

6 Evolution Strategies

- Basic introduction: Representation, strategy parameters, mutation, recombination, selection.
- Self-adaptation of strategy parameters.
- The (1+1)-evolution strategy.
- Convergence velocity theory.
- Application examples.



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 (μ, λ) -selection.
- *Self-adaptation* of strategy parameters: First self-adaptive EA.
- Generation of an offspring surplus $\lambda \gg \mu$.



Representation (1)

Spaces:

- Phenotype space:

$$\mathbb{R}^n$$

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

$$\mathcal{S} = \mathbb{R}_+^{n_\sigma} \times [-\pi, \pi]^{n_\alpha}$$

- Individual space (genotype):

$$I = \mathbb{R}^n \times \mathcal{S}$$

One individual:

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



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 set ($s = (\vec{\sigma}, \vec{\alpha}) \in \mathcal{S}$):

- Is part of an individual.
- Represents the probability density function (p.d.f.) for its mutation.

n_σ	n_α	Remark
1	0	standard mutation
n	0	standard mutations
n	$n \cdot (n - 1) / 2$	correlated mutations
$1 \leq n_\sigma \leq n$	$(n - \frac{n_\sigma}{2})(n_\sigma - 1)$	general case (correlated mutations)

Possible settings of n_σ and n_α .



Genetic operators: mutation (1)

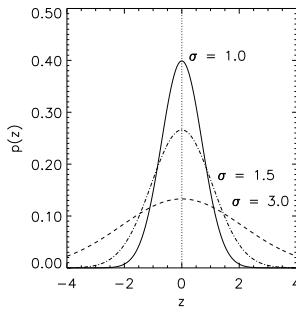
Simple mutation I

- Simple mutation makes use of normally distributed variations, $N(\xi, \sigma)$.

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

- Expectation (ξ) is assumed to equal 0.
- Standard deviation (σ) must be adapted.

The one-dimensional case.



Genetic operators: mutation (2)

Simple mutation I (continued)

- x_i is mutated by adding some Δx_i from a normal probability distribution.
- σ is mutated by multiplying by e^Γ , with Γ from a normal probability distribution.

$$\begin{aligned} I &= \mathbb{R}^n \times \mathbb{R}_+ \\ m'_{\{\tau_0\}}(\vec{x}, \sigma) &= (\vec{x}', \sigma') \\ \tau_0 &\sim 1/\sqrt{n} \end{aligned}$$

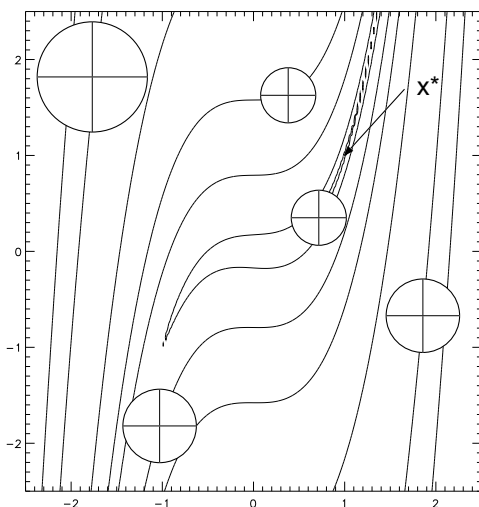
$$\begin{aligned} \sigma' &= \sigma \cdot \exp(\tau_0 \cdot N(0, 1)) \\ x'_i &= x_i + \sigma' \cdot N_i(0, 1) \end{aligned}$$



Genetic operators: mutation (3)

Simple mutation I (continued)

⊕ equal probability to place an offspring



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



Genetic operators: mutation (4)

Simple mutation II

- x_i is mutated by adding some Δx_i from a normal probability distribution.
- σ_j is mutated by multiplying by e^{Γ_j} with Γ_j from a normal probability distribution.

$$\begin{aligned} I &= \mathbb{R}^n \times \mathbb{R}_+^n \\ m'_{\{\tau, \tau'\}}(\vec{x}, \vec{\sigma}) &= (\vec{x}', \vec{\sigma}') \\ \tau &\sim 1/\sqrt{2\sqrt{n}} \\ \tau' &\sim 1/\sqrt{2n} \end{aligned}$$

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

Boundary rule for preserving standard deviations larger than zero:

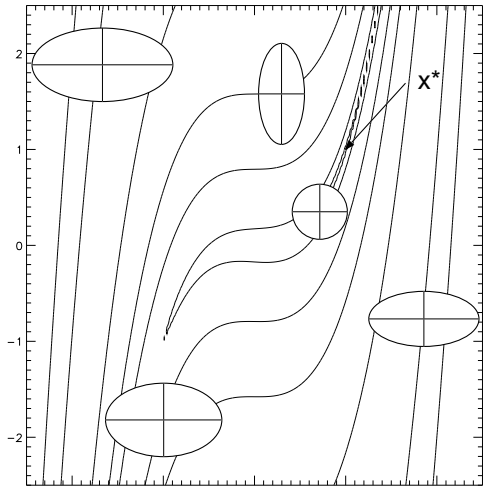
$$\sigma'_i < \varepsilon_\sigma \Rightarrow \sigma'_i := \varepsilon_\sigma$$



Genetic operators: mutation (5)

Simple mutation II (continued)

⊕ equal probability to place an offspring



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



Genetic operators: mutation (6)

Correlated mutation

- Correlated mutation uses the following probability distribution function for Δx :

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

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

$$c_{ii} = \sigma_i^2$$

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



Genetic operators: mutation (7)

Correlated mutation (continued)

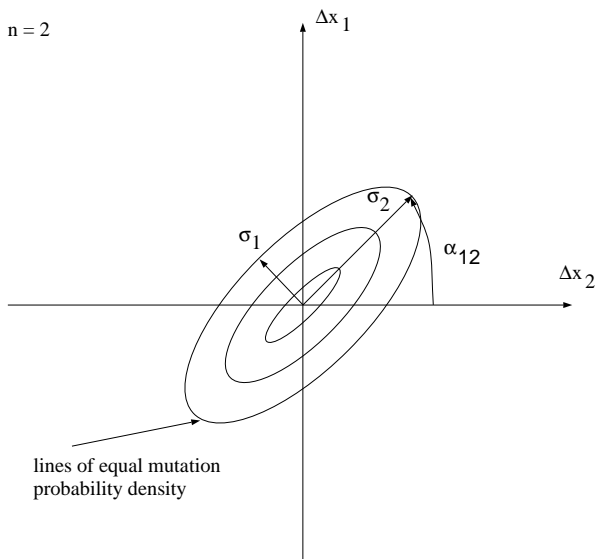


Illustration of the mutation ellipsoid for the case $n = 2, n_\sigma = 2, n_\alpha = 1$.



Genetic operators: mutation (8)

Correlated mutation (continued)

- \vec{x} is mutated by adding some $\vec{\Delta}_x$ from an n-dimensional normal distribution.
- σ_i is mutated by multiplying by e^{Γ_i} with Γ_i from a normal probability distribution.
- α_j is mutated by adding some $\Delta\alpha_j$ from a normal probability distribution.



Genetic operators: mutation (9)

Correlated mutation (continued)

$$\begin{aligned}
 n_\alpha &= n \cdot (n - 1) / 2 \\
 I &= \mathbb{R}^n \times \mathbb{R}_+^n \times [-\pi, \pi]^{n_\alpha} \\
 m_{\{\tau, \tau', \beta\}}^l(\vec{x}, \vec{\sigma}, \vec{\alpha}) &= (\vec{x}', \vec{\sigma}', \vec{\alpha}') \\
 \tau &\sim 1 / \sqrt{2\sqrt{n}} \\
 \tau' &\sim 1 / \sqrt{2n} \\
 \beta &\approx 5^\circ
 \end{aligned}$$

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

Boundary rule for keeping rotation angles feasible:

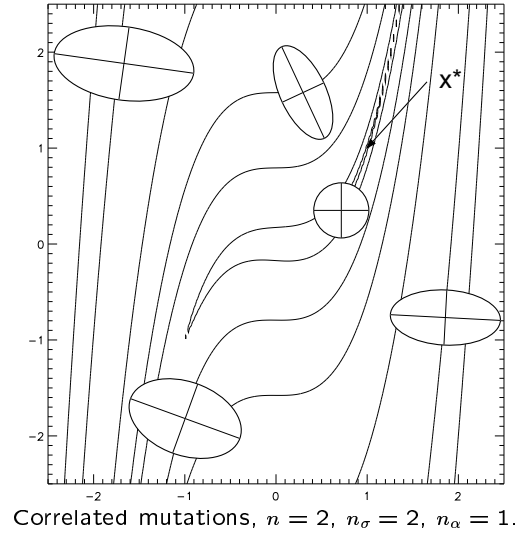
$$|\alpha'_j| > \pi \Rightarrow \alpha'_j := \alpha'_j - 2\pi \cdot \text{sign}(\alpha'_j)$$



Genetic operators: mutation (10)

Correlated mutation (continued)

\oplus equal probability to place an offspring



Genetic operators: mutation (11)

Some remarks:

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



Genetic operators: mutation (12)

Some remarks (continued):

- Standard strategy: $n_\sigma = n, n_\alpha = 0$.
- For correlated mutations:
 - $\vec{\sigma}_c \sim \vec{N}(\vec{0}, C)$ is generated by a multiplication of the uncorrelated random vector $\vec{\sigma}_u$ by n_α rotation matrices (Schwefel 1981, Rudolph 1992).

$$\vec{\sigma}_c = \prod_{i=1}^{n-1} \prod_{j=i+1}^n R(\alpha_{ij}) \cdot \vec{\sigma}_u$$

- Exactly the feasible (positive definite) correlation matrices C can be created this way (Rudolph 1992).



Genetic operators: recombination (2)

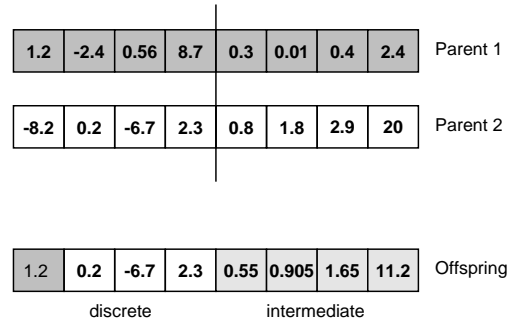
The operator:

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.



Genetic operators: recombination (3)

Recombination illustrated



Discrete recombination on x_i , intermediate on σ_i .



Selection (1)

- Strictly deterministic, rank-based.
- The μ best ranks are handled equally.
- The μ best offspring ($P''(t)$) survive.
 - Important for self-adaptation.
 - Applicable also for noisy objective functions, moving optima.
- N.B. μ selected from λ ; notation: (μ, λ) .
- Selective pressure: very high.



Selection (2)

Takeover time τ^* :

Definition:

number of generations until repeated application of selection completely fills the population with copies of the best individual.

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 GAs: $\tau^* \approx \lambda \ln \lambda = 460$ generations!



Other components

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



Reproduction Cycle

Generational ES model:

```

t := 0;
initialize P(t);
evaluate P(t);
while not terminate 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
  
```

- *recombine*: Recombination applied to all individuals.
- $P'(t)$ has size $\lambda > \mu$, $P(t)$ size μ .
- *mutate*: Normally distributed variations, all individuals.
- *select*: $(\mu + \lambda)$ or (μ, λ) .



Self-adaptation principles

- Biological model: Repair enzymes, mutator genes.
- No deterministic control: strategy parameters *evolve*.
- *Indirect* link between fitness and useful strategy parameter settings.
- Strategy parameters are conceivable as an *internal model* of the local topology.
- Individual space:

$$I = M \times S$$

- M : Search space.
- S : Strategy parameter space.

**The crucial claim (Schwefel 1987, 1992):**

Self-adaptation of strategy parameters works

- Without exogenous control.
- By recombining/mutating the strategy parameters.
- By exploiting the implicit link between fitness and useful internal model.

**Necessary conditions (found by experiments):**

- Generation of a surplus, $\lambda > \mu$
- (μ, λ) -selection (to guarantee extinction of misadapted individuals).
- A not too strong selective pressure e.g., (15,100) where $\lambda/\mu \approx 7$, but clearly $\mu > 1$ is necessary.
- Recombination also on strategy parameters (especially: intermediate recombination).

**Empirical Test Design**

- With simple functions (with predictable optimal σ_i values), check whether it works.
- Investigate impact of selection.
- Compare with optimal behavior (if known).



Test functions for experiments

- One common step size ($n_\sigma = 1$): Sphere model.

$$f_1(\vec{x}) = \sum_{i=1}^n x_i^2$$

- Appropriate scaling of variables ($n_\sigma = n$):

$$f_2(\vec{x}) = \sum_{i=1}^n i \cdot x_i^2$$

- A metric ($n_\sigma = n, n_\alpha = n \cdot (n - 1)/2$):

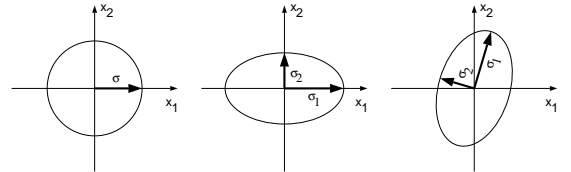
$$f_3(\vec{x}) = \sum_{i=1}^n \left(\sum_{j=1}^i x_j \right)^2$$



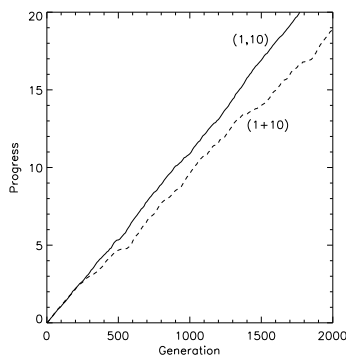
Experiments

Sketch of the lines of equal probability density

- Left: Standard mutations, $n_\sigma = 1$.
- Middle: Standard mutations, $n_\sigma = 2$.
- Right: Correlated mutations, $n_\sigma = 2, n_\alpha = 1$.



Experimental Results on Sphere Model (1)



Convergence velocity of a (1, 10)-ES vs. that of a (1 + 10)-ES (sphere model f_1 with $n = 30$ and $n_\sigma = 1$).



Experimental Results on Sphere Model (2)

Progress measure:

$$P_g = \log \sqrt{\frac{f_{\min}(0)}{f_{\min}(g)}}$$

- Counterintuitive: Elitist strategy is a bad choice.
- Misadapted σ might survive in an elitist strategy.
- Forgetting is necessary to prevent stagnation periods.

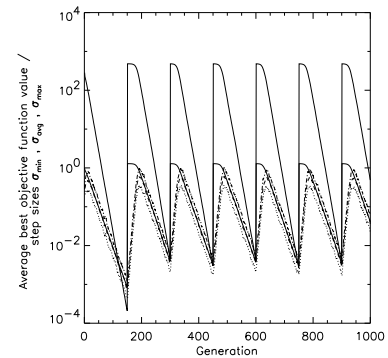


Time-Varying Sphere Model (1)

- Sphere model, $f(\vec{x}) = \|\vec{x} - \vec{x}^*\|^2 = R^2$.
- Optimum location \vec{x}^* is shifted every 150 generations.
- (15,100)-ES, $n_\sigma = 1$, $n = 30$, no recombination.
- Simple model of a dynamic environment (with “catastrophes”).



Time-varying Sphere Model (2)



Best objective function value and minimum, average, maximum and optimal standard deviation.



Time-varying Sphere Model (3)

- Standard deviation σ adapts to the optimum value

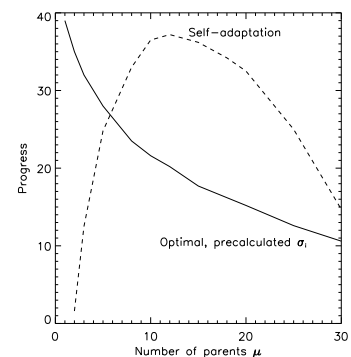
$$\sigma_{opt} = c_{\mu,\lambda} \frac{R}{n} = c_{\mu,\lambda} \frac{\sqrt{f(\vec{x})}}{n}$$

- Transition time is $g \propto n$ (Beyer 1995).

⇒ The principle *learns* the optimal setting of the mutation rate (“internal strategy”) without exogenous control.



Self-Adaptation is Collective Learning (1)



Average convergence velocity on f_2

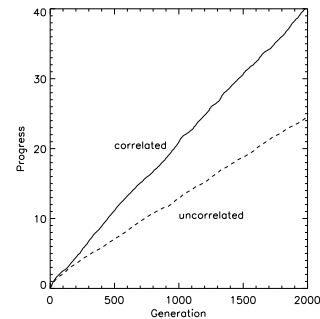


Self-Adaptation is Collective Learning (2)

- $(\mu, 100)$ -ES with $\mu \in \{1, \dots, 30\}$
- $n_\sigma = n = 30$, and the optimum $\sigma_i \propto 1/\sqrt{i}$ is known.
- Optimum setting of σ_i : $\mu = 1$ performs best.
- Self-adaptation: $\mu = 12$ imperfect, diverse parents are as good as the optimal strategy.
- Individuals exchange information about their "internal models" by recombination.



Self-Adaptation of Covariances (1)



Convergence velocity of ES with correlated mutations vs. one with self-adaptation of standard deviations only, on f_3 .



Self-Adaptation of Covariances (2)

- $(15, 100)$ -ES, $n = n_\sigma = 10$, $n_\alpha = 45$.
- Recombination:
 - Intermediary on x_i .
 - Global intermediary on σ_i .
 - None on α_j (covariances).

Covariances increase effectiveness in case of rotated coordinate systems.



Other Variants for Continuous Search Spaces

- Original EP:

$$\sigma' = \sigma \cdot (1 + \alpha \cdot N(0, 1))$$

Equivalent to log-normal with $n_\sigma = 1$, $\tau_0 = \alpha$ (Beyer 1995).

- Two-point distribution:

$$\sigma' = \begin{cases} \sigma \cdot \alpha & \text{if } u \sim U(0, 1) \leq 1/2 \\ \sigma/\alpha & \text{if } u \sim U(0, 1) > 1/2 \end{cases}$$

(Mutational step size control after Rechenberg, $\alpha = 1.3$).

- Substitution of $N(0, 1)$ by other distributions (e.g., one-dimensional Cauchy, Yao and Liu 1996).



Self-Adaptation: Conclusions

- Powerful & robust parameter control scheme.
- Optimal conditions concerning selection, population size, etc.?
- Perfect adaptation vs. useful diversity — or a mixture ?
- Optimal speed of self-adaptation (i.e., learning rate settings) ?
- Few theoretical results.



Self-Adaptation: Individuals as Agents

- Individuals are *autonomous*; internal control of their behavior (mutation).
- Individuals *communicate* by exchanging partial information (recombination).
- Individuals are *reactive* to their environment (objective function).
- Further possibilities:
 - Spatial communication structure (graph).
 - Parallel implementation.
 - More complex internal strategies; including symbolic representation.



Basic Theory: The (1+1)-Strategy

- Properties of the mutation vector.
- The (1+1)-evolution strategy.
- Convergence velocity: Sphere model, corridor model.
- 1/5-success rule.
- Evolution window.
- ES vs. gradient strategy.



The mutation vector (1)

$$\Delta \vec{x} = \vec{z} = (z_1, \dots, z_n)$$

Z_1, \dots, Z_n : $(0, \sigma)$ -normally distributed random variables.

$$\Rightarrow S^2 = \sum_{i=1}^n Z_i^2 \text{ is } \chi^2\text{-distributed.}$$

Random variable $S = \sqrt{S^2}$:

Length of the mutation vector \vec{z} .

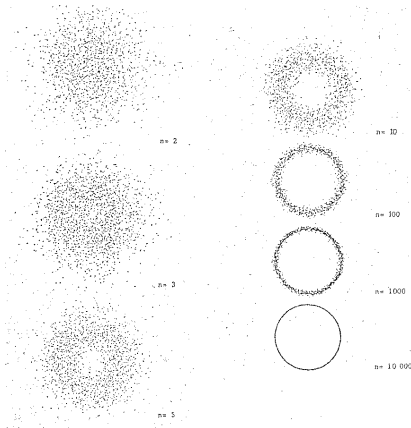
After some math:

$$E(S) \approx \sigma \sqrt{n} \quad , \quad V(S) = \frac{1}{2} \sigma$$



The mutation vector (2)

- Variance $V(S)$ is independent of n .
- For large n : Offspring located on hypersphere of radius $E(S) \approx \sigma \sqrt{n}$.



The simple (1+1)-ES (1)

Algorithm:

```

t := 0;
initialize P(0) := {x̄(0)} ∈ I, I = IR^n, x̄ = (x1, ..., xn);
evaluate P(0) := {f(x̄(0))}
while not terminate(P(t)) do
  mutate: x̄'(t) := m(x̄(t))
  where x'_i := x_i + σ(t) · N_i(0, 1) ∀ i ∈ {1, ..., n}
  evaluate: P'(t) := {x̄'(t)} : {f(x̄'(t))}
  select: P(t+1) := s_{(1+1)}(P(t) ∪ P'(t));
  t := t + 1;
  if (t mod n = 0) then
    σ(t) := { σ(t-n)/c, if p_s > 1/5
             σ(t-n) · c, if p_s < 1/5
             σ(t-n), if p_s = 1/5
    where p_s is the relative frequency of successful mutations, measured over intervals of, say, 10 · n trials;
    and 0.817 ≤ c ≤ 1;
  else
    σ(t) := σ(t-1);
  fi
od
    
```



(1+1)-ES: Convergence velocity (1)

Convergence velocity: Expectation of the distance towards the optimum covered per generation.

$$\varphi = E(\|\vec{x}^* - \vec{x}_t\| - \|\vec{x}^* - \vec{x}_{t+1}\|)$$

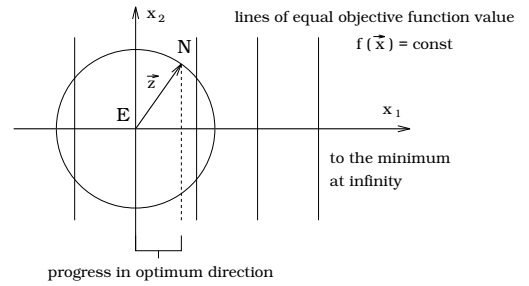
Alternatively:

$$\tilde{\varphi} = E(|f(\vec{x}^*) - f(\vec{x}_t)| - |f(\vec{x}^*) - f(\vec{x}_{t+1})|)$$



(1+1)-ES: Convergence velocity (2)

Example: The linear model.



$$\begin{aligned} \varphi &= E(Z_1) \\ &= \sigma \cdot E(Z'_1) \\ &= \sigma \int_0^\infty z_1 \phi(z_1) dz_1 \\ &= \frac{\sigma}{\sqrt{2\pi}} \int_0^\infty z_1 \cdot \exp\left(-\frac{z_1^2}{2}\right) dz_1 \\ &= \frac{\sigma}{\sqrt{2\pi}} \end{aligned}$$

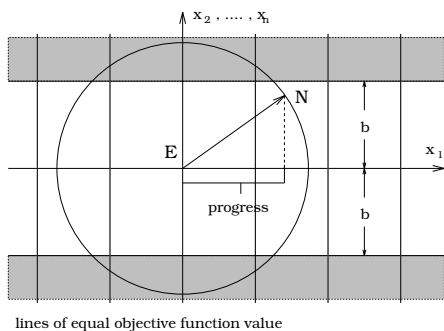
Notice: $Z_1 \sim N(0, \sigma)$, $Z'_1 \sim N(0, 1)$



(1+1)-ES: Convergence velocity (3)

Example: The corridor model.

$$f(\vec{x}) = c \cdot x_1 \quad -b \leq x_2, \dots, x_n \leq b$$



(1+1)-ES: Convergence velocity (4)

Result of analysis:

$$\varphi = \frac{\sigma}{\sqrt{2\pi}} \left(1 - \frac{\sigma}{\sqrt{2\pi}b}\right)^{n-1}$$

After normalization of variables $\varphi' = \frac{\varphi}{b}$, $\sigma' = \frac{\sigma}{b}$:

$$\varphi' \approx \frac{\sigma'}{\sqrt{2\pi}} \exp\left(-\frac{\sigma'}{\sqrt{2\pi}}\right) \quad \text{for } n \gg 1$$

Success probability: $w_e = P(f(\vec{x}') \leq f(\vec{x}))$

$$w_e \approx \frac{1}{2} \exp\left(-\frac{\sigma'}{\sqrt{2\pi}}\right) \quad \text{for } n \gg 1$$

From these results:

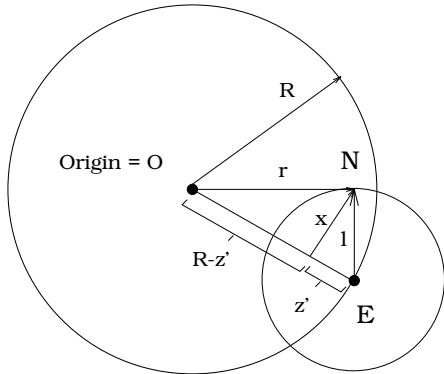
- Optimal standard deviation: $\sigma'_{opt} = \sqrt{2\pi}$
- Maximum convergence velocity: $\varphi'_{max} = \frac{1}{e}$
- Optimal success probability: $w_{e,opt} = \frac{1}{2e}$



(1+1)-ES: Convergence velocity (5)

Example: The sphere model.

$$f(\vec{x}) = \sum_{i=1}^n x_i^2 = r^2$$



Geometry: $z'^2 + x^2 = l^2$, $x^2 + (R - z')^2 = r^2$, i.e.,
 $r^2 = l^2 + R^2 - 2Rz'$



(1+1)-ES: Convergence velocity (6)

Derivation:

$$\begin{aligned} \tilde{\varphi} &= E(R^2 - r^2) = E(2RZ' - l^2) \\ &= E(2R\sigma Z - \sigma^2 n) \\ &= \dots \\ &= 2R\sigma \int_{z_{\min}}^{\infty} z\phi(z)dz - \sigma^2 n \int_{z_{\min}}^{\infty} \phi(z)dz \\ &= \frac{2R\sigma}{\sqrt{2\pi}} \exp\left(-\frac{\sigma^2 n^2}{8R^2}\right) - \sigma^2 n \left(1 - \Phi\left(\frac{\sigma n}{2R}\right)\right) \end{aligned}$$



(1+1)-ES: Convergence velocity (7)

Normalization of variables:

$$\varphi' = \frac{\sigma'}{\sqrt{2\pi}} \exp\left(-\frac{\sigma'^2}{8}\right) - \frac{\sigma'^2}{2} \left(1 - \Phi\left(\frac{\sigma'}{2}\right)\right)$$

Success probability:

$$w_e = 1 - \Phi(\sigma'/2)$$

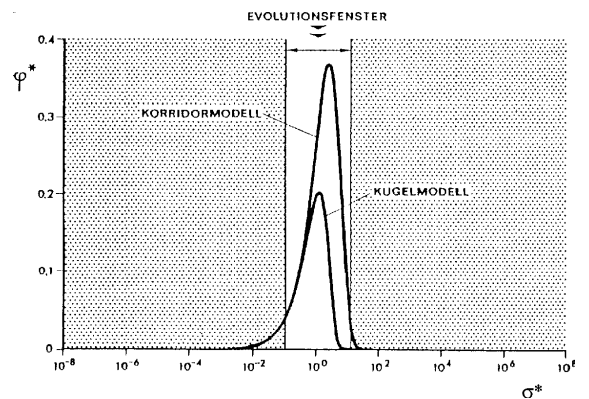
From these results:

- Optimal standard deviation: $\sigma'_{opt} \approx 1.224$
- Maximum convergence velocity: $\varphi'_{max} \approx 0.2025$
- Optimal success probability: $w_{e,opt} \approx 0.270$



(1+1)-ES: 1/5-success rule (1)

Progress window:



(1+1)-ES: 1/5-success rule (2)

Formulation of the 1/5-success rule:

$w_{e_{opt}}$ should be about 1/5. If w_e — measured during execution of the (1+1)-ES — is larger than 0.2, increase σ . If it is smaller than 0.2, decrease σ .

Algorithmically:

$$\sigma(t+n) = \begin{cases} \sigma(t) & , w_e = 0.2 \\ \sigma(t) \cdot k & , w_e < 0.2 \\ \sigma(t)/k & , w_e > 0.2 \end{cases}$$

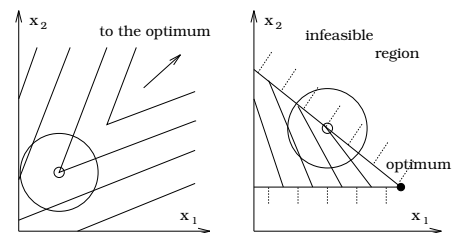
Choice of k : $k \approx 0.82$ (theoretical arguments, Schwefel 1977).



(1+1)-ES: 1/5-success rule (2)

Disadvantages of the (1+1)-ES:

- Certainly a more local search method.
- 1/5 success rule may fail.



(1+1)-ES vs. gradient method (1)

Gradient method:

- n sampling steps (partial derivative determinations).
- 1 working step, length $s = \sqrt{\sum_{i=1}^n z_i^2}$.

$$\varphi_{gradient} = \frac{s}{n+1} \sim \frac{1}{n}$$

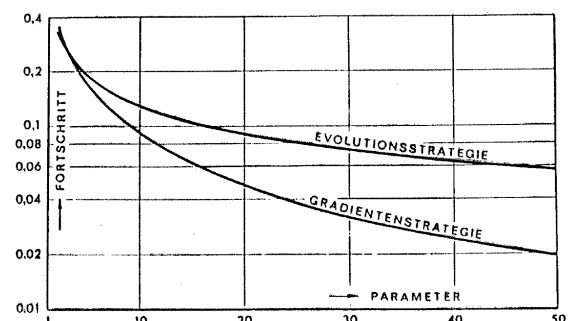
(1+1)-ES:

$$\varphi_{(1+1)} = \frac{\sigma}{\sqrt{2\pi}} = \frac{E(S)}{\sqrt{2\pi n}} \sim \frac{1}{\sqrt{n}}$$

(since $E(S) = \sigma\sqrt{n}$).



(1+1)-ES vs. gradient method (2)



Progress rate as a function of n , ES vs. gradient.



Extended Theory

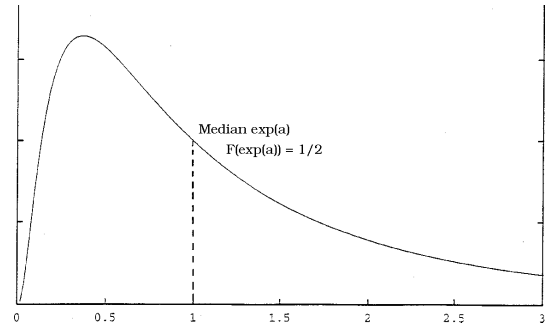
- The lognormal distribution (self-adaptation).
- $(1+\lambda)$ and $(1,\lambda)$ -strategies.
- (μ,λ) -strategies with recombination:
Discrete recombination, intermediary recombination.



Log-normal distribution (1)

- Probability density function:

$$f_X(x) = \frac{1}{\sigma x \sqrt{2\pi}} \cdot \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right)$$



log-normal distribution, $\sigma = 1, \mu = 0$.



Log-normal distribution (2)

- Expectation:

$$E(X) = \exp\left(\mu + \frac{\sigma^2}{2}\right)$$

- Median:

$$\exp(\mu) \quad \left(F_X(\exp(\mu)) = \frac{1}{2}\right)$$

Characteristics:

- Identical probability to sample x and $\frac{1}{x}$.
- Small changes more likely than large ones.
- σ_i are guaranteed to remain positive.

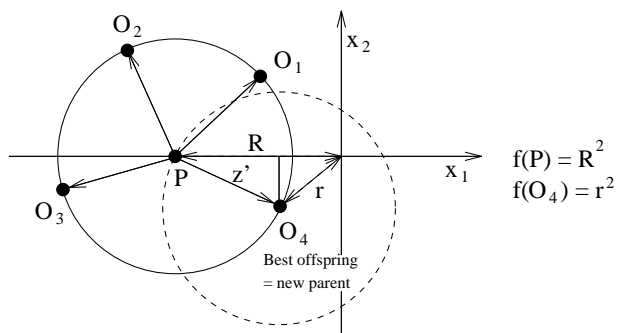


Convergence velocity of multi-membered ESs (1)

Simplifications:

- No self-adaption.
- One step-size.
- No recombination.
- $\mu = 1$

$\Rightarrow (1 + \lambda)$ -strategies, $(1, \lambda)$ -strategies.



$(1,4)$ -strategy, sphere model



Convergence velocity of multi-membered ESs (2)

Definition:

$Z_1, Z_2, \dots, Z_\lambda$ i.i.d. random variables with p.d.f. $p(z)$.

$$Z_{1:\lambda} \leq Z_{2:\lambda} \leq \dots \leq Z_{\lambda:\lambda}$$

is called order statistics of the Z_i . $p_{v:\lambda}(z)$ denotes the p.d.f. of $Z_{v:\lambda}$.

Idea:

Best of offspring individual has

- smallest value of $r \Rightarrow r_{1:\lambda}$
- largest value of $z' \Rightarrow z'_{\lambda:\lambda}$

Z' : projection into direction of origin.

$$Z'_{v:\lambda} \sim N(0, \sigma)$$

$$Z_{v:\lambda} \sim N(0, 1)$$



Convergence velocity of multi-membered ESs (3)

$$\tilde{\varphi}_{(1:\lambda)} = E(R^2 - r_{1:\lambda}^2)$$

$$r_{v:\lambda}^2 = l^2 + R^2 - 2R \cdot Z'_{\lambda-v+1:\lambda}$$

Some math:

$$\tilde{\varphi}_{(1:\lambda)} = E(2R \cdot Z'_{\lambda:\lambda} - \sigma^2 n)$$

$$= E(2R\sigma \cdot Z_{\lambda:\lambda} - \sigma^2 n)$$

$$= \int_{z_{min}}^{\infty} (2R\sigma \cdot z - \sigma^2 n) \cdot p_{\lambda:\lambda}(z) dz$$

$$= 2R\sigma \int_{z_{min}}^{\infty} z \cdot p_{\lambda:\lambda}(z) dz -$$

$$\sigma^2 n \int_{z_{min}}^{\infty} p_{\lambda:\lambda}(z) dz$$



Convergence velocity of multi-membered ESs (4)

With:

$$p_{\lambda:\lambda}(z) = \lambda \phi(z) (\Phi(z))^{\lambda-1} = \frac{d}{dz} (\Phi(z))^\lambda$$

It follows that:

$$\tilde{\varphi}_{(1:\lambda)} = 2R\sigma \int_{z_{min}}^{\infty} z \cdot \frac{d}{dz} (\Phi(z))^\lambda dz -$$

$$\sigma^2 n \int_{z_{min}}^{\infty} \frac{d}{dz} (\Phi(z))^\lambda dz$$



Convergence velocity of (1, λ)-ESs (1)

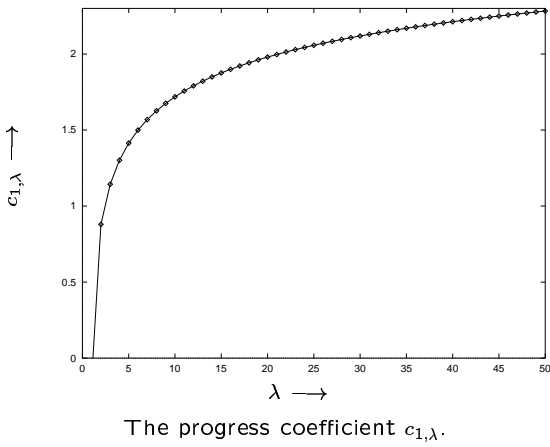
When accepting everything (non-elitist), $z_{min} = -\infty$.

$$\tilde{\varphi}_{(1,\lambda)} = 2R\sigma \cdot c_{1,\lambda} - \sigma^2 n$$

$$c_{1,\lambda} := E(Z_{\lambda:\lambda}) \begin{cases} \text{progress coefficient (Rechenberg)} \\ \text{selection intensity (Mühlenbein)} \end{cases}$$



Convergence velocity of $(1, \lambda)$ -ESs (2)



- Asymptotic behaviour: $c_{1,\lambda} \approx \sqrt{2 \ln \lambda}$.



Convergence velocity of $(1, \lambda)$ -ESs (3)

- Normalisation of $\tilde{\varphi}$, with

$$\varphi \approx \frac{\tilde{\varphi}}{2R}, \varphi' = \frac{\varphi_n}{R}, \sigma' = \frac{\sigma n}{R}$$

$$\varphi'_{1,\lambda} = c_{1,\lambda} \sigma' - \frac{1}{2} \sigma'^2$$

- Optimal standard deviation:

$$\sigma'_{opt} = c_{1,\lambda}$$

- Maximum convergence velocity:

$$\varphi'_{max} = \frac{1}{2} c_{1,\lambda}^2 \approx \ln \lambda$$

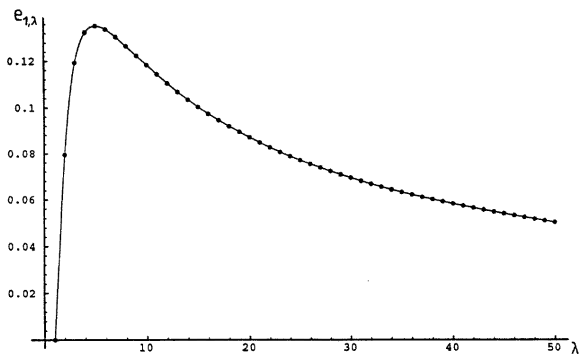
- Evolution condition: $\sigma' < 2c_{1,\lambda}$ (Guarantees $\varphi' > 0$).



Evolution efficiency

Maximum progress per individual

$$e_{1,\lambda} = \frac{\varphi'_{max}}{\lambda}$$



Convergence velocity of $(1 + \lambda)$ -ESS

- From $r \leq R$ it follows that

$$z_{min} = \frac{\sigma n}{2R}$$

- Thus:

$$\varphi'_{(1+\lambda)} = \sigma' c_{1+\lambda}(\sigma') - \frac{\sigma'^2}{2} (1 - \Phi^\lambda(\frac{\sigma'}{z}))$$

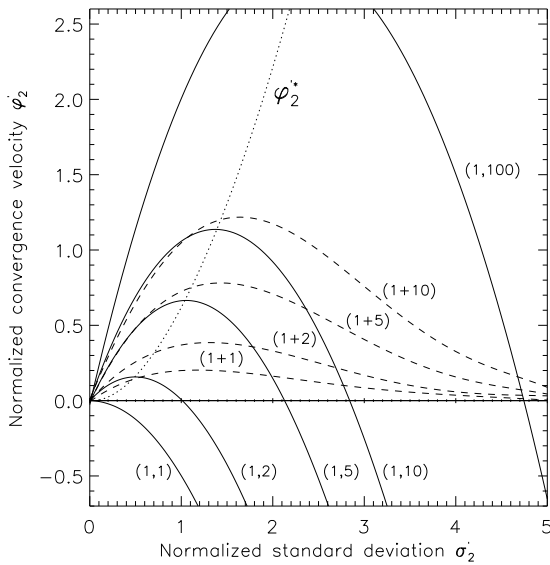
- Where

$$c_{1+\lambda}(x) = \int_{\frac{x}{2}}^{\infty} z \frac{z}{dz} \Phi^\lambda(z) dz$$

No further analytical simplifications are possible.



Convergence velocity: illustration



Normalized convergence velocity φ' as a function of normalized standard deviation σ' for $(1, \lambda)$ - and $(1 + \lambda)$ -evolution strategies.



Convergence velocity of (μ, λ) -ESs (1)

Simplifications:

- No self-adaptation.
- One step-size.
- Recombination:
 - center of mass recombination μ/μ_I (intermediary), or
 - global discrete recombination μ/μ_D .

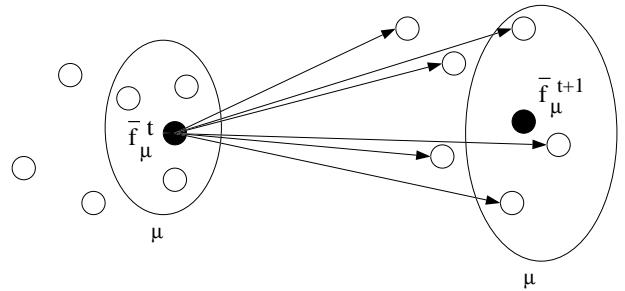


Illustration of center of mass recombination



Convergence velocity of (μ, λ) -ESs (2)

$$\begin{aligned} \varphi_{\mu,\lambda} &= \langle R \rangle - E(\langle \tilde{R} \rangle_{\mu,\lambda}) \\ &= \frac{1}{\mu} \sum_{v=1}^{\mu} R_v - \frac{1}{\mu} \sum_{v=1}^{\mu} r_{v:\lambda} \end{aligned}$$

Where:

- $\langle R \rangle$: Average distance to the optimum of parents.
- $\langle \tilde{R} \rangle_{\mu,\lambda}$: Average distance to the optimum of the μ best offspring.



Convergence velocity of $(\mu/\mu_I, \lambda)$ -ESs (1)

Without derivation (Rechenberg '94, Beyer '96):

$$\varphi'_{\mu/\mu_I,\lambda} = c_{\mu,\lambda} \cdot \sigma' - \frac{\sigma'^2}{2\mu}$$

(For $\sigma' \ll n, \mu^2 \ll n$)

- Optimal standard deviation:

$$\sigma'_{opt} = \mu \cdot c_{\mu,\lambda}$$

- Maximum convergence velocity:

$$\varphi'_{max} = \frac{1}{2} \mu \cdot c_{\mu,\lambda}^2$$



Convergence velocity of $(\mu/\mu_I, \lambda)$ -ESs (2)

Progress coefficient ($Z_{v:\lambda} \sim N(0, 1)$):

$$\begin{aligned} c_{\mu,\lambda} &= \frac{1}{\mu} \sum_{v=\lambda-\mu+1}^{\lambda} E(Z_{v:\lambda}) \\ &\approx \frac{\lambda}{\mu} \cdot \phi(\Phi^{-1}(1 - \frac{\mu}{\lambda})) \\ &\approx O\left(\sqrt{\ln \frac{\lambda}{\mu}}\right) \end{aligned}$$

Conjecture:

$$\varphi'_{max} \approx \mu \cdot \ln \frac{\lambda}{\mu}$$

Convergence velocity of $(\mu/\mu_D, \lambda)$ -ESs (1)

Without derivation (Rechenberg '94, Beyer '96):

$$\varphi'_{(\mu/\mu_D, \lambda)} = \sqrt{\mu} \cdot c_{\mu,\lambda} \sigma' - \frac{\sigma'^2}{2}$$

(For $\sigma' \ll n, \mu^2 \ll n$)

- Optimal standard deviation:

$$\sigma'_{opt} = \sqrt{\mu} \cdot c_{\mu,\lambda}$$

- Maximum convergence velocity:

$$\varphi'_{max} = \frac{1}{2} \mu \cdot c_{\mu,\lambda}^2$$

Again:

$$\varphi'_{max} \approx \mu \cdot \ln \frac{\lambda}{\mu}$$

Interpretation of results

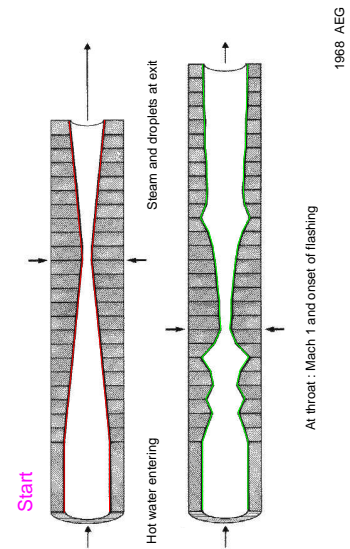
- Genetic repair (Beyer '96):
 μ/μ_I -recombination decreases the harmful part of mutation.
- Incest taboo:
 μ/μ_I -recombination is only useful, if parents are different from each other.
- Implicit genetic repair:
 μ/μ_D -recombination estimates the center of mass corresponding to a species centered around the wild-type.

Application Examples of ESs

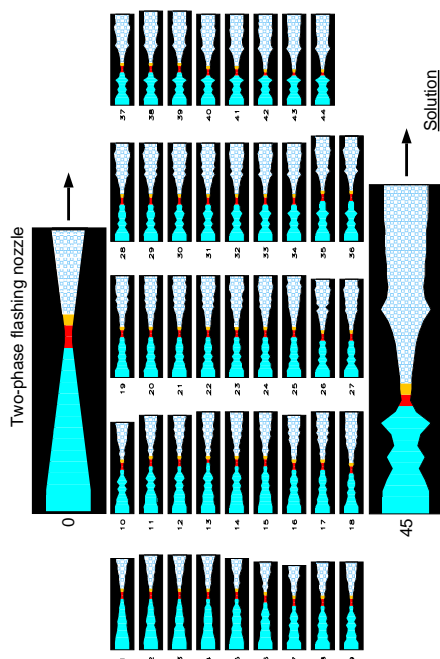
- Hot water / steam turbulent nozzle.
- Truss layout.
- Concrete shell roof.
- Dipole magnet structures.
- Optical multilayer design.



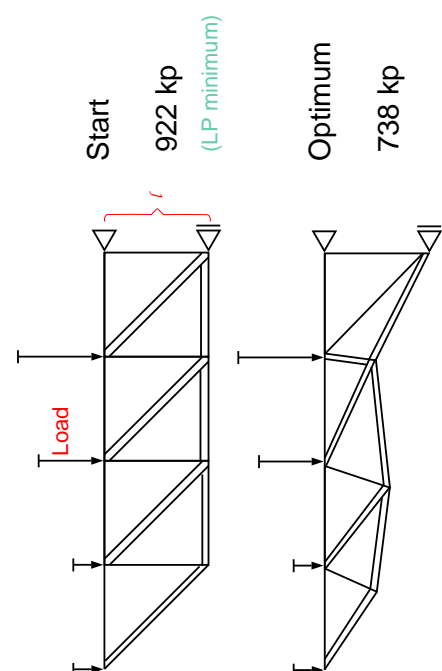
Hot Water flashing nozzle (1)



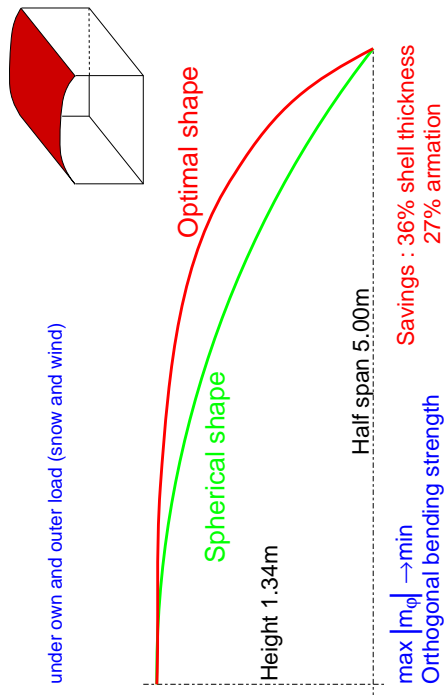
Hot Water flashing nozzle (2)



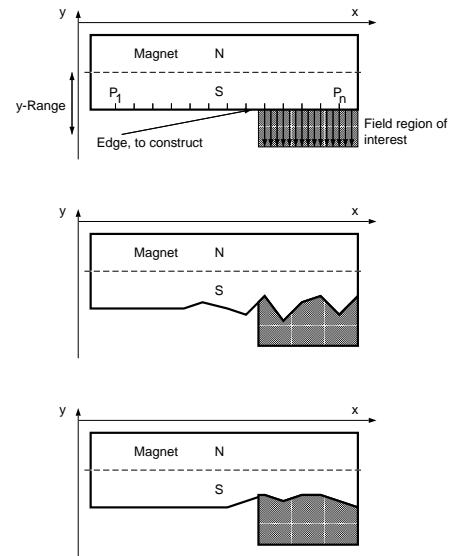
Minimal weight truss layout



Concrete shell roof



Dipole Magnet Structure (1)

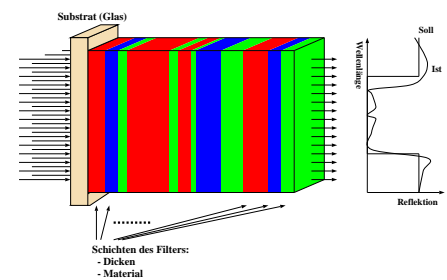


Dipole Magnet Structure (2)

- Analysis of the magnetic field by Finite Element Analysis (FEM).
- Minimize sum of squared deviations from the ideal.
- Individuals: Vectors of positions (y_1, \dots, y_n) .
- Middle: 9.82% better than upper graphic; bottom: 2.7% better.



Optical Multilayers (1)



Problem parameters:

- Thicknesses $\vec{d} = (d_1, \dots, d_n)$ of layers.
- Layer materials $\vec{\eta} = (\eta_1, \dots, \eta_n)$ (integer values).
- Number of layers n .

⇒ Mixed-integer, variable-dimensional problem.



Optical Multilayers (2)

Objective function:

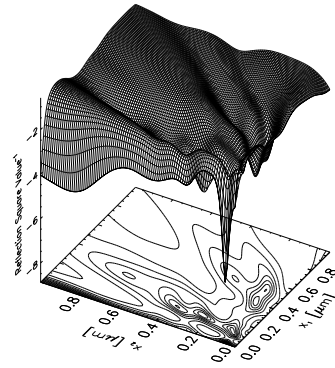
$$f(\vec{d}, \vec{\eta}) = \int_{\lambda_d}^{\lambda_u} [R(\vec{d}, \vec{\eta}, \lambda) - \tilde{R}(\lambda)]^2 d\lambda$$

- $R(\vec{d}, \vec{\eta}, \lambda)$: Reflection of the actual filter for wavelength λ . Calculation according to *matrix method*.
- $\tilde{R}(\lambda)$: Desired reflection value.

Optical Multilayers (3)

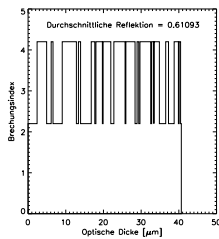
Example topology:

- $n = 2$.
- Only layer thicknesses vary.

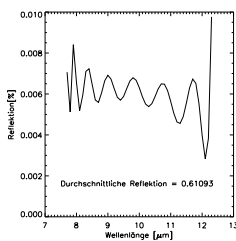


Optical Multilayers (4)

Example structure:



Example reflection:



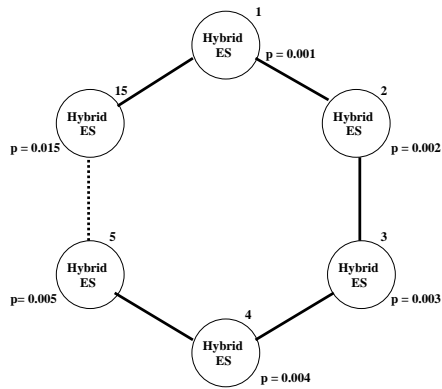
Optical Multilayers (5)

Existing Methods:

- Refinement methods:
 - Initial design constructed by an expert.
 - Local optimization of the initial design.
- Synthesis methods:
 - Without initial design (random start).
 - Automatical global optimization.

Optical Multilayers (6)

Parallel evolutionary algorithm:

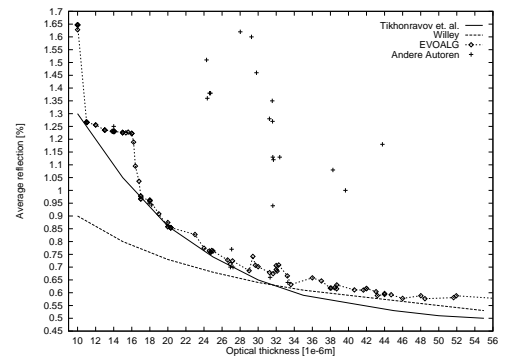


- Per node: EA for mixed-integer representation.
- *Isolation* and *migration* of best individuals.
- Mutation of discrete variables: Fixed p_m per population.



Optical Multilayers (7)

Reference Results:



Comparison of literature results, theoretical predictions, and EA results.

⇒ Excellent algorithm for the synthesis of filters.

