Evolutionäre Algorithmen: Überblick über die Arbeitsprinzipien

Karsten Weicker

HTWK Leipzig

5. Januar 2013

Überblick

1. Wechselspiel Variation – Selektion
2. Populationskonzept
3. Verknüpfen mehrerer Individuen
4. Selbstanpassende Algorithmen
5. Zusammenfassung

Überblick

Einfaches binäres Beispiel

$\textsc{Ein-Bit-Binäre-Mutation}$(Individuum A mit $A.G \in \mathbb{B}^l$)

$B \leftarrow A$
$i \leftarrow$ wähle zufällig gemäß $U(\{1, \ldots, l\})$
$B_i \leftarrow 1 - A_i$
return B

Einfaches binäres Beispiel

$\textsc{Binäres-Hillclimbing}$(Zielfunktion F)

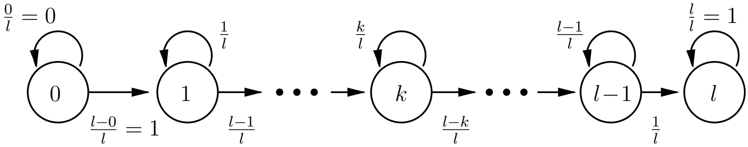
$t \leftarrow 0$
$A(t) \leftarrow$ erzeuge Lösungskandidat
bewerte $A(t)$ durch F
while Terminierungsbedingung nicht erfüllt
do $t \leftarrow t + 1$
  $B \leftarrow \textsc{Ein-Bit-Binäre-Mutation}(A(t-1))$
  bewerte B durch F
  if $B.F \succeq A(t-1).F$
  then $A(t) \leftarrow B$
  else $A(t) \leftarrow A(t-1)$
return $A(t)$

Endliche Markovkette

Definition

- *endliche Markovkette* $(\mathit{Zustände}, \mathit{Start}, \mathit{Übergang})$
- mögliche Zustände $\mathit{Zustände} = \{0, \ldots, k\}$
- Anfangswahrscheinlichkeit pro Zustand
  $\mathit{Start} \in [0, 1]^{k+1}$
  es gilt: $\sum_{0 \leq i \leq k} \mathit{Start}_i = 1$
- Übergangswahrscheinlichkeit
  $\mathit{Übergang} : \{0, \ldots, k\} \times \{0, \ldots, k\} \rightarrow [0, 1]$
  für alle i gilt: $\sum_{0 \leq j \leq k} \mathit{Übergang}(i, j) = 1$

Optimierung als Markovprozess



Endliche Markovkette

Satz

BINÄRES-HILLCLIMBING *erreicht das Optimum eines Musterabgleichs mit l Bits in $\mathcal{O}(l \log l)$ Schritten (als Erwartungswert).*

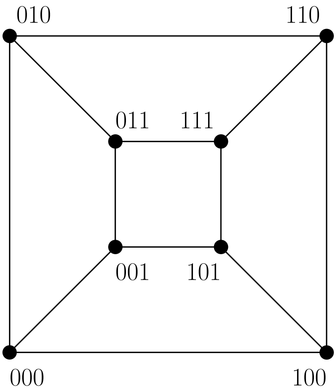
Gütelandschaften: Nachbarschaft

Definition

- Mutationsoperator $Mut^{\xi}: \ \mathcal{G} \times \mathcal{Z} \to \mathcal{G} \times \mathcal{Z}$ mit $\mathcal{Z} = \{\perp\}$
- *Nachbarschaftsgraph* ist gerichteter Graph $G = (V, E)$ mit $V = \mathcal{G}$ und

$$E = \left\{ (A.G, B.G) \in V \times V \,\middle|\, \exists \xi \in \Xi: \ Mut^{\xi}(A) = B \right\}$$

Gütelandschaften: Nachbarschaft

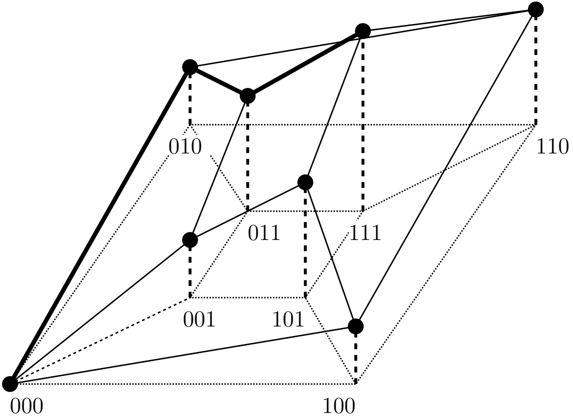


Gütelandschaft

Definition

- Geg.: Nachbarschaftsgraphen $G = (\mathcal{G}, E)$, induzierte Bewertungsfunktion $F : \mathcal{G} \to \mathbb{R}$
- *Gütelandschaft* (G, F) durch Zuordnung der Güte zu den Punkten
- *Weg in der Landschaft*

Gütelandschaften: Hillclimbing

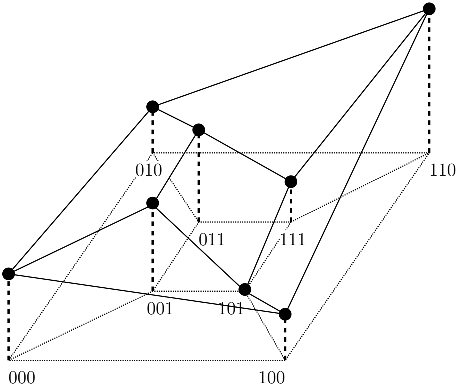


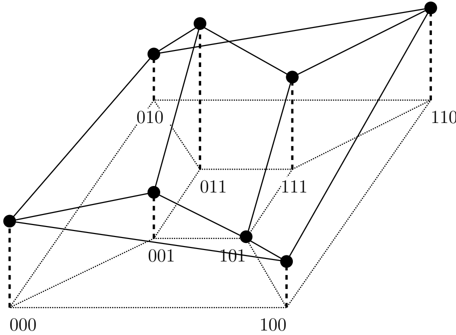
Lokales Optimum

Definition

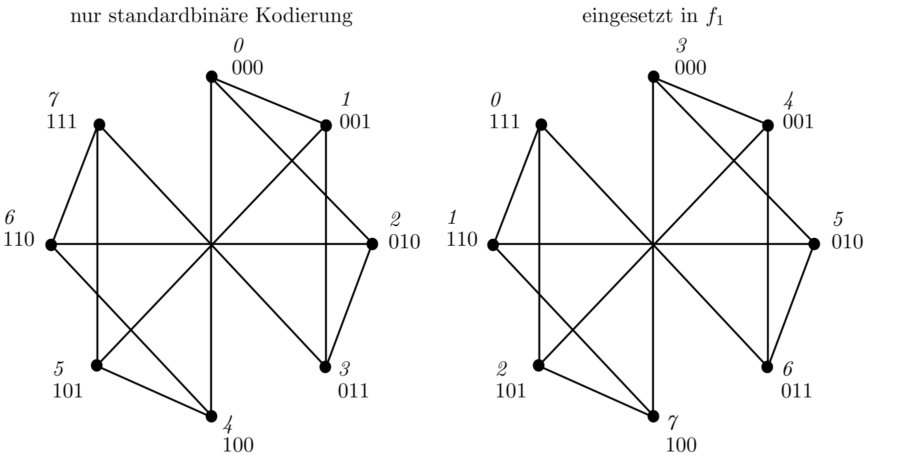
- Geg.: Mutationsoperator, dazugehörige Gütelandschaft $((\mathcal{G}, E), F)$
- A ist *lokales Optimum*, falls alle möglichen Mutanten $B = Mut(A)$ nicht besser sind und Wege zu besseren Lösungskandidaten über schlechtere führen.
- A ist *Plateau-Punkt*, falls alle möglichen Mutanten $B = Mut(A)$ nicht besser sind und wenigstens einer gleich gut ist.

Gütelandschaften: Hillclimbing

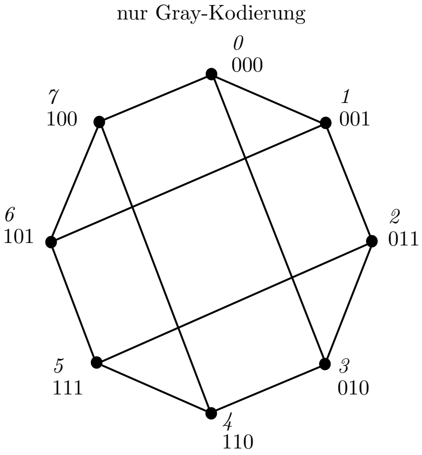


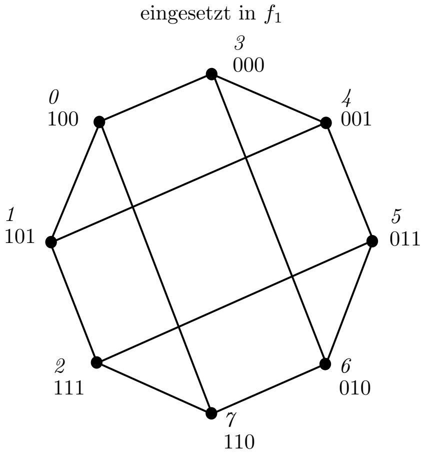


Kodierung: standardbinär



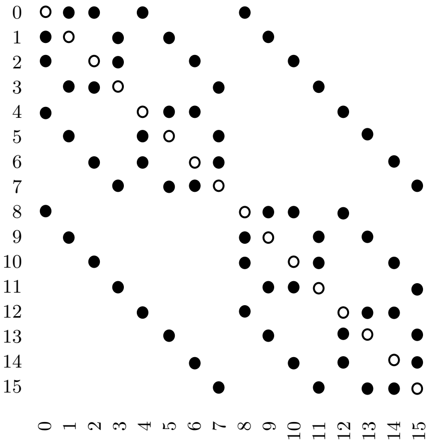
Kodierung: Gray



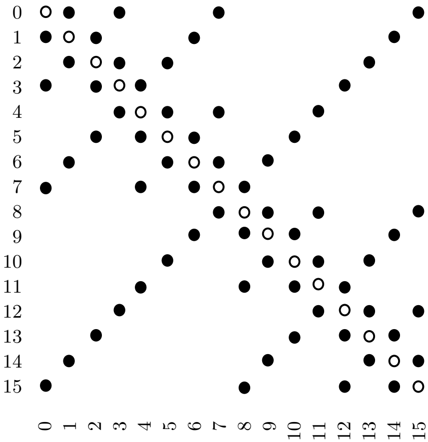


Kodierung: Vergleich der Nachbarschaft

standardbinäre Kodierung



Gray-Kodierung



Rollen der Mutation

Exploration oder Erforschung

- stichprobenartiges Erkunden
- auch: weiter entfernte Regionen des Suchraums

Exploitation oder Feinabstimmung

- lokale Verbesserung eines Lösungskandidaten
- wichtig: Einbettung der phänotypischen Nachbarschaft

Binäre Mutation

BINÄRE-MUTATION(Individuum A mit $A.G \in \mathbb{B}^l$)

$B \leftarrow A$
for $i \in \{1, \ldots, l\}$
do $u \leftarrow$ wähle zufällig gemäß $U([0,1))$
    if $u \leq p_m$ ⦇Mutationswahrscheinlichkeit⦈
    then $B.G_i \leftarrow 1 - A.G_i$
return B

Animation

Binärer Hillclimber

Eine alternative reellwertige Mutation

Gauß-Mutation

- direkt auf den reellwertigen Zahlen
- Addition einer normalverteilten Zufallszahl auf jede Komponente

Eine alternative reellwertige Mutation

Gauß-Mutation

- direkt auf den reellwertigen Zahlen
- Addition einer normalverteilten Zufallszahl auf jede Komponente

$\textsc{Gauss-Mutation}$(Individuum A mit $A.G \in \mathbb{R}^l$)

for $i \in \{1, \ldots, l\}$
do $u_i \leftarrow$ wähle zufällig gemäß $\mathcal{N}(0, \sigma$ (Standardabweichung))
$\quad B_i \leftarrow A_i + u_i$
$\quad B_i \leftarrow \max\{B_i, ug_i$ (untere Wertebereichsgrenze)$\}$
$\quad B_i \leftarrow \min\{B_i, og_i$ (obere Wertebereichsgrenze)$\}$
return B

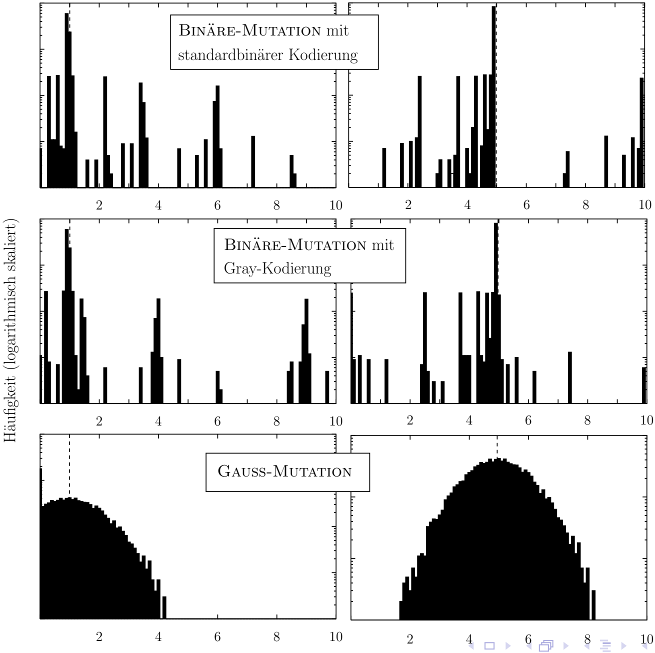
Vergleich der Mutationen

Ansatz

- Optimierung der einfachen Funktion

$$f_2(x) = \begin{cases} x & \text{falls } x \in [0,10] \subset \mathbb{R} \\ \text{undef.} & \text{sonst} \end{cases}$$

- zwei Elternindividuen (1,0 und 4,99)
- Ermittlung der Nachkommensverteilung mit jeweils 10 000 Mutationen


Binäre-Mutation mit standardbinärer Kodierung
Binäre-Mutation mit Gray-Kodierung
Gauss-Mutation
Häufigkeit (logarithmisch skaliert)
2
4
6
8
10

Vergleich der Mutationen

Binäre Mutation: Ergebnis

- rastert den Suchraum
- weitgestreute Erforschung
- nur beim Graykode: akzeptable Feinabstimmung

Gauß-Mutation: Ergebnis

- regelmäßiges Verhalten
- gute Feinabstimmung
- schlechte, weit springende Erforschung

Überblick

Populationskonzept

Idee

- mehr als ein Individuum existiert in einer Population

Nutzen

- gleichzeitige Betrachtung mehrerer Individuen
- parallele Suche an verschiedenen Stellen im Suchraum
- Hoffnung: Bedeutung der lokalen Optima verringern

Diversitätsmaße

Definition

- Population $P = \langle A^{(i)} \rangle_{1 \leq i \leq s}$, Genotyp $\mathcal{G} = G^l$
- *mittlerer Abstand* der Individuen in der Population

$$Divers_{\text{Abstand},d}(P) = \frac{1}{s \cdot (s-1)} \sum_{1 \leq i,j \leq s} d(A^{(i)}.G, A^{(j)}.G)$$

wobei $d : \mathcal{G} \times \mathcal{G} \to \mathbb{R}$ ein beliebiges Abstandsmaß ist.

Diversitätsmaße

Definition

- Population $P = \langle A^{(i)} \rangle_{1 \leq i \leq s}$, Genotyp $\mathcal{G} = \mathbb{B}^l$
- *Shannon-Entropie* als positionsorientierte Diversität

$$Divers_{\text{Entropie}}(P) =$$

$$\frac{1}{l} \sum_{k=1}^{l} \left(-\#_0(P,k) \log \#_0(P,k) - \#_1(P,k) \log \#_1(P,k) \right),$$

$$\text{mit } \#_1(P,k) = \frac{\#\{1 \leq i \leq s \mid A^{(i)}.G_k = 1\}}{s}$$

$$\text{und } \#_0(P,k) = \frac{\#\{1 \leq i \leq s \mid A^{(i)}.G_k = 0\}}{s}.$$

Diversitätsmaße

Definition

- Population $P = \langle A^{(i)} \rangle_{1 \leq i \leq s}$, Genotyp $\mathcal{G} = \mathbb{B}^l$
- *teilstringorientierte Diversität*

$$Divers_{\text{Teilstring}}(P) = \frac{s \cdot \# \left(\bigcup_{1 \leq i \leq s} Teil(A^{(i)}) \right)}{\sum_{1 \leq i \leq s} \# Teil(A^{(i)})},$$

$$\text{wobei } Teil(A) = \bigcup_{1 \leq i \leq j \leq l} \{A.G_i \ldots A.G_j\}.$$

Konvergenz

Definition

Eine Population $P = \langle A^{(i)} \rangle_{1 \leq i \leq s}$ heißt *konvergiert*, wenn alle Individuen identisch sind, d. h. für alle $1 \leq i, j \leq s$ gilt $A^{(i)}.G = A^{(j)}.G$.

Ein vergleichendes Experiment

POPULATIONSBASIERTER-BINÄRER-HILLCLIMBER(Zielfunktion F)

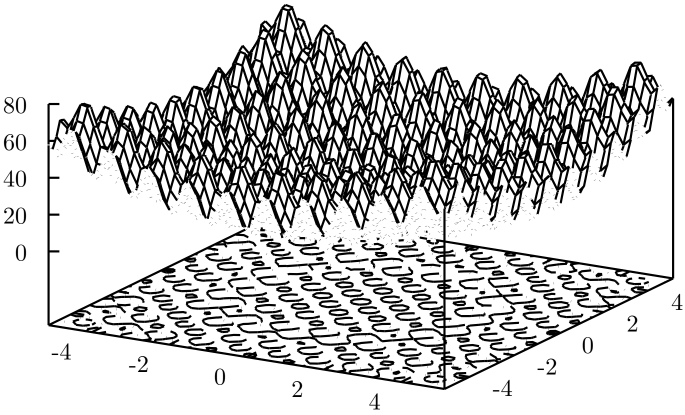
$t \leftarrow 0$
$P(t) \leftarrow$ erzeuge Population mit μ (Populationsgröße) Individuen
bewerte $P(t)$ durch F
while Terminierungsbedingung nicht erfüllt
do ⌜ $P' \leftarrow P(t)$
    for $i \in \{1, \ldots, \mu\}$
    do ⌜ $B \leftarrow$ EIN-BIT-BINÄRE-MUTATION($A^{(i)}$) wobei $P(t) = \langle A^{(k)} \rangle_{1 \leq i \leq \mu}$
        bewerte B durch F
        ⌞ $P' \leftarrow P' \circ \langle B \rangle$
    $t \leftarrow t + 1$
  ⌞ $P(t) \leftarrow$ Selektion aus P' mittels BESTEN-SELEKTION
return bestes Individuum aus $P(t)$

Selektion der Besten

$\textsc{Besten-Selektion}$(Gütewerte $\langle A.F^{(i)} \rangle_{i=1,\dots,r}$)

```
I ← ⟨⟩
for 1 ≤ j ≤ s (|Anzahl der zu wählenden Individuen|)
do ⌜ index_j ← derjenige Index aus {1, ..., r} \ I mit dem besten Gütewert
     ⌞ I ← I ∘ ⟨index_j⟩
return I
```

Rastrigin-Funktion:
$f(X) = 10 \cdot n + \sum_{i=1}^{n} (X_i^2 - 10 \cdot \cos(2\pi \cdot X_i))$ mit $n = 2$
und $-5,12 \leq X_1, X_2 \leq 5,12$



HC vs. Populationsbasiertes HC

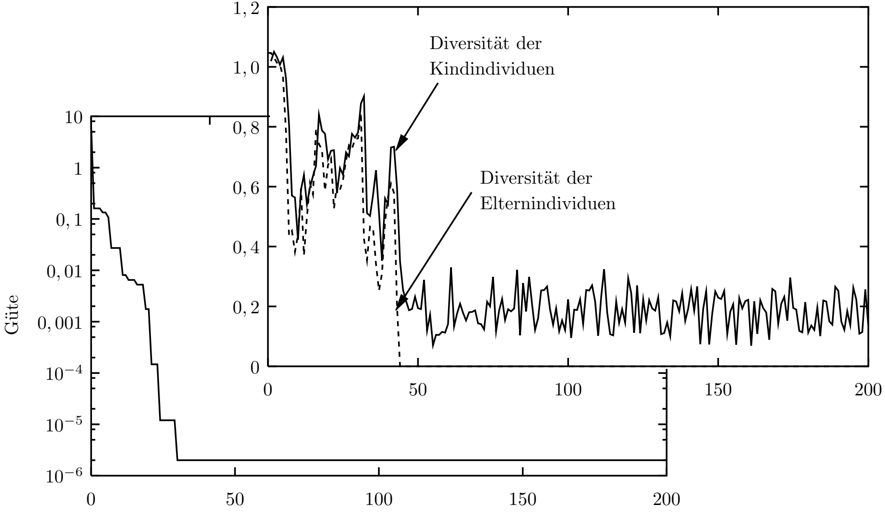
Rahmenbedingungen

- 100 Optimierungsläufe
- Populationsgröße $\mu = 50$
- 10000 evaluierte Individuen

Ergebnis

	Hillclimbing	Populationsbas. Hillclimbing
erfolgreich	63%	76%
durchschnittliche Endgüte	0,429	0,297

Diversität in der Population



Folgerungen für die Selektion

Elternselektion

- alle Individuen haben Chance, gewählt zu werden
- entweder: jedes Individuum ist Elter für $m > 0$ Kinder
- oder: Eltern mit einer Wahrscheinlichkeit wählen

Umweltselektion

- Vielfalt erhalten UND bessere Individuen übernehmen
- entweder: reine Auswahl der Besten
- oder: zufällige Auswahl, die bessere Individuen bevorzugt

Eigenschaften der Selektion

Definition

Sei $IS^{\xi} : \mathbb{R}^r \to \{1, \ldots, r\}^s$ Indexselektion.

- *deterministisch*, falls
  $\forall v \in \mathbb{R}^r\ \forall \xi, \xi' \in \Xi:\ \ IS^{\xi}(v) = IS^{\xi'}(v)$,
- *probabilistisch* genau dann, wenn nicht deterministisch, und
- *duplikatfrei*, falls
  $\forall v \in \mathbb{R}^r\ \forall \xi \in \Xi\ \forall 1 \leq i < j \leq r:\ \ (IS^{\xi}(v))_i \neq (IS^{\xi}(v))_j$

Zwei Minuten

Welche Aussagen sind richtig?

- ☐ Bestenselektion (TSP) ist duplikatfrei
- ☐ Uniforme Zufallswahl jedes Elters ist duplikatfrei
- ☐ Selektion des Hillclimbers ist probabilistisch
- ☐ zufällige Zahl $0 < k < \mu$ wählen, dann die k besten und die $\mu - k$ schlechtesten wählen ist probabilistisch
- ☐ Probabilistische Wahl kann die Übernahme des Besten nicht garantieren

Zwei Minuten

Welche Aussagen sind richtig?

- ☒ Bestenselektion (TSP) ist duplikatfrei
- ☐ Uniforme Zufallswahl jedes Elters ist duplikatfrei
- ☐ Selektion des Hillclimbers ist probabilistisch
- ☒ zufällige Zahl $0 < k < \mu$ wählen, dann die k besten und die $\mu - k$ schlechtesten wählen ist probabilistisch
- ☒ Probabilistische Wahl kann die Übernahme des Besten nicht garantieren

Überlappende Populationen

Definition

- Eltern $P = \langle A^{(i)} \rangle_{1 \le i \le \mu}$, Kinder $P' = \langle A^{(i)} \rangle_{1 \le i \le \lambda}$
- Sei $IS^{\xi} : \mathbb{R}^r \to \{1, \ldots, r\}^s$ Umweltsel. auf $P \circ P'$
- *überlappend*, falls wenigstens ein Tupel $x \in \mathbb{R}^r$ existiert, so dass $IS^{\xi}(x)$ einen Wert aus $\{1, \ldots, \mu\}$ enthält.
- *überlappend mit einem Überlappungsgrad* $g \in \{1, \ldots, \mu - 1\}$, falls für alle $x \in \mathbb{R}^r$ genau g Werte in $IS^{\xi}(x)$ aus $\{1, \ldots, \mu\}$
- *elitär*, falls immer ein Wert k mit $A^{(k)}.F \succeq A^{(i)}.F$ für alle $1 \le i \le \mu$ in $IS^{\xi}(x)$ enthalten ist.

Turnier als Umweltselektion

Selektion-Q-Stufiges-Turnier(Gütewerte $\langle A^{(i)}.F\rangle_{i=1,\dots,r}$)

$Scores \leftarrow \langle\rangle$
for $i = 1, \dots, r$
do ⌜ $Siege \leftarrow 0$
    for $j = 2, \dots, q$ ⦇Anzahl der direkten Turniere⦈
    do ⌜ $u \leftarrow$ wähle Zufallszahl gemäß $U(\{1, \dots, r\})$
        if $A^{(i)}.F \succ A^{(u)}.F$
        ⌞ **then** [$Siege \leftarrow Siege + 1$
    ⌞ $Scores \leftarrow Scores \circ \langle Siege\rangle$
$I \leftarrow \langle\rangle$
for $1 \leq j \leq s$ ⦇Anzahl der zu wählenden Individuen⦈
do ⌜ $index \leftarrow$ derjenige Index aus $\{1, \dots, r\} \setminus I$ mit maximalem Wert $Score^{(index)}$
    ⌞ $I \leftarrow I \circ \langle index\rangle$
return I

Ein Beispiel zur Turnierselektion

Individuum	Turniere gegen Gegner			Siege	Wahl
$A^{(1)}.F = 3,1$	3	8	5		
$A^{(2)}.F = 1,0$	1	2	9		
$A^{(3)}.F = 4,5$	10	4	7		
$A^{(4)}.F = 2,4$	6	9	10		
$A^{(5)}.F = 3,6$	1	8	7		
$A^{(6)}.F = 2,1$	3	6	4		
$A^{(7)}.F = 2,7$	2	5	8		
$A^{(8)}.F = 1,8$	3	9	1		
$A^{(9)}.F = 2,2$	6	7	4		
$A^{(10)}.F = 3,5$	2	10	5		

Ein Beispiel zur Turnierselektion

Individuum	Turniere gegen Gegner			Siege	Wahl
$A^{(1)}.F = 3,1$	3	8✓	5		
$A^{(2)}.F = 1,0$	1	2	9		
$A^{(3)}.F = 4,5$	10✓	4✓	7✓		
$A^{(4)}.F = 2,4$	6✓	9✓	10		
$A^{(5)}.F = 3,6$	1✓	8✓	7✓		
$A^{(6)}.F = 2,1$	3	6	4		
$A^{(7)}.F = 2,7$	2✓	5	8✓		
$A^{(8)}.F = 1,8$	3	9	1		
$A^{(9)}.F = 2,2$	6✓	7	4		
$A^{(10)}.F = 3,5$	2✓	10	5		

Ein Beispiel zur Turnierselektion

Individuum	Turniere gegen Gegner			Siege	Wahl
$A^{(1)}.F = 3,1$	3	8✓	5	1	
$A^{(2)}.F = 1,0$	1	2	9	0	
$A^{(3)}.F = 4,5$	10✓	4✓	7✓	3	
$A^{(4)}.F = 2,4$	6✓	9✓	10	2	
$A^{(5)}.F = 3,6$	1✓	8✓	7✓	3	
$A^{(6)}.F = 2,1$	3	6	4	0	
$A^{(7)}.F = 2,7$	2✓	5	8✓	2	
$A^{(8)}.F = 1,8$	3	9	1	0	
$A^{(9)}.F = 2,2$	6✓	7	4	1	
$A^{(10)}.F = 3,5$	2✓	10	5	1	

Ein Beispiel zur Turnierselektion

Individuum	Turniere gegen Gegner			Siege	Wahl
$A^{(1)}.F = 3,1$	3	8✓	5	1	✓
$A^{(2)}.F = 1,0$	1	2	9	0	
$A^{(3)}.F = 4,5$	10✓	4✓	7✓	3	✓
$A^{(4)}.F = 2,4$	6✓	9✓	10	2	✓
$A^{(5)}.F = 3,6$	1✓	8✓	7✓	3	✓
$A^{(6)}.F = 2,1$	3	6	4	0	
$A^{(7)}.F = 2,7$	2✓	5	8✓	2	✓
$A^{(8)}.F = 1,8$	3	9	1	0	
$A^{(9)}.F = 2,2$	6✓	7	4	1	
$A^{(10)}.F = 3,5$	2✓	10	5	1	

Selektionsintensität

Definition

- Sei $\overline{F}$ durchschnittliche Güte einer Population,
- σ deren Standardabweichung und
- $\overline{F}_{sel}$ die durchschnittliche Güte nach der Selektion
- Dann ist die Selektionsintensität definiert als

$$Intensität = \begin{cases} \frac{\overline{F}_{sel} - \overline{F}}{\sigma} & \text{falls } \succ \equiv > \\ \frac{\overline{F} - \overline{F}_{sel}}{\sigma} & \text{sonst.} \end{cases}$$

Fitnessproportionale Selektion

Idee

- ein Fitnesswert der Individuen bestimmt die Auswahlwahrscheinlichkeit (und damit die Anzahl der Kinder)
- Biologie: Ursache und Wirkung vertauscht
- zunächst: Fitness = Güte

Definition

Auswahlwahrscheinlichkeiten der *fitnessproportionalen Selektion*:

$$Pr[A^{(i)}] = \frac{A^{(i)}.F}{\sum_{k=1}^{r} A^{(k)}.F}$$

Fitnessproportionale Selektion

Selektion-Fitnessproportional(Gütewerte $\langle A^{(i)}.F \rangle_{1 \le i \le r}$)

$\langle \mathit{Fitness}^{(i)} \rangle_{1 \le i \le r} \leftarrow$ berechne Fitnesswerte aus $\langle A^{(i)}.F \rangle_{1 \le i \le r}$
$I \leftarrow \langle\rangle$
for $i \in \{1, \ldots, s$ (Anzahl der zu wählenden Individuen) $\}$
do $j \leftarrow 1$
    $\mathit{sum} \leftarrow \mathit{Fitness}^{(j)}$
    $u \leftarrow$ wähle Zufallszahl gemäß $U([0,1))$
    while $\mathit{sum} < u$
    do $j \leftarrow j + 1$
        $\mathit{index} \leftarrow j$
        $\mathit{sum} \leftarrow \mathit{sum} + \mathit{Fitness}^{(j)}$
    $I \leftarrow I \circ \langle \mathit{index} \rangle$
return I

Verbesserungen

Skalieren

- Idee: nur den interessanten Gütebereich heranziehen
- $schlechtesteF_W^{(t)}$ = schlechtester Gütewert der letzten W Generationen
- $besteF_W^{(t)}$ = bester Gütewert der letzten W Generationen
- *lineare Skalierung:*

$$Fitness = \frac{A.F - schlechtesteF_W^{(t)}}{besteF_W^{(t)} - schlechtesteF_W^{(t)}}.$$

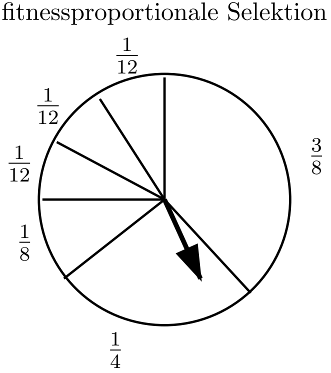
Verbesserungen

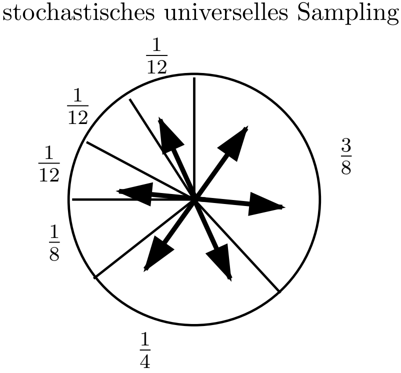
Rangbasierte Selektion

- Idee: Individuen nach ihrem Gütewert sortieren, Fitness nur nach Rang vergeben
- Sei $A^{(1)}.F \succeq A^{(2)}.F \succeq \cdots \succeq A^{(r)}.F$
- Wahrscheinlichkeiten werden linear zugewiesen

$$Pr[A^{(i)}] = \frac{2}{r}\left(1 - \frac{i-1}{r-1}\right)$$

Stochastische Universelles Sampling





Stochastische Universelles Sampling

STOCHASTISCHES-UNIVERSELLES-SAMPLING(Gütewerte $\langle A^{(i)}.F \rangle_{1 \le i \le}$

$Summe_1 \leftarrow$ berechne Fitnesswert aus $A^{(1)}.F$
for $i \in \{2, \ldots, r\}$
do $\ulcorner$ $Fitness \leftarrow$ berechne Fitnesswert aus $A^{(i)}.F$
$\quad\llcorner$ $Summe_i \leftarrow Summe_{i-1} + Fitness$
$u \leftarrow$ wähle Zufallszahl gemäß $U([0, \frac{Summe_r}{s}))$
$j \leftarrow 1$
$I \leftarrow \langle\rangle$
for $i \in \{1, \ldots, s\}$
do $\ulcorner$ **while** $Summe_j < u$
$\qquad$ **do** $\ulcorner$ $j \leftarrow j + 1$
$\qquad\quad\llcorner$ $sum \leftarrow sum + Fitness^{(j)}$
$\qquad u \leftarrow u + \frac{Summe_r}{s}$
$\quad\llcorner$ $I \leftarrow I \circ \langle j \rangle$
return I

Eine Alternative: Turnierselektion (Eltern)

$\textsc{Turnier-Selektion}(\text{Gütewerte } \langle A^{(i)}.F \rangle_{1 \le i \le r})$

$I \leftarrow \langle\rangle$

for $i = 1, \ldots, s$ $(\!|$Anzahl der zu wählenden Individuen$|\!)$

do ⌜ $\mathit{index} \leftarrow$ wähle Zufallszahl gemäß $U(\{1, \ldots, r\})$

  for $j = 2, \ldots, q$ $(\!|$Anzahl der Gegner$|\!)$

  do ⌜ $u \leftarrow$ wähle Zufallszahl gemäß $U(\{1, \ldots, r\})$

    if $A^{(u)}.F \succ A^{(\mathit{index})}.F$

    ⌞ **then** ⌈ $\mathit{index} \leftarrow u$

  ⌞ $I \leftarrow I \circ \langle \mathit{index} \rangle$

return I

Kombinationen Eltern-/Umweltselektion

Umweltselektion ↓ Elternsel. →	uniforme Auswahl	Identität	probabilistisch
Identität	kein Sel.druck	kein Sel.druck	GA
duplikatfrei prob.	?	$\times$	?
(überlappend)	?	EP (90er)	?
prob. überlappend	?	SA	steady state GA
deterministisch	ES (Komma)	kein Sel.druck	?
(überlappend)	ES (Plus)	EP (60er)	steady state GA

Parametrierung von μ und λ

- großer Einfluss auf Erfolg und/oder Geschwindigkeit der Optimierung
- Extrembeispiel: Selektion der Besten
  $\Rightarrow$ Selektionsdruck hängt von $\frac{\mu}{\lambda}$ ab
- Populationsgrößen abstimmen mit
  - Schwierigkeit und der Charakter des Optimierungsproblems,
  - involvierter Selektionsoperator und
  - Erforschung und Feinabstimmung durch Mutation und Rekombination.

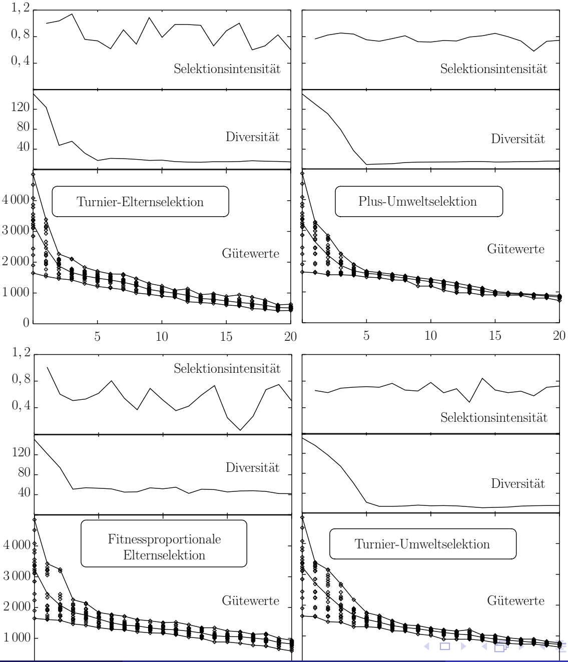
Experimenteller Vergleich

Selektionsszenarien

- Eltern: 3-faches Turnier, Umwelt: Identität
- Eltern: fitnessproportional, Umwelt: Identität
- Eltern: Identität, Umwelt: Besten-Selektion
- Eltern: Identität, 5-tufige 2-fache Turnierselektion

Experimenteller Rahmen

- $\mu = \lambda = 20$
- Sphäre $f(X) = \sum_{i=1}^{n} X_i^2$ mit $n = 2$ (Minimierung)
- Gauß-Mutation


1,2
0,8
0,4
Selektionsintensität
120
80
40
Diversität
4 000
3 000
2 000
1 000
0
Turnier-Elternselektion
Gütewerte
5
10
15
20
Selektionsintensität
Diversität
Plus-Umweltselektion
Gütewerte
5
10
15
20
1,2
0,8
0,4
Selektionsintensität
120
80
40
Diversität
4 000
3 000
2 000
1 000
Fitnessproportionale Elternselektion
Gütewerte
Selektionsintensität
Diversität
Turnier-Umweltselektion
Gütewerte

Überblick

Drittes Prinzip: Verknüpfen mehrerer Individuen

- andere mögliche Wege im Suchraum entstehen durch das Verknüpfen mehrerer Individuen
- drei verschiedene Arbeitsweisen:
  - sprichwörtliche „Rekombination"
  - interpolierende Operatoren
  - extrapolierende Operatoren

Kombinierende Operatoren

Idee

- Bestandteile verschiedener Individuen werden neu zusammengestellt
- es entsteht nichts eigentlich Neues

Beobachtungen

- Arbeitsweise hängt von der hohen Diversität ab
- systematisches Durchforsten des Suchraums

Uniformer Crossover

Uniformer-Crossover(Individuum A, Individuum B)

for $i \in \{1, \ldots, l\}$
do $b \leftarrow$ wähle zufällig gemäß $U(\mathbb{B})$
    if b
    then $C.G_i \leftarrow A.G_i$
        $D.G_i \leftarrow B.G_i$
    else $C.G_i \leftarrow B.G_i$
        $D.G_i \leftarrow A.G_i$
return C, D

Interpolierende Operatoren

Idee

- vermischen die Charakteristika der Eltern
- neue Eigenschaften entstehen

Beobachtungen

- Stabilität statt systematischer Erforschung
- konzentriert Population auf gemeinsamen Nenner (Schwerpunkt der Population)
- fördert Feinabstimmung
- Gegengewicht: diversitätserhaltende Mutation

Arithmetischer Crossover

$\textsc{Arithmetischer-Crossover}$(Individuen A, B mit $A.G,$

```
u ← wähle zufällig aus U([0, 1])
for i ∈ {1, ..., l}
do ⌊ C.G_i ← u · A.G_i + (1 − u) · B.G_i
return C
```

Extrapolierende Operatoren

Idee

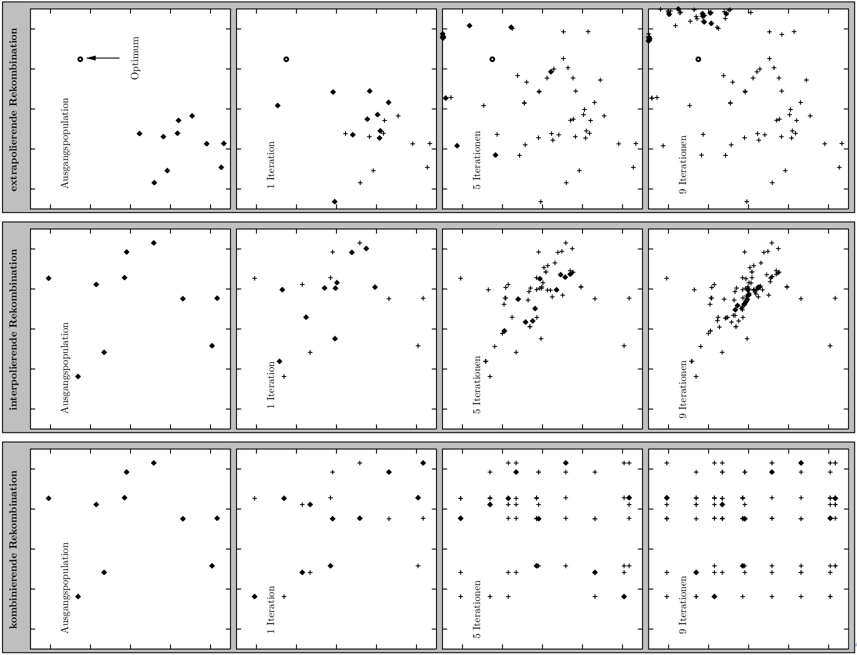
- gezielt Information aus mehreren Individuen ableiten
- Prognose, wo Güteverbesserungen zu erwarten sind

Beobachtungen

- können bisheriges Suchgebiet verlassen
- einzige Rekombination, die Gütewerte benutzt!

Algorithmus

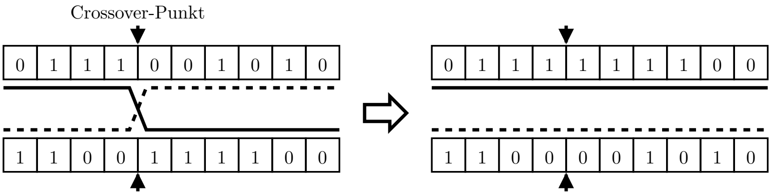
- Arithmetischer Crossover mit u aus $U([1,\ 2])$



extrapolierende Rekombination
Optimum
Ausgangspopulation
1 Iteration
5 Iterationen
9 Iterationen
interpolierende Rekombination
Ausgangspopulation
1 Iteration
5 Iterationen
9 Iterationen
kombinierende Rekombination
Ausgangspopulation
1 Iteration
5 Iterationen
9 Iterationen

Schema-Theorie

- zunächst: genetischer Algorithmus
- 1-Punkt-Crossover
- Frage: was für eine Suchdynamik ergibt sich durch die kombinierende Rekobmination?



GA: 1-Punkt-Crossover

Ein-Punkt-Crossover(Individuen A, B)

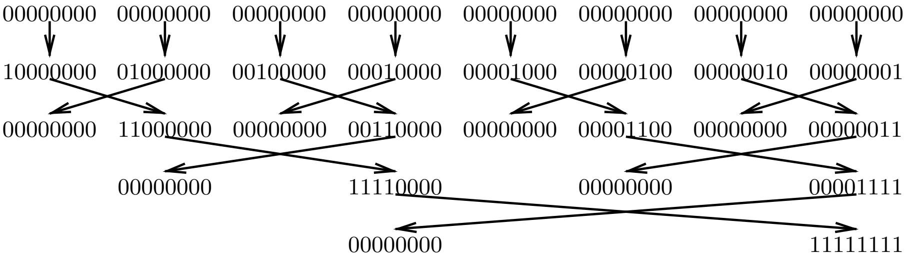
$j \leftarrow$ wähle zufällig gemäß $U(\{1, \ldots, l-1\})$
for $i \in \{1, \ldots, j\}$
do $C.G_i \leftarrow A.G_i$
$D.G_i \leftarrow B.G_i$
for $i \in \{j+1, \ldots, l\}$
do $C.G_i \leftarrow B.G_i$
$D.G_i \leftarrow A.G_i$
return C, D

GENETISCHER-ALGORITHMUS(Zielfunktion F)

$t \leftarrow 0$
$P(t) \leftarrow$ erzeuge Population mit μ ⦇gerade Populationsgröße⦈ Individuen
bewerte $P(t)$ durch F
while Terminierungsbedingung nicht erfüllt
do $P' \leftarrow$ Selektion aus $P(t)$ mittels SELEKTION-FITNESSPROPORTIONAL
    ⦇ Es sei: $P' = \langle A^{(1)}, \ldots, A^{(\mu)} \rangle$ ⦈
    $P'' \leftarrow \langle\rangle$
    for $1 \leq i \leq \frac{\mu}{2}$
    do $u \leftarrow$ wähle Zufallszahl gemäß $U([0,\ 1))$
        if $u \leq p_x$ ⦇Rekombinationswahrscheinlichkeit⦈
        then $\langle B, C\rangle \leftarrow$ EIN-PUNKT-CROSSOVER$(A^{(2i-1)}, A^{(2i)})$
        else $\langle B, C\rangle \leftarrow \langle A^{(2i-1)}, A^{(2i)}\rangle$
        $B \leftarrow$ BINÄRE-MUTATION(B)
        $C \leftarrow$ BINÄRE-MUTATION(C)
        $P'' \leftarrow P'' \circ \langle B, C\rangle$
    bewerte P'' durch F
    $t \leftarrow t + 1$
    $P(t) \leftarrow P''$
return bestes Individuum aus $P(t)$

Motivation

Wie kann sich die Laufzeit durch Kombination ändern?



Grundbegriffe

Def.: Schema

- Sei $\mathcal{G} = \mathbb{B}^l$
- Dann: $H \in \{0, 1, *\}^l$ heißt Schema und beschreibt

$$\mathcal{I}(H) = \left\{ A.G_1 \cdots A.G_l \in \mathcal{G} \;\middle|\; \forall 1 \le i \le l : \; (H_i \ne *) \Rightarrow (A.G_i = H_i) \right\}.$$

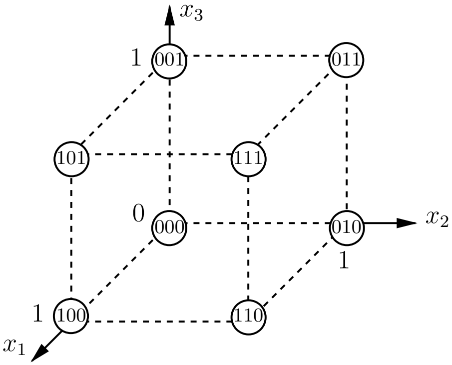
- Ordnung von H:

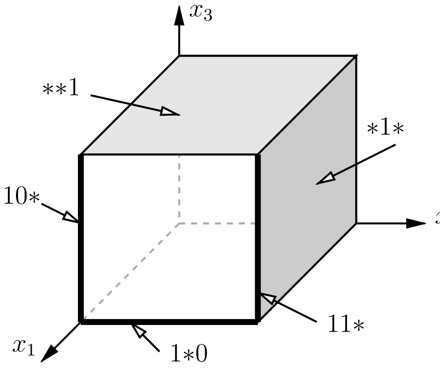
$$o(H) = \# \left\{ i \;\middle|\; (1 \le i \le l) \;\wedge\; (H_i \ne *) \right\}.$$

- definierende Länge von H:

$$\delta(H) = \max \left\{ |i - j| \;\middle|\; (1 \le i, j \le l) \;\wedge\; (H_i \ne *) \;\wedge\; (H_j \ne *) \right\}$$

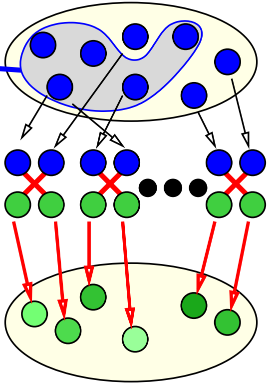
Hyperebenen





Schema-Theorem: Überblick

Ein bestimmter Anteil der Population hat ein Merkmal (d.h. diese Individuen sind Vertreter des Schemas H)



Wie verändert sich dieser Anteil durch Anwendung einer Generation?

- welche Eigenschaften (Schemata) vermehren sich besonders stark?

Schema-Theorem

- GENETISCHER-ALGORITHMUS auf $\mathcal{G} = \mathbb{B}^l$
- für beliebiges Schema H und Population $P(t) = \langle A^{(t,i)} \rangle_{1 \leq i \leq \mu}$ gilt:

$$\mathrm{Erw}\left[p_H^{(t+1)}\right] \geq p_H^{(t)} \cdot \frac{\overline{F}_H^{(t)}}{\overline{F}^{(t)}} \cdot (1 - p_m)^{o(H)} \cdot \left(1 - p_x \cdot \frac{\delta(H)}{l-1} \cdot \left(1 - p_H^{(t)} \cdot \frac{\overline{F}_H^{(t)}}{\overline{F}^{(t)}}\right)\right)$$

$$\text{mit } p_H^{(t)} = \frac{\#\{1 \leq i \leq \mu \mid A^{(t,i)}.G \in \mathcal{I}(H)\}}{\mu},$$

- mit durchschnittlicher Güte $\overline{F}^{(t)}$ in der Population und $\overline{F}_H^{(t)}$ der Schema-Vertreter

Beispiel

Individuum	Güte
10101...	3
01101...	3
01100...	2
11101...	4
11000...	2

Individuum	Güte
00001...	1
10001...	2
01001...	2
11001...	3
01110...	3

Korollar: Einfaches Schema-Theorem

Einfaches Schema-Theorem

Durch einige einfache Abschätzungen erhält man die folgende bekanntere Form des Schema-Theorems:

$$\mathrm{Erw}\left[p_H^{(t+1)}\right] \geq p_H^{(t)} \cdot \frac{\overline{F}_H^{(t)}}{\overline{F}^{(t)}} \cdot \left(1 - o(H) \cdot p_m - p_x \cdot \frac{\delta(H)}{l-1}\right).$$

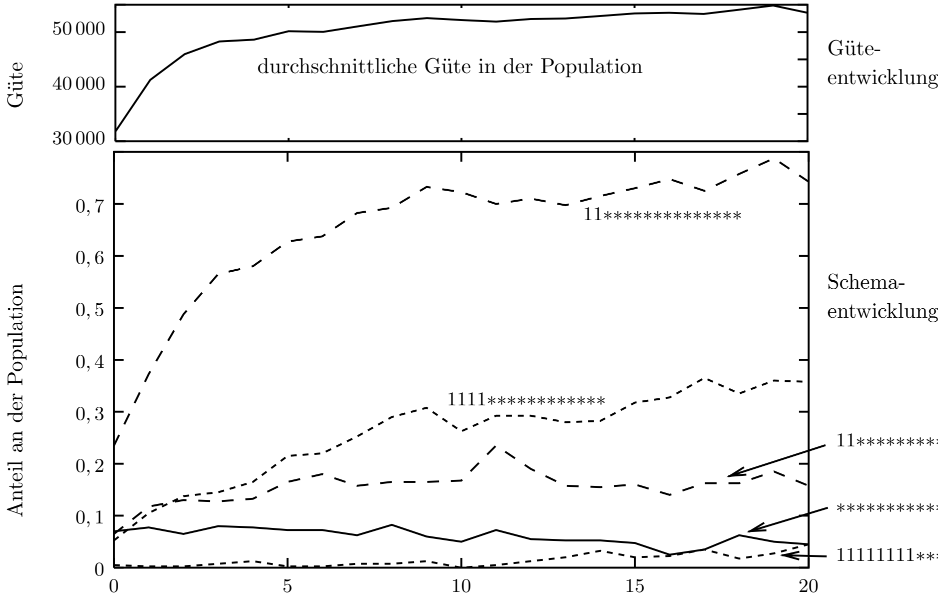
Bausteine

Schemata mit

- überdurchschnittlicher Güte,
- kleiner definierender Länge und
- geringer Ordnung

Beispiel

- Rekombinationswahrscheinlichkeit $p_x = 1.0$
- Mutationsrate $p_m = \frac{1}{16}$
- Genotyp: mit 16 Bit enkodierte Zahl (Maximum: 65536)
- Populationsgröße 400


Güte
50 000
40 000
30 000
durchschnittliche Güte in der Population
Güte-
entwicklung
Anteil an der Population
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
0
5
10
15
20
11**************
1111************
Schema-
entwicklung

Kritik am Schema-Theorem

- Relevanz der Aussage
  - Ergebnis lässt sich nicht iterativ anwenden
    $\Rightarrow$ Bausteinhypothese (exponentielles Wachstum von Schemata) gilt nicht
  - implizite Annahme: Kombination guter Schemata verbessert die Güte
    $\Rightarrow$ muss nicht gelten!
- Randbedingungen der evolutionären Algorithmen
  - Wahrscheinlichkeitsaussagen in sehr kleinen Mengen sind fragwürdig
  - beobachtete Güte eines Schemas muss nicht der tatsächlichen entsprechen
    Stichpunkt: Konvergenz einzelner Bits

„Fehlendes" Schema-Theorem

- laut Altenberg: Es fehlt ein Schema-Theorem mit einer Aussage zum Suchfortschritt
- Ausgangssituation:
  - Genetischer-Algorithmus mit fitnessprop. Selektion und ohne Mutation
  - allgemeine Rekombination $Merk \subseteq \{1, \ldots, l\}$
  - mögliche Eltern von Individuum A: $H_{Merk}(A.G)$ und $H_{\widetilde{Merk}}(A.G)$ mit $\widetilde{Merk} = \{1, \ldots, l\} \setminus Merk$
  - kommen mit Wahrscheinlichkeiten p_{Merk} vor

Verknüpfung Schema/Suchfortschritt

Auf zwei Ebenen:

- am Ende im Ergebnis: Wie ändert sich die durchschnittliche Güte?
- durch Verknüpfen von Eltern und Kindgüte

$$\mathrm{Cov}\left[A.F, \frac{\overline{F}^{(t)}_{H_{Merk}(A.G)} \cdot \overline{F}^{(t)}_{H_{\widetilde{Merk}}(A.G)}}{\left(\overline{F}^{(t)}\right)^2}\right]$$

Lemma

Es gilt:

$$\mathrm{Cov}\left[A.F, \frac{\overline{F}^{(t)}_{H_{Merk}(A.G)} \cdot \overline{F}^{(t)}_{H_{\widetilde{Merk}}(A.G)}}{(\overline{F}^{(t)})^2}\right]$$

$$= \sum_{A.G \in \mathcal{G}} \left(A.F - \overline{F}^{(t)}\right) \cdot \frac{\overline{F}^{(t)}_{H_{Merk}(A.G)} \cdot \overline{F}^{(t)}_{H_{\widetilde{Merk}}(A.G)}}{(\overline{F}^{(t)})^2} \cdot p_{A.G} \ ,$$

mit Häufigkeit $p_{A.G}$ von Individuum A in der Population

Beispiel:

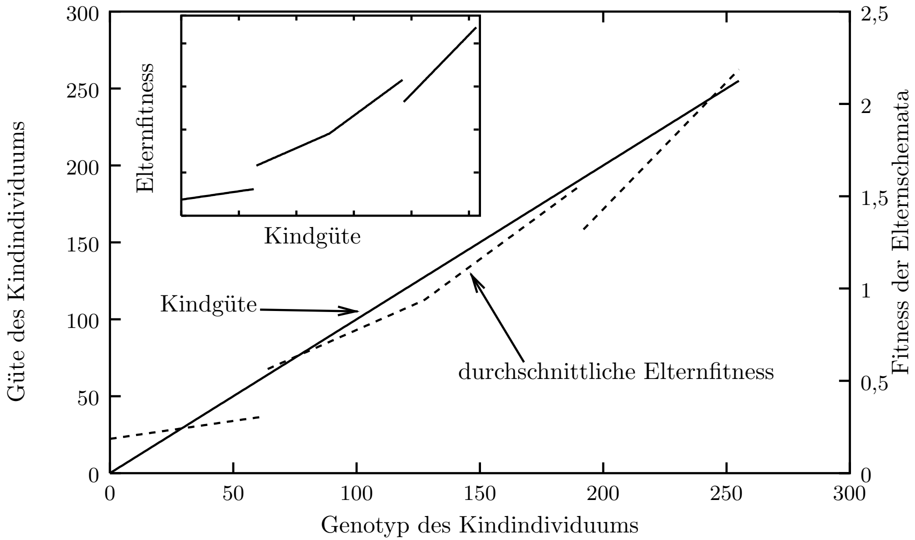
- $\mathcal{G} = \mathbb{B}^8$ enkodiert standardbinär $\{0, \ldots, 255\}$
- Rekombination:

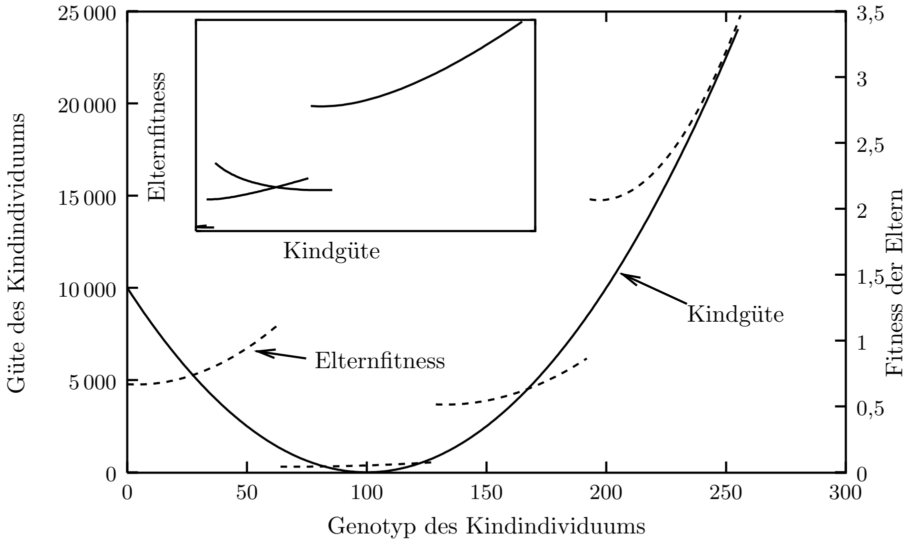
Crossover-Punkt

1	0	0	1	0	1	1	0

Kindindividuum

$$\left.\begin{array}{cccccccc} 1 & 0 & * & * & * & * & * & * \\ * & * & 0 & 1 & 0 & 1 & 1 & 0 \end{array}\right\} \text{mögliche Eltern}$$


Güte des Kindindividuums
300
250
200
150
100
50
0
Elternfitness
Kindgüte
Kindgüte
durchschnittliche Elternfitness
0
50
100
150
200
250
300
Genotyp des Kindindividuums
2,5
2
1,5
1
0,5
0
Fitness der Elternschemata


Elternfitness
Kindgüte
Elternfitness
Kindgüte
Güte des Kindindividuums
Fitness der Eltern
Genotyp des Kindindividuums
25 000
20 000
15 000
10 000
5 000
0
3,5
3
2,5
2
1,5
1
0,5
0
0
50
100
150
200
250
300

Satz: „Fehlendes“ Schema-Theorem

Für einen genetischen Algorithmus nur mit Rekombination gilt:

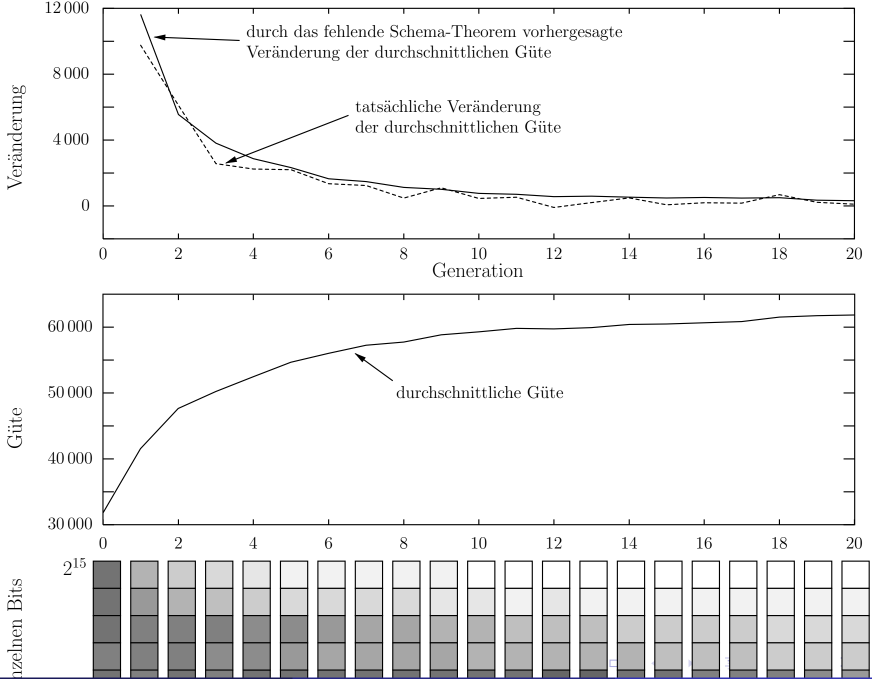
$$\mathrm{Erw}\left[\overline{F}^{(t+1)}\right]-\overline{F}^{(t)} = \sum_{Merk\subseteq\{1,\dots,l\}} p_{Merk}\cdot\left(\mathrm{Cov}\left[A.F, \frac{\overline{F}_{H_{Merk}(A.G)}\cdot\overline{F}_{H_{\widetilde{Merk}}(A.G)}}{(\overline{F}^{(t)})^2}\right]\right.$$

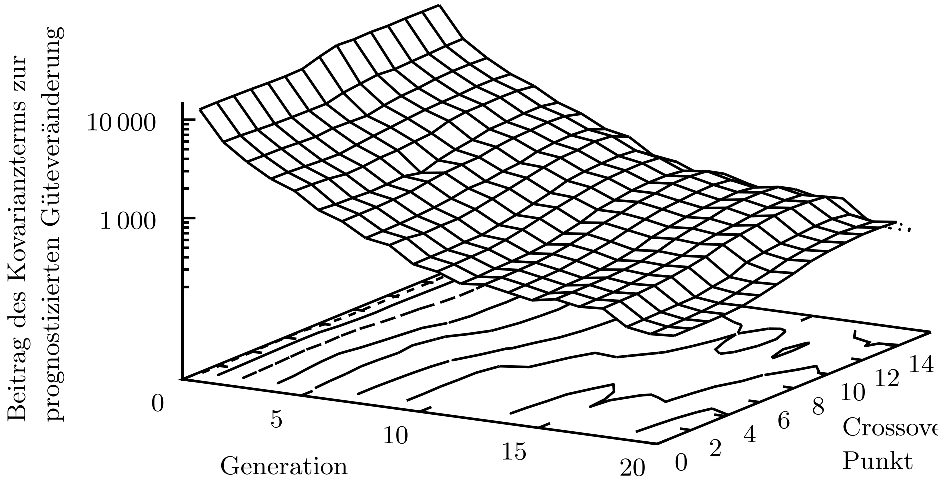
$$\left. - \sum_{A.G\in\mathcal{G}} \left(p_{A.G} - p_{H_{Merk}(A.G)}\cdot p_{H_{\widetilde{Merk}}(A.G)}\right)\cdot\left(A.F - \overline{F}^{(t)}\right)\cdot\frac{\overline{F}_{H_{Merk}(A.G)}\cdot\overline{F}_{H_{\widetilde{Merk}}(A.G)}}{(\overline{F}^{(t)})^2}\right)$$

mit Häufigkeiten $p_{H_{Merk}(A.G)}$ und $p_{H_{\widetilde{Merk}}(A.G)}$ der erzeugenden Schemata von $A.G$ und durchschnittlichen Schematagütewerten $\overline{F}_{H_{Merk}(A.G)}$ und $\overline{F}_{H_{\widetilde{Merk}}(A.G)}$.

Beispiel

- 1-Punkt-Crossover
- Rekombinationswahrscheinlichkeit $p_x = 1.0$
- keine Mutation
- Genotyp: mit 16 Bit enkodierte Zahl (Maximum: 65536)
- Populationsgröße 400


12 000
8 000
4 000
0
Veränderung
durch das fehlende Schema-Theorem vorhergesagte
Veränderung der durchschnittlichen Güte
tatsächliche Veränderung
der durchschnittlichen Güte
0 2 4 6 8 10 12 14 16 18 20
Generation
60 000
50 000
40 000
30 000
Güte
durchschnittliche Güte
0 2 4 6 8 10 12 14 16 18 20
2^{15}
nzelnen Bits


Beitrag des Kovarianzterms zur
prognostizierten Güteveränderung
10 000
1 000
0
5
10
15
20
Generation
0
2
4
6
8
10
12
14
Crossove
Punkt

Überblick

1. Wechselspiel Variation – Selektion
2. Populationskonzept
3. Verknüpfen mehrerer Individuen
4. Selbstanpassende Algorithmen
5. Zusammenfassung

Einfluss des Stands der Suche

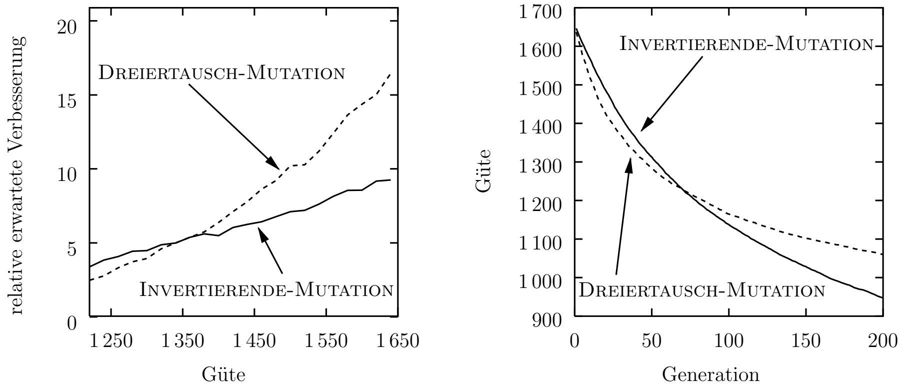
Grundgedanke

- bisher: argumentiert, dass Mutation eine kleine Veränderung bzgl. des Phänotyps vornehmen soll
- jetzt: gilt das immer zu jedem Zeitpunkt der Optimierung?

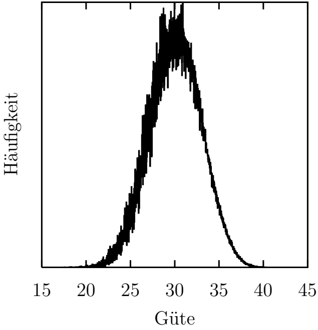
Kontrollexperiment

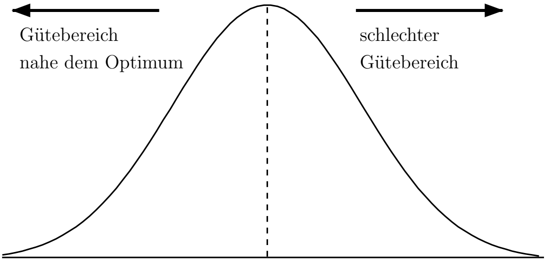
- Hillclimbing-Szenario
- Handlungsreisendenproblem
- unterschiedlich lokale Mutationsoperatoren:
    - Invertieren eines Teilstücks
    - Tausch dreier Städte

Vergleich: Geschwindigkeit

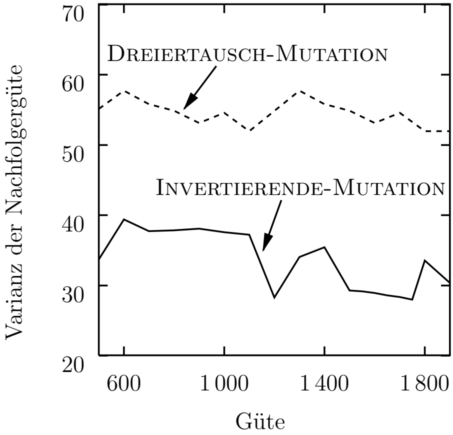


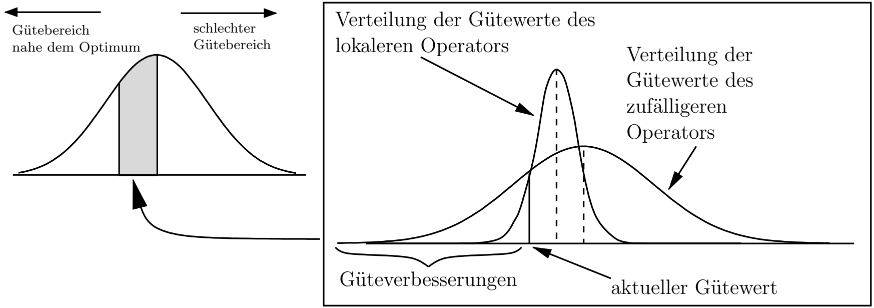
Gesamter Suchraum

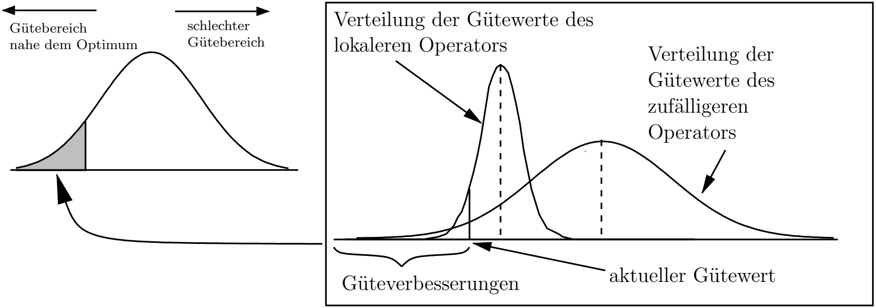




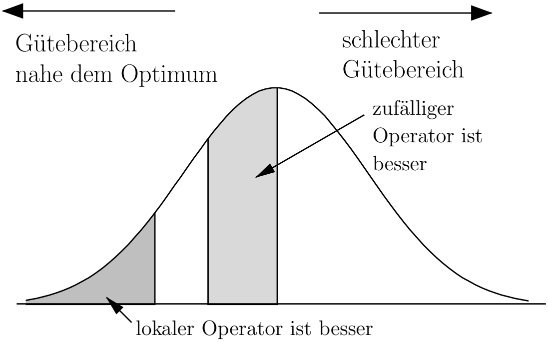
Varianz der erzeugten Gütewerte




Gütebereich
nahe dem Optimum
schlechter
Gütebereich
Verteilung der Gütewerte des
lokaleren Operators
Verteilung der
Gütewerte des
zufälligeren
Operators
Güteverbesserungen
aktueller Gütewert


Gütebereich
nahe dem Optimum
schlechter
Gütebereich
Verteilung der Gütewerte des
lokaleren Operators
Verteilung der
Gütewerte des
zufälligeren
Operators
Güteverbesserungen
aktueller Gütewert

(Beweisbares) Ergebnis der Überlegungen



Anpassungstechniken

Vordefiniert

- die Veränderung wird vorab festgelegt

Adaptiv

- es werden Maßzahlen für die Angepasstheit erhoben
- aufgrund von Regeln, wird daraus die Anpassung abgeleitet

Selbstadaptiv

- Zusatzinformationen im Individuum werden genutzt
- zufallsbasiert sollen sich Parameter individuell einstellen

Vordefinierte Anpassung

Betrachteter Parameter

- reellwertige Gauss-Mutation
- σ bestimmt die durchschnittliche Schrittweite

Umsetzung

Vordefinierte-Anpassung(Standardabweichung σ)

$\sigma' \leftarrow \sigma \cdot \alpha$ ⟦Modifikationsfaktor⟧

return σ'

- z.B. mit $\alpha = 0,98$

Adaptation

Umsetzung

- Maß: Anteil der verbessernden Mutationen in den letzten k Generationen
- falls dieser „hoch“ ist, σ vergrößern

ADAPTIVE-ANPASSUNG(Standardabweichung σ, Erfolgsrate p_s)

if $p_s > \Theta$ ⦇Schwellwert⦈
then $\sigma' \leftarrow \sigma \cdot \alpha$ ⦇Modifikationsfaktor($\alpha > 1$)⦈
else **if** $p_s < \Theta$
    then $\sigma' \leftarrow \frac{\sigma}{\alpha}$
    else $\sigma' \leftarrow \sigma$
return σ'

Selbstadaptation

Umsetzung

- die Standardabweichung σ bei der Erzeugung des Individuums wird in den Zusatzinformationen gemerkt
- sprich: *Strategieparameter*
- wird beim Mutieren wieder leicht zufällig variiert
- „gute“ Werte für σ setzen sich durch bessere Güte der Kinder durch

Experimenteller Vergleich

Testumgebung

- 10-dimensionale Sphäre
- Hillclimber
- aber: pro Generation werden $\lambda = 10$ Kindindividuen erzeugt
- reellwertige Gauß-Mutation mit $\sigma = 1$
- Umweltselektion des Besten aus Elter und Kindern

Selbstadaptation

Selbstadaptive-Gauss-Mutation(Individuum A)

$u \leftarrow \mathcal{N}(0,1)$

$B.S \leftarrow A.S\ e^{\frac{1}{\sqrt{l}}u}$

for $i \in \{1, \ldots, l\}$

do $u_i \leftarrow$ wähle zufällig gemäß $\mathcal{N}(0, B.S)$

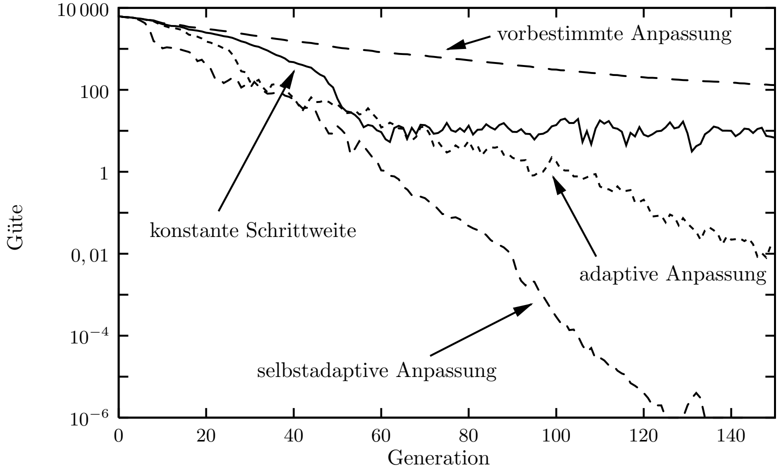
$B_i \leftarrow A_i + u_i$

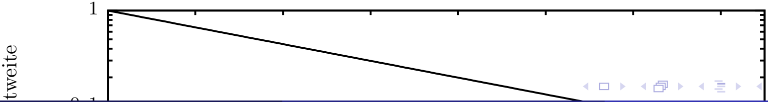
$B_i \leftarrow \max\{B_i, ug_i$ (untere Wertebereichsgrenze)$\}$

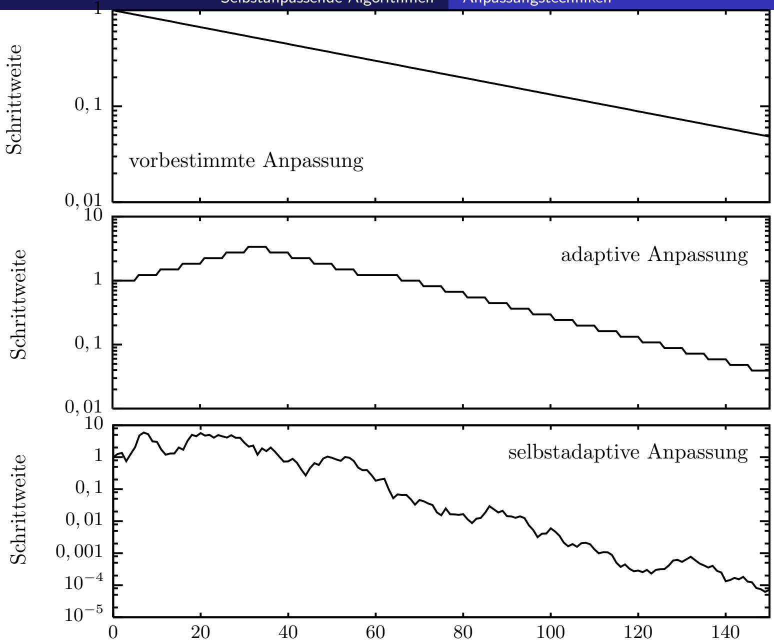
$B_i \leftarrow \min\{B_i, og_i$ (obere Wertebereichsgrenze)$\}$

return B

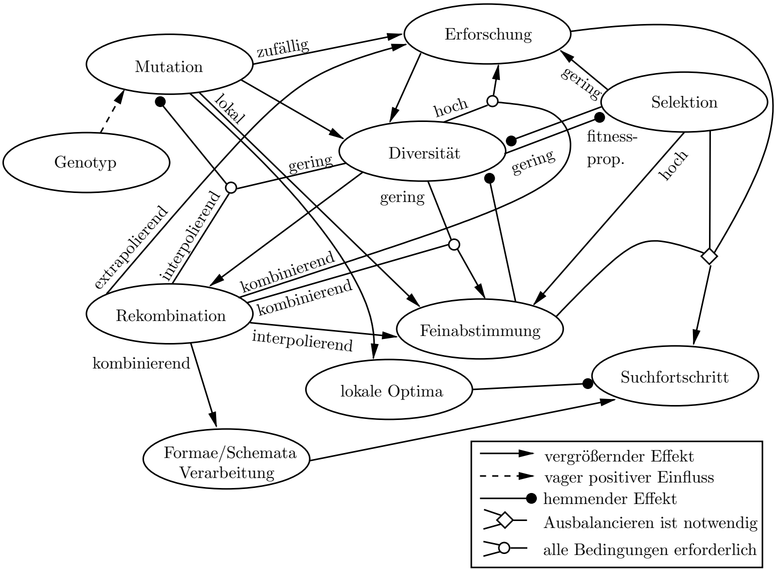
Ergebnis des Vergleichs






1
0, 1
0, 01
Schrittweite
vorbestimmte Anpassung
10
1
0, 1
0, 01
Schrittweite
adaptive Anpassung
10
1
0, 1
0, 01
0, 001
10^{-4}
10^{-5}
Schrittweite
selbstadaptive Anpassung
0
20
40
60
80
100
120
140

Überblick


Erforschung
Mutation
zufällig
Selektion
gering
lokal
hoch
Genotyp
Diversität
gering
gering
fitness-
prop.
gering
hoch
extrapolierend
interpolierend
kombinierend
kombinierend
Rekombination
Feinabstimmung
interpolierend
kombinierend
Suchfortschritt
lokale Optima
Formae/Schemata
Verarbeitung
vergrößernder Effekt
vager positiver Einfluss
hemmender Effekt
Ausbalancieren ist notwendig
alle Bedingungen erforderlich

Zusammenhänge (1)

Bedingung	Zielgröße	Erwarteter Effekt
Genotyp	Mutation	Nachbarschaft des Mutationsoperators wird beeinflusst
Mutation	Erforschung	zufälligere Mutationen unterstützen die Erforschung
Mutation	Feinabst.	gütelokale Mutationen unterstützen die Feinabstimmung
Mutation	Diversität	die Mutation vergrößert die Diversität
Mutation	lokale Optima	gütelokale Mutationen erhalten lokale Optima des Phänotyps, zufälligere Mutation können sogar mehr einführen
Rekombination	Erforschung	extrapolierende Operatoren stärken die Erforschung
Rekombination	Feinabst.	interpolierende Operatoren stärken die Feinabstimmung

Zusammenhänge (2)

Bedingung	Zielgröße	Erwarteter Effekt
Div./Rekomb.	Mutation	geringe Diversität und interpolierende Rekombination dämpft Ausreisser der Mutation
Diversität	Rekombination	hohe Diversität unterstützt die Funktionsweise der Rekombination
Selektion	Erforschung	geringer Selektionsdruck stärkt die Erforschung
Selektion	Feinabst.	hoher Selektionsdruck stärkt die Feinabstimmung
Selektion	Diversität	Selektion verringert meist die Diversität
Div./Rekomb.	Erforschung	kombinierende Rekombination stärkt die Erforschung bei hoher Diversität
Div./Rekomb.	Feinabst.	kombinierende Rekombination stärkt die Feinabstimmung bei geringer Diversität

Zusammenhänge (3)

Bedingung	Zielgröße	Erwarteter Effekt
Erforschung	Diversität	erforschende Operationen erhöhen die Diversität
Feinabst.	Diversität	feinabstimmende Operationen verringern die Diversität
Diversität	Selektion	geringe Diversität verringert den Selektionsdruck der fitnessproportionalen Selektion
Rekombination	Forma-Verarb.	Rekombination gemäß den Forma-Regeln unterstützt das Schema-Theorem
Forma-Verarb.	Suchfortschritt	Erfolgreiche Forma-Verarbeitung unterstützt den Suchfortschritt
lokale Optima	Suchfortschritt	viele lokale Optima hemmen den Suchfortschritt
Erf./Fein./Sel.	Suchfortschritt	Ausbalancieren der drei Faktoren ist notwendig