Evolutionary Programming (EP)

anfangs: endliche Automaten zur Symbolketten-Vorhersage

heute: numerische Optimierung (ähnlich ES)

wichtige Eigenschaften:

- evolvierende Einheit ist die Art (Spezies)
- keine Rekombination; kein Geburtenüberschuß ($\lambda = \mu$)
- normalverteilte Mutationen (ähnlich ES)
- Turnierselektion zwischen alten und neuen Spezies (Majoritätsauswahl)
- Selbstanpassung der Mutations-Varianzen (ähnlich ES)

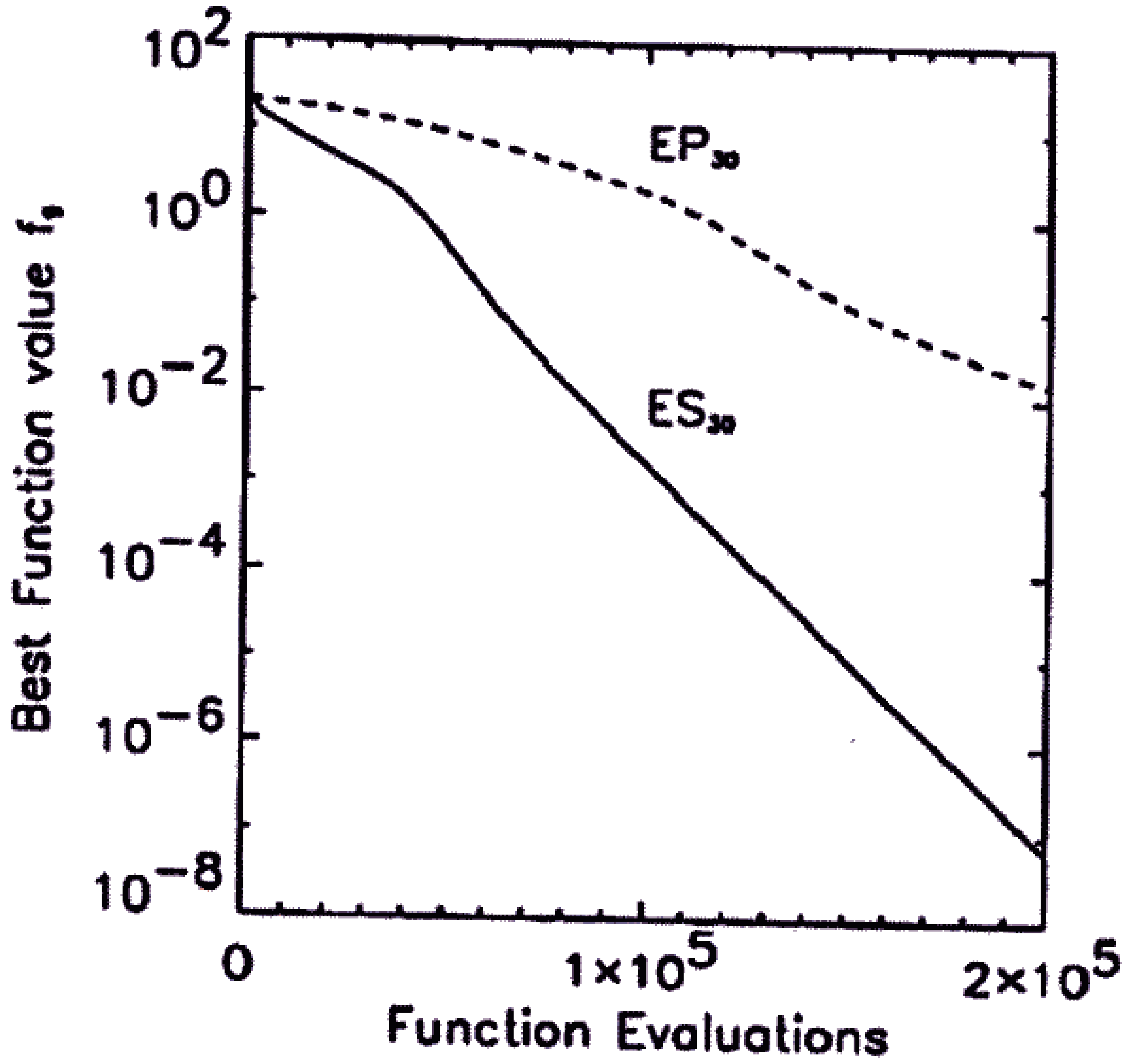
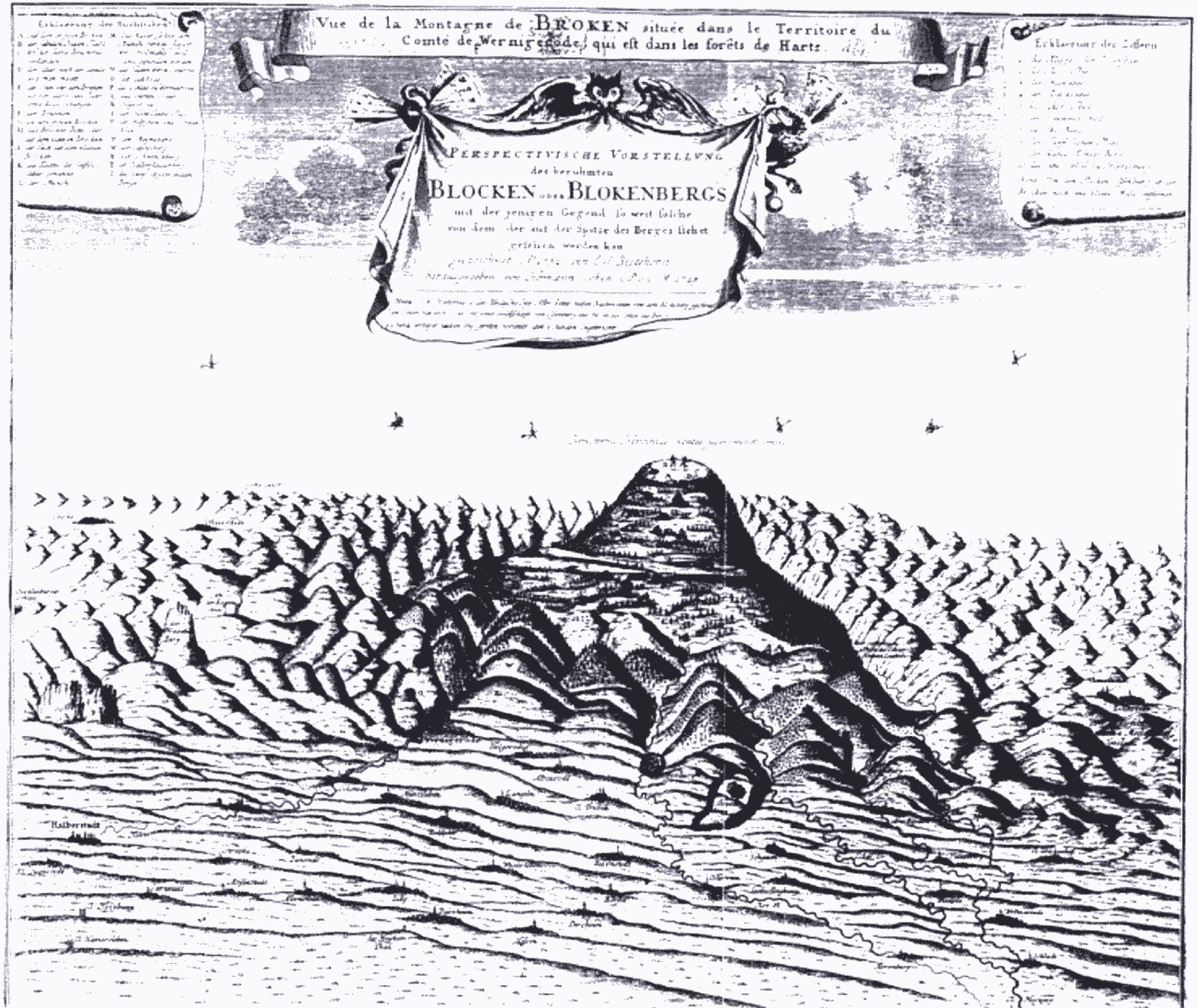


Figure 3: Experimental runs of an evolution strategy and evolutionary programming on f_9.

Ackley-Funktion


Vue de la Montagne de BROKEN située dans le Territoire du Comté de Wernigerode, qui est dans les forêts de Hartz.
PERSPECTIVISCHE VORSTELLUNG
des berühmten
BLOCKEN oder BLOKENBERGS
mit der jenigen Gegend so weit solche
von dem der auf der Spitze des Berges stehet
gesehen werden kan
Erklaerung der Zieffern
Halberstadt

Evolutionsstrategien (ES)

anfangs: experimentelles Optimierverfahren

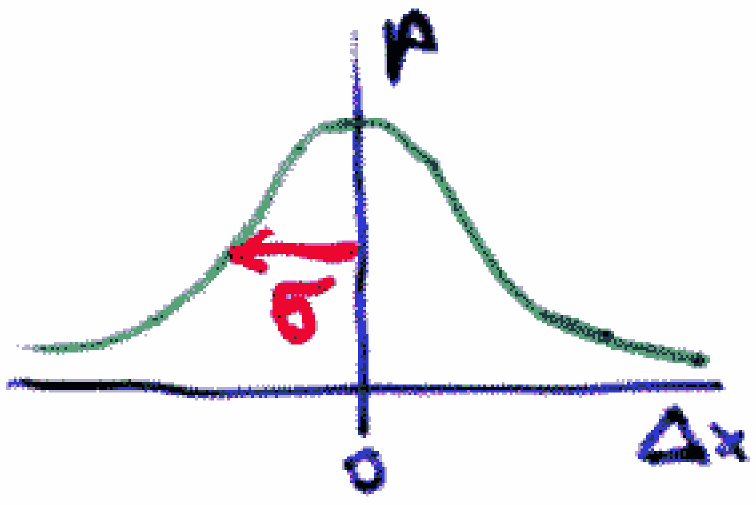
später: numerische Parameteroptimierung in Simulationsmodellen

wichtige Eigenschaften:

- evolvierende Einheit ist das Individuum (Phänotyp)
- Reproduktion mit Geburtenüberschuß ($\lambda > \mu$); rein zufällige Paarung
- Rekombination und Gaußsche Mutation als Quellen der Variation / Innovation
- Umweltselektion: Elimination der $\lambda - \mu$ Schlechtesten
- Selbstadaptation der (internen) Strategieparameter (Varianzen, Kovarianzen)
- gemischtes Diffusions-/Migrationsmodell für skalierbare Parallelität

ES Evolutionsstrategien

Mutationen: Änderungen normalverteilt (für reellwertige Variable)



$$p(\Delta x_i) = \frac{1}{\sqrt{2\pi}\,\sigma_i}\, e^{-\frac{\Delta x_i^2}{2\sigma_i^2}}$$

Rekombination:

a) intermediär — Mittelwert $\bar{x}$ aus Elternpaar.

b) diskret — Werte komponentenweise von verschied. Eltern

Reproduktion: $\lambda > \mu$ Geburtenüberschuß (à la Darwin/Malthus)

Selektion: deterministisch

μ beste aus λ — (μ,λ) ES

bzw. $\lambda+\mu$ — $(\mu+\lambda)$ ES

Selbstadaptation der σ_i

↑ werden ebenfalls mutiert!

Individuum:

$x_1, x_2, \ldots, x_n, \sigma_1, \sigma_2, \ldots, \sigma_n, \ldots$

Evolutionsstrategie (kanonisch)

Individuen : Vektor aus Objektvariablen und Strategievariablen (internes Modell)

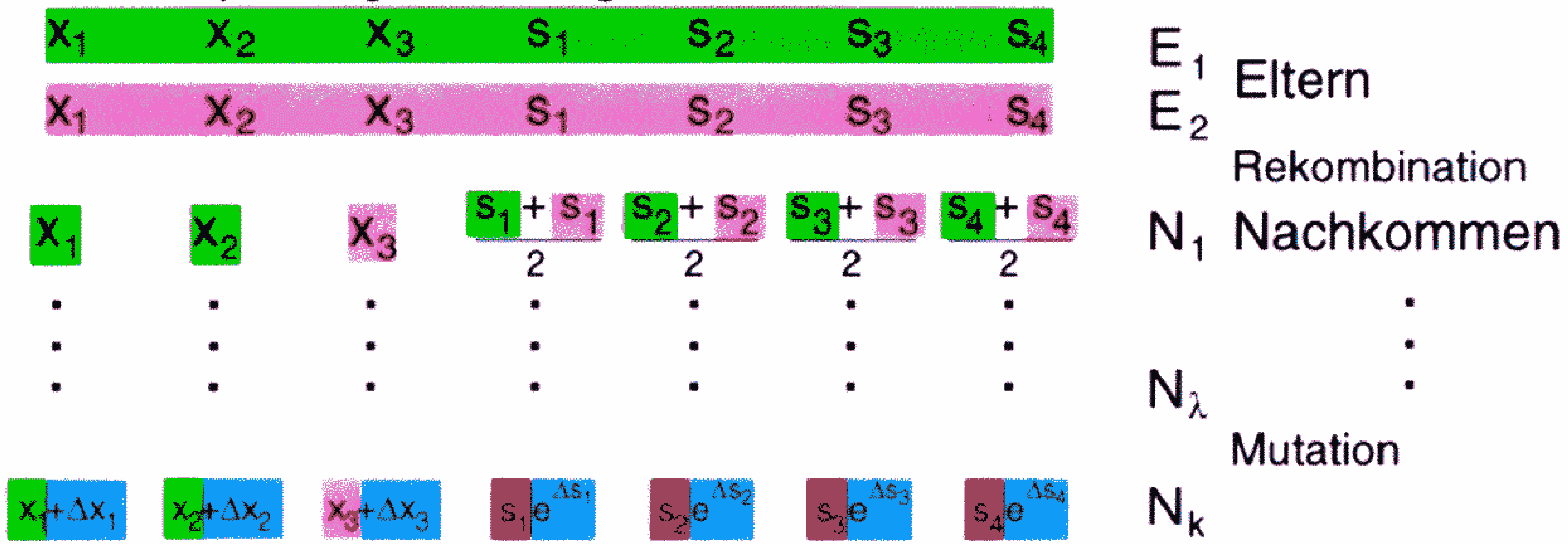
Mutation : zufällige Änderung aller Variablen

Rekombination : mittelnd oder komponentenweise

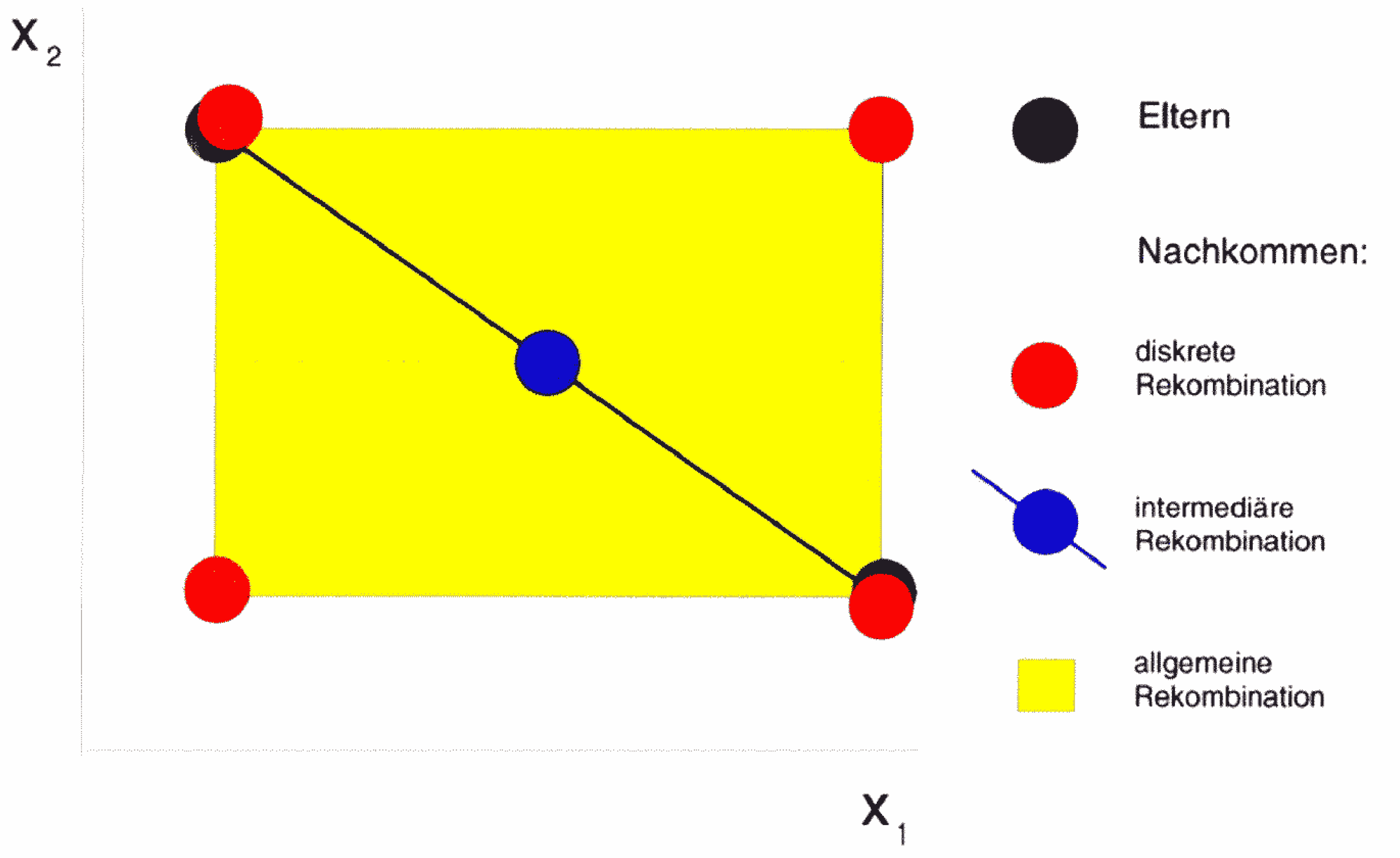
Reproduktion : $\lambda > \mu$, Geburtenüberschuß

Selektion : Die μ Besten von $\lambda(+\mu)$ werden Eltern mit gleicher Wahrscheinlichkeit (environmental selection)

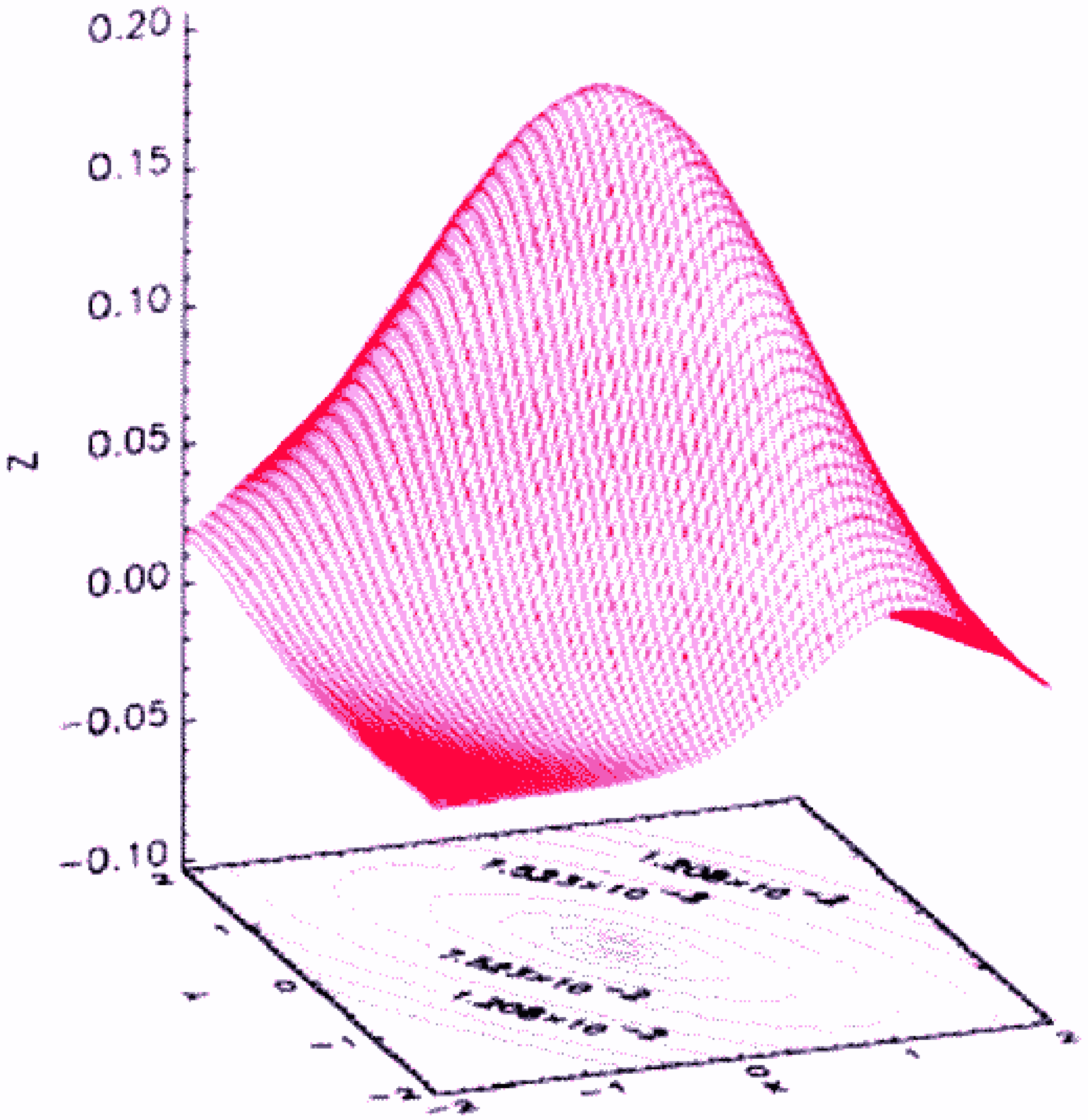
Selbstanpassung der Strategievariablen !



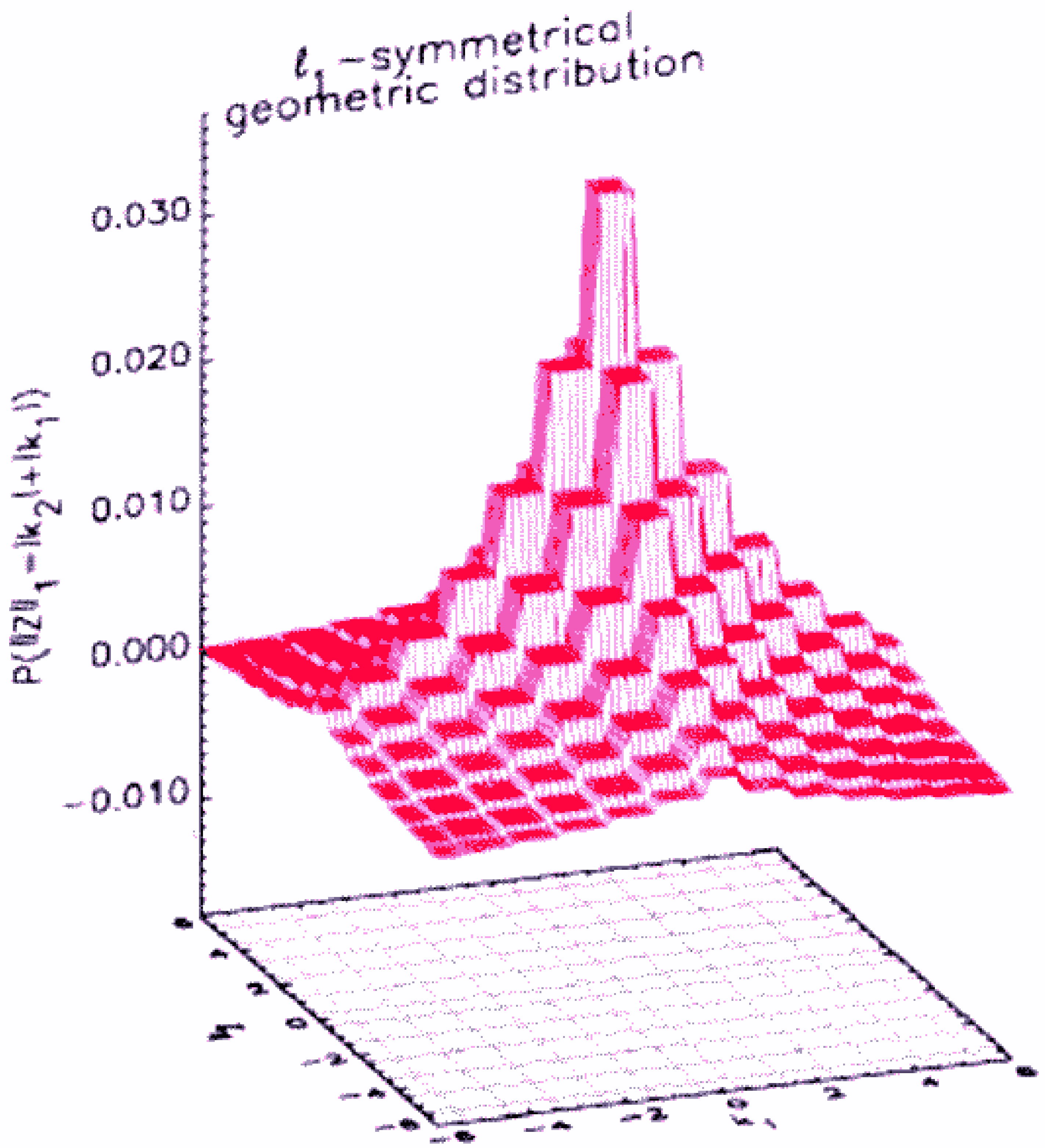
ES Rekombination



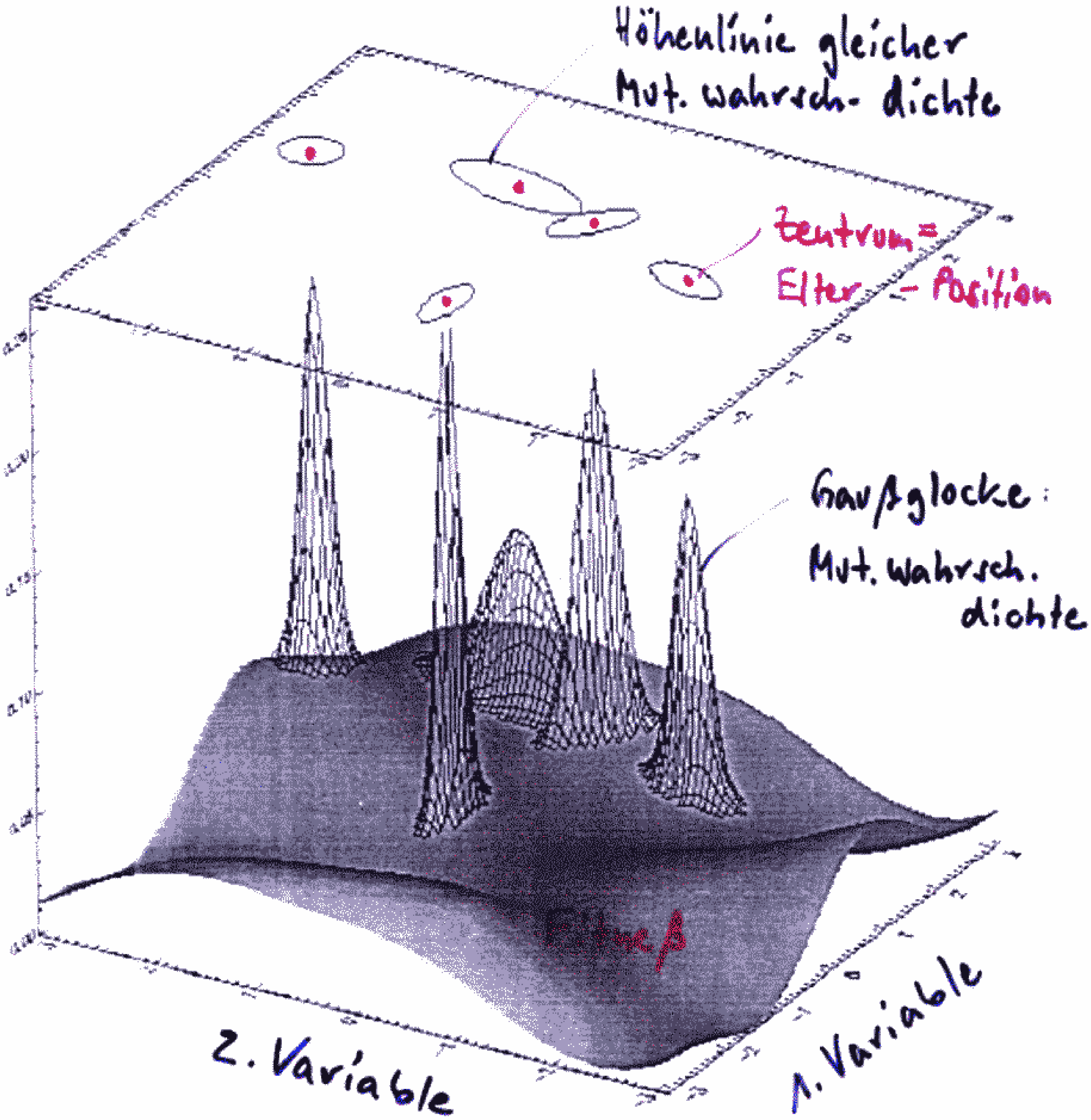
Mutationsdichteverteilung



Normalverteilung für kontinuierliche Variable



geometrische Verteilung für diskrete Variable



normalverteilte Mutationen kontinuierlich

Erwartungswert für Abweichung : $\emptyset$

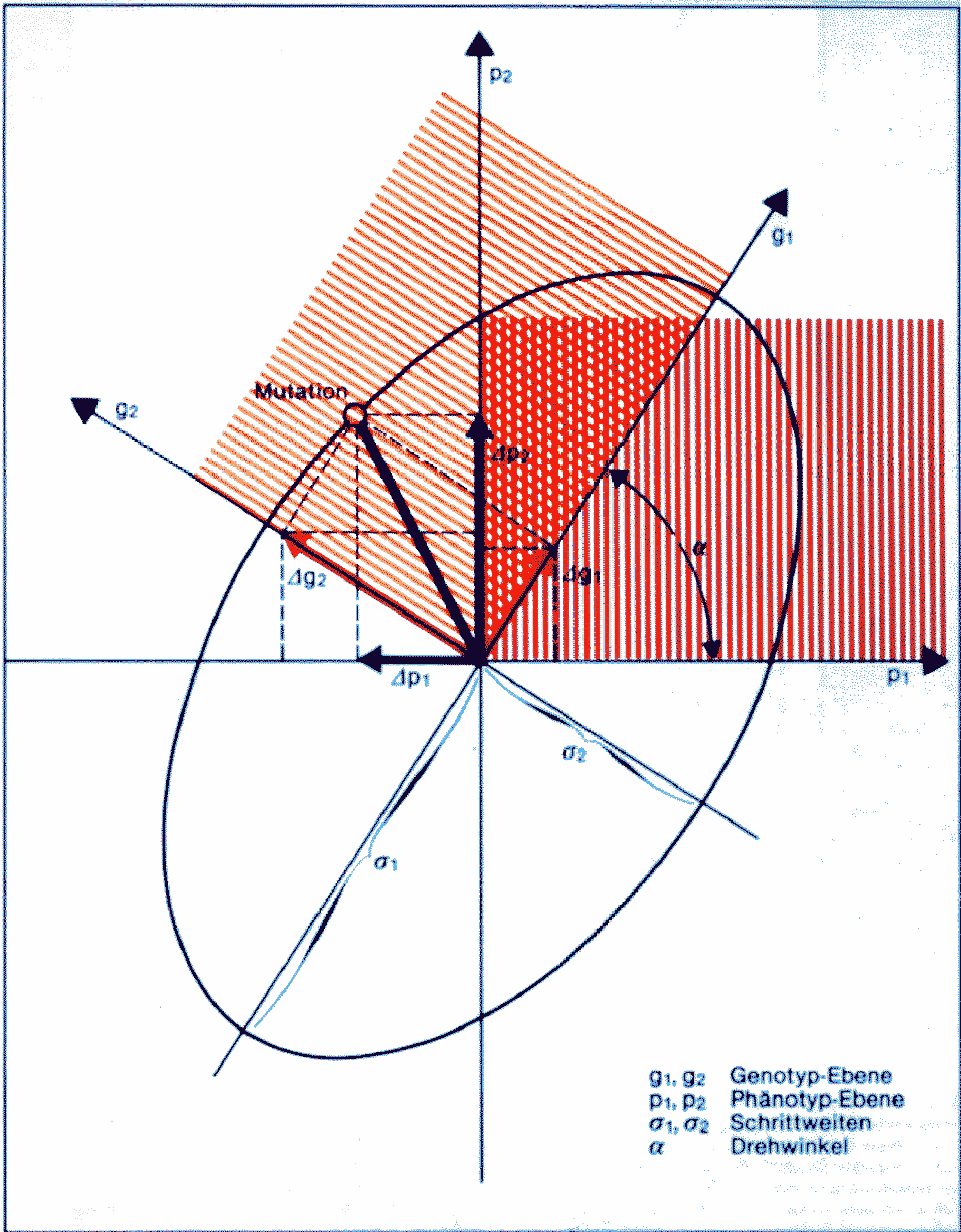
Freiheitsgrade : Varianzen σ_i^2

Kovarianzen c_{ij}

einfachster Fall : $c_{ij} = 0 \quad \forall\, i \neq j$

$\sigma_i = \sigma \quad \forall\, i = 1, 2, \ldots, n$

Polygenie, Pleiotropie: korrelierte phänotypische Veränderungen



The evolution of Evolution Strategies

Cons		Pros
two individuals	$(1+1)$ ES	June 12, 1964
just creeping random search		worked in experiments (as opposed to other methods)
μ ancestors	$(\mu+1)$ ES
rediculous waste of storage capacity		speedup through recombination
λ descendants	$(1+\lambda)$ ES
rediculous not to use new info. immediately		if parallel: speedup $\sim \log \lambda$
non-elitist	$(1,\lambda)$ ES
rediculous to forget good intermed. solution		self-adaptation of mutability works (one common σ)
$\kappa = 1$	(μ,λ) ES
rediculous to conserve inferior solutions		recombination helps auto-scaling, variable metric, and speeds up
$\kappa = \infty$	$(\mu+\lambda)$ ES
hampers self-adaptation, does not work for moving goals		elitist version – no loss of good results

$(\mu,\kappa,\lambda,\rho)$ ES — contemporary standard

$1 \le \kappa \le \infty$ max. life span (reproduction cycles)

$2 \le \rho \le \mu$ recombination uses ρ ancestors

Evolution der ES

12.6.1964	Geburt der (1+1)ES: experimentelle Optimierung wichtig: alle Variablen zugleich etwas (zufällig) verändern
1965	meine Diplomarbeit zeigt: Stagnationsgefahr bei Binomialvertlg.
1968	Düsenexperimente für die AEG
1971	Diss. Rechenberg: erste Theorie der (1+1)ES für $\mathbf{x} \in \mathbb{R}^n$ und erste ES mit Rekombination: $(\mu+1)$ES
1972-1975	DFG-Projekte: numerische ES mit autoadaptiven (Ko-)Varianzen $(1+\lambda)$ / $(1,\lambda)$ / $(\mu+\lambda)$ / (μ,λ) ES-Simulationen, z.B. (10,100)ES
1975	Diss. Schwefel: erste Theorie der $(1+\lambda)$ / $(1,\lambda)$ ES

Nomenklatur:

μ: Zahl der Eltern, λ: Zahl der Nachkommen einer Generation
Plus (+): max. Lebensdauer unendlich, Komma (,): max. 1 Generation

heute wesentlich verallgemeinerte ES-Versionen

$$(\mu,\kappa,\lambda,\rho)ES := (P^{(0)},\mu,\kappa,\lambda,\mathbf{rec},p_r,\rho,\gamma,\omega,\mathbf{mut},p_m,\tau,\tau_0,\delta,\beta,\mathbf{sel},\zeta,t,\varepsilon)$$

with

$P^{(0)} := (\mathbf{a}_1,\ldots,\mathbf{a}_\mu)^{(0)} \in I^\mu$	$I := \mathbb{N}_0 \times \mathbb{R}^n \times \times \mathbb{R}_+^{n_\sigma} \times [-\pi,\pi]^{n_\alpha}$	start population
$\mu \in \mathbb{N}$	$\mu \geq 1$	number of parents
$\kappa \in \mathbb{N}$	$\kappa \geq 1$	upper limit for life span
$\lambda \in \mathbb{N}$	$\lambda > \mu$ if $\kappa = 1$	number of offspring
$\mathbf{rec} : I^\mu \to I$		recombination operator
$p_r \in \mathbb{R}_+^3$	$0 \leq p_r \leq 1$	recombination probability
$\rho \in \mathbb{N}^3$	$1 \leq \rho \leq \mu$	number of ancestors for each descendant
$\gamma \in \mathbb{N}^3$	$1 \leq \gamma \leq n_x - 1$ $\gamma \geq \rho - 1$	number of crossover sites in a string of n_x elements
$\omega \in \{0,1,2,3,\ldots\}^3$		type of recombination
$\mathbf{mut} : I \to I$		mutation operator
$p_m \in \mathbb{R}_+^3$	$0 < p_m \leq 1$	mutation probability
$\tau, \tau_0, \delta \in \mathbb{R}_+$	$0 \leq \delta \leq 1$	step length variabilities
$\beta \in \mathbb{R}_+$	$0 \leq \beta \leq \frac{\pi}{4}$	correlation variability
$\mathbf{sel} : I^{\mu+\lambda} \to I^\mu$		selection operator
$\zeta \in \mathbb{N}$	$2 \leq \zeta \leq \mu + \lambda$	tournament participators
$t : I^{2\mu} \to \{0,1\}$		termination criterion
$\varepsilon \in \mathbb{R}_+^4$		accuracies required.

Weitere ES-Varianten

parallel

- mit Diploidie, Dominanz + Rezessivität
  für Vektoroptimierung (mehrfache Zielsetzung)
- mit Annidation / Migration
  für globale Optimierung (grobkörnig)
- mit lokaler Interaktion (Diffusion)
  für globale Optimierung (feinkörnig)

- geschachtelte ES
  für (sequentielle) globale Optimierung

$$[\mu'/\varrho', \lambda'(\mu/\varrho, \lambda)_\gamma]\text{ ES}$$

$(\mu/\varrho, \lambda)$: λ' Subpopulationen agieren γ Generationen lang unabh. voneinander

$[\mu'/\varrho', \lambda'$: Auslese unter Subpopulationen

Notation

- Individual: $\vec{a} \in I$.
- Individual space: $I = I\!R^n \times S$.
- Strategy parameters: S.
- Population: $P^{(t)} = \{\vec{a}_1, \ldots, \vec{a}_n\} \in I^k$, $k \in \{\mu, \lambda\}$.
- Recombination: $\mathbf{rec} : I^\mu \to I$.
- Mutation: $\mathbf{mut} : I \to I$.
- Selection: $\mathbf{sel}_\mu^k : I^k \to I^\mu$, $k \in \{\lambda, \mu + \lambda\}$.
- Generation transition: $opt_{ES}(P^{(t)}) : I^\mu \to I^\mu$

$$\boxed{opt_{ES}(P^{(t)}) = \mathbf{sel}_\mu^k(\sqcup_{i=1}^{\lambda}\{\mathbf{mut}(\mathbf{rec}(P^{(t)}))\} \sqcup Q)}$$

  $(k = \lambda + |Q|,\ Q \in \{P^{(t)}, \emptyset\})$.

- Realization of a normally distributed random variable: $z \sim \mathrm{N}(\zeta, \sigma^2)$.

Random number generation:

$$z_1 = \sqrt{-2\ln u_1}\,\sin(2\pi u_2) \quad , \quad z_2 = \sqrt{-2\ln u_1}\,\cos(2\pi u_2)$$

$u_1, u_2 \sim \mathrm{U}((0,1])$, $z_1, z_2 \sim \mathrm{N}(0,1)$.

- $\mu = \lambda = 1$, no recombination.
- Individual structure: $\vec{a} = (\vec{x}, \sigma) \in I\!R^n \times I\!R_+$.
- Mutation: 1/5-success rule

$$\tilde{\sigma} := \mathbf{mu}_\sigma(\sigma) = \begin{cases} \sigma / \sqrt[n]{c} & , \text{ if } p > 1/5 \\ \sigma \cdot \sqrt[n]{c} & , \text{ if } p < 1/5 \\ \sigma & , \text{ if } p = 1/5 \end{cases}$$

$$\tilde{\vec{x}} := \mathbf{mu}_x(\vec{x}) = (x_1 + z_1, \ldots, x_n + z_n) .$$

  - p denotes the measured relative frequency of successful mutations.
  - $c = 0.817$ proposed by Schwefel.
  - $z_i \sim N_i(0, \tilde{\sigma}^2)$.

- Selection:

$$\mathbf{sel}_1^2(\{\vec{a}, \tilde{\vec{a}}\}) = \begin{cases} \{\tilde{\vec{a}}\} & , \text{ if } f(\tilde{\vec{x}}) \le f(\vec{x}) \\ \{\vec{a}\} & , \text{ otherwise} \end{cases} .$$

- Generation transition:

$$\boxed{opt_{(1+1)-ES}(\{\vec{a}\}) = \mathbf{sel}_1^2(\{\mathbf{mut}(\vec{a})\} \sqcup \{\vec{a}\})}$$

Pseudocode for a (1 + 1)-ES

ALGORITHM 1 ((1 + 1)-ES)

$t := 0;$
initialize $P^{(t)} = \{(\vec{x}, \sigma)\};$
evaluate $f(\vec{x});$
while $(T(P^{(t)}) = 0)$ **do**
    $(\tilde{\vec{x}}, \tilde{\sigma}) :=$ **mut**$((\vec{x}, \sigma));$
    evaluate $f(\tilde{\vec{x}});$
    if $(f(\tilde{\vec{x}}) \leq f(\vec{x}))$
        then $P^{(t+1)} := \{(\tilde{\vec{x}}, \tilde{\sigma})\};$
        else $P^{(t+1)} := P^{(t)};$
    $t := t + 1;$
od

σ-modification according to Schwefel:

> After every n mutations, check how many successes have occurred over the preceding $10 \cdot n$ mutations. If this number is less than $2 \cdot n$, multiply the step length by the factor $c = 0.85$; divide it by 0.85 if more than $2 \cdot n$ successes occurred.

The (μ,λ)-ES

One generation transition function ((μ,λ)-selection):

$$opt_{(\mu,\lambda)-ES}(P^{(t)}) = \mathbf{sel}_\mu^\lambda(\sqcup_{i=1}^\lambda \{\mathbf{mut}(\mathbf{rec}(P^{(t)}))\})$$

ALGORITHM 2 ((μ, λ)-ES)

$t := 0$;
initialize $P^{(0)} = \{\vec{a}_1, \ldots, \vec{a}_\mu\} \in I^\mu$;
evaluate $f(\vec{x}_1), \ldots, f(\vec{x}_\mu)$;
while $(T(P^{(t)}) = 0)$ **do**
  $\tilde{P} := \emptyset$;
  for $i := 1$ **to** λ **do**
    $(\tilde{\vec{x}}, \tilde{\vec{\sigma}}, \tilde{\vec{\alpha}}) :=$ **mut(rec(**$P^{(t)}$**))**;
    evaluate $f(\tilde{\vec{x}})$;
    $\tilde{P} := \tilde{P} \sqcup \{(\tilde{\vec{x}}, \tilde{\vec{\sigma}}, \tilde{\vec{\alpha}})\}$;
  od
  $P^{(t+1)} :=$ $\mathbf{sel}_\mu^\lambda(\tilde{P})$;
  $t := t + 1$;
od

For $(\mu + \lambda)$-selection:

$$opt_{(\mu+\lambda)-ES}(P^{(t)}) = \mathbf{sel}_\mu^{\mu+\lambda}(\sqcup_{i=1}^\lambda \{\mathbf{mut}(\mathbf{rec}(P^{(t)}))\} \sqcup P^{(t)}) \quad .$$

Mutations

Normally distributed variations:

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

- Expectation $\xi = 0$ is assumed.
- Standard deviation σ must be adapted.

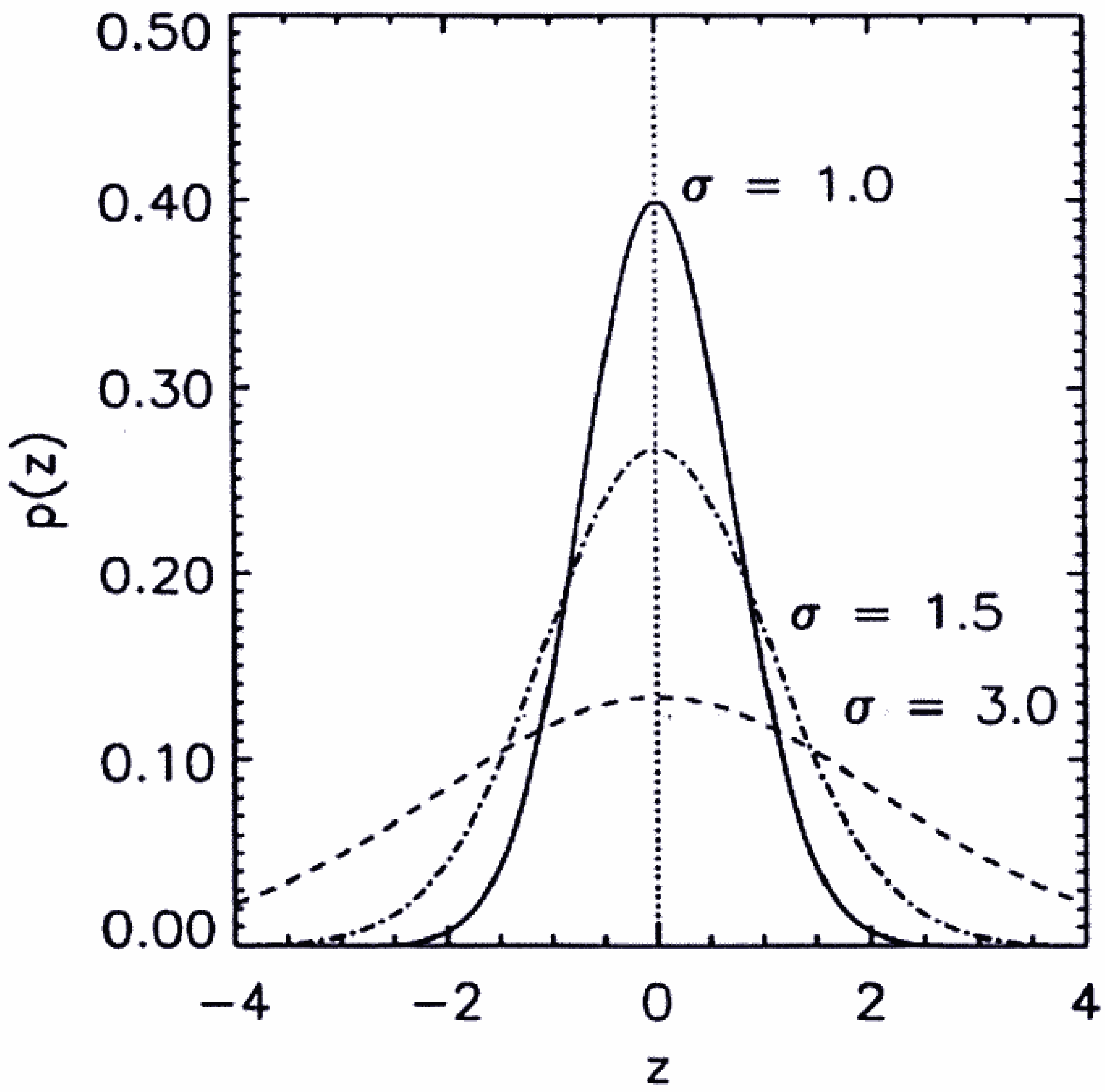


Figure 6: Probability density function $p(z)$ for a normal distribution with $\xi = 0$ and standard deviations $\sigma \in \{1, 1.5, 3\}$.

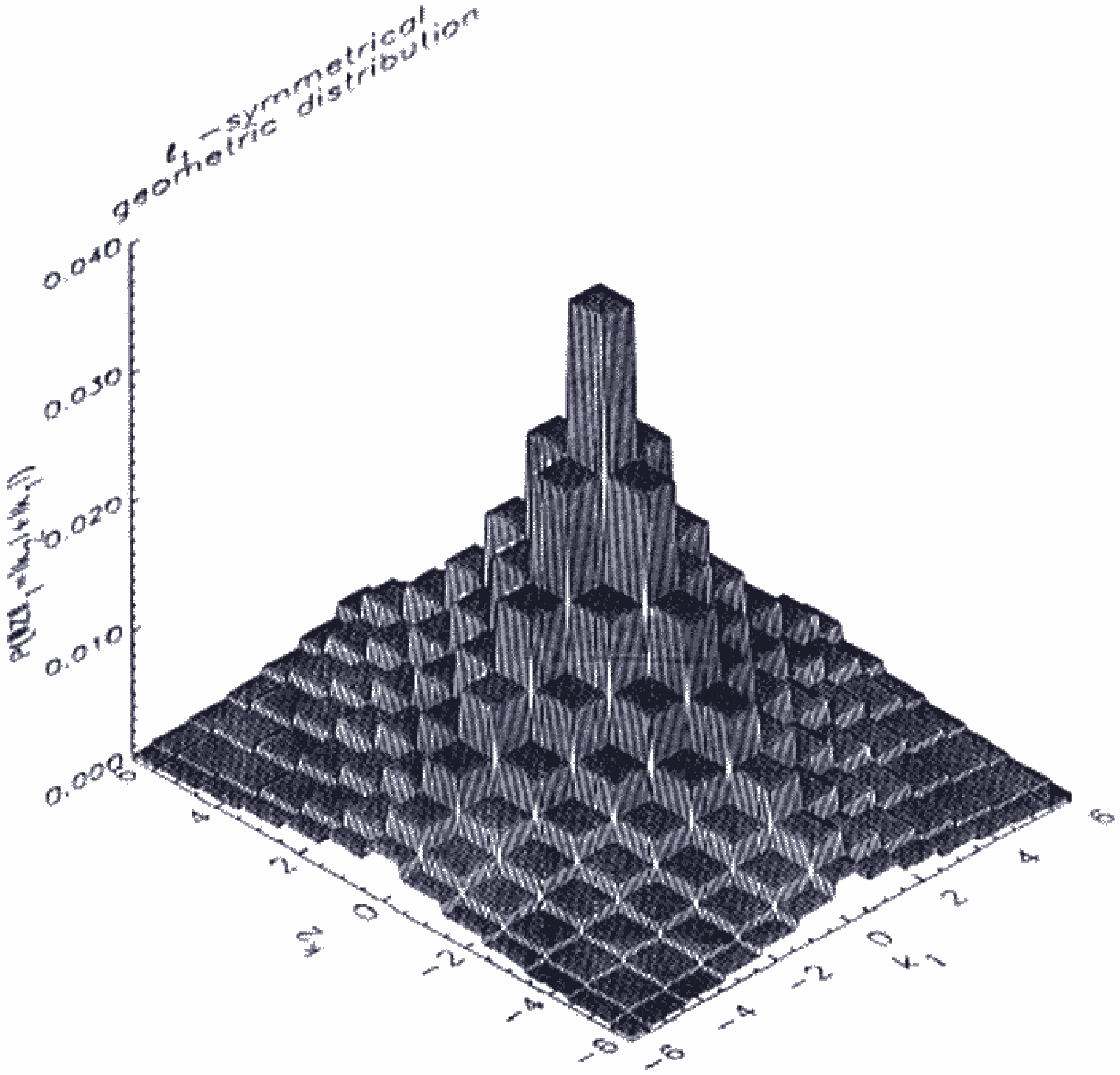


Figure 1: Maximum-entropy probability distribution in $\mathbb{Z}^2$ for $E[\|Z\|_1] = 5$.

Simple Mutations: $n_\sigma = 1$

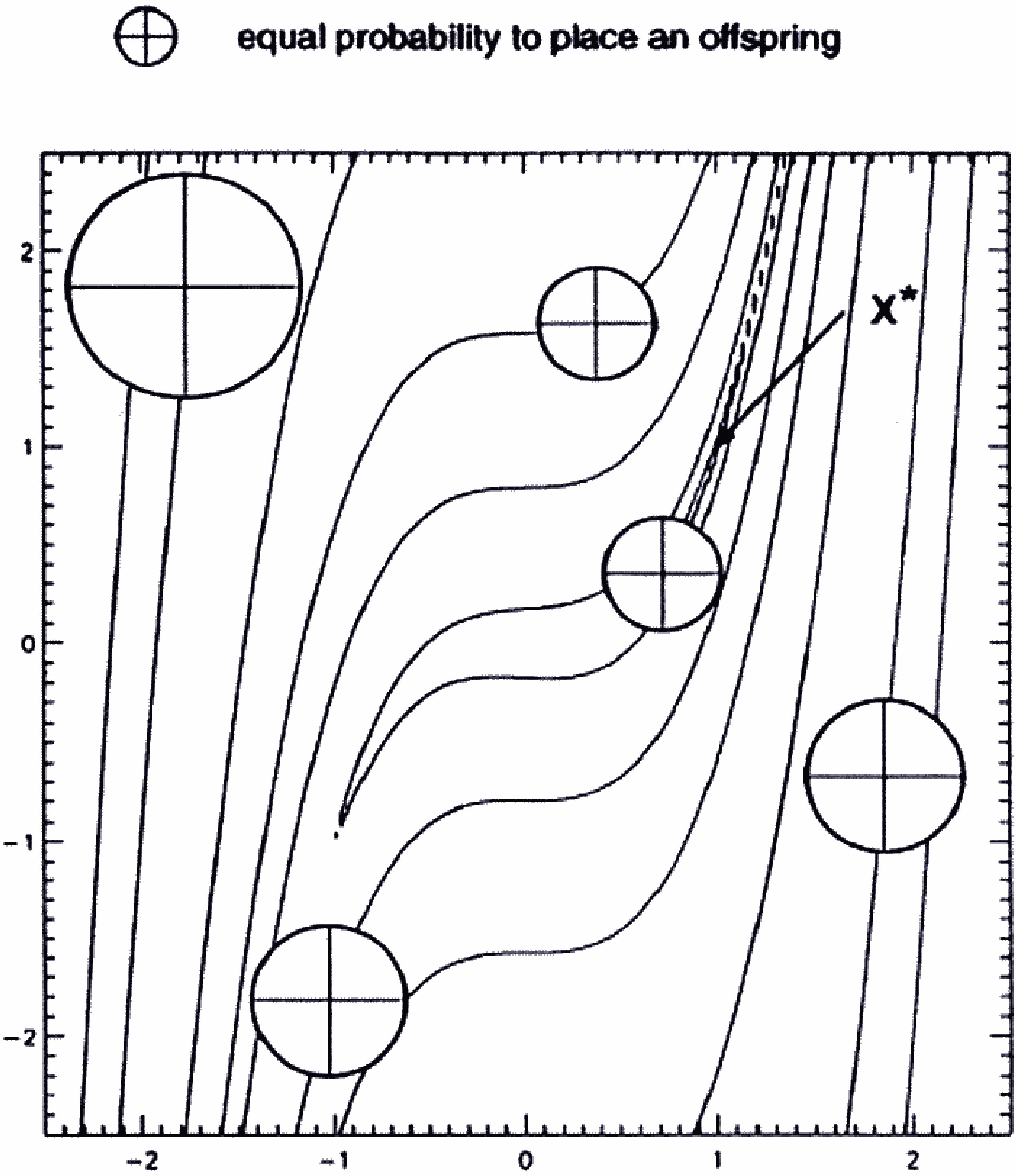
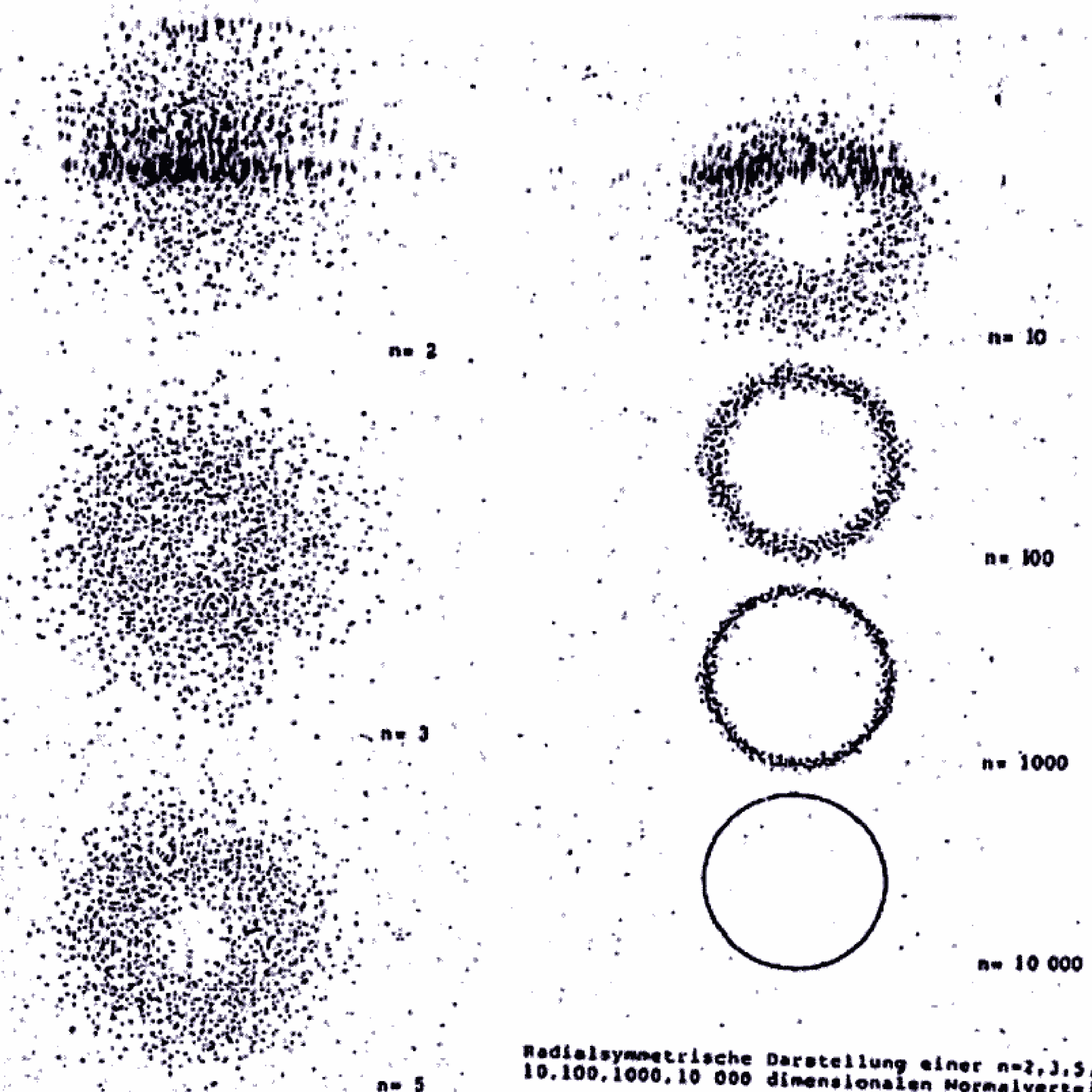


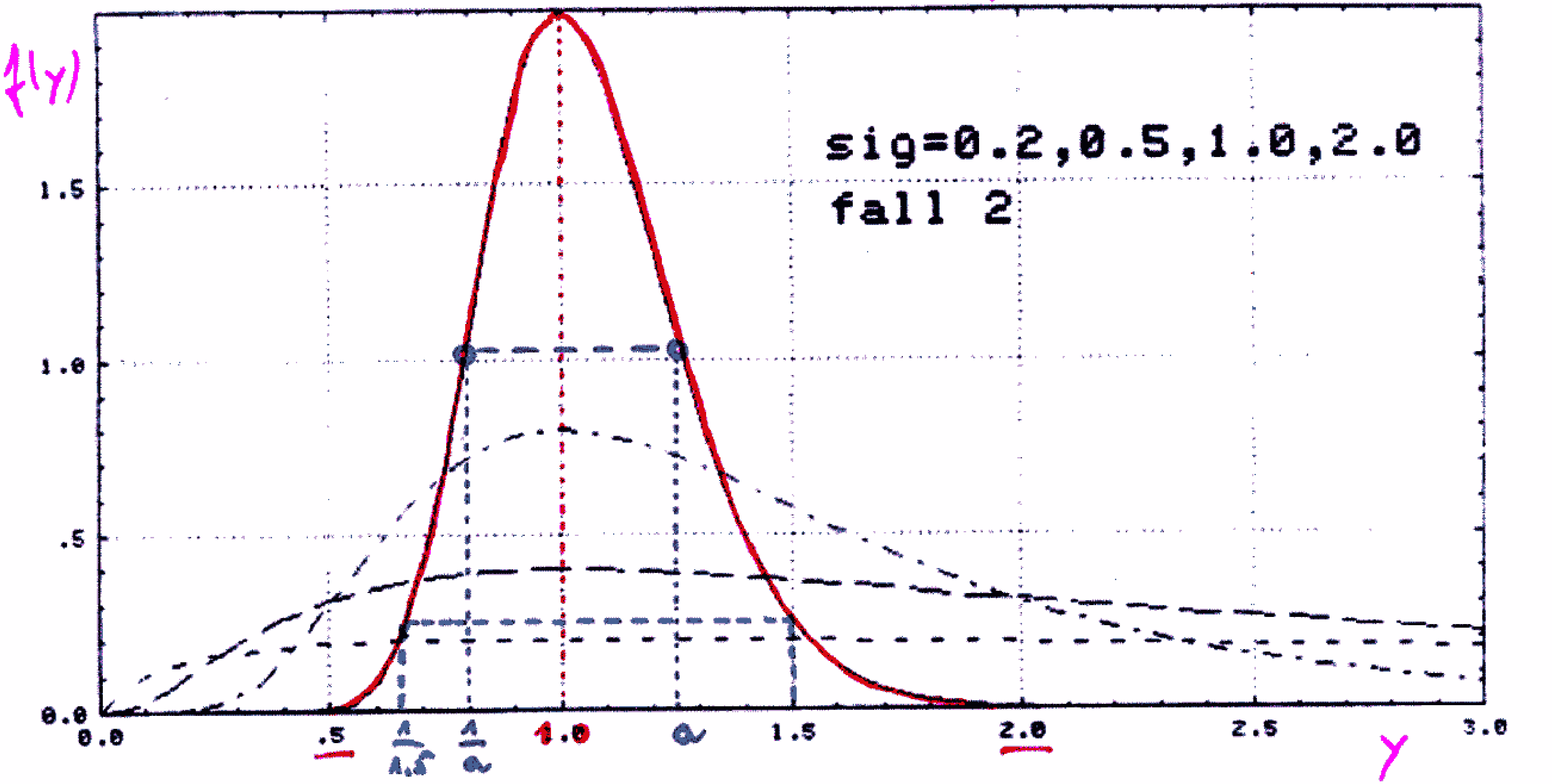
Figure 7: Simple mutations, $n = 2$, $n_\sigma = 1$, ($\Rightarrow n_\alpha = 0$).

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

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



Radialsymmetrische Darstellung einer n=2,3,5, 10,100,1000,10 000 dimensionalen Normalverteilung 'neutraler' Zufallsmultiplikator



log. norm. y · f(y)

z sei (0, σ) normalverteilt

dann ist

$y = e^z$ logarithmisch normalverteilt, wahrscheinlichster Wert $e^0 = 1$

An alternative method for $n_\sigma = 1$:

(proposed by Rechenberg; see e.g. (Ostermeier 1992)).

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

$$\begin{aligned} \sigma' &= \sigma \cdot u \\ x'_i &= x_i + \frac{\sigma'}{\sqrt{n}} \cdot N_i(0,1) \end{aligned}$$

- $u \in \{1, \alpha, 1/\alpha\}$ with equal probability for each of the three outcomes $\Rightarrow$
  - One-third of the offspring tries larger step-sizes,
  - one-third of the offspring tries smaller step-sizes, and
  - one-third of the offspring keeps step-sizes constant.
- Normally, $\alpha = 1.5$ is used (Ostermeier, 1992).
- Terminology: *Mutational stepsize control* (Rechenberg).
- Extension proposed by (Ostermeier 1992): *Momentum adaptation*.

  (Allow for an adaptation of expected values *different from zero* for the normally distributed random numbers).

Standard Mutations: $n_\sigma = n$

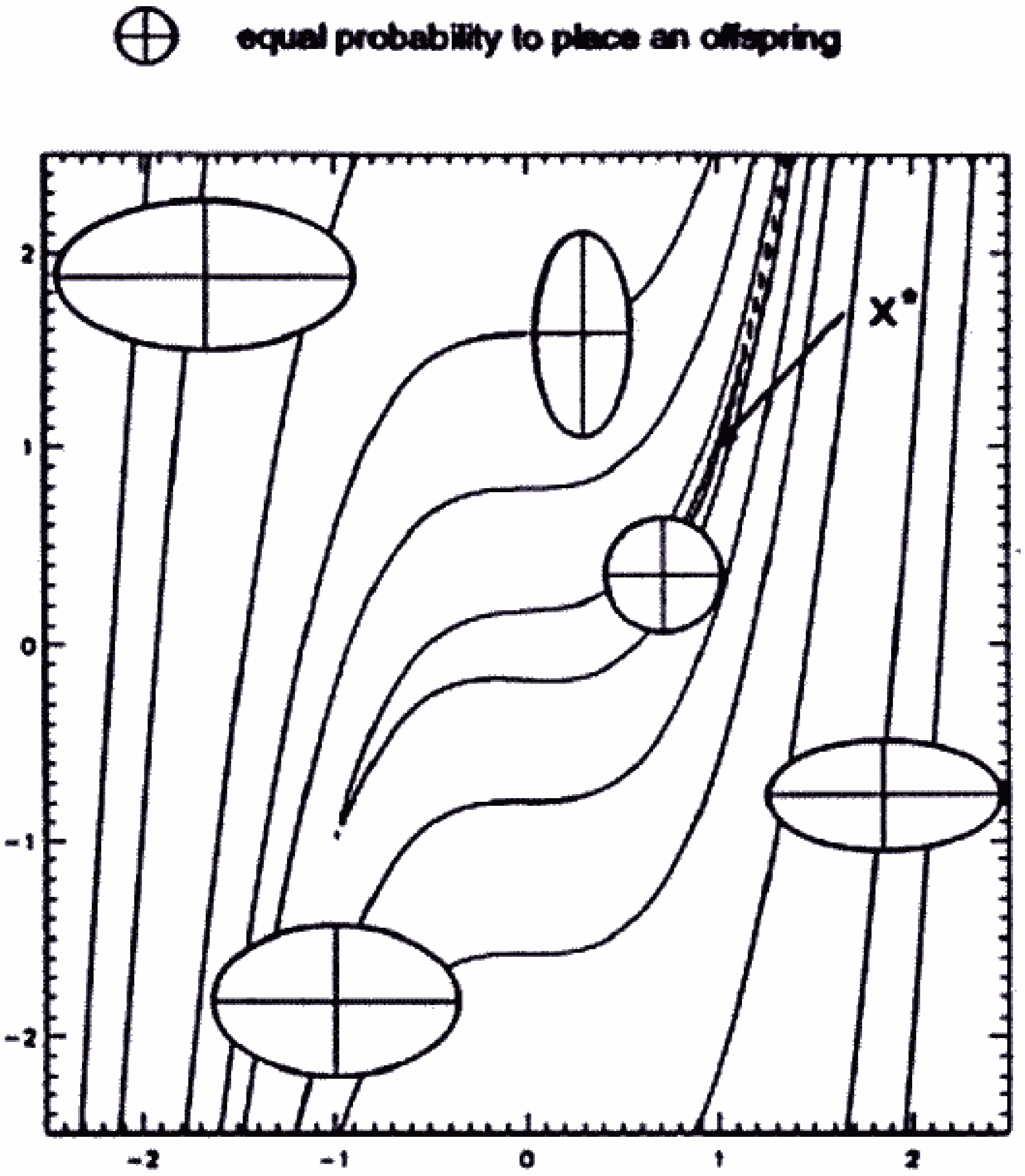


Figure 8: Simple mutations, $n = 2$, $n_\sigma = 2$, $(n_\alpha = 0)$.

$$
\begin{aligned}
I &= I\!R^n \times I\!R^n_+ \\
m'_{\{\tau,\tau'\}}(\vec{x},\vec{\sigma}) &= (\vec{x}',\vec{\sigma}') \\
\tau &\sim 1/\sqrt{2\sqrt{n}} \quad \cdot c_{\mu,\lambda,\varrho,\kappa} \\
\tau' &\sim 1/\sqrt{2n} \quad \cdot c_{\mu,\lambda,\varrho,\kappa}
\end{aligned}
$$

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

Boundary rule for preserving standard deviations larger than zero:

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

Correlated Mutations

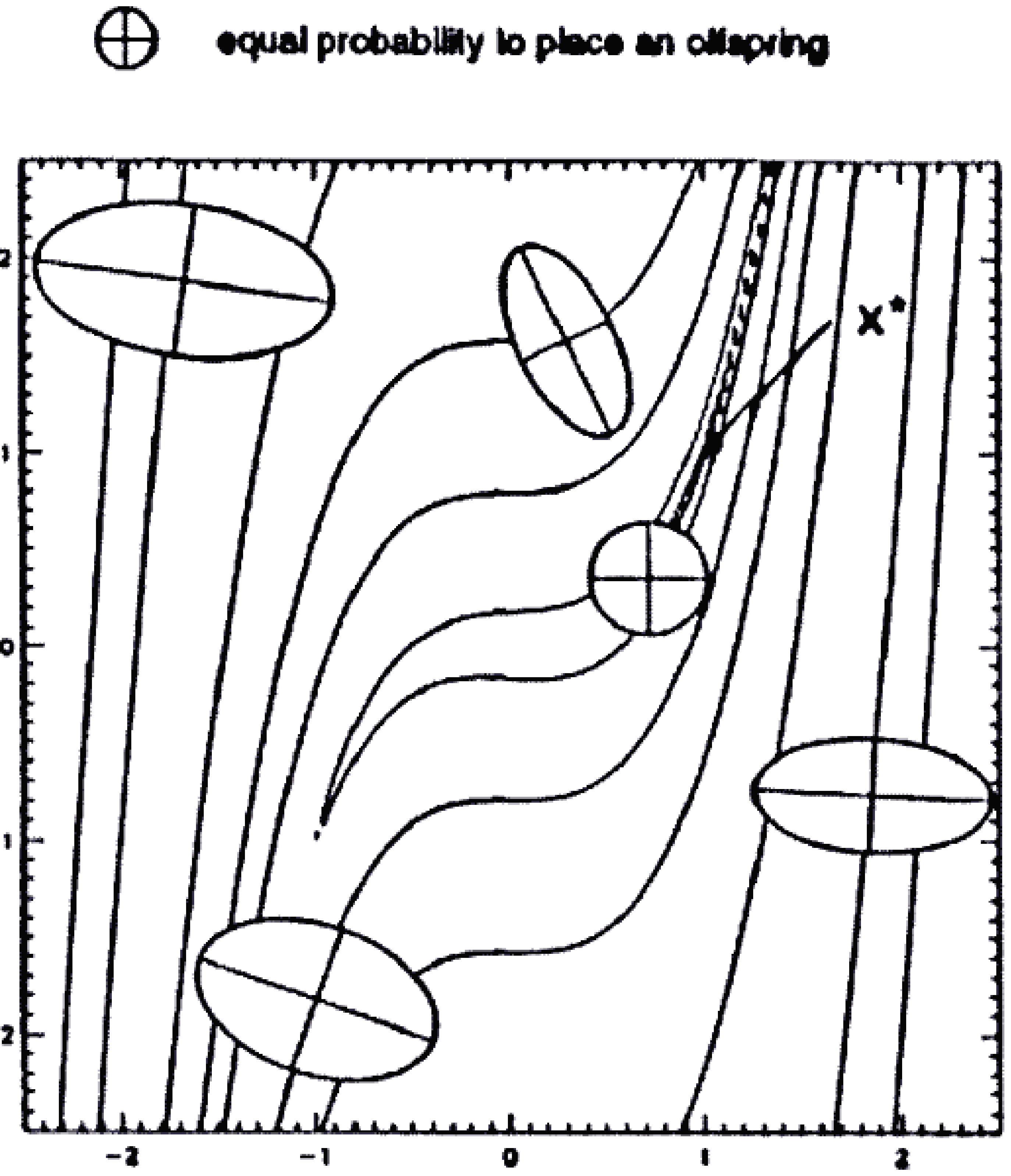


Figure 9: Correlated mutations, $n = 2$, $n_\sigma = 2$, $n_\alpha = 1$.

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

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

Boundary rule for keeping rotation angles feasible:

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

⇒ A strategy parameter set

- is part of each individual,
- represents the p.d.f. for mutation of the individual:

$$p(\vec{z}) = \sqrt{\frac{\det\mathbf{C}}{(2\pi)^n}} \exp\left(-\frac{1}{2}\vec{z}^T\mathbf{C}\vec{z}\right) \quad .$$

- $\mathbf{C}^{-1}$: Covariance matrix:

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

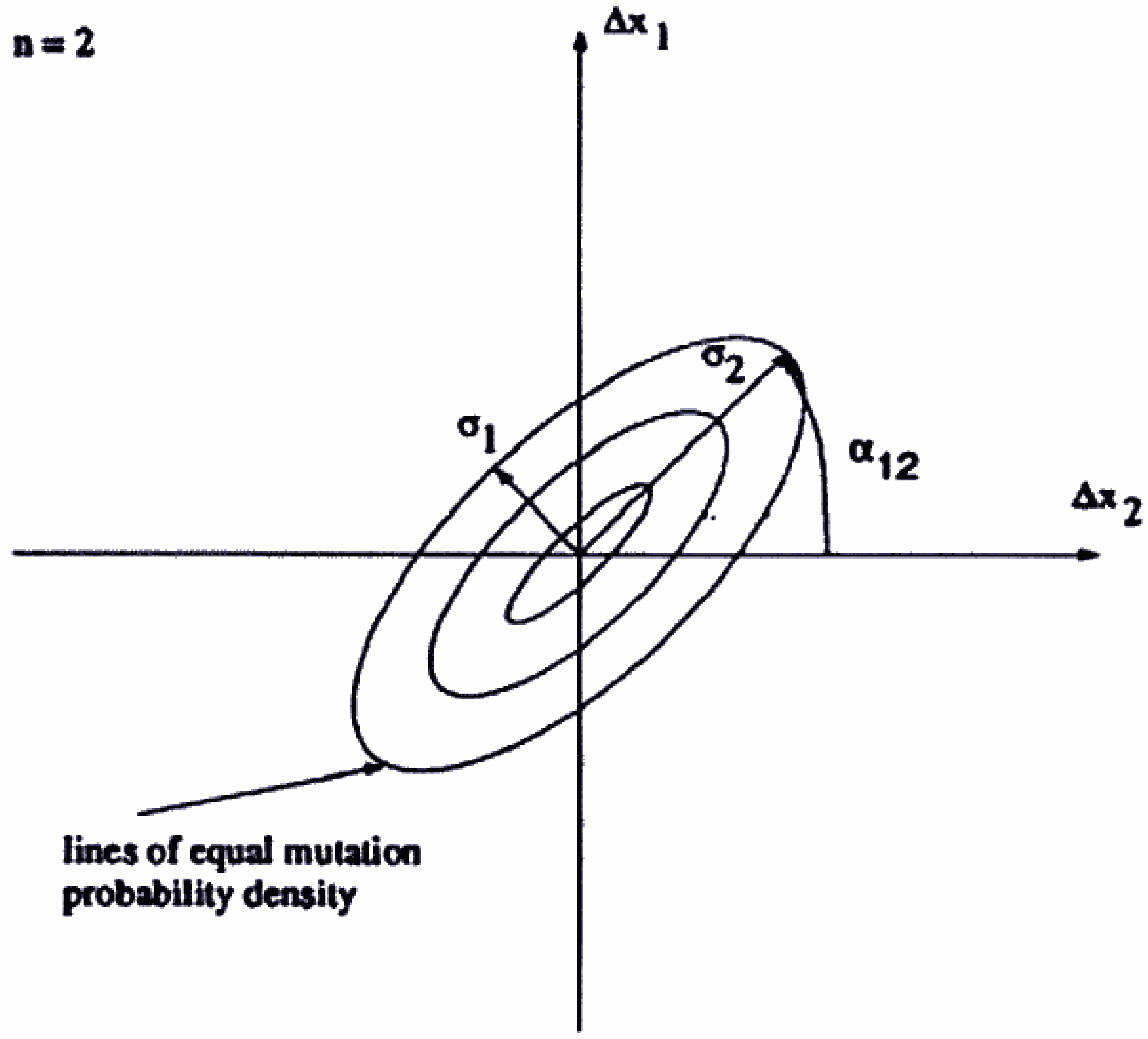


Figure 5: Illustration of the mutation ellipsoid for the case $n = 2$, $n_\sigma = 2$, $n_\alpha = 1$.

Structure of Individuals

- Individual space

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

- Strategy parameters

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

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

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

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

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

n_σ	n_α	Remark
1	0	standard mutations
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)

Table 1: Possible settings of n_σ and n_α.

- If $1 < n_\sigma < n$: All x_i ($i > n_\sigma$) are mutated according to σ_{n_σ}.
- E.g.: $n_\sigma = 2$, $n_\alpha = n - 1$ facilitates learning of one preference direction.

Some remarks:

- Biological model:
  Repair enzymes, mutator genes.
- There exists *no deterministic control*.
  Instead, strategy parameters evolve as object variables do.
- There is only an *indirect link* between fitness and useful strategy parameter settings.
- $\vec{\sigma}$, $\vec{\alpha}$ are conceivable as an *internal model* of the local topology.
- Standard strategy: $n_\sigma = n$, $n_\alpha = 0$.
- For correlated mutations:
  - $\vec{\sigma}_c \sim \vec{N}(\vec{0}, \mathbf{C})$ is generated by a multiplication of the uncorrelated random vector $\vec{\sigma}_u$ by n_α rotation matrices (Schwefel 1981, Rudolph 1992).

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

  - Exactly the feasible (positive definite) correlation matrices $\mathbf{C}$ can be created this way (Rudolph 1992).

rec : $I^\mu \to I$, **rec** = **re** ∘ **co**

- Creates one individual per application.
- Works according to:
  - **co** : $I^\mu \to I^\varrho$
    Chooses ϱ parents at random. e.g. $\varrho = 2$
  - **re** : $I^\varrho \to I$
    Creates one offspring individual.
- Common cases:
  - $\varrho = 2$ bisexual recombination.
  - $\varrho = \mu$ global recombination ("gene pool recombination").

Recombination types:

- No recombination, i.e., **rec** performs just a random choice of an individual.
- Global intermediary recombination:

$$b'_i = \frac{1}{\varrho} \sum_{k=1}^{\varrho} b_{k,i}$$

  (Averaging over all parents).
- Local intermediary recombination:

$$b'_i = u_i b_{k_1,i} + (1 - u_i) b_{k_2,i}$$

  $u_i \sim \mathrm{U}([0,1])$ or $u_i = 1/2$
  $k_1, k_2 \sim \mathrm{U}(\{1, \ldots, \varrho\})$ for each offspring.
- Discrete recombination:

$$b'_i = b_{k_i,i}$$

  $k_i \sim \mathrm{U}(\{1, \ldots, \varrho\})$ at random for each i.

b stands for $\{x, \sigma, \alpha\}$

Recombination example

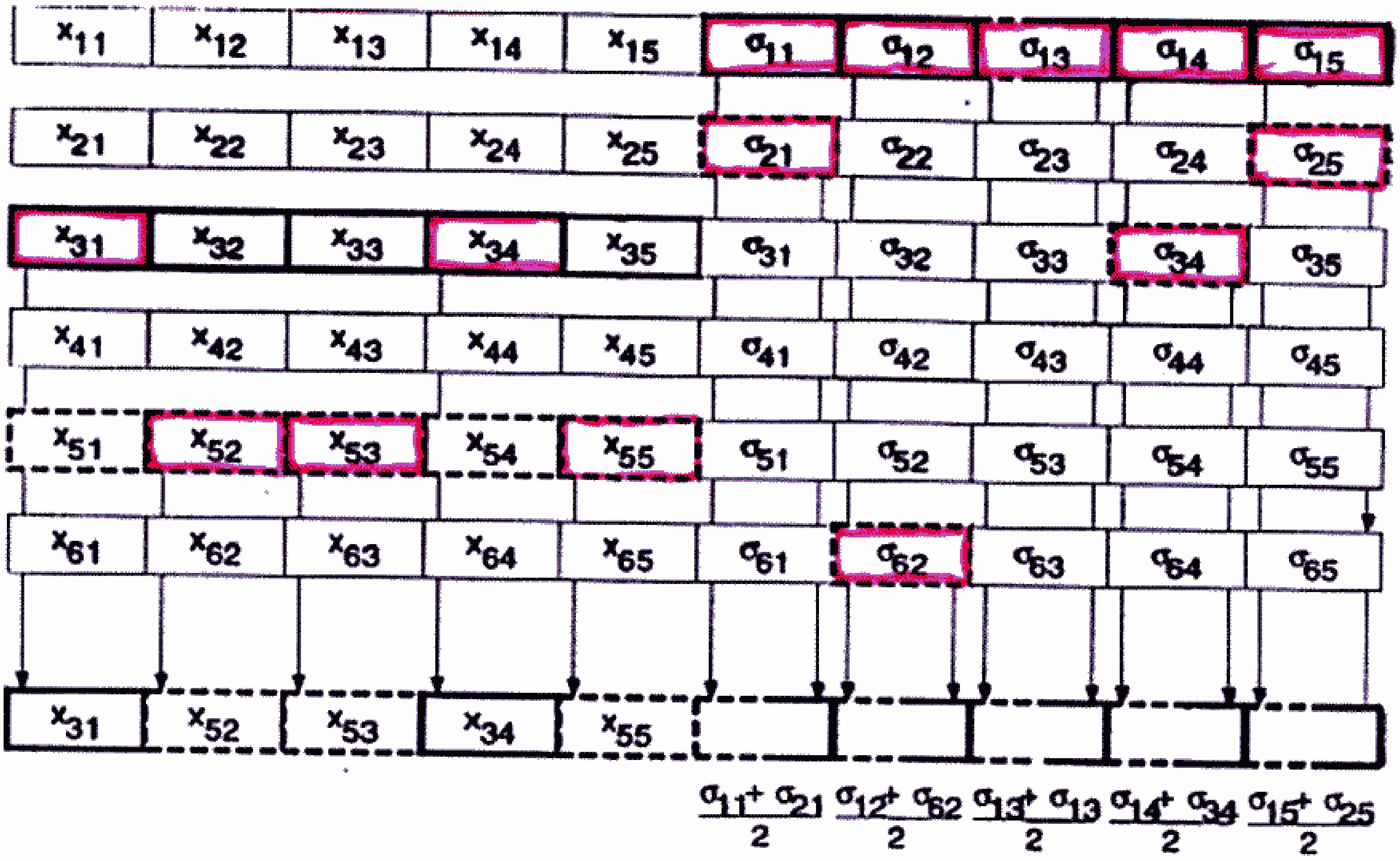


Figure 1: Recombination example for $n = 5$, $\mu = 6$, discrete recombination on x_i and global intermediate recombination on σ_i. The process is performed λ times for the creation of a complete offspring population.

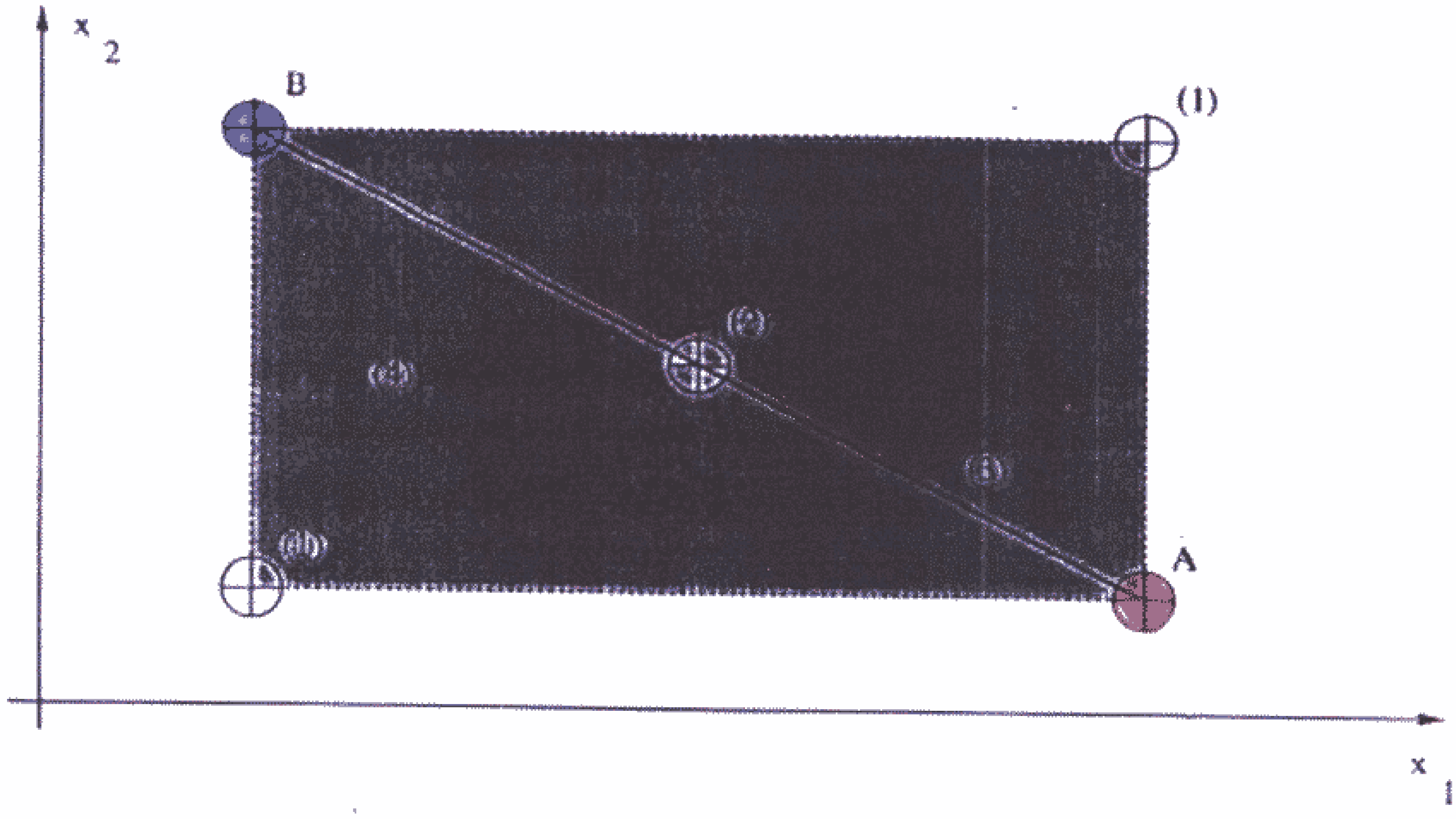
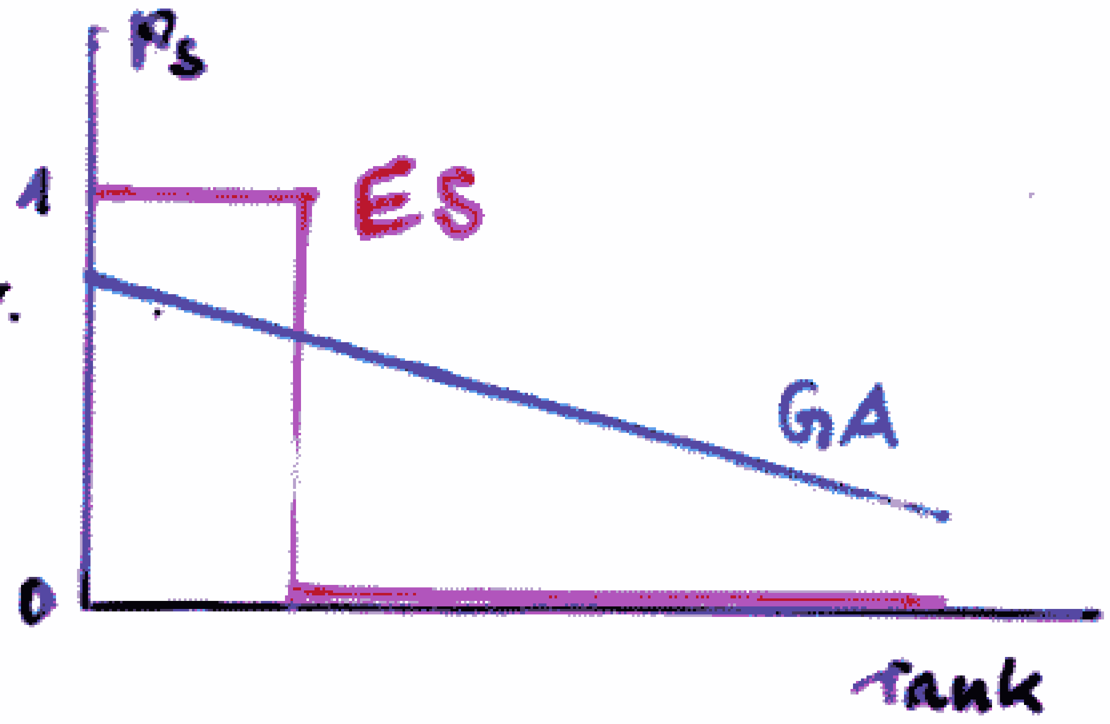


Figure 2: Illustration of the possible locations of recombination results according to the selected recombination scheme ($n = 2$). Notice that global random intermediate recombination is the most general mechanism.

Selection



- Strictly deterministic, rank-based.
- The μ best ranks are handled equally.
- Two forms:
  - $(\mu+\lambda)$-selection:

    The μ best of offspring and parents survive.
  - (μ,λ)-selection:

    The μ best offspring survives.

Formally: $sel_{\mu}^{k}(P) = \tilde{P}$, where $|\tilde{P}| = \mu$, $|P| = k \geq \mu$, and

$$\forall \tilde{\vec{a}} \in \tilde{P} : \nexists \vec{a} \in P - \tilde{P} : \quad f(\vec{x}) \leq f(\tilde{\vec{x}})$$

- Default: (μ,λ)-selection (e.g., (15,100)).
  - Important for self-adaptation.
  - Applicable also for noisy objective functions, moving optima.
- Selective pressure: Very high.

In terms of the *takeover time* τ^* (number of generations required until repeated application of selection completely fills the population with copies of the best individual; Goldberg & Deb 1991):

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

$\tau^* \approx 2$ generations for a (15,100)-ES / $\tau^* \approx 458$ for a (99,100)-ES

(cf. proportional selection in GAs: $\tau^* \approx \lambda \ln \lambda = 460$ generations !).

Self-Adaptation

The critical claim (Schwefel 1987, 1992): 1974/75

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:

- 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); $\lambda/\mu \approx 7$ turns out to be useful, but μ has to be clearly larger than 1),
- recombination also on strategy parameters (especially: intermediate recombination).

An illustration of self-adaptation:

Sphere model, (15,100)-ES with $n = 30$, $n_\sigma = 1$, and the optimum location is changed every 200 generations.

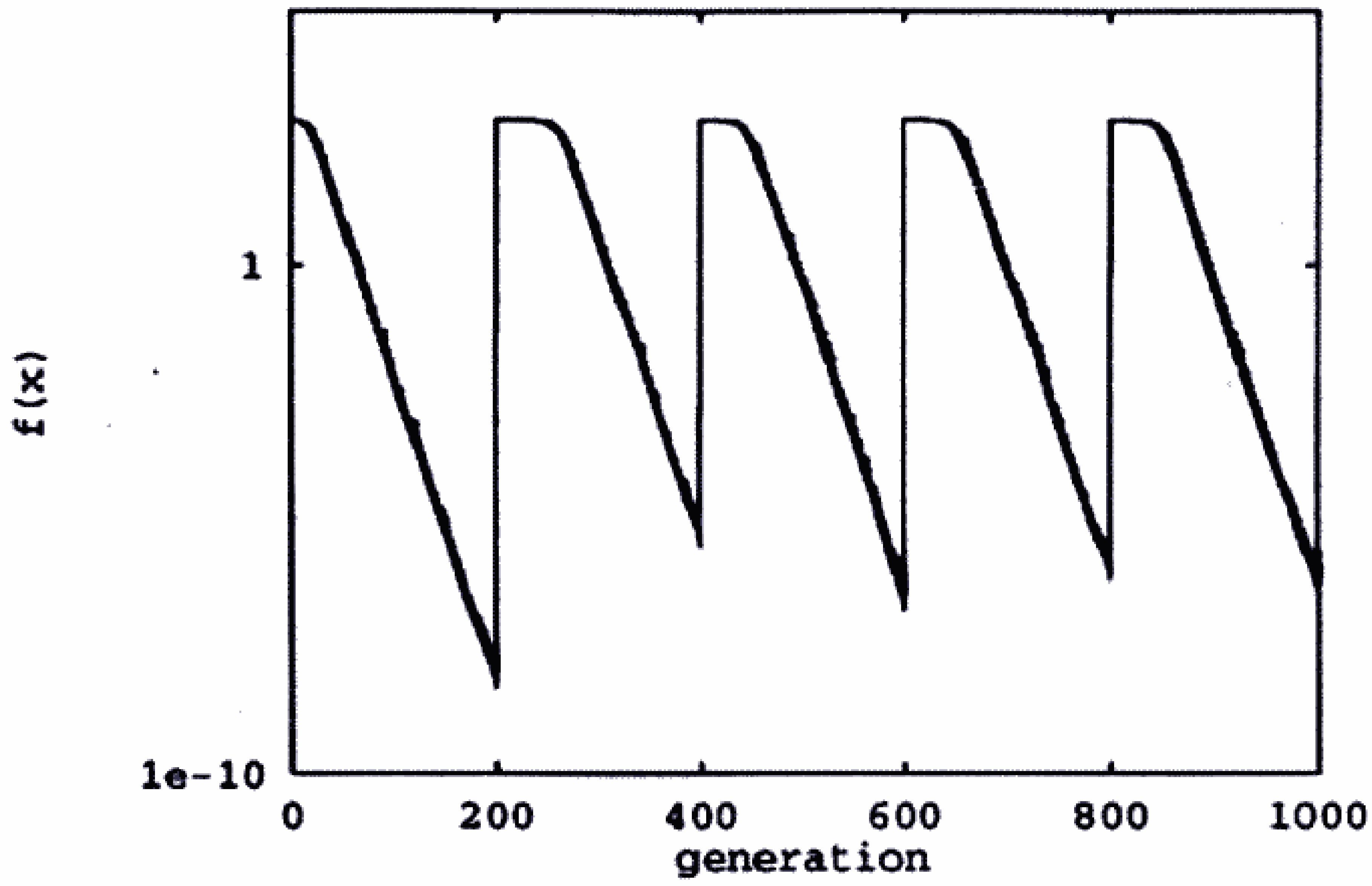


Figure 12: Behaviour of the objective function value.

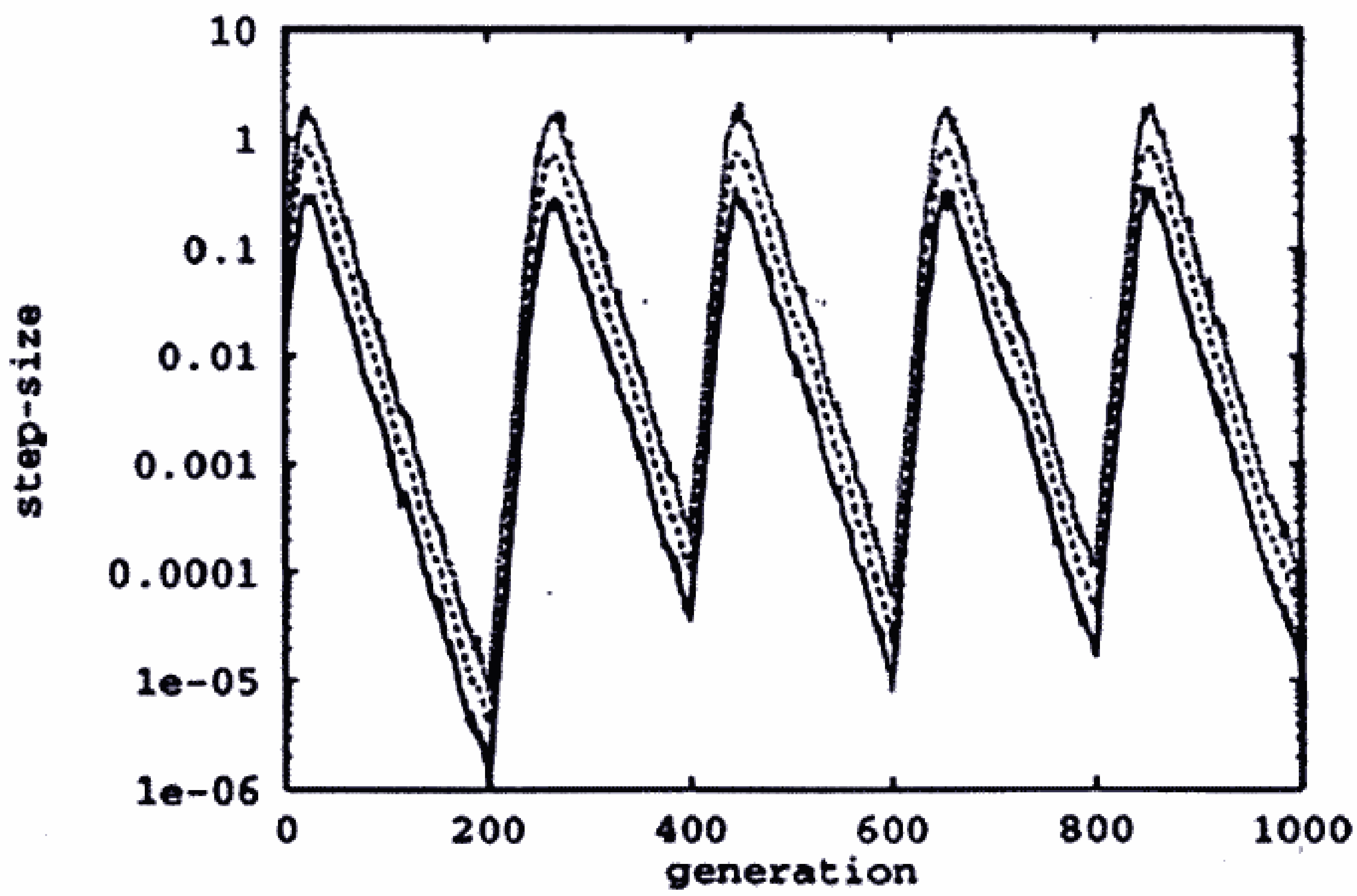


Figure 13: Behaviour of the standard deviations (minimum, average, and maximum within the population).

Collective learning ($n = 30$)

- of one common step size ($\Rightarrow$ 1 step size):

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

- of proper scalings ($\Rightarrow$ n step sizes):

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

- of a metric ($\Rightarrow$ correlated mutations):

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

(Schwefel, 1987) compares the progress rate for f_2 and a (μ,100)-ES for

- optimum prefixed scaling, i.e., *perfect information* ($\sigma_i = c/\sqrt{i}$) (A),
- prefixed arbitrary scaling ($\sigma_i = c'$) (B),
- adaptive scaling (C).

$F(x) = \sum_{i=1}^{n=30} i \cdot x_i^2$ (μ, 100) ES

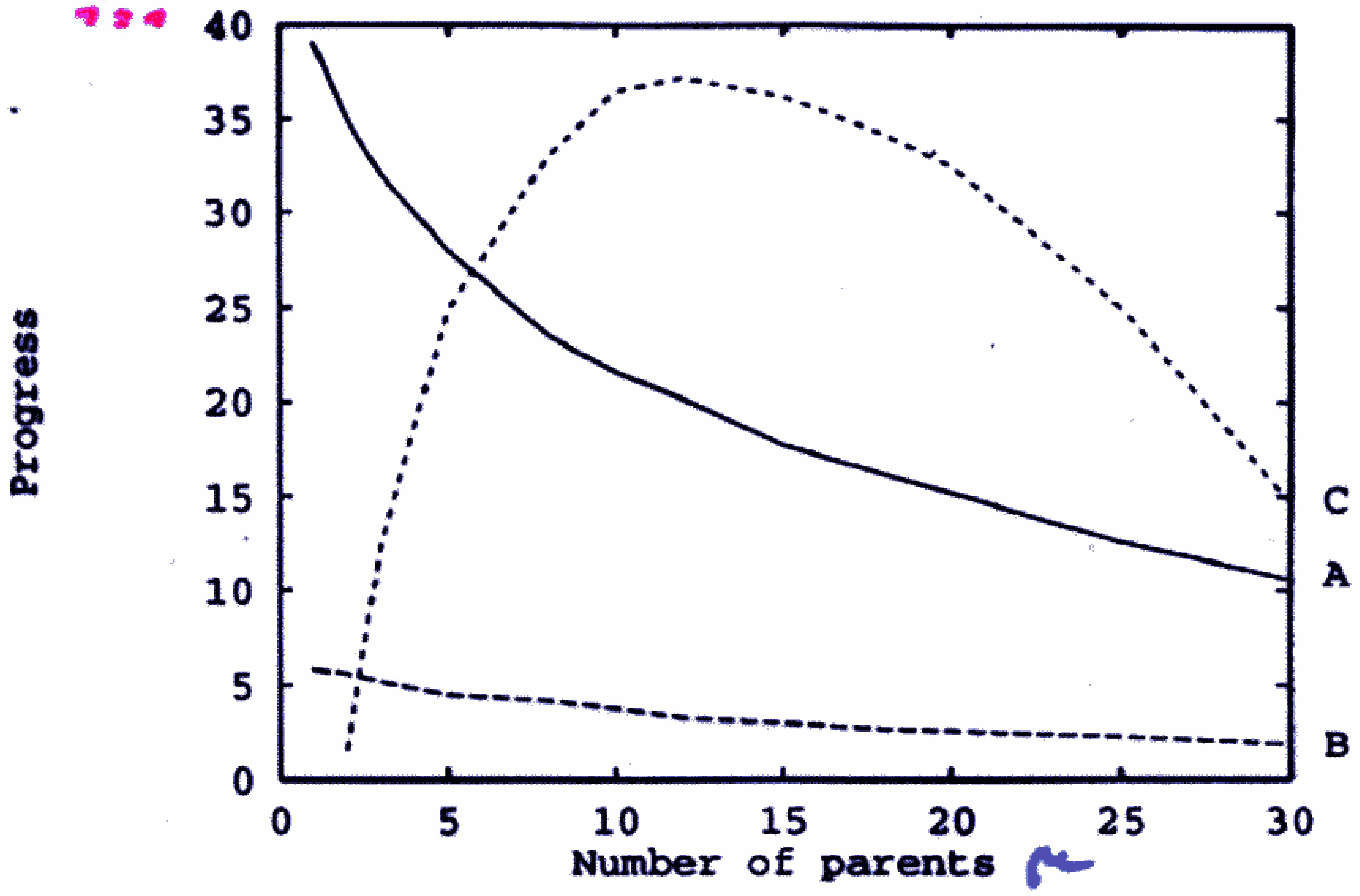


Figure 14: Comparison of progress rates. 15 imperfect, but different individuals (C) perform collectively better than the same number of cloned specialists (A) and almost as good as possible in case of perfect information (A, (1,100)).

Progress measure:

$$P = \ln \sqrt{\frac{f_0^{min}}{f_{1000}^{min}}}$$

Outlook: "Contemporary ESs"

Extensions of the $(\mu,\ \kappa,\ \lambda,\ \varrho)$-ES:

- Life span limited to $\kappa \geq 1$ generations.
  - $\kappa = 1$: $(\mu,\ \lambda)$-selection.
  - $\kappa = \infty$: $(\mu + \lambda)$-selection.
- Tournament selection as an alternative method.
- Further recombination types (e.g., crossover from GAs).
- Introduction of operator application probabilities p_m, p_c.

other ES variants

- with diploidy / dominance + recessivity
  for vector optimization (MCDM)
- with niching / migration
  for multimodal optimization (coarse grained)
- with local interaction (diffusion)
  for multimodal optimization (fine grained)
- with predator-prey selection
  for multimodal optimization (tournament-like)

- nested ESs [Rechenberg & Co.]

$$[\mu'/\varrho', \lambda' (\underbrace{\mu/\varrho, \lambda}_{\text{}})^{\gamma}] ES$$

λ' subpopulations (demes) act in parallel for γ generations independently

meta-selection among λ' subpopulations

(μ, λ) ES good for dynamic opt. (changing environment)

References

[1] Th. Bäck. *Evolutionary Algorithms in Theory and Practice*. Oxford University Press, New York, 1995.

[2] R. K. Belew and L. B. Booker, editors. *Proceedings of the 4th International Conference on Genetic Algorithms*. Morgan Kaufmann Publishers, San Mateo, CA, 1991.

[3] J. Branke. Evolutionary algorithms for neural network design and traning. In *Proceedings of the 1st Nordic Workshop on Genetic Algorithms and its Applications*. Vaasa, Finland, 1995.

[4] Y. Davidor. *Genetic Algorithms and Robotics. A Heuristic Strategy for Optimization*, volume 1 of *World Scientific Series in Robotics and Automated Systems*. World Scientific, 1991.

[5] L. Davis, editor. *Handbook of Genetic Algorithms*. Van Nostrand Reinhold, New York, 1991.

[6] D. B. Fogel. *Evolutionary Computation: Toward a New Philosophy of Machine Intelligence*. IEEE Press, Piscataway, NJ, 1995.

[7] S. Forrest, editor. *Proceedings of the 5th International Conference on Genetic Algorithms*. Morgan Kaufmann Publishers, San Mateo, CA, 1993.

[8] D. E. Goldberg. *Genetic algorithms in search, optimization and machine learning*. Addison Wesley, Reading, MA, 1989.

[9] J. J. Grefenstette, editor. *Proceedings of the 1st International Conference on Genetic Algorithms and Their Applications*. Lawrence Erlbaum Associates, Hillsdale, NJ, 1985.

[10] J. J. Grefenstette, editor. *Proceedings of the 2nd International Conference on Genetic Algorithms and Their Applications*. Lawrence Erlbaum Associates, Hillsdale, NJ, 1987.

[11] Frédéric Gruau. Cellular encoding of genetic neural networks. Technical Report 92-21, Ecole Normale Superieure de Lyon, May 1992.

[12] Frederic Gruau. *Neural Network Synthesis Using Cellular Encoding and the Genetic Algorithm*. PhD thesis, PhD Thesis, Ecole Normale Supérieure de Lyon, 1994. anonymous ftp: lip.ens-lyon.fr (140.77.1.11) directory pub/Rapports/PhD file PhD94-01-E.ps.Z (english) PhD94-01-F.ps.Z (french).

[13] R. F. Hartl. A global convergence proof for a class of genetic algorithms. Technische Universität Wien, 1990.

[14] J. H. Holland. *Adaptation in natural and artificial systems*. The University of Michigan Press, Ann Arbor, MI, 1975.

[15] K. A. De Jong. *An analysis of the behaviour of a class of genetic adaptive systems*. PhD thesis, University of Michigan, 1975. Diss. Abstr. Int. 36(10), 5140B, University Microfilms No. 76–9381.

[16] R. Männer and B. Manderick, editors. *Parallel Problem Solving from Nature 2*. Elsevier, Amsterdam, 1992.

[17] Z. Michalewicz. *Genetic Algorithms + Data Structures = Evolution Programs*. Artificial Intelligence. Springer, Berlin, 1992.

[18] G.F. Miller, P.M. Todd, and S.U. Hedge. Designing neural networks using genetic algorithms. In *International Conference on Genetic Algorithms*, pages 379–384. Morgan Kaufman, 1989.

[19] G. J. E. Rawlins, editor. *Foundations of Genetic Algorithms*. Morgan Kaufmann Publishers, San Mateo, CA, 1991.

[20] I. Rechenberg. Cybernetic solution path of an experimental problem. Royal Aircraft Establishment, Library translation No. 1122, Farnborough, Hants., UK, August 1965.

[21] I. Rechenberg. *Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*. Frommann–Holzboog, Stuttgart, 1973.

[22] J. D. Schaffer, editor. *Proceedings of the 3rd International Conference on Genetic Algorithms*. Morgan Kaufmann Publishers, San Mateo, CA, 1989.

[23] H.-P. Schwefel. *Numerische Optimierung von Computer-Modellen mittels der Evolutionsstrategie*, volume 26 of *Interdisciplinary Systems Research*. Birkhäuser, Basel, 1977.

[24] H.-P. Schwefel. *Numerical Optimization of Computer Models*. Wiley, Chichester, 1981.

[25] H.-P. Schwefel. Collective phenomena in evolutionary systems. In *Preprints of the 31st Annual Meeting of the International Society for General System Research, Budapest*, volume 2, pages 1025–1033, June 1987.

[26] H.-P. Schwefel. *Evolution and Optimum Seeking*. Sixth-Generation Computer Technology Series. Wiley, New York, 1995.

[27] H.-P. Schwefel and R. Männer, editors. *Parallel Problem Solving from Nature — Proceedings 1st Workshop PPSN I*, volume 496 of *Lecture Notes in Computer Science*. Springer, Berlin, 1991.

[28] G. Weiss. Combining neural and evolutionary learning: Aspects and approaches. Technical Report FKI-132-90, Technische Universität München, 1990.

[29] D. Whitley, editor. *Foundations of Genetic Algorithms 2*. Morgan Kaufmann Publishers, San Mateo, CA, 1993.

[30] D. Whitley, T. Starkweather, and C. Bogart. Genetic algorithms and neural networks: optimizing connections and connectivity. volume 14, pages 347–361. 1990.

Strategien - Test

1. Konvergenzgeschwindigkeit
   bzw. benötigte Rechenzeit (n) für geg. Approx
   zwei quadratische Probleme

   a) $F = \sum_{i=1}^{n} x_i^2$ 

   b) $F = \sum_{i=1}^{n} \left(\sum_{j=1}^{i} x_j \right)^2$ 

   $= x^T A x$

   ↑ Kondition der Matrix $\sim n^2$

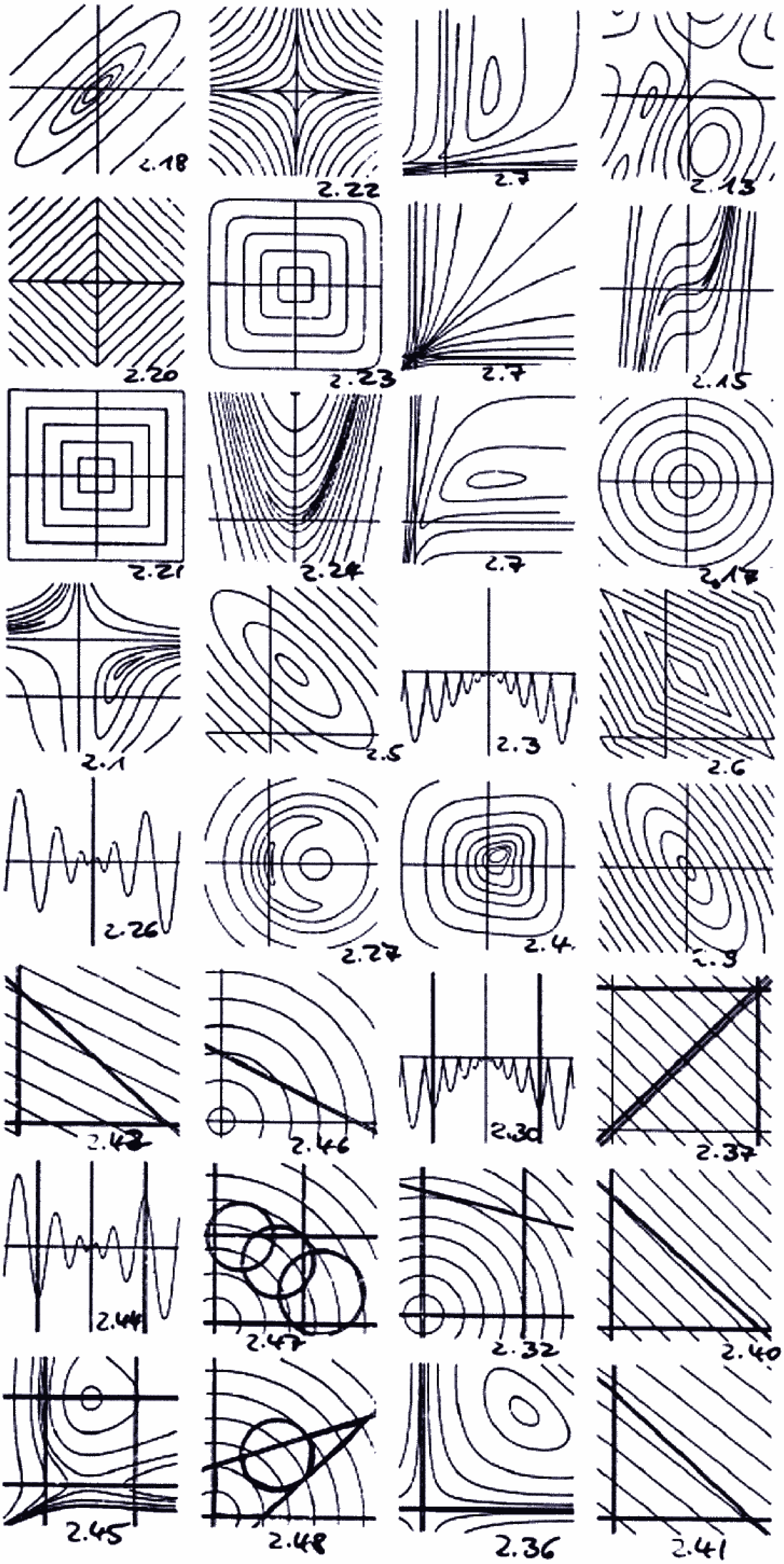
2. Zuverlässigkeit
   erreichbare Approximation
   50 Testprobleme mit $n \leq 6$

   28 ohne Nebenbedingungen
   22 mit Nebenbedingungen

   8 davon multimodal

getestete Strategien

Koordinatenstrategie mit linearer Suche	
Fibonacci	FIBO
Gold. Schnitt	GOLD
Lagrange	LAGR
Hooke + Jeeves	HOJE
Rosenbrock	ROSE
Davies, Swann, Campey	DSCG / DSCP
Powell (konjug. R.)	POWE
variable Metrik	DFPS
Simplex	SIMP
Complex	COMP
Evolutionsstrategien	
(1+1)	EVOL
(10,100)	GRUP
(10,100) mit Rekomb.	REKO


2.18
2.22
2.7
2.13
2.20
2.23
2.7
2.15
2.21
2.24
2.7
2.17
2.1
2.5
2.3
2.6
2.26
2.27
2.4
2.9
2.42
2.46
2.30
2.37
2.44
2.47
2.32
2.40
2.45
2.48
2.36
2.41

Kennzeichen / Problem Nr. \ Strategie	FIBO	GOLD	LAGR	HOJE	DSCG	DSCP	POWE	DFPS	SIMP	ROSE	COMP	EVOL	GRUP	REKO
2.1	3	3	3	1	1	1	2	1e	2n	1	2	1	1	1
2.2	3	3	3	1	1	1	1	2	2a	1a	2	1	1	1
2.3	2	2	3	1	1	1	1	1	1n	1a	4	1	1	1
2.4	3	3	3	1	1	1	2	2e	3	1	3	1	1	1
2.5	2	2	2	2	1	1	2	1e	2a	1	2	1	1	1
2.6	5	5	5	5	2	2	5	3	2a	1	3	1	1	1
2.7	5	5	4	2	5ea	5ea	5e	5	3	5	3	2	1	1
2.8	5	5	4	3	5ea	5ea	5e	5	3	5	2	1	1	1
2.9	3	3	3	2	2	2	2e	2	1a	3	1	1	2	1
2.10	5	5	4	3	2	2	5a	4e	4n	2	2	4	3	1
2.11	5	5	5	3	2	2	2	4	4n	2	2	4	3	1a
2.12	5	5	5	3	2	3	2	4e	2	2	2	4	3	1a
2.13	3	3	3	2	2	1	3	2	3	3	3	2	1	1
2.14	3	3	3	2	2	2	2a	5e	2n	2	2	3r	3r	3r
2.15	3	3	3	2	2	2	2ea	3	2	2	2	3r	3r	3r
2.16	2	1	2	2	1	1	2	2	2n	2	3	3	2	2
2.17	2	2	1	2	2	2	2e	2	1a	2	1	1	1	1
2.18	5	5	5	2	2	2	2e	2	1an	1	1	1	1	1
2.19	5	5	5	5	2	2	5	2e	2	2	3	3	2	3
2.20	2	2	2	2	3	2	2	1e	3n	2	2	1	1	1
2.21	5	5	5	2	4	2	5	2e	5a	2	5	1	1	1
2.22	2	2	2	2	2	2	2	5	5a	2	5	1	1	1
2.23	1	1	1	1	1	1	5a	5e	1a	1a	1	1	1	1
2.24	3	3	3	2	1	1	2	2	2	2	2	2	2	1
2.25	3	3	3	2	1	1	2e	2e	3	1	3	1	1	1
2.26	1	1	2	1	1	1	1e	1	1n	1	4	1	1	1
2.27	1	1	5	2	1	1	1	5	1	1	2	1	1	1
2.28	4	4	4	3	4	3	2	4e	2	3	1	4	4	3

(FIBO, GOLD, LAGR: Koord. strat.)

e: fataler Fehler

a: schaltet nicht ab

r: sehr langsam; Zeitabschaltung

n: Neustart

1 Genauigkeit besser 10^{-38}

2 ~ 10^{-8}

3 ~ 10^{-4}

4 ~ 10^{-2}

5 schlechter 10^{-2}

Kennzeichen / Problem Nr.	Strategie	ROSE	COMP	EVOL	GRUP	REKO
2.29		3	1	4	3	3
2.30		1	5	1	1	1
2.31		3v	3	1	1	1
2.32		3v	3	1	1	1
2.33		3	2	5	4	1
2.34		1	2	3	3	2
2.35		3v	1	4	4	4
2.36		1	1	1	1	1
2.37		3	1	1	1	1
2.38		3v	3	1	1	1
2.39		3	3	4	4	3
2.40		5	5	5	5	5
2.41		5	5	5	5	5
2.42		3	3	2	2	1
2.43		3v	3	2	2	1
2.44		1	5	1	1	1
2.45		3	2	4	2	1
2.46		3	2	3	3	1
2.47		3v	1	1	1	1
2.48		3v	3	1	1	1
2.49		3	2	4	3	1
2.50		3	1	1	1	1

Strategie	Anzahl der exakt gelösten Aufgaben ohne Nebenbedingungen	Anzahl der nicht gelösten Aufgaben ohne Nebenbedingungen	Anzahl der exakt gelösten Aufgaben mit Nebenbedingungen	Anzahl der nicht gelösten Aufgaben mit Nebenbedingungen
FIBO	3	9	nicht direkt anwendbar	
GOLD	4	9	nicht direkt anwendbar	
LAGR	2	7	nicht direkt anwendbar	
HOJE	6	2	nicht direkt anwendbar	
DSCG	11	2	nicht direkt anwendbar	
DSCP	12	2	nicht direkt anwendbar	
POWE	4	7	nicht direkt anwendbar	
DFPS	5	6	nicht direkt anwendbar	
SIMP	7	2	nicht direkt anwendbar	
ROSE	11	2	4	2
COMP	5	2	6	4
EVOL	17	0	10	3
GRUP	18	0	10	2
REKO	23	0	16	2
	von 28		von 22	

Zusammengefaßte Ergebnisse des Zuverlässigkeitstests im Hinblick auf lokale Konvergenz (28 Probleme ohne und 22 mit Nebenbedingungen; Strategie-Namen wie in Tabelle 1)

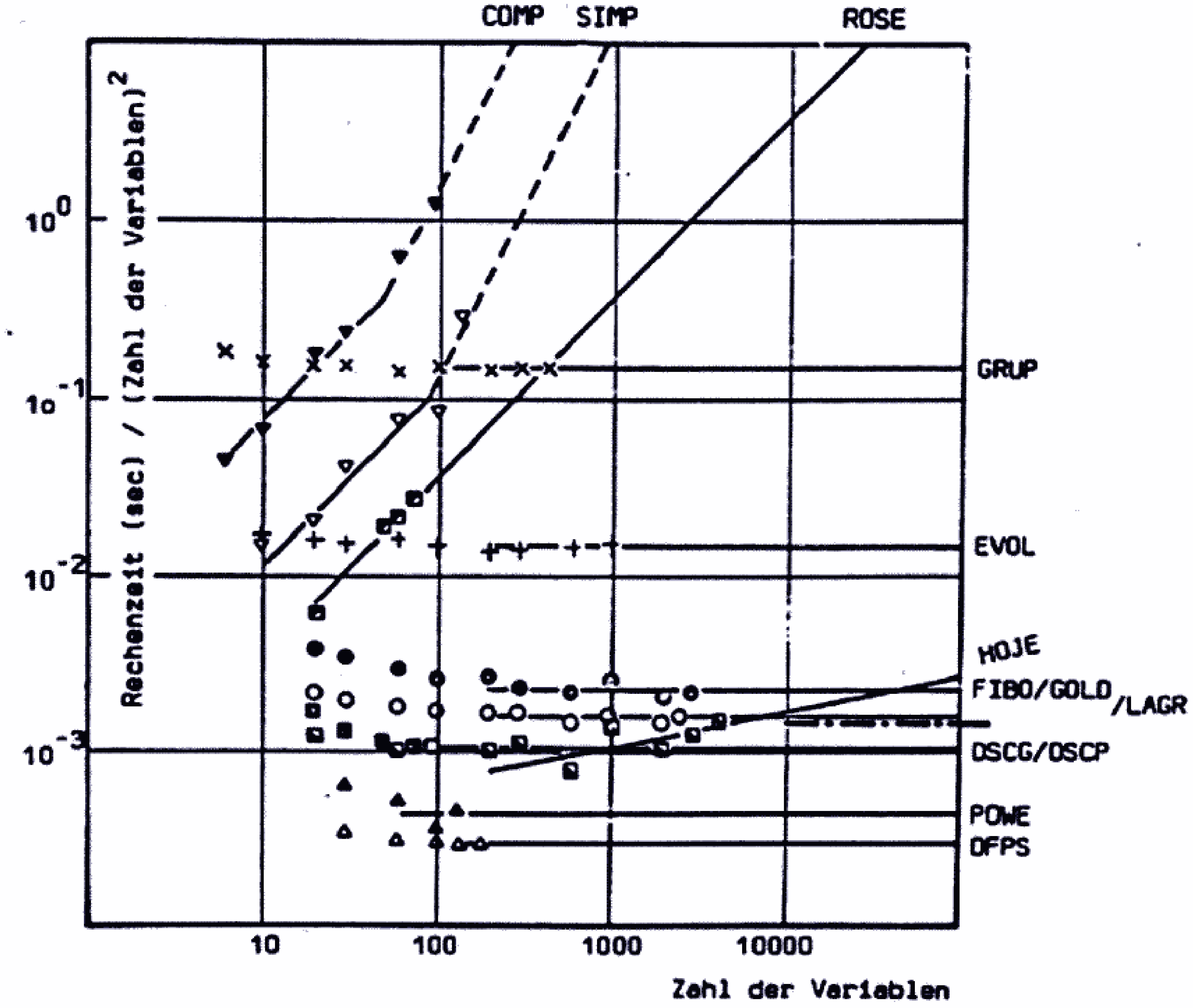
Strategie / Problem Nr.	FIBO	GOLD	LAGR	HOJE	DSCG	DSCP	POWE	DFPS	SIMP	ROSE	COMP	EVOL	GRUP	REKO
1	L1	L1	L3	L1	L7	L7	L1	L3	L1	L6	L1	Lm	G	G
2	L1	L1	L1	L1	L1	L1	L1	L1	L1	L1	L1	L1	G	G
3										L4	L1	Lm	G	G
4										G	G	G	G	G
5										L1	L1	L1	G	G
6										L	G	G	G	G
7										L3	L1	L2	G	G
8										L2	Lm	Lm	GL	GL

Ergebnisse des Tests auf globale Konvergenz
(Strategienamen wie in Tabelle 1)

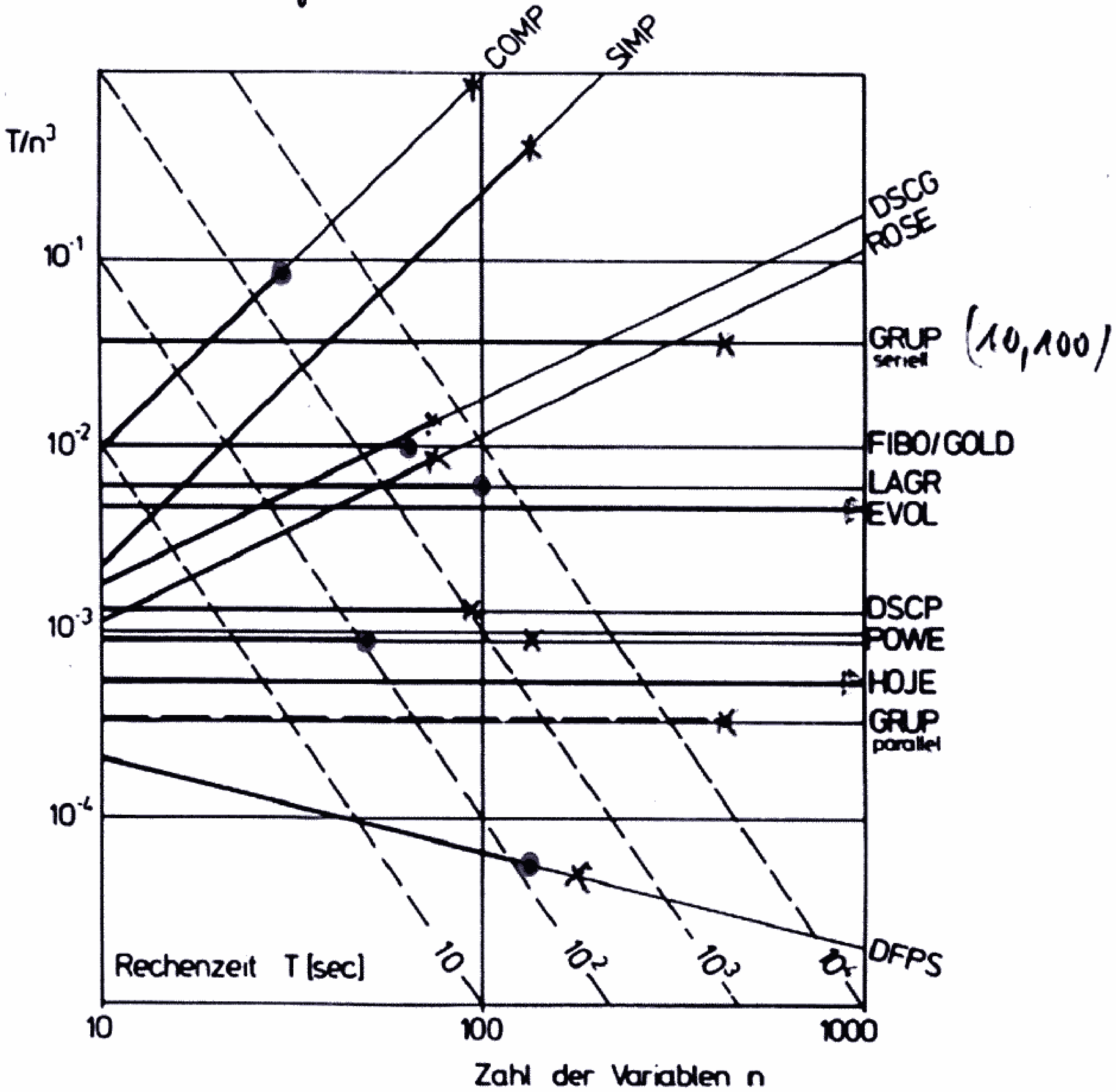
Bedeutung der Symbole:

- L Suche konvergiert gegen lokales Minimum
- L3 Suche konvergiert gegen das 3. lokale Minimum (Zählung der Reihenfolge nach absteigenden Zielfunktionswerten)
- Lm Suche konvergiert gegen verschiedene lokale Minima je nach Zufallszahlen
- G Suche konvergiert gegen globales Minimum
- GL Suche konvergiert gegen lokales oder globales Minimum je nach Zufallszahlen

$$F = \sum_{i=1}^{n} x_i^2 \rightarrow \min$$



$$F = \sum_{i=1}^{m} \Big(\sum_{j=1}^{i} x_j \Big)^2 \to \min$$



● fatale Fehler

x Speicherplatzbegrenzung

Benötigte Rechenzeit in Abhängigkeit von der Parameterzahl für die Erzielung einer vorgegebenen Approximation an die Lösung eines quadratischen Optimierproblems

Bedeutung der Symbole:

o   *Strategie versagt bei höherer Parameterzahl*

x   *Kernspeicher-Platzbedarf überschreitet 30 K Worte*