = \frac{\ln(T_0 - T_N)}{\ln(N)}$$

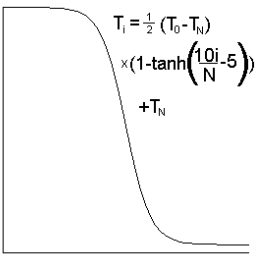


Example of cooling schedules (Brian T. Luke)

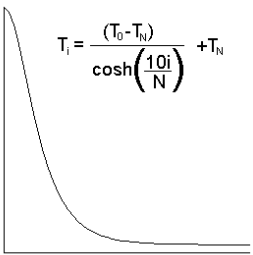




Cooling Schedule 6



Cooling Schedule 7

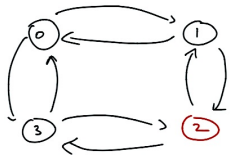


Numerical example

Traveling salesman problem (click here)

Hazard of fast annealing schedule

- $\mathcal{X} = \{0, 1, 2, 3\}$
- $(f(0), f(1), \underline{\underline{f(2)}}, f(3)) = (1, 2, \underline{\underline{0}}, 2)$



Q is a symmetric RW

$$P = \begin{bmatrix} 1 - e^{-t/T} & \frac{1}{2}e^{-t/T} & 0 & \frac{1}{2}e^{-t/T} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2}e^{-2/T} & 1 - e^{-2/T} & \frac{1}{2}e^{-2/T} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{bmatrix}$$

$X_0 = 0$

What is probability of getting stuck at 0?

$$A = \{X_t = 0 \ \forall \ t\} \qquad \mathbb{P}(A) = \prod_{t \geq 1} P(X_t = 0 \mid X_{t-1} = 0) = \prod_{t \geq 1} \left[1 - e^{-t/T(t)}\right]$$

$$\mathbb{P}(A) = \prod_{t \geq 1} \left[1 - e^{-t/T(t)}\right]$$

Equal to 0 IFF $\sum_t e^{-t/T(t)} = \infty$

If $T(t) \downarrow$ s.t. $\sum_t e^{-t/T(t)} < \infty$ [$T(t)$ decays rapidly]

e.g.

$$T(t) = \frac{1}{2 \log t} \Longrightarrow \sum_{t \geq 1} e^{-1/T(t)} = \sum_{t \geq 1} t^{-2} = \frac{\pi^2}{6} < \infty$$

Failure!

Two facts create this situation

(1) Annealing schedule is too fast

(2) $x = 0$ is a local minimum