Simulated annealing and Gibbs sampler

Leonid Zhukov

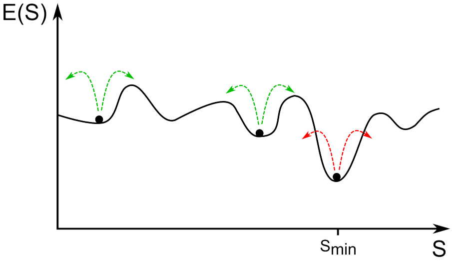
School of Applied Mathematics and Information Science
National Research University Higher School of Economics

12.12.2013

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

Simulated annealing

- Global optimization method (vs greedy strategies - local minimum)
- Works for both continues and discrite optimization problems
- Intuition from thermodynamics, physical annealing process
- Algorithm:
  1. generate trial point and evaluate function at that location.
  2. accept new location if it reduces "energy"(improves solution)
  3. accept some new location even not improving the solution
  4. probability accepting non-improving location decreases with lowering "temperture"



Simulated annealing

- Compute change of energy for a k-step $\Delta E_k = E_k - E_{k-1}$:
  - If $\Delta E_k \leq 0$, accept the step
  - if $\Delta E_k > 0$, accept the step with probability $P(\Delta E_k) = e^{-\Delta E_k / T_k}$
- Cooling schedule:
  - $T_k = \alpha T_0 / k$
  - $T_k = \alpha T_0 / \log k$
  - $T_k = T_0 / \alpha^k$

Traveling salesman problem

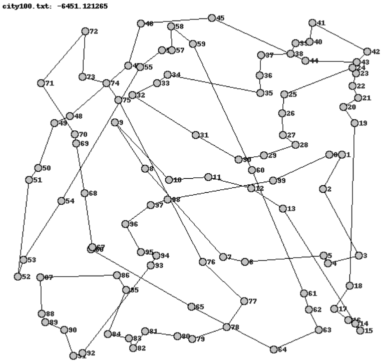
Travelling salesman problem: given a list of cities and the distances between each pair of cities, what is the shortest possible route that visits each city exactly once and returns to the origin city? (NP-hard problem in combinatorial optimization)
Simulated annealing algorithm:

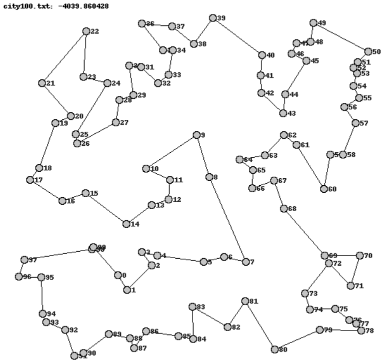
- Define energy cost function:

$$E = \sum_{i}^{N} \sqrt{(x_{i+1} - x_i^2)^+ (y_{i+1} - y_i)^2}$$

- Select initial route (sequence of city labeles)
- Iteratively imporve the route by trying local changes:
  swap a pair of cities + reverse the section between them
- always accept lower enegergy swap, sometimes higher enegery
- reduce temperature with cooling schedule

Traveling salesman problem





MCMC connection

- MCMC constructs Markov chain that generates random samples distributed according to $P(x)$
- Metropolis algorithm uses symmetric "candidate"distribution $Q(x, x^*) = Q(x^*, x)$
- Probabiity of "forward move"of the chain:

$$\alpha(x, x^*) = \min\left[1, \frac{P(x^*)}{P(x)}\right]$$

- Consider Boltzman (Gibbs) distribution

$$P(x) = \frac{1}{Z} e^{-\frac{E(x)}{T}}$$

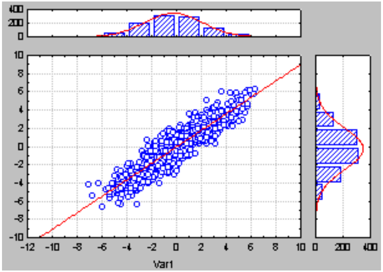
- Probability of move

$$\alpha(x, x^*) = \min\left[1, e^{-\frac{E(x^*)-E(x)}{T}}\right] = \min\left[1, e^{-\frac{\Delta E}{T}}\right]$$

$\Delta E \leq 0$ - always move, $\Delta E > 0$ move with probability $\exp\left(-\frac{\Delta E}{T}\right)$

The Gibbs sampler

- Goal: get random samples from joint multivariate density $p(x_1, ...x_n)$ if it is not known explicitely or hard to sample from
- Given conditional univatiate distributions $p(x_1|x_2...x_n)$, $p(x_2|x_1...x_n)$, $..p(x_n|x_1...x_{n-1})$



- Bayesian inference, posterior distributions
- Approximate joint distribution (histogram), compute averages

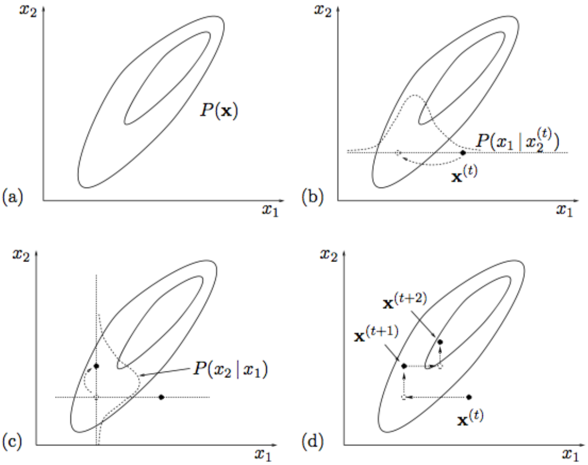
Bivariate case

- joint bivariate distribution
  $p(x, y)$
- marginal distribution
  $p(x) = \int p(x, y)dy$
  $p(y) = \int p(x, y)dx$
- conditional probability
  $p(x|y) = p(x, y)/p(y)$
  $p(y|x) = p(x, y)/p(x)$
- marginal from conditional distribution
  $p(x) = \int p(x|y)p(y)dy = E_{p(y)}[p(x|y)]$
  $p(y) = \int p(y|x)p(x)dx = E_{p(x)}[p(y|x)]$

Gibbs sampler: bivariate case

- Given: $p(x|y), p(y|x)$
- Choose y_0, $t = 0$
- do "sampler scan"
  $x_t \sim p(x|y = y_t)$
  $y_{t+1} \sim p(y|x = x_t)$
- repeat k-times, Gibbs sequence $(x_0, y_0), (x_1, y_1)...(x_k, y_k)$

Gibbs sampler



Gibbs sampler

Algorithm: Gibbs sampler

Input: all marginals $p(x_i|x_1..x_{i_1}, x_{i+1}, x_n)$

initialize $x^{(0)} = (x_1^{(0)}, ...x_k^{(0)})$, $t = 0$

while $t < T$ **do**

  for $i = 1$ *to* n **do**

    $x_i^{(t+1)} \sim p(x_i|x_1^{(t+1)}, .., x_{i-1}^{(t+1)}, x_{i+1}^{(t)}, ..., x_n^{(t)})$

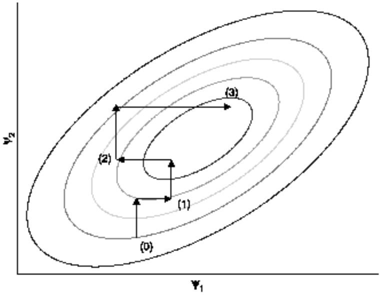
  end

  $t = t + 1$

end

return $\{\mathbf{x}^0, \mathbf{x}^1, ...\mathbf{x}^N\}$

Gibbs sampler



MCMC connection

- Metropolis-Hastings

$$\alpha(x, x^*) = \min\left[1, \frac{p(x^*)Q(x|x^*)}{p(x)Q(x^*|x)}\right]$$

- use conditional $p(x|x^*)$ as a candidate density $Q(x|x^*)$
  $(x^t, y^t) \to (x^{t+1}, y^t)$

$$\alpha = \frac{p(x^{t+1}, y^t)p(x^t|y^t)}{p(x^t, y^t)p(x^{t+1}|y^t)} = \frac{p(x^{t+1}, y^t)}{p(x^t, y^t)}\frac{p(x^t, y^t)}{p(y^t)}\frac{p(y^t)}{p(x^{t+1}, y^t)} = 1$$

- MCMC algorithm with acceptance probability 1

References

- Explaining the Gibbs Sampler. G. Casella, E.I. Georrge. The American Statistician, V 46, N3, 1992, pp 167-174
- Markov Chain Monte Carlo and Gibbs Sampling. Lecture notes. B. Walsh
- Introduction to Monte Carlo methods. Notes. D.J.C. Mackay.