NEW EXAMPLE: Consider Another Case: Example 2.5 and Figure 2.8

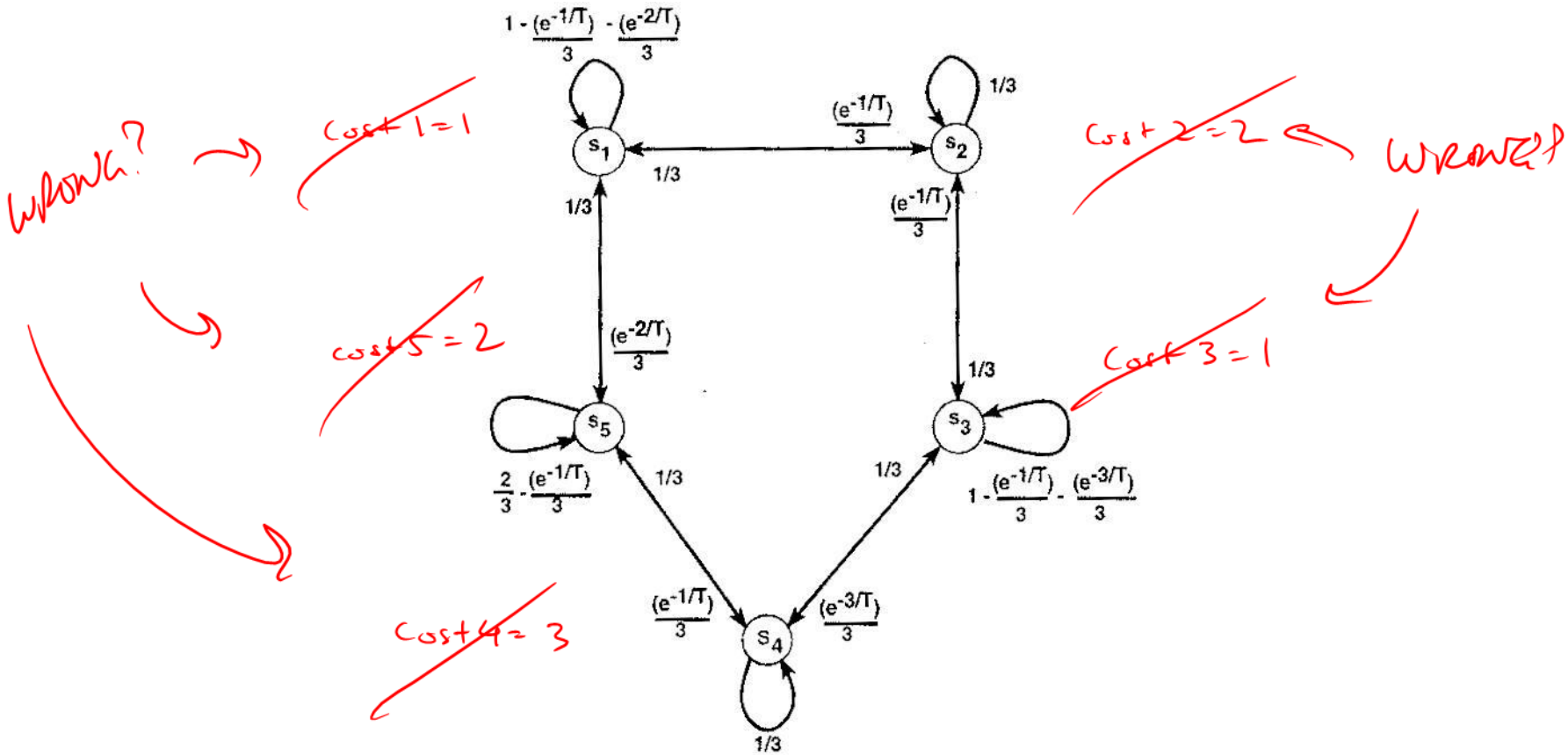


Figure 2.8: Configuration graph for Example 2.5.

define t so that $t = e^{-1/T}$, then $t^j = e^{-j/T}$

not iteration #

Handout 10-26-11

Double limit analysis for Example 2.5/Fig. 2.8

To find the final solution of the eventual SA search, we do the following:

1. Holding *T* constant, let *t* (number of iterations) go in the limit to infinity)
2. Given the limit in 2 (which is a function of *T*), let *T* go in the limit to 0.
3. In the book they do the substitution t=exp(-1/T). Don't confuse this t with the same symbol used to mean the iteration number.

The book makes the substitution **$t=e^{-1/T}$.** Don't confuse this t, with the number of iterations. The transition matrix **(with SA and T>0) is**

$\frac{1}{3}e^{-1/t}$

$$
D = \begin{array}{c|ccccc}
 & 1 & 2 & 3 & 4 & 5 \\
\hline
1 & \boxed{1-\frac{t}{3}-\frac{t^2}{3}} & \frac{t}{3} & 0 & 0 & \frac{t^2}{3} \\
2 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\
3 & 0 & t/3 & 1-\frac{t}{3}-\frac{t^3}{3} & \frac{t^3}{3} & 0 \\
4 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & 1/3 \\
5 & \frac{1}{3} & 0 & 0 & t/3 & \frac{2}{3}-\frac{t}{3}
\end{array}
$$

Solving for the stationary distribution for this matrix (as $t \to \infty$)

$\uparrow$ iteration #

$$\pi^* = \left(\overset{\pi_1}{\frac{1}{N}}, \overset{\pi_2}{\frac{t}{N}}, \overset{\pi_3}{\frac{1}{N}}, \overset{\pi_4}{\frac{t^3}{N}}, \overset{\pi_5}{\frac{t^2}{N}}\right) \quad \text{for } N = 2 + t + t^2 + t^3$$

$$\text{so } \pi_1 = \frac{1}{2 + t + t^2 + t^3} = \pi_3$$

$$t = e^{-1/\tau} \quad \text{so} \quad \text{as } \tau \to 0, \ t \to 0$$

$$\Rightarrow \lim_{t \to 0} \pi_1 = 1/2 = \lim_{t \to 0} \pi_3$$

$$\lim_{t \to 0} \pi_2 = \lim_{t \to 0} \frac{t}{2 + t + t^2 + t^3} = \frac{0}{2} = 0$$

Similarly π_4 + π_5 have t^3 and t^2 in numerator

$$\text{so } \lim_{t \to 0} \pi_4 = 0, \ \lim_{t \to 0} \pi_5 = 0$$

$$\text{so } \lim_{\tau \to 0} \pi(\tau) = (1/2, 0, 1/2, 0, 0)$$

Meaning of limiting π in Figure 2.8

- Lim as T goes to zero is $\pi(T)=(1/2,0,1/2,0,0)$

- What does that mean? Is it consistent with what you would expect by looking at the Figure 2.8?

- Is This consistent with the General SA Theory results given earlier? What is Ω for Figure 2.8?

Usefulness of Markov Chain Theory for larger problems

- All of these cases (on previous slides) can be analyzed by the theory to determine if they will they converge (with enough iterations) to the global optimium.

- How big is the matrix P in each of these cases?

- Can you compute the numerical value π* for all these cases?

  Why or why not?

$$\lim_{T \to 0} \left[\lim_{t \to \infty} \pi(z, T) \right]$$

do you reach this?

So what good is the theory?

$$\lim_{T \to 0} \left[\pi^*(t) \right]$$

SA GENERAL THEORETICAL RESULTS

Lim t goes to infinity and then Lim T goes to zero

(**Repeat Slide**)

- **From Ergodic Markov Chain Theory for Simulated Annealing we know that**

if $\pi^* = \lim_{t \to \infty} \pi(t)$, then

$$t \to \infty \quad \pi_i^* = \frac{e^{-Cost_i / T}}{N(T)}, \quad N(T) = \sum_{k=1}^{n} e^{-\Delta Cost_{ik} / T} \qquad (2.14)$$

$$T \to 0 \quad \lim_{T \to 0} \pi_i^*(T) = \begin{cases} \dfrac{1}{|\Omega_0|} & \text{if } S_i \in \Omega_0 \\ 0 & \text{if } S_i \notin \Omega_0 \end{cases}$$

where Ω_0 is the set of optimal solutions

(so $Cost_i$ is equal for all members of Ω_0)

End handout 10-20-10

Theory as it relates Numerical Application of SA

- Are the assumptions of the theory satisfied when you do the calculations ?
- The theory says that finally we want to take

Lim as T goes to zero of $\pi^*(T)$

How do we get $\pi^*(T)$, which is itself a limit?

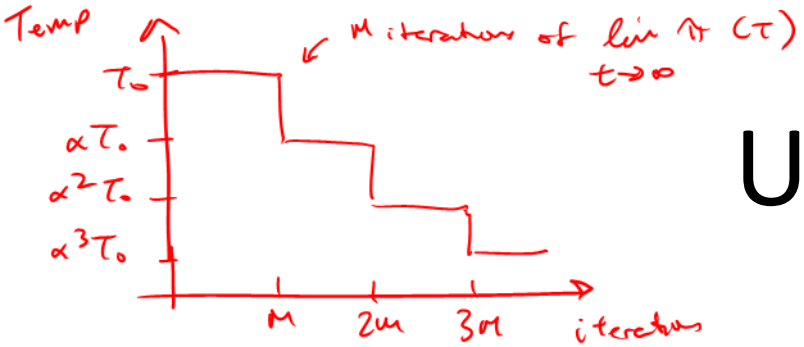
Theory as it relates Numerical Application of SA

- Are the assumptions of the theory satisfied when you do the calculations ?
- The theory says that finally we want to take

"Lim as T goes to zero of $\pi^*(T)$"

How do we get $\pi^*(T)$, which is itself a limit?

Is this what the SA code is doing?



Use of M

- Recall in the description of SA there was an “M” parameter.
- The idea with the M parameter is that you run the SA algorithm with T =T1 for M iterations.
- Then you lower T to T2 for M more iterations, etc.
- This is an attempt to mimic more of what the theory is requiring, that you try to converge to π*(Tj) for each of a series of decreasing values of Tj.

Differences between Theory and Numerical Application

- However, your numerical approach is not identical to the theoretical approach because

  might never reach $\pi^*(\tau)$ or low $\pi^*(\tau)$?

  $\tau \to 0$

- So what is the value of the theoretical approach?
- What about other methods, e.g. derivative based methods that converge linearly? Does the numerical algorithm match the theoretical ?

Bottom Line on SA Theory

- Among Heuristic Algorithms, the SA theory is the among the most Mathematically sophisticated
- SA theory does give us a great deal of insight about how the algorithm works even when it is searching in discrete domains with millions or billions of points .
- SA theory does indicate that under the assumptions of ergodic Markov Chains, the probability of finding the best answer is increasing with more iterations.

Bottom Line on SA Theory

- However, the theory does <u>not</u> give us information about what we would really like to know, which is how many iterations do I have to do to get an answer that is within x% of the correct answer?
- Most non heuristic methods for nonlinear don't give you that information either since you don't know how large your initial error is.
- (However in the special case of continuous optimization where you have accurate second derivatives and a convex function, the convergence can be very quick since error drops quadratically .)

Next Topic is Theory for Genetic Algorithms

- A basic definition in Genetic Algorithms is the SCHEMA.

- A schema is a set of genes that make up a partial solution to our optimization problem.
- Plural of schema is “schemata”.
- Schemata defining subsets of similar chromosomes.
- We denote a building block as {1*0***} where the * indicates it can be either 0 or 1.

Let H_1, H_2, H_3, H_4, be four schemata represent by the following strings

$H_1 = [1\ 1\ *\ *\ *\ *]$

$H_2 = [1\ *\ *\ *\ *\ 0]$

$H_3 = [1\ *\ *\ *\ *\ *]$

$H_4 = [0\ *\ *\ *\ *\ *]$

	Fitness
$s_1 = [0\ 11\ 001]$	625
$s_2 = [1\ 0\ 11\ 00]$	1936
$s_3 = [1\ 10\ 101]$	2809
$s_4 = [111\ 000]$	3136

So H_1 "matches" s_3 + s_4

H_2 matches ___ + ___ ?

H_4 matches ___

Order and Length of Schema

- The order of a schema H denoted by o(H) is the number of non-* symbols it contains
- The defining length denoted by δ(H) is the distance from the first non* to the last non*position.
- So for *1*0*110, the order is 5 and its defining length is 6 (=8-2)
- During a crossover a schema may be cut, so that is why we the GA theory depends on schema.
- The text calls the symbols in the decision variable “metasymbols”, which for binary strings are 0’s and 1’s.