Optimization

Dr. Md. Aminul Haque Akhand
Dept. of CSE, KUET

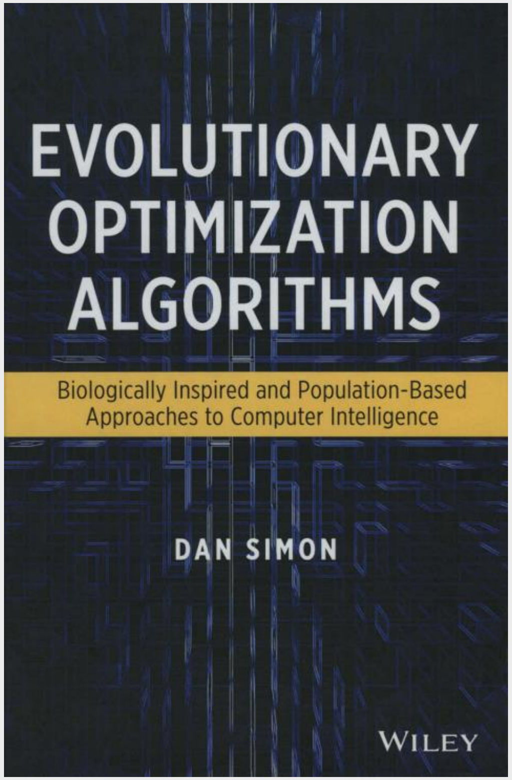
Contents

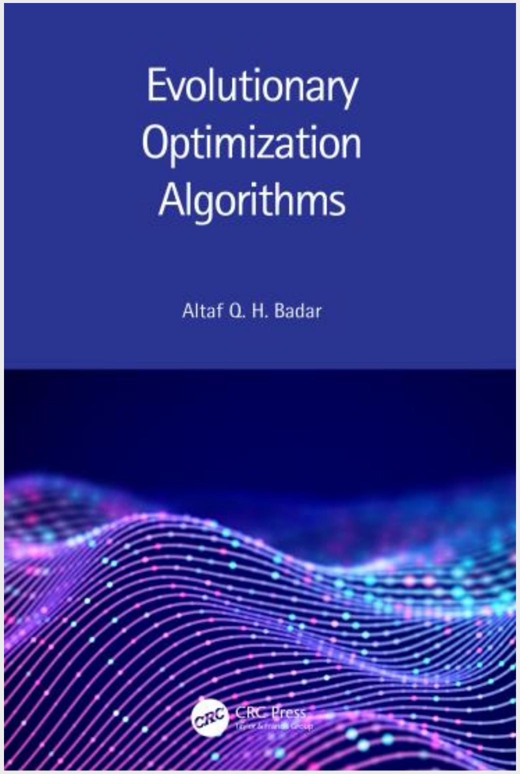
- Optimization Basics
- Unconstrained, Multi-Objective, Multimodal, Combinatorial Optimizations
- Hill Climbing
- Intelligence

Optimization Basics

Optimization saturates what we do and drives almost every aspect of engineering.

—Dennis Bernstein [Bernstein, 2006]





Unconstrained Optimization

- Optimization is applicable to virtually all areas of life. Optimization algorithms can be applied to everything from aardvark breeding to zygote research.
- The possible applications of EAs are limited only by the engineer's imagination, which is why EAs have become so widely researched and applied in the past few decades.

An optimization problem can be written as a minimization problem or as a maximization problem. [Sometimes we try to minimize a function and sometimes we try to maximize a function.]

These two problems are easily converted to the other form:

$$\begin{aligned} \min_x f(x) \quad &\Longleftrightarrow \quad \max_x [-f(x)] \\ \max_x f(x) \quad &\Longleftrightarrow \quad \min_x [-f(x)]. \end{aligned}$$

The number of elements in *x* is called the dimension of the problem. The function *f(x)* is called the objective function, and the vector *x* is called the independent variable, or decision variable.

$$\begin{aligned} \min_x f(x) \quad &\Rightarrow \quad f(x) \text{ is called ``cost'' or ``objective''} \\ \max_x f(x) \quad &\Rightarrow \quad f(x) \text{ is called ``fitness'' or ``objective.''} \end{aligned}$$

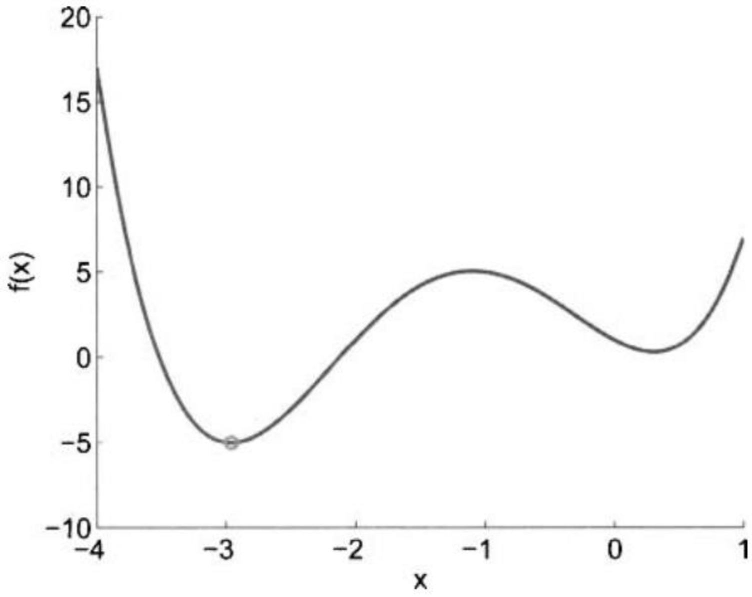
Unconstrained Optimization (Example)

$$f(x, y, z) = (x - 1)^2 + (y + 2)^2 + (z - 5)^2 + 3. \qquad (2.3)$$

- *f*(x, *y, z)* is called the objective function or the cost function.
- We can change the problem to a maximization problem by defining *g(x,y,z)* = -*f(x,y,z)* and trying to maximize *g(x,y,z)*. The function *g(x,y,z)* is called the objective function or the fitness function.
- The solution to the problem *min f*(*x,y,z)* is the same as the solution to the problem *max g(x,y,z),* and is $x = 1$, $y = -2$, and $z = 5$.
- However, the optimal value of *f*(x,*y,z)* is the negative of the optimal value of *g(x,y,z)*.

Unconstrained Optimization (Example)

$$\min_{x} f(x), \text{ where } f(x) = x^4 + 5x^3 + 4x^2 - 4x + 1. \qquad (2.4)$$



$$f'(x) = 4x^3 + 15x^2 + 8x - 4,$$

$f'(x) = 0$ at $x = -2.96$, $x = -1.10$, and $x = 0.31$.

$$f''(x) = 12x^2 + 30x + 8 = \begin{cases} 24.33, & x = -2.96 \\ -10.48, & x = -1.10 \\ 18.45, & x = 0.31 \end{cases}$$

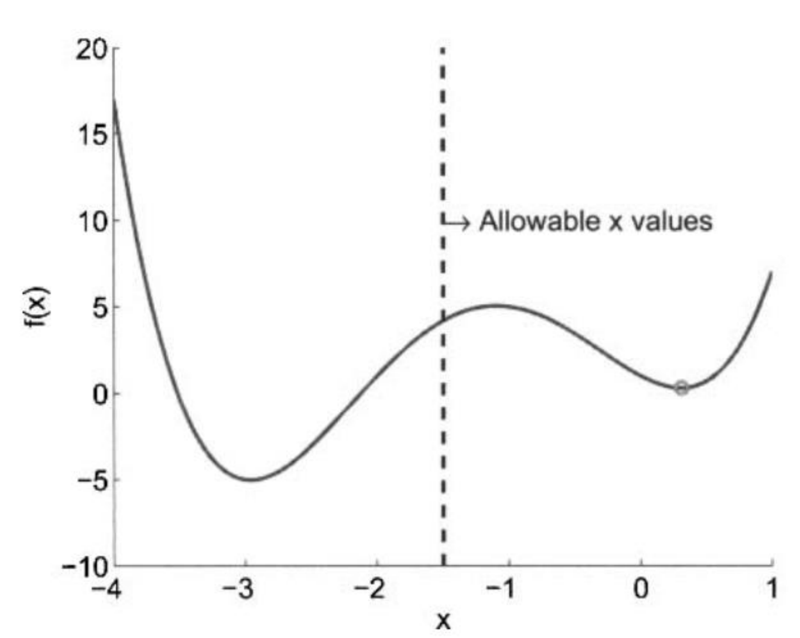
Figure 2.1 Example 2.2: A simple minimization problem. $f(x)$ has two local minima and one global minimum, which occurs at $x = -2.96$.

Recall that the second derivative of a function at a local minimum is positive, and the second derivative of a function at a local maximum is negative. The values of $f''(x)$ at the stationary points therefore confirm that $x = -2.96$ is a local minimum, $x = -1.10$ is a local maximum, and $x = 0.31$ is another local minimum.

A local minimum x^* can be defined as $f(x^*) < f(x)$ for all x such that $||x - x^*|| < \epsilon$

A global minimum x^* can be defined as $f(x^*) \leq f(x)$ for all x.

Constrained Optimization



$$\min_x f(x) \quad \text{where} \quad f(x) = x^4 + 5x^3 + 4x^2 - 4x + 1$$
$$\text{and} \quad x \geq -1.5.$$

$$f(x) = \begin{cases} 4.19 & \text{for } x = -1.50 \\ 0.30 & \text{for } x = 0.31 \end{cases}.$$

$f'(x) = 0$ at $x = -2.96$, $x = -1.10$, and $x = 0.31$.

Figure 2.2 Example 2.3: A simple constrained minimization problem. The constrained minimum occurs at $x = 0.31$.

Multi-Objective Optimization

$$\min_x [f(x) \text{ and } g(x)], \quad \text{where} \quad f(x) = x^4 + 5x^3 + 4x^2 - 4x + 1$$
$$\text{and} \quad g(x) = 2(x+1)^2.$$

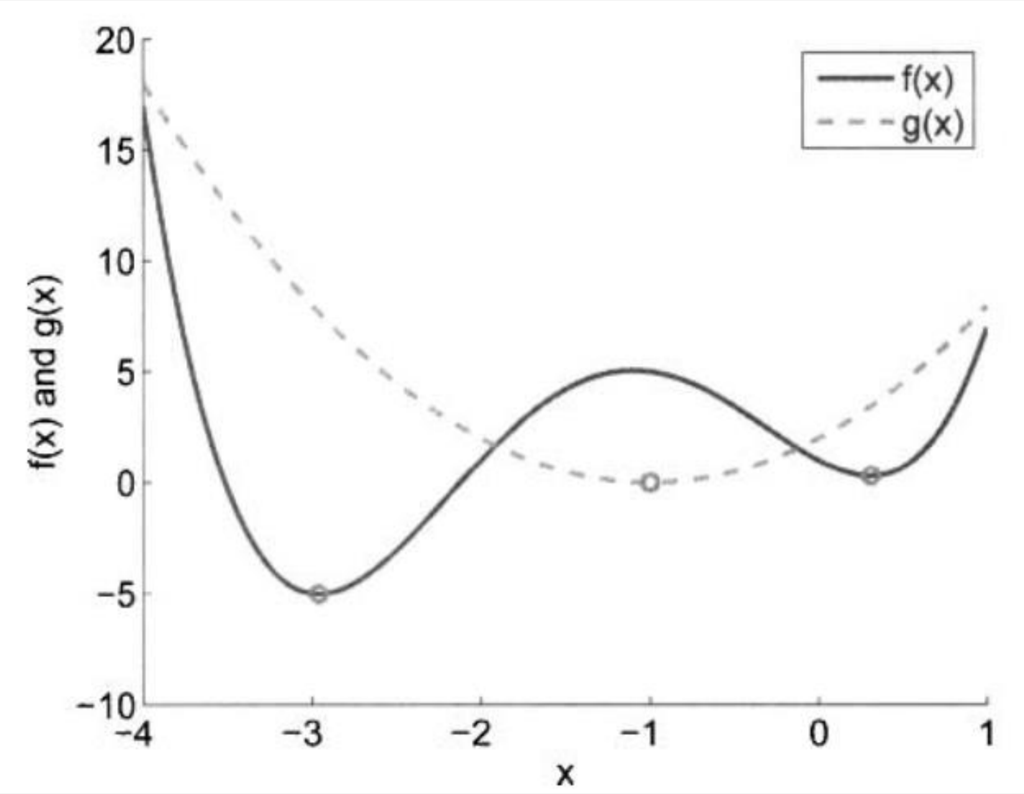
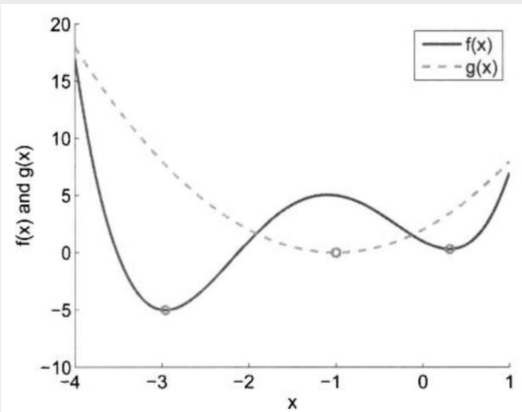


Figure 2.3 Example 2.4: A simple multi-objective minimization problem. $f(x)$ has two minima and $g(x)$ has one minimum. The two objectives conflict.

Multi-Objective Optimization (Cont.)

$$\min_x [f(x) \text{ and } g(x)], \quad \text{where} \quad f(x) = x^4 + 5x^3 + 4x^2 - 4x + 1$$
$$\text{and} \quad g(x) = 2(x+1)^2.$$



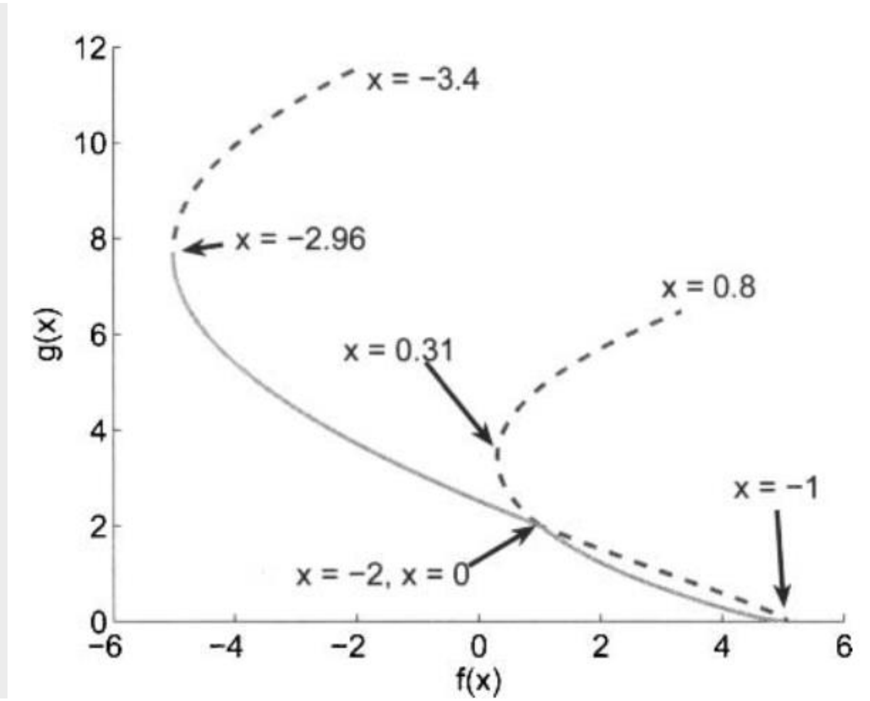


Figure 2.4 Example 2.4: This figure shows $g(x)$ as a function of $f(x)$ for a simple multi-objective minimization problem as x varies from -3.4 to 0.8. The solid line is the Pareto front.

The Pareto front gives a set of reasonable choices, but any choice of *x* from the Pareto set still entails a tradeoff between the two objectives.

- A typical real-world optimization problem involves many more objectives, and Pareto front is difficult to obtain.
- Even if we could obtain the Pareto front, we would not be able to visualize it because of its high dimensionality.

Multimodal Optimization

A multimodal optimization problem is a problem that has more than one local minimum.

$$\min_{x,y} f(x,y), \ \text{where}$$

$$f(x,y) = e - 20\exp\left(-0.2\sqrt{\frac{x^2+y^2}{2}}\right) - \exp\left(\frac{\cos(2\pi x)+\cos(2\pi y)}{2}\right)$$

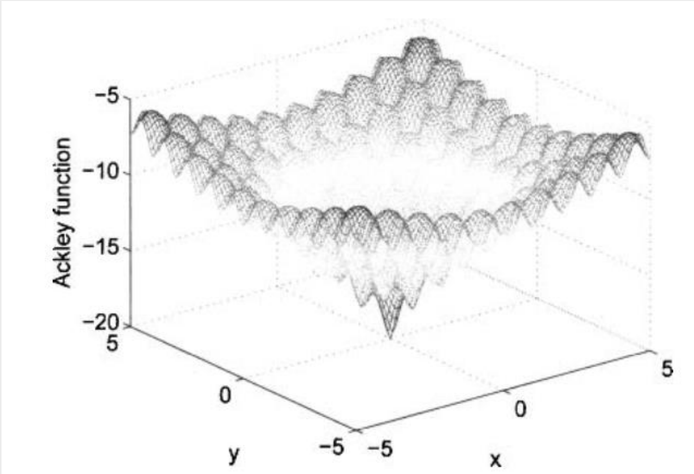
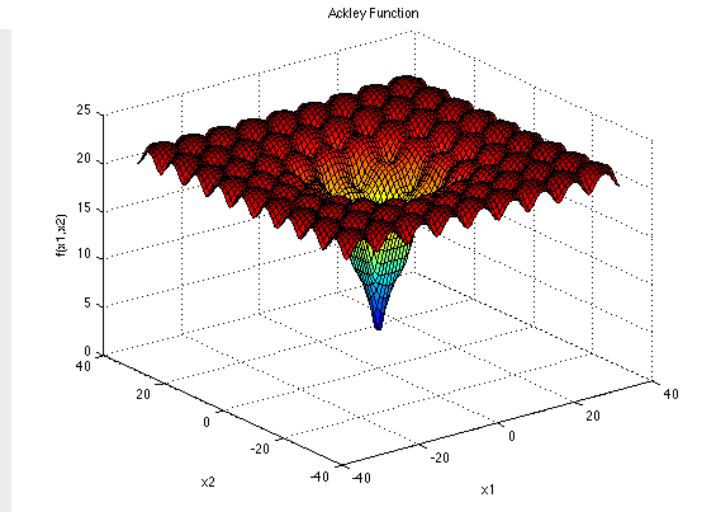


Figure 2.5 Example 2.5: The two-dimension Ackley function.



https://www.sfu.ca/~ssurjano/ackley.html

- We were able to solve shown previous examples using graphical methods or calculus
- But many real-world problems are more like this with more independent variables, with multiple objectives, and with constraints.
- Evolutionary Algorithms (EAs) are a good choice for real-world problems.

Combinatorial Optimization : Traveling Salesman Problem (TSP)

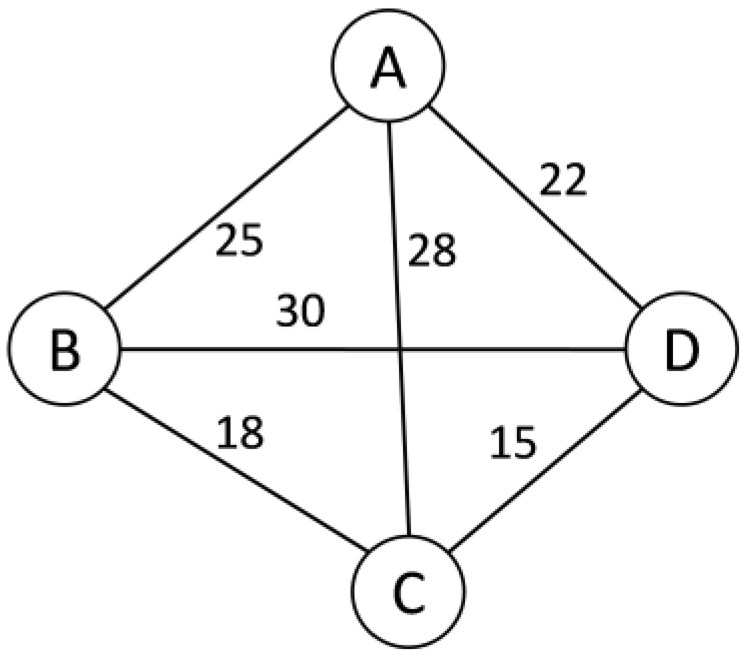


Figure 2.7: Simple Traveling Salesman Problem

There are four cities, A, B, C and D. Thus, the different possible routes would be (A as the hometown):
[A - B - C - D - A], [A - B - D - C - A],
[A - C - B - D - A], [A - C - D - B - A],
[A - D - C - B - A], [A - D - B - C - A]

If the salesman has to visit *n* cities there would be (*n*-1)! possibilities

- This number grows very rapidly, and for modest values of *n* it is not possible to calculate all possible solutions.
- Suppose the business person needs to visit one city in each of the 50 states in the USA. The number of possible solutions is 49! = 6.1×10^{62}.

Hill Climbing

Hill climbing (HC) is a simple optimization algorithm; actually, it is a family of algorithms with many variations.

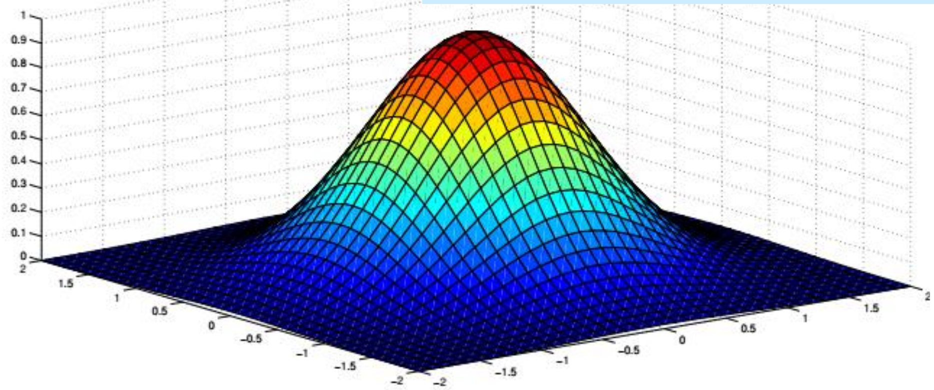


Figure 2.8: Simple Hill Climbing Problem

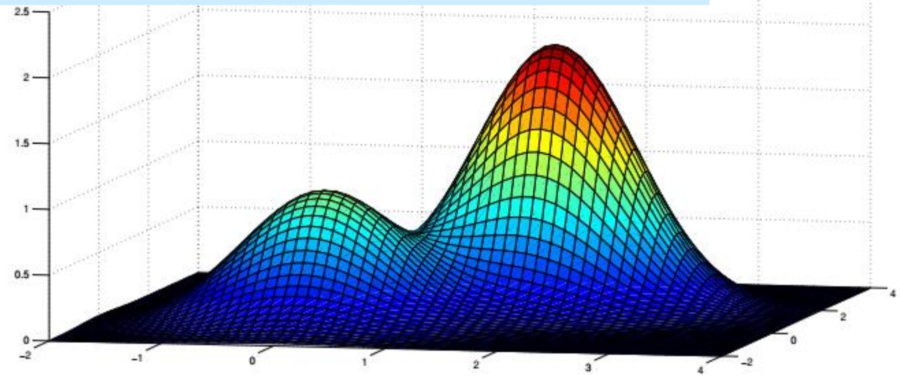


Figure 2.9: Multiple Hills in an Hill Climbing Problem

Figure 2.8 shows a simple hill-climbing surface and is the realization of the equation $e^{-(x^2+y^2)}$, where the value of x and y vary from -2 to 2.

Fig 2.9 is given by $e^{-(x^2+y^2)} + 2e^{-((x-1.7)^2+(y-1.7)^2)}$ in the range of -2 to 4.

- Some researchers consider HC to be a simple EA, while others consider it a non-evolutionary algorithm.
- HC is often a good first choice for solving a new optimization problem because it is simple, surprisingly effective
- The idea of HC is so straightforward that it must have been invented many times, and long ago, so it is difficult to determine its origin.

Hill Climbing

The different Hill Climbing solution techniques are:

1. **Simple Hill Climbing**: In a simple hill climbing algorithm, neighboring nodes are examined one after another. The first node, which gives a better objective function value than the existing position, is selected.

2. **Steepest Ascent Hill Climbing**: In the steepest ascent hill climbing technique all the neighboring positions are evaluated. The position which gives the largest improvement in the objective function value is selected as the next node.

3. **Iterative Hill Climbing**: In this technique, the process starts with a simple random guess. Increments are done for one input variable of the random solution.

4. **Stochastic Hill Climbing**: This method chooses a neighboring position randomly and evaluates its objective function value.

5. **Random Restart/Shotgun Hill Climbing**: In the random restart hill climbing method, the hill climbing is started with a random initial position.

Intelligence

- Computers were very good at some things that humans did poorly, like calculating the trajectory of a ballistic missile.
- But computers were (and still are) not effective at tasks that humans can do well, like recognizing a face.
- This led to attempts to mimic biological behavior in an effort to make computers better at such tasks.
- These efforts resulted in technologies like fuzzy systems, neural networks, genetic algorithms, and other EAs. EAs are therefore considered to be a part of the general category of computer intelligence.
- EAs should be *intelligent.* But what does it mean to be intelligent? Does it mean that our EAs can score high on an IQ test?
- Characteristics of Intelligence: adaptation, randomness, communication, feedback, exploration, and exploitation.

Intelligence Characteristics

- **Adaptation:**

  Adaptation to changing environments is a feature of intelligence.

- **Randomness:**
  - ✓ Some degree of randomness is a necessary component of intelligence.
  - ✓ Too much randomness will be counterproductive.
  - ✓ So randomness is a feature of intelligence, but only within limitations.

- **Communication:**
  - ✓ Communication is a feature of intelligence.
  - ✓ Intelligence not only involves communication, but it is also emergent. A single individual cannot be intelligent.
  - ✓ Communication is required to develop intelligence, and intelligence is required to communicate.
  - ✓ Intelligence and communication form a positive feedback loop.

Intelligence Characteristics

- **Feedback:**
  - Feedback is a fundamental characteristic of intelligence.
  - When we make mistakes, we change so that we don't repeat those mistakes. Feedback is also the basis for many natural phenomena.
  - Designed EA will have incorporate positive and negative feedback. An EA without feedback will not be very effective, but an EA with feedback has satisfied this necessary condition for intelligence.

- **Exploration and Exploitation:**
  - Exploration is the search for new ideas or new strategies.
  - Exploitation is the use of existing ideas and strategies that have proven successful in the past. Exploration is high-risk; a lot of new ideas waste time and lead to dead ends. However, exploration can also be high-return; a lot of new ideas pay off in ways that we could not have imagined.
  - Intelligence includes the proper balance of exploration and exploitation.

Exercises

2.1 Consider the problem $\min f(x)$, where

$$f(x) = 40 + \sum_{i=1}^{4} x_i^2 - 10\cos(2\pi x_i).$$

Note that $f(x)$ is the Rastrigin function – see Section C.1.11.

- **a)** What are the independent variables of $f(x)$? What are the decision variables of $f(x)$? What are the solution features of $f(x)$?
- **b)** What is the dimension of this problem?
- **c)** What is the solution to this problem?
- **d)** Rewrite this problem as a maximization problem.

2.2 Consider the function $f(x) = \sin x$.

- **a)** How many local minima does $f(x)$ have? What are the function values at the local minima, and what are the locally minimizing values of x?
- **b)** How many global minima does $f(x)$ have? What are the function values at the global minima, and what are the globally minimizing values of x?

2.3 Consider the function $f(x) = x^3 + 4x^2 - 4x + 1$.

- **a)** How many local minima does $f(x)$ have? What are the function values at the local minima, and what are the locally minimizing values of x?
- **b)** How many local maxima does $f(x)$ have? What are the function values at the local maxima, and what are the locally maximizing values of x?
- **c)** How many global minima does $f(x)$ have?
- **d)** How many global maxima does $f(x)$ have?

2.4 Consider the same function as in Problem 2.3, $f(x) = x^3 + 4x^2 - 4x + 1$, but with the constraint $x \in [-5, 3]$.

- **a)** How many local minima does $f(x)$ have? What are the function values at the local minima, and what are the locally minimizing values of x?
- **b)** How many local maxima does $f(x)$ have? What are the function values at the local maxima, and what are the locally maximizing values of x?
- **c)** How many global minima does $f(x)$ have? What is the function value at the global minimum, and what is the globally minimizing values of x?
- **d)** How many global maxima does $f(x)$ have? What is the function value at the global maximum, and what is the globally maximizing values of x?

Exercises

2.6 How many unique closed paths exist through N cities? By *unique* we mean that the starting city does not matter, and the direction of travel does not matter. For example, in a four-city problem with cities A, B, C, and D, we consider route $A \to B \to C \to D \to A$ equivalent to routes $D \to C \to B \to A \to D$ and $B \to C \to D \to A \to B$.

2.7 Consider the closed TSP with the cities in Table 2.2.

City	x	y
A	5	9
B	9	8
C	−6	−8
D	9	−2
E	−5	9
F	4	−7
G	−9	1

Table 2.2 TSP coordinates of cities for Problem 2.7.

a) How many closed routes exist through these seven cities?
b) Is it easy to see the solution by looking at the coordinates in Table 2.2?
c) Plot the coordinates. Is it easy to see the solution from the plot? What is the optimal solution? This problem shows that looking at a problem in a different way might help us find a solution.

Open Discussion