CSE/ECE 848 Introduction to Evolutionary Computation

Module 3 - Lecture 10 - Part 2

Evolutionary Strategies - Mutation and Recombination

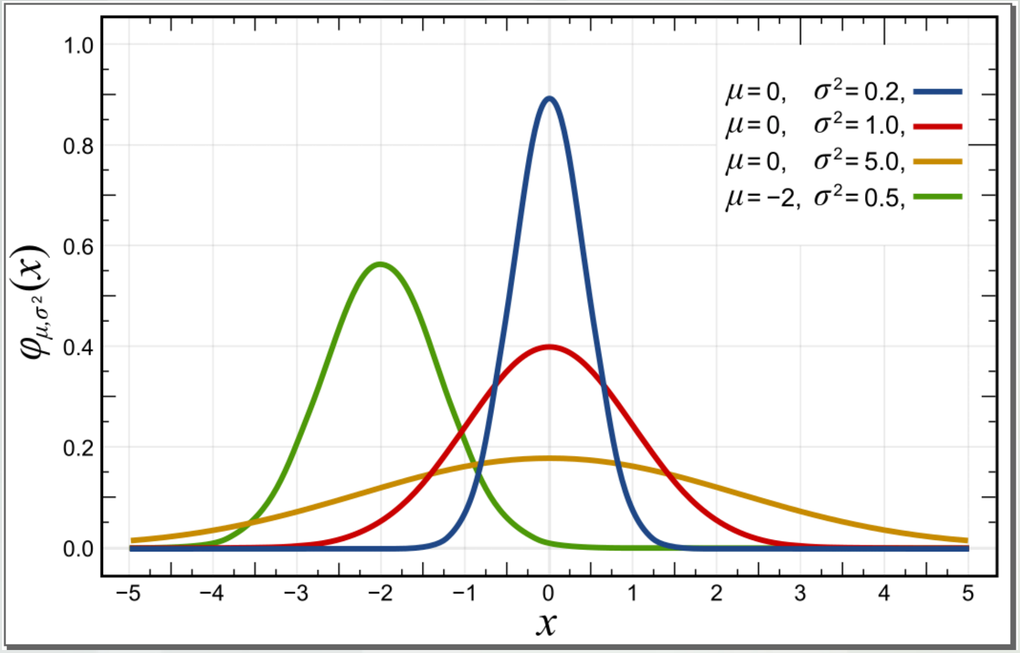
Wolfgang Banzhaf, CSE
John R. Koza Chair in Genetic Programming

Mutation

- As mentioned previously, mutation is the driving force
- x is the individual parameter value
- $x_i' = x_i + \sigma * N(0,1)$

- where σ is the stepsize, and N(0,1) represents a single standard Gaussian random variable

Normal Distribution, modified by σ



Mutation Strategies

- Early ES were concerned with the concept of stepsize σ
- “Stepsize” meant the strength of mutations
- However, the rate of change of each element of vector x should change over time as the element “converges” to better and better answers
- The issue of stepsize is the same as for any gradient hillclimber: If the step size is too small, little progress is made in exploring the space. If the stepsize is too big, the answer is missed because of overstepping in the space

Heuristic Mutation Strategy

- Rechenberg used two common functions to estimate optimal step sizes:
- Linear corridor

$$f_1(x) = F(x_1) = c_0 + c_1 x_1$$
$$\forall_i \in \{2, \dots, n\} : -b/2 \le x_i \le b/2$$

- The sphere

$$f_2(x) = c_0 + c_1 * \sum_{i=1}^{n} (x_i - x_i^*)^2$$

The 1/5 Rule

- He solved the optimal expected convergence rates stepsizes for those two (which were respectively)

$$p_{opt} \approx 0.184 \quad p_{opt} \approx 0.270$$

- and decided that the best rate of successful mutations to failed mutations should be about 0.20, or one fifth.

1/5 Rule II

This led to the basic mutation rule:

- If more than 1/5th of the mutations cause an improvement (in the objective function) then multiply σ by a factor 1+s
- If less than 1/5th of the mutations cause an improvement, then multiply σ by (1+s)^(-1/4)

Algorithm: (1+1)-ES with 1/5 success-rule	
1. Initialize $\boldsymbol{X}_0, \sigma_0$	
2. **repeat**	
3. $\widetilde{\boldsymbol{X}}_n = \boldsymbol{X}_n + \sigma_n \mathcal{N}(\mathbf{0}, \boldsymbol{I})$	Sample one offspring
4. **if** $f(\widetilde{\boldsymbol{X}}_n) \leq f(\boldsymbol{X}_n)$ **then**	If f(offsp.) $\leq f$(parent)
5. $\boldsymbol{X}_{n+1} = \widetilde{\boldsymbol{X}}_n$	New parent = offsp.
6. $\sigma_{n+1} = 1.5\,\sigma_n$	Step-size is increased
7. **else**	If offspring strictly worse
8. $\boldsymbol{X}_{n+1} = \boldsymbol{X}_n$	New parent = old parent
9. $\sigma_{n+1} = 1.5^{-1/4}\,\sigma_n$	Step-size is decreased
10. **until** stopping criteria is met	

After one successful and 4 unsuccessful mutations, this results in:

$$E(\sigma_{n+1}|\sigma_n) = \left((1.5)^{-1/4}\right)^{4/5} (1.5)^{1/5} \sigma_n$$

$$E(\sigma_{n+1}|\sigma_n) = \sigma_n$$

First ES

- The first ES was based on a (1+1)-ES strategy, where the child competes with the parent, and mutation is driven by the 1/5 rule.
- Note, population size is 1, and σ is global
- This is a kind of stochastic gradient method, best characterized as a local hill climber

$1/5^{th}$ rule and what it's good for

The 1/5 rule was derived for unimodal linear functions

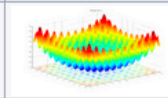
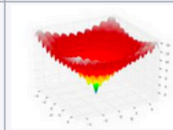
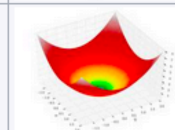
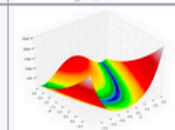
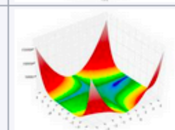
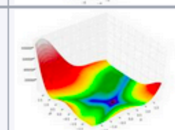
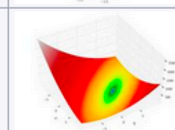
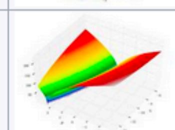
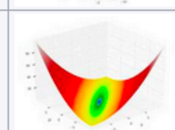
Thus not particularly useful for practical problems but indicative of a requirement:

One should adapt as the problem difficulty changes!

Test Functions

http://en.wikipedia.org/wiki/Test_functions_for_optimization

Test functions for single-objective optimization [edit]

Name	Plot	Formula	Global minimum	Search domain
Rastrigin function		$f(\mathbf{x}) = An + \sum_{i=1}^{n}\left[x_i^2 - A\cos(2\pi x_i)\right]$ where: $A = 10$	$f(0,\dots,0) = 0$	$-5.12 \le x_i \le 5.12$
Ackley's function		$f(x,y) = -20\exp\left[-0.2\sqrt{0.5\left(x^2+y^2\right)}\right] - \exp\left[0.5\left(\cos 2\pi x + \cos 2\pi y\right)\right] + e + 20$	$f(0,0) = 0$	$-5 \le x,y \le 5$
Sphere function		$f(\boldsymbol{x}) = \sum_{i=1}^{n} x_i^2$	$f(x_1,\dots,x_n) = f(0,\dots,0) = 0$	$-\infty \le x_i \le \infty$, $1 \le i \le n$
Rosenbrock function		$f(\boldsymbol{x}) = \sum_{i=1}^{n-1}\left[100\left(x_{i+1} - x_i^2\right)^2 + \left(x_i - 1\right)^2\right]$	$\text{Min} = \begin{cases} n=2 \rightarrow f(1,1)=0, \\ n=3 \rightarrow f(1,1,1)=0, \\ n>3 \rightarrow f(\underbrace{1,\dots,1}_{n \text{ times}})=0 \end{cases}$	$-\infty \le x_i \le \infty$, $1 \le i \le n$
Beale's function		$f(x,y) = \left(1.5 - x + xy\right)^2 + \left(2.25 - x + xy^2\right)^2 + \left(2.625 - x + xy^3\right)^2$	$f(3,0.5) = 0$	$-4.5 \le x,y \le 4.5$
Goldstein–Price function		$f(x,y) = \left[1 + (x+y+1)^2\left(19 - 14x + 3x^2 - 14y + 6xy + 3y^2\right)\right]\left[30 + (2x-3y)^2\left(18 - 32x + 12x^2 + 48y - 36xy + 27y^2\right)\right]$	$f(0,-1) = 3$	$-2 \le x,y \le 2$
Booth's function		$f(x,y) = (x + 2y - 7)^2 + (2x + y - 5)^2$	$f(1,3) = 0$	$-10 \le x,y \le 10$
Bukin function N.6		$f(x,y) = 100\sqrt{\left\lvert y - 0.01x^2\right\rvert} + 0.01\left\lvert x + 10\right\rvert.$	$f(-10,1) = 0$	$-15 \le x \le -5$, $-3 \le y \le 3$
Matyas function		$f(x,y) = 0.26\left(x^2+y^2\right) - 0.48xy$	$f(0,0) = 0$	$-10 \le x,y \le 10$

Recombination in (μ+1)-ES

- (μ+1)-ES allows the possibility of creating new individuals based on a combination of features of the parents, where $\mu > 1$
- Choose p parent vectors (1 <= p <= n), and mix characters from these p parent vectors to create a child
- Thus p=2 is similar to GA crossover

Recombination II

- Global intermediary recombination: Position i is average over all p parents
- Local intermediary recombination: Select two out of p parents for each child position i and take a weighted average
- Discrete recombination: copy a value from a randomly chosen parent for each child position
- Other scenarios are possible, too
- While recombination is used in ES, it is not the primary driving force, which is the previously described mutation