Evolutionary Strategies

- Evolutionary Strategies (ES) are Evolutionary algorithms specifically developed for real-valued, semi-continuous, parameter optimization
- Key characteristic: ES use an advanced mutation operator which controls its own mutability → **self-adaptation**
- Genotype representation also includes a set of strategy parameters encoding the mutation probability distribution

ES representation

- Fitness function: $f(x_1, \ldots, x_n) : \Re^n \rightarrow \Re$
- Genotype representation of an individual solution:

$$(x_1, ..., x_n, \sigma_1^2, ..., \sigma_n^2, c_{12}, ..., c_{n-1,n})$$

  Parameters $(x_1, \ldots, x_n)$ need to be optimized
- Individual solution consists of 3 parts:
  1. $\vec{x}$: problem variables ⇒ Fitness $f(\vec{x})$
  2. $\vec{\sigma}$: standard deviations ⇒ variances
  3. $\vec{\alpha}$: rotation angles ⇒ covariances

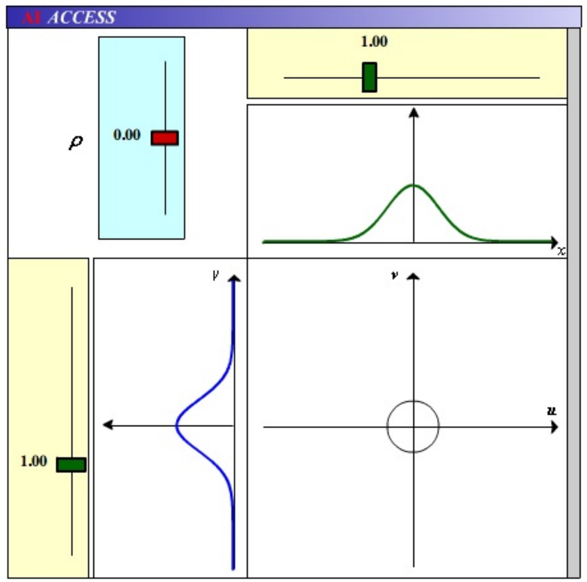
ES representation

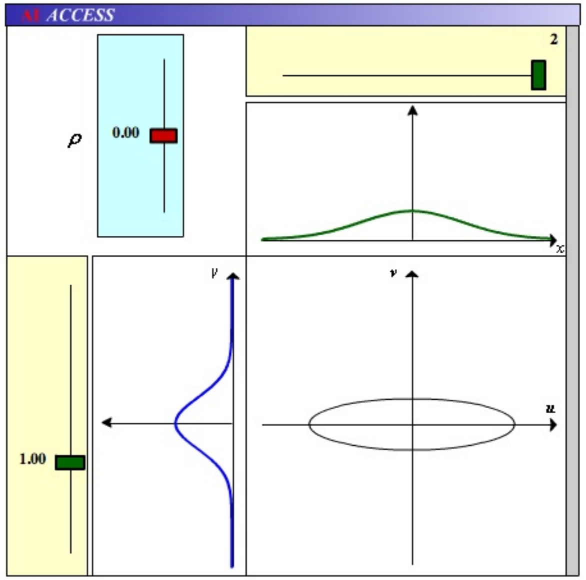
- The strategy parameter set $(\vec{\sigma}, \vec{\alpha})$ is part of the individual and represents the probability function for its mutation
- Strategy parameters $(\sigma_1^2, ..., \sigma_n^2, c_{12}, ..., c_{n-1,n})$ specify the n-dimensional normal distribution describing how X is mutated
- The n-dimensional normal probability density function:

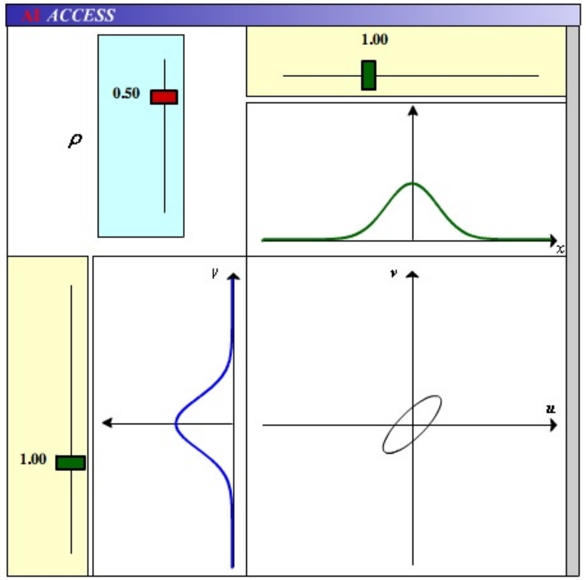
$$p(X = x_1, \ldots, x_n) = \frac{\exp(-\frac{1}{2} X^T \mathbf{C}^{-1} X)}{\sqrt{(2\pi)^n |\mathbf{C}|}}$$

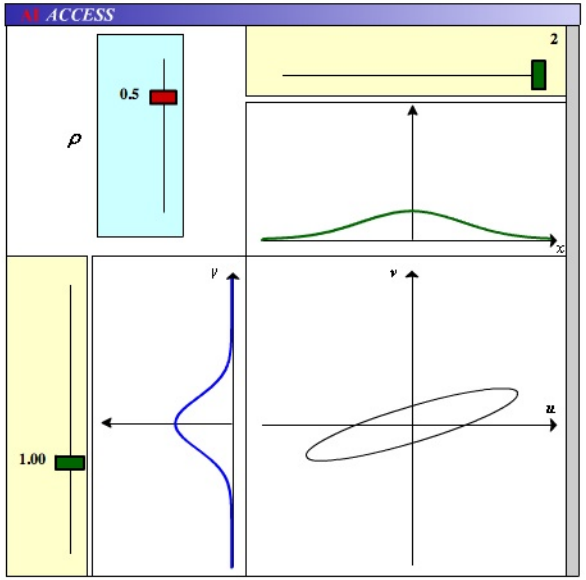
  C: correlation matrix (c_{ij}); $|\mathbf{C}|$ determinant
  ⇒ rotation angles $\alpha_{ij} : \tan 2\alpha_{ij} = 2c_{ij}/(\sigma_i^2 - \sigma_j^2)$
- cfr. 1-dimensional Gaussian function:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp^{\frac{(x-\mu)^2}{2\sigma^2}}$$









ES representation

- Amount of strategy parameters decided by the user: global search reliability and robustness increases at the cost of computing time when number of strategy parameters increases
- Commonly used settings:
  1. only single standard deviation controlling the mutation of all problem parameters x_i (no correlated mutations):

     $$\sigma_1 = \ldots = \sigma_n; \quad c_{ij} = 0 \; (i \neq j)$$

  2. individual standard deviations controlling the mutation of all problem parameters x_i (no correlated mutations)::

     $$\sigma_1, \ldots, \sigma_n; \quad c_{ij} = 0 \; (i \neq j)$$

  3. complete covariance matrix: $\sigma_1, \ldots, \sigma_n; \quad c_{ij} \neq 0 \; (i \neq j)$

ES mutation I

1. Case 1: one single standard deviation controls the mutation of all problem parameters x_i (no correlated mutations):

$$\sigma = \sigma_1 = \ldots \sigma_n; \quad c_{ij} = 0 \ (i \neq j)$$

2. First, the strategy parameters are mutated. $\mathbb{N}(0,1)$ = a normally distributed random number (mean = 0, variance = 1):

$$\sigma' = \sigma e^{\frac{\mathbb{N}(0,1)}{\sqrt{n}}}$$

lower limit ϵ : if $\sigma' < \epsilon \Rightarrow \sigma' := \epsilon$

3. Second, problem parameters are mutated with the new strategy parameter:

$$x_i' = x_i + \sigma' \mathbb{N}_i(0,1)$$

ES mutation II

1. Case 2: individual standard deviations controlling the mutation of all problem parameters x_i (no correlated mutations)::

$$\sigma_1, \ldots, \sigma_n; \quad c_{ij} = 0 \ (i \neq j)$$

2. First, the strategy parameters are mutated:

$$\sigma_i' = \sigma_i e^{\frac{\mathbb{N}(0,1)}{\sqrt{2n}} + \frac{\mathbb{N}_i(0,1)}{\sqrt{2\sqrt{n}}}}$$

lower limit ϵ : if $\sigma_i' < \epsilon \Rightarrow \sigma_i' := \epsilon$

3. Second, problem parameters are mutated with new strategy parameters:

$$x_i' = x_i + \sigma_i' \mathbb{N}_i(0,1)$$

ES mutation III

1. case 3: complete covariance matrix: $\sigma_1, \ldots, \sigma_n; \quad c_{ij} \neq 0 \ (i \neq j)$
2. First, the strategy parameters are mutated:

$$\sigma_i' = \sigma_i e^{\frac{\mathbb{N}(0,1)}{\sqrt{2n}} + \frac{\mathbb{N}_i(0,1)}{\sqrt{2\sqrt{n}}}}$$

$$\alpha_j' = \alpha_j + \beta \mathbb{N}_j(0,1)$$

$\beta \approx 0.0873$ (5^o in radians), $\mathbb{N}(0,1)$: standard normal distribution

3. Second, problem parameters are mutated with new strategy parameters:

$$\vec{x}' = \vec{x} + \vec{N}(\vec{0}, \vec{\sigma}', \vec{\alpha}')$$

$\vec{N}$: n-dimensional normal distribution

ES recombination

- Creates one offspring from several parents that are selected at random from the parent population
- Problem parameters and strategy parameters are differently recombined:
  1. *problem parameters:* select at random 2 parents of the μ parents for each parameter x_i and take their average

$$x_i^{offspring} = \frac{1}{2}\,(x_i^{parent_1^i} + x_i^{parent_2^i})$$

  2. *standard deviations:* select at random 2 parents of the μ parents and take at random one of the two parent values

$$\sigma_i^{offspring} = \sigma_i^{parent_1} \quad \text{or} \quad \sigma_i^{parent_2}$$

  3. *rotation angles:* not recombined

ES selection

- ES applies a high selection pressure: from μ parents λ offspring are generated with $\lambda >> \mu$ (typically, $\lambda \approx 5$ to 10 times μ)
- Common 'standard' values: $\mu = 15, \lambda = 100$
- The best μ solutions of the λ offspring are selected for the next generation - this is, (μ, λ)-selection - or, the best μ solutions of the μ parents and the λ offspring are selected for the next generation - this is, $(\mu + \lambda)$-selection
- Experimental results: self-adaptation works better with (μ, λ) selection

Self-adaptation: necessary conditions

Necessary conditions found by experiments to let self-adaptation work well:

- Generation of a surplus offspring: $\lambda > \mu$
- (μ, λ)-selection to guarantee extinction of misadapted individuals (as opposed to $(\mu + \lambda)$
- Intermediate selective pressure, eg. $(\mu, \lambda) = (15, 100)$
- Multiple parents necessary: $\mu > 1$
- Recombination also applied on strategy parameters (more specifically the use of intermediate recombination)

Self-adaptation is collective learning

- Test function: $F(\vec{x}) = \sum_{i=1}^{30} i x_i^2$
- Optimal scaling of variances can formally be derived: $\sigma_i \propto \frac{1}{\sqrt{i}}$
- Test function is unimodal $\Rightarrow$ when using optimal variances a single parent solution gives best performance, ie. $\mu = 1$.
- Experimental results for $(\mu, 100) - ES$ with $\mu \in \{1, \ldots, 30\}$ show that using self-adaptation (and without knowledge of optimal variances) a parent population of size $\mu = 12$ performs (nearly) as good as the optimal strategy
- Self-adaptation: imperfect, diverse set of parents cooperate by exchanging information (= recombination) about their "internal models" (ie. their strategy parameters)