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 $t^{ij} = e^{-ij/T}$
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)

\[
\begin{pmatrix}
1 & 2 & 3 & 4 & 5 \\
1 & \frac{1}{3} e^{-1/t} & \frac{t}{3} & 0 & 0 \\
2 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\
3 & 0 & \frac{1}{3} & 1 - \frac{T}{3} - \frac{t^3}{3} & \frac{t^3}{3} \\
4 & 0 & 0 & \frac{1}{3} & \frac{1}{3} \\
5 & \frac{1}{3} & 0 & 0 & \frac{1}{3} - \frac{T}{3} - \frac{t^3}{3}
\end{pmatrix}
\]
Solving for the stationary distribution for this matrix (as $t \to \infty$)

$\pi_1^* = \left( \frac{1}{N}, \frac{1}{N}, \frac{1}{N}, \frac{1}{N}, \frac{1}{2} \right)$ for $N = 2 + t + t^2 + t^3$

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

$t = -\frac{1}{t}$ so as $T \to 0$, $t \to 0$

\[ \lim_{t \to 0} \pi_1 = \frac{1}{2} = \lim_{t \to 0} \pi_3 \]

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

Similarly $\pi_4$ and $\pi_5$ have $t^3$ and $t^2$ in numerator

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

So $\lim_{T \to 0} \pi(T) = \left( \frac{1}{2}, 0, \frac{1}{2}, 0, 0 \right)$
Meaning of limiting $\pi$ 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 $\Omega$ 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 converge (with enough iterations) to the global optimum.

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

- Can you compute the numerical value $\pi^*$ for all these cases?
  Why or why not?

So what good is the theory?
SA GENERAL THEORETICAL RESULTS

Lim \( t \) goes to infinity and then \( \text{Lim } T \) goes to zero
(Repeat Slide)

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

\[
\lim_{t \to \infty} \pi_i^* = \frac{e^{-\frac{\text{Cost}_i}{T}}}{N(T)}
\]

\[
N(T) = \sum_{\varphi=1}^{n} e^{-\frac{\Delta \text{Cost}_i^\varphi}{T}}
\]  

\[
\lim_{T \to 0} \pi_i^*(T) = \frac{1}{1 - \Omega_0} \quad \text{if } S_i \in \Omega_0
\]

\[
0 \quad \text{if } S_i \notin \Omega_0
\]

(Where \( \Omega_0 \) is the set of optimal solution)

so \( \text{Cost}_i \) is equal for all members of \( \Omega_0 \)

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 \( \lim_{T \to 0} \pi^*(T) \)

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 $T = T_1$ for M iterations.

• Then you lower $T$ to $T_2$ for M more iterations, etc.

• This is an attempt to mimic more of what the theory is requiring, that you try to converge to $\pi^*(T_j)$ for each of a series of decreasing values of $T_j$. 
Differences between Theory and Numerical Application

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

\[
\lim_{\tau \to 0} \text{reach } \Pi^*(\tau) \text{ or } \lim_{\tau \to 0} \Pi^U(\tau) = \text{________}\n\]

• 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 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 not 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.)
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 represented by the following strings:

$H_1 = [1 \ 1 \ 1 \ \ldots\ h]
$ $H_2 = [1 \ 1 \ 1 \ \ldots\ 0]
$ $H_3 = [1 \ 1 \ 1 \ \ldots\ h]
$ $H_4 = [0 \ 0 \ 0 \ \ldots\ h]

So $H_1$ "matches" $S_3 + S_4$

$H_2$ matches ___ + ___

$H_4$ matches ___

$s_1 = [0 1 1 0 0 1]$ Fitness 625

$s_2 = [1 0 1 1 0 0]$ Fitness 1936

$s_3 = [1 1 0 1 0 1]$ Fitness 2809

$s_4 = [1 1 1 0 0 0]$ Fitness 3136
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 \( \delta(H) \) 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.