Exploration vs. Exploitation

• Evolutionary search is commonly viewed as a process of exploiting the information of the current search points to explore other points in the search space.
• To be effective, evolutionary search requires a balance between exploitation (continuous population fitness improvement) and exploration (maintaining population diversity).

Exploration and Exploitation

• Selection is the primary source of exploitation.
• The level of selection pressure is commonly measured by the speed of population fitness improvement (see your Assignment 1).
• Variation is the primary source of exploration.
• You will analyze how variation operators affects population diversity in Assignment 2.
• Based on these experimental results, you will chose suitable selection and variation methods to design an EA that finds the global optimum of 2-D Rosenbrock function.
Other Population Diversity Maintenance Techniques

• Fitness Sharing (Goldberg and Richardson, 1987)
• Crowding Technique (De Jong, 1975)
• Heterogeneous crossover (Eshelman and Schaffer, 1991)
• Seduction selection (Ronald, 1995)
• Island Model Population Structure

Population Diversity Measures

• Fitness Diversity:
  1. Measured by the number of unique fitness values in the population, divided by the size of the population.
  2. Measured by the standard deviation of fitness values in a population.

\[ \text{stddev}(P) = \sqrt{\frac{\sum_{i=1}^{N} (f_i - \bar{f})^2}{N-1}} \]

• Is maintaining population fitness diversity beneficial for evolutionary search?

Genotype Distance Measurements

• Hamming distance (for discrete values)
  – E.g. HD (0011,1001) = 2
  – E.g. HD (ACDEF, ADDFF) = 2

• Euclidean distance (for real values)
  – vector difference between two individuals
  – E.g. a = (1,1), b = (4,4),

\[ d((1,1),(4,4)) = \sqrt{(4-1)^2 + (4-1)^2} = 4.243 \]

Population Diversity Measures—Continued

• Genotype (discrete-value) diversity:
  1. Measured by the sum of the Hamming Distance between any two genotypes in the population.

\[ \text{diversity}(P) = \sum_{i \neq j} \text{Ham} \min g(P[i], P[j]) \]

• The number of pairs is \( \frac{n(n-1)}{2} \)
  2. Measured by the number of unique genotype in the population, divided by the size of the population.

\[ \text{diversity}(P) = \frac{n}{n^2} \]
Population Diversity Measure - Continued

- Genotype (continuous-value) diversity:
  1. Measured by the sum of the Euclidean distance between any two genotypes in the population.
\[
\text{diversity}(P) = \sum_{i,j} \text{Euclidean}(P[i], P[j])
\]
- The number of pairs is \( \binom{n}{2} = \frac{n!}{(n-2)!2!} \)
  2. Measured by population dispersion: the average Euclidean distance of population members from the centroid of the population (see code).

Niching

- In ecology, a niche refers to the total of an organism’s use of the living and nonliving environment, e.g. the location of its dwelling and its food source etc.
- In EC, niching means competing solutions cannot coexist in the same niche. If two competing solutions are within the same range of the search space (similar individuals), they have to share their fitness.
- The idea is to discourage similar individuals in the population so that the genetic diversity is maintained.

Fitness Sharing Technique (Goldberg and Richardson, 1987)

- Individual i share its fitness \( f(x_i) \) with all individuals in the population who are within the distance threshold \( \sigma_{sh} \) from i.
- \( \sigma_{sh} \) is the predefined distance threshold.
\[
f(x_i) = \frac{f(x_i)}{\sum_{j=1}^{n} sh(d_{ij})}
\]
\[
sh(d_{ij}) = \begin{cases} 
1 - \left( \frac{d_{ij}}{\sigma_{sh}} \right)^{\alpha} & \text{if } d_{ij} < \sigma_{sh} \\
0 & \text{otherwise}
\end{cases}
\]

Sharing function calculates sharing counts

Triangular Sharing Function (\( \alpha=1 \))

- The Sharing function of an individual to itself is \( sh(d_{ii}) = 1 - \frac{0}{\sigma_{sh}} = 1 \)
How it works

• When many individuals are in the same neighborhood, they contribute to one another’s sharing count, thereby derating one another’s fitness value.
• No individual receives fitness from any other individual.

\[ f'(x) \leq f(x) \]

• This mechanism limits the uncontrolled growth of particular species (similar individuals) within a population.

Crowding Technique
(De Jong, 1975)

• When a new solution is created, it replaces the one in the population that is most similar and has a worse fitness.
• “Similarity” is defined by the same distance measures used in the fitness sharing technique (Hamming and Euclidean).
• In this way, population contains individuals spread across the search space to avoid converging to local optima.

Examples

<table>
<thead>
<tr>
<th>x1</th>
<th>x2</th>
<th>f(x)</th>
<th>σ=0.5</th>
<th>f'(x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>3.4</td>
<td>2.6</td>
<td>18.32</td>
<td>1.88</td>
</tr>
<tr>
<td>b</td>
<td>3.1</td>
<td>2.4</td>
<td>15.37</td>
<td>1.28</td>
</tr>
<tr>
<td>c</td>
<td>2.3</td>
<td>2.4</td>
<td>11.05</td>
<td>1</td>
</tr>
<tr>
<td>d</td>
<td>3.4</td>
<td>2.8</td>
<td>19.4</td>
<td>1.6</td>
</tr>
</tbody>
</table>

\[ x \text{ shares fitness with } y \]

Implementation Details

• A steady-state GA with generational gap \( G \) (% of overlapping between parents and offspring in a population of size \( N \)).
• Fitness proportional selection is used to select parents and applied with crossover/mutation to produce offspring.
• Replace individual in the population per following procedures:
  – Define crowding factor \( CF \) (the sampling size)
  – Among the randomly selected CF individuals, the one that is most similar to the new offspring and has a worse fitness is replaced.
Other Methods

• Heterogeneous crossover (Eshelman and Schaffer, 1991): Strings with binary encoding must be of a certain Hamming distance (or Euclidean distance) away from each another to be allowed to crossover.
• Incest prevention

Other Method - Continued

• Seduction selection (Ronald, 1995): Select the first parent as normal but selecting the second parent according to the preferences of the first parent.
  – After the first chromosome involved in a recombination operation is selected, all other individuals in the population are provided with a secondary fitness according to certain rules (a seduction function) that reflects the preferences of the first parent. Then, the second parent is chosen according to the secondary fitness.

Island Model Population Structure

• In biological evolution, “most species contain numerous small, random breeding local populations (demes)”-Wright, 1964
• “Small populations are suited to rapid adaptive change following the onset of reproductive isolation”- Eldredge and Gould,1972
• In EC, island model uses a population structure that involves sub-populations which have their isolated evolution, occasionally punctuated by inter-population communication [Cohon et al, 1991].

Island Model Structure

• The population is divided into $N$ sub-populations.
• The evolution process is divided into $E$ major iterations called epochs.
• During an epoch, each sub-population process independently for $G$ generations.
• After each epoch, there is a communication phase during which individuals migrate between neighboring subpopulations.
Pseudocode

Island_Model(E,N,µ)
{
    Concurrently for i ← 1 to N subpopulation
    Initialize (P_i,µ);
    For epoch ← 1 to E do
        Concurrently for each i ← 1 to N subpopulations
            Sequential_EA(P_i,G_i);
        For i ← to N do
            For each neighbor j of i
                Migration (P_i, P_j);
                Assimilate (P_i);
    }

Island Model Discussions

• The size of each sub-population P_i can vary. The smaller the subpopulation size, the finer the granularity.
• The communication topology decides the migration rate; the higher the degree of connectivity, the closer it becomes the single population model (ring topology is the most commonly used topology).
• After migration, the subpopulation may re-adjust the contents by deleting old individuals or by increasing the subpopulation size.