## Evolution strategies

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



## The famous jet nozzle experiment Rechenberg-Schwefel, 1964

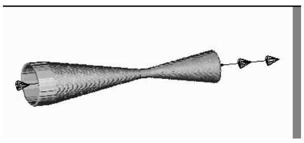
Task: to optimize the shape of a jet nozzle
Approach: random mutations to shape + selection

Initial shape

Final shape



## The famous jet nozzle experiment (movie)





## Main characteristics at a glance

- Often continuous search spaces, $\mathbb{R}^n$
- Emphasis on mutation: *n*-dimensionally normally distributed, expectation zero
- Various recombination operators
- Deterministic ($\mu$, $\lambda$)-selection
- *self-adaptation* of strategy parameters: first self-adaptive EA
- Generation of an offspring surplus $\lambda >> \mu$



## Reproduction cycle (1)

Evolution Strategy main procedure:

```
t := 0;
initialize P(t);
evaluate P(t);
WHILE NOT termination DO
     P'(t)  := recombine(P(t));
     P''(t) := mutate(P'(t));
     evaluate(P''(t));
     P(t + 1)  := select(P''(t) ∪ P(t));
//   P(t + 1)  := select(P''(t));
     t := t + 1;
OD
```

Either of these two is used



## Reproduction cycle (2)

- P(t) size $\mu$
- *recombination*: applied to all individuals
- P'(t) has size $\lambda > \mu$
- *mutation*: normally distributed variations, all individuals
- P''(t) has size $\lambda > \mu$,
- *selection*
  - ($\mu + \lambda$ ) selection: from P''(t) $\cup$ P(t)
  - ($\mu$, $\lambda$) selection: from P''(t) only

## Representation (1)

Spaces:

– Phenotype space:

$$\mathbb{R}^n$$

– Strategy parameter space (standard deviations and rotation angles of mutation):

$$S = IR_+^{n_\sigma} \cdot [-\pi, \pi]^{n_\alpha}$$

– Individual space (genotype):

$$I = IR^n \cdot S$$

One individual:

$$\boxed{\vec{a} = (\underbrace{(x_1,\dots,x_n)}_{\vec{x}}, \underbrace{(\sigma_1,\dots,\sigma_{n_\sigma})}_{\vec{\sigma}}, \underbrace{(\alpha_1,\dots,\alpha_{n_\alpha})}_{\vec{\alpha}}) \in I}$$



## Representation (2)

The three parts of an individual:

$\vec{x}$ : Object variables $\Rightarrow$ Fitness $f(\vec{x})$

$\vec{\sigma}$ : Standard deviations $\Rightarrow$ Variances

$\vec{\alpha}$ : Rotation angles $\Rightarrow$ Covariances

A strategy parameter sets $(s = (\vec{\sigma}, \vec{\alpha}) \in S)$ :

– Is part of an individual

– Represents the **p**robability **d**ensity **f**unction (**p.d.f.**) for its mutation

| $n_\sigma$ | $n_\alpha$ | Remark |
|---|---|---|
| 1 | 0 | standard mutation |
| $n$ | 0 | standard mutations |
| $n$ | $n \cdot (n-1)/2$ | correlated mutations |
| $1 \le n_\sigma \le n$ | $(n - \frac{n_\sigma}{2})(n_\sigma - 1)$ | general case (correlated mutations) |

Possible settings of $n_\sigma$ and $n_\alpha$



## Genetic operators: mutations (1)

Simple mutation I

– Simple mutation makes use of normally distributed variations, thus

– It requires a given a normal (Gaussian) distribution N(ξ, σ) with the corresponding pdf

$$p(\Delta x_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(\Delta x_i - \xi)^2}{2\sigma^2}\right)$$

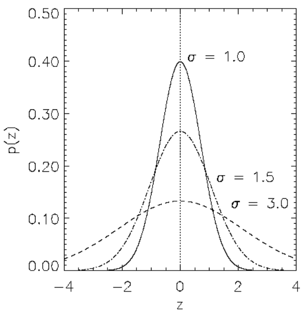
– Expectation (ξ) is assumed to equal 0

– Standard deviation (σ) must be adapted



## Genetic operators: mutations (2)

The one dimensional case






## Genetic operators: mutations (3)

Simple mutation I (continued)

– $x_i$ is mutated by adding some $\Delta x_i$ from a normal probability distribution

– σ is mutated by multiplying by $e^\Gamma$, with Γ from a normal probability distribution

$$I = IR^n \cdot IR_+$$

$$m'_{\{\tau_0\}}(\bar{x}, \sigma) = (\bar{x}', \sigma')$$

$$\tau_0 \sim 1/\sqrt{n}$$

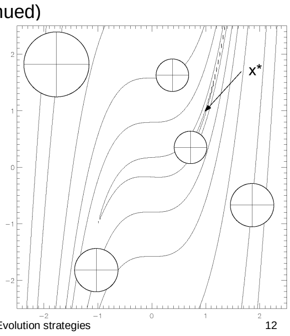
$$\boxed{\begin{aligned} \sigma' &= \sigma \cdot \exp(\tau_0 \cdot N(0,1)) \\ x'_i &= x_i + \sigma' \cdot N_i(0,1) \end{aligned}}$$



## Genetic operators: mutations (4)

Simple mutation I (continued)

Equal probability to place an offspring




Simple mutations, n = 2, $n_\sigma$ = 1, ($n_\alpha$ = 0)

## Genetic operators: mutations (5)

Simple mutation II

- $x_i$ is mutated by adding some $\Delta x_i$ from a normal probability distribution
- $\sigma_j$ is mutated by multiplying by $e^{\Gamma_j}$, with $\Gamma_j$ from a normal probability distribution

$$\begin{aligned} I &= IR^n \cdot IR^n_+ \\ m'_{\{\tau,\tau'\}}(\bar{x},\bar{\sigma}) &= (\bar{x}',\bar{\sigma}') \\ \tau &\sim 1/\sqrt{2\sqrt{n}} \\ \tau' &\sim 1/\sqrt{2n} \end{aligned}$$



## Genetic operators: mutations (6)

Simple mutation II (continued)

$$\begin{aligned} \sigma_i' &= \sigma_i \cdot \exp(\tau' \cdot N(0,1) + \tau \cdot N_i(0,1)) \\ x_i' &= x_i + \sigma_i' \cdot N_i(0,1) \end{aligned}$$

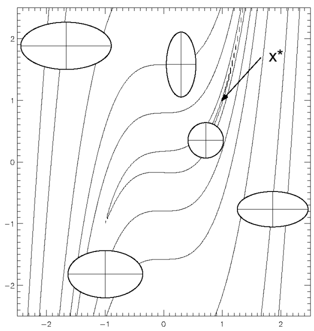
Boundary rule for preserving standard deviations larger than zero: $\sigma_i' < \varepsilon_\sigma \Rightarrow \sigma_i' := \varepsilon_\sigma$



## Genetic operators: mutations (7)

Simple mutation II (continued)

Equal probability to place an offspring




Simple mutations, n = 2, $n_\sigma$ = 2, ($n_\alpha$ = 0)



## Genetic operators: mutations (8)

Correlated mutation

- Correlated mutation uses following probability distribution function for $\Delta x$:

$$p(\overline{\Delta x}) = \sqrt{\frac{\det C}{(2\pi)^n}} \cdot \exp\left(-\tfrac{1}{2}\overline{\Delta x}^T \cdot C\overline{\Delta x}\right)$$

- Where $C^{-1}$ is the covariance matrix:

$$\begin{aligned} c_{ii} &= \sigma_i^2 \\ c_{ij,(i\neq j)} &= \begin{cases} 0 & \text{no correlations} \\ \frac{1}{2}\left(\sigma_i^2 - \sigma_j^2\right)\tan(2\alpha_{ij}) & \text{correlations} \end{cases} \end{aligned}$$



## Genetic operators: mutations (9)

Correlated mutation (continued)

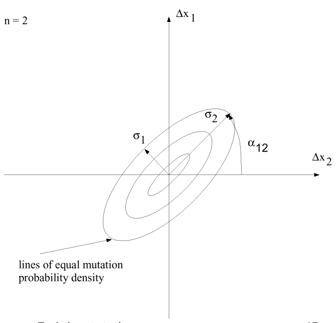



Illustration of the mutation ellipsoid for the case n = 2, $n_\sigma$ = 2, $n_\alpha$ = 1



## Genetic operators: mutations (10)

Correlated mutation (continued)

- $\bar{x}$ is mutated by adding some $\overline{\Delta x}$ from an n-dimensional normal distribution
- $\sigma_i$ is mutated by multiplying by $e^{\Gamma_i}$, with $\Gamma_i$ from a normal probability distribution
- $a_j$ is mutated by adding some $\Delta a_j$ from a normal probability distribution

## Genetic operators: mutations (11)

Correlated mutation (continued)

$$
\begin{aligned}
n_\alpha &= n \cdot (n-1)/2 \\
I &= IR^n \cdot IR_+^n \cdot [-\pi, \pi]^{n_\alpha} \\
m'_{\{\tau,\tau',\beta\}}(\bar{x},\bar{\sigma},\bar{\alpha}) &= (\bar{x}',\bar{\sigma}',\bar{\alpha}') \\
\tau &\sim 1/\sqrt{2\sqrt{n}} \\
\tau' &\sim 1/\sqrt{2n} \\
\beta &\approx 5^\circ
\end{aligned}
$$

Evolutionary Computing Evolution strategies 19

## Genetic operators: mutations (12)

Correlated mutation (continued)

$$
\begin{aligned}
\sigma'_i &= \sigma_i \cdot \exp(\tau' \cdot N(0,1) + \tau \cdot N_i(0,1)) \\
\alpha'_j &= \alpha_j + \beta \cdot N_j(0,1) \\
\vec{x}' &= \vec{x} + \vec{N}(\vec{0}, C')
\end{aligned}
$$

Boundary rule for keeping rotation angles feasible:

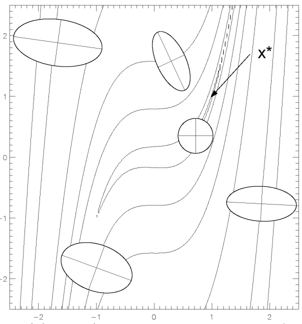
$$\left|\alpha'_j\right| > \pi \;\Rightarrow\; \alpha'_j := \alpha'_j - 2\pi \cdot sign(\alpha'_j)$$

Evolutionary Computing Evolution strategies 20

## Genetic operators: mutations (13)

Correlated mutation (continued)

Equal probability to place an offspring




Correlated mutations, n = 2, $n_\sigma$ = 2, $n_\alpha$ = 1)

Evolutionary Computing Evolution strategies 21

## Genetic operators: mutations (14)

Some remarks:

- Biological model: Repair enzymes, mutator genes
- No deterministic control: strategy parameters evolve
- Indirect link between fitness and useful strategy parameter settings
- $\bar{\sigma}, \bar{\alpha}$ are conceivable as an internal model of the local topology
- Standard strategy: $n_\sigma = n, n_\alpha = 0$

Evolutionary Computing Evolution strategies 22

## Genetic operators: recombination (1)

Basic ideas:

- $I^\mu \rightarrow I$, μ parents yield 1 offspring
- Is applied λ times, typically λ >> μ
- Is applied to object variables as well as strategy parameters
- Per offspring gene two corresponding parent genes are involved
- Two ways to recombine two parent alleles:
  - Discrete recombination: choose one randomly
  - Intermediate recombination: average the values
- Might involve two or μ parents (global recombination)

Evolutionary Computing Evolution strategies 23

## Genetic operators: recombination (2)

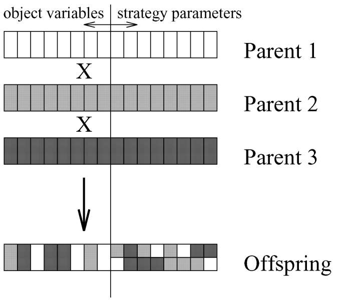
The operator:

1. For each object variable:
   a. Choose two parents
   b. Apply discrete recombination on the corresponding variables
2. For each strategy parameter:
   a. Choose two parents
   b. Apply intermediate recombination on the corresponding parameters

Evolutionary Computing Evolution strategies 24

## Genetic operators: recombination (3)

Recombination illustrated




$\mu = 3$



## Selection (1)

- Strictly deterministic, rank based
- The $\mu$ best ranks are handled equally
- The $\mu$ best offspring (P”(t)) survive
  - Important for self-adaptation
  - Applicable also for noisy objective functions, moving optima
- N.B. $\mu$ selected from $\lambda$; notation: $(\mu, \lambda)$
- Selective pressure: very high



## Selection (2)

Take-over time $\tau^*$:

Definition:
Number of generations until application of selection completely fills the population with copies of the best individual



## Selection (3)

Remarks:

- Goldberg and Deb 1991:

$$\tau^* = \frac{\ln \lambda}{\ln(\lambda / \mu)}$$

- $\tau^* \approx 2$ generations for a (15, 100)-ES
  (15 and 100 are typical values for the standard ES)
- Proportional selection in GA's:
  $\tau^* \approx \lambda \ln\lambda = 460$ generations!



## Other components

- Initialisation:
  - $x_i$, $\alpha_i$: randomly
  - $\sigma_i$: $\delta x_i / \sqrt{n}$, with $\delta x_i$ a very rough measure for the distance to the optimum
- Termination:
  - Termination after a number of generations
  - Or iff $\max\{f(\bar{x}_i(t))\} - \min\{f(\bar{x}_i(t))\} \le c(P(t))$
    - c(P(t)) absolute (= $\varepsilon_1 > 0$), or
    - c(P(t)) relative (= $\varepsilon_2 \cdot |\bar{f}|$)
- Constraints:
  - Handled by repeating creation and evaluation of individuals