Artificial Intelligence

Evolutionary Algorithms

# Lesson 7: Evolutionary Strategies

**Vincenzo Piuri**

Università degli Studi di Milano

# Contents

# **Evolutionary Strategies** (1)

- Do not only optimizes organisms but the entire evolution process
  - Reproduction
  - Mortality rate
  - Life span
- Susceptibility to
  - Mutation
  - Mutation step width
  - Speed of evolution
  - Etc.

# **Evolutionary Strategies** (2)

Numerical optimization

- Given: function $f$: $\mathbb{R}^n \rightarrow \mathbb{R}$
- Wanted: minimum or maximum of $f$
- Chromosomes: vectors of real numbers

# **Evolutionary Strategies** (3)

- Mutation: adding a normal-distributed random vector $r$
  - Each element $r_i \in r$ is a sample of a normal-distributed random number
  - Expected value 0 (independent of element index $i$)
  - Standard deviation $\sigma_i$
- Crossover: not used

# Selection (1)

**Elite principle**

- Parents are allowed to evolve
- Selection is performed on offsprings
- Only the best individuals enter the next generation
- $\mu$ – number of individuals in the parent generation
- $\lambda$ – number of offspring individuals that were created by ES

# **Selection** (2)

Selection strategies

- **plus-strategy** (+strategy)
  - Selection works on $(\mu + \lambda)$ individuals
  - The best $\mu$ chromosomes are selected for the next generation
- **comma-strategy** (,strategy)
  - Generates offspring of size $\lambda > \mu$
  - The best $\mu$ chromosomes are selected
  - Chromosomes of parent generation are lost

# **Selection** (3)

- Advantage of the +strategy:
  - Due to strict elite principle: only improvements
- Disadvantages of the +strategy:
  - Risk of getting stuck in local optima
  - If no improvement over further generations, temporary helpful to switch off to the comma strategy to overcome local minima (increases diversity in population)

# Adaptation

Adaptation of random vector's variance
(mutation step width)

• Small variance ⇒ small changes of chromosomes
⇒ local search (exploitation)

• High variance ⇒ big changes of chromosomes
⇒ global search (exploration)

# Global Adaptation

- Choose $\sigma$ in such a way that the mean convergence rate is (approximately) optimized
  - Determining the probabilities for a successful mutation
  - A successful mutation improves the result
- **1/5 success rule**
  - Mutation step size is appropriate if approximately 1/5 of the offspring are better than the parents
  - Heuristically inferred under +strategy
  - If more than 1/5 of the children are better than the parents, the variance should be increased
  - If less than 1/5 of the children are better than the parents, the variance should be reduced
  - In large populations the 1/5 success rule is sometimes too optimistic: define a function that increases the threshold over time

# **Local Adaptation** (1)

- Variance (standard deviation) is used as additional genetic information
  - A variance for all vector entries
  - An individual variance for every entry of the vector (double vector length)
- Chromosomes with “bad” variances generate more “bad” offspring
  - Too small: chromosomes are growing too slowly
  - Too big: chromosomes clearly diverge from their respective parents
  - Genes (and so the variances) become extinct

# Local Adaptation (2)

- Element-specific mutation step widths: step width are mutated

$$\sigma_i' = \sigma_i \cdot \exp(r_1 \cdot N(0,1) + r_2 \cdot N_i(0,1)).$$

$N(0,1)$: one normal distributed number per chromosome

$N_i(0,1)$: one normal distributed number per gene

recommended values for $r_1, r_2$

$$r_1 = \frac{1}{\sqrt{2n}}, \qquad r_2 = \frac{1}{\sqrt{2\sqrt{n}}},$$

where $n$ is number of vector entries

$$r_1 = 0.1, \qquad r_2 = 0.2$$

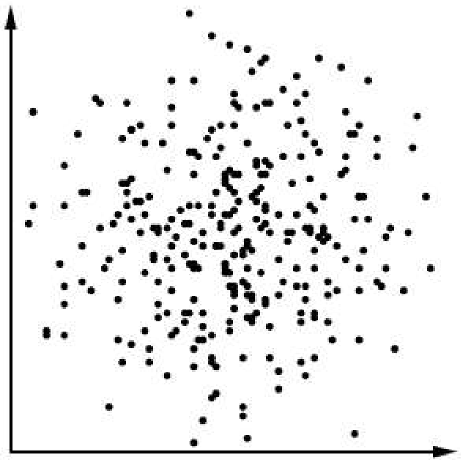
# **Local Adaptation** (3)

- Standard form of the local variance adaptation: variances of the different vector entries are independent of each other
  - Covariance matrix is a diagonal matrix
- **Correlated variance**
  - Variants of chromosomes favored in certain directions
  - Adaptation with single variance if and only if orthogonal directions
  - Use covariance matrix with a high covariance

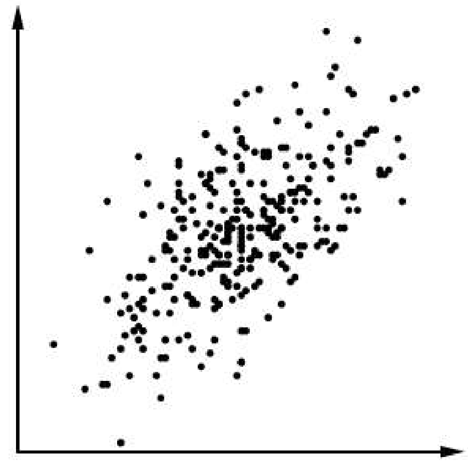
$$\Sigma = \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}$$

# Local Adaptation (4)

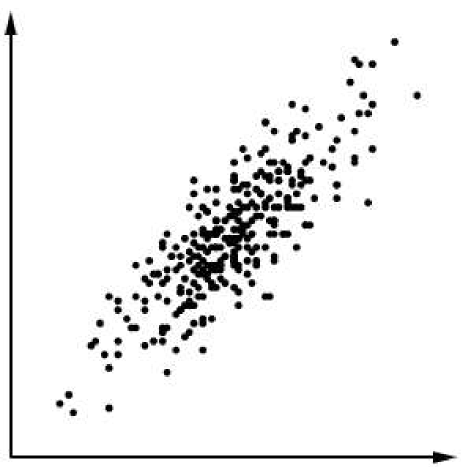
no correlation



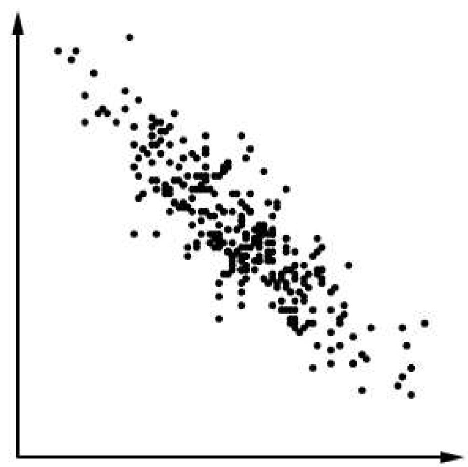
weak positive correlation



strong positive correlation



strong negative correlation



## **Local Adaptation** (5)

- Disadvantages of the correlated mutation
  - More parameters have to be adapted
  - Variances and rotation angles have no direct influence on the fitness function
  - Its adaptation is performed rather randomly

# Crossover/Recombination

- Uniform crossover
  - Random selection of components of the parents
- Averaging (blending, intermediary recombination)

$$\begin{array}{l}(x_1, \ldots, x_n) \\ (y_1, \ldots, y_n)\end{array} \quad \Rightarrow \quad \frac{1}{2}(x_1 + y_1, \ldots, x_n + y_n)$$

  - Disadvantage: total disappearance of any diversity in a population