Sommersemester 2014, Struktur und Simulation

Revision Part 2

1. Show that the Metropolis Monte Carlo method fulfils the detailed balance equation

$$p_i \, \pi(i \to j) = p_j \, \pi(j \to i)$$

where p_i is the equilibrium probability of being in state i and $\pi(i \to j)$ is the transition probability of going from state i to state j.

2. (a) Genetic algorithms are often used as a general optimisation method. The general idea is that you have a population of n solutions. For some number of generations, you select the least fit individuals and remove them. From the remaining individuals, you apply "mutations" and maybe some events like "crossovers". You then copy individuals to bring the number back up to n.

   If you are working with protein coordinates, you may have the internal angles stored in an array and you have n copies of the array. A mutation may mean you change an internal angle. If you are working with binding a rigid ligand to a protein, a mutation might consist of a small change in the ligand coordinates. Consider the example of binding a ligand to a protein:

```
generate n random poses (drug + protein). (initial set {m})
while (not_finished)
    foreach member m of {m}
        apply a small random change to m
    calculate energy of each of the n poses
    rank set of candidates
    discard 1/2 of the candidates with the highest energy
    duplicate the remaining candidates
```

   Why is this likely to lead to low energy configurations? Compare this with simulated annealing where you have a temperature parameter and gradually cool the system. What could you change in the genetic algorithm to mimic the effect of high or low temperature?

   (b) You wish to implement a similar scheme with Monte Carlo / simulated annealing. You have a population / set of size n and you think it is a good idea to discard some individuals at each generation. Describe a method to generate a new generation of individuals which

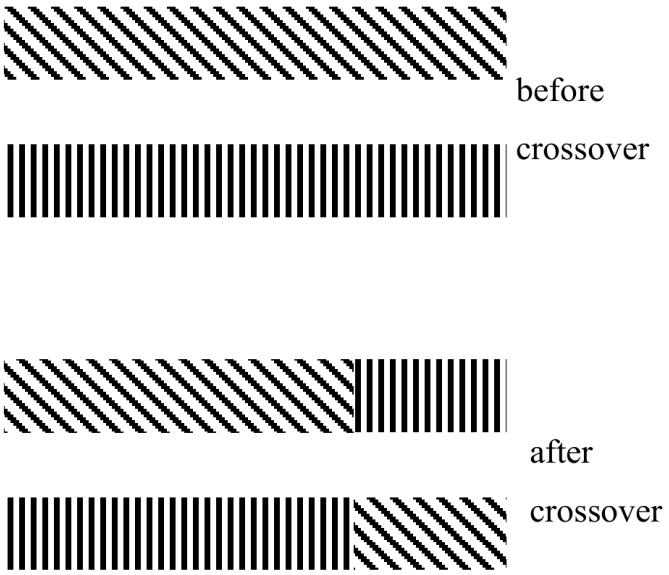
   i. discards some individuals according to their probability
   ii. maintains a Boltzmann distribution

   There is more than one possible scheme which will do this. You keep some fraction f of the population at each step or this may be determined each step probabilistically.

(c) How would you incorporate simulated annealing into this scheme?

(d) What are advantages of the Monte Carlo based approach?

3. (a) Somebody says that the key to a genetic algorithm is not mutation, but crossover events. A possible scheme would be at each generation, first throw away some number of individuals (maybe 12 of them). Of the remainder, pick 12, copy them and apply a mutation. Then pick pairs randomly. Copy them, but swap some characteristics within the pair.



This move is easy to implement if you can define your individuals as an array, such as a set of internal angles. How would you implement crossover as a move in Monte Carlo? Describe a simple method where there is no change in the number of individuals. If you start with two copies, you finish with two copies. To answer the question,

i. Write down the probability of the system (pair of individuals) before the crossover

ii. Write down the probability of the system after the crossover

iii. Write down an acceptance criterion

(b) What is the disadvantage of simply using crossover events in a genetic algorithm, without any real acceptance criterion?