Computational Intelligence
Master in Artificial Intelligence
2019-20

Introduction to Evolutionary Computation (III)
Evolution Strategies

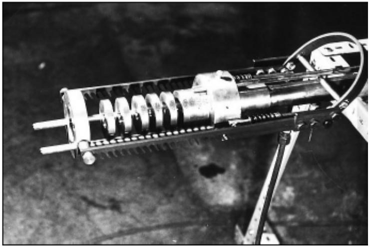
Soft Computing Research Group

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
School of Professional & Executive Development

Nozzle experiment (I)

collection of conical nozzle parts

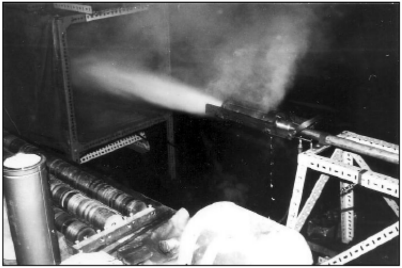
device for clamping nozzle parts



Nozzle experiment (II)

Hans-Paul Schwefel
while changing nozzle parts

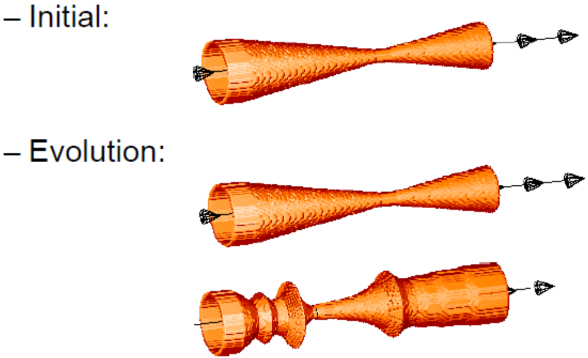
Nozzle experiment (III)



the nozzle in operation …

… while measuring degree of efficiency

Nozzle experiment (IV)



32% of increase in efficiency!

J. Klockgether and H.-P. Schwefel, "Two-phase nozzle and hollow core jet experiments". Proceedings of the 11th Symposium on Engineering Aspects of Magneto-Hydrodynamics, Caltech, Pasadena, California, USA, 1970.

A continuous d-variate random vector $\mathbf{X} = (X_1, \ldots, X_d)^{\mathrm{T}}$ is **normally distributed**, written $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$, when its joint pdf is:

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^{\mathrm{T}} \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\}$$

where $\boldsymbol{\mu}$ is the *mean vector* and $\Sigma_{d \times d} = (\sigma_{ij}^2)$ is the (real symmetric and p.d.) *covariance matrix*.

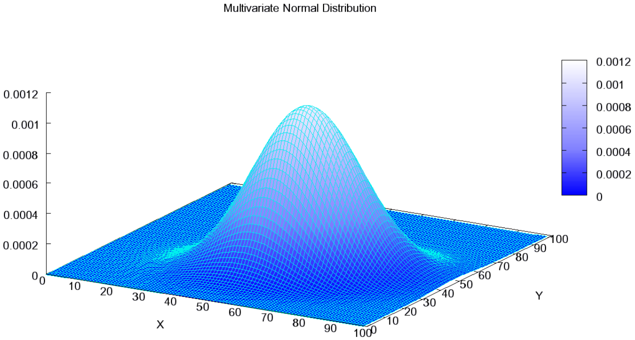
- $\mathbb{E}[\mathbf{X}] = \boldsymbol{\mu}$ and $\mathbb{E}[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^{\mathrm{T}}] = \Sigma.$
- $CoVar[X_i, X_j] = \sigma_{ij}^2$ and $Var[X_i] = \sigma_{ii}^2$

---

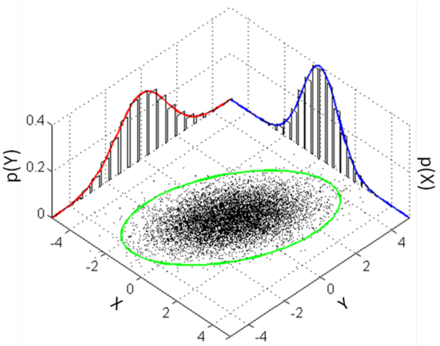
if $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$, then X_i, X_j are independent $\iff$ $CoVar[X_i, X_j] = 0$

(in general, only the left-to-right implication holds)

The Gaussian Distribution ($d = 2$)

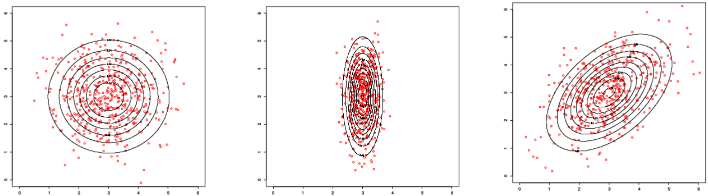


The Gaussian Distribution ($d = 2$)



Observations from a bivariate normal distribution, a contour ellipsoid, the two marginal distributions, and their histograms (images from the Wikipedia)

The Gaussian Distribution ($d = 2$)



$$\boldsymbol{\mu} = \begin{bmatrix} 3 \\ 3 \end{bmatrix} \quad \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \boldsymbol{\Sigma} = \begin{bmatrix} 0.1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$$

- The principal directions (a.k.a. PCs) of the hyperellipsoids are given by the *eigenvectors* $\mathbf{u}_i$ of Σ, which satisfy $\Sigma \mathbf{u}_i = \lambda_i \mathbf{u}_i$.
- The lengths of the hyperellipsoids along these axes are proportional to $\sqrt{\lambda_i}$ (note $\lambda_i > 0$)

Conceptual view

- What is behind the choice of a **multivariate Gaussian**?

  Examples from a class are noisy versions of an ideal class member (a *prototype*):

  - Prototype: modeled by the mean vector
  - Noise: modeled by the covariance matrix

- The quantity

$$d(\mathbf{x}) := \sqrt{(\mathbf{x} - \boldsymbol{\mu})^{\mathrm{T}} \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})}$$

  is called the **Mahalanobis distance** for $\mathbf{x}$

- Very important! the number of parameters is $\frac{d(d+1)}{2} + d$

Positive definiteness

For a Gaussian distribution to be well-defined, Σ has to be real symmetric and positive definite (p.d.): for all non-null vectors $\mathbf{x} \in \mathbb{R}^d$, $\mathbf{x}^{\mathrm{T}}\Sigma\mathbf{x} > 0$ must hold true

Examples: are these matrices p.d.?

$$a.\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad b.\begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix}$$

$$c.\begin{pmatrix} 3 & -1 \\ -1 & 2 \end{pmatrix} \qquad d.\begin{pmatrix} 1 & 4 \\ \frac{1}{2} & 1 \end{pmatrix}$$

a. YES; b. YES
c. YES; d. NO

Mathematical view

Positive definiteness

For a Gaussian distribution to be well-defined, Σ has to be real symmetric and positive definite (p.d.): for all non-null vectors $\mathbf{x} \in \mathbb{R}^d$, $\mathbf{x}^{\mathrm{T}}\Sigma\mathbf{x} > 0$ must hold true

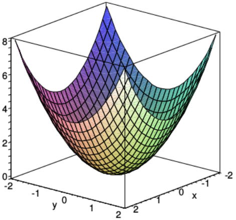
Examples: are these matrices p.d.?

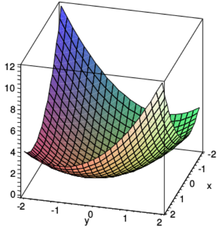
$$a. \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad b. \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix}$$

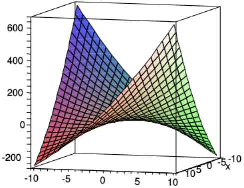
$$c. \begin{pmatrix} 3 & -1 \\ -1 & 2 \end{pmatrix} \qquad d. \begin{pmatrix} 1 & 4 \\ \frac{1}{2} & 1 \end{pmatrix}$$

a. YES; b. YES
c. YES; d. NO

Mathematical view







$a.\ z_1^2 + z_2^2;$ $\qquad b.\ z_1^2 + z_1 z_2 + z_2^2;$ $\qquad d.\ z_1^2 + \frac{9}{2} z_1 z_2 + z_2^2$

Evolution Strategies: main characteristics

- Continuous search space $\mathbb{R}^n$
- Various recombination operators
- Deterministic (μ,λ)-replacement
- Emphasis on mutation: n-dimensional Gaussian, zero expectation
- Self-adaptation of mutation parameters (first self-adaptive EA!)
- Generation of an offspring surplus $\lambda \gg \mu$

Evolution Strategies: Representation

1. **objective** space: $\mathcal{O} := \mathbb{R}^n$
2. **strategy** space (stdevs and angles of mutation):

$$\mathcal{P} := \mathbb{R}_+^{n_\sigma} \times (-\pi, \pi]^{n_\alpha}$$

The three parts of an individual

1. object variables $\mathbf{x} \in \mathbb{R}^n$ to compute fitness $F(\mathbf{x})$
2. standard deviations $\boldsymbol{\sigma} \in \mathbb{R}_+^{n_\sigma}$ to express variances
3. rotation angles $\boldsymbol{\alpha} \in (-\pi, \pi]^{n_\alpha}$ to express covariances

Evolution Strategies: Mutation (I)

Simple Self-Adaptive Mutation

$n_\sigma = 1, n_\alpha = 0$ (one mutation parameter per individual)

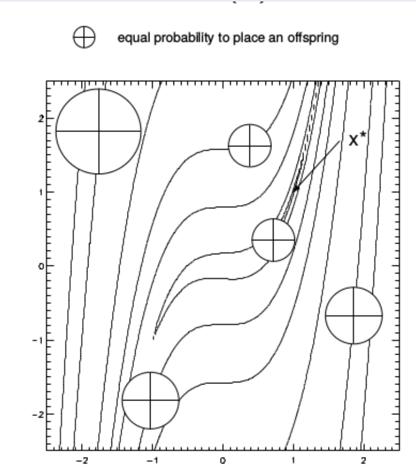
1. $\sigma := \sigma \cdot \exp(\mathcal{N}(0, \tau_0))$
2. $x_i := x_i + \mathcal{N}_i(0, \sigma^2), \; 1 \leq i \leq n$

where

$$\tau_0 \propto \frac{1}{n}$$

Evolution Strategies: Mutation (I)

Simple self-adaptive Mutation ($n = 2$)



Diagonal self-adaptive Mutation

$n_\sigma = n, n_\alpha = 0$ (one mutation parameter per individual and variable)

1. $\sigma_i := \sigma_i \cdot \exp(\mathcal{N}(0, \tau') + \mathcal{N}_i(0, \tau))$
2. $x_i := x_i + \mathcal{N}_i(0, \sigma_i^2), \;\; 1 \leq i \leq n$

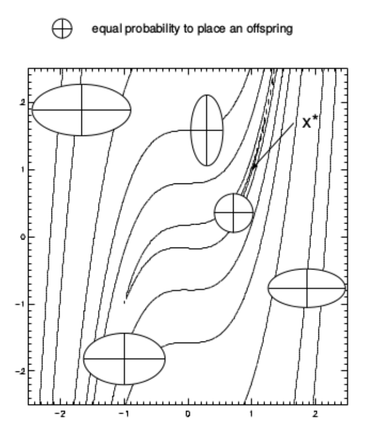
where

$$\tau \propto \frac{1}{2\sqrt{n}}$$

$$\tau' \propto \frac{1}{2n}$$

Evolution Strategies: Mutation (II)

Diagonal Self-Adaptive Mutation ($n = 2$)



Evolution Strategies: Mutation (III)

Correlated self-adaptive Mutation

$n_\sigma = n, n_\alpha = \frac{n(n-1)}{2}$ (one covariance matrix per individual)

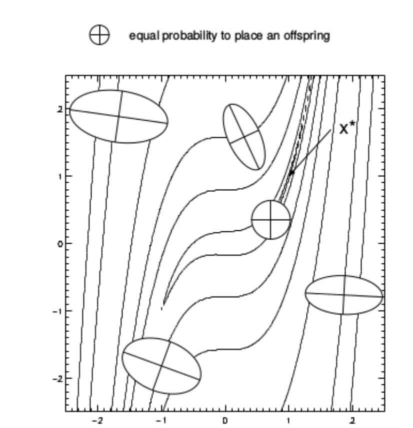
1. $\sigma_i := \sigma_i \cdot \exp(\mathcal{N}(0, \tau') + \mathcal{N}_i(0, \tau))$
2. $\alpha_i := \alpha_i + \mathcal{N}_i(0, \beta^2), \; 1 \leq i \leq n$
3. Build Σ using the $\boldsymbol{\sigma}$ and $\boldsymbol{\alpha}$ for individual $\mathbf{x}$
4. $\mathbf{x} := \mathbf{x} + \mathcal{N}(0, \Sigma)$

where

$$\tau \propto \frac{1}{2\sqrt{n}}$$
$$\tau' \propto \frac{1}{2n}$$
$$\beta \propto 5^\circ \; (\pi/36 \text{ radians})$$

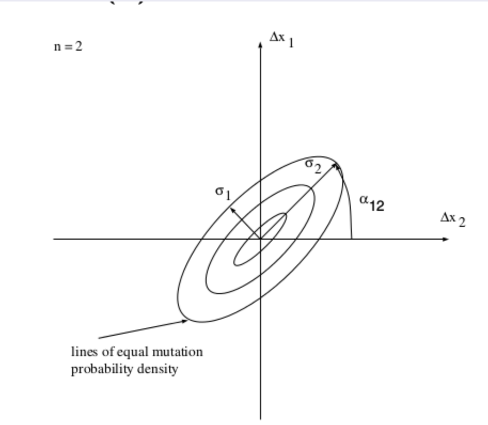
Evolution Strategies: Mutation (III)

Correlated Self-Adaptive Mutation ($n = 2$)



Evolution Strategies: Mutation (III)

Illustration of the mutation ellipsoid for the case ($n = 2$)



Evolution Strategies: Mutation (III)

Theorem

A real symmetric matrix $A_{n\times n}$ is P.D. iff it can be decomposed as $A = T^{\mathrm{T}}DT$, with T orthogonal and D diagonal with positive entries:

$$T = \prod_{i=1}^{n-1} \prod_{j=i+1}^{n} R_{ij}(\alpha_{ij})$$

- T is the product of $\frac{n(n-1)}{2}$ elementary rotation matrices
- α_{ij} are the rotation angles (between axes i and j)
- $R_{ij}(\alpha)$ is build as the identity matrix and modified as:

$$[R_{ij}(\alpha)]_{ii} = [R_{ij}(\alpha)]_{jj} := cos(\alpha)$$
$$[R_{ij}(\alpha)]_{ij} = -[R_{ij}(\alpha)]_{ji} := -sin(\alpha)$$

Evolution Strategies: log-normal mutation for the σ_i

log-normal distribution

It is a continuous probability distribution whose logarithm is normally distributed. A random variable which is log-normally distributed takes only positive real values.

$$\sigma_i := \sigma_i \cdot \exp(\mathcal{N}(0, \tau'))$$

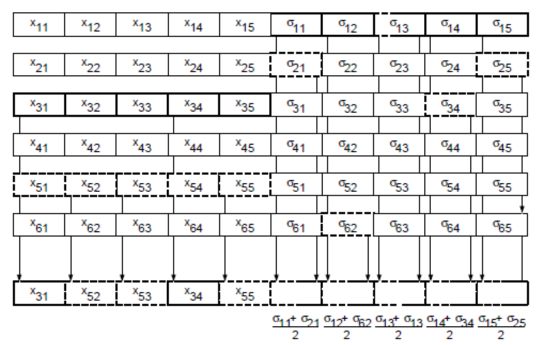
1. Multiplication by positive values preserves positivity
2. $Pr\{x\} = Pr\{\frac{1}{x}\}, x > 0$
3. Small modifications are more probable than larger ones

Evolution Strategies: recombination (I)

- Usually introduced as the *first* operator
- Generates an intermediate population size of λ by generating one individual at a time out of ξ parents by looping $\lambda \gg \mu$ times (generation of a **surplus**)
- Typically $\xi = 2$ (**dual**) or $\xi = \mu$ (**global** recombination):
  - dual: the two parents are chosen at random, per individual
  - global: one parent is held fixed and the other is chosen anew per each gene
- Applied to both objective and strategy parameters (and often differently)
- Two basic ways: choose randomly (**discrete**) and average (**intermediate**)

Evolution Strategies: recombination (II)

recombination example



- $\mu = 6, n = 5, n_\sigma = n, n_\alpha = 0$ (one mutation parameter per individual and gene)
- dual discrete recombination on x_i; global intermediate on σ_i (first parent held fixed, second chosen anew)

Evolution Strategies: replacement

- Strictly deterministic, rank-based
- The μ best are treated equally
- (μ, λ) selection:
  - offspring surplus $\lambda \gg \mu$
  - important (necessary?) for self-adaptation
  - useful for moving optima, noisy F, ...

$\Rightarrow$ Very strong selective pressure

Evolution Strategies: summary

The crucial claim (Schwefel '87 '92)

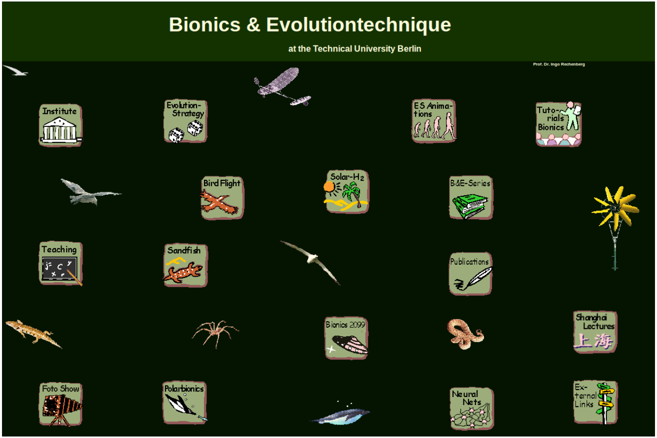
Self-adaptation of strategy parameters works

- without exogenous or centralized control
- needs mutation of all parameters
- needs generation of a surplus and (μ, λ) replacement
- needs recombination of all parameters

default (recommended) settings:

- $\mu = 15, \lambda \propto 7\mu = 105$
- dual discrete recombination on objective parameters
- global intermediate on strategy parameters

Evolution Strategies: demos



Prof. Dr. Ingo Rechenberg `http://www.bionik.tu-berlin.de/`

Evolution Strategies: Modern developments

The CMA-ES (Covariance Matrix Adaptation Evolution Strategy) is the more recent development of ESs

- It uses a sophisticated method to update the covariance matrix, particularly useful if the fitness function is complex
- Adaptation of the covariance matrix amounts to learning a second order model of the underlying function (similar to the approximation of the inverse Hessian matrix in quasi-Newton methods)

Resources:

- *A short introduction to CMA-ES* by N. Hansen: `http://www.lri.fr/~hansen/cmaesintro.html`
- Matlab code: `http://www.lri.fr/~hansen/cmaes_inmatlab.html`
- R package `parma`