

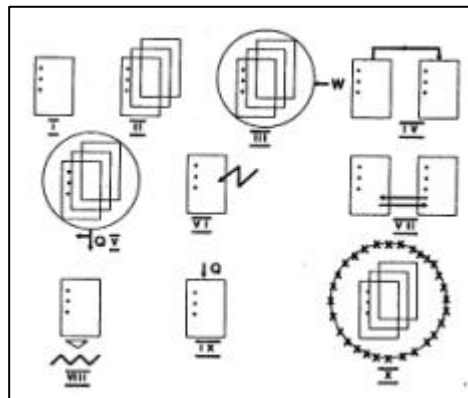
## Evolutionsstrategien

### Ursprünge der Evolutionsstrategien

Die Entwicklung der Evolution resultiert aus den Überlegungen die biologischen Evolution auf nichtbiologische Probleme zu übertragen. Dabei haben sich zwei Modelle herausentwickelt, die sich für Computersimulationen und Anwendungen in der Informatik eignen: Die Evolutionsstrategien und die genetischen Algorithmen. Die Evolutionsstrategie, die im folgenden beschrieben werden soll, wurde in den sechziger Jahren von Ingo Rechenberg an der TU Berlin entwickelt. Sie wurden später von ihm, sowie auch von anderen Forschern, insbesondere von Hans-Paul Schwefel, verbessert und erweitert. Am Anfang waren die Modelle noch sehr abstrakt und vernachlässigten viele Details der biologischen Evolution. Dies wurde später behoben, doch Rechenberg hielt eine naturgetreue Nachbildung der Evolution nicht für sinnvoll. Genauer schrieb er in seinem Buch: „Nun ist es falsch zu glauben, ein natürliches Vorbild müsse möglichst genau kopiert werden, um höchste Vollkommenheit in der technischen Ebene zu erreichen. (...) Das Verstehen des jeweiligen biologischen Vorganges kann ebenso wertvoll sein, wenn es darauf gelingt, ein idealisiertes Schema zu entwerfen, das die gleiche Wirkung hervorbringt.“ Es gibt demnach verschiedene Varianten der ES, die jeweils unterschiedliche Details der biologischen Evolution stärker berücksichtigen. Daraus resultiert jedoch auch ein Dilemma: Wie exakt muß oder sollte man die Evolution nachahmen; oder kann es sogar passieren, daß man sie überrepräsentiert. Diese Frage spaltete die Forscher dann auch in zwei Lager, auf der einen Seite die deutsche Schule um Rechenberg, die die biologische Evolution nur als Richtschnur zur Entwicklung von Optimierungsverfahren benutzt und auf der anderen Seite die amerikanische Schule der Verfechter der genetischen Algorithmen um Holland und Goldberg, die sich eher dafür interessieren, wie es der Evolution gelingt, Informationen zu codieren und zu verarbeiten und diese über die Generationen hinweg weiterzureichen. Leider arbeiten die beiden Schulen nicht zusammen, sondern gehen völlig getrennte Wege.

### Rechenbergsche Grafik-Notation

Rechenberg gelingt es mit nur 10 Symbolen auszukommen, um alle Vorgänge des Evolutionsmodells darzustellen. Um Individuen darzustellen, verwendet er Karten mit Punkten darauf, die die Gene darstellen. Mehrere Karten hintereinander stellen eine Population dar. Ein Kreis um die Karten herum stellt eine Urne für eine Auswahl dar. Ein Pfeil, der mit «w» markiert auf eine Urne zeigt, bedeutet eine zufällige Auswahl. Führt ein mit «Q» markierter Pfeil von der Urne weg, findet die Auswahl mit Hilfe einer Qualitätsfunktion statt. Zeigt ein mit «Q» markierter Pfeil auf eine Karte, so bedeutet dies eine Bewertung des Individuums. Die Umwandlung abstrakter Erbinformationen in ein Individuum ist durch eine Zickzacklinie angedeutet. Auf den nachfolgenden Phänotypen zeigt



dann vielfach ein mit «Q» beschrifteter Pfeil, der andeutet, daß er nun bewertet werden kann. Symbol IV stellt eine einfache Duplikation eines Individuums dar. Ein blitzartiger Pfeil, der auf ein Individuum zeigt stellt eine Mutation dar, Symbol VII schließlich eine Rekombination. Ist die Urne gekreuzt (Symbol X), so handelt es sich um eine isolierte Population.

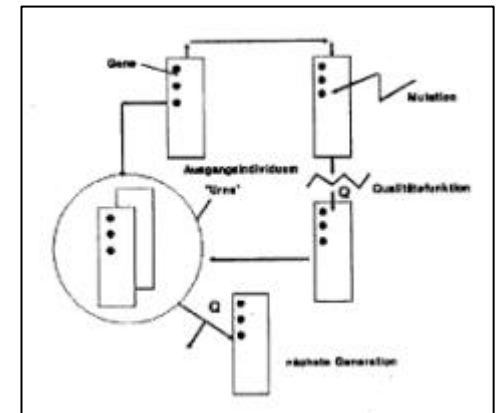
### Codierung

Bei der Anwendung von Evolutionsstrategien geht es vielfach um eine Optimierung von Parametern. Die Erbinformation, die ein Individuum besitzt, wird deshalb als Vektor von reellen Zahlen dargestellt. Die komplizierte Genstruktur und Codierung der Chromosomen wurde also sehr vereinfacht. Für die frühen Evolutionsstrategen gab es daher nur zwei wesentliche Informationen: die Werte der zu optimierenden Parameter und die Werte der zu optimierenden Qualitätsfunktion.

### Modelle der Evolutionsstrategie

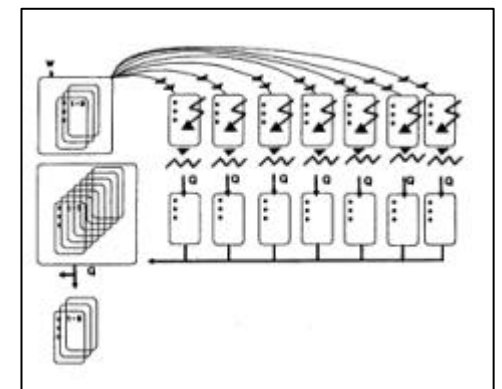
#### (1+1)-Evolutionsstrategie

Das Modell (1+1) ist die einfachste Form der Evolutionsstrategie. Hier wird ein „UR-Individuum“, ein Elter, dupliziert; dieser Nachkomme wird dann mutiert. Bei der Darstellung als Vektor bedeutet dies, daß zu jeder Komponente des Vektors ein zufälliger kleiner positiver oder negativer Wert addiert wird. Elter und Kind kommen zusammen in eine Urne. Das Individuum, das den besseren Wert für die Qualitätsfunktion liefert, wird als neuer Elter ausgewählt. (survival of the fittest) Dies wiederholt sich immerfort bis ein definiertes Abbruchkriterium erreicht ist. Darauf wird später noch genauer eingegangen. Dieses Modell läßt die sexuelle Rekombination außer acht. Trotz der Einfachheit findet man sinnvolle Anwendungen, Dank der Möglichkeiten der Schrittweitenregelung der Mutation.



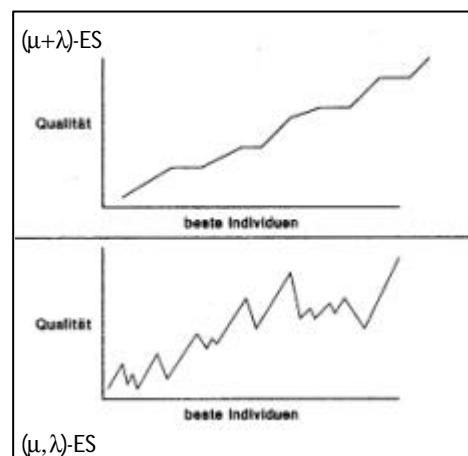
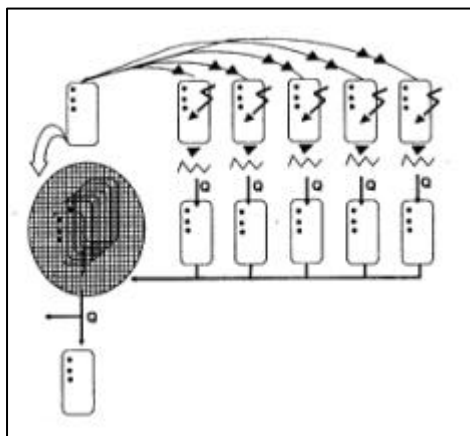
#### (μ-λ)-Evolutionsstrategie

Dieses Modell ist die Verallgemeinerung der (1+1)-Evolutionsstrategie. Man verwendet  $\mu$  Eltern, die  $\lambda$  Kinder erzeugen. Dabei muß gelten:  $\lambda \geq \mu \geq 1$ . Die Eltern, die die Kinder Erzeugen, werden zufällig mit gleicher Wahrscheinlichkeit ausgewählt. Von den  $\mu$  Eltern und  $\lambda$  Kindern, die zusammen in die Urne geworfen werden, werden die  $\mu$  besten Individuen als neue Eltern herausgenommen. Die Qualität der besten Individuen wird so von Generation zu Generation nie schlechter.



**(m)**-Evolutionsstrategie

Bei der  $(\mu+\lambda)$ -Evolutionsstrategie überleben die  $\mu$  besten aus den Eltern und Kindern. Das hat zur Folge, daß Individuen mehrere Generationen hindurch überleben können, wenn ihre Qualitätsfunktion besser ist als die aller anderen. In gewissem Maße ist das gewollt und es kommt in der Natur auch durchaus vor. Dies kann jedoch von Nachteil sein, z.B., wenn man sich mit einem solchen Individuum in einem lokalen Optimum befindet. Die Komma-Notation, 1975 von Schwefel eingeführt, behebt dieses Manko. Die  $\mu$  besten, die überleben, werden nur noch aus den  $\lambda$  Nachkommen ausgewählt. Man kommt damit der biologischen Evolution näher; es gibt keine unsterblichen Individuen mehr; jedes Individuum lebt nur noch für die Dauer einer Generation. Auch wenn in der Realität einige Lebewesen mehrere Generationen überleben, so ist dieser Ansatz trotzdem plausibler als Unsterblichkeit.

**(m/#)**-Evolutionsstrategie

Bislang wurde die sexuelle Rekombination außer Acht gelassen. Es wurde deshalb die  $(\mu/p\#\lambda)$ -Evolutionsstrategie eingeführt. «#» steht dabei für «+» oder «,». Das Erzeugen der Duplikate geht anders als bei der  $(\mu\#\lambda)$ -Evolutionsstrategie vonstatten. Die sexuelle Rekombination wird in diesem Modell zusätzlich auch noch gegenüber dem biologischen Vorbild verallgemeinert. Die Nachkommen können nämlich nicht nur von 2 Eltern (Mutter und Vater) erzeugt werden, sondern von  $p$  Eltern.

Es werden also  $p$  Vektoren rekombiniert um einen Nachkommen zu erzeugen. Dazu gibt es zwei Möglichkeiten:

- Entsprechende Werte (an gleicher Position) der  $p$  Vektoren werden gemittelt. Der Vektor der gemittelten Werte wird dem Kind zugeordnet. Die Erbinformation eines jeden Elter geht also mit  $1/p$  ein.
- Der Wert eines Elter geht komplett in den Nachkommen ein, dh. daß für jede Position des Nachkommen-Vektors zufällig ein Elter ausgesucht wird, dessen Wert der entsprechenden Position übernommen wird.

Bei einer  $(5/3,7)$ -Evolutionsstrategie werden aus 5 Eltern  $E_1$ - $E_5$  7 Mal zufällig 3 ausgewählt, die jeweils ein Kind erzeugen. Aus diesen 7 werden wieder 5 Eltern ausgesucht. Ein solcher Schritt sei im folgenden beschrieben.

Beispiel:  $E_1=[a_1,b_1,c_1]$ ,  $E_2=[a_2,b_2,b_2]$ ,  $E_3=[a_3,b_3,c_3]$ ,  $AW=[2,3,1] \Rightarrow K_1=[a_2,b_3,c_1]$

Aus den 5 Eltern seien zufällig  $E_1$ ,  $E_2$  und  $E_3$  ausgewählt worden. Im Auswahl-Vektor  $AW$  sei zufällig festgelegt, welcher Elter eine bestimmte Position des Kind-Vektors bestimmt.  $E_2$  gibt hier seinen ersten Wert,  $E_3$  den zweiten und  $E_1$  den dritten. Jede dieser Auswahlen erfolgt per Zufall mit Gleichverteilung.

**Parameter der Evolutionsstrategie**

Das Verhalten der Evolution kann variiert werden. Ein Parameter, dessen man sich dazu bedienen kann, ist der Selektionsdruck. Er ist definiert durch den Quotienten  $s=\mu/\lambda$ . Aus den  $\lambda$  Kindern gehen  $\mu$  neue Eltern hervor.  $s$  liegt zwischen 0 und 1 und gibt an wie stark die Auslese ist. Je kleiner der Wert, desto stärker ist die Auslese.

Es lassen sich auch Populationswellen simulieren, indem man  $\mu$  systematisch oder zufallsgesteuert verändert. Soll der Selektionsdruck gleich bleiben, muß man nur  $\mu$  und  $\lambda$  im gleichen Maße verändern. In der Natur kann der Selektionsdruck zum Beispiel durch das Vorhandensein von natürlichen Feinden verändert werden. Bei gleichbleibendem  $\lambda$  (Nachwuchs) variiert  $\mu$  durch das Verändern von  $s$ . Bleibt der Feind aus ( $s$  groß), so ändert sich in erster Näherung das Populationswachstum, die Individuen vermehren sich. Biologisch interessant sind periodische und zyklische Variationen von  $s$ .

Ein weiterer Parameter ist die Mutationsschrittweite. Sie ist die Größe des Zahlenwertes, der bei der Mutation zu den Komponenten des Vektors hinzugezählt wird. Kleine Werte verändern in der Regel die Qualitätsfunktion nur wenig, große hingegen verändern sie stark und können eine Lösungsfindung sogar verhindern. Hier muß man also einen angemessenen Wertebereich finden.

Schließlich muß noch festgelegt werden, wie lange die Evolution simuliert werden muß. Es müssen also Abbruchkriterien definiert werden. Im großen und ganzen gibt es drei verschiedene solcher Kriterien:

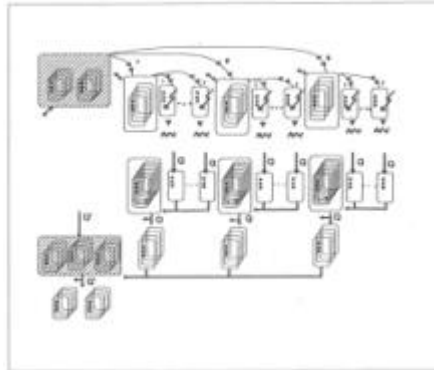
- Am einfachsten ist es, abzubrechen, wenn die Qualitätsfunktion eines Nachkommen hinreichend gut ist. Voraussetzung ist natürlich, daß man ein Optimum definieren kann. Zudem muß sichergestellt sein, daß man sich nicht in einem lokalen Optimum wiederfindet, da dieses oft nutzlos ist.
- Wenn die Optimale Lösung nicht bekannt ist, verwendet man oft das Überschreiten einer gewissen Zeit als Abbruchkriterium.
- Ein anderes, doch ähnliches, Kriterium stellt der Abbruch nach einer gewissen Anzahl erzeugter Generationen dar. Man gibt also sozusagen die Anzahl der Iterationsschritte vor.

Bei den beiden zuletzt genannten Kriterien besteht jedoch die Gefahr, daß man das Finden der optimalen Lösung durch zu frühzeitigen Abbruch verhindert.

### Evolutionsstrategien mit Populationen

Die bislang kennengelernten ES arbeiten auf der Basis einzelner Individuen. Ihre Theorie und Notation können jedoch auch leicht auf ES mit Populationen übertragen werden. Rechenberg benutzt hierfür neben den runden Klammern zusätzliche eckige Klammern.

Während die runden Klammern sich weiterhin auf die Entwicklung von Individuen beziehen, sollen die eckigen Klammern jene von Populationen beschreiben: [2,3(4,7)] Dieser Ausdruck beschreibt eine ES, bei der zwei unabhängige Populationen jeweils drei neue Populationen erzeugen. Dies erfolgt völlig analog zur Erzeugung der Duplikate von Individuen. Die nebenstehende Abbildung zeigt die entsprechende graphische Darstellung.



Man beachte, daß auch bei einer auf Populationen basierenden ES jede beliebige Kombination von «+» und «,» innerhalb und außerhalb der runden Klammern sinnvoll ist.

Zur Bestimmung der (in diesem Fall zwei) besten Populationen muß ein Qualitätsmaß festgelegt werden. Mehrere Möglichkeiten sind denkbar, eine naheliegende ist die mittlere Qualität aller Individuen einer entsprechenden Population.

Mithilfe der «/»-Notation ist es möglich – ähnlich der sexuellen Rekombination von Individuen, bei der Gene ausgetauscht werden – ganze Individuen zwischen Populationen auszutauschen, die sich parallel entwickeln. Bei einer [4/2#3(5#7)]-ES zum Beispiel werden aus jeweils 4 Populationen 2 zufällig zur Erzeugung von 3 neuen Populationen ausgewählt, welche dann zu je 3 Zweiergruppen zusammengefaßt und gemischt werden, indem die Individuen der beiden Populationen jeweils zufällig untereinander ausgetauscht werden. Jede einzelne der auf diese Weise erzeugten neuen Populationen werden anschließend nach dem Schema einer (5#7)-ES weiterverarbeitet.

Eine hochgestellte Zahl beschreibt die Dauer der Isolation einer Population. Hiermit wird versucht, die Entwicklung von abgeschotteten, isolierten Populationen nachzubilden. [1,2(4,7)<sup>30</sup>] beschreibt demnach eine ES-Variante, bei der die beiden Populationen von 4 Individuen, die jeweils 7 Nachkommen erzeugen, 30 Generationen lang isoliert werden. Dabei nutzt dies Modell zweifach die implizite Parallelität der ES aus (parallele Optimierung von Populationen, in denen jeweils wiederum die Individuen optimiert werden).

Auf diese Weise lassen sich beliebig iterierte ES darstellen. {y'/x'#z"[y'/x'#z'(y/x#z)<sup>1</sup>]<sup>m</sup>}<sup>n</sup>-ES und noch tiefere Schachtelungen sind möglich, haben sich jedoch in der Praxis nicht als sinnvoll erwiesen. Aus dieser Idee wurde die Standard-ES geboren, die bei der zweiten Iterationsstufe endet und demnach folgende Form hat: [u/v#w(x/y#z)<sup>n</sup>]-ES

Angesichts der schier unendlichen Vielfalt der Variationsmöglichkeiten der Strategieparameter kommt die Frage auf, ob es eventuell optimale Einstellungen der Parameter  $\mu$ ,  $\rho$ ,  $\lambda$  für eine  $(\mu/\rho/\lambda)$ -ES gibt. Rechenberg empfiehlt je nach Ziel folgende Strategien:

„glatte“ Qualitätsfunktionen	gleichmäßige Konvergenz	hohe Zuverlässigkeit
1#5-ES	1#10-ES	1/5 < $\mu$ / $\lambda$ < 1/3

Diese „optimalen“ Strategieparameter sind jedoch lediglich Richtwerte. Es sei hier außerdem noch erwähnt, daß die Notation der Evolutionsstrategien in mehrerer Hinsicht unvollständig und mehrdeutig ist. Es ist zum Beispiel nicht möglich, innerhalb einer [u#v(w'x)<sup>n</sup>]-ES auszudrücken, daß sich die v isolierten Populationen zwar simultan aber dennoch unterschiedlich entwickeln sollen.

### Wahrscheinlichkeitstheorie

Bevor später auf Mutationen eingegangen wird, ist es sinnvoll, ein paar einfache Begriffe der Wahrscheinlichkeitstheorie aufzufrischen.

Es seien n nicht notwendigerweise verschiedene Werte  $a_1, \dots, a_n$  gegeben.

Mittelwert	Varianz	Standardabweichung
$m := \frac{1}{n} \sum_{i=1}^n a_i$	$s^2 := \frac{1}{n} \sum_{i=1}^n (a_i - m)^2$	$s := \sqrt{s^2} = \sqrt{\frac{1}{n} \sum_{i=1}^n (a_i - m)^2}$

Oft ist es sinnvoll, mit den standardisierten Reihenwerten zu arbeiten ( $m=0, s^2=s=1$ ):

$$a_1^*, a_2^*, \dots, a_n^* \quad a_i^* := \frac{a_i - m}{s}$$

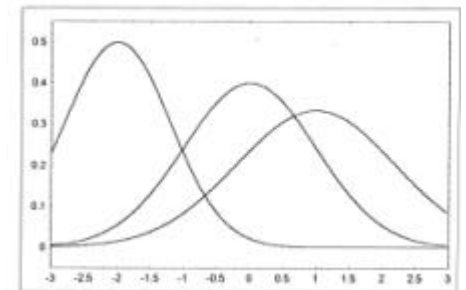
Die Wahrscheinlichkeitsfunktion einer Zufallsvariable X ist definiert als:  $f(x_i) = W(X = x_i)$

Hiermit läßt sich die Verteilungsfunktion angeben:  $F(x) = W(X \leq x)$  Die Ableitung der Verteilungsfunktion wird als Dichtefunktion bezeichnet:  $F'(x) = f(x)$

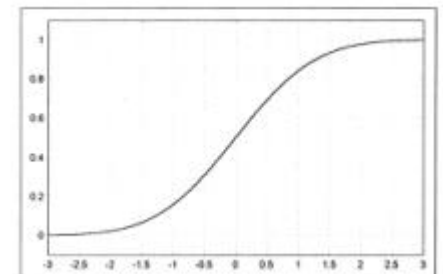
Eine diskrete Zufallsvariable heißt gleichverteilt, wenn gilt:  $W(X = x_i) = 1/k$  für  $i = 1, 2, \dots, k$

Die Normalverteilung ist die wichtigste statistische Verteilung, da viele Eigenschaften in der Natur normalverteilt auftreten. Ihre Dichtefunktion wird oft als Gauß'sche Glockenkurve bezeichnet (Abb. rechts). Sie ist definiert zu:

$$f(x, m, s) = \frac{1}{\sqrt{2\pi s^2}} e^{-\frac{(x-m)^2}{2s^2}}$$



Die Standardnormalverteilung ist gekennzeichnet durch die Parameter  $m=0$  und  $s=1$ . Die Abbildung unten zeigt ihre Verteilungsfunktion.



### Mutationen

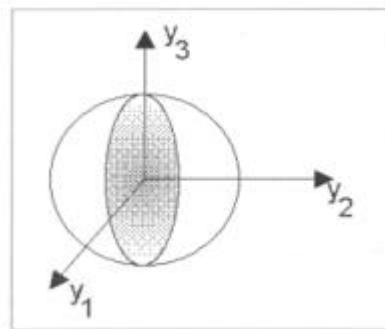
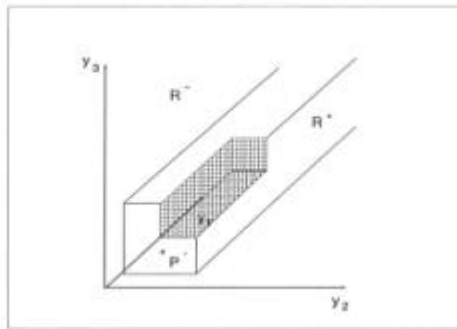
Die Evolutionsstrategien interessieren sich insbesondere für die Mutationsstrategie. Ein neues Individuum wird erzeugt, indem das Duplikat eines Ausgangsindividuums nach dem Prinzip der statistischen Normalverteilung mutiert wird. Die Selektion der Ausgangsindividuen, deren Mischung und Rekombination erfolgt hingegen nach dem Prinzip der statistischen Gleichverteilung.

Formal läßt sich die Erzeugung mutierter Nachkommen wie folgt beschreiben:  $x_{neu} = x_{alt} + N(0, \sigma)$  mit  $\sigma$  fest, jedoch nicht notwendigerweise gleich für alle Vektorkomponenten. Dabei wird keine bestimmte Suchrichtung bevorzugt; dieser Mutationsprozeß erzeugt potentiell um jeden Elter ein Sphäre (Hyperkugel) von Nachkommen, die mit wachsendem radialen Abstand zu den Eltern lichter wird.

**Optimierung der Parameter**

Da bei einer kleinen Streuung der Zufallszahlen die Fortschrittsgeschwindigkeit der Suche häufig sehr langsam ist, bei einer großen Streuung hingegen die Gefahr unnötiger Sprünge im Suchraum besteht, stellt sich die Frage, wie die optimale Streuung der Zufallszahlen aussehen soll. Rechenberg beantwortet diese Frage analytisch für zwei einfache Qualitätsfunktionen:

Korridormodell	Sphärenmodell
$Q_{korr}(x_1, \dots, x_n) := a + cx_i \quad x_i \in [-b, +b] \forall i > 1$	$Q_{sphär}(x_1, \dots, x_n) := x_1^2 + \dots + x_n^2$
$Q_{korr}(x_1, \dots, x_n) := -\infty \quad sonst$	



$s_{korr} = \sqrt{2p \cdot b} / n$	$s_{sphär} = 1,224 \cdot r / n$
$p_{korr} = 0,184$	$p_{sphär} = 0,27$

Bei den von Rechenberg ermittelten optimalen Streuungen  $\sigma$  liegen die Erfolgswahrscheinlichkeiten  $p$  für den Fortschritt in der Suche in den Hyperräumen durch Mutation nahe bei 1/5. Hieraus hat er die berühmte 1/5-Erfolgsregel für die Anpassung der Mutationsschrittweiten (adaptive Schrittweitenregelung) formuliert:

Der Quotient aus den erfolgreichen Mutationen (also den Mutationen, die eine Verbesserung der Qualität bewirken) zu allen Mutationen sollte mindestens 1/5 betragen. Ist der Quotient größer als 1/5 so sollte die Streuung der Mutationen erhöht werden; ist der Quotient geringer (also die Mutationen seltener erfolgreich), so sollte die Streuung verringert werden.

Dies läßt sich zusammenfassen zu:

$p > 1/5 \rightarrow$ Streuung erhöhen, $p < 1/5 \rightarrow$ Streuung verringern
---

Schwefel hat anhand von Simulationen einen algorithmischen Ansatz für die Veränderung der Schrittweiten nach  $n$  Generationen gefunden:

$\sigma(t+n) = 1,22 \cdot \sigma(t)$ wenn $p > 1/5$
$\sigma(t+n) = 0,82 \cdot \sigma(t)$ wenn $p < 1/5$
$\sigma(t+n) = \sigma(t)$ wenn $p = 1/5$

Die 1/5-Erfolgsregel ist hart codiert und wird nicht problemabhängig automatisch angepaßt. Schwefel schlägt deshalb vor, ein Individuum als Tupel  $\langle x, s \rangle$  aufzufassen, wobei  $x$  der gewöhnliche Vektor und  $\sigma$  die komponentenbezogene Standardabweichung ist. Damit wird der Steuerungsparameter selbst Objekt der Evolution. Die neuen Werte berechnen sich wie folgt:

$$s_{neu} := s_{neu} e^{N(0, \Delta)} \quad x_{neu} := x_{neu} + e^{N(0, s_{neu})}$$

wobei  $\Delta$  ein frei wählbarer Parameter ist, der die Größe der Anpassung der Schrittweite bestimmt.

Rechenberg hat einen ähnlichen Mechanismus für die Mutation des Individuums vorgeschlagen:

$$x_{neu} = x_{alt} + \text{Schrittweite} \cdot N(0, 1/\sqrt{n}) \quad \text{mit Schrittweite} := x \cdot y$$

$x$  ist konstant und die eigentliche Schrittweite, der Faktor  $y$  wird adaptiert und entweder deterministisch oder zufällig gewählt. Bei einer  $(\mu, \lambda)$ -ES schlägt Rechenberg für jedes Individuum folgende Bestimmung von  $y$  vor:

$$y_{neu} = y_{alt} \cdot 1,5 \quad \text{oder} \quad y_{neu} = y_{alt} / 1,5$$

Neben der zufälligen Bestimmung der Schrittweitenadaptation sind auch deterministische Ansätze denkbar.

Es muß jedoch noch einmal betont werden, daß es für die Wahl dieser Parameter und der Schrittweite bislang keine fundierte Theorie gibt. Alle hier angegebenen Werte können deshalb nur als Richtwerte dienen.

Wichtig ist außerdem, nicht zu vergessen, daß die Mutationen und die adaptive Schrittweitenanpassung die ausgefeiltesten, jedoch auch die kritischsten Teile der Evolutionsstrategien sind. Eine geeignete Mutationsstrategie und eine gute Schrittweitenanpassung garantieren den Erfolg. Um diesen Erfolg abschätzen zu können, hat Rechenberg den Begriff des Evolutionsfensters geprägt. Das Evolutionsfenster ist das in der Regel extrem schmale Band der für die Evolution sinnvollen und effektiven Schrittweiten. Liegen die Mutationen rechts von diesem Bereich, kann es zu einem evolutiven Rückschritt kommen, da große, unkontrollierte Änderungen die Fortschrittsgeschwindigkeit empfindlich hemmen; liegen die Mutationen links vom Evolutionsfenster wird die Suche nach dem Optimum aufgrund der zu geringen Änderungen stagnieren. Die Schwierigkeit bei der Anwendung der Evolutionsstrategien besteht demnach darin, mit den Mutationen möglichst oft in das Evolutionsfenster zu treffen.

