Chapter 8
Evolution Strategies and Other Methods

History

Year	Inventor	Technique	Individual
1958	Friedberg	learning machine	virtual assembler
1959	Samuel	mathematics	polynomial
1965	Fogel, Owens and Walsh	evolutionary programming	automaton
1965	Rechenberg, Schwefel	evolutionary strategies	real-numbered vector
1975	Holland	genetic algorithms	fixed-size bit string
1978	Holland and Reitmann	genetic classifier systems	rules
1980	Smith	early genetic programming	var-size bit string
1985	Cramer	early genetic programming	tree
1986	Hicklin	early genetic programming	LISP
1987	Fujiki and Dickinson	early genetic programming	LISP
1987	Dickmanns, Schmidhuber and Winklhofer	early genetic programming	assembler
1992	Koza	genetic programming	tree

Table 4.1

Evolution Strategies (ESs)

Evolution Strategies

- Developed: Germany in the 1970's
- Inventors: I. Rechenberg, H.-P. Schwefel
- Typically applied to:
  - numerical optimisation
- Attributed features:
  - fast
  - good optimizer for real-valued optimisation
  - relatively much theory
- Special:
  - self-adaptation of (mutation) parameters standard

ES technical summary

Representation	Real-valued vectors
Recombination	Discrete or intermediary
Mutation	Gaussian perturbation
Parent selection	Uniform random
Survivor selection	(μ,λ) or $(\mu+\lambda)$
Specialty	Self-adaptation of mutation step sizes

Evolution Strategies (ESs)

- *Two-membered ES (earliest)* $((1+1)-ES)$
  - Population size: one
  - Number of offspring: one
  - Representation: floating point $v=(x,\sigma)$
  - Alter operator: only mutation

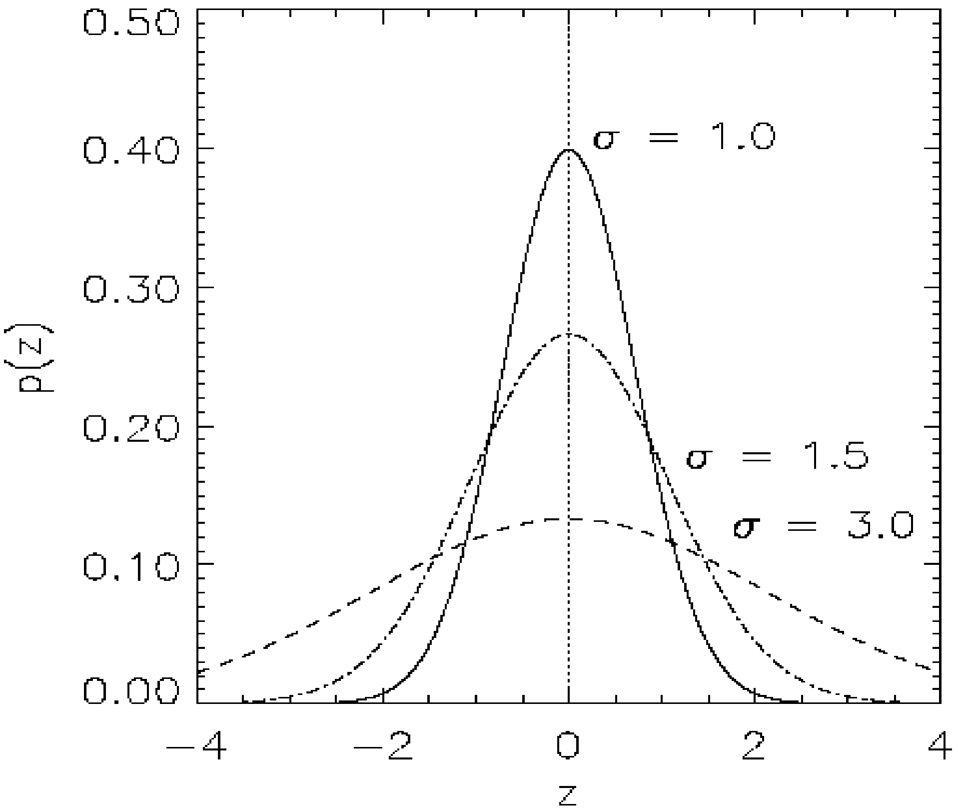
$$(x^{t+1},\sigma)=\begin{cases}(x^t+N(0,\sigma),\sigma) & iff \quad f(x^{t+1})>f(x^t)\\(x^t,\sigma) & otherwise\end{cases}$$

  - 1/5 success rule: σ

$$\sigma^{t+1}=\begin{cases}c_d\cdot\sigma^t, & if \quad \varphi(k)<1/5\\c_i\cdot\sigma^t, & if \quad \varphi(k)>1/5\\\sigma^t, & if \quad \varphi(k)=1/5\end{cases}$$

$$where \quad c_i>1 \text{ and } c_d<1$$

Normal Distribution



Example

Maximize

$$f(x_1, x_2) = 21.5 + x_1 \cdot \sin(4\pi x_1) + x_2 \cdot \sin(20\pi x_2)$$

where

$$4.1 \le x_2 \le 5.8$$
$$-3.0 \le x_1 \le 12.1$$

At some time t,

$$(X^t, \sigma) = ((5.3, 4.9), (1.0, 1.0))$$

The result of the mutation,

$$x_1^{t+1} = x_1^t + N(0, 1.0) = 5.3 + 0.4 = 5.7$$

$$x_2^{t+1} = x_2^t + N(0, 1.0) = 4.9 - 0.3 = 4.6$$

Since,

$$f(X^{t+1}) = f(5.7, 4.6) = 24.849532 > f(X^t) = f(5.3, 4.9) = 18.383705$$

So,

$$(X^{t+1}, \sigma) = ((5.7, 4.6), (1.0, 1.0))$$

- *Multi-membered ES* $((\mu + 1) - ES)$
  - Population size: $\mu > 1$
  - Number of offspring: one
  - Representation: floating point $v = (x, \sigma)$
  - Alter operators: crossover & mutation
    - *uniform crossover*:

$$v_1 = (x^1, \sigma^1) = ((x_1^1, \ldots, x_n^1), (\sigma_1^1, \ldots, \sigma_n^1))$$
$$v_2 = (x^2, \sigma^2) = ((x_1^2, \ldots, x_n^2), (\sigma_1^2, \ldots, \sigma_n^2))$$

then

$$v' = (x, \sigma) = ((x_1^{q_1}, \ldots, x_n^{q_n}), (\sigma_1^{q_1}, \ldots, \sigma_1^{q_n}))$$

where $q_i = 1 \text{ or } 2$

    - *mutation*:

(same as (1+1)-ES)

  - Competition: The weakest individual is eliminated.

$(\mu + 1 \rightarrow \mu)$

- *Multi-membered ES* $((\mu+\lambda)-ES)$ and $(\mu,\lambda)-ES))$
  - Population size: $\mu > 1$
  - Number of offspring: λ
  - Representation: floating point $v = (x,\sigma)$
  - Alter operators: crossover & mutation
    - *Uniform crossover*:

$$v_1 = (x^1,\sigma^1) = ((x_1^1,\ldots,x_n^1),(\sigma_1^1,\ldots,\sigma_n^1))$$
$$v_2 = (x^2,\sigma^2) = ((x_1^2,\ldots,x_n^2),(\sigma_1^2,\ldots,\sigma_n^2))$$

discrete:

$$v' = (x,\sigma) = ((x_1^{q_1},\ldots,x_n^{q_n}),(\sigma_1^{q_1},\ldots,\sigma_1^{q_n}))$$

intermediate:

$$v' = (x,\sigma) = ((x_1^1+x_1^2)/2,\ldots,(x_n^1+x_n^2)/2),$$
$$((\sigma_1^1+\sigma_1^2)/2,\ldots,(\sigma_n^1+\sigma_n^2)/2))$$

- *Mutation*:

$$\sigma' = \sigma \cdot e^{N(0,\Delta\sigma)}$$
$$x' = x + N(0,\sigma')$$

- Competition:

$$(\mu+\lambda)-ES \quad \Rightarrow \quad \mu+\lambda \to \mu$$
$$(\mu,\lambda)-ES \quad \Rightarrow \quad \lambda \to \mu \quad (\lambda > \mu)$$

ESs and GAs

- represent the individuals
- selection
- relative order of the procedures selection and recombination
- reproduction parameters
- simultaneously

Genetic Programing (GP)

Genetic Programing

- Developed: USA in the 1990's
- Inventors: J. Koza
- Typically applied to:
  - machine learning tasks (prediction, classification…)
- Attributed features:
  - competes with neural nets and alike
  - needs huge populations (thousands)
  - slow
- Special:
  - non-linear chromosomes: trees, graphs
  - mutation possible but not necessary (disputed!)

GP technical summary

Representation	Tree structures
Recombination	Exchange of subtrees
Mutation	Random change in trees
Parent selection	Fitness proportional
Survivor selection	Generational replacement

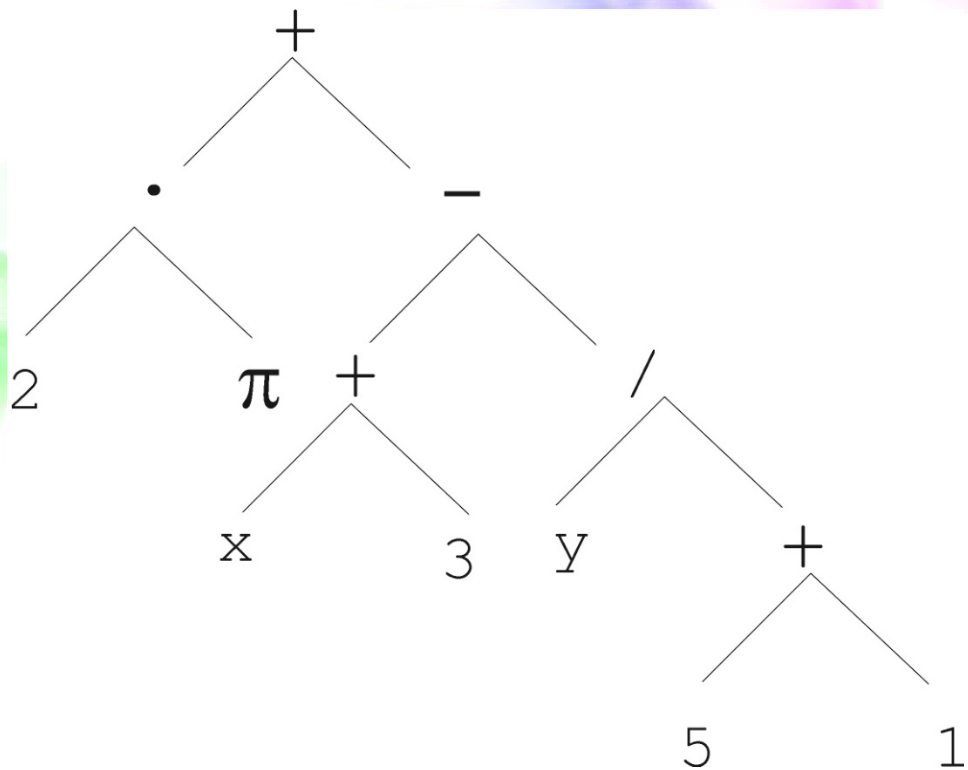
Representation

- Trees are a universal form
  - Terminal set (T)
  - Function set (F)

- Examples
  - Arithmetic formula
  - Logical formula
  - Program
  - …

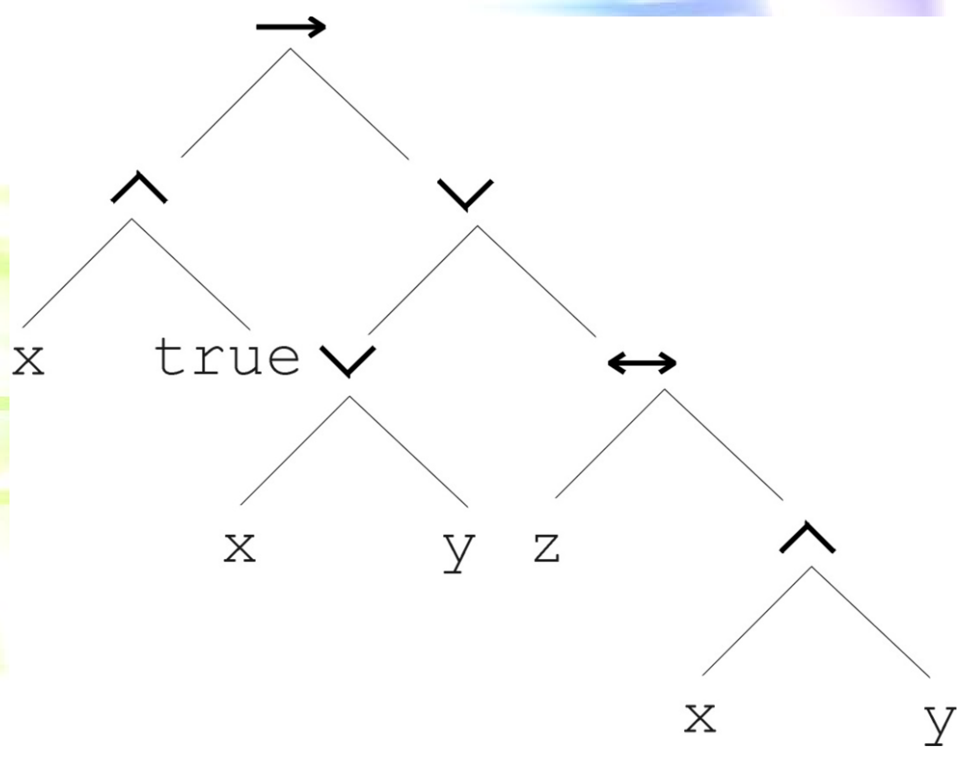
- Arithmetic formula

$$2 \cdot \pi + \left((x+3) - \frac{y}{5+1} \right)$$



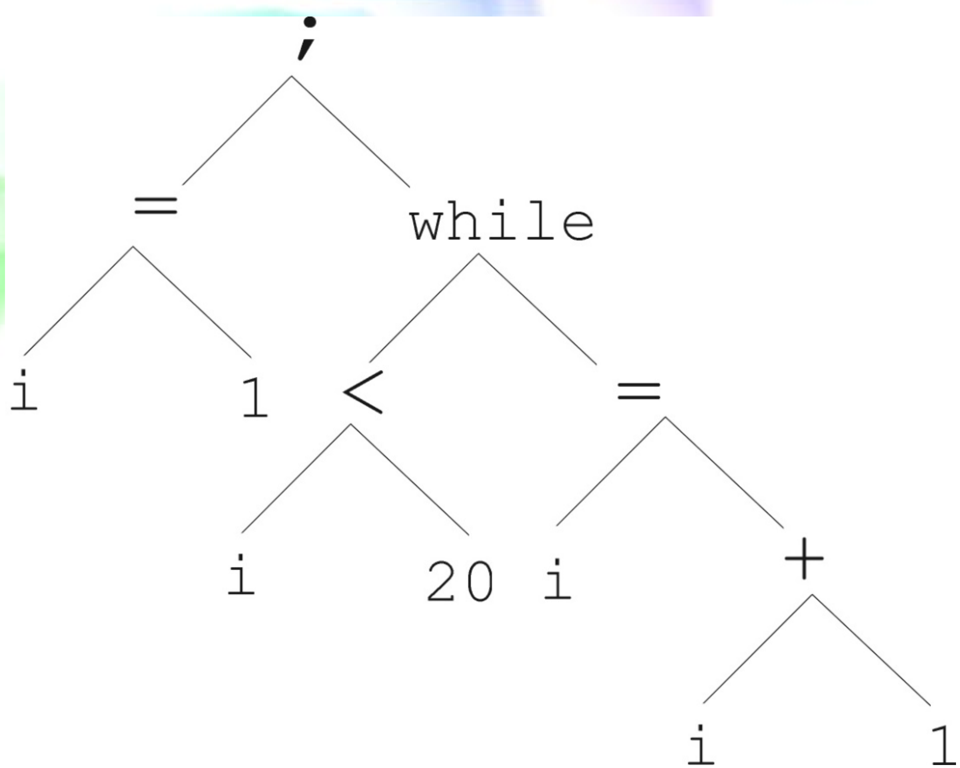
- Logical formula

$$(x \wedge \text{true}) \rightarrow ((x \vee y) \vee (z \leftrightarrow (x \wedge y)))$$



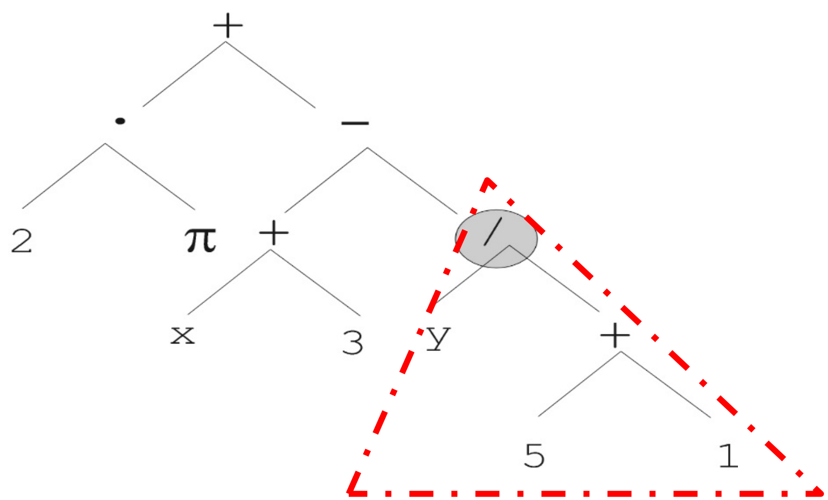
- Program

```
i =1;
while (i < 20)
{
        i = i +1
}
```

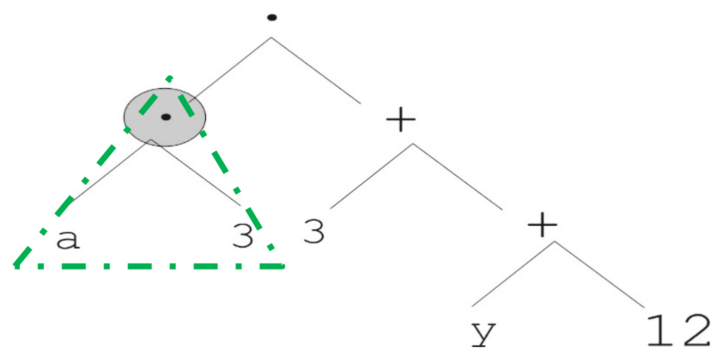


Crossover

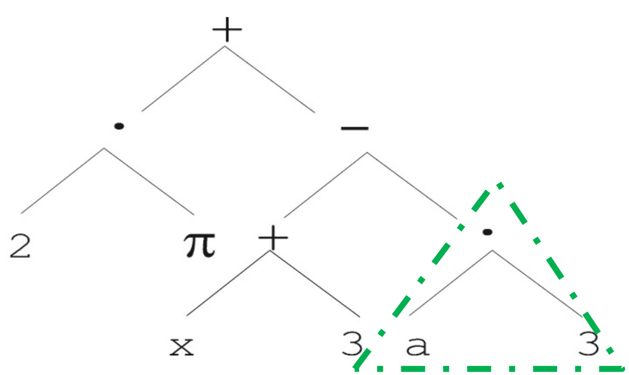
Parent 1



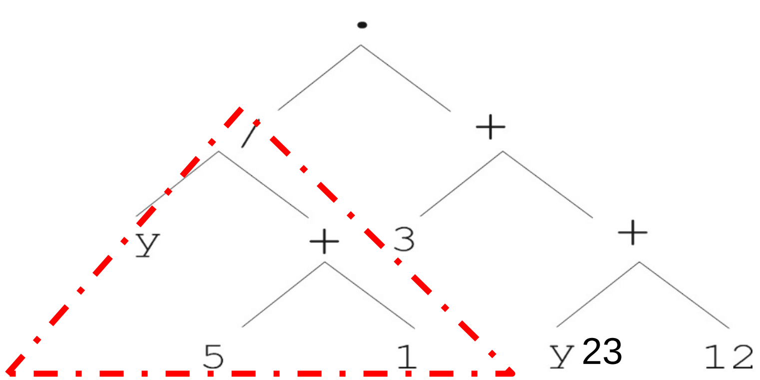
Parent 2



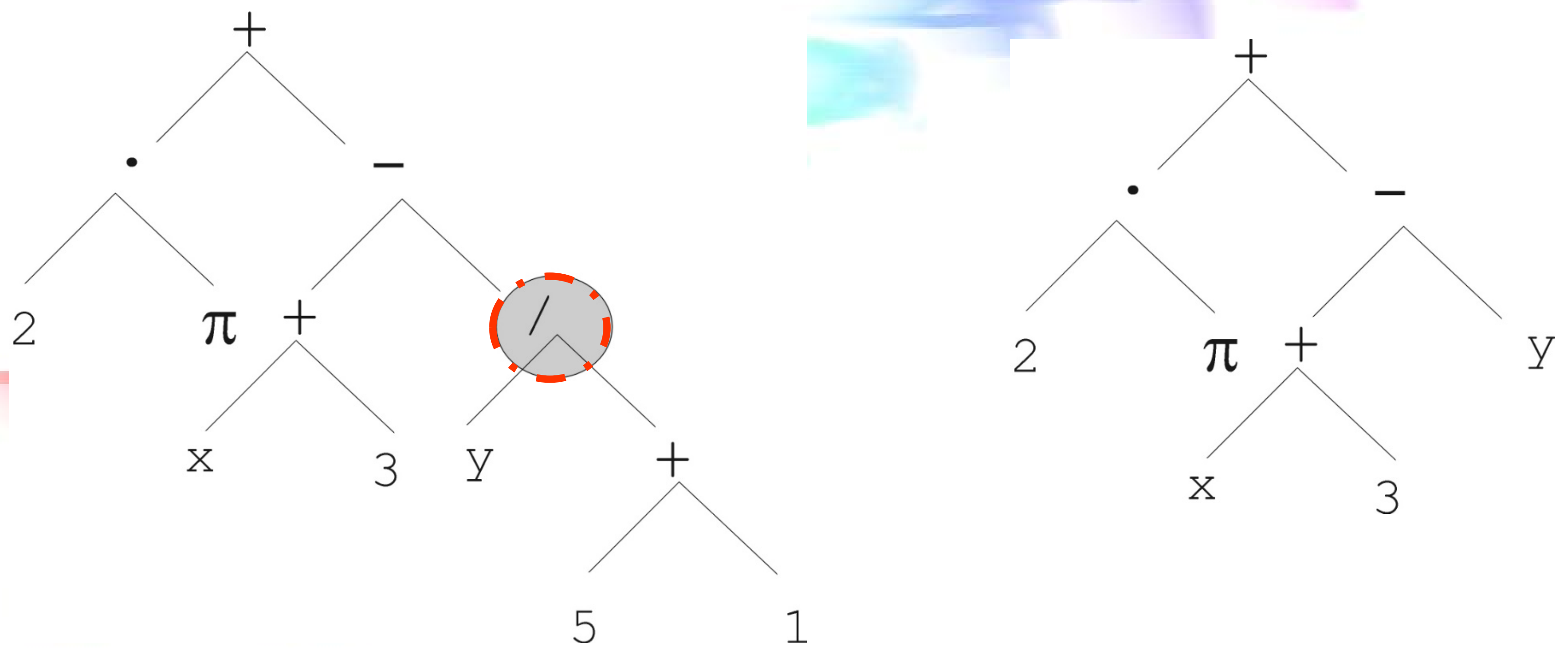
Offspring 1



Offspring 2



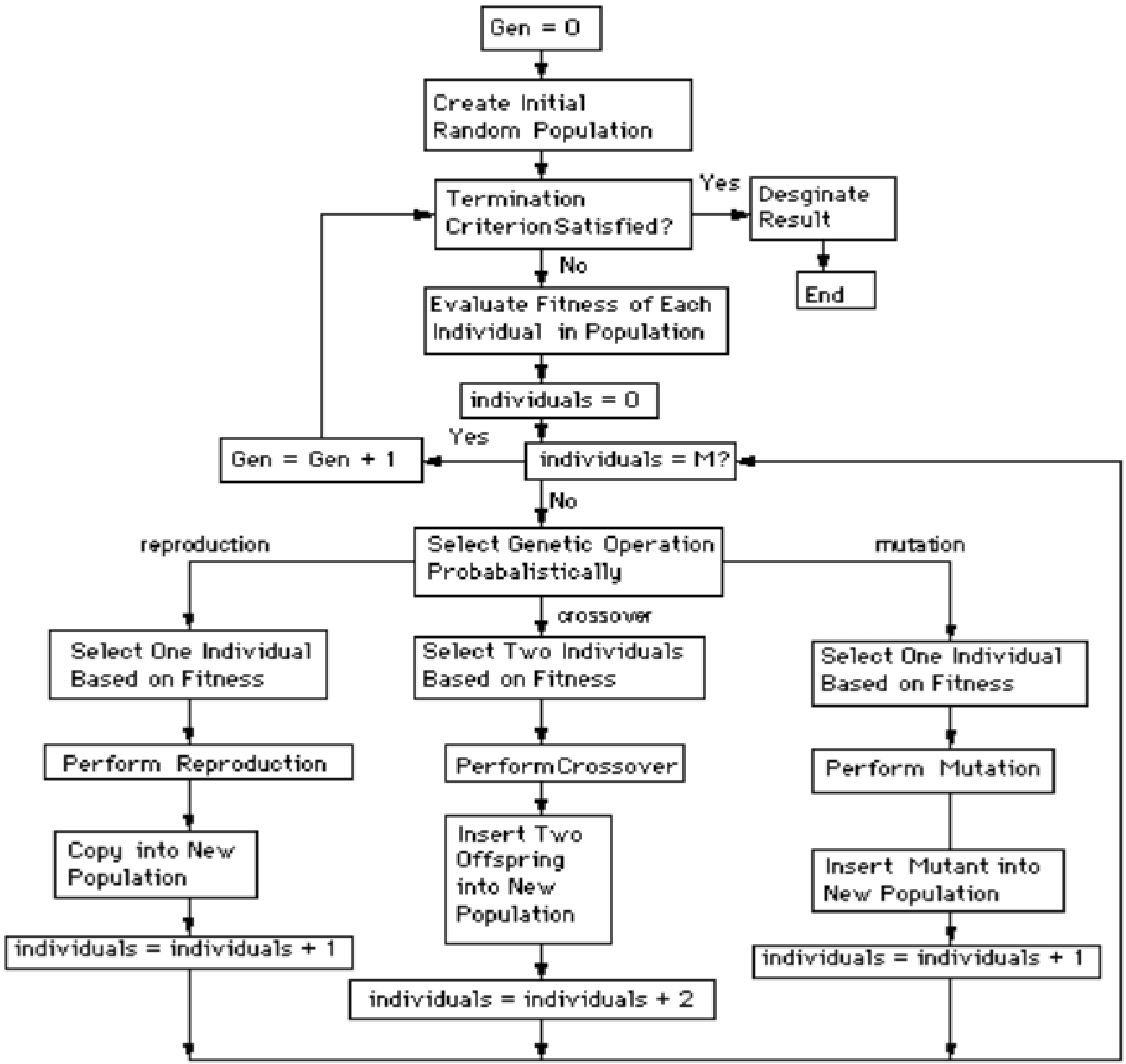
Mutation



Selection

- Parent selection typically fitness proportionate
- Over-selection in very large populations
  - rank population by fitness and divide it into two groups: group 1: best x% of population, group 2: other (100-x)%
  - 80% of selection operations chooses from group 1, 20% from group 2
  - motivation: to increase efficiency, %'s come from rule of thumb
- Survivor selection:
  - Typical: generational scheme (thus none)
  - Recently steady-state is becoming popular for its elitism

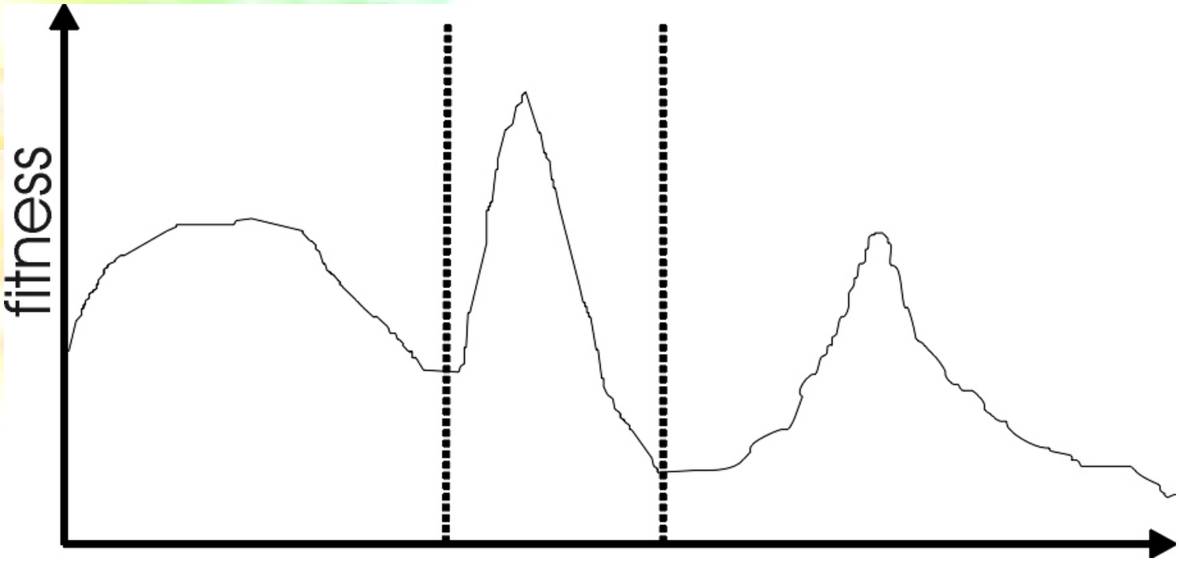
GP Flowchart

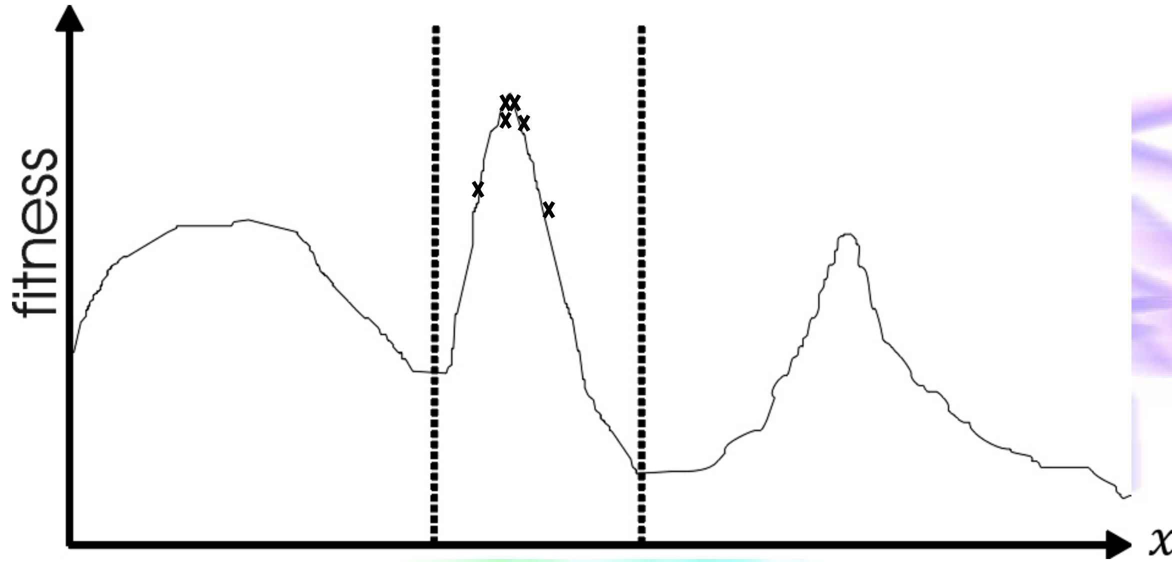


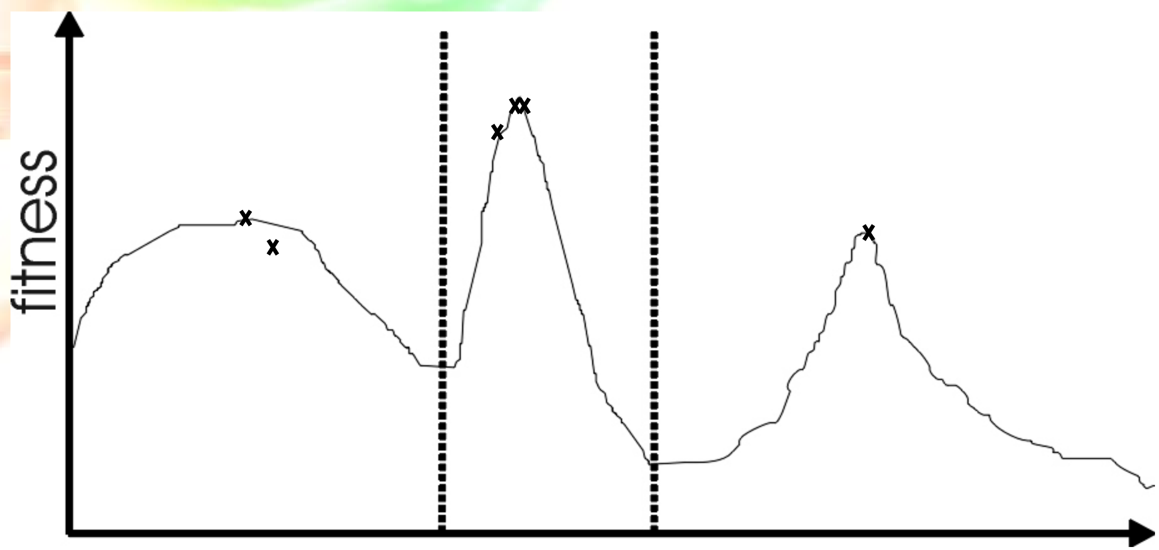
Multimodal Optimization

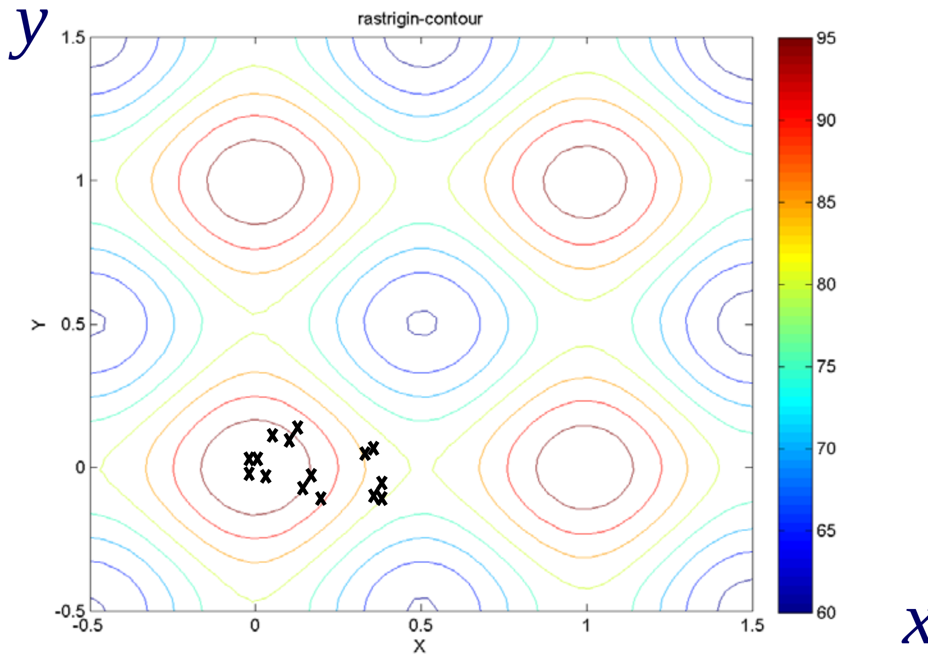
Multimodal Optimization Problems

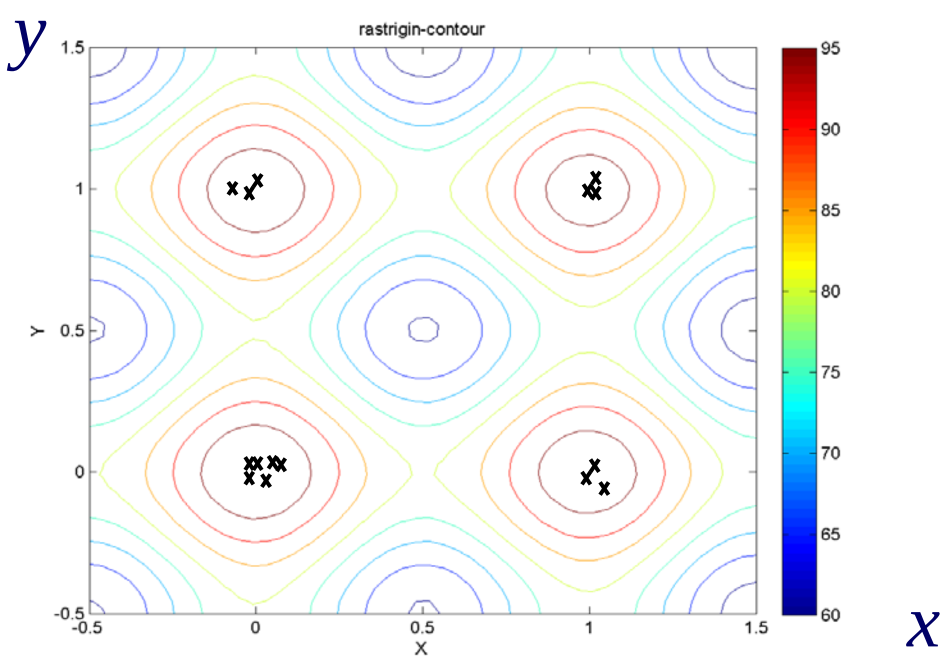
Most interesting problems have more than one <u>locally optimal solution</u>.




fitness
x


fitness


y
rastrigin-contour
Y
X
x


y
rastrigin-contour
Y
X
x

Methods

1. Repeat several runs

$$p\sum_{i=1}^{p}\frac{1}{i} \approx p(\gamma + \log p)$$

2. Parallel implementation of the iterative method
3. Investigate many peaks in parallel (Niching method)

Niching Method

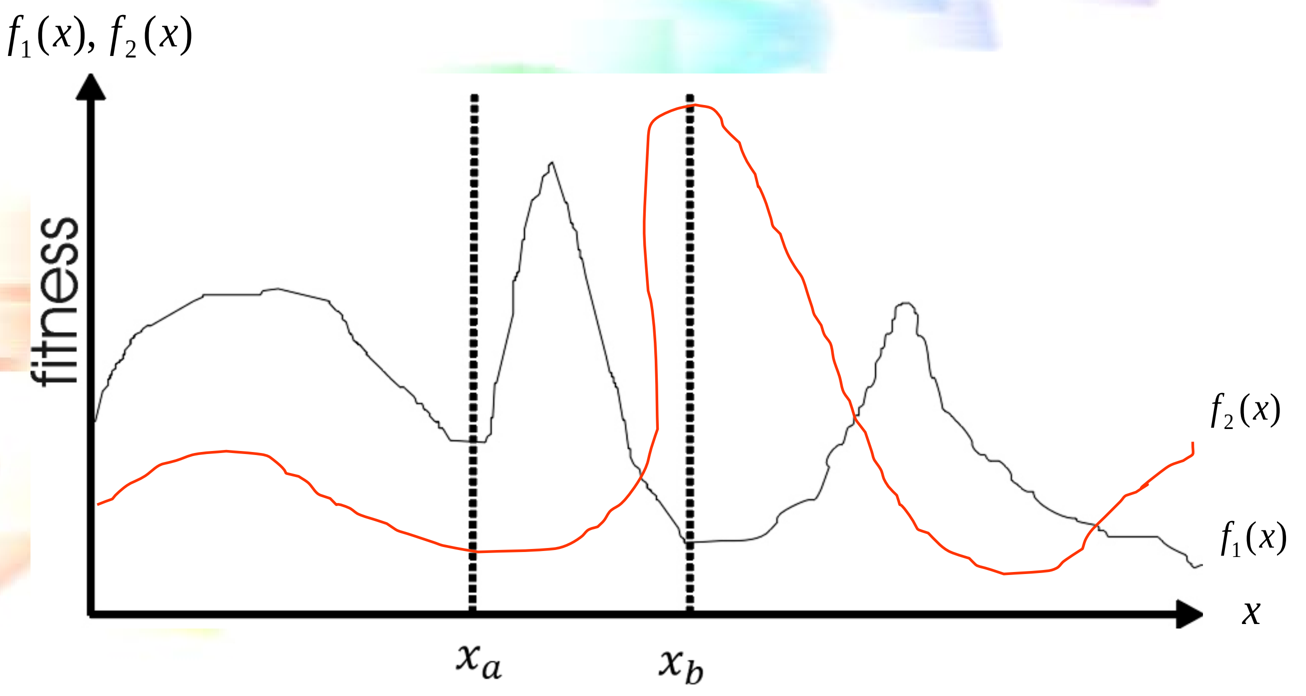
Investigate many peaks in parallel

- Sharing function: the degradation of an individual's fitness
  - $$sh(dist(x,y)) = \begin{cases} 1 - \left(\dfrac{dist(x,y)}{\sigma_{sh}} \right)^{\alpha} & \text{if} \quad dist < \sigma_{sh} \\ 0 & otherwise \end{cases}$$

    $where \quad \sigma_{sh}$ and α are constants
  - $0 \le sh(dist) \le 1$
  - $sh(0) = 1$
  - $sh(\lim_{dist \to \infty}) = 0$
- Niche count
  $$m(x) = \sum_y sh(dist(x,y))$$
- New shared fitness of an individual x
  $$eval'(x) = eval(x) / m(x)$$

Multi-objective Optimization

Multi-Objective Problems (MOPs)



Multi-Objective Problems (MOPs)

$$\min(f_1(x), f_2(x), f_3(x), \cdots, f_k(x)) \quad \text{s.t. } x \in X \quad \textit{where} \;\; k \geq 2$$

- Wide range of problems can be categorised by the presence of a number of *n* possibly <u>conflicting objectives</u>:
  - buying a car: speed vs. reliability…
  - Engineering design: lightness vs strength
- Two part problem:
  - choice of best for particular application
  - finding set of good solutions

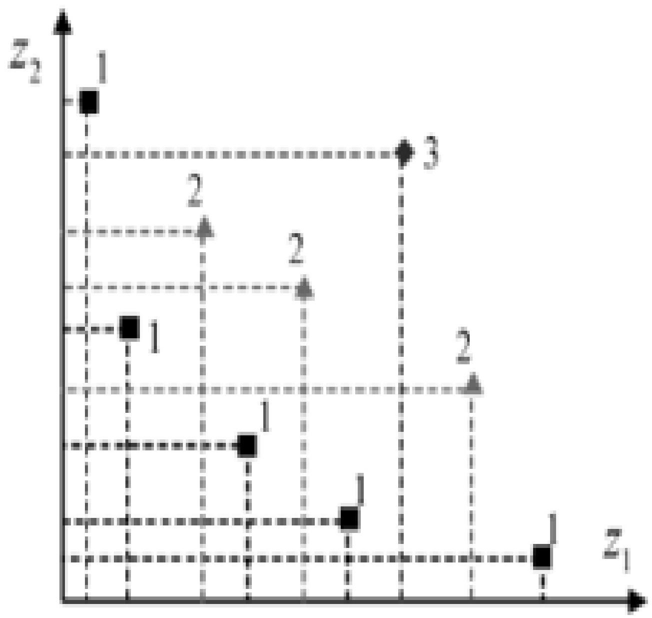
Methods

- Objective Weighting

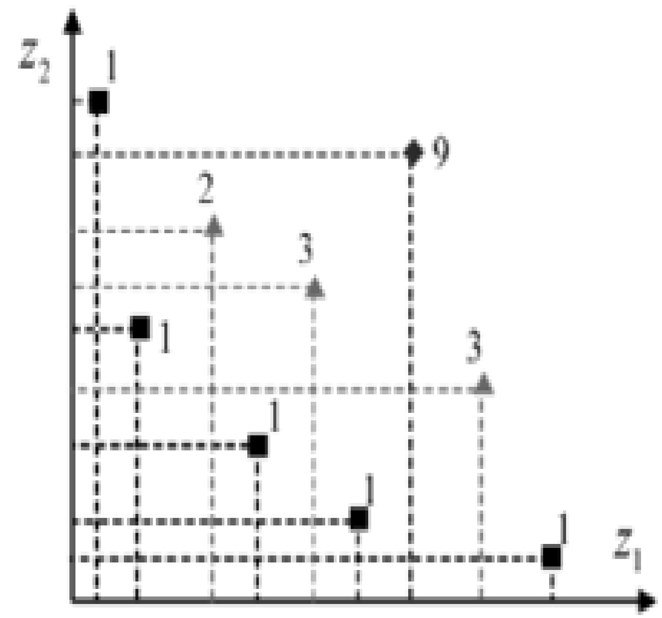
$$F(x) = \sum_{i=1}^{k} w_i f_i(x) \qquad \text{where } w_i \in [0..1], \quad \sum_{i=1}^{k} w_i = 1$$

- Pareto Ranking
  - Goldberg
  - Fonseca and Fleming

- Pareto Ranking
  - Goldberg
  - Fonseca and Fleming



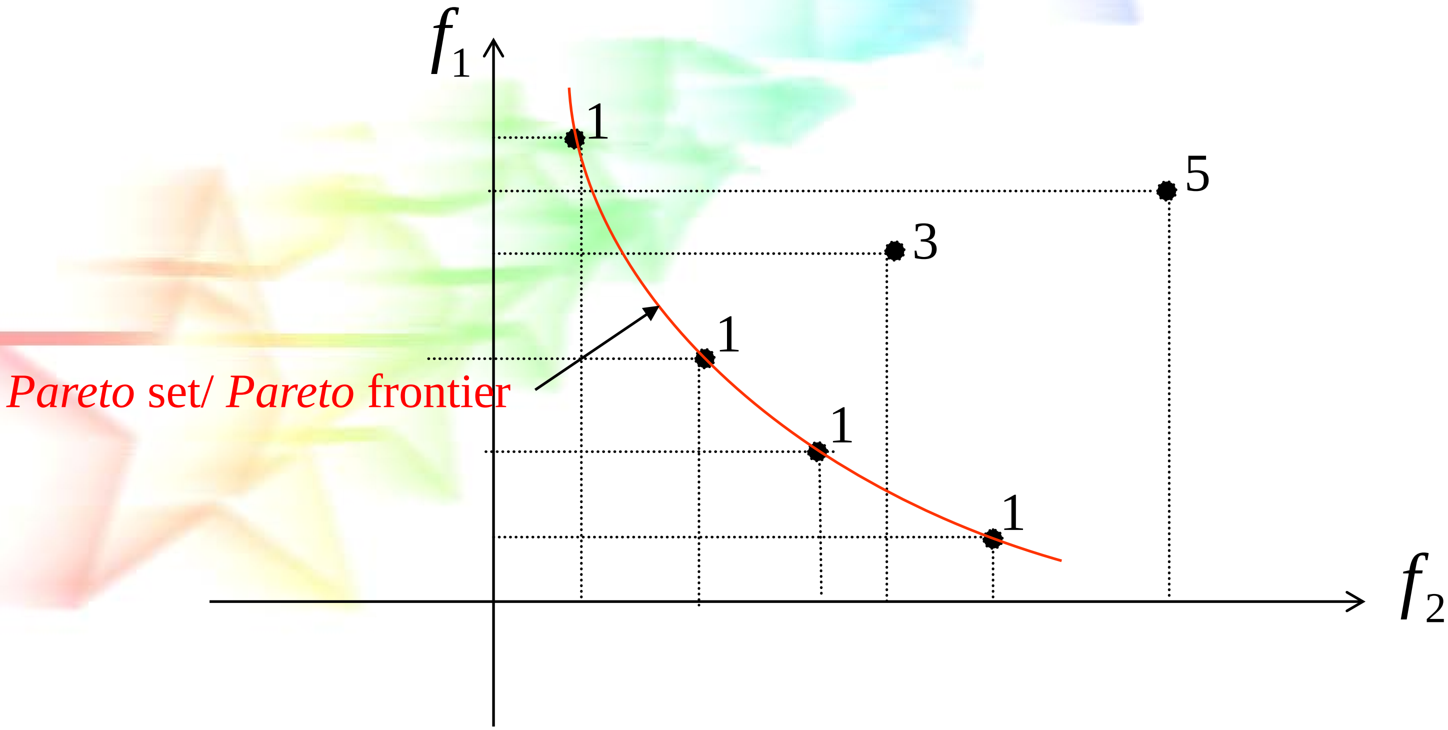
(Goldberg)



(Fonseca and Fleming)

Non-dominated/Pareto-optimal Solutions

- Pareto frontier
  - $\min(f_1(x), f_2(x))$

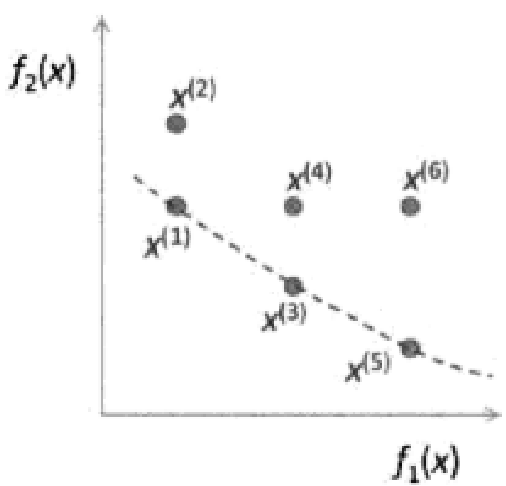


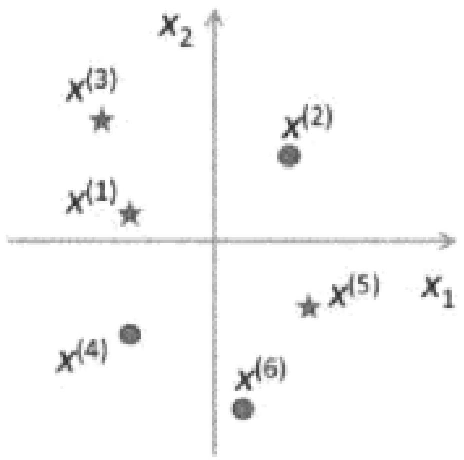
$rank(x) = 1 + p \qquad (p \text{ chromosomes dominate } x.)$

Example

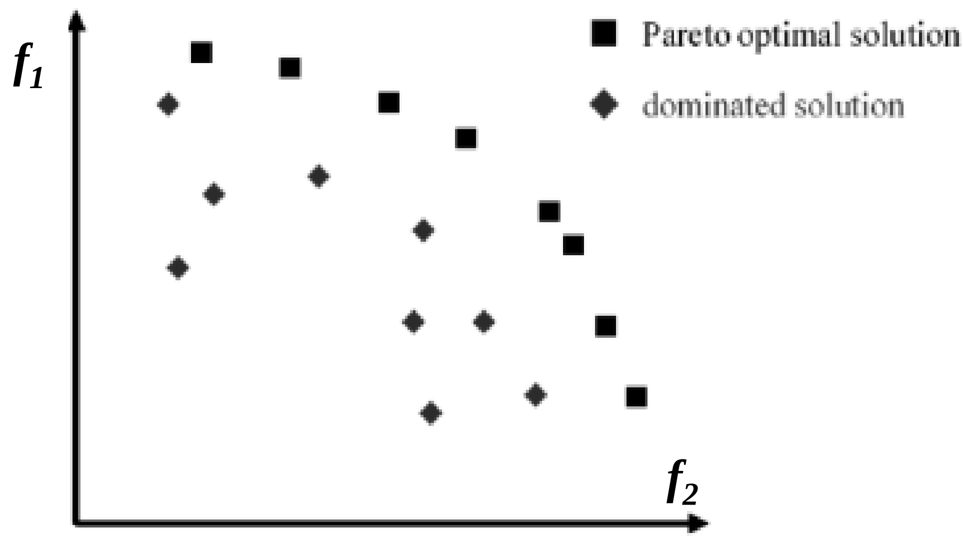
$$\min(f_1(x), f_2(x))$$

$$f_1\left(x^{(1)}\right) = 1 \qquad f_2\left(x^{(1)}\right) = 3$$
$$f_1\left(x^{(2)}\right) = 1 \qquad f_2\left(x^{(2)}\right) = 4$$
$$f_1\left(x^{(3)}\right) = 2 \qquad f_2\left(x^{(3)}\right) = 2$$
$$f_1\left(x^{(4)}\right) = 2 \qquad f_2\left(x^{(4)}\right) = 3$$
$$f_1\left(x^{(5)}\right) = 3 \qquad f_2\left(x^{(5)}\right) = 1$$
$$f_1\left(x^{(6)}\right) = 3 \qquad f_2\left(x^{(6)}\right) = 3$$

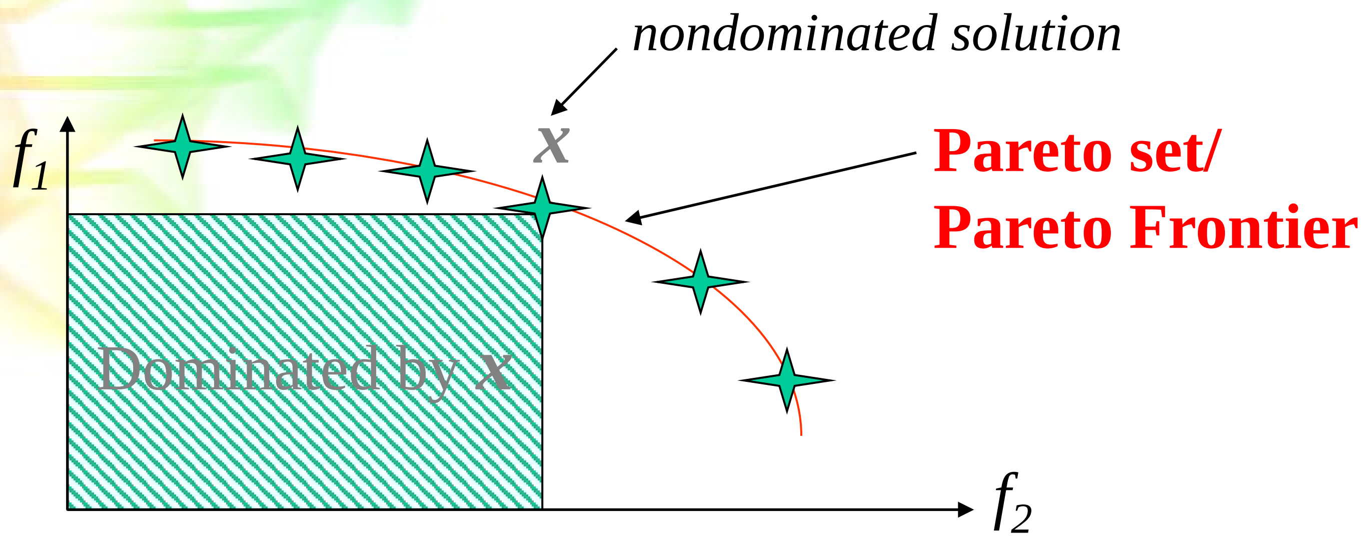




$$\max(f_1(x), f_2(x))$$



f_1
Pareto optimal solution
dominated solution
f_2



nondominated solution
x
f_1
Pareto set/
Pareto Frontier
Dominated by x
f_2