
Asymmetrische Evolutionstrategien

Dissertation
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
der Universität Dortmund
am Fachbereich Informatik
von

Lars Hildebrand

Dortmund
2001

Tag der mündlichen Prüfung: 22. 2. 2002

Dekan: Prof. Dr. Th. Herrmann

Gutachter: Prof. Dr. B. Reusch, Prof. Dr. W. Banzhaf

Inhaltsverzeichnis

Kapitel 1	Motivation & Einführung	9
1.1	Motivation	9
1.2	Einführung	9
Kapitel 2	Symmetrische Evolutionsstrategien	13
2.1	Grundlagen der Evolutionsstrategien	14
2.1-1	Stochastische Grundlagen	14
2.1-2	Repräsentation	17
2.1-3	Fitness	18
2.1-4	Mutation	18
2.1-5	Rekombination	18
2.1-6	Selektion	20
2.2	(1+1)-Evolutionsstrategie	21
2.2-1	Grundlegender Algorithmus	21
2.2-2	Konvergenzverhalten der (1+1)-Evolutionsstrategie	23
2.2-3	Die 1/5-Erfolgsregel	24
2.3	(μ +1)-Evolutionsstrategie	25
2.4	(μ + λ)-Evolutionsstrategie	27
2.4-1	Grundlegender Algorithmus	27
2.4-2	Repräsentation	28
2.4-3	Rekombination	29
2.4-4	Mutation der Strategiekomponenten – Selbstadaptation	29
2.4-5	Adaptation einer Schrittweite	29
2.4-6	Adaptation von n Schrittweiten	31
2.4-7	Adaptation von n Schrittweiten und $(n(n-1))/2$ Lagewinkeln	32
2.4-8	Mutation der Objektkomponenten	33
2.5	Zusammenfassung	34

Kapitel 3	Erweiterungen der Evolutionsstrategien	35
3.1	Variation der Mutationsfunktion	35
3.2	Kumulative Schrittweitenregelung (KSA)	37
3.2-1	Grundidee	37
3.2-2	$(\mu/1, \lambda)$ -KSA-ES	38
3.2-3	$(\mu/1\mu, \lambda)$ -KSA-ES	40
3.2-4	Eigenschaften der $(\mu/1, \lambda)$ - und $(\mu/1\mu, \lambda)$ -KSA-ES	41
3.3	Kovarianzmatrixadaptation (CMA)	41
3.3-1	Grundidee	41
3.3-2	$(\mu/1\mu, \lambda)$ -CMA-ES	42
3.3-3	Eigenschaften der $(\mu/1\mu, \lambda)$ -CMA-ES	43
3.4	Differentielle Evolution (DE)	44
3.4-1	Tastschritt	44
3.4-2	Rekombination	45
3.4-3	Lokale Differentielle Evolution	46
3.4-4	Eigenschaften der DE	46
3.5	Evolutionary Gradient Search (EGS)	46
3.5-1	Tastschritte	47
3.5-2	Individuelle Steuerung der Schrittweiten	48
3.5-3	Eigenschaften der EGS	49
3.6	Robust Evolution Strategies (RES)	49
3.6-1	Cauchy Mutation	49
3.6-2	Robuste Adaptation	50
3.6-3	Eigenschaften der RES	52
3.7	Lokale Selektion - Diffusionsmodelle (LES)	52
3.7-1	Lokale Rekombination	53
3.7-2	Lokale Selektion	53
3.7-3	Eigenschaften der LES	53
3.8	Adaptation einer nicht-zentralen Normalverteilung	54
3.8-1	Mutation	54
3.8-2	Eigenschaften des Verfahrens	55
3.9	Simulated Binary Crossover (SBX)	55
3.9-1	Grundidee	57
3.9-2	Eigenschaften des SBX	58
3.10	Estimation of Distribution Algorithms (EDA)	59
3.11	Zusammenfassung	61

Kapitel 4	Asymmetrische Verteilungsfunktionen	63
4.1	Eigenschaften von Wahrscheinlichkeitsverteilungen	64
4.1-1	Modus	65
4.1-2	Quantile	65
4.1-3	Momente	66
4.1-4	Erwartungswert (1. Moment)	67
4.1-5	Varianz (2. Moment)	68
4.1-6	Schiefe (3. Moment)	69
4.2	Mutationsfunktionen	70
4.3	Asymmetrische Verteilungsfunktion $\Xi_c(x)$	72
4.3-1	Verteilungsfunktion einer $\Xi_c(x)$ -verteilten Zufallsvariablen	76
4.3-2	$\Xi_c(x)$ als Mutationsfunktion	78
4.3-3	Inverse Verteilungsfunktion einer $\Xi_c(x)$ -verteilten Zufallsvariablen	80
4.3-4	Erwartungswert einer $\Xi_c(x)$ -verteilten Zufallsvariablen	83
4.3-5	Varianz einer $\Xi_c(x)$ -verteilten Zufallsvariablen	84
4.3-6	Schiefe einer $\Xi_c(x)$ -verteilten Zufallsvariablen	88
4.4	Asymmetrische Verteilungsfunktion $\Xi_{c,\sigma}(x)$	90
4.4-1	Verteilungsfunktion einer $\Xi_{c,\sigma}(x)$ -verteilten Zufallsvariablen	92
4.4-2	$\Xi_{c,\sigma}(x)$ als Mutationsfunktion	94
4.4-3	Inverse Verteilungsfunktion einer $\Xi_{c,\sigma}(x)$ -verteilten Zufallsvariablen	97
4.4-4	Erwartungswert einer $\Xi_{c,\sigma}(x)$ -verteilten Zufallsvariablen	98
4.4-5	Varianz einer $\Xi_{c,\sigma}(x)$ -verteilten Zufallsvariablen	99
4.4-6	Schiefe einer $\Xi_{c,\sigma}(x)$ -verteilten Zufallsvariablen	100
4.5	Asymmetrische Verteilungsfunktion $\Xi_{c,\sigma}^\gamma(x)$	102
4.5-1	Verteilungsfunktion einer $\Xi_{c,\sigma}^\gamma(x)$ -verteilten Zufallsvariablen	103
4.5-2	$\Xi_{c,\sigma}^\gamma(x)$ als Mutationsfunktion	105
4.5-3	Inverse Verteilungsfunktion einer $\Xi_{c,\sigma}^\gamma(x)$ -verteilten Zufallsvariablen	108
4.5-4	Erwartungswert einer $\Xi_{c,\sigma}^\gamma(x)$ -verteilten Zufallsvariablen	109
4.5-5	Varianz einer $\Xi_{c,\sigma}^\gamma(x)$ -verteilten Zufallsvariablen	111
4.5-6	Schiefe einer $\Xi_{c,\sigma}^\gamma(x)$ -verteilten Zufallsvariablen	112
4.6	Anmerkungen zur Wahl der Grundfunktion	114
4.6-1	Alternative Verteilungsfunktionen	114
4.6-2	Verteilungssysteme	115
4.6-3	Alternative Grundfunktionen	127
4.7	Spezielle Eigenschaften der Ξ -Verteilungen	129
4.7-1	Symmetrie	129
4.7-2	Vollständige Äquivalenz zur Normalverteilung	130
4.7-3	Fallweise Äquivalenz zur Normalverteilung	131
4.7-4	Stetigkeit	131
4.7-5	Differenzierbarkeit	133
4.7-6	Summe von Ξ -verteilten Zufallsvariablen	134
4.7-7	Eigenschaften der Summe	141
4.7-8	Zentraler Grenzwertsatz	144
4.8	Zusammenfassung	148

Kapitel 5	Implementierung	149
5.1	Eigenschaften von Zufallszahlengeneratoren	149
5.1-1	Genauigkeit	150
5.1-2	Effizienz	152
5.1-3	Robustheit	153
5.1-4	Reproduzierbarkeit	153
5.2	Typen von Zufallszahlengeneratoren	153
5.2-1	Transformation der Inversen	154
5.2-2	Kompositionsmethode	155
5.2-3	Konvolutionsmethode	156
5.2-4	Akzeptanz/Ablehnungs-Methode	156
5.3	Approximation der inversen Gaußschen Fehlerfunktion	158
5.4	Der Zufallszahlengenerator $A_{\gamma}(c, \sigma)$	161
5.4-1	Realisierung	161
5.4-2	Komplexität	164
5.5	Zusammenfassung	164
Kapitel 6	Asymmetrische Evolutionsstrategien	167
6.1	Asymmetrische Evolutionsstrategie	167
6.1-1	Grundlegender Algorithmus	167
6.1-2	Repräsentation	168
6.1-3	Rekombination	168
6.1-4	Mutation	169
6.1-5	Selbstadaptation der Strategieparameter	170
6.2	Eigenschaften der asymmetrischen Evolutionsstrategien	173
6.2-1	Klassische Beweisstrategien	173
6.2-2	Asymmetrischer Konvergenzmechanismus	174
6.2-3	Reversibilität	177
6.3	Empfehlungen zur Parametrisierung	181
6.4	Zusammenfassung	183
Kapitel 7	Experimentelle Resultate	185
7.1	Varianten des Kugelmodells	185
7.1-1	f_1 - Kugelmodell	187
7.1-2	f_2 - alternierendes Kugelmodell	189
7.1-3	f_3 - steigendes Kugelmodell	190
7.1-4	f_4 - Doppelsumme	192
7.1-5	f_5 - alternierende Doppelsumme	193
7.1-6	f_6 - steigende Doppelsumme	195
7.1-7	Bewertung der Resultate	196

7.2	Problemklassen nach Hansen und Ostermeier	197
7.2-1	B_1 - Ebenenmodell	197
7.2-2	B_2 - Kugelmodell	198
7.2-3	B_4 - Doppelsumme	199
7.2-4	B_5 - Rosenbrock-Funktion	200
7.2-5	B_6 - Ellipse	201
7.2-6	B_7 - Zigarre	203
7.2-7	B_8 - Tablette	204
7.2-8	B_9 - Summe verschiedener Potenzen	206
7.2-9	B_{10} - Parabelgrat	207
7.2-10	B_{11} - Spitzer Grat	208
7.2-11	Bewertung der Resultate	210
7.3	Hoch-dimensionale Probleme	210
7.3-1	B_1 - Ebenenmodell	210
7.3-2	f_1 - Kugelmodell	211
7.3-3	f_4 - Doppelsumme	212
7.4	Optimierung von unscharfen Mengen zur Beschreibung von Farben	213
7.4-1	Problembeschreibung	213
7.4-2	Qualitätsmerkmale	214
7.4-3	Unscharfe Farbverarbeitung	215
7.4-4	Transformation der Farbmodelle	216
7.4-5	Generierung unscharfer Farbbeschreibungen	217
7.4-6	Repräsentation	218
7.4-7	Fitnessfunktion	219
7.4-8	Resultate	220
7.4-9	Bewertung der Resultate	221
7.5	Unscharfe Inferenzsysteme als Funktionsapproximatoren	221
7.5-1	Problembeschreibung	221
7.5-2	Repräsentation	224
7.5-3	Fitnessfunktion	225
7.5-4	Resultate	226
7.5-5	Bewertung der Resultate	228
7.6	Optimierung von Schraubenrotorprofilen	229
7.6-1	Problembeschreibung	229
7.6-2	Repräsentation	230
7.6-3	Resultate	231
7.7	Zusammenfassung	231
Kapitel 8	Zusammenfassung & Ausblick	233
8.1	Zusammenfassung	233
8.2	Ausblick	236

Kapitel 9	Literatur	239
Anhang A	Verwendete Symbole	253
Anhang B	Abkürzungsverzeichnis	267
Anhang C	Abbildungsverzeichnis	269
Anhang D	Algorithmenverzeichnis	275
Anhang E	Vollständige Übersicht über die Experimente	277
E.1	B1 - Ebenenmodell in hohen Dimensionen	277
E.2	f_1 - Kugelmodell in hohen Dimensionen	281
E.3	f_4 - Doppelsumme in hohen Dimensionen	285
Anhang F	Stichwortverzeichnis	289

Kapitel 1 Motivation & Einführung

*Es gibt keine Klasse im Leben für Anfänger, es ist immer
gleich das Schwierigste, was von einem verlangt wird.
(Rainer Maria Rilke)*

1.1 Motivation

Während meiner Zeit als wissenschaftlicher Mitarbeiter am Fachbereich Informatik der Universität Dortmund habe ich an unterschiedlichen ingenieur-wissenschaftlichen Problemen gearbeitet. Viele dieser Problemstellungen konnten auf teilweise sehr komplexe Optimierprobleme zurückgeführt werden. Bedingt durch die Struktur dieser Optimierprobleme und durch die Spezialisierung im Studium fanden Evolutionsstrategien bereits früh ihren Einsatz zur Lösung dieser Problemstellungen. Eines der ersten Beispiele aus diesem Bereich ist im Abschnitt 7.5 (Unschärfe Inferenzsysteme als Funktionsapproximatoren, Seite 221) ausführlich beschrieben. Dieses Beispiel stammt aus den Bereichen Wissenstransformation und Expertensysteme. Evolutionsstrategien haben hier eine bemerkenswerte Leistungsfähigkeit bei der Umwandlung von numerischem, exaktem Wissen in eine unscharfe, auf Fuzzy-Logik basierende, Darstellungsform gezeigt. Gleichzeitig wurde jedoch eine hohe Sensibilität bezüglich einer geeigneten Initialisierung der Startpopulation festgestellt. Grund für diese Sensibilität und damit auch für das Scheitern des eingesetzten Optimierverfahrens im Falle einer ungeeigneten Initialisierung schien die verbesserungswürdige Fähigkeit zur Richtungsadaptation der Evolutionsstrategie zu sein. Eine ausführliche Beschreibung dieses Phänomens findet sich in dem o. g. Abschnitt dieser Arbeit. Als Lösung dieses Problems ist die asymmetrische Evolutionsstrategie entstanden, die über eine ausgeprägte Fähigkeit zur Richtungsadaptation verfügt. Sie ist das Thema dieser Arbeit.

1.2 Einführung

Evolutionsstrategien sind Optimierverfahren, die zur Lösung komplexer Optimierprobleme verwendet werden, bei denen klassische Verfahren aufgrund der Problemkomplexität oder fehlender mathematischer Eigenschaften des Optimierproblems nicht eingesetzt werden können. Das grundlegende Einsatzgebiet von Evolutionsstrategien ist die Klasse der reellwertigen Optimierprobleme. Evolutionsstrategien werden erfolgreich in vielen Teilbereichen der Ingenieurwissenschaften und Informatik eingesetzt.

Ausgehend von den klassischen Evolutionsstrategien, die Ende der 60'er und Anfang der 70'er Jahre von Rechenberg und Schwefel entwickelt wurden, sind die Evolutionsstrategien in vielen

Richtungen erweitert worden. Mutationsoperationen wurden verändert bzw. neue Operatoren entwickelt. Evolutionsstrategien, die Eigenschaften der letzten Generationen verwenden, um Informationen über mögliche Gradienten oder geeignete Optimierrichtungen zu gewinnen, gehören ebenfalls zu diesen Neuentwicklungen. Ein Teil dieser Arbeiten beschäftigte sich mit einer speziellen Eigenschaft der Evolutionsstrategien, der Selbstadaptation. Selbstadaptation ermöglicht es einer Evolutionsstrategie, ein inneres Modell des Optimierproblems zu erlernen, um so einen möglichst optimalen Fortschritt bei der Optimierung zu erzielen. Von Bedeutung ist hierbei die Anzahl der Strategieparameter sowie die Skalierung der Parameter mit steigender Dimensionalität des Optimierungsproblems. Viele Varianten der Evolutionsstrategien haben quadratischen oder kubischen Zuwachs der Strategieparameter in Relation zur Problemdimensionalität. Die Gesamtzahl an Veränderungen, Erweiterungen und Neuentwicklungen auf dem Gebiet Evolutionsstrategien zeigt, dass die klassischen Varianten Verbesserungen zulassen.

Untersucht man diese neuen Varianten der Evolutionsstrategien, stellt man jedoch fest, dass es keine einheitliche Entwicklungsrichtung bei diesen Verbesserungen gibt. Es existieren zwar Varianten, die bei speziellen Optimierproblemen gute Ergebnisse liefern, da Steuerparameter geeignet auf diese Problemstellungen angepasst werden können. Es fällt jedoch auf, dass jede neue Variante der Evolutionsstrategien einen eigenen Satz an Steuerparametern mit sich bringt, diese Parameter unter den einzelnen Varianten nicht austauschbar sind und die einzelnen Varianten zudem für spezielle Problemklassen optimiert wurden und damit bei anderen Problemklassen versagen können.

Das Ziel dieser Arbeit ist die Entwicklung eines Optimierverfahrens, das folgende Eigenschaften besitzt:

- geringe Anforderungen an mathematische Eigenschaften
Das zu entwickelnde Optimierverfahren soll auch dann anwendbar sein, wenn die zu optimierende Funktion nicht differenzierbar ist oder eventuell überhaupt nicht in geschlossener Form darstellbar ist.
- Robustheit
Bei vielen Leistungsbewertungen wird das zu bewertende Optimierverfahren günstig, d. h. in der Nähe des Optimums initialisiert. Bei realen Anwendungen ist dies oft nicht möglich, da eine initiale Abschätzung über die Lage des Optimums nicht möglich ist. Das zu entwickelnde Optimierverfahren soll auch bei ungünstiger Initialisierung in der Lage sein, das Optimum zuverlässig zu finden.
- Verzicht auf externe Steuerparameter
Viele Optimierverfahren benötigen problemspezifische Sätze und Belegungen von Steuerparametern. Diese Parameter sowie deren Belegungen müssen zum Teil aufwendig bestimmt werden und gelten in der Regel nur für wenige, gleichartige Probleme. Das zu entwickelnde Optimierverfahren soll vollständig auf externe Steuerparameter verzichten.
- lineare Steigerung des systembedingten Rechenaufwandes
Der systembedingte Rechenaufwand, d. h. der Anteil am Gesamtrechenaufwand, der nicht durch die Komplexität der zu optimierenden Funktion bedingt ist, soll maximal linear mit der Problemdimensionalität steigen. So soll vermieden werden, dass das zu entwickelnde Optimierverfahren für geringe Problemdimensionalität gute Ergebnisse liefert, für höhere Problemdimensionalität jedoch bereits an der Komplexität des Optimierverfahrens scheitert.

Das Haupteinsatzgebiet des zu entwickelnden Optimierverfahrens wird die reellwertige Optimierung sein. Aus diesem Grund wird als Ausgangspunkt für die Entwicklung die Evolutionsstrategie bzw. eine ihrer Varianten gewählt. Evolutionsstrategien erfüllen von sich aus bereits die Anforderungen bezüglich der mathematischen Eigenschaften sowie in Teilen die Forderung nach linearer Steigerung des systembedingten Rechenaufwandes. In Bezug auf Robustheit und Verzicht

auf externe Steuerparameter erfüllen sie die Anforderungen nur eingeschränkt. Die in dieser Arbeit vorgestellte Variante der Evolutionsstrategien, die asymmetrische Evolutionsstrategie, verwendet ausschließlich eine neuartige Mutationsfunktion, die durch ihre Parameter eine differenzierte Selbstadaptation ermöglicht, ohne den rechentechnischen Aufwand wesentlich zu erhöhen. Die Parameter der Mutationsfunktion werden durch die asymmetrische Evolutionsstrategie automatisch adaptiert, es existieren keine externen Steuerparameter, die vom Benutzer vorgegeben werden müssen.

Grundsätzlich wird jede Arbeit über die Entwicklung neuer Optimierverfahren durch ein Theorem von Wolpert und Macready überschattet [Wol97]. In ihrem *No Free Lunch*-Theorem (NFL) zeigen sie, dass jeder Algorithmus, der für bestimmte Klassen von Problemen ein besseres Optimierverhalten aufweist als die Zufallssuche für alle anderen Problemklassen ein schlechteres Verhalten als die Zufallssuche aufweist. Sie verweisen daher auf die geringe Aussagekraft, die die Angabe von Konvergenzeigenschaften für eine ausgewählte Menge von Beispielpunkten mit sich bringt. Glücklicherweise (zumindest aus Sicht der Anhänger natur-analoger Verfahren) handelt es sich bei den Untersuchungen zum *No Free Lunch*-Theorem nicht um spezielle Untersuchungen, die ausschließlich für evolutionäre Algorithmen gelten, sondern um die Untersuchung von Eigenschaften, die in dieser Form für alle Optimierverfahren gelten. Da im Vorfeld bekannt ist, dass jedes Optimierverfahren nur für bestimmte Problemklassen gute Optimiereigenschaften aufweist, ist in dieser Arbeit der Versuch unternommen worden, die untersuchten Beispiele möglichst breit zu streuen.

Diese Arbeit untergliedert sich in die folgenden Kapitel:

1. Einführung

2. Symmetrische Evolutionsstrategien

Dieses Kapitel stellt die klassischen Varianten der Evolutionsstrategien, die von Rechenberg und Schwefel entwickelt wurden, vor. In diesem Kapitel werden die grundlegenden Begriffe wie Individuum, Population und Generation sowie die evolutionären Operationen Mutation, Rekombination und Selektion eingeführt. Die Selbstadaptationsmechanismen werden erläutert sowie das Konvergenzverhalten dieser Varianten.

3. Erweiterungen der Evolutionsstrategien

In diesem Kapitel werden Erweiterungen der klassischen Varianten beschrieben. Diese Varianten stammen aus unterschiedlichen Entwicklungsrichtungen. Ein Schwerpunkt dieses Kapitels sind die Arbeiten von Ostermeier und Hansen, die mit ihrer entstochastisierten Adaptation ein grundlegend anderes Modell der Selbstadaptation entwickelt haben. Die Auswahl der beschriebenen Arbeiten soll das breite Spektrum der Weiterentwicklungen, gleichzeitig jedoch auch die Problematik der Steuerparameter aufzeigen.

4. Asymmetrische Verteilungsfunktionen

Die in dieser Arbeit vorgestellte Variante der Evolutionsstrategien benötigt eine spezielle Wahrscheinlichkeitsverteilungsfunktion zur Realisierung der asymmetrischen Mutationsfunktion. Dieses Kapitel beschreibt die Entwicklung einer solchen Verteilungsfunktion im Detail. Da keine bekannte Verteilungsfunktion die geforderten Eigenschaften aufweist, kann hier nicht auf Literatur zurückgegriffen werden. Es werden in diesem Kapitel insgesamt drei Verteilungsfunktionen entwickelt die aufeinander aufbauen und in unterschiedlichen Eigenschaften der asymmetrischen Evolutionsstrategie resultieren. Zusätzlich zu den stochastischen Eigenschaften der entwickelten Verteilungsfunktionen werden Eigenschaften aus dem Bereich der Analysis bestimmt, die ein vollständiges Bild der mathematischen Eigenschaften der asymmetrischen Verteilungsfunktionen liefern.

5. Implementierung

Die Implementierung einer asymmetrischen Evolutionsstrategie ist grundsätzlich nicht verschieden von der Implementierung anderer Evolutionsstrategien. Die Realisierung eines effizienten Zufallszahlengenerators, der Zufallszahlen erzeugt, die entsprechend der asymmetrischen Verteilungsfunktion des vorhergehenden Kapitels verteilt sind, gehört jedoch nicht zum Standardrepertoire der evolutionären Algorithmen. Daher beschäftigt sich dieses Kapitel mit den Implementierungsdetails eines Zufallszahlengenerators, der zur Realisierung einer asymmetrischen Evolutionsstrategie notwendig ist.

6. Asymmetrische Evolutionsstrategien

An dieser Stelle sind alle theoretischen Vorarbeiten und Implementierungsdetails bearbeitet, sodass die asymmetrische Evolutionsstrategie beschrieben werden kann. Dieses Kapitel beschreibt auf Grundlage der im zweiten Kapitel eingeführten Notation alle Rechenschritte der asymmetrischen Evolutionsstrategien sowie den Selbstadaptationsmechanismus. In diesem Kapitel werden zusätzlich grundlegende Eigenschaften des Konvergenzverhaltens sowie spezielle Problematiken bei der Übertragung existierender Konvergenzbeweise dargestellt.

7. Experimentelle Resultate

Die Konvergenzeigenschaften der asymmetrischen Evolutionsstrategie werden in diesem Kapitel am Beispiel einer Vielzahl Testfunktionen sowie am Beispiel von zwei realen Anwendungen dargestellt. Die Auswahl der Testfunktionen orientiert sich an Arbeiten von Bäck, Ostermeier und Hansen. Die Resultate werden mit den Resultaten anderer Varianten der Evolutionsstrategien in Relation gestellt. Die realen Anwendungen stammen aus der Praxis und beziehen sich auf Problemstellungen, die aus dem Bereich der Ingenieurwissenschaften kommen.

8. Zusammenfassung & Ausblick

In diesem abschließenden Kapitel werden die Eigenschaften der asymmetrischen Evolutionsstrategie zusammengefasst und bewertet. Zusätzlich werden in Form eines Ausblickes weitere Forschungsmöglichkeiten über die asymmetrischen Evolutionsstrategie dargestellt, deren Bearbeitung das Thema dieser Arbeit sprengen würde.

Zusätzlich zu den in Kapiteln strukturierten Inhalten dieser Arbeit erleichtert ein Verzeichnis der Formelsymbole, ein Abbildungs-, Abkürzungs- und Algorithmenverzeichnis sowie ein umfangreicher Index dem Leser das Finden von Informationen in dieser Arbeit.

Kapitel 2 Symmetrische Evolutionstrategien

*Der Brei wird kalt, wenn man sich erst einen Löffel machen soll, nein,
überhaupt erst lernen soll, wie Löffel gemacht werden.
(Rainer Maria Rilke)*

In den 60er Jahren wurden unabhängig voneinander mehrere Optimierverfahren entwickelt, die die Prinzipien der Evolution als Grundlage für technische Optimierungsaufgaben einsetzten. Die bekanntesten Verfahren sind die Genetischen Algorithmen, die von Holland entwickelt und von DeJong verfeinert wurden sowie die Evolutionsstrategien, die von Rechenberg entwickelt und u. a. von Schwefel erweitert wurden [Hol75, DeJ75, Rec73, Sch75]. Diese Verfahren beruhen auf der Tatsache, dass die Evolution ein nahezu perfektes Verfahren darstellt, mit dem sich Individuen ihrer Umwelt anpassen. Die Mechanismen der Evolution sind eine ständige, jedoch geringe Veränderung der Individuen (Mutation), ein Austausch von Erbinformationen zwischen einzelnen Individuen (Rekombination) und eine Selektion, die es nur den optimal an die Umwelt angepassten Individuen ermöglicht, zu überleben. Diese drei Prinzipien sind auf Problemstellungen technischer Disziplinen übertragen worden. Aktuelle Übersichten hierzu finden sich in [Bäc96, Mic00].

Genetische Algorithmen und Evolutionsstrategien unterscheiden sich hauptsächlich in der Repräsentation der Individuen. Holland arbeitete bei der Entwicklung der Genetischen Algorithmen mit diskreten Problemstellungen und hat als Kodierung der Individuen eine Folge von Bits gewählt. Die Einsatzgebiete der Evolutionsstrategien waren reellwertig, sodass Evolutionsstrategien Vektoren reeller Zahlen als Kodierung verwenden. Die Unterschiede in den Kodierungen gelten nur für die Ursprungsverfahren. Inzwischen existieren Mischformen, wie reellwertige Genetische Algorithmen [Gar99, Her98] und diskrete bzw. gemischt-ganzzahlige Evolutionsstrategien [Bäc95, Rud94].

Zusammen mit weiteren Verfahren, die Ideen der Evolution nachahmen, z. B. dem Evolutionären Programmieren [Fog99] und dem Genetischen Programmieren [Ban98], bilden die Genetischen Algorithmen und die Evolutionsstrategien die Methodenklasse der evolutionären Algorithmen. Die evolutionären Algorithmen sind ein Teil einer umfassenderen Klasse von Algorithmen, der naturanalogen Algorithmen. Unter diesem Begriff werden Verfahren zusammengefasst, die Mechanismen der Natur aufgreifen und zur Lösung von Problemen einsetzen. Beispiele für solche Verfahren sind Ameisensysteme, die das Gruppenverhalten von Ameisenkolonien nachbilden [Leg99], das Simulated Annealing, bei dem Abkühlprozesse in Metallen nachgebildet werden [Kir83] oder künstliche neuronale Netze, die die Neuronenstruktur von Gehirnen nachbilden [Hec88]. Natur-analoge Verfahren werden erfolgreich zur Optimierung, Klassifizierung oder Vorhersage von Ereignissen eingesetzt [Mic00].

In diesem Kapitel werden die Grundlagen der von Rechenberg und Schwefel entwickelten Evolutionsstrategien erläutert. Da diese Verfahren auf symmetrischen Mutationsoperatoren aufbauen, werden sie als symmetrische Evolutionsstrategien bzw. als Evolutionsstrategien (ohne weiteren Zusatz) bezeichnet. Die benötigten stochastischen Grundlagen werden an dieser Stelle nur kurz erläutert. Die für die Entwicklung einer asymmetrischen Evolutionsstrategie notwendigen Grundlagen werden ausführlich im Kapitel 4 (Asymmetrische Verteilungsfunktionen, Seite 63) behandelt. Vollständigere Abhandlungen über den hier verwendeten Teilbereich der Stochastik finden sich z. B. in [Bey88, Bro90].

Diese Arbeit verzichtet auf die Darstellung der Grundlagen und Mechanismen der biologischen Evolution. Stattdessen sei für diesen Themenkomplex auf das Kapitel 2 (Natürliche Evolution) der Dissertation von Kursawe hingewiesen, die eine ausführliche Übersicht über die Evolution gibt [Kur99]. Zusätzlich zu dieser Literatur existiert eine Vielzahl weiterer Arbeiten, in denen der Zusammenhang zwischen biologischer Evolution und den evolutionären Algorithmen, zum Teil auch kritisch, dargestellt wird [Bae74, Daw76, Daw86].

2.1 Grundlagen der Evolutionsstrategien

Die biologische Evolution hat die Entwicklung der Evolutionsstrategien in der Hinsicht inspiriert, dass die Idee der kleinen Veränderung mit darauffolgender Selektion Einzug in die mathematisch orientierten Optimierverfahren gehalten hat. Die Begriffsbildung der Evolution ist hierbei übernommen worden. Dies drückt einerseits die Nähe zum Vorbild aus, andererseits besteht jedoch die Gefahr, dass Begriffe, die in anderen Zweigen der Wissenschaft etabliert sind, hier in veränderter Form verwendet werden.

Im Bereich der Evolutionsstrategien werden einzelne, potentielle Lösungen eines Optimierproblems als Individuum bezeichnet. Eine Menge von Individuen, die sich im gleichen Optimierschritt befinden, bildet eine Population. Zur Unterscheidung einzelner Populationen werden diese als eine bestimmte Generation bezeichnet. Bei dem Übergang einer Generation in die nächste finden evolutionäre Operationen, wie z. B. Rekombination, Mutation und Selektion statt. In dieser Arbeit wird für alle Betrachtungen das Auffinden eines Minimums als Optimierziel unterstellt. Sämtliche Aussagen gelten selbstverständlich auch für das Auffinden eines Maximums.

2.1-1 Stochastische Grundlagen

Für die weitere Betrachtung der Evolutionsstrategien müssen die informell eingeführten Begriffe mathematisch exakt definiert werden. Da es sich bei den Evolutionsstrategien um ein probabilistisches Optimierverfahren handelt, werden zunächst einige Begriffe aus der Wahrscheinlichkeitstheorie eingeführt. Die Auswahl der Definitionen beschränkt sich hier auf das Notwendige. Eine ausführliche Abhandlung über den Themenkomplex Stochastik findet sich zum Beispiel in [Bey88, Bro90].

Definition 2.1: Zufallsvariable

Eine Variable, die in Abhängigkeit vom Zufall verschiedene Werte annimmt, heißt Zufallsvariable. Als Verteilungsfunktion $\Phi_X(x)$ der Zufallsgröße X bezeichnet man die Funktion

$$\Phi_X(x) = P(X \leq x). \quad (2.1)$$

Die Verteilungsfunktion $\Phi_X(x)$ gibt die Wahrscheinlichkeit P an, dass die Realisierung einer Zufallsvariable X einen Wert kleiner oder gleich x annimmt.

Für die allgemeine Praxis sind zwei Typen von Zufallsvariablen von Bedeutung, diskrete und stetige Zufallsvariablen. Für die Anwendung im Bereich der Evolutionsstrategien sind jedoch die stetigen Zufallsvariablen von größerer Bedeutung.

Definition 2.2: Stetige Zufallsvariable

Eine Zufallsvariable heißt stetig, wenn sich ihre Verteilungsfunktion durch

$$\Phi_X(x) = \int_{-\infty}^x \varphi_X(t) dt \quad (2.2)$$

darstellen lässt. Die Funktion $\varphi_X(x)$ heißt Dichtefunktion der Verteilung.

Von besonderer Bedeutung für die Evolutionsstrategien ist eine spezielle, stetige Verteilung: die Normalverteilung:

Definition 2.3: Normalverteilte Zufallsvariable

Eine stetige Zufallsvariable X heißt normalverteilt, wenn ihre Dichtefunktion sich folgendermaßen beschreiben lässt:

$$\varphi_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} \quad (2.3)$$

Ist X normalverteilt mit den Parametern a und σ , so spricht man von einer $N(a, \sigma)$ verteilten Zufallsvariablen.

Definition 2.4: Normierte und zentrale Normalverteilung

Eine Normalverteilung mit den Parametern $a = 0$ und $\sigma = 1$ wird als normierte und zentrale Normalverteilung $\varphi_X(x)$ bezeichnet. Ihre Dichtefunktion reduziert sich zu:

$$\varphi_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \quad (2.4)$$

Abbildung 2.1 gibt eine graphische Darstellung für einige Belegungen des Parameters σ . Der Parameter a ist in diesen Beispielen konstant 0.

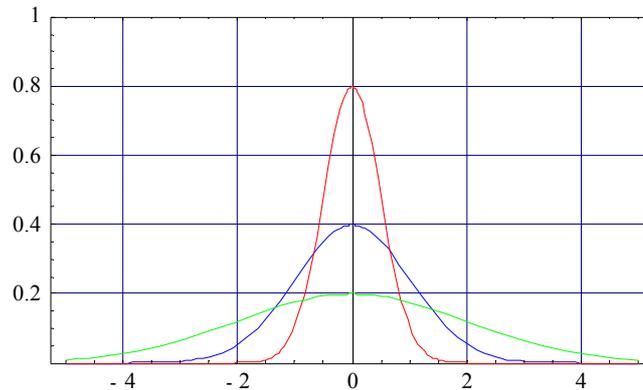


Abb. 2.1: Normalverteilung für unterschiedliche σ

Charakteristische Maße einer Zufallsvariable sind Erwartungswert, Varianz und Standardabweichung. Im Rahmen des Kapitels 4 (Asymmetrische Verteilungsfunktionen, Seite 63) werden diese Maße ausführlich behandelt. An dieser Stelle sind daher nur die Definitionen angeführt.

Definition 2.5: Erwartungswert

Der Erwartungswert E_X einer stetigen Zufallsvariablen X berechnet sich durch

$$E_X = \int_{-\infty}^{\infty} x \varphi(x) dx \quad (2.5)$$

Definition 2.6: Varianz

Die Varianz V_X einer stetigen Zufallsvariablen X berechnet sich durch

$$V_X = \int_{-\infty}^{\infty} (x - E_X)^2 \varphi(x) dx \quad (2.6)$$

Definition 2.7: Standardabweichung

Die Standardabweichung SD_X einer stetigen Zufallsvariablen X berechnet sich durch

$$SD_X = \sqrt{V_X} \quad (2.7)$$

Für den Fall einer nach $N(a, \sigma)$ normalverteilten Zufallsvariable X ergeben sich für diese drei charakteristischen Werte folgende Vereinfachungen:

$$E_X = a \quad (2.8)$$

$$V_X = \sigma^2 \quad (2.9)$$

$$SD_X = \sigma \quad (2.10)$$

Die Definitionen 2.1 bis 2.7 lassen sich auf Tupel von Zufallsvariablen, sogenannte Zufallsvektoren übertragen.

Definition 2.8: Normalverteilter Zufallsvektor

Ein n -dim. Zufallsvektor $\mathbf{X} = (x_1, \dots, x_n)$ heißt normalverteilt, wenn für seine Dichtefunktion gilt:

$$\varphi_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det(\mathbf{C})}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{a})^T \mathbf{C}^{-1}(\mathbf{x}-\mathbf{a})} \quad (2.11)$$

Die Matrix $\mathbf{C} = (c_{ij})$, mit $i, j \in \{1, \dots, n\}$ wird als Kovarianzmatrix bezeichnet und es gilt:

$$c_{ij}(x) = \text{cov}(x_i, x_j) = E_{X_{ij}} - E_{X_i} E_{X_j}, \text{ mit} \quad (2.12)$$

$$E_{X_{ij}} = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} x_i x_j \varphi_{\mathbf{X}}(x_1, \dots, x_n) dx_1 \dots dx_n \quad (2.13)$$

Die Diagonalkomponenten $c_{ii} = \sigma_i^2$ stellen die Varianzen der Verteilung, die Komponenten c_{ij} mit $i \neq j$ die Kovarianzen dar.

2.1-2 Repräsentation

Nachdem die Grundlagen der normalverteilten Zufallsvariablen und -vektoren eingeführt sind, sollen diese im folgenden auf die Evolutionsstrategien übertragen werden. Von Bedeutung für alle weiteren Betrachtungen ist die Repräsentation bzw. die Kodierung eines Individuums. Gegeben sei ein n -dimensionales Optimierproblem im \mathbb{R}^n . Ein Individuum \mathbf{I} wird als Punkt eines $(n + s + w)$ -dimensionalen Raumes beschrieben, mit

$$\mathbf{I} = (\mathbf{x}, \boldsymbol{\sigma}, \boldsymbol{\alpha})^T \in \mathbb{R}^n \times \mathbb{R}^s \times [-\pi, \pi]^w \quad (2.14)$$

für $s \in \{1, \dots, n\} \subset \mathbb{N}$ und $w \in \{0, \dots, (n(n-1))/2\} \subset \mathbb{N}$. Hierbei ist $\mathbf{x} = (x_1, \dots, x_n)^T \in \mathbb{R}^n$ die Objektkomponente, $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_s)^T \in \mathbb{R}^s$ und $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_w)^T \in [-\pi, \pi]^w$ bezeichnen die Strategiekomponente. Beide werden in der Literatur auch häufig als Objektvariable bzw. Strategievariable bezeichnet.

Die Objektkomponente \mathbf{x} gibt die Position des Individuums \mathbf{I} im Suchraum an. Zusätzlich zur Position eines Individuums existieren weitere Informationen zu jedem Individuum \mathbf{I} , die Strategiekomponenten $\boldsymbol{\sigma}$ und $\boldsymbol{\alpha}$. Die Strategiekomponente $\boldsymbol{\sigma}$ gibt die Schrittweite eines Individuums in jede Koordinatenrichtung des Suchraumes an, mit der Strategiekomponente $\boldsymbol{\alpha}$ können die Lagewinkel der einzelnen Schrittweiten zueinander eingestellt werden. Die Anzahl der Elemente in $\boldsymbol{\sigma}$ und $\boldsymbol{\alpha}$ sind durch die oberen Grenzen s und w vorgegeben. Konkrete Realisierungen bestimmter Varianten der Evolutionsstrategien können eine geringere Anzahl an Elementen in den Strategiekomponenten verwenden. Als Beispiel sei hier die (1+1)-ES zu nennen, die lediglich den Schrittweitenvektor $\boldsymbol{\sigma}$, jedoch keine Lagewinkel $\boldsymbol{\alpha}$ verwendet. Die Bedeutung der Strategiekomponenten wird nach den grundlegenden Definitionen an einem Beispiel erläutert.

2.1-3 Fitness

Einem Individuum \mathbf{I} muss ein Fitnesswert zugeordnet werden, der seine Güte bezüglich des Optimierproblems angibt und damit Einfluss auf seine Überlebensfähigkeit nimmt.

Definition 2.9: Fitness

Jedem Individuum \mathbf{I} wird durch die Fitnessfunktion $\Psi: \mathbb{R}^n \rightarrow \mathbb{R}$ ein Fitnesswert

$$\Psi(\mathbf{I}) = \Psi(\mathbf{x}) \in \mathbb{R} \quad (2.15)$$

zugeordnet.

Die Funktion Ψ wird nur durch die Objektkomponente \mathbf{x} eines Individuums bestimmt. Der Einfluss von σ und α auf die Fitness eines Individuums kann nur indirekt über die Güte der Objektvariablen \mathbf{x} bestimmt werden. Die Fitness eines Individuums wird nicht nur durch den Funktionswert, sondern zusätzlich durch weitere Eigenschaften wie Erfüllung aller Nebenbedingungen bestimmt. Wird im folgenden der Fitnesswert Ψ verwendet, sind solche zusätzlichen Bedingungen bereits berücksichtigt.

2.1-4 Mutation

Der Mutationsoperator erzeugt aus einem Individuum \mathbf{I} ein Nachfolgeindividuum \mathbf{I}' .

$$\mathbf{I}' = mut(\mathbf{I}) \quad (2.16)$$

Die Informationen über die Art der Veränderungen können durch drei verschiedene Mechanismen definiert werden:

1. Der Benutzer gibt eine feste Variationsbreite vor, die im Laufe der Optimierung nicht verändert wird. Ein Beispiel hierfür ist die im folgenden beschriebene (1+1)-Evolutionstrategie.
2. Es existiert ein deterministisches Schema, das in Abhängigkeit von dem Verlauf der Optimierung die Variationsbreite beeinflusst. Ein Beispiel für diese Art der Veränderung ist die 1/5 Erfolgsregel von Rechenberg bzw. die Präzisierung dieser Regel durch Schwefel, die in [Sch94, Seite 112] beschrieben wird.
3. Die Variationsbreite unterliegt einer stochastischen Veränderung und wird durch die Fitnessfunktion bewertet. Beispiele für diese Art der Veränderung sind die Selbstadaptationsvarianten der Evolutionstrategien sowie die Adaptation der in dieser Arbeit vorgestellten asymmetrischen Evolutionstrategie.

Für alle drei Varianten werden in dem folgenden Abschnitt Beispiele diskutiert, in denen der hier informell eingeführte Mutationsoperator formalisiert wird.

2.1-5 Rekombination

Durch die Mutation werden Nachkommen erzeugt, deren Position von der Position des Elterindi-

duums abweicht. Ein Maß für diese Positionsveränderung ist die Standardabweichung der zugehörigen Verteilungsfunktion. In Analogie zu anderen Abhandlungen wird auch hier die Standardabweichung als Schrittweite bezeichnet. Am Anfang einer Optimierung ist es notwendig, dass der Suchraum möglichst breit gestreut mit Individuen belegt wird, um mögliche (lokale) Optima schnell zu entdecken. Die Mutationsoperation ist dazu nicht in idealer Form in der Lage, da die begrenzte Standardabweichung die Wahrscheinlichkeit für große Schrittweiten senkt und eine Mutation der Schrittweiten hin zu großen Werten relativ lange dauern kann.

Die Mutation verwendet zur Erzeugung eines Nachkommen ausschließlich die Informationen eines einzelnen Individuums. Diese Art der Veränderung ist die einzige Möglichkeit, bei einer Populationsgröße von 1 Nachkommen zu erzeugen. Für Populationsgrößen größer 1 können die Informationen mehrerer Individuen durch den Rekombinationsoperator zu einem Nachfolger kombiniert werden. Für die Erzeugung von λ Nachkommen stehen somit maximal alle μ Individuen der Elterngeneration zur Verfügung. Die Rekombination ist im Zuge der Entwicklung dieser mehrgliedrigen Evolutionsstrategien eingeführt worden. Die Rekombination für Objekt- und Strategiekomponente wird getrennt durchgeführt:

$$\begin{aligned} \mathbf{I}' &= rek(P^{(t)}) \\ &= rek(\mathbf{I}_1, \dots, \mathbf{I}_\mu) \\ &= (rek(\mathbf{x}_1, \dots, \mathbf{x}_\mu), rek(\boldsymbol{\sigma}_1, \dots, \boldsymbol{\sigma}_\mu), rek(\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_\mu)) \end{aligned} \quad (2.17)$$

Bevor die einzelnen Rekombinationsoperationen vorgestellt werden, wird eine Auswahlfunktion zur Indizierung der Individuen eingeführt.

Definition 2.10: Auswahlfunktion

Die Zufallsfunktion G_n liefert zu einem gegebenen $n > 0$ einen zufälligen Wert aus dem Intervall $[1, n] \subset \mathbb{N}$. Die Wahl dieses Wertes wird durch eine diskrete, gleichverteilte Verteilung über $[1, n]$ bestimmt.

Im allgemeinen werden vier Rekombinationsarten benutzt, für spezielle Anwendungsfälle sind jedoch eine Vielzahl weiterer Rekombinationsarten entwickelt worden [Rec94]:

1. Diskrete Rekombination

Zwei Elternindividuen \mathbf{I}_1 und \mathbf{I}_2 werden gleichverteilt aus der Menge aller Elternindividuen bestimmt. Ein Nachkomme \mathbf{I}' wird erzeugt, indem für alle Elemente aus \mathbf{I}' gleichverteilt das entsprechende Element aus \mathbf{I}_1 oder \mathbf{I}_2 gewählt wird¹:

$$\mathbf{I}' = rek(\mathbf{I}_1, \mathbf{I}_2) = rek \begin{pmatrix} \mathbf{I}_{1,1}, \mathbf{I}_{2,1} \\ \mathbf{I}_{1,2}, \mathbf{I}_{2,2} \\ \dots, \dots \\ \mathbf{I}_{1,n}, \mathbf{I}_{2,n} \end{pmatrix} = \begin{pmatrix} \mathbf{I}_{G_2,1} \\ \mathbf{I}_{G_2,2} \\ \dots \\ \mathbf{I}_{G_2,n} \end{pmatrix} \quad (2.18)$$

mit $\mathbf{I}_1 = \mathbf{I}_{G_\mu}$ und $\mathbf{I}_2 = \mathbf{I}_{G_\mu}$ nach Definition 2.10.

1. Die Schreibweise $\mathbf{I}_{j,i}$ bezeichnet hier die i -te Komponente des j -ten Individuums.

2. Global diskrete Rekombination

Für jedes Element des Nachkommen \mathbf{I}' wird ein Elternindividuum \mathbf{I} gleichverteilt aus der Menge aller Elternindividuen bestimmt. Im Unterschied zur diskreten Rekombination ist die Menge der möglichen Eltern nicht auf zwei Eltern beschränkt:

$$\mathbf{I}' = \text{rek}(\mathbf{I}_1, \dots, \mathbf{I}_\mu) = \text{rek} \begin{pmatrix} \mathbf{I}_{1,1}, \dots, \mathbf{I}_{\mu,1} \\ \mathbf{I}_{1,2}, \dots, \mathbf{I}_{\mu,2} \\ \dots, \dots, \dots \\ \mathbf{I}_{1,n}, \dots, \mathbf{I}_{\mu,n} \end{pmatrix} = \begin{pmatrix} \mathbf{I}_{G_\mu,1} \\ \mathbf{I}_{G_\mu,2} \\ \dots \\ \mathbf{I}_{G_\mu,n} \end{pmatrix} \quad (2.19)$$

3. Intermediäre Rekombination

Zwei Elternindividuen \mathbf{I}_1 und \mathbf{I}_2 werden gleichverteilt aus der Menge aller Elternindividuen bestimmt. Ein Nachkomme \mathbf{I}' wird erzeugt, indem für alle Elemente aus \mathbf{I}' der Mittelwert von \mathbf{I}_1 und \mathbf{I}_2 gewählt wird:

$$\mathbf{I}' = \text{rek}(\mathbf{I}_1, \mathbf{I}_2) = \text{rek} \begin{pmatrix} \mathbf{I}_{1,1}, \mathbf{I}_{2,1} \\ \mathbf{I}_{1,2}, \mathbf{I}_{2,2} \\ \dots, \dots \\ \mathbf{I}_{1,n}, \mathbf{I}_{2,n} \end{pmatrix} = \begin{pmatrix} (\mathbf{I}_{1,1} + \mathbf{I}_{2,1})/2 \\ (\mathbf{I}_{1,2} + \mathbf{I}_{2,2})/2 \\ \dots \\ (\mathbf{I}_{1,n} + \mathbf{I}_{2,n})/2 \end{pmatrix} \quad (2.20)$$

mit $\mathbf{I}_1 = \mathbf{I}_{G_\mu}$ und $\mathbf{I}_2 = \mathbf{I}_{G_\mu}$.

4. Global intermediäre Rekombination

Für jedes Element des Nachkommen \mathbf{I}' werden zwei Elternindividuen \mathbf{I}_1 und \mathbf{I}_2 gleichverteilt aus der Menge aller Elternindividuen bestimmt. Das Element aus \mathbf{I}' ist dann der Mittelwert der entsprechenden Elemente aus \mathbf{I}_1 und \mathbf{I}_2 .

$$\mathbf{I}' = \text{rek}(\mathbf{I}_1, \dots, \mathbf{I}_\mu) = \text{rek} \begin{pmatrix} \mathbf{I}_{1,1}, \dots, \mathbf{I}_{\mu,1} \\ \mathbf{I}_{1,2}, \dots, \mathbf{I}_{\mu,2} \\ \dots, \dots, \dots \\ \mathbf{I}_{1,n}, \dots, \mathbf{I}_{\mu,n} \end{pmatrix} = \begin{pmatrix} (\mathbf{I}_{G_\mu,1} + \mathbf{I}_{G_\mu,1})/2 \\ (\mathbf{I}_{G_\mu,2} + \mathbf{I}_{G_\mu,2})/2 \\ \dots \\ (\mathbf{I}_{G_\mu,n} + \mathbf{I}_{G_\mu,n})/2 \end{pmatrix} \quad (2.21)$$

Durch die Wahl eines Parameters $\gamma \neq 1/2$ (siehe (2.22)) ist es möglich, intermediär rekombinierte Individuen zu einem Elternteil zu verlagern. Der Einfluss eines jeden Elternteils wird dann durch

$$\mathbf{I}' = \gamma \mathbf{I}_{1,i} + (1 - \gamma) \mathbf{I}_{2,i}, \text{ mit } \gamma \in [0,1] \subset \mathbb{R} \quad (2.22)$$

bestimmt.

2.1-6 Selektion

Der Selektionsoperator wählt aus den in der aktuellen Generation erzeugten Nachkommen $P^{(t)}$ die Individuen aus, die in die nächste Generation übernommen werden und dann die Elternindividuen $P^{(t+1)}$ der Generation $t+1$ bilden.

$$P^{(t+1)} = sel(P^{(t)}) \quad (2.23)$$

Der Selektionsoperator der Evolutionsstrategien ist in den grundlegenden Arbeiten von Rechenberg und Schwefel ein deterministischer Operator. Aus einer Population mit μ Elternindividuen werden die besten λ entsprechend ihres Fitnesswertes ausgewählt. Hier unterscheiden sich die Evolutionsstrategien von den Genetischen Algorithmen, die eine probabilistische Selektion verwenden, bei der die Wahrscheinlichkeit eines Individuums, selektiert zu werden, mit steigender Fitness steigt [Dav91, Jul99, Mic96]. Grundsätzlich wird bei Evolutionsstrategien zwischen Selektionsverfahren unterschieden, die ausschließlich die Nachkommen in die Selektion einbeziehen und Selektionsverfahren, die aus der gesamten Population, Eltern und Nachkommen gemeinsam, auswählen. Neuere Varianten erlauben zudem eine zeitlich begrenzte Lebensspanne der Individuen sowie zusätzliche Mechanismen wie lokale Selektion. Eine Auswahl neuerer Verfahren wird in Kapitel 3 (Erweiterungen der Evolutionsstrategien, Seite 35) diskutiert. Die unterschiedlichen Selektionsmechanismen führen zu einer ersten Notation der Evolutionsstrategien [Sch94]:

Definition 2.11: Plus/Komma-Strategie

Mit $(\mu+\lambda)$ wird eine Evolutionsstrategie mit μ Elternindividuen und λ Nachkommen bezeichnet. Werden bei der Selektion Elterngeneration und Nachkommen berücksichtigt, handelt es sich um eine $(\mu+\lambda)$ -Strategie, werden nur die Nachkommen berücksichtigt, handelt es sich um eine (μ,λ) -Strategie.

Im nächsten Abschnitt werden drei unterschiedliche Typen von Evolutionsstrategien erläutert.

2.2 (1+1)-Evolutionsstrategie

Die $(1+1)$ -Evolutionsstrategie wurde in den ersten Anwendungen, die von Rechenberg Mitte der 60'er Jahre durchgeführt wurden, verwendet. Die $(1+1)$ -Evolutionsstrategie besteht aus einer Population mit einem Elternindividuum, aus dem durch Addition einer normalverteilten Zufallsvariablen ein Nachkomme erzeugt wird. Die Nachfolgegeneration wird durch das Individuum mit der höheren Fitness gebildet.

2.2-1 Grundlegender Algorithmus

Nach Hoffmeister und Bäck kann eine $(1+1)$ -ES wie folgt beschrieben werden [Hof92]:

$$(1+1)\text{-ES} = (P^{(0)}, mut, sel, c_d, c_i, f, g, \kappa) \text{ , mit} \quad (2.24)$$

$P^{(0)}$	$= \mathbf{I}^{(0)} = (\mathbf{x}^{(0)}, \boldsymbol{\sigma}^{(0)}) \in I$, mit $I = \mathbb{R}^n \times \mathbb{R}_+^n$	Startpopulation
mut	$: I \rightarrow I$	Mutationsoperator
sel	$: I \times I \rightarrow I$	Selektionsoperator
c_d, c_i	$: c_d, c_i \in \mathbb{R}_+$	Schrittweitenkontrolle
f	$: \mathbb{R}^n \rightarrow \mathbb{R}$	Zielfunktion
g_j	$: \mathbb{R}^n \rightarrow \mathbb{R}$, mit $j \in \{1, \dots, q\}$	Nebenbedingungen
κ	$: I \rightarrow \{0, 1\}$	Abbruchkriterium

Die Parameter c_d und c_i bekommen erst mit Einführung der 1/5-Erfolgsregel, die im Abschnitt 2.2-3 (Die 1/5-Erfolgsregel, Seite 24) beschrieben wird, eine Bedeutung. Der folgende Algorithmus gibt den Ablauf einer solchen (1+1)-Evolutionstrategie an.

```

1   t := 0;
2   initialisiere  $P^{(t)} := (\mathbf{x}^{(t)}, \boldsymbol{\sigma}^{(t)})$ ;           //Initialisierung
3   solange  $\kappa(P^{(t)}) \neq 1$                                    //Abbruchkriterium
4        $\mathbf{I}^{(t)} = mut(\mathbf{I}^{(t)})$ ;                               //Mutation
5       berechne  $\Psi(\mathbf{I}^{(t)})$ ;
6       wenn  $\Psi(\mathbf{I}^{(t)}) \leq \Psi(\mathbf{I}^{(t)})$                  //Selektion
7           dann  $P^{(t+1)} := \mathbf{I}^{(t)}$ ;
8           sonst  $P^{(t+1)} := \mathbf{I}^{(t)}$ ;
9       t := t+1;
10  ende

```

Alg. 2-1: Ablauf einer (1+1)-Evolutionstrategie

Das Individuum $\mathbf{I}^{(t)}$ besteht aus zwei Komponenten $\mathbf{x}^{(t)}$ und $\boldsymbol{\sigma}^{(t)}$, die wiederum n -elementige Vektoren darstellen. Die Objektkomponente $\mathbf{x}^{(t)}$ enthält die Koordinaten des Individuums im Suchraum, die Strategiekomponente $\boldsymbol{\sigma}^{(t)}$ enthält die Schrittweiten, die in Verbindung mit dem Mutationsoperator die Position für den Nachkommen bestimmen. Der Nachkomme $\mathbf{I}^{(t)}$ wird durch Anwendung der Mutation bestimmt:

$$\mathbf{I}^{(t)} = mut(\mathbf{I}^{(t)}) \quad (2.25)$$

Der Mutationsoperator $mut(\cdot)$ berechnet für jede Koordinate $x_i^{(t)}$ aus $\mathbf{x}^{(t)}$ eine neue Koordinate $x_i'^{(t)}$ mittels der Vorschrift

$$x_i'^{(t)} = x_i^{(t)} + N(0, \sigma_i^{(t)}) \quad (2.26)$$

Die Wahl einer $N(0, \sigma)$ -verteilten Zufallsvariable liegt in der Tatsache begründet, dass kleine Änderungen häufiger auftreten sollen als große. Dieses Verhalten ist an die Evolution angelehnt, bei der zu beobachten ist, dass Nachkommen ihren Eltern ähnlich sind.

Der Selektionsoperator bewertet Elter und Nachkomme und selektiert das Individuum mit der besten Fitness unter Berücksichtigung der Nebenbedingungen g :

$$P^{(t+1)} = sel(\mathbf{I}'^{(t)}, \mathbf{I}^{(t)}) = \begin{cases} \mathbf{I}'^{(t)}, & \text{falls } \Psi(\mathbf{x}'^{(t)}) \leq \Psi(\mathbf{x}^{(t)}) \wedge g_j \geq 0 \quad \forall j \in \{1, \dots, q\} \\ \mathbf{I}^{(t)}, & \text{sonst} \end{cases} \quad (2.27)$$

Dieser Iterationsprozess von Generation t zu Generation $t+1$ stoppt, wenn das Abbruchkriterium $\kappa(P^{(t)}) = 1$ gilt. In der oben dargestellten Variante der (1+1)-ES wird der Schrittweitenvektor $\boldsymbol{\sigma}^{(t)}$ nicht verändert, d. h. es findet keine Selbstadaptation statt. Für diese Variante und reguläre Optimierprobleme ist in [Bor78] die globale Konvergenz bewiesen worden. Für die Definition der Regularität eines Optimierproblems siehe [Hof92].

2.2-2 Konvergenzverhalten der (1+1)-Evolutionstrategie

Am Beispiel der (1+1)-ES wird hier erläutert, wie die Ansätze von Rechenberg und Schwefel lauten, mit denen Fortschrittsgeschwindigkeit und optimale Schrittweite berechnet werden können. Im Abschnitt 6.2-1 (Klassische Beweisstrategien, Seite 173) wird erläutert, wie diese Beweistechniken auf asymmetrische Evolutionsstrategien übertragen werden können. Zusätzlich zu diesen Resultaten existieren Konvergenzsätze für boolesche Funktion [Dro98].

Für die Dichte der Wahrscheinlichkeit eines bestimmten Zufallsvektors \mathbf{z} gilt hier:

$$\begin{aligned} p(\mathbf{z}) &= \prod_{i=1}^n p(z_i) \\ &= \frac{1}{\sqrt{(2\pi)^n \prod_{i=1}^n \sigma_i}} e^{-\frac{1}{2} \sum_{i=1}^n \left(\frac{z_i}{\sigma_i}\right)^2} \end{aligned} \quad (2.28)$$

Unter der Vereinfachung, dass alle Elemente in $\boldsymbol{\sigma}$ identisch sind, d. h. $\sigma = \sigma_i$ für $1 \leq i \leq n$ lässt sich (2.28) vereinfachen:

$$p(\mathbf{z}) = \left(\frac{1}{\sqrt{2\pi\sigma}}\right)^n e^{-\frac{1}{2} \frac{\mathbf{z}\mathbf{z}^T}{\sigma^2}} \quad (2.29)$$

Die Länge $\|\mathbf{z}\|$ des Zufallsvektors \mathbf{z} wird durch

$$\|\mathbf{z}\| = \sqrt{\sum_{i=1}^n z_i^2} \quad (2.30)$$

bestimmt. Die Summe der Quadrate von n unabhängigen $N(0, 1)$ -verteilten Zufallszahlen x_i ist χ^2 -verteilt, mit n Freiheitsgraden [Eva93]:

$$\sum_{i=1}^n x_i^2 \sim \chi^2(n) \quad (2.31)$$

Angewandt auf (2.30) unter Berücksichtigung der Schrittweite σ folgt daraus für $\|\mathbf{z}\|$:

$$\|\mathbf{z}\| \sim \sigma \sqrt{\chi^2(n)} \quad (2.32)$$

Die χ^2 -Verteilung nähert sich für große n einer $N(\sigma\sqrt{n-1/2}, \sigma^2/2)$ -verteilten Normalverteilung [Sch94]. Daraus folgt, dass die Länge des Zufallsvektors, bei konstantem σ mit der reziproken Quadratwurzel der Dimension n steigt.

Für das Korridor- und das Kugelmodell hat Rechenberg den Erwartungswert für die Fortschrittsgeschwindigkeit ϕ sowie die optimale Schrittweite σ_{opt} berechnet. Die vollständige Herleitung dieser Gleichungen findet sich in [Rec70]. An dieser Stelle sind nur die Endergebnisse dargestellt. Es gilt für das Korridormodell mit Breite b des Korridors:

$$f_{Korridor}(\mathbf{x}) = c_0 + c_1 x_i, \text{ mit} \quad (2.33)$$

$$-b/2 \leq x_i \leq b/2 \text{ für } 2 \leq i \leq n \quad (2.34)$$

$$\Phi_{Korridor} = \frac{\sigma}{\sqrt{2\pi}} \left(1 - \frac{1}{\sqrt{2\pi}} \frac{\sigma}{b}\right)^{n-1} \text{ und} \quad (2.35)$$

$$\sigma_{opt} = b \frac{\sqrt{2\pi}}{n} \quad (2.36)$$

Für das Kugelmodell

$$f_{Kugel}(\mathbf{x}) = \sum_{i=1}^n x_i^2, \text{ für } 1 \leq i \leq n \quad (2.37)$$

gilt:

$$\Phi_{Kugel} = \frac{\sigma}{\sqrt{2\pi}} \left(e^{-\left(\frac{n\sigma}{\sqrt{8}r}\right)^2} - \sqrt{\pi} \frac{n\sigma}{\sqrt{8}r} \left(1 - \operatorname{erf}\left(\frac{n\sigma}{\sqrt{8}r}\right)\right) \right) \text{ und} \quad (2.38)$$

$$\sigma_{opt} = 1.224 \frac{r}{n} \quad (2.39)$$

2.2-3 Die 1/5-Erfolgsregel

Aus den Resultaten der theoretischen Betrachtungen über Konvergenzgeschwindigkeit ist von Rechenberg die sog. 1/5-Erfolgsregel entwickelt worden. Sie besagt, dass das Verhältnis ζ von erfolgreichen Mutationen zu allen Mutationen 1/5 betragen sollte. Falls es größer als 1/5 ist, muss die Schrittweite erhöht werden, falls es kleiner als 1/5 ist, muss die Schrittweite verringert werden. Rechenberg setzt für seine 1/5-Erfolgsregel die Schrittweiten der einzelnen Koordinaten gleich, d. h. für jede Generation existiert nur eine Schrittweite $\sigma^{(t)}$, die für alle n Dimensionen identisch ist. Formal beschrieben lautet diese Regel:

$$\sigma^{(t+1)} = \begin{cases} c_d \sigma^{(t)}, & \text{falls } \zeta < 1/5 \\ c_i \sigma^{(t)}, & \text{falls } \zeta > 1/5 \\ \sigma^{(t)}, & \text{sonst} \end{cases} \quad (2.40)$$

Für die Parameter c_d und c_i gibt Rechenberg folgende Empfehlung:

$$c_d = 0.82 \text{ und } c_i = 1/0.82 \quad (2.41)$$

In regelmäßigen Abständen wird das Überlebensverhältnis von Elternindividuum und Nachkomme überprüft (2.40). Sollte die Fitness der Nachkommen zu hoch werden $\zeta > 1/5$, wird die Schrittweite vergrößert. Hierbei wird angenommen, dass eine große Distanz zum Optimum Ursache für die hohe Anzahl der erfolgreichen Mutationen ist. Sollte dagegen ein Elternindividuum zu lange überleben ($\zeta < 1/5$), wird die Schrittweite erniedrigt, um so den Suchbereich einzugrenzen. Von Nachteil ist bei dieser Regel, dass bei Optimierproblemen, die aufgrund ihrer Topologie keine günstige Erfolgsrate ermöglichen, die Schrittweite mit der deterministischen Schrittweitenanpassung ständig verringert wird und die Evolutionsstrategie in einem lokalen Optimum stagnieren kann.

2.3 $(\mu+1)$ -Evolutionsstrategie

Da die (1+1)-Evolutionsstrategie nur ein Elternindividuum in jeder Generation besitzt, können die Ideen der Rekombination aus der biologischen Evolution nicht angewendet werden. Um Rekombination anwendbar zu machen, ist von Rechenberg die $(\mu+1)$ -Evolutionsstrategie mit $\mu > 1$ entwickelt worden. Diese Variante besitzt μ Elternindividuen und erzeugt aus diesen einen Nachkommen. Mittels der +-Selektion wird ein Elternindividuum (in der Regel das Individuum mit schlechtestem Fitnesswert Ψ) durch den Nachkommen ersetzt, wenn mindestens eines der Elternindividuen einen schlechteren Fitnesswert besitzt als der Nachkomme. Die formale Notation einer $(\mu + 1)$ -ES lautet [Sys92]:

$$(\mu + 1)\text{-ES} = (P^{(0)}, \mu, rek, mut, sel, c_d, c_i, f, g, \kappa), \text{ mit} \quad (2.42)$$

$P^{(0)}$	$= (\mathbf{I}_1^{(t)}, \dots, \mathbf{I}_\mu^{(t)}) \in I^\mu$, mit $I = \mathbb{R}^n \times \mathbb{R}_+^n$	Startpopulation
μ	$: \mu > 1$	Anzahl der Eltern
rek	$: I^\mu \rightarrow I$	Rekombinationsoperator
mut	$: I \rightarrow I$	Mutationsoperator
sel	$: I^{\mu+1} \rightarrow I^\mu$	Selektionsoperator
c_d, c_i	$: c_d, c_i \in \mathbb{R}_+$	Schrittweitenkontrolle
f	$: \mathbb{R}^n \rightarrow \mathbb{R}$	Zielfunktion
g_j	$: \mathbb{R}^n \rightarrow \mathbb{R}$, mit $j \in \{1, \dots, q\}$	Nebenbedingungen
κ	$: I^\mu \rightarrow \{0, 1\}$	Abbruchkriterium

Auch bei der $(\mu+1)$ -Evolutionsstrategie werden die Parameter c_d und c_i als Schrittweitenkontrolle für die 1/5-Erfolgsregel verwendet. Die 1/5-Erfolgsregel gilt analog zu der (1+1)-Evolutionsstrategie. Der folgende Algorithmus gibt den Ablauf einer $(\mu+1)$ -Evolutionsstrategie wieder.

```

1  t:=0;
2  initialisiere  $P^{(t)} := \{\mathbf{I}_1^{(t)}, \dots, \mathbf{I}_\mu^{(t)}\}$ ; //Initialisierung
3  solange  $\kappa(P^{(t)}) \neq 1$  //Abbruchkriterium
4       $\mathbf{I}^{(t)} = rek(P^{(t)})$ ; //Rekombination
5       $\mathbf{I}''^{(t)} = mut(\mathbf{I}^{(t)})$ ; //Mutation
6      berechne  $\Psi(\mathbf{I}''^{(t)})$ ;

```

```

7            $P^{(t+1)} := \text{sel}(\mathbf{I}^{(t)} \cup P^{(t)});$  //Selektion
8            $t := t+1;$ 
9           ende

```

Alg. 2-2: Ablauf einer $(\mu+1)$ -Evolutionstrategie

Der neu eingeführte Rekombinationsoperator verändert sowohl die Objekt-, als auch die Strategieparameter eines Individuums:

$$\text{rek}(P^{(t)}) = \mathbf{I}^{(t)} = (\mathbf{x}^{(t)}, \boldsymbol{\sigma}^{(t)}) \in I. \quad (2.43)$$

Zu Beginn der Rekombination werden gleichverteilt¹ zwei Indizes $a = G_n$ und $b = G_n$ bestimmt, die die Individuen $\mathbf{I}_a^{(t)}$ und $\mathbf{I}_b^{(t)}$ bestimmen, die als Rekombinationspartner dienen.

$$x_i^{(t)} = \begin{cases} x_{a,i}^{(t)}, & \text{falls } G_2=1 \\ x_{b,i}^{(t)}, & \text{falls } G_2=2 \end{cases}, \text{ für } 1 \leq i \leq n \text{ und} \quad (2.44)$$

$$\sigma_i^{(t)} = \begin{cases} \sigma_{a,i}^{(t)}, & \text{falls } G_2=1 \\ \sigma_{b,i}^{(t)}, & \text{falls } G_2=2 \end{cases}, \text{ für } 1 \leq i \leq n \quad (2.45)$$

Da die einzelnen Koordinaten der Individuen ohne Veränderung übernommen werden, wird diese Art der Rekombination als diskrete Rekombination bezeichnet. Der Mutationsoperator wird analog zur $(1+1)$ -ES durchgeführt. Die Mutationsfunktion $\text{mut}(\cdot)$ berechnet für jede Koordinate $x_i^{(t)}$ aus $\mathbf{x}^{(t)}$ eine neue Koordinate $x_i^{''(t)}$ mittels der Vorschrift

$$x_i^{''(t)} = x_i^{(t)} + N(0, \sigma_i^{(t)}). \quad (2.46)$$

Das so durch Mutation erzeugte Individuum wird mit $\mathbf{I}^{''(t)}$ bezeichnet. Es gilt:

$$\mathbf{I}^{''(t)} = (\mathbf{x}^{''(t)}, \boldsymbol{\sigma}^{(t)}) \quad (2.47)$$

Der Selektionsoperator verwirft das Individuum mit dem schlechtestem Fitnesswert aus der Population $P^{(t)}$ und übernimmt die restlichen μ Individuen sowie das neu erzeugte Individuum $\mathbf{I}^{''(t)}$ in die folgende Population $P^{(t+1)}$:

$$\begin{aligned} P^{(t+1)} &= \text{sel}(\mathbf{I}^{''(t)} \cup P^{(t)}) \\ &= \{\mathbf{I}_i^{(t)} \in \mathbf{I}^{''(t)} \cup P^{(t)} : \forall \mathbf{I}_i^{(t)} \in P^{(t+1)} \neg \exists \mathbf{I}_j^{(t)} : \Psi(\mathbf{I}_i^{(t)}) < \Psi(\mathbf{I}_j^{(t)})\} \forall i \in \{1, \dots, \mu\} \end{aligned} \quad (2.48)$$

Auch bei dieser Variante der Evolutionstrategien kann die 1/5-Erfolgsregel eingesetzt werden, um die Schrittweiten σ_i anzupassen.

1. Für die Definition der gleichverteilten Zufallsvariablen G_n siehe Definition 2.10.

2.4 (μ, λ) -Evolutionstrategie

Schwefel hat die Idee der $(\mu + 1)$ -ES weiterverfolgt und eine neue Variante der Evolutionstrategien entwickelt, die zusätzlich zu einer Population mit μ Individuen in jeder Generation λ Nachkommen erzeugt. Gründe für diese Erweiterung liegen zum einen in einer möglichen Ausnutzung von Parallelrechnern und zum anderen in der Integration eines Selbstadaptationsmechanismus, der in der Lage ist, ein inneres Modell des Optimierproblems zu lernen, ohne dabei auf externe Heuristiken, wie die 1/5-Erfolgsregel zurückgreifen zu müssen. Die im Vergleich zu der $(1+1)$ - und der $(\mu+1)$ -Evolutionstrategie neu eingeführte Komma-Selektion erlaubt es nur den Nachkommen, in die Nachfolgeneration übernommen zu werden. Jedes Individuum hat damit eine maximale Lebensspanne von einer Generation. Evolutionstrategien mit Komma-Selektion können daher im Verlaufe einer Optimierung Generationen aufweisen, deren Fitness schlechter ist als die der vorhergehenden Generation. Vorteilhaft ist dieses Verhalten, wenn eine Evolutionstrategie ein lokales Optimum erreicht hat und nun versucht, aus diesem zu entkommen. In praktischen Einsätzen wird u. a. aus diesem Grund nahezu ausschließlich die Komma-Selektion eingesetzt.

2.4-1 Grundlegender Algorithmus

(μ, λ) -ES mit Plus- oder Komma-Selektionen können formal wie folgt beschrieben werden:

$$(\mu, \lambda)\text{-ES} = (P^{(0)}, \mu, \lambda, rek, mut, sel, \Sigma, f, g, \kappa), \text{ mit} \quad (2.49)$$

$P^{(0)}$	$= (\mathbf{I}_1^{(t)}, \dots, \mathbf{I}_\mu^{(t)}) \in I^\mu$, mit $I = \mathbb{R}^n \times \mathbb{R}_+^n$	Startpopulation
μ	: $\mu > 1$	Anzahl der Eltern
λ	: $\lambda > \mu$	Anzahl der Nachkommen
rek	: $I^\mu \rightarrow I$	Rekombinationsoperator
mut	: $I \rightarrow I$	Mutationsoperator
sel	: strategieabhängiger Selektionsoperator	
	$I^\lambda \rightarrow I^\mu$	Komma-Selektionsoperator
	$I^{\lambda+\mu} \rightarrow I^\mu$	Plus-Selektionsoperator
Σ	: adaptationsabhängige Menge Steuerparameter	
	$\Sigma = \{\tau_0\}$	Adaptation einer Schrittweite
	$\Sigma = \{\tau, \tau'\}$	Adaptation von n Schrittweiten
	$\Sigma = \{\tau, \tau', \beta\}$	Adaptation von n Schrittweiten und $n((n-1)/2)$ Lagewinkeln
f	: $\mathbb{R}^n \rightarrow \mathbb{R}$	Zielfunktion
g_j	: $\mathbb{R}^n \rightarrow \mathbb{R}$, mit $j \in \{1, \dots, q\}$	Nebenbedingungen
κ	: $I^\mu \rightarrow \{0, 1\}$	Abbruchkriterium

Der folgende Algorithmus gibt den Ablauf einer (μ, λ) -Evolutionstrategie wieder.

```

1   t:=0;
2   initialisiere  $P^{(t)} := \{\mathbf{I}_1^{(t)}, \dots, \mathbf{I}_\mu^{(t)}\}$ ;
3   solange  $\kappa(P^{(t)}) \neq 1$                                      //Abbruchkriterium
4        $\forall(k \in \{1, \dots, \lambda\}) : \mathbf{I}_k^{(t)} := rek(P^{(t)})$ ;         //Rekombination//
5        $\forall(k \in \{1, \dots, \lambda\}) : \mathbf{I}_k^{(t)} := mut(\mathbf{I}_k^{(t)})$ ;         //Mutation//
6        $P^{(t)} := \{\mathbf{I}_1^{(t)}, \dots, \mathbf{I}_\lambda^{(t)}\}$ ;
7       berechne  $\Psi(P^{(t)}) = (\Psi(\mathbf{I}_1^{(t)}), \dots, \Psi(\mathbf{I}_\lambda^{(t)}))$ ;
8a       $P^{(t+1)} := sel(P^{(t)})$ ;                                     //Komma-Selektion//
8b       $P^{(t+1)} := sel(P^{(t)} \cup P^{(t)})$ ;                       //Plus-Selektion//
9       t:= t + 1;
10  ende

```

Alg. 2-3: Ablauf einer (μ, λ) -Evolutionstrategie

Der Algorithmus 2-3 beschreibt sowohl die Komma-Variante der Selektion, falls die Zeile 8a verwendet wird, als auch die Plus-Variante, falls die Zeile 8b verwendet wird. Die (μ, λ) -Evolutionstrategien verwenden eine Mutation, die nicht mehr von einem Meta-Algorithmus gesteuert wird, sondern eine Steuerung, die ihre wesentlichen Informationen innerhalb eines Individuums speichert und so Einfluss auf die zur Mutation gehörende Verteilungsfunktion nimmt. Diese Informationen beziehen sich auf die Schrittweiten sowie auf Lagewinkel, die in der Lage sind, eine n -dimensionale, korrelierte Normalverteilung im Raum zu orientieren. Dadurch, dass diese Informationen im Individuum gespeichert sind, unterliegen sie der allgemeinen Mutation und werden in jeder Generation verändert.

Die Qualität der Mutation der Strategiekomponenten wird indirekt gemessen, indem lediglich die Fitness der Objektkomponente des Nachkommen bestimmt wird. Die Idee, die diese Art der indirekten Bewertung motiviert ist die, dass eine gute Mutation der Strategiekomponenten bewirkt, dass die hieraus erzeugten Nachkommen eine höhere Fitness aufweisen, als die Nachkommen, die durch schlecht mutierte Strategiekomponenten erzeugt werden. Somit erlangen gute Strategiekomponenten einen Vorteil gegenüber anderen und erlauben eine Adaptation an die zu optimierende Funktion. Im folgenden werden die notwendigen Schritte der Evolutionstrategie sowie die zugehörigen Operatoren formal beschrieben.

2.4-2 Repräsentation

Ein Individuum $\mathbf{I}^{(t)}$ besteht aus zwei Komponenten, der Objektkomponente $\mathbf{x}^{(t)}$ sowie der Strategiekomponente, die durch die Matrix $\mathbf{C}^{(t)}$ beschrieben wird:

$$\mathbf{I}^{(t)} = (\mathbf{x}^{(t)}, \mathbf{C}^{(t)}) \quad (2.50)$$

Für die Darstellung der Strategiekomponente wird hier die allgemeinste Form, die Form einer Matrix, gewählt. Die von Schwefel entwickelten Adaptationsarten können alle durch Veränderung dieser Matrix ausgedrückt werden. Für einige Adaptationsarten lässt sich die Matrix stark vereinfachen, da dann nur Vektoren bzw. Skalare verwendet werden.

2.4-3 Rekombination

Der erste Schritt innerhalb einer Generation ist die Rekombination (Algorithmus 2-3, Zeile 4):

$$\forall(k \in \{1, \dots, \lambda\}) : \mathbf{I}_k^{(t)} := \text{rek}(P^{(t)}) \text{ und} \quad (2.51)$$

$$P^{(t)} := \{\mathbf{I}_1^{(t)}, \dots, \mathbf{I}_\lambda^{(t)}\} \quad (2.52)$$

Die Rekombination kann auf unterschiedliche Arten ausgeführt werden. Eine formale Definition der möglichen Rekombinationsarten ist bereits im Abschnitt 2.1-5 (Rekombination, Seite 18) aufgeführt worden. Die Grundarten der Rekombination sind (global) diskrete bzw. (global) intermediäre Rekombination. Eine gute Übersicht über weitere Rekombinationsarten beschreibt Mandischer in [Man00].

2.4-4 Mutation der Strategiekomponenten – Selbstadaptation

Durch die Rekombination entsteht aus den μ Elternindividuen eine Zwischenpopulation $P^{(t)}$, die aus λ Nachkommen besteht. Die Individuen aus $P^{(t)}$ werden in dem folgenden Schritt mutiert (Zeile 5) und bilden damit die Zwischenpopulation $P''^{(t)}$:

$$\forall(k \in \{1, \dots, \lambda\}) : \mathbf{I}_k''^{(t)} := \text{mut}(\mathbf{I}_k^{(t)}) \quad (2.53)$$

$$P''^{(t)} := \{\mathbf{I}_1''^{(t)}, \dots, \mathbf{I}_\lambda''^{(t)}\} \quad (2.54)$$

Die Matrix $\mathbf{C}_i^{(t)}$ eines Individuums \mathbf{I}_i der Generation t enthält alle Informationen über Lagewinkel und Schrittweiten. Die Selbstadaptation erlaubt die Anpassung von $\mathbf{C}_i^{(t)}$ während der Optimierung, um so eine möglichst optimale Belegung der Komponenten von $\mathbf{C}_i^{(t)}$ zu erlangen. $\mathbf{C}_i^{(t)}$ wird während der Optimierung permanent durch den Mutationsoperator verändert. Die indirekte Bewertung von $\mathbf{C}_i^{(t)}$ durch die Fitness der erzeugten Individuen garantiert dann die Selektion der Nachkommen, deren Strategiematrix die erfolgreichsten Veränderungen bewirkt hat. Die Mutation ist in (2.50) in der allgemeinsten Form dargestellt. Es existieren drei Mutationsvarianten, die die Strategiekomponente eines Individuums mutieren. Mit dieser mutierten Strategiekomponente wird dann die Objekt-komponente mutiert. Die drei Mutationsvarianten der Strategiekomponente wird im folgenden erläutert.

2.4-5 Adaptation einer Schrittweite

Für den Fall, dass nur eine Schrittweite pro Individuum den gesamten Optimierprozess steuert, ist die Struktur Strategiematrix $\mathbf{C}_i^{(t)}$ sehr einfach, da sie nur von der Schrittweite $\sigma_i^{(t)}$ des Individuums \mathbf{I}_i der Generation t bestimmt wird:

$$\mathbf{C}_i^{(t)} = \sigma_i^{(t)} \quad (2.55)$$

Die Individuen $\mathbf{I}_i^{(t)}$ der Population $P^{(t)}$, d. h. die Individuen der Zwischenpopulation, die nach der Rekombination entsteht, haben folgenden Aufbau:

$$\mathbf{I}_i^{(t)} = \left(\mathbf{x}_i^{(t)}, \underbrace{\sigma_i^{(t)}}_{\mathbf{C}_i^{(t)}} \right) \quad (2.56)$$

Die Mutation der Strategiekomponente findet wie folgt statt:

$$\sigma_i^{n(t)} = \text{mut}(\sigma_i^{(t)}) \quad (2.57)$$

mit nachstehender Realisierung des Mutationsoperators:

$$\text{mut}(\sigma_i^{(t)}) = \sigma_i^{(t)} \cdot z \quad (2.58)$$

Für die Zufallsvariable z gilt:

$$z \sim LN(0, \tau_0) \quad (2.59)$$

Für die Mutation der Strategiematrix wird eine logarithmisch normalverteilte Zufallsvariable verwendet. Durch Multiplikation mit der Ursprungsschrittweite kann so sichergestellt werden, dass die neue Schrittweite immer positiv ist. Bei Addition einer normalverteilten Zufallsvariablen ist dies nicht garantiert. Für den Steuerparameter τ_0 schlägt Schwefel den aus dem Kugelmodell abgeleiteten Wert $1/\sqrt{n}$ vor. Die Höhenlinien der n -dimensionalen Dichtefunktion, die durch die mutierte Strategiematrix $\mathbf{C}_i^{(t)}$ definiert wird, beschreiben für den Fall der Adaptation einer Schrittweite eine n -dimensionale Hyperkugel. Ein typischer Verlauf einer solchen Dichtefunktion ist für ein 2-dimensionales Beispiel in der folgenden Abbildung 2.2 gegeben.

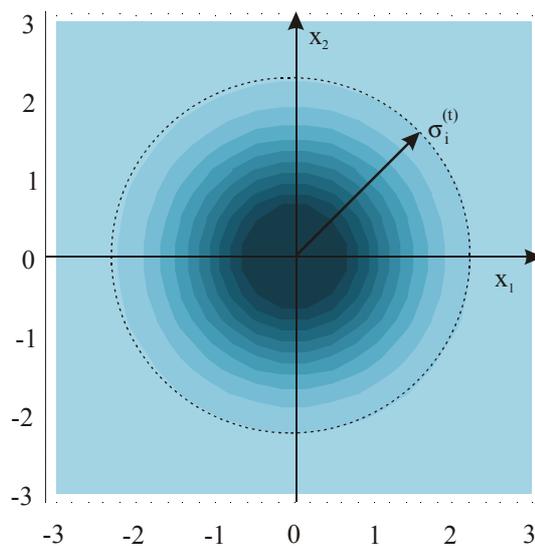


Abb. 2.2: Adaptation einer Schrittweite

Die so generierte, neue Schrittweite ist Basis für die Mutation der Objektkomponenten, die im Abschnitt 2.4-8 (Mutation der Objektkomponenten, Seite 33) beschrieben wird.

2.4-6 Adaptation von n Schrittweiten

Die Adaptation von n Schrittweiten erfolgt ähnlich zur Adaptation einer Schrittweite mit der Ausnahme, dass σ nicht mehr als Skalar verwendet wird, sondern als n -dimensionaler Vektor σ . Die Strategiematrix $\mathbf{C}_i^{(t)}$ hat dann folgende Struktur:

$$\mathbf{C}_i^{(t)} = \sigma_i^{(t)} \mathbf{1} \text{ bzw.} \quad (2.60)$$

$$\mathbf{C}_i^{(t)} = \begin{bmatrix} \sigma_1^{(t)} & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & \sigma_n^{(t)} \end{bmatrix} \quad (2.61)$$

Die Individuen $\mathbf{I}_i^{(t)}$ der Population $P^{(t)}$ haben folgenden Aufbau:

$$\mathbf{I}_i^{(t)} = \left(\mathbf{x}_i^{(t)}, \underbrace{\sigma_i^{(t)} \cdot \mathbf{1}}_{\mathbf{C}_i^{(t)}} \right) \quad (2.62)$$

Die Mutation der Strategiekomponente findet wie folgt statt:

$$\sigma_i'^{(t)} = \text{mut}(\sigma_i^{(t)}) \quad (2.63)$$

mit nachstehender Realisierung des Mutationsoperators, die die Komponenten $\sigma_{i,j}^{(t)}$ des Vektors $\sigma_i^{(t)}$ getrennt verarbeitet

$$\text{mut}(\sigma_{i,j}^{(t)}) = \sigma_{i,j}^{(t)} \cdot z \cdot z_i \quad (2.64)$$

mit der Zufallsvariablen

$$z \sim LN(0, \tau') \quad (2.65)$$

die einmal generiert für die gesamte Generation gilt und der Zufallsvariablen

$$z_i \sim LN(0, \tau), \text{ mit } 1 \leq i \leq n \quad (2.66)$$

die für jede Koordinate neu realisiert wird. Für die Steuerparameter τ und τ' schlägt Schwefel die Werte $\tau = 1/\sqrt{2\sqrt{n}}$ und $\tau' = 1/\sqrt{2n}$ vor, die aus dem Kugelmodell abgeleitet wurden.

Die Höhenlinien der n -dimensionalen Dichtefunktion beschreiben für den Fall der Adaptation von n Schrittweiten n -dimensionale Hyperellipsoide. Die Abbildung 2.3 stellt diesen Mechanismus am Beispiel einer 2-dimensionalen Adaptation dar.

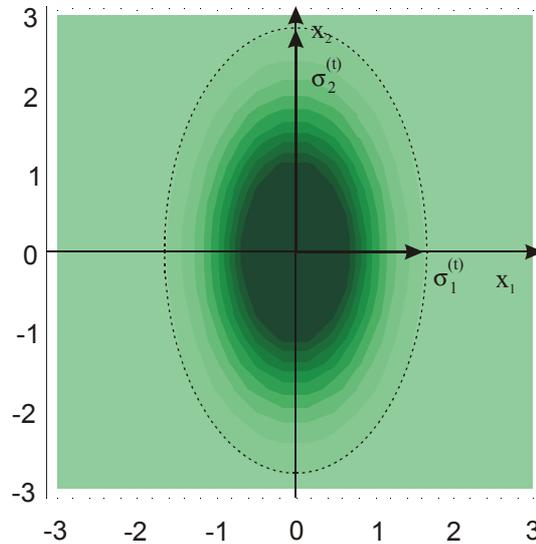


Abb. 2.3: Adaptation n Schrittweiten

Auch hier gilt, dass die so generierte, neue Schrittweite die Basis für die Mutation der Objektkomponenten, die im Abschnitt 2.4-8 (Mutation der Objektkomponenten, Seite 33) beschrieben wird, ist.

2.4-7 Adaptation von n Schrittweiten und $(n(n-1))/2$ Lagewinkeln

Die zuvor vorgestellten Adaptationsarten sind in der Lage, Dichtefunktionen zu realisieren, deren Höhenlinien Hyperkugeln bzw. Hyperellipsoide sind. Die Orientierung der Hyperellipsoide entspricht hierbei dem Koordinatensystem des Optimierproblems, d.h. die Symmetrieachsen der Hyperellipsoide liegen entlang der Achsen des Koordinatensystems. Aus Sicht der Stochastik wird dies durch den Einsatz von Mutationsoperatoren erreicht, die unabhängige Zufallsvariablen verwenden. Bereits in (2.11) wurde gezeigt, dass es möglich ist, n -dimensionale Normalverteilungen zu erzeugen, bei der die Komponenten voneinander abhängig, d. h. korreliert sind. Die Korrelation kann erreicht werden, wenn die Kovarianzen c_{ij} der Kovarianzmatrix $\mathbf{C}_i^{(t)}$ für $i \neq j$ nicht verschwinden. Schwefel merkt jedoch an, dass die Veränderung der Kovarianzen durch Mutation die Gefahr birgt, dass die Höhenlinien der resultierenden n -dimensionalen Normalverteilung keine Hyperellipsoide darstellen [Sch94]. Eine solche entartete Normalverteilung wäre im Bereich der Evolutionsstrategien nicht einsetzbar. Als Alternative schlägt Schwefel die Generierung einer unkorrelierten n -dimensionalen Normalverteilung vor und darauf aufbauend eine Rotation der Koordinatenachsen mittels $(n(n-1))/2$ Lagewinkeln α_{ij} anstelle der Kovarianzen [Sch94]. Den Zusammenhang zwischen der Lagewinkelmatrix und der Kovarianzmatrix gibt Bäck [Bäc94]:

$$\tan(2\alpha_{ij}) = \frac{2c_{ij}}{\sigma_i^2 - \sigma_j^2} \quad (2.67)$$

Durch die Einführung der Lagewinkelmatrix α sind Schrittweiten und Lagewinkel entkoppelt, d. h. beide Strategiekomponenten können in eigenständigen Schritten verändert werden.

$$\sigma_i^{(t)} = \text{mut}(\sigma_i^{(t)}) \text{ und} \quad (2.68)$$

$$\alpha_i^{(t)} = \text{mut}(\alpha_i^{(t)}) \quad (2.69)$$

Die Realisierung der Mutationsfunktion für den Schrittweitenvektor ist identisch zu (2.64) - (2.66). Schwefel schlägt zur Veränderung der Lagewinkelmatrix α eine additive Mutation vor:

$$\alpha_i^{(t)} = \alpha_i^{(t)} + z, \text{ mit} \quad (2.70)$$

$$z \sim N(0, \beta) \quad (2.71)$$

Die Standardabweichung β soll für alle Winkel identisch gewählt werden. Durch Einsatz der Adaptation von Schrittweiten und Lagewinkeln ist es nun möglich, die Hyperellipsoide beliebig im Raum zu orientieren.

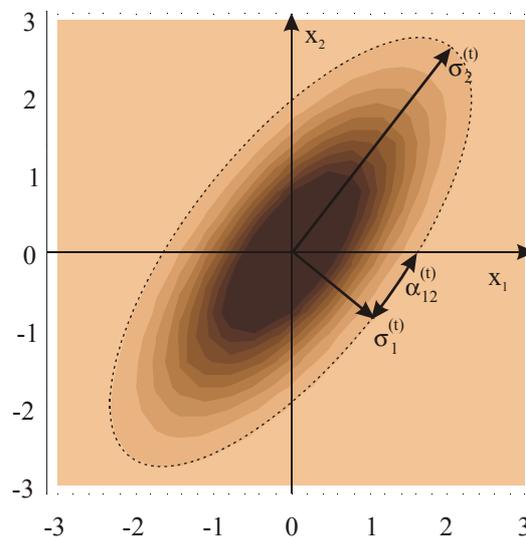


Abb. 2.4: Adaptation von Schrittweiten und Lagewinkeln

2.4-8 Mutation der Objektkomponenten

Ist durch eine der drei genannten Varianten die Strategiekomponente mutiert worden, kann mit dieser Strategiekomponente eine neue Objektkomponente generiert werden. Dies geschieht im Prinzip durch Addition einer zentrierten Normalverteilung mit Kovarianzmatrix $\mathbf{C}^{(t)}$. Da eine direkte Mutation der Kovarianzmatrix im Fall der korrelierten Adaptation nicht sinnvoll ist, wird hier zunächst die unkorrelierte Mutation vorgestellt.

Die Komponenten $x_{i,j}^{(t)}$ des i -ten Nachkommen der Generation t werden wie folgt aus der Zwischenpopulation $P^{(t)}$ mit den Individuenkomponenten $x_{i,j}^{(t)}$ erzeugt:

$$x_{i,j}^{(t)} = x_{i,j}^{(t)} + N(\mathbf{0}, \mathbf{C}_i^{(t)}) \quad (2.72)$$

Für den Fall der korrelierten Adaptation werden die ebenfalls temporären Individuenkomponenten analog zu (2.72) bestimmt. Die Mutation der Komponenten $x_{i,j}^{(t)}$ basiert jedoch nur auf den Diago-

nalelementen der Kovarianzmatrix $\mathbf{C}^{(t)}$, d. h. die Mutation berücksichtigt für diesen Schritt nur die Varianzen. Die Korrelation der Koordinaten und damit die Drehung des Hyperellipsoids um den Koordinatenursprung wird erreicht, indem die Komponenten der Lagewinkelmatrix $\boldsymbol{\alpha}$ verwendet werden, um sukzessive Drehungen um jeweils zwei Koordinatenachsen durchzuführen. Die Komplexität dieser Drehungen ist $O(n^2)$, falls die maximale Anzahl an Lagewinkeln verwendet wird. n bezeichnet hier die Dimensionalität des Optimierproblems. Der Algorithmus zur Drehung der Objektkomponente \mathbf{x} in Abhängigkeit von $\boldsymbol{\alpha}$ lautet wie folgt [Sch94]:

```

1  Funktion Drehung: Eingabe= $\mathbf{x}, \boldsymbol{\alpha}$ , Ausgabe gedrehter Vektor  $\mathbf{x}'$ 
2   $k \leftarrow |\boldsymbol{\alpha}|$ 
3   $n_x \leftarrow |\mathbf{x}|$ 
4  Schleife über  $1 \leq i \leq |\boldsymbol{\alpha}|$ 
5       $n_1 \leftarrow n_x - i$ 
6       $n_2 \leftarrow n_x$ 
7      Schleife über  $j \leq 1 \leq i$ 
8           $x_{n_2} \leftarrow \sin(\alpha_k) \cdot x_{n_1} + \cos(\alpha_k) \cdot x_{n_2}$ 
9           $x_{n_1} \leftarrow \cos(\alpha_k) \cdot x_{n_1} - \sin(\alpha_k) \cdot x_{n_2}$ 
10          $n_2 \leftarrow n_2 - 1$ 
11          $k \leftarrow k - 1$ 
12     Ende Schleife über  $j$ 
13 Ende Schleife über  $i$ 
14 Rückgabe  $\leftarrow \mathbf{x}$ 

```

Alg. 2-4: Drehung der Objektkomponente

2.5 Zusammenfassung

Die von Rechenberg und Schwefel entwickelten klassischen Varianten der Evolutionsstrategien stellen ein leistungsstarkes Werkzeug für die Optimierung zur Verfügung. In seiner Dissertation vergleicht Schwefel das Verhalten der Varianten an unterschiedlichen Zielfunktionen und setzt die Resultate zu weiteren Optimierverfahren in Relation [Sch75]. Hierzu zählen die Koordinatenstrategie, das Complex Verfahren von Box, das Verfahren der konjugierten Richtungen sowie die Methode von Hooke und Jeeves [Fle87, Hof71, Kel99, Pow98]. Hierbei hat sich gezeigt, dass eine (10, 100)-Evolutionstrategie nahezu alle Testaufgaben lösen konnte. Das Versagen dieser Variante lag meist in einer Degenerierung des zur Mutation erlaubten Bereiches, sodass kaum noch gültige Mutationen erzeugt werden konnten (Problem 2.35) bzw. in einer sehr kleinen Schrittweite und damit bedingter, geringer Konvergenzgeschwindigkeit (Problem 2.14). Die zweigliedrige (1 + 1)-Evolutionstrategie konnte deutlich weniger Testaufgaben lösen, da sie häufig nur lokale Optima fand. Deterministische Verfahren, wie das Complex Verfahren von Box, das Verfahren der konjugierten Richtungen sowie die Methode von Hooke und Jeeves konnten einige Testaufgaben zufriedenstellend lösen, andere Aufgaben jedoch nicht. Zum Teil versagten sie völlig, d. h. selbst lokale Optima konnten nicht gefunden werden.

Bei hochdimensionalen Testproblemen zeigt sich, dass die Rekombination für eine sichere Konvergenz notwendig ist. Deterministische Verfahren müssen bei den hochdimensionalen Testaufgaben häufig neu starten bzw. konvergieren nicht. Eine vollständige Bewertung der Verfahren findet sich in [Sch75].

Kapitel 3 Erweiterungen der Evolutionstrategien

*Ich glaube nicht, dass das zerstörte Alte
schon etwas Neues ist.
(Rainer Maria Rilke)*

In den letzten Jahren sind die klassischen Evolutionstrategien nach Rechenberg und Schwefel erweitert bzw. neue Operatoren für Mutation, Rekombination und Selektion entwickelt worden. Ziel war es dabei, die Konvergenzeigenschaften der Evolutionstrategien zu verbessern bzw. Selbstadaptationsmechanismen zu entwickeln, die eine weitgehende Anpassung an die Problemtopologie ermöglichen, jedoch die Anzahl der Strategieparameter und die mathematische Komplexität möglichst minimal zu halten. In diesem Kapitel werden Untersuchungen von Rudolph [Rud98], Salomon [Sal98a, Sal98b], Ohkura, Matsumura und Ueda [Ohk98], Storn und Price [Sto95, Sto96], Gorges-Schleuter [Gor98] sowie Ostermeier und Hansen [Han98] und Weicker [Wei01] beschrieben und diskutiert.

Zusätzlich zu diesen Erweiterungen der Evolutionstrategien werden in diesem Kapitel neue Ansätze aus dem Bereich der reellwertig kodierten Genetischen Algorithmen (GA) vorgestellt. Sie zeigen, dass auch im Bereich der Genetischen Algorithmen selbstadaptive Verfahren an Bedeutung gewinnen. Im Einzelnen konzentriert sich dieser Abschnitt auf *Simulated Binary Crossover* (SBX) von Deb und Beyer [Deb99] sowie die Familie der *Estimation of Distribution Algorithms* (EDA) nach u. a. Mühlenbein, et al und Larrañaga, et. al. [Sys93, Bal94, Kva96, Bal97, Har97, Müh97, Pel98, Müh99a - Müh99c, Pel99].

3.1 Variation der Mutationsfunktion

In [Rud98] untersucht Rudolph das asymptotische Konvergenzverhalten der $(1+1)$ - und der $(1, \lambda)$ -Evolutionstrategien für bestimmte Verteilungsfunktionen am Beispiel einer quadratischen Zielfunktion. Die untersuchten Verteilungen eines Zufallvektors \mathbf{X} müssen drei Bedingungen erfüllen:

1. Bei der Verteilungsfunktion von \mathbf{X} muss es sich um eine Mutationsfunktion handeln, d. h.

$$E_{\mathbf{X}} = 0 \tag{3.1}$$

2. Die Verteilungsfunktion von \mathbf{X} muss faktorisierend sein, d. h. für die zugehörige n -dimensionale Dichtefunktion muss gelten:

$$\varphi_{\mathbf{X}}(\mathbf{x}) = \varphi_{\mathbf{X}}(x_1, \dots, x_n) = \prod_{i=1}^n \varphi_{X_i}(x_i) \quad (3.2)$$

mit

$$\varphi_{X_1}(\cdot) = \dots = \varphi_{X_n}(\cdot) \quad (3.3)$$

3. Die vierten absoluten Momente¹ der betrachteten Verteilungen müssen endlich sein, d. h. bei gegebenen Dichtefunktionen $\varphi_{X_i}(x_i)$ muss gelten

$$\int_{-\infty}^{\infty} |x_i - E_{X_i}|^4 \varphi_{X_i}(x_i) dx_i < \infty \quad (3.4)$$

Der von Rudolph verwendete Begriff faktorisierend ist damit hier gleichbedeutend mit der Bezeichnung unabhängig und identisch verteilt. Übertragen auf den hier diskutierten Fall bedeutet dies, dass die n Komponenten z_i des Zufallsvektors \mathbf{z} unabhängig und identisch verteilt sind.

Verteilungsfunktionen, die die Eigenschaften nach (3.2) bis (3.4) erfüllen und von Rudolph untersucht wurden, sind die Normalverteilung, die logistische Verteilung, die Laplace-Verteilung und die Student's t -Verteilung. Die Abbildung 3.1 gibt einen Überblick über die zugehörigen Dichtefunktionen dieser Verteilungen. In der Reihenfolge steigender Maxima sind hier die Normalverteilung, die logistische Verteilung, die Student's t -Verteilung sowie die Laplace-Verteilung dargestellt.

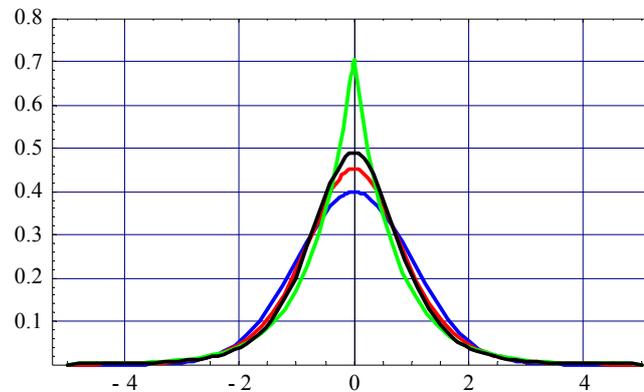


Abb. 3.1: Verlauf der von Rudolph untersuchten Dichtefunktionen

Rudolph untersucht diese Verteilungen auf Basis zentraler Grenzwertsätze und kann das asymptotische Verhalten der Fortschrittsrate für eine (1+1)- sowie eine (1, λ)-Evolutionstrategie am Beispiel einer quadratischen Zielfunktion bestimmen. Er stellt dabei fest, dass alle untersuchten Verteilungsfunktionen eine nahezu identische, schnelle Annäherung an das Optimum gewährleisten. Rudolph folgert aus seinen Resultaten zusätzlich, dass Selbstadaptationsmechanismen, die ursprünglich für normalverteilte Mutation entwickelt wurden, für andere faktorisierende Verteilungsfunktionen nicht modifiziert werden müssen.

Zu einer ähnlichen Aussage kommt auch Kappler [Kap96]. Sie untersucht den Einfluss einer auf der Cauchy-Verteilung basierenden Mutationsfunktion, auf Robustheit und Fortschrittsrate am Beispiel

1. Für die Definition der Momente siehe auch Abschnitt 4.1-3 (Momente, Seite 66)

der geeigneten Ebene und des geeigneten Korridormodells für eine $(1 + 1)$ -Evolutionstrategie. Für den 1-dimensionalen Fall verhält sich eine Evolutionsstrategie mit Cauchy-Mutation ähnlich zu einer normalverteilten Mutation. Für den mehrdimensionalen Fall existieren Bedingungen (geringe Schrittweite und große Anzahl Nachkommen), unter denen die Cauchy-Mutation eine höhere Konvergenzrate als eine normalverteilte Mutation ermöglicht. Eine allgemeine Aussage über diese speziellen Bedingungen hinaus kann Kappler auf Basis ihrer Untersuchungen nicht treffen.

Konvergenzgeschwindigkeiten sind ebenfalls von Rudolph für den Fall der Cauchy-Mutation untersucht worden [Rud97]. Seine Arbeit auf diesem Gebiet führt die Arbeiten von Kappler weiter, indem er die Eignung sphärischer und nicht-sphärischer multivariater Cauchy-Verteilungen als Mutationsfunktionen bewertet. Beispielhafte Zielfunktion ist auch bei dieser Untersuchung eine quadratische Funktion. Rudolph stellt dabei fest, dass die Anzahl der Iterationen, die zum Erreichen eines festgelegten Zielfunktionswertes notwendig ist, für sphärische Cauchy-Verteilungen linear und für nicht-sphärische Cauchy-Verteilungen quadratisch mit der Anzahl der Problemdimension steigt. Zudem merkt er an, dass Adaptationsparameter, die für normalverteilte Mutation ermittelt wurden, für nicht normalverteilte Mutation überprüft werden müssen.

3.2 Kumulative Schrittweitenregelung (KSA)

Von Ostermeier und Hansen sind Evolutionsstrategien entwickelt worden, bei denen die Strategieparameter auf die gleiche Art und Weise mutiert werden wie die Objektparameter. So wird die indirekte Bewertung der Strategieparameter durch eine Bewertung auf Basis der Auswertung der Objektparameter ersetzt. Als Begründung für diese Ersetzung geben sie an, dass eine Mutation, die eine geringe Veränderung einer Koordinate hervorruft, nicht zwangsläufig durch eine kleine Schrittweite bedingt sein muss, sondern auch durch eine große Schrittweite entstehen kann, falls die Realisierung der Zufallsvariablen zufällig einen kleinen Wert ergibt. Um diese Kopplung von Mutation und Realisierung der zugehörigen Zufallsvariablen aufzulösen, werden Mutationen eingeführt, bei denen die Realisierung der Zufallsvariablen gleichzeitig auf Objektvariable und (in transformierter Form) auf die Strategievariable wirkt. Die so entstehenden Verfahren nennen sie *entstochastisiert*. Die Grundlagen und Beweggründe für die Einführung der Entstochastisierung sowie die Übertragung auf eine $(1, \lambda)$ -Evolutionstrategie werden ausführlich in [Ost94] behandelt. In diesem und dem folgenden Abschnitt werden zwei weitere, entstochastisierte Verfahren, die kumulative Schrittweitenregelung und die Kovarianzmatrixadaptation, beschrieben.

3.2-1 Grundidee

Die kumulative Schrittweitenregelung basiert auf den Arbeiten von Ostermeier und Hansen [Ost97, Han98]. Die Grundidee ist die Nutzung von generationsübergreifenden Informationen zur Adaptation der (globalen) Schrittweite, indem ein sog. Evolutionspfad im Suchraum konstruiert wird. Der Evolutionspfad ist dabei der Differenzvektor zwischen dem Individuum und seinem mehrere Generationen zurückliegenden Vorgänger. Die Anzahl der Generationen, die zwischen dem Individuum und dem Vorgänger liegen, definiert dabei den Bereich der generationsübergreifenden Speicherung von Informationen. Die Länge des Evolutionspfades bestimmt dann die Veränderung der Schrittweite. Zur Verdeutlichung dieser Idee gibt Hansen folgendes Beispiel, bei dem der Evolutionspfad farblich hervorgehoben ist.

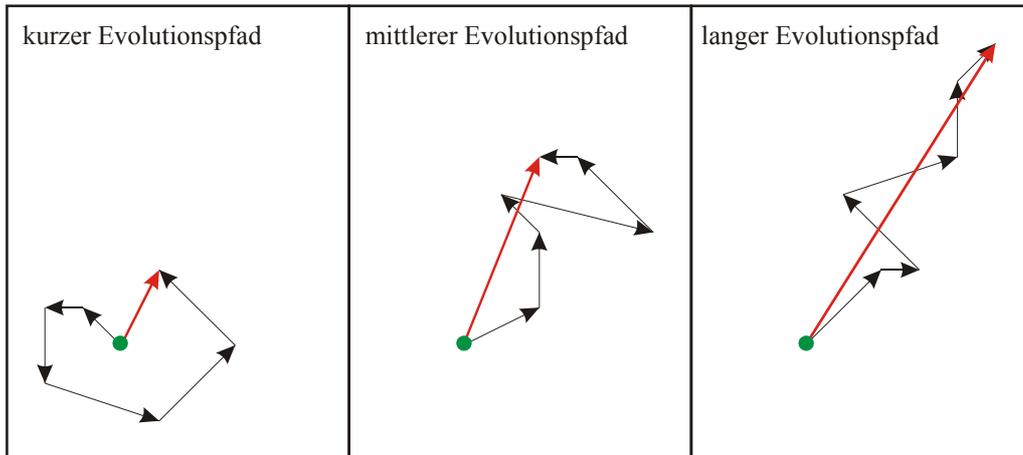


Abb. 3.2: Typische Evolutionspfade bei der kumulativen Schrittweitenadaptation

Die einzelnen Mutationsvektoren sind bezogen auf ihre Länge ähnlich. Die Länge des resultierenden Evolutionspfades ist jedoch sehr unterschiedlich. Ist der Evolutionspfad kurz, besteht der Verdacht, dass die Evolutionsstrategie bereits in der Nähe eines Optimums angelangt ist die Schrittweite wird verringert. Der gegensätzliche Fall besteht, falls der Evolutionspfad lang ist. In diesem Fall wird die Schrittweite vergrößert. Für den Fall einer mittleren Länge des Evolutionspfades, die ungefähr der Länge des Evolutionspfades bei zufälliger Selektion entspricht, wird die Schrittweite nicht verändert.

3.2-2 $(\mu/1, \lambda)$ -KSA-ES

Die $(\mu/1, \lambda)$ -KSA-ES ist eine Evolutionsstrategie mit μ Eltern und λ Nachkommen. Die Bezeichnung der Evolutionsstrategie entspricht hier dem Schema $(\mu/\rho, \lambda)$ ¹, bei dem der zusätzliche Parameter $\rho \leq \mu$ die Anzahl der für die Rekombination zur Verfügung stehenden Elternindividuen bezeichnet. Bei der $(\mu/1, \lambda)$ -KSA-ES findet demnach keine Rekombination statt.

Der folgende Algorithmus gibt den Ablauf einer $(\mu/1, \lambda)$ -KSA-Evolutionsstrategie wieder:

```

1   t := 0;
2   initialisiere  $P^{(t)} := \{\mathbf{I}_1^{(t)}, \dots, \mathbf{I}_\mu^{(t)}\}$ ;           //Initialisierung
3   solange  $\kappa(P^{(t)}) \neq 1$                                    //Abbruchkriterium
4        $\forall (k \in \{1, \dots, \lambda\}) : \mathbf{I}_k^{(t)} := \text{mut}(\mathbf{I}_k^{(t)})$ ;           //Mutation//
5        $P^{(t)} := \{\mathbf{I}_1^{(t)}, \dots, \mathbf{I}_\lambda^{(t)}\}$ ;
6       berechne  $\Psi(P^{(t)}) = (\Psi(\mathbf{I}_1^{(t)}), \dots, \Psi(\mathbf{I}_\lambda^{(t)}))$ ;
7        $P^{(t+1)} := \text{sel}(P^{(t)})$ ;                               //Selektion//
8       t := t + 1;
9   ende
```

Alg. 3-1: Ablauf einer $(\mu/1, \lambda)$ -KSA-ES

1. Für weitere Angaben zu den Notationen siehe Definition 2.11 in Abschnitt 2.1-6 (Selektion, Seite 20) und Abschnitt 3.7 (Lokale Selektion - Diffusionsmodelle (LES), Seite 52)

Der Algorithmus 3-1 entspricht dem Algorithmus der (μ, λ) -Evolutionsstrategie aus dem Abschnitt 2.4 ((μ, λ) -Evolutionsstrategie, Seite 27) ohne Rekombination unter Komma-Selektion. Ein Individuum $\mathbf{I}_i^{(t)}$ hat dabei folgenden Aufbau:

$$\mathbf{I}_i^{(t)} = (\mathbf{x}_i^{(t)}, \mathbf{s}_i^{(t)}, \delta_i^{(t)}) \quad (3.5)$$

mit der Objektkomponente $\mathbf{x}_i^{(t)} \in \mathbb{R}^n$, dem Evolutionspfad $\mathbf{s}_i^{(t)} \in \mathbb{R}^n$ und der Schrittweite $\delta_i^{(t)} \in \mathbb{R}$. Die Realisierung des Mutationsoperators $mut(\cdot)$ unterscheidet sich wesentlich von der klassischen Variante. Die Objektkomponente wird mittels der zuvor bestimmten Schrittweite wie folgt berechnet:

$$\mathbf{x}_i^{(t+1)} = \mathbf{x}_{\zeta_i}^{(t)} + \delta_{\zeta_i}^{(t)} \mathbf{z}_i^{(t+1)} \quad (3.6)$$

$\mathbf{x}_i^{(t+1)}$ ist hier der Nachkomme, der aus $\mathbf{x}_i^{(t)}$ durch Addition des durch die Schrittweite $\delta_{\zeta_i}^{(t)}$ skalierten n -dimensionalen Zufallsvektor $\mathbf{z}_i^{(t+1)}$ bestimmt wird. $\mathbf{z}_i^{(t+1)}$ ist normalverteilt mit Erwartungswert $\mathbf{0}$ und Standardabweichung $\mathbf{1}$. Alle Zufallsvektoren $\mathbf{z}_i \in \mathbb{R}^n$ werden bei der kumulativen Schrittweitenadaptation unabhängig voneinander generiert. Nach der Mutation der Objektkomponente wird die Strategiekomponente mutiert. Im ersten Schritt wird für jeden Nachkommen der Evolutionspfad $\mathbf{s}_i^{(t+1)}$ bestimmt:

$$\mathbf{s}_i^{(t+1)} = (1 - c)\mathbf{s}_{\zeta_i}^{(t)} + c_u \mathbf{z}_i^{(t+1)} \quad (3.7)$$

mit

$$c_u = \sqrt{c(2 - c)} \quad (3.8)$$

Der Parameter c , mit $0 < c \leq 1$ bestimmt den Kumulationszeitraum. Für den Fall $c = 1$ findet keine Kumulation statt, es gilt dann: $\mathbf{s}_i^{(t+1)} = \mathbf{z}_i^{(t+1)}$. Die Zufallszahl ζ_i ist die Realisierung einer im Intervall $[1, \mu] \in \mathbb{N}$ gleichverteilten Zufallsvariable, die den Index eines selektierten Individuums bestimmt. ζ_i entspricht damit der Funktion G_μ , die im Abschnitt 2.1-5 (Rekombination, Seite 18) eingeführt wurde. Für jeden Nachkommen der Generation t wird ζ_i einmal bestimmt und dann ohne Veränderung in den Gleichungen (3.6), (3.7) und (3.9) verwendet. Dieses Vorgehen sichert den entstochastisierten Anspruch dieser Variante der Evolutionsstrategien. Mittels des Evolutionspfades $\mathbf{s}_i^{(t+1)}$ wird nun die Schrittweite $\delta_i^{(t+1)}$ des i -ten Nachkommens berechnet:

$$\delta_i^{(t+1)} = \delta_{\zeta_i}^{(t)} e^{\frac{\|\mathbf{s}_i^{(t+1)}\| - \hat{\chi}_n}{D \cdot \chi_n}} \quad (3.9)$$

Hier ist D der Dämpfungswert, der die Geschwindigkeit der Änderung der Schrittweite in der Generationenfolge bestimmt. Hansen empfiehlt den Standardwert $D = \sqrt{n}$ und gibt in [Han98] Empfehlungen, wie D in Abhängigkeit von c und der Zielfunktion gewählt werden sollte. $\hat{\chi}_n$ ist der Erwartungswert der Länge einer n -dimensionalen Normalverteilung mit Erwartungswert $\mathbf{0}$ und Standardabweichung $\mathbf{1}$. In der Arbeit von Hansen wird dieser Wert mit

$$\hat{\chi}_n = \sqrt{n} \left(1 - \frac{1}{4n} + \frac{1}{21n^2} \right) \quad (3.10)$$

angenähert. Durch (3.7) und (3.9) ist der Mutationsoperator der Schrittweiten definiert. Da für jedes Individuum lediglich eine Schrittweite existiert, ist der Mutationsoperator dem Mutationsoperator der Adaptation einer Schrittweite entfernt ähnlich (siehe 2.4-5 (Adaptation einer Schrittweite, Seite 29)). Von Bedeutung ist jedoch, dass bei dieser Variante die Reihenfolge von Strategie- und Objekt-komponentenmutation umgekehrt wurde. Es werden zuerst die Objektkomponenten und darauf aufbauend, die Strategiekomponenten mutiert.

3.2-3 $(\mu/\mu, \lambda)$ -KSA-ES

Die $(\mu/\mu, \lambda)$ -KSA-ES unterscheidet sich von der $(\mu/1, \lambda)$ -KSA-ES in erster Linie durch die Einführung der intermediären Rekombination sowie den daraus resultierenden Veränderungen der Operatoren. Ein Individuum $\mathbf{I}_i^{(t)}$ hat auch hier den Aufbau¹:

$$\mathbf{I}_i^{(t)} = (\mathbf{x}_i^{(t)}, \mathbf{s}^{(t)}, \delta^{(t)}) \quad (3.11)$$

mit der Objektkomponente $\mathbf{x}_i^{(t)} \in \mathbb{R}^n$, dem Evolutionspfad $\mathbf{s}^{(t)} \in \mathbb{R}^n$ und der Schrittweite $\delta^{(t)} \in \mathbb{R}$. Die Objektkomponente der Nachkommen wird mutativ durch folgende Vorschrift erzeugt:

$$\mathbf{x}_i^{(t+1)} = \langle \mathbf{x} \rangle_{\mu}^{(t)} + \delta^{(t)} \mathbf{z}_i \quad (3.12)$$

Für $\langle \mathbf{x} \rangle_{\mu}^{(t)}$ gilt:

$$\langle \mathbf{x} \rangle_{\mu}^{(t)} = \frac{1}{\mu} \sum_{i \in \mathbf{I}_{sel}^{(t)}} \mathbf{x}_i^{(t)} \quad (3.13)$$

mit der Indexmenge der in Generation t selektierten Individuen $\mathbf{I}_{sel}^{(t)}$. $\langle \mathbf{x} \rangle_{\mu}^{(t)}$ stellt damit den Schwerpunkt der in Generation t selektierten Individuen dar. $\delta^{(t)}$ beschreibt die Schrittweite der Generation t . Analog zur $(\mu/1, \lambda)$ -KSA-ES wird in einem zweiten Schritt die Schrittweite der Generation $t+1$ auf Basis der Berechnung des Evolutionspfades bestimmt. Für den Evolutionspfad $\mathbf{s}_i^{(t+1)}$ eines Nachkommens gilt:

$$\mathbf{s}_i^{(t+1)} = (1-c)\mathbf{s}_i^{(t)} + c_u \frac{\sqrt{\mu}}{\delta^{(t)}} (\langle \mathbf{x} \rangle_{\mu}^{(t+1)} - \langle \mathbf{x} \rangle_{\mu}^{(t)}) \quad (3.14)$$

Auch hier gilt $c_u = \sqrt{c(2-c)}$. Analog zu (3.7) bestimmt der Parameter c mit $0 < c \leq 1$ den Kumulationszeitraum. In seiner Dissertation zeigt Hansen, dass der Term

$$\frac{\sqrt{\mu}}{\delta^{(t)}} (\langle \mathbf{x} \rangle_{\mu}^{(t+1)} - \langle \mathbf{x} \rangle_{\mu}^{(t)}) \quad (3.15)$$

für zufällige Selektion $(\mathbf{0}, \mathbf{1})$ -normalverteilt ist.

1. Pro Generation existiert eine Schrittweite sowie ein Evolutionspfad. Die Strategieparameter $\mathbf{s}^{(t)}$ und $\delta^{(t)}$ müssen daher streng genommen nicht in jedem Individuum gespeichert werden, sondern nur in einer generationsbegleitenden Strategiekomponente.

Nachdem der Evolutionspfad $\mathbf{s}_i^{(t+1)}$ bestimmt wurde, kann die Schrittweite $\delta^{(t+1)}$ der Generation $t+1$ berechnet werden:

$$\delta^{(t+1)} = \delta^{(t)} e^{\frac{\|\mathbf{s}^{(t+1)}\| - \hat{\chi}_n}{D \cdot \chi_n}} \quad (3.16)$$

Für den Dämpfungsparameter D und den Wert $\hat{\chi}_n$ gelten die Bemerkungen zur Gleichung (3.9).

3.2-4 Eigenschaften der $(\mu/1, \lambda)$ - und $(\mu/\mu, \lambda)$ -KSA-ES

Durch Einsatz der kumulativen Schrittweitenadaptation ist es möglich, eine Degenerierung der Schrittweite zu verhindern. Diese Degenerierung findet bei konventioneller Schrittweitenadaptation (siehe Abschnitt 2.4-4 (Mutation der Strategiekomponenten – Selbstadaptation, Seite 29)) häufig statt und lässt sich in an einer systematisch zu kleinen Schrittweite festmachen. Für das Kugelmodell zeigt Ostermeier, dass die durch Kumulation gewonnenen Schrittweiten optimal sind [Ost97]. Alle Varianten der kumulativen Schrittweitenadaptation besitzen zwei Steuervariablen c und D , die die Lebensdauer der kumulierten Informationen sowie die Dämpfung der Schrittweitenadaptation festlegen. Für die Belegung dieser Parameter können Ostermeier und Hansen nur empirische Aussagen treffen. Alle Varianten der KSA-ES sind lediglich auf ihr Verhalten bei unimodale Testfunktionen untersucht worden. Es existieren keinerlei Angaben, ob die Varianten der KSA-ES auch für praxis-relevante Probleme eingesetzt werden können oder bereits wurden.

3.3 Kovarianzmatrixadaptation (CMA)

Die Kovarianzmatrixadaptation basiert auf einer Speicherung der erfolgreichen Mutation und der Einflussnahme der gespeicherten Informationen auf die Verteilungsfunktion, mit der neue Individuen erzeugt werden. Die erfolgreichen Mutationen sind genau die, die zu einem selektierten Individuum geführt haben. Der Einfluss der gespeicherten Mutationen wird mit dem Alter, d. h. mit der Anzahl der bereits erzeugten Nachfolgenerationen, gedämpft, sodass der Einfluss zurückliegender Mutationen mit Fortschreiten der Optimierung geringer wird. Auf diese Weise kann sich die Verteilungsfunktion der lokalen Topologie der Zielfunktion anpassen, indem bevorzugt Mutationen erzeugt werden, die den in der jüngeren Vergangenheit erfolgreichen Mutationen ähneln [Han98].

3.3-1 Grundidee

Die Realisierung dieses Verfahrens basiert auf einer speziellen Methode, n -dimensionale normalverteilte Zufallszahlen zu erzeugen. Diese Methode wird in [Han98] ausführlich diskutiert. Hier werden nur das Verfahren sowie der Einsatz in der CMA beschrieben. Die Vektoren $\mathbf{x}_1, \dots, \mathbf{x}_m \in \mathbb{R}^n$, mit $m \geq n$ seien so gewählt, dass sie den \mathbb{R}^n aufspannen. Die Zufallsvariablen $\mathbf{X}_1, \dots, \mathbf{X}_m$ seien unabhängig normalverteilte Zufallsvariablen mit Erwartungswert 0 und Standardabweichung 1. Dann ist der Vektor

$$\mathbf{x} = X_1 \mathbf{x}_1 + \dots + X_m \mathbf{x}_m \quad (3.17)$$

ebenfalls ein normalverteilter Zufallsvektor.

Initial werden die Vektoren \mathbf{x}_i so gewählt, dass sie eine Orthonormalbasis bilden, d. h.

$$\mathbf{x}_i^{(0)} = \mathbf{e}_i, \text{ für } 1 \leq i \leq n \quad (3.18)$$

Im folgenden bezeichnet $\mathbf{x}_{sel}^{(t)}$ den Vektor, der in der Generation t das Individuum erzeugt hat, das selektiert wurde. Dann kann ein neuer Mutationsvektor durch die Vorschrift

$$\mathbf{x}^{(t+1)} = \sum_{i=1}^n X_i q^t \mathbf{x}_i^{(0)} + \sum_{i=1}^t X_{(n+i)} q^{t-i} \mathbf{x}_{sel}^{(i)} \quad (3.19)$$

erzeugt werden. Für den Dämpfungsparameter q gilt: $0 < q < 1$. Das folgende Beispiel stammt aus der Dissertation von Hansen und zeigt für einen 2-dimensionalen Fall die Konstruktion der Mutationsverteilung nach vier Generationen. Die Einheitsvektoren sind rot, die jeweils selektierten Mutationsvektoren sind gelb gekennzeichnet.

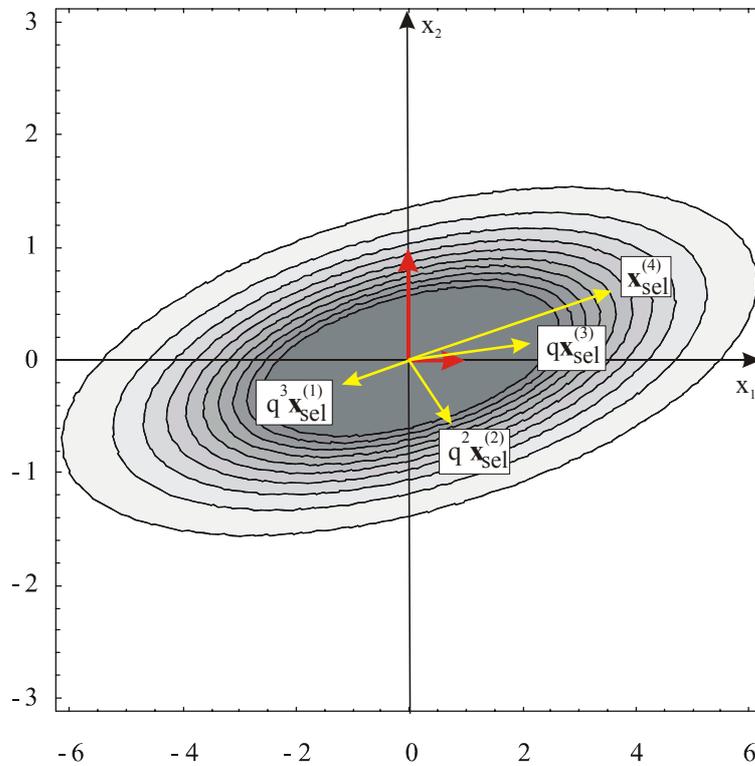


Abb. 3.3: Konstruktion der Mutationsverteilung für Generation vier

3.3-2 $(\mu/\lambda, \lambda)$ -CMA-ES

Dieser Abschnitt beschreibt den Algorithmus einer $(\mu/\lambda, \lambda)$ mit Kovarianzmatrixadaptation, wie sie von Hansen in [Han98] eingeführt wurde. Die Mutation erfolgt in mehreren Schritten. Zu Beginn wird für jeden Nachkommen eine neue Objektkomponente $\mathbf{x}_i^{(t+1)}$ durch

$$\mathbf{x}_i^{(t+1)} = \langle \mathbf{x} \rangle_{\mu}^{(t)} + \delta^{(t)} \mathbf{B}^{(t)} \mathbf{D}^{(t)} \mathbf{z}_k \quad (3.20)$$

bestimmt. Für die Symbole und Variablen gelten die Bemerkungen zu 3.2-2 ($(\mu/1, \lambda)$ -KSA-ES, Seite 38) und 3.2-3 ($(\mu/1\mu, \lambda)$ -KSA-ES, Seite 40). Die neue Variable $\mathbf{B}^{(t)}$ ist eine orthogonale $n \times n$ -Matrix, die das achsenparallele Verteilungsellipsoid $\mathbf{D}^{(t)} \mathbf{z}_k^{(t)}$ umorientiert. Die normierten Eigenvektoren der Kovarianzmatrix $\mathbf{C}^{(t)}$ stellen die Spalten von $\mathbf{B}^{(t)}$. Die Variable $\mathbf{D}^{(t)}$ ist eine $n \times n$ -Diagonalmatrix. Die Diagonalelemente $D_{ii}^{(t)}$ sind die Quadratwurzeln der Eigenwerte der Kovarianzmatrix $\mathbf{C}^{(t)}$. Die i -te Spalte von $\mathbf{B}^{(t)}$ ist ein mit $D_{ii}^{(t)}$ korrespondierender Eigenvektor. Die Matrizen $\mathbf{B}^{(t)}$ und $\mathbf{D}^{(t)}$ sind durch die Kovarianzmatrix $\mathbf{C}^{(t)}$ bis auf Spaltenvertauschung und Vorzeichen bestimmt. $\mathbf{C}^{(t+1)}$ wird mit Hilfe des Evolutionspfades $\mathbf{s}^{(t+1)}$ und der Kovarianzmatrix $\mathbf{C}^{(t)}$ bestimmt:

$$\mathbf{s}^{(t+1)} = (1 - c) \mathbf{s}^{(t)} + c_u \frac{\sqrt{\mu}}{\delta^{(t)}} (\langle \mathbf{x} \rangle_{\mu}^{(t+1)} - \langle \mathbf{x} \rangle_{\mu}^{(t)}) \quad (3.21)$$

$$\mathbf{C}^{(t+1)} = (1 - c_{cov}) \mathbf{C}^{(t)} + c_{cov} \mathbf{s}^{(t+1)} (\mathbf{s}^{(t+1)})^T \quad (3.22)$$

Der Vektor $\mathbf{s}^{(t+1)}$ ist auch hier der durch Kumulation erzeugte Evolutionspfad. Es gilt $\mathbf{s}^{(0)} = \mathbf{0}$. Der Parameter c_{cov} steuert die Mittelungszeit für die Adaptation der Kovarianzmatrix. Er wird zu $c_{cov} = 2/(n^2 + n)$ festgesetzt. Es gilt somit $0 < c_{cov} \leq 1$. Die Variable $\delta^{(t)}$ dient der kumulativen Regelung der Schrittweite. Der Summationsvektor $\mathbf{s}_{\delta}^{(t+1)}$ wird in einem Koordinatensystem berechnet, das nicht durch $\mathbf{D}^{(t)}$ skaliert wurde. Ostermeier und Hansen sichern so die Möglichkeit, die Richtungsinformation getrennt von der Skalierungsinformation auswerten zu können. Für den Summationsvektor $\mathbf{s}_{\delta}^{(t+1)}$ gilt:

$$\mathbf{s}_{\delta}^{(t+1)} = (1 - c_{\delta}) \mathbf{s}_{\delta}^{(t)} + c_{\delta_u} \mathbf{B}^{(t)} (\mathbf{D}^{(t)})^{-1} (\mathbf{B}^{(t)})^{-1} \frac{\sqrt{\mu}}{\delta^{(t)}} (\langle \mathbf{x} \rangle_{\mu}^{(t+1)} - \langle \mathbf{x} \rangle_{\mu}^{(t)}) \quad (3.23)$$

Für die Schrittweite $\delta^{(t+1)}$ gilt dann:

$$\delta^{(t+1)} = \delta^{(t)} e^{\frac{\|\mathbf{s}_{\delta}^{(t+1)}\| - \hat{\chi}_n}{D \cdot \chi_n}} \quad (3.24)$$

Der initiale Evolutionspfad $\mathbf{s}_{\delta}^{(0)}$ der Generation $t = 0$ wird auf den Startwert 0 gesetzt. Der Steuerparameter c_{δ} bestimmt den Kumulationszeitraum für $\mathbf{s}_{\delta}^{(t)}$, analog zu dem Parameter c der Gleichung (3.23). Es gilt auch hier: $0 < c_{\delta} \leq 1$. Für c_{δ_u} gilt:

$$c_{\delta_u} = \sqrt{c_{\delta}(2 - c_{\delta})} \quad (3.25)$$

Es gilt $\mathbf{B}^{-1} = \mathbf{B}^T$. Die Matrix \mathbf{D}^{-1} wird durch Invertierung der Diagonalelemente der Matrix \mathbf{D} bestimmt: $(D_{ii}^{-1})^{-1} = D_{ii}^{-1}$. Für alle weiteren Parameter gelten die Bemerkungen zu 3.2-2 ($(\mu/1, \lambda)$ -KSA-ES, Seite 38) und 3.2-3 ($(\mu/1\mu, \lambda)$ -KSA-ES, Seite 40).

3.3-3 Eigenschaften der $(\mu/1\mu, \lambda)$ -CMA-ES

Generell verfügt eine CMA-ES über mehrere Steuerparameter: c_{cov} steuert die Mittelungszeit für die Adaptation der Kovarianzmatrix. Der Parameter c_{δ} bestimmt den Kumulationszeitraum für $\mathbf{s}_{\delta}^{(t)}$.

Zusätzlich existiert ein Dämpfungsparameter D , der Einfluss auf die Schrittweite $\delta^{(t+1)}$ nimmt. Ostermeier gibt allgemeine Empfehlungen für die Belegung dieser Steuerparameter [Ost97]. Die erfolgreiche Einsatzmöglichkeit der CMA-ES wird von Ostermeier am Beispiel unimodaler Testfunktionen gezeigt [Ost97]. Auch für die CMA-ES gilt, dass sie lediglich auf ihr Verhalten bei unimodalen Testfunktionen untersucht wurde. Es existieren keinerlei Angaben, ob die Varianten der CMA-ES auch für praxis-relevante Probleme eingesetzt werden können oder bereits wurden.

3.4 Differentielle Evolution (DE)

Bei der Differentiellen Evolution handelt es sich um ein Optimierverfahren, das 1995 von Storn und Price entwickelt wurde [Sto95]. Die DE wurde mit der Maßgabe entwickelt, über ein Optimierverfahren zu verfügen, das reellwertige Zielfunktionen lösen kann und über eine möglichst geringe Anzahl von externen, d. h. vom Anwender einzustellenden Steuerparameter verfügt. Als Konsequenz verwendet die DE nur einen Skalierungsparameter F sowie einen Rekombinationsparameter c_r . DE basiert auf Populationen von Individuen, die mögliche Lösungen im Suchraum darstellen. Die Veränderung dieser Individuen findet bei dem Übergang einer Generation t zur Generation $t+1$ statt. Somit ähnelt die DE den Evolutionsstrategien. Im Gegensatz zur Evolutionsstrategie besteht ein Individuum \mathbf{I} jedoch nur aus der Objektkomponente:

$$\mathbf{I} = (x_1, \dots, x_n)^T \quad (3.26)$$

Die Veränderung eines Individuums basiert auf der Addition (hier als Rekombination bezeichnet) von Differenzvektoren, die dem Verfahren den Namen gegeben hat sowie der Durchführung von Tastschritten.

3.4-1 Tastschritt

Für jedes Individuum $\mathbf{I}_i^{(t)} = (x_{i,1}, \dots, x_{i,n})^T$ der Generation t wird ein Tastindividuum $\tilde{\mathbf{I}}$ wie folgt erzeugt:

$$\tilde{\mathbf{I}}^{(t)} = \tilde{\mathbf{x}}^{(t)} = \mathbf{x}_{r_1}^{(t)} + F(\mathbf{x}_{r_2}^{(t)} - \mathbf{x}_{r_3}^{(t)}) \quad (3.27)$$

$\mathbf{x}_{r_1}^{(t)}$, $\mathbf{x}_{r_2}^{(t)}$ und $\mathbf{x}_{r_3}^{(t)}$ sind hier die Objektkomponenten von drei unterschiedlichen Individuen $\mathbf{I}_{r_1}^{(t)}$, $\mathbf{I}_{r_2}^{(t)}$ und $\mathbf{I}_{r_3}^{(t)}$ einer Generation t , die gleichverteilt aus der Menge der Individuen der Generation t bestimmt werden. Die Zufallsvariablen r_1 , r_2 und r_3 werden in jedem Tastschritt neu bestimmt.

Die Abbildung 3.4 verdeutlicht die Generierung des Vektors $\tilde{\mathbf{I}}^{(t)}$ an einem 2-dimensionalen Beispiel.

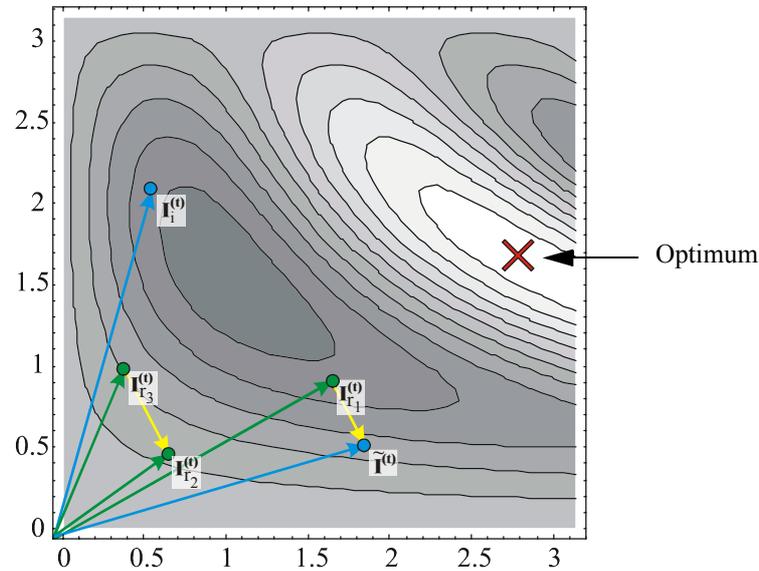


Abb. 3.4: Tastschritt der Differentiellen Evolution

3.4-2 Rekombination

In einem anschließenden Schritt werden die Koordinaten der neu gebildeten Objektkomponente $\tilde{\mathbf{x}}^{(t)}$ rekombiniert. Hierzu wird ein Vektor $\mathbf{u}^{(t)}$ bestimmt, mit

$$\mathbf{u}^{(t)} = (u_1^{(t)}, \dots, u_n^{(t)})^T \quad (3.28)$$

und

$$u_i^{(t)} = \begin{cases} \tilde{x}_i^{(t)} & \text{für } i = \langle r \rangle_n, \langle r+1 \rangle_n, \dots, \langle r+L-1 \rangle_n \\ x_i^{(t)} & \text{sonst} \end{cases} \quad (3.29)$$

Der Operator $\langle \cdot \rangle_n$ bezeichnet hier die Modulo-Funktion mit Modulus n . Der Startindex r wird gleichverteilt aus dem Intervall $[0, n-1]$ bestimmt, die ganze Zahl L aus dem Intervall $[0, n-1]$ wird zufällig bestimmt, mit sinkender Wahrscheinlichkeit für steigende L . Die Steuerung dieser Wahrscheinlichkeit erfolgt mit dem Parameter $c_r \in [0, 1] \subset \mathbb{R}$. Sollte die Zielfunktion an der Stelle $\mathbf{u}^{(t)}$ einen kleineren Funktionswert als das Individuum $\mathbf{I}_i^{(t)}$ besitzen, so wird $\mathbf{I}_i^{(t)}$ durch $\mathbf{u}^{(t)}$ ersetzt. Die Werte für n und L werden für jedes Individuum neu bestimmt. Hat das Individuum $\tilde{\mathbf{I}}^{(t)}$ eine höhere Fitness als das Individuum $\mathbf{I}_i^{(t)}$, so wird es in die nachfolgende Generation $t+1$ aufgenommen. Im anderen Fall überlebt das Individuum $\mathbf{I}_i^{(t)}$. Die Abbildung 3.5 verdeutlicht die Rekombina-

tion an einem Beispiel. Hierbei ist die Dimensionalität des Optimierungsproblems $n = 7$, für den Startindex gilt $r = 2$, die Länge des Rekombinationsbereichs L ist hier als $L = 3$ gewählt.

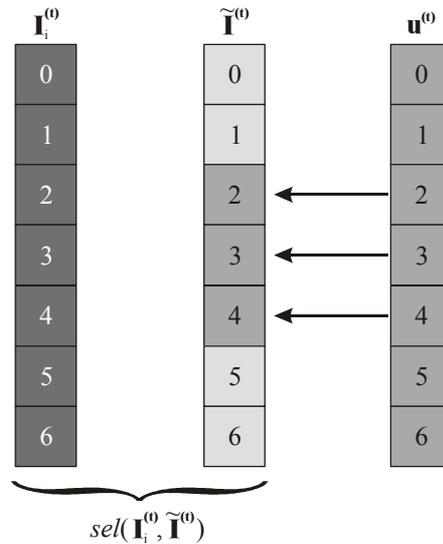


Abb. 3.5: Rekombination der Differentiellen Evolution

3.4-3 Lokale Differentielle Evolution

Storn und Price geben eine Erweiterung der Differentiellen Evolution an, bei der durch Einführung eines zusätzlichen Steuerparameters λ und der Berücksichtigung des aktuell besten Individuums $\mathbf{I}^{*(t)}$ der Generation t die Generierung neuer Nachkommen auf die Position von $\mathbf{I}^{*(t)}$ fokussiert werden kann [Sto95]. Der Tastschritt nach (3.27) wird wie folgt abgewandelt:

$$\tilde{\mathbf{I}}^{(t)} = \tilde{\mathbf{x}}^{(t)} = \mathbf{x}_i^{(t)} + \lambda(\mathbf{x}^{*(t)} - \mathbf{x}_i^{(t)}) + F(\mathbf{x}_{r_2}^{(t)} - \mathbf{x}_{r_3}^{(t)}), \text{ mit } 0 < \lambda < \infty \quad (3.30)$$

3.4-4 Eigenschaften der DE

Die Differentielle Evolution ist erfolgreich auf Problemstellungen aus den Bereichen Filterentwicklung und Robotik angewendet worden [Sto96]. Parameter, die im Vorfeld der Optimierung eingestellt werden müssen, sind der Skalierungsfaktor F und die Rekombinationsrate c_r . Zusätzlich werden die Größe der Startpopulation sowie eine maximale Lebensspanne der einzelnen Individuen als kritisch herausgearbeitet [Sto96]. Insbesondere von der Wahl der maximalen Lebensspanne hängt eine mögliche Konvergenz in lokalen Minima ab. Empfehlungen für eine gute Wahl dieser Parameter werden in der Literatur nicht gegeben.

3.5 Evolutionary Gradient Search (EGS)

Die *Evolutionary Gradient Search* ist ein Optimierverfahren, das versucht, den Gradienten der Zielfunktion durch eine Vielzahl von Tastschritten zu schätzen [Sal98a]. Sie ähnelt den Evolutionsstrategien insoweit, dass sie ausgehend von dem Elternindividuum $\mathbf{I}^{(t)}$ der Generation t λ Nachkommen bestimmt, populationsbasiert ist und Ansätze zur Steuerung der Schrittweiten enthält.

3.5-1 Tastschritte

Die Nachkommen, die bei diesen Tastschritten erzeugt werden, werden wie folgt berechnet:

$$\mathbf{I}_i^{(t)} = \mathbf{I}^{(t)} + \mathbf{z}_i, \text{ mit } 1 \leq i \leq \lambda \quad (3.31)$$

Bei den \mathbf{z}_i handelt es sich um Zufallsvektoren, jede Koordinate ist normalverteilt mit Erwartungswert 0 und normierter Standardabweichung $\sigma^{(t)}/\sqrt{n}$. Für $n \gg 1$ sind die $\mathbf{I}_i^{(t)}$ annähernd auf einer Hyperkugel mit dem Radius $\sigma^{(t)}$ verteilt. Zur Schätzung des Gradienten wird der Wert $\tilde{\mathbf{g}}^{(t)}$ wie folgt bestimmt:

$$\tilde{\mathbf{g}}^{(t)} = \sum_{i=1}^{\lambda} (f(\mathbf{I}_i^{(t)}) - f(\mathbf{I}^{(t)})) \cdot (\mathbf{I}_i^{(t)} - \mathbf{I}^{(t)}) \quad (3.32)$$

Hieraus resultiert dann der Einheitsvektor

$$\mathbf{e}^{(t)} = \frac{\tilde{\mathbf{g}}^{(t)}}{\|\tilde{\mathbf{g}}^{(t)}\|} \quad (3.33)$$

Der Iterationsschritt, der zu der Position des neuen Elternindividuums $\mathbf{I}^{(t+1)}$ führt, erfolgt im einfachsten Fall wie folgt:

$$\mathbf{I}^{(t+1)} = \mathbf{I}^{(t)} - \sigma^{(t+1)} \mathbf{e}^{(t)} \quad (3.34)$$

mit

$$\sigma^{(t+1)} = \begin{cases} \sigma^{(t)}\zeta & \text{für } f(\mathbf{I}^{(t)} - \sigma^{(t)}\zeta \mathbf{e}^{(t)}) \leq f(\mathbf{I}^{(t)} - (\sigma^{(t)}/\zeta) \mathbf{e}^{(t)}) \\ \sigma^{(t)}/\zeta & \text{sonst} \end{cases} \quad (3.35)$$

Der Wert für ζ wird von Salomon als nicht kritisch betrachtet, er schlägt $\zeta = 1.8$ vor [Sal98a]. Die Abbildung 3.6 veranschaulicht dieses Vorgehen graphisch.

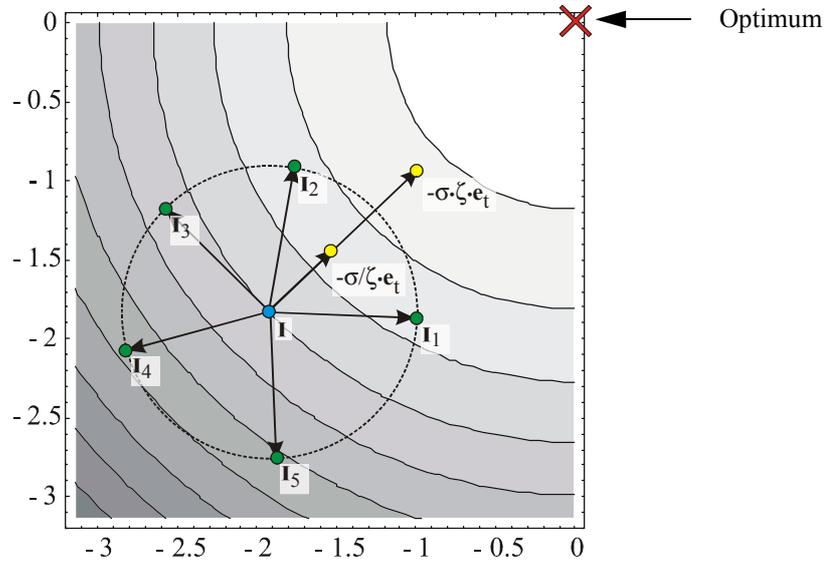


Abb. 3.6: Schätzung des Gradienten in der Evolutionary-Gradient-Search

3.5-2 Individuelle Steuerung der Schrittweiten

Die im vorhergehenden Abschnitt erläuterte EGS versagt bei Problemen mit stark unterschiedlich skalierten Koordinatenachsen. Um hier Abhilfe zu schaffen, führt Salomon in [Sal98b] eine individuelle Schrittweitensteuerung ein. Zusätzlich zu der innerhalb einer Generation t global gültigen Schrittweite $\sigma^{(t)}$ existiert nun ein Schrittweitenvektor $\boldsymbol{\eta}$, der Einfluss auf die Generierung der λ Nachkommen nimmt. Die Nachkommen werden wie folgt erzeugt:

$$I_{j,i}^{(t)} = I_i^{(t)} + \eta_i z_i, \text{ mit } 1 \leq j \leq \lambda \text{ und } 1 \leq i \leq n \quad (3.36)$$

Der Term $I_{j,i}^{(t)}$ bezeichnet hierbei die i -te Koordinate des j -ten Nachkommen der Generation t , während der Term $I_i^{(t)}$ die i -te Koordinate des Elternindividuums der Generation t bezeichnet. Die Adaptation des Schrittweitenvektors $\boldsymbol{\eta}$ findet komponentenweise wie folgt statt:

$$\eta_i = \begin{cases} \eta_i \beta & \text{für } \mathbf{e}^{(t)T} \mathbf{e}^{(t-1)} > 0 \\ \eta_i & \text{für } \mathbf{e}^{(t)T} \mathbf{e}^{(t-1)} = 0 \\ \eta_i / \beta & \text{für } \mathbf{e}^{(t)T} \mathbf{e}^{(t-1)} < 0 \end{cases} \quad (3.37)$$

Für den Parameter β schlägt Salomon Werte nahe 1 vor, um die Adaptation der Schrittweite $\sigma^{(t)}$ nicht störend zu beeinflussen. Gleichzeitig merkt er jedoch auch an, dass für $\beta = 1$ die individuelle Schrittweitensteuerung keinen Einfluss auf die Optimierung hat [Sal98]. Somit existiert keine begründbare Empfehlung für eine geeignete Wahl des Parameters β .

3.5-3 Eigenschaften der EGS

Die *Evolutionary Gradient Search* benötigt eine geringe Anzahl an Iterationen, wenn die Koordinatenachsen keine allzu unterschiedliche Skalierung aufweisen. In [Sal98] werden Testergebnisse für die Testfunktion Zigarre angegeben (siehe auch Abschnitt 7.2-6 (B7 - Zigarre, Seite 203)). Sollten die Koordinatenachsen stark unterschiedlich skaliert sein, soll die individuelle Schrittweitensteuerung eine Lösung des Optimierproblems ermöglichen. Problematisch scheinen hier die widersprüchlichen Empfehlungen für den Parameter β zu sein. Zusätzlich verfügt die EGS über einen externen Parameter ζ , der jedoch unkritisch belegt werden kann.

3.6 Robust Evolution Strategies (RES)

Ausgangspunkt für die Entwicklung der robusten Evolutionsstrategie RES von Ohkura, Matsumura und Ueda [Ohk98] war die Beobachtung, dass das Versagen von Evolutionsstrategien in vielen Fällen auf die Degeneration der Schrittweiten zurückzuführen ist. In solchen Fällen konvergiert die Schrittweite gegen 0 und lässt somit keine Veränderung der Individuen mehr zu. Häufig wird in solchen Fällen eine untere Schranke für die Schrittweite vorgegeben, sodass eine Mindestschrittweite erhalten bleibt. Die Wahl dieser unteren Schranke ist jedoch problemabhängig und daher für jedes Problem experimentell neu zu ermitteln. Zur Lösung dieser Problematik schlagen Ohkura, Matsumura und Ueda zwei Veränderungen der klassischen Evolutionsstrategien vor: Zum einen soll durch die Wahl eines nicht auf der Normalverteilung basierenden Mutationsoperators die Wahrscheinlichkeit für grosse Veränderungen dauerhaft erhöht werden. Zum anderen soll eine neue Art der Adaptation das Auftreten zu geringer Schrittweiten verhindern.

3.6-1 Cauchy Mutation

Ohkura, Matsumura und Ueda folgen bei der Wahl der Verteilungsfunktion für den Mutationsoperator einer Empfehlung von Yao und Liu, die vorschlagen, den auf Normalverteilungen basierenden Mutationsoperator durch einen Mutationsoperator zu ersetzen, der statt dessen Cauchy-verteilte Zufallsvariablen verwendet [Yao97]. Die so entstehende Variante der Evolutionsstrategien werden als *Fast Evolution Strategies* (FES) oder als *Fast Evolutionary Programming* (FEP) bezeichnet. Von Yao und Liu wurde hier insbesondere die Eignung einer (μ, λ) -Evolutionsstrategie mit Cauchy-Mutation untersucht. Im Gegensatz zur Untersuchung von Kappler (3.1 (Variation der Mutationsfunktion, Seite 35)) konnte in diesem Fall eine deutliche Verbesserung der Konvergenzgeschwindigkeit, zumindest für die untersuchten Testfunktionen, festgestellt werden. Eine vollständige Übersicht der Resultate findet sich in [Yao99]. Die konkrete Realisierung verwendet eine Variante der allgemeinen Cauchy-Verteilung

$$\Phi_t(x) = \frac{1}{2} + \frac{\arctan(x/t)}{\pi} \quad (3.38)$$

bei der der Parameter $t = 1$ gesetzt ist. Die Gleichung (3.38) reduziert sich damit zu

$$\Phi_1(x) = \frac{1}{2} + \frac{\arctan(x)}{\pi} \quad (3.39)$$

Die zugehörige Dichtefunktion lautet im allgemeinen:

$$\varphi_t(x) = \frac{1}{\pi t \left(1 + \left(\frac{x}{t} \right)^2 \right)} \quad (3.40)$$

bzw.

$$\varphi_1(x) = \frac{1}{\pi(1+x^2)} \quad (3.41)$$

Zum Vergleich stellt die Abbildung 3.7 den Verlauf der $N(0, 1)$ Normalverteilung (blau) sowie den Verlauf der Cauchy-Verteilung $\varphi_1(x)$ (rot) dar.

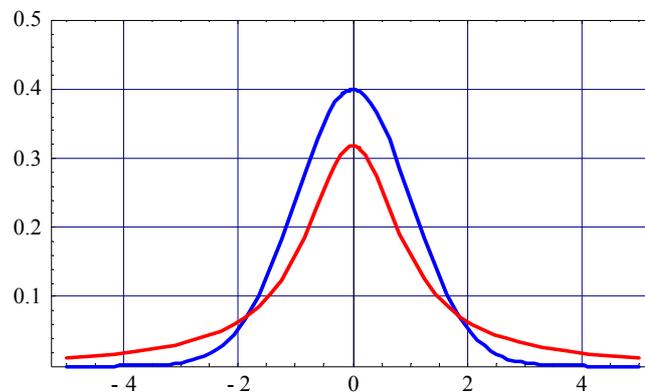


Abb. 3.7: Verlauf von Normalverteilung (blau) und Cauchy-Verteilung (rot)

3.6-2 Robuste Adaptation

Zur Vermeidung der Degeneration der Schrittweiten schlagen Ohkura, Matsumura und Ueda eine Änderung der Repräsentation vor. Ein Individuum $\mathbf{I}^{(t)}$ der Generation t wird wie folgt kodiert:

$$\mathbf{I}^{(t)} = (\mathbf{x}^{(t)}, \boldsymbol{\sigma}^{(t)}) \quad (3.42)$$

mit

$$\mathbf{x}^{(t)} = (x_1^{(t)}, \dots, x_n^{(t)})^T \quad (3.43)$$

und

$$\boldsymbol{\sigma}^{(t)} = \left\{ \left(\begin{array}{c} \sigma_{1,1}^{(t)} \\ \dots \\ \sigma_{1,m}^{(t)} \end{array} \right), \dots, \left(\begin{array}{c} \sigma_{n,1}^{(t)} \\ \dots \\ \sigma_{n,m}^{(t)} \end{array} \right) \right\} \quad (3.44)$$

Aus dieser Definition folgt, dass jede Koordinate der Objektkomponente $x_i^{(t)}$ m zugehörige Koordinaten in der Strategiekomponente besitzt. Diese mehrfach redundante Speicherung, in Verbindung

mit den folgenden Operationen, soll die Adaptation der Schrittweiten auf zu kleine Werte verhindern. Es existiert immer nur eine *aktive* Strategiekomponente $\sigma^{*(t)}$, die aus den Schrittweiten

$$\sigma_i^{*(t)} = (\sigma_{1,1}^{(t)}, \dots, \sigma_{n,1}^{(t)})^T, \text{ mit } 1 \leq i \leq n \quad (3.45)$$

besteht. Alle weiteren Strategiekomponenten werden als *inaktiv* bezeichnet. Nur die *aktive* Strategiekomponente wird für die Mutation der Objektkomponenten verwendet. Die Mutation erfolgt analog zu der Mutation, die in 2.4 ((μ +, λ)-Evolutionsstrategie, Seite 27) beschrieben ist. Die *inaktiven* Strategiekomponenten dienen zur Veränderung der *aktiven* Strategiekomponente, indem drei verschiedene Strategiemutationen angewendet werden. Aus Gründen der Übersichtlichkeit wird in den folgenden Gleichungen (3.47) - (3.49) auf die Mitführung des Generationenzählers t als Exponent verzichtet.

$$\begin{aligned} O_{dub}: \tilde{\sigma}_{i,1} &= \sigma_{i,1} \\ \tilde{\sigma}_{i,p} &= \sigma_{i,p-1} \forall p \in \{2, \dots, m\} \\ \sigma'_{i,p} &= mut_{\sigma}(\tilde{\sigma}_{i,p}) \forall p \in \{1, \dots, m\} \end{aligned} \quad (3.46)$$

$$\begin{aligned} O_{del}: \tilde{\sigma}_{i,p} &= \sigma_{i,p+1} \forall p \in \{1, \dots, m-1\} \\ \tilde{\sigma}_{i,m} &= \min \left(\sigma_{max}, \sum_{p=1}^{m-1} \tilde{\sigma}_{i,p} \right) \\ \sigma'_{i,p} &= mut_{\sigma}(\tilde{\sigma}_{i,p}) \forall p \in \{1, \dots, m\} \end{aligned} \quad (3.47)$$

$$\begin{aligned} O_{inv}: \tilde{\sigma}_{i,1} &= \sigma_{i,p} \exists p \in f\{2, \dots, d\} \\ \tilde{\sigma}_{i,p} &= \sigma_{i,1} \\ \sigma'_{i,p} &= mut_{\sigma}(\tilde{\sigma}_{i,p}) \forall p \in \{1, \dots, m\} \end{aligned} \quad (3.48)$$

Diese Operatoren bilden Duplikation, Deletation und Inversion der biologischen Evolution nach. Die Mutationsfunktion mut_{σ} lautet:

$$\sigma_i^{(t)} = \max(\varepsilon, \sigma_i^{(t)} \cdot e^{(\tau'z + \tau z_i)}) \quad (3.49)$$

Die Zufallsvariable z wird als $N(0, 1)$ normalverteilte Zufallsvariable für jedes Individuum einmal realisiert, die Zufallsvariable z_i , die ebenfalls $N(0, 1)$ verteilt ist, wird für jede Koordinate der Objektkomponente neu realisiert. Die Parameter τ und τ' werden analog zu den Definitionen von Schwefel verwendet, siehe 2.4 ((μ +, λ)-Evolutionsstrategie, Seite 27). Die kritischen Parameter σ_{max} und ε , die maximale bzw. minimale Schranken für die Operationen darstellen, müssen problemabhängig gewählt werden.

3.6-3 Eigenschaften der RES

Die *Robust Evolution Strategy* wird in [Ohk98] mit der *Fast Evolution Strategy* (FES) von [Yao97] verglichen. Die FES verwendet Cauchy-Mutation und eine untere Schranke für die Schrittweiten. Im Vergleich zeigt sich, dass die RES weniger Generationen benötigt, um das Optimum der Testfunktionen zu finden. Die Qualität einer RES hängt jedoch stark von der Wahl der Parameter σ_{max} und ε ab. Bei der Wahl ungeeigneter Werte für σ_{max} und ε findet die RES das Optimum der Testfunktionen nicht.

3.7 Lokale Selektion - Diffusionsmodelle (LES)

Der Selektionsmechanismus, der abhängig von der Fitness der Nachkommenpopulation die nachfolgende Elternpopulation bestimmt, ist bei den klassischen Versionen der Evolutionsstrategien deterministisch und global. In der Evolution ist die globale Selektion nicht typisch, da dort nur Individuen miteinander konkurrieren, die sich in einem lokal begrenzten Gebiet aufhalten. Zusätzlich wird die Auswahl der Individuen für die Rekombination lokal durchgeführt. Verfahren, die diese räumlichen Aspekte widerspiegeln werden, in der Literatur auch als Diffusionsmodelle bezeichnet, da einzelne Informationen der Individuen ähnlich einer Diffusion durch die Population wandern. Diese Idee der räumlich begrenzten Selektionen wird von Gorges-Schleuter aufgegriffen [Gor98].

Zur Beschreibung der lokalen Selektion verwendet sie folgende von Bäck vorgeschlagene Notation, die eine Erweiterung der Notation nach Definition 2.11 darstellt:

$$(\mu, \kappa, \lambda/\mu, \rho, \nu)\text{-LES} \quad (3.50)$$

mit

μ	:	$\mu \geq 1$	Anzahl der Eltern
κ	:	$\kappa \geq 1$	maximale Lebenserwartung
λ/μ	:	$\lambda/\mu \in \mathbb{N}$	Anzahl der Nachkommen, die um die Selektion in die Folgegeneration konkurrieren
ρ	:	$1 \leq \rho \leq \nu$	Anzahl der Eltern pro Nachkomme
ν	:	$\nu \geq 1$	Größe der Nachbarschaft

Die Menge $\nu_i \subseteq I^\nu$ beschreibt die Nachbarschaft eines Individuums $\mathbf{I}_i \in I^\mu$ und gibt damit die Menge der Individuen an, die für den Rekombinationsprozess herangezogen werden. In der Arbeit von Gorges-Schleuter wurden folgende Parameter festgehalten [Gor98]:

$$\kappa = 1 \text{ und } \rho = 2 \quad (3.51)$$

d. h. bei der Selektion handelt es sich um eine Komma-Selektion und aus jeweils zwei Elternindividuen wird ein Nachkomme erzeugt.

Zur Modellierung der lokalen Eigenschaften werden sowohl der Rekombinationsoperator, als auch der Selektionsoperator erweitert. Der Mutationsoperator wird in der Arbeit von Gorges-Schleuter nicht verändert. Es handelt sich um den Mutationsoperator nach 2.4 ((μ +, λ)-Evolutionsstrategie, Seite 27) für n Schrittweiten. Es findet keine Adaptation von Lagewinkeln statt.

3.7-1 Lokale Rekombination

Der Rekombinationsoperator rek , der lokale Eigenschaften modelliert ist wie folgt definiert:

$$rek^v = re \circ co_i, \text{ mit } 1 \leq i \leq \mu \quad (3.52)$$

Für die Operatoren re und co_i gilt:

$$re: I^\rho \rightarrow I \quad (3.53)$$

$$co_i: v_i \rightarrow I^\rho \quad (3.54)$$

Der Operator co_i wählt $1 \leq \rho \leq v$ Eltern aus der Nachbarschaft v_i des Individuums \mathbf{I}_i aus, mit denen der Operator re dann ein neues Individuum erzeugt. Die Auswahl der Eltern kann auf unterschiedliche Weisen erfolgen. In [Gor98] werden zwei Mechanismen vorgeschlagen: *local parent selection* und *centric parent selection*. Bei der *local parent selection* werden alle Eltern aus der Nachbarschaft gewählt, bei der *centric parent selection* wird ein Elternindividuum global gewählt, alle weiteren aus der Nachbarschaft.

3.7-2 Lokale Selektion

Der Selektionsoperator wählt aus der aktuellen Generation die Nachkommen aus. Bei den grundlegenden Varianten der Evolutionsstrategien werden entweder alle Nachkommen (Komma-Selektion) oder alle Eltern und alle Nachkommen (Plus-Selektion) einer Generation in den Selektionsprozess einbezogen. Um die Lokalität zu erreichen, wird in [Gor98] folgende Variante vorgeschlagen:

$$sel_i: I^{\lambda/\mu} \rightarrow I \quad (3.55)$$

Die Selektion wird μ mal unabhängig voneinander durchgeführt. Es werden jedoch nur λ/μ Individuen in die jeweiligen Selektionsoperationen einbezogen.

3.7-3 Eigenschaften der LES

Die in [Gor98] präsentierten experimentellen Resultate zeigen, dass sich eine LES bei der Optimierung anders verhält als eine Standard Evolutionsstrategie mit Adaptation von n Schrittweiten. Problematisch ist, dass in Abhängigkeit des Parameters v sowohl bessere als auch schlechtere Resultate bezüglich der Konvergenzgeschwindigkeit erzielt werden. Zusätzlich zu dem Parameter v existieren weitere Parameter, die Einfluss auf das Verhalten der LES nehmen. Hierzu zählen die Wahl des Selektionsmechanismus für die Rekombination, das Zulassen von zwei identischen Eltern sowie die Topologie der Population, die auf die Nachbarschaft Einfluss nimmt. Da in [Gor98] keine Richtlinien für die Wahl geeigneter Parameter angegeben werden, scheint die Wahl der optimalen Parameter nur durch Experimentieren zu finden zu sein.

3.8 Adaptation einer nicht-zentralen Normalverteilung

Zur Lösung dynamischer Optimierprobleme, bei denen das Optimum der Zielfunktion zeitabhängig ist, schlägt Weicker eine Abwandlung eines Verfahrens vor, das von Ghozeil und Fogel 1996 entwickelt wurde [Gho96, Wei01]. Dieses Verfahren ist speziell für dynamische Problemstellungen entwickelt worden. Eine Anwendung auf statische Problemstellungen ist noch nicht durchgeführt worden. Trotzdem soll dieses Verfahren hier kurz vorgestellt werden, da es einen gänzlichen anderen Ansatz zur Adaptation von Richtungen darstellt, als der in dieser Arbeit untersucht.

Weickers Idee ist es, anstelle einer $N(0, \sigma)$ -verteilten Zufallsvariablen eine $N(a_i, \sigma)$ -verteilte Zufallsvariable zu verwenden und den Lagevektor

$$\mathbf{a} = (a_1, \dots, a_n)^T \quad (3.56)$$

während der Optimierung zu adaptieren. Eine Richtungsadaptation wird also nicht durch Asymmetrie der Verteilung sondern durch deren Translation erreicht. Weicker verwendet hier n unabhängige Zufallsvariablen, wobei n die Dimensionalität des Optimierproblems bezeichnet. In einer im Vorfeld durchgeführten Untersuchung hat Weicker den Einsatz der 1/5-Erfolgsregel sowie die Kovarianzmatrixadaptation untersucht und festgestellt, dass sich diese Adaptationsvarianten nicht für dynamische Probleme eignen [Wei99].

3.8-1 Mutation

Die Mutation wird in drei Schritten durchgeführt. Im ersten Schritt wird die Schrittweite $\sigma_i^{(t)}$ des Individuums i der Generation t gemäß folgender Vorschrift mutiert:

$$\sigma_i^{(t+1)} = \sigma_i^{(t)} e^{(z/\sqrt{n})} \quad (3.57)$$

Die Variable z bezeichnet hier eine $N(0, 1)$ verteilte Zufallsvariable. z wird für jedes Individuum einer Generation neu realisiert. Im zweiten Schritt wird koordinatenweise ein neuer Lagevektor $\mathbf{a}_i^{(t+1)}$ bestimmt:

$$a_{i,j}^{(t+1)} = a_{i,j}^{(t)} + z_i' \quad (3.58)$$

Die Zufallsvariable z_i' bezeichnet eine $N(0, 1)$ -verteilte Zufallsvariable. Sie wird für jede Koordinate neu realisiert. Basierend auf der mutierten Schrittweite $\sigma_i^{(t+1)}$ und des mutierten Lagevektors $\mathbf{a}_i^{(t+1)}$ wird nun in dem abschließenden Schritt die Objektkomponente mutiert. Dies geschieht ebenfalls koordinatenweise laut der folgenden Vorschrift:

$$x_{i,j}^{(t+1)} = x_{i,j}^{(t)} + a_{i,j}^{(t+1)} + z_i'' \quad (3.59)$$

Bei der Zufallsvariablen z_i'' handelt es sich um eine $N(0, \sigma_i^{(t+1)})$ -verteilte Zufallsvariable, die für jede Koordinate der Objektkomponente neu realisiert wird. Von Weicker wurden keine Untersuchungen bezüglich lokaler und globaler Lernparameter der einzelnen Mutationsschritte vorgenommen.

3.8-2 Eigenschaften des Verfahrens

Hauptziel der Entwicklung dieses Verfahrens war es, eine Adaptationsvariante der Evolutionsstrategien zu entwickeln, mit der das Optimum dynamischer Zielfunktionen verfolgt werden kann. Als beispielhafte Zielfunktionen hat Weicker die Funktionen *moving circle* und *moving corridor* ausgewählt. Bei der Zielfunktion *moving circle* wandert eine kreisförmige Region durch den Suchraum. Die Fitness innerhalb dieser Region ist direkt abhängig vom Abstand zwischen der Position eines Individuums und dem Mittelpunkt des Kreises. Einem Individuum, das außerhalb der kreisförmigen Region liegt, wird eine konstant schlechte Fitness zugewiesen. Als einziger Parameter der Zielfunktion dient der Überlappungsindex k_{circle} . Dieser bestimmt den Abstand sukzessiver Wanderungen der kreisförmigen Region. Als zweite Zielfunktion hat Weicker die Funktion *moving corridor* ausgewählt, bei der sich ein Korridor durch den Suchraum bewegt. Für den Überlappungsindex $k_{corridor}$ gelten ähnliche Anmerkungen, wie zu dem Index k_{circle} . Eine vollständige Beschreibung der Zielfunktionen findet sich in [Wei01].

Die vollständige Auswertung der unterschiedlichen Varianten der Zielfunktionen ist an dieser Stelle von zweitrangiger Bedeutung, da der Schwerpunkt dieser Untersuchung auf den dynamischen Zielfunktionen lag. Als Zusammenfassung kann resümiert werden, dass für große Überlappungsindizes die Adaptation des Lagevektors \mathbf{a} der Adaptation einer Schrittweite und der Adaptation von n Schrittweiten überlegen ist. Bei einigen Testläufen musste jedoch ein Reparaturalgorithmus degenerierende Schrittweiten korrigieren. Wesentlich interessanter als diese Ergebnisse ist jedoch Weickers Untersuchung über das Verhältnis von erfolgreichen zu nicht erfolgreichen Testläufen. Hier zeigt es sich, dass mit zunehmender Generationenzahl die Evolutionsstrategie mit Adaptation des Lagevektors vermehrt ungültige, d. h. nicht in der wandernden Region befindliche, Nachkommen erzeugt. In seiner Untersuchung sind bei der Zielfunktion *moving corridor* bereits nach ca. 60 Generationen 50% der erzeugten Nachkommen ungültig. Nach 200 Generationen wurden ausschließlich ungültige Nachkommen erzeugt. Grundsätzlich zeigen Evolutionsstrategien mit Adaptation einer Schritte oder Adaptation von n Schrittweiten ein ähnliches, jedoch weit schwächer ausgeprägtes Verhalten. In dem Abschnitt 6.2-3 (Reversibilität, Seite 177) wird dieses Phänomen in Zusammenhang mit der asymmetrischen Evolutionsstrategie erneut diskutiert.

3.9 Simulated Binary Crossover (SBX)

Selbstadaptation als ein wichtiger Mechanismus zur Steuerung eines evolutionären Algorithmus ist Bestandteil der Evolutionsstrategien und des evolutionären Programmierens. Die Grundprinzipien für die Evolutionsstrategien sind in den Abschnitten 2.4-5 (Adaptation einer Schrittweite, Seite 29), 2.4-6 (Adaptation von n Schrittweiten, Seite 31) und 2.4-7 (Adaptation von n Schrittweiten und $(n(n-1))/2$ Lagewinkeln, Seite 32) beschrieben. Die Grundprinzipien für das evolutionäre Programmieren haben Fogel, Angeline und Fogel 1995 vorgestellt [Fog95]. Für die Genetischen Algorithmen existieren nur wenige Arbeiten, die selbstadaptives Verhalten einführen bzw. untersuchen [Bäc92]. Der folgende Abschnitt beschreibt selbst-adaptive Genetische Algorithmen nach Deb und Beyer, die ausschließlich Crossover¹, hier *Simulated Binary Crossover* (SBX), verwenden und vollständig auf einen Mutationsoperator verzichten [Deb99]. Dieses Verfahren stammt aus dem Bereich

1. Den Empfehlungen zur Verwendung der Fachbegriffe im Bereich der evolutionären Algorithmen zu Folge, soll der Terminus Rekombination nur für Operatoren verwendet werden, die einen Nachkommen erzeugen. Operatoren, die zwei oder mehr Nachkommen erzeugen sollen mit Crossover bezeichnet werden. Da der SBX-Operator zwei Nachkommen erzeugt wird hier die korrekte, jedoch englische Bezeichnung verwendet.

der reellwertig kodierten Genetischen Algorithmen, d. h. es findet keine Kodierung der Variablen in Form einer Bitkette statt. Statt dessen werden die Variablen direkt verwendet.

Es existieren eine Vielzahl unterschiedlicher Crossover Operatoren für reellwertig kodierte Genetische Algorithmen. Einer der ersten Operatoren aus diesem Bereich wurde von Wright entwickelt [Wri91]. Ausgehend von zwei Elternindividuen $x_a^{(t)} = (x_{a,1}^{(t)}, \dots, x_{a,n}^{(t)})^T$ und $x_b^{(t)} = (x_{b,1}^{(t)}, \dots, x_{b,n}^{(t)})^T$ werden drei neue Individuen durch koordinatenweise Linearkombination erzeugt:

$$x_{1,i}^{(t)} = x_{a,i}^{(t)} + x_{b,i}^{(t)}, \text{ für } 1 \leq i \leq n \quad (3.60)$$

$$x_{2,i}^{(t)} = 1.5x_{a,i}^{(t)} - 0.5x_{b,i}^{(t)}, \text{ für } 1 \leq i \leq n \quad (3.61)$$

$$x_{3,i}^{(t)} = -0.5x_{a,i}^{(t)} + 1.5x_{b,i}^{(t)}, \text{ für } 1 \leq i \leq n \quad (3.62)$$

Aus diesen drei Individuen werden die beiden besten als Nachkommen ausgewählt.

Eshelman und Schaffer haben eine intervallbasierte Erzeugung neuer Nachkommen vorgeschlagen [Esh93]. Ihr *Blend-Crossover Operator* (BLX- α) definiert in einem ersten Schritt ein Intervall V , das durch zwei Elternindividuen

$$x_a^{(t)} = (x_{a,1}^{(t)}, \dots, x_{a,n}^{(t)})^T \text{ und} \quad (3.63)$$

$$x_b^{(t)} = (x_{b,1}^{(t)}, \dots, x_{b,n}^{(t)})^T \quad (3.64)$$

definiert ist. Unter der Annahme, dass für alle Koordinaten die Bedingung $x_{a,i}^{(t)} < x_{b,i}^{(t)}$ für alle $1 \leq i \leq n$ erfüllt ist, gilt für das Intervall:

$$V = [x_{a,i}^{(t)} - \alpha(x_{b,i}^{(t)} - x_{a,i}^{(t)}), x_{b,i}^{(t)} + \alpha(x_{b,i}^{(t)} - x_{a,i}^{(t)})] \quad (3.65)$$

In dem zweiten Schritt wird für jede Koordinate eine gleichverteilte Zufallszahl $u_i = G(0, 1)$ erzeugt. Mittels der u_i wird ein neues Individuum $x_{1,i}^{(t+1)}$ nach folgender Vorschrift erzeugt:

$$x_{1,i}^{(t+1)} = (1 - \gamma_i)x_{a,i}^{(t)} + \gamma_i x_{b,i}^{(t)}, \text{ mit} \quad (3.66)$$

$$\gamma_i = (1 + 2\alpha)u_i - \alpha \quad (3.67)$$

Von besonderer Bedeutung ist die Tatsache, dass sich die Position des neuen Individuums im Suchraum an der Position der beiden Elternindividuen orientiert: Ist die Differenz der beiden Elternindividuen, bezogen auf den Suchraum, gering, ist auch die Differenz des neuen Individuums zu den Elternindividuen in Erwartung gering. Mit steigender Differenz der beiden Elternindividuen, steigt auch die Wahrscheinlichkeit, dass das neue Individuum eine Position entfernt von den Eltern hat. Die Ausbreitung der Nachkommen wird so durch die Ausbreitung der Eltern beeinflusst. Deb und Beyer weisen in [Deb99] auf die Bedeutung dieser Tatsache für Selbstadaptation hin.

3.9-1 Grundidee

Der SBX-Operator wurde 1995 von Deb, Agrawal und Kumar entwickelt [Deb95a, Deb95b]. Hauptaufgabe des SBX-Operators ist die Ausweitung des 1-Punkt Crossovers für Bit-Ketten auf reellwertig kodierte Genetische Algorithmen. Unter Berücksichtigung sog. expandierender und kontrahierender Crossovers bestimmen Deb und Agrawal die Wahrscheinlichkeitsdichte für die Änderung eines Nachkommen bezogen auf seine Eltern. Die so bestimmte Dichtefunktion, die in dieser Analyse nur für Bit-Ketten gültig ist, wird dann auf reellwertige Genetische Algorithmen übertragen [Deb95a]. Der so entstehende SBX-Operator ist wie folgt aufgebaut [Deb99]: Ausgehend von zwei Elternindividuen

$$x_a^{(t)} = (x_{a,1}^{(t)}, \dots, x_{a,n}^{(t)})^T \text{ und} \quad (3.68)$$

$$x_b^{(t)} = (x_{b,1}^{(t)}, \dots, x_{b,n}^{(t)})^T \quad (3.69)$$

werden zwei neue Individuen

$$x_a^{(t+1)} = (x_{a,1}^{(t+1)}, \dots, x_{a,n}^{(t+1)})^T \text{ und} \quad (3.70)$$

$$x_b^{(t+1)} = (x_{b,1}^{(t+1)}, \dots, x_{b,n}^{(t+1)})^T \quad (3.71)$$

erzeugt. In dem ersten Schritt wird der Ausbreitungsfaktor β_i bestimmt

$$\beta_i = \left| \frac{x_{b,i}^{(t+1)} - x_{a,i}^{(t+1)}}{x_{b,i}^{(t)} - x_{a,i}^{(t)}} \right| \quad (3.72)$$

Daran anschließend wird eine kontinuierliche Zufallszahl $u_i = G(0, 1)$ bestimmt. Die folgende Wahrscheinlichkeitsdichtefunktion (3.73) ist von Deb und Agrawal entwickelt worden und bildet das Suchverhalten des 1-Punkt-Crossovers der binären Genetischen Algorithmen nach:

$$\varphi(\beta) = \begin{cases} 0.5(\eta + 1)\beta^\eta & \text{für } \beta \leq 1 \\ 0.5(\eta + 1)\frac{1}{\beta^{\eta+1}} & \text{für } \beta > 1 \end{cases} \quad (3.73)$$

Der Parameter β_{q_i} wird so bestimmt, dass er dem Ordinatenwert entspricht, für den die Fläche zwischen 0 und β_{q_i} den Wert u_i annimmt. Für β_{q_i} gilt damit:

$$\beta_{q_i} = \begin{cases} (2u_i)^{\frac{1}{\eta+1}} & \text{für } u_i \leq 0.5 \\ \left(\frac{1}{2(1-u_i)}\right)^{\frac{1}{\eta+1}} & \text{für } u_i > 0.5 \end{cases} \quad (3.74)$$

Nachdem β_{q_i} berechnet wurde, können zwei neue Individuen nach (3.75) und (3.76) bestimmt werden:

$$x_{a,i}^{(t+1)} = 0.5[(1 + \beta_{q_i})x_{a,i}^{(t)} + (1 - \beta_{q_i})x_{b,i}^{(t)}] \quad (3.75)$$

$$x_{b,i}^{(t+1)} = 0.5[(1 - \beta_{q_i})x_{a,i}^{(t)} + (1 + \beta_{q_i})x_{b,i}^{(t)}] \quad (3.76)$$

Die folgende Abbildung 3.8 stellt für den 1-dimensionalen Fall die Wahrscheinlichkeitsdichte für die Position der Nachkommen dar. Die Positionen der Eltern sind in diesem Beispiel $x_{a,1}^{(t)} = 2$ und $x_{b,1}^{(t)} = 4$.

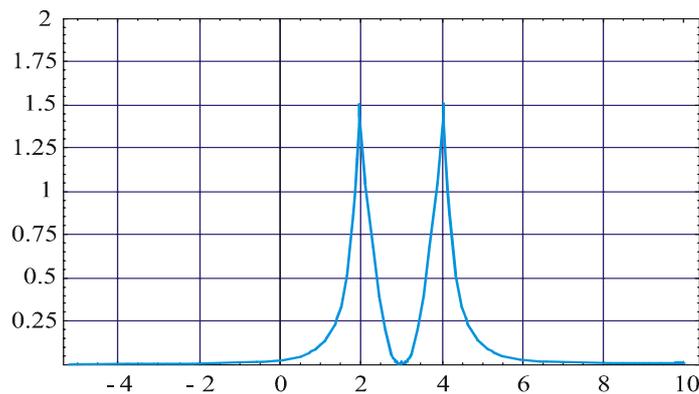


Abb. 3.8: Arbeitsweise des SBX-Operators

3.9-2 Eigenschaften des SBX

Deb und Beyer weisen darauf hin, dass der SBX-Operator in der Notation des BLX- α -Operators beschrieben werden kann, indem der Wert γ aus Gleichung (3.66) und (3.67) auf

$$\gamma = 0.5(1 \mp \beta_{q_i}) \quad (3.77)$$

gesetzt wird [Deb99]. Es ist jedoch zu beachten, dass die Zufallsvariable u in (3.66) gleichverteilt bestimmt wird, die Variable β_{q_i} jedoch durch (3.74) nicht gleichverteilt ist. Der SBX-Operator hat zwei weitere, wichtige Eigenschaften:

1. Die Entfernung der erzeugten Nachkommen ist proportional zum Abstand der Elternindividuen
2. Die Wahrscheinlichkeit für die Position eines neuen Individuums sinkt monoton mit der Entfernung zu den Elternindividuen

Der BLX- α -Operator besitzt die erste Eigenschaft, jedoch nicht die zweite. Deb und Beyer zeigen in [Bey00, Deb99], dass ein Genetischer Algorithmus, der über einen Rekombinationsoperator mit beiden Eigenschaften verfügt, ein ähnliches Optimierverhalten aufweist, verglichen mit einer Evolutionsstrategie mit Adaptation einer bzw. n Schrittweiten.

Deb und Beyer beschreiben in [Deb99] einige experimentelle Resultate. Bei den untersuchten Testfunktionen handelt es sich um Varianten des Kugelmodells, die Doppelsumme sowie eine multimodale Funktion. Alle Experimente wurden an 30-dimensionalen Zielfunktionen durchgeführt. Die Experimente mit dem Kugelmodell haben gezeigt, dass der SBX-Operator ein Verhalten aufweist, das dem Verringern der Schrittweite in Evolutionsstrategien ähnlich ist. Ein Genetischer Algorithmus mit dem SBX-Operator ist in der Lage, das Optimum in beliebiger Annäherung zu finden. Die Anzahl der Zielfunktionsauswertungen ist jedoch ein Vielfaches höher als bei selbstadaptiven Varianten der Evolutionsstrategien. Bei korrelierten Zielfunktionen konvergiert ein Genetischer Algorithmus mit SBX-Operator schneller als eine Evolutionsstrategie mit Adaptation von n Schrittweiten. Eine Evolutionsstrategie mit Adaptation von Schrittweiten und Lagewinkeln benötigt bei korrelierten Zielfunktionen, wie der Doppelsumme jedoch um Größenordnungen weniger Zielfunktionsauswertungen. Deutlich besseres Konvergenzverhalten, verglichen zu den unterschiedlichen Varianten der Evolutionsstrategien zeigt ein Genetischer Algorithmus mit SBX-Operator am Beispiel der multi-modalen Zielfunktion. Deb und Beyer weisen darauf hin, dass bei einer besseren Wahl der initialen Schrittweite auch eine Evolutionsstrategie ähnliches Konvergenzverhalten zeigt. Eine vollständige Übersicht über die von Deb und Beyer durchgeführten Experimente findet sich in [Deb99].

3.10 Estimation of Distribution Algorithms (EDA)

Die Idee der *Estimation of Distribution Algorithms* stammt, analog zu der Idee des *Simulated Binary Crossovers*, aus dem Bereich der Genetischen Algorithmen. Ihre ersten Anwendungen bezogen sich somit auf diskret repräsentierte Individuen, d. h. auf diskret repräsentierte mögliche Lösungen des Optimierproblems. EDA versuchen Abhängigkeiten zwischen den Variablen des Optimierproblems zu erkennen und explizit in die Generierung weiterer Lösungen einzubeziehen. Die Theorie der Genetischen Algorithmen beschäftigt sich zu einem großen Teil mit den sog. *building blocks*. *Building blocks* bezeichnen Teillösungen, die sich in der diskreten Repräsentation bereits entwickelt haben. Die Grundidee der Genetischen Algorithmen ist nun, dass die Anzahl der *building blocks* im Zuge der Optimierung zunimmt und aus mehreren *building blocks* größere Teillösungen entstehen. Durch die genetischen Operatoren Mutation und Crossover können gute *building blocks* jedoch auch zerstört werden, ein Phänomen, das in der Theorie der Genetischen Algorithmen *linkage problem* genannt wird. EDA versuchen dieses Problem zu umgehen, indem sie statistische Informationen der selektierten Individuen einer Generation auswerten, um aus diesen Informationen ein stochastisches Modell zu lernen und daraus möglichst vielversprechende neue Individuen zu generieren. EDA verzichten aus diesem Grund auf den Einsatz von Mutation und Crossover.

Die grundlegenden Arbeiten, die zu den EDA führten haben bereits Anfang der 90er Jahre begonnen. Syswerda hat mit dem *Bit Based Simulated Crossover* (BBSC) einen Genetischen Algorithmus eingeführt, der statistische Informationen einer Population verwendet [Sys93]. Ein anderer Ansatz, das sog. *Population Based Incremental Learning* (PBIL), wurde von Baluja gewählt, indem er als Alternative zu einer diskret repräsentierten Population einen reellwertigen Vektor einsetzt [Bal94, Kva96]. PBIL geht dabei von einer Unabhängigkeit der Variablen aus und aggregiert in dem Vektor die Position der besten Individuen. Die Selektion eines weiteren Individuums bewirkt dabei die leichte Verschiebung des Vektors in Richtung dieses Individuums. Diese Verschiebung ist im Bereich der künstlichen neuronalen Netze als Hebbsche Lernregel bekannt [Heb49, Rec99]. Der von Harik, Lobo und Goldberg entwickelte *Compact Genetic Algorithm* (cGA) verwendet denselben Mechanismus [Har97]. Der von Mühlenbein entwickelte *Univariate Marginal Distribution Algorithm* (UMDA) unterstellt ebenfalls eine Unabhängigkeit der Variablen verwendet jedoch wieder eine Population von Lösungen anstelle eines reellwertigen Vektors [Müh97].

Pelikan, Goldberg und Lobo geben in [Pel00] eine graphische Notation der unterschiedlichen Modellierungsformen von Variablenabhängigkeiten. Bei dieser Notation werden Variablen durch Knoten und Abhängigkeiten durch Kanten dargestellt. Gruppen von Variablen, die voneinander abhängig sind, jedoch zu keiner weiteren Gruppe in Abhängigkeit stehen, werden durch eine unterbrochene Linie als Gruppe dargestellt. Mit dieser Notation kann die Klasse der EDA, die von unabhängigen Variablen ausgehen, wie folgt charakterisiert werden.

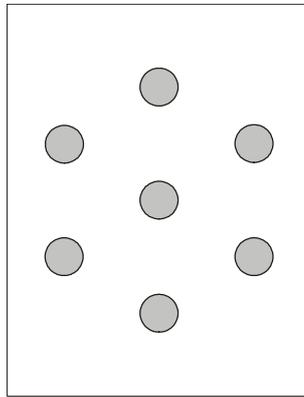


Abb. 3.9: Stochastisches Modell mit unabhängigen Variablen

PBIL, cGA und UMDA arbeiten zuverlässig, wenn sie zur Lösung von Optimierproblemen eingesetzt werden, bei denen es keine signifikante Abhängigkeit der Variablen gibt. Untersuchungen hierzu finden sich in [Müh97, Har97, Pel99]. Sind die Variablen abhängig, versagen diese Algorithmen. Um auch diese Art von Optimierproblemen lösen zu können sind weitere Verfahren entwickelt worden, die die Idee der statistischen Auswertung bestimmter Teile einer Population weiter verfolgen. Die ersten Versuche, Abhängigkeiten zwischen den Variablen in die Optimierung einzubeziehen stammen von Baluja und Davies [Bal97] sowie Pelikan und Mühlenbein [Pel98]. Sowohl der Algorithmus *Mutual Information Maximizing Input Clustering* (MIMIC) von Baluja und Davies, als auch der *Bivariate Marginal Distribution Algorithm* (BMDA) von Pelikan und Mühlenbein sind in der Lage Optimierprobleme zu lösen, bei denen die Variablen paarweise abhängig sind. In der Notation Pelikans kann die Struktur der stochastischen Modelle wie folgt beschrieben werden:

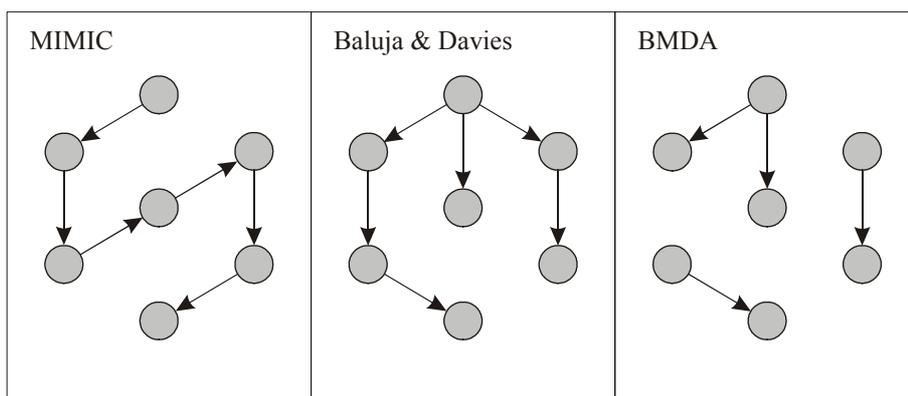


Abb. 3.10: Stochastische Modelle mit paarweise abhängigen Variablen

Einen anderen Ansatz verfolgt der *Extended Compact Genetic Algorithm* (ECGA) von Harik [Har99]. Die Grundannahme hierbei ist, dass alle Variablen in eine Menge von Variablengruppen aufgeteilt werden können. Die Variablen in einer Gruppe sind abhängig, die Variablengruppen untereinander sind jedoch unabhängig. Die Menge der Variablengruppen wird daraufhin, analog zu einem UMDA, als Menge unabhängiger Variablen betrachtet. Um die Menge der Variablengruppen zu

bestimmen wird ein gieriger Algorithmus verwendet, bei dem durch eine Straffunktion weniger komplexe Mengen von Variablengruppen bevorzugt werden.

Um auch höherwertige Abhängigkeiten modellieren zu können, ist von Mühlenbein, Rodriguez und Mahnig der *Factorized Distribution Algorithm* (FDA) entwickelt worden [Müh99a - Müh99c]. FDA zeigt gute Resultate bei Problemen die additiv zerlegbar sind. Von Nachteil ist, dass die Zerlegung und die resultierende Faktorisierung des Optimierproblems im Vorfeld bekannt sein muss. Ein FDA ist somit nicht in der Lage, ein stochastisches Modell des Optimierproblems automatisch zu erkennen bzw. zu lernen. Ein Verfahren, das ohne diese Information auskommt und dennoch höherwertige Abhängigkeiten erkennen kann ist der *Bayesian Optimization Algorithm* (BOA) von Pelikan, Goldberg und Cantú-Paz [Pel98].

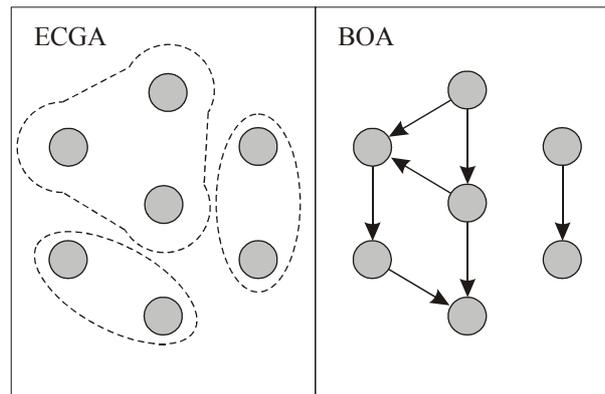


Abb. 3.11: Stochastische Modelle mit mehrfach abhängigen Variablen

Bayessche Netzwerke in Verbindung mit Lernalgorithmen sind von Etxeberria, Larrañaga und Lonzo ausführlich untersucht worden [Etx99, Lar99, Lar00]. In ihren Arbeiten verwenden sie zwei unterschiedliche stochastische Modell, die denen von PBIL und BOA ähneln.

EDAs sind in nahezu allen Anwendungsgebieten den einfachen Genetischen Algorithmen überlegen. Viele Varianten zeigen eine Invarianz gegenüber einer Neuordnung bzw. Umbenennung der einzelnen Variablen. Dies ist von Bedeutung, da durch diese Eigenschaft die Position einzelner Variablen innerhalb der Kodierung bedeutungslos wird. Die Kodierung wird so robuster. Problematisch ist, dass nicht alle Varianten der EDA in der Lage sind, die Abhängigkeiten der Variablen automatisch zu erkennen und daher Vorwissen über das Optimierungsproblem benötigen.

3.11 Zusammenfassung

Die Vielzahl an Erweiterungen der klassischen Evolutionsstrategien zeigt, dass zumindest für spezielle Probleme die klassischen Varianten Nachteile aufzeigen. Ein typisches Problem, das durch diese Erweiterungen gelöst werden soll, ist die frühzeitige Konvergenz der Schrittweite gegen 0. Ein Beispiel zur Lösung dieses Problems ist die *Robust Evolution Strategy* (RES). Andere Erweiterungen integrieren zusätzliche Merkmale der Evolution, wie lokale Selektion und Diffusionsmodelle. Zusätzlich zu den hier näher erläuterten Verfahren existiert eine Vielzahl weiterer Varianten. Zu nennen sind hier das *Improved Fast Evolutionary Programming* (IFEP) von Yao, et al., bei dem sowohl Cauchy- als auch normalverteilte Mutation stattfindet, die Verbindung von Delta-Evaluation mit gerichteter Mutation von Ross, et. al oder die Verwendung von Polarkoordinaten von Ghozeil und Fogel [Yao99, Ros94, Gho96].

Ein Nachteil vieler Erweiterungen ist jedoch die Einführung zusätzlicher Steuerparameter, die nicht durch die Evolutionsstrategie adaptiert werden, sondern vom Anwender im Vorfeld ermittelt werden

müssen. Können keine Regeln für die Wahl dieser Parameter gegeben werden, ist ein experimentelles Herantasten an geeignete Belegungen der Steuerparameter nötig. Problematisch ist zudem, dass einige Verfahren über Steuerparameter verfügen, deren Bedeutungen sich widersprechen wie im Falle des *Evolutionary-Gradient-Search* (EGS). Ein weiteres Problem kann durch den zusätzlichen Aufwand der integrierten Operationen und Datenstrukturen entstehen. Beispiele hierfür sind die Zeitkomplexität der kumulativen Schrittweitenregelung (KSA) sowie der Kovarianzmatrixadaptation (CMA), für die $O(n^3)$ gilt. Die Robust Evolution Strategy (RES) speichert mehrfache Kopien der Strategiekomponenten, sodass der benötigte Speicherplatz im Extremfall quadratisch mit der Dimensionalität des Optimierproblems steigt. Gleichzeitig arbeitet die Strategiemutation auf allen Kopien und benötigt daher ebenfalls mehr Rechenzeit. Hohe Anforderungen an den Speicherplatz sind im allgemeinen nicht mehr von Bedeutung, können aber den Einsatz dieser Varianten im Bereich der eingebetteten Systeme unmöglich machen. Abwägungen über die Rechenzeit können pauschal nicht gemacht werden. Sollte die Bestimmung einer optimalen Anordnung von Fertigungseinheiten in einer Fabrik bestimmt werden, die mit dieser Anordnung ein Jahrzehnt arbeitet, ist eine Laufzeit des Optimierverfahrens von einem Monat durchaus gerechtfertigt. Sollte jedoch die optimale Auslastung von Ressourcen in Telekommunikationsnetzen bestimmt werden, deren Last sich minütlich ändert, darf die Laufzeit eines Optimieralgorithmus auf keinen Fall länger sein als die Gültigkeit der erfassten Kapazitätsdaten.

Kapitel 4 Asymmetrische Verteilungsfunktionen

*Jeder Zufall ist Gesetz.
(Rainer Maria Rilke)*

Alle im Kapitel 3 (Erweiterungen der Evolutionsstrategien) vorgestellten Varianten der Evolutionsstrategie beruhen auf symmetrischen Verteilungsfunktionen. Zum Teil sind sehr aufwendige Verfahren entwickelt worden, um die Adaptationsfähigkeit, d. h. die automatische Erkennung von guten Strategieparametern, zu verbessern. Mit Ausnahme des Verfahrens von Weicker, das in Abschnitt 3.8 (Adaptation einer nicht-zentralen Normalverteilung, Seite 54) vorgestellt wurde, ist allen Varianten jedoch gemeinsam, dass sie symmetrische Verteilungsfunktionen um den Ursprung verwenden und damit grundsätzlich Veränderungen an den Objektvariablen vornehmen, deren Erwartungswert 0 ist. Bereits in der Einleitung dieser Arbeit wurde dargestellt, dass eine Motivation für Verteilungsfunktionen, deren Erwartungswert von 0 verschieden ist, darin liegt, dass bevorzugte Koordinaten**richtungen** adaptiert werden können. Im Vergleich zu allen anderen ES-Varianten, die unterschiedliche Schrittweiten für die einzelnen Koordinaten**achsen** des Optimierproblems adaptieren können¹, ermöglicht der Einsatz einer asymmetrischen Verteilungsfunktion unterschiedliche Schrittweiten für die Koordinaten**halbachsen**.

In diesem Kapitel wird die Entwicklung einer asymmetrischen Verteilungsfunktion in drei Entwicklungsabschnitten dargestellt. Beginnend mit der Verteilungsfunktion Ξ_c , die es erlaubt, in Abhängigkeit des Parameters c den Grad der Asymmetrie einzustellen, wird im zweiten Schritt die Funktion $\Xi_{c,\sigma}$ entwickelt, die zusätzlich einen richtungsunabhängigen Skalierungsfaktor σ besitzt. Den Abschluss bildet die Verteilungsfunktion $\Xi_{c,\sigma}^\gamma$, die einen globalen Skalierungsfaktor γ enthält, der das globale Verhalten einer asymmetrischen ES beeinflusst. Im Gegensatz zu den Parametern c und σ , wird der Parameter γ während einer Optimierung nicht adaptiert. Hieraus resultiert die Notation $\Xi_{c,\sigma}^\gamma$, die adaptierte Parameter als Index bezeichnet. Globale, nicht adaptierte Parameter werden als Exponent notiert. Die asymmetrischen Verteilungsfunktionen werden abschnittsweise definiert, mit Abschnittswechsel an der Stelle $x = 0$. Die Herleitungen für die Verteilungsfunktionen werden am Beispiel rechtsschiefer Verteilungen ($c \geq 0$)² durchgeführt. Alle Ableitungen gelten in ähnlicher Form auch für linksschiefe Verteilungen. Es wird während der Herleitung angegeben, ob es sich bei dem entsprechenden Wert bzw. der entsprechenden Funktion um den Teil handelt, der

-
1. Im Falle der korrelierten Normalverteilung können zusätzlich Schrittweiten bezüglich eines gedrehten Koordinatensystems adaptiert werden.
 2. Für $c = 0$ ist die Verteilungsfunktion symmetrisch. Dieser Fall wird in der Definition einer rechtsschiefen Verteilungsfunktion subsummiert.

für $c < 0$ oder für $c \geq 0$ gilt. Die vollständige Notation, dargestellt am Beispiel der Verteilungsfunktion, lautet daher:

$$\text{Definitionsbereich} \longrightarrow (c \geq 0) \overset{\text{globaler Parameter}}{\Xi^{\gamma}}_{c, \sigma} \longleftarrow \text{adaptierbare Parameter}$$

Je nach Variante der Verteilungsfunktion und des Definitionsbereiches, können einzelne Parameter als Skalare oder Vektoren Verwendung finden oder auch entfallen. Die Arbeitsweise einer asymmetrischen Evolutionsstrategie, die die in diesem Kapitel beschriebenen Verteilungsfunktionen verwendet, ist Gegenstand des Kapitels 6 (Asymmetrische Evolutionsstrategien, Seite 167), in dem der Ablauf einer asymmetrischen Evolutionsstrategie und die zugehörige Selbstadaptation ausführlich behandelt wird.

Bevor die Entwicklung der Verteilungsfunktion $\Xi^{\gamma}_{c, \sigma}$ beschrieben wird, stellt der folgende Abschnitt eine kurze Übersicht über die Eigenschaften von Verteilungsfunktionen sowie benötigte Grundlagen dar.

4.1 Eigenschaften von Wahrscheinlichkeitsverteilungen

Die Funktion Φ_X , die zu einer Zufallsvariablen X gehört¹, heißt Verteilungsfunktion und ist wie folgt definiert (siehe auch Def. 2.1, Seite 14):

$$\Phi_X(x) = P(X \leq x) \quad (4.1)$$

$\Phi_X(x)$ ist eine monoton steigende Funktion auf dem Intervall $[0, 1]$. Es gilt zusätzlich:

$$\lim_{x \rightarrow \infty} \Phi_X(x) = 1 \quad (4.2)$$

$$\lim_{x \rightarrow -\infty} \Phi_X(x) = 0 \quad (4.3)$$

$\Phi_X(x)$ ist rechtsseitig stetig, d. h.

$$\lim_{\delta \rightarrow 0} \Phi_X(x + \delta) = \Phi_X(x) \quad (4.4)$$

Hat eine Zufallsvariable X eine Verteilungsfunktion Φ_X , die durch

$$\Phi_X(x) = \int_{-\infty}^x \varphi_X(t) dt \quad (4.5)$$

dargestellt werden kann, so ist X eine Zufallsvariable mit stetiger Verteilung. Die Funktion $\varphi_X(x)$ wird als Dichtefunktion der Zufallsvariablen X bezeichnet.

1. Zentrales Thema dieser Arbeit ist die Entwicklung einer stetigen Verteilungsfunktion. Diskrete Verteilungen sind hier nur untergeordnet zu sehen. Wenn eine Zufallsvariable oder Verteilung ohne weiteren Zusatz beschrieben wird, handelt es sich um eine stetige Zufallsvariable oder stetige Verteilung.

Für $\varphi_X(x)$ gilt:

$$\varphi_X(x) = \frac{d}{dx}\Phi_X(x). \quad (4.6)$$

Zusätzlich gilt:

$$P(x_1 \leq X \leq x_2) = \Phi_X(x_2) - \Phi_X(x_1) = \int_{x_1}^{x_2} \varphi_X(t) dt \quad (4.7)$$

Eine stetige Zufallsvariable X mit der Verteilung Φ_X wird durch die im folgenden beschriebenen Kenngrößen charakterisiert.

4.1-1 Modus

Der Modus MD einer Menge von Zahlen ist der Wert, der mit der größten Häufigkeit in der Menge vorkommt. Für eine beliebige Menge von Zahlen ist es möglich, dass kein Modus, ein Modus oder beliebig viele Modi existieren. Übertragen auf Zufallsvariablen ist der Modus damit der Wert, bei dem die Dichtefunktion ein Maximum hat. Eine Zufallsvariable ist unimodal, wenn nur ein globales Maximum der Dichtefunktion existiert bzw. multimodal, wenn mehrere Maxima¹ existieren. Alle in dieser Arbeit betrachteten Zufallsvariablen die als Mutationsoperatoren für Evolutionsstrategien eingesetzt werden, sind unimodal, mit Modus an der Stelle $x = 0$.

4.1-2 Quantile

Das α -Quantil q_α einer Zufallsvariablen X ist wie folgt definiert:

Definition 4.1: Quantile

Sei X eine stetige Zufallsvariable mit Verteilungsfunktion Φ_X und Dichtefunktion φ_X . Für ein beliebiges $\alpha \in (0, 1) \subset \mathbb{R}$ ist das α -Quantil q_α als der Wert des Wertebereichs definiert, für den

$$P(X \leq q_\alpha) \geq \alpha, \text{ d. h.} \quad (4.8)$$

$$\Phi_X(q_\alpha) = \int_{-\infty}^{q_\alpha} \varphi_X(t) dt = \alpha \quad (4.9)$$

gilt.

1. Der Funktionswert der Dichtefunktion an der Stelle der Maxima muss identisch sein, da lokale Maxima im allgemeinen keine Modi sind.

Das 0.5-Quantil wird auch Median genannt. Für symmetrische Verteilungsfunktionen ist der Median identisch mit dem Erwartungswert. Weitere spezielle Quantile sind das untere bzw. obere Dezil mit $\alpha = 0.1$ und $\alpha = 0.9$ sowie das untere und obere Quartil, mit $\alpha = 0.25$ und $\alpha = 0.75$. Quantile werden u. a. als Streuungsmaße verwendet.

4.1-3 Momente

Momente stellen eine umfassende Klasse von Kenngrößen für Zufallsvariablen dar. Zu ihnen gehören Erwartungswert, Varianz und Schiefe. Allgemein gilt für das Moment k -ter Ordnung:

Definition 4.2: Momente

Sei X eine stetige Zufallsvariable mit Dichtefunktion φ_X . Das Moment k -ter Ordnung m_k ist als

$$m_k = \int_{-\infty}^{\infty} x^k \varphi_X(x) dx \quad (4.10)$$

definiert, vorausgesetzt, dass das Integral

$$\int_{-\infty}^{\infty} |x|^k \varphi_X(x) dx \quad (4.11)$$

existiert.

Für die zentralen Momente k -ter Ordnung, die auf den Erwartungswert (Zentrum) bezogen sind, gilt:

Definition 4.3: Zentrale Momente

Sei X eine stetige Zufallsvariable mit Dichtefunktion φ_X . Das zentrale Moment k -ter Ordnung μ_k ist als

$$\mu_k = \int_{-\infty}^{\infty} (x - m_1)^k \varphi_X(x) dx \quad (4.12)$$

definiert, vorausgesetzt, dass das Integral

$$\int_{-\infty}^{\infty} |x - m_1|^k \varphi_X(x) dx \quad (4.13)$$

existiert.

Zusätzlich zu den Momenten k -ter Ordnung und den zentralen Momenten k -ter Ordnung existiert eine weitere Klasse von Momenten, die bezüglich der Standardabweichung normiert werden:

Definition 4.4: Standardisierte Momente

Sei X eine stetige Zufallsvariable mit Dichtefunktion φ_X . Das standardisierten Moment k -ter Ordnung η_k ist als

$$\eta_k = \int_{-\infty}^{\infty} \left(\frac{x - m_1}{\sqrt{\mu_2}} \right)^k \varphi_X(x) dx \quad (4.14)$$

definiert, vorausgesetzt, dass das Integral

$$\int_{-\infty}^{\infty} \left| \frac{x - m_1}{\sqrt{\mu_2}} \right|^k \varphi_X(x) dx \quad (4.15)$$

existiert.

Die standardisierten Momente η_k der Ordnung k werden in der Literatur häufig verwendet, um Eigenschaften der zugehörigen Verteilungen darzustellen. Von Bedeutung sind hier der Koeffizient der Schiefe η_3 sowie der Koeffizient der Kurtosis η_4 . Abschließend seien an dieser Stelle die absoluten Momente erwähnt, die entstehen, wenn als Faktor die Beträge verwendet werden. Somit entsprechen die absoluten Momente der Gleichung (4.11), die absoluten zentralen Momente der Gleichung (4.13) und die absoluten standardisierten Momente der Gleichung (4.15), vorausgesetzt die Integrale existieren.

4.1-4 Erwartungswert (1. Moment)

Der Erwartungswert E_X beschreibt das durchschnittliche Verhalten einer Zufallsvariablen X . Er ist wie folgt definiert¹:

$$E_X = m_1 = \int_{-\infty}^{\infty} x \varphi_X(x) dx \quad (4.16)$$

Für den Erwartungswert gelten folgende Rechenregeln: Ist eine Zufallsvariable $Y = g(X)$ als Funktion der Zufallsvariablen X definiert, so folgt für den Erwartungswert:

$$E_Y = \int_{-\infty}^{\infty} g(x) \varphi_X(x) dx \quad (4.17)$$

1. Es gibt Verteilungen, für die kein Erwartungswert existiert. Bei nahezu allen in dieser Arbeit betrachteten Verteilungen existiert der Erwartungswert, daher wird dieser Sachverhalt nicht in jeder Aussage über Erwartungswerte wiederholt. Sollte der Erwartungswert einer Zufallsvariablen nicht existieren, wird dies explizit erwähnt.

Für lineare Abbildungen $Y = \alpha X + \beta$ gilt:

$$E_Y = \alpha E_X + \beta \quad (4.18)$$

Für die Summe von n Zufallsvariablen X_i mit $1 \leq i \leq n$ gilt:

$$E_{\sum_i X_i} = \sum_i E_{X_i} \quad (4.19)$$

Sind die Zufallsvariablen X_i unabhängig, so gilt zusätzlich:

$$E_{\prod_i X_i} = \prod_i E_{X_i} \quad (4.20)$$

Für die spätere Berechnung der Erwartungswerte abschnittsweise definierter, asymmetrischer Verteilungsfunktionen wird der Begriff des partiellen Erwartungswertes benötigt, für den folgende Definition gilt:

Definition 4.5: Partielle Erwartungswerte

Ein partieller Erwartungswert $E_X|_a^b$ ist über dem Intervall $[a, b] \subseteq \mathbb{R}$ definiert, und es gilt:

$$E_X|_a^b = \int_a^b x \varphi_X(x) dx \quad (4.21)$$

Für den Erwartungswert E_X von abschnittsweise definierten, asymmetrisch verteilten Zufallsvariablen, wie sie in dieser Arbeit behandelt werden, gilt damit insbesondere:

$$\begin{aligned} E_X &= E_X|_{-\infty}^{\infty} \\ &= E_X|_{-\infty}^0 + E_X|_0^{\infty} \end{aligned} \quad (4.22)$$

Das Konzept der partiellen Erwartungswerte kann auf alle Momente erweitert werden.

4.1-5 Varianz (2. Moment)

Die Varianz V_X einer Zufallsvariablen X ist das zweite zentrale Moment. Sie ist ein Maß für die Streuung der Realisierung der Zufallsvariablen X um den Erwartungswert E_X . Für die Varianz V_X gilt:

$$V_X = \mu_2 = E_{(X-E_X)^2} \text{ bzw.} \quad (4.23)$$

$$V_X = \int_{-\infty}^{\infty} (x - E_X)^2 \varphi_X(x) dx \quad (4.24)$$

Der eng mit der Varianz V_X verbundene Begriff der Standardabweichung SD_X ist wie folgt definiert:

$$SD_X = \sqrt{V_X} \quad (4.25)$$

Gleichung (4.23) lässt sich durch folgende Umformung in eine Form bringen, die eine einfache Berechnung der Varianz durch die Momente erlaubt:

$$\begin{aligned} V_X &= E_{(X-E_X)^2} \\ &= E_{(X^2 - 2E_X X + E_X^2)} \\ &= E_{X^2} - 2m_1 E_X + m_1^2 \\ &= m_2 - m_1^2 \end{aligned} \quad (4.26)$$

In Gleichung (4.26) fließt die Additionsregel (4.19) und die Regel für lineare Abbildungen (4.18) für Erwartungswerte ein. Für die Terme $E(X^2)$ und $E(X)^2$ in Gleichung (4.26) gilt:

$$E_{X^2} = m_2 = \int_{-\infty}^{\infty} x^2 \varphi_X(x) dx \quad (4.27)$$

und

$$E_X^2 = m_1^2 = \left(\int_{-\infty}^{\infty} x \varphi_X(x) dx \right)^2 \quad (4.28)$$

4.1-6 Schiefe (3. Moment)

Die Schiefe basiert auf dem 3. Moment. Sie bezeichnet den Grad der Asymmetrie bezüglich der y -Achse einer Dichtefunktion bzw. der Grad der Asymmetrie bezüglich des y -Abschnittes $\Phi(0)$ der Verteilungsfunktion. Gilt für den y -Abschnitt der Verteilungsfunktion $\Phi(0) > 0.5$, so wird Φ als rechtsschief bezeichnet. Gilt $\Phi(0) < 0.5$, wird Φ als linksschief bezeichnet. Für die Schiefe existieren mehrere Definitionen, die sich darin unterscheiden, ob sie auf den Erwartungswert bezogen oder mittels der Varianz normiert werden. Häufig wird der Koeffizient der Schiefe η_3 verwendet, der als das standardisierte Moment 3-ter Ordnung definiert ist. Es gilt:

$$\eta_3 = \int_{-\infty}^{\infty} \left(\frac{x - m_1}{\sqrt{\mu_2}} \right)^3 \varphi_X(x) dx = \frac{\mu_3}{\mu_2^{3/2}} = \frac{E_{(X-E_X)^3}}{(E_{(X-E_X)^2})^{3/2}} \quad (4.29)$$

Das 3. zentrale Moment μ_3 kann ähnlich zu (4.26) hergeleitet werden:

$$\begin{aligned}
 \mu_3 &= E_{(X-E_X)^3} \\
 &= E[X^3 - 3X^2E_X + 3XE_X^2 - E_X^3] \\
 &= E_{X^3} - 3m_1E_{X^2} + 3m_1^2E_X - m_1^3 \\
 &= m_3 - 3m_1m_2 + 2m_1^3
 \end{aligned} \tag{4.30}$$

Für die Momente m_1 , m_2 , m_3 , und m_1^3 gilt:

$$m_1 = E_X = \int_{-\infty}^{\infty} x \varphi_X(x) dx \tag{4.31}$$

$$m_2 = \int_{-\infty}^{\infty} x^2 \varphi_X(x) dx \tag{4.32}$$

$$m_3 = \int_{-\infty}^{\infty} x^3 \varphi_X(x) dx \tag{4.33}$$

$$m_1^3 = \left(\int_{-\infty}^{\infty} x \varphi_X(x) dx \right)^3 \tag{4.34}$$

Die in dieser Arbeit entwickelten asymmetrischen Verteilungs- und Dichtefunktionen sind abschnittsweise definiert. Daher gestaltet sich die Berechnung der Momente, besonders des Momentes m_1^3 aufwendig. Das Schiefemaß m_3 lässt sich mit weniger Aufwand berechnen, liefert aber ein Maß, dass nicht auf den Erwartungswert bezogen ist. In dieser Arbeit wird die Schiefe der entwickelten Verteilungsfunktion nicht mit anderen asymmetrischen Verteilungen verglichen, da solche aus der Literatur bekannten Verteilungen nicht für die Verwendung in Evolutionsstrategien geeignet sind (siehe hierfür auch den Abschnitt 4.6-1 (Alternative Verteilungsfunktionen, Seite 114)). Die Wahl des konkret verwendeten Maßes in dieser Arbeit ist damit beliebig.

4.2 Mutationsfunktionen

Verteilungsfunktionen bzw. Zufallsvariablen, die passend zu einer bestimmten Verteilungsfunktion erzeugt werden und sich für den Einsatz in Evolutionsstrategien eignen, sind Kernpunkt dieser Arbeit. Häufig werden solche Verteilungsfunktionen auch als Mutationsfunktionen bezeichnet [Rud98].

Beyer gibt eine umfassende Charakterisierung des Begriffs Mutationsfunktion. Demzufolge sollte eine Mutationsfunktion über folgende Eigenschaften verfügen [Bey01]:

1. Erreichbarkeit

Die Mutationsfunktion muss es jedem beliebigen Individuum erlauben, in endlicher Zeit jede Stelle des Suchraumes zu erreichen. Diese Bedingung muss notwendigerweise erfüllt sein.

2. Skalierbarkeit

Die Schrittweite, mit der Individuen im Laufe einer Optimierung ihre Position im Suchraum verändern, soll sich den Erfordernissen der Zielfunktion anpassen. So kann eine lokal optimale Schrittweite erreicht werden.

3. Verzicht auf systematische Tendenzen

Die zugrunde liegende Verteilungsfunktion sollte dem Maximum-Entropie-Prinzip gehorchen. Es besagt, dass bei unvollständigen Informationen eine Verteilung gewählt werden sollte, die in maximaler Entropie bei Einhaltung aller Restriktionen resultiert. Übertragen auf Evolutionsstrategien bedeutet dies, dass für den Fall, dass die Informationen über die optimale Verteilungsfunktion unvollständig sind, die allgemeinsten, tendenzfreien Annahmen gemacht werden sollten. Dies resultiert dem zentralen Grenzwertsatz der Stochastik zufolge in einer Normalverteilung.

4. Symmetrie

Die Mutationsfunktion sollte symmetrisch sein und keine Koordinatenrichtung bevorzugen. Diese Forderung resultiert in einem Erwartungswert der Mutationsfunktion von 0.

Diese Liste von Forderungen ist von Beyer als Richtlinie für den Entwurf von Mutationsfunktionen vorgestellt worden. Mit Ausnahme der Forderung 1 nach Erreichbarkeit einer beliebigen Stelle im Suchraum, müssen diese Forderungen nicht zwangsweise eingehalten werden. Die Forderung nach Symmetrie scheint sich jedoch so stark in der Gedankenwelt der Evolutionsstrategen verfestigt zu haben, dass asymmetrische Mutationsfunktionen bis zum heutigen Zeitpunkt nicht in die engere Betrachtung gezogen wurden.

Der Begriff Mutationsfunktion fordert in der Literatur eine symmetrische Verteilungsfunktion. Mutationsfunktionen, die diese Eigenschaft erfüllen, werden in dieser Arbeit als Mutationsfunktion im engeren Sinne bezeichnet. Alle im Abschnitt 3.1 (Variation der Mutationsfunktion, Seite 35) diskutierten Verteilungsfunktionen gehören zu diesen Mutationsfunktionen. Da diese Definition die asymmetrischen Verteilungsfunktionen ausschließt, wird in dieser Arbeit eine weitere Definition verwendet:

Definition 4.6: Mutationsfunktionen im weiteren Sinne

Die Verteilungsfunktion $\Phi_X(x)$ einer stetigen Zufallsvariable X ist eine Mutationsfunktion im weiteren Sinne, wenn folgende Bedingungen gelten:

1. Die Verteilungsfunktion gewährleistet die Erreichbarkeit, d. h.:

$$\Phi_X(x) \neq 0 \text{ ist streng monoton für alle } x \in \mathbb{R}^1 \text{ bzw.} \quad (4.35)$$

$$\varphi_X(x) > 0 \text{ für alle } x \in \mathbb{R} \quad (4.36)$$

-
1. Rudolph verwendet in seinen Arbeiten eine strengere Definition des Begriffs Erreichbarkeit [Rud98, Rud99]. Da die hier verwendete Verteilungsfunktion stochastisch unabhängige Zufallszahlen erzeugt und für die Schrittweiten σ_i per Definition $\sigma_i \in \mathbb{R}^+$ gilt, kann hier diese vereinfachte Definition der Erreichbarkeit gewählt werden.

2. Die Verteilungsfunktion ist skalierbar.
3. Bei allen richtungsbestimmenden Parametern handelt es sich um endogene Parameter.
4. Für den Modus gilt:

$$MD(X) = 0. \quad (4.37)$$

Der Grund, dass für den Modus $MD(X) = 0$ gelten muss, liegt in der Forderung, dass geringfügige Veränderungen häufiger auftreten sollen als große Veränderungen. Für den Fall, dass symmetrische Verteilungsfunktionen untersucht werden, fällt die Forderung $MD(X) = 0$ mit der Forderung $E_X = 0$ zusammen. Daraus folgt, dass zwischen dem Modus und dem Erwartungswert bei symmetrischen Verteilungen nicht unterschieden werden muss, wenn die wahrscheinlichste Veränderung, die durch die Mutationsfunktion erzielt wird, charakterisiert werden soll. Für den asymmetrischen Fall gilt dies nicht. Zusätzlich hat eine Verschiebung des Modus Auswirkungen auf die Reversibilität der Mutationsfunktion. Diese Auswirkung wird in dem Abschnitt 6.2-3 (Reversibilität, Seite 177) ausführlich behandelt. Die Forderung nach tendenzfreien Annahmen über unvollständige Informationen wird abgeschwächt, indem die Annahmen über diese Informationen endogen sind. Sie unterliegen der Selbstadaptation und werden nicht fest vorgegeben oder über eine Heuristik gesteuert.

4.3 Asymmetrische Verteilungsfunktion $\Xi_c(x)$

Zur Herleitung einer asymmetrisch verteilten Zufallsvariablen wird die Dichtefunktion verwendet. Die zugehörige Verteilungsfunktion kann dann durch Anwendung von (4.5) bestimmt werden [Hil96]. Als Grundlage soll die Exponentialfunktion $e^{-x^2/2}$ dienen. Mit dem zusätzlichen Parameter $c \in \mathbb{R}$ wird auf die Symmetrie der Dichtefunktion Einfluss genommen. Folgende Eigenschaften sollen, in Abhängigkeit des Parameters c , gelten:

1. symmetrisch für $c = 0$
Der Verlauf der Dichtefunktion soll für $c = 0$ symmetrisch zur y -Achse sein. Dies bedeutet, dass die Wahrscheinlichkeit für einen positiven Wert identisch ist mit der Wahrscheinlichkeit, dass ein negativer Wert erzeugt wird. Zusätzlich gilt, dass die Wahrscheinlichkeit für einen konkreten Wert x identisch ist mit der Wahrscheinlichkeit, dass der Wert $-x$ erzeugt wird.
2. rechtsschief für $c > 0$
Für Parameter $c > 0$ soll die Wahrscheinlichkeit für Werte auf der positiven Halbachse ($x > 0$) zunehmen. Das Maß der Steigerung soll proportional mit dem Betrag des Parameters c steigen.
3. linksschief für $c < 0$
Analog zu dem o. a. Fall soll die Wahrscheinlichkeit für Werte auf der negativen Halbachse ($x < 0$) zunehmen. Das Maß der Steigerung soll ebenfalls proportional mit dem Betrag des Parameters c steigen.

Zur Steuerung dieses Verlaufs wird die Grundfunktion wie folgt erweitert:

$${}_{(c \geq 0)}f(x) = e^{\frac{1}{2} \frac{x^2}{(1+c)}} \text{ bzw.} \quad (4.38)$$

$${}_{(c < 0)}f(x) = e^{\frac{1}{2} \frac{x^2}{(1-c)}} \quad (4.39)$$

Da die Funktionen (4.38) und (4.39) symmetrisch zur y -Achse ist, bietet sich für die Einführung der Asymmetrie eine abschnittsweise Definition an:

$${}_{(c \geq 0)}f(x) = \begin{cases} {}_{(c \geq 0)}f_1(x) & \text{für } x < 0 \\ {}_{(c \geq 0)}f_2(x) & \text{für } x \geq 0 \end{cases} \quad (4.40)$$

$${}_{(c \geq 0)}f_1(x) = e^{-\frac{1}{2}x^2} \quad (4.41)$$

$${}_{(c \geq 0)}f_2(x) = e^{\frac{1}{2} \frac{x^2}{(1+c)}} \quad (4.42)$$

Für $c < 0$ werden die abschnittswisen Definitionen vertauscht und c mit -1 multipliziert. Die zu (4.40) - (4.42) analogen Definitionen lauten dann:

$${}_{(c < 0)}f(x) = \begin{cases} {}_{(c < 0)}f_1(x) & \text{für } x < 0 \\ {}_{(c < 0)}f_2(x) & \text{für } x \geq 0 \end{cases} \quad (4.43)$$

$${}_{(c < 0)}f_1(x) = e^{\frac{1}{2} \frac{x^2}{(1-c)}} \quad (4.44)$$

$${}_{(c < 0)}f_2(x) = e^{-\frac{1}{2}x^2} \quad (4.45)$$

Mittels ${}_{(c < 0)}f(x)$ und ${}_{(c \geq 0)}f(x)$ lässt sich die Funktion $f(x)$ definieren:

$$f(x) = \begin{cases} {}_{(c < 0)}f(x) & \text{für } c < 0 \\ {}_{(c \geq 0)}f(x) & \text{für } c \geq 0 \end{cases} \quad (4.46)$$

Der Verlauf von $f(x)$ ist in Abbildung 4.1 für $c_1 = 5.0$ (blau) und $c_2 = -8.0$ (rot) dargestellt. Bei diesen Funktionen handelt es sich nicht um Dichtefunktionen. Die Normierung der Fläche sowie die Einhaltung der notwendigen Eigenschaften wird in den folgenden Abschnitten durchgeführt.

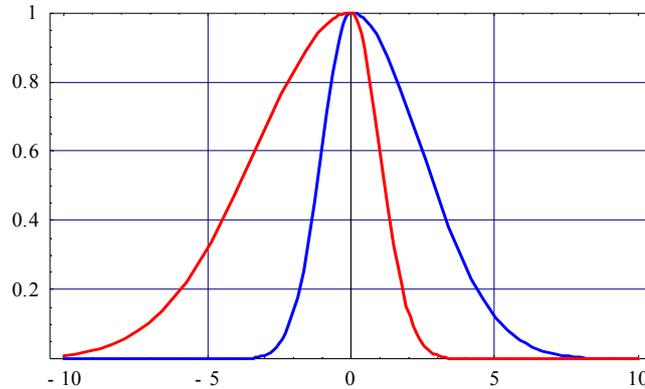


Abb. 4.1: Verlauf der asymmetrischen Funktion $f(x)$

Um die Funktion $f(x)$ nach (4.46) auf eine Fläche von 1.0 normieren zu können, muss zuerst die Fläche in Abhängigkeit des Parameters c bestimmt werden. Diese wird hier für den Fall $c \geq 0$ durchgeführt. Für $c < 0$ gilt die Herleitung analog¹. Die Fläche der Einzelfunktionen f_1 und f_2 lautet wie folgt:

$$\int_{-\infty}^0 (c \geq 0) f_1(x) dx = \int_{-\infty}^0 e^{-\frac{1}{2}x^2} dx \quad (4.47)$$

$$= \sqrt{\frac{\pi}{2}}$$

$$\int_0^{\infty} (c \geq 0) f_2(x) dx = \int_0^{\infty} e^{-\frac{1}{2(1+c)}x^2} dx \quad (4.48)$$

$$= \sqrt{\frac{\pi(1+c)}{2}}$$

Mit diesen Resultaten lautet die Fläche unter der Gesamtkurve von f :

$$\int_{-\infty}^{\infty} f(x) dx = \int_{-\infty}^0 (c \geq 0) f_1(x) dx + \int_0^{\infty} (c \geq 0) f_2(x) dx \quad (4.49)$$

$$= \frac{\sqrt{\pi} (1 + \sqrt{1+c})}{\sqrt{2}}$$

1. Für die analogen Herleitungen in diesem Kapitel gilt, dass sie durch Vertauschen der abschnittswisen Definitionen und Multiplizieren von c mit -1 durchgeführt werden müssen. Als Beispiel für diese analogen Herleitungen sei der Übergang der Gleichungen (4.40) - (4.42) auf die Gleichungen (4.43) - (4.45) angeführt.

Mit dem Normierungsfaktor nach (4.49) kann nun aus der Funktion $f(x)$ eine Wahrscheinlichkeitsdichtefunktion $\varphi(x)$ bestimmt werden. Diese Funktion ist, um Asymmetrie zu ermöglichen, ebenfalls abschnittsweise definiert.

$${}_{(c \geq 0)} \varphi(x) = \begin{cases} {}_{(c \geq 0)} \Phi_1(x) & \text{für } x < 0 \\ {}_{(c \geq 0)} \Phi_2(x) & \text{für } x \geq 0 \end{cases} \quad (4.50)$$

$${}_{(c \geq 0)} \Phi_1(x) = \frac{\sqrt{2}}{\sqrt{\pi}(1 + \sqrt{1+c})} e^{-\frac{1}{2}x^2} \quad (4.51)$$

$${}_{(c \geq 0)} \Phi_2(x) = \frac{\sqrt{2}}{\sqrt{\pi}(1 + \sqrt{1+c})} e^{-\frac{1}{2} \frac{x^2}{(1+c)}} \quad (4.52)$$

Für die analoge Dichtefunktion ${}_{(c < 0)} \varphi(x)$ gilt:

$${}_{(c < 0)} \varphi(x) = \begin{cases} {}_{(c < 0)} \Phi_1(x) & \text{für } x < 0 \\ {}_{(c < 0)} \Phi_2(x) & \text{für } x \geq 0 \end{cases} \quad (4.53)$$

$${}_{(c < 0)} \Phi_1(x) = \frac{\sqrt{2}}{\sqrt{\pi}(1 + \sqrt{1-c})} e^{-\frac{1}{2} \frac{x^2}{(1-c)}} \quad (4.54)$$

$${}_{(c < 0)} \Phi_2(x) = \frac{\sqrt{2}}{\sqrt{\pi}(1 + \sqrt{1-c})} e^{-\frac{1}{2}x^2} \quad (4.55)$$

Auf Grundlage der Dichtefunktion ${}_{(c \geq 0)} \varphi(x)$ und der analogen Dichtefunktion ${}_{(c < 0)} \varphi(x)$ lässt sich nun die Dichtefunktion $\xi_c(x)$ definieren:

$$\xi_c(x) = \begin{cases} {}_{(c < 0)} \varphi(x) & \text{für } c < 0 \\ {}_{(c \geq 0)} \varphi(x) & \text{für } c \geq 0 \end{cases} \quad (4.56)$$

Der Verlauf von $\xi_c(x)$ ist in Abbildung 4.2 für $c_1 = 5.0$ (blau) und $c_2 = -10$ (rot) und $c_3 = 0$ (grün) dargestellt.

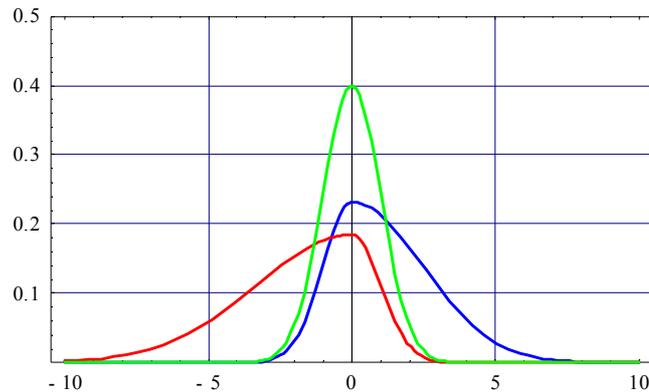


Abb. 4.2: Verlauf der Wahrscheinlichkeitsdichtefunktion φ für unterschiedliche c

Das Diagramm 4.3 zeigt den Verlauf von $\xi_c(x)$ für variierende $c \in [-5, 5] \subset \mathbb{R}$:

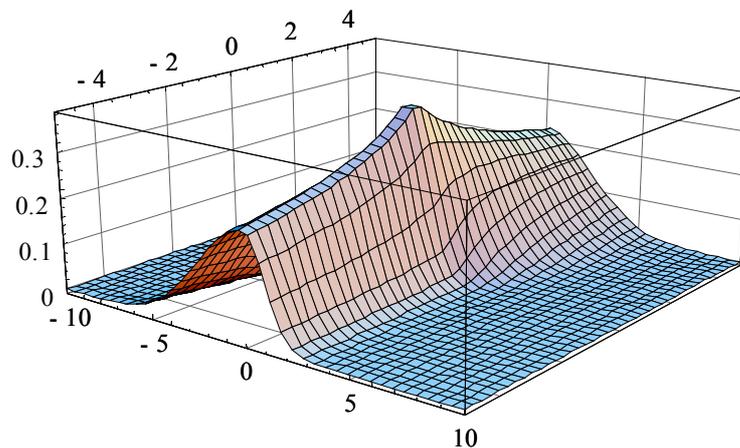


Abb. 4.3: Dichtefunktion $\xi_c(x)$ für variierende c

4.3-1 Verteilungsfunktion einer $\Xi_c(x)$ -verteilten Zufallsvariablen

In den vorhergehenden Abschnitten wurde eine asymmetrische Dichtefunktion sowie die daraus resultierenden Eigenschaften abgeleitet. Der folgende Abschnitt beschreibt die Entwicklung der zu $\xi_c(x)$ gehörenden Verteilungsfunktion $\Xi_c(x)$. Darauf aufbauend wird gezeigt, dass $\Xi_c(x)$ eine Verteilungsfunktion bzw. $\xi_c(x)$ tatsächlich eine Dichtefunktion ist. Die dann folgenden Abschnitte beschreiben die Bestimmung der inversen Verteilungsfunktion $\bar{\Xi}_c(x)$ sowie die Berechnung der charakteristischen Kenngrößen.

Nach (4.5) gilt für beliebige Dichtefunktionen φ_X :

$$\Phi_X(x) = \int_{-\infty}^x \varphi_X(t) dt \quad (4.57)$$

Angewandt auf die abschnittsweise Definition von $\xi_c(x)$ folgt daraus:

$$\Xi_c(x) = \begin{cases} x & \\ \int_{-\infty}^x \xi_c(t) dt & \text{für } x < 0 \\ 0 & \\ \int_{-\infty}^x \xi_c(t) dt + \int_0^x \xi_c(t) dt & \text{für } x \geq 0 \\ -\infty & 0 \end{cases} \quad (4.58)$$

Das Einsetzen der Definitionen für $\xi_c(x)$ führt im Fall $c \geq 0$ zu folgender Verteilungsfunktion ($c \geq 0$) $\Xi_c(x)$:

$${}_{(c \geq 0)} \Xi_c(x) = \begin{cases} \frac{1 + \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right)}{1 + \sqrt{1+c}} & \text{für } x < 0 \\ \frac{1 + \sqrt{1+c} \operatorname{erf}\left(\frac{x}{\sqrt{2+2c}}\right)}{1 + \sqrt{1+c}} & \text{für } x \geq 0 \end{cases} \quad (4.59)$$

Für den analogen Fall gilt:

$${}_{(c < 0)} \Xi_c(x) = \begin{cases} \frac{\sqrt{1-c} \left(1 + \operatorname{erf}\left(\frac{x}{\sqrt{2-2c}}\right)\right)}{1 + \sqrt{1-c}} & \text{für } x < 0 \\ \frac{\sqrt{1-c} + \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right)}{1 + \sqrt{1-c}} & \text{für } x \geq 0 \end{cases} \quad (4.60)$$

Für die allgemeine Verteilungsfunktion $\Xi_c(x)$ folgt daraus:

$$\Xi_c(x) = \begin{cases} {}_{(c < 0)} \Xi_c(x) & \text{für } c < 0 \\ {}_{(c \geq 0)} \Xi_c(x) & \text{für } c \geq 0 \end{cases} \quad (4.61)$$

Für die verwendete Fehlerfunktion $\operatorname{erf}(x)$ gilt:

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \quad (4.62)$$

Den Verlauf der Verteilungsfunktion $\Xi_c(x)$ für $c_1 = 5$ (blau), $c_2 = -10$ (rot) und $c_3 = 0$ (grün) stellt die Abbildung 4.4 dar.

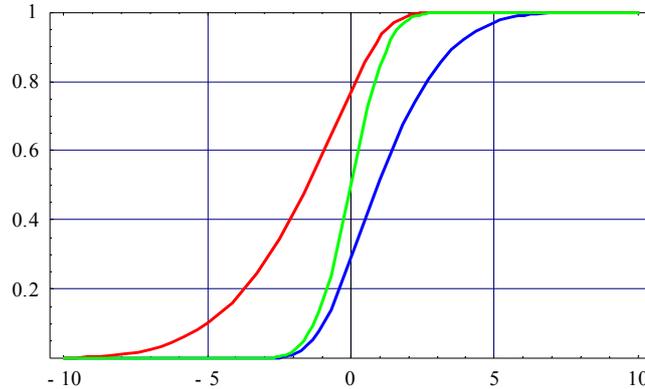


Abb. 4.4: Verlauf der Verteilungsfunktion $\Xi_c(x)$ für variierende c

4.3-2 $\Xi_c(x)$ als Mutationsfunktion

In den beiden vorhergehenden Abschnitten wurden die Verteilungsfunktion $\Xi_c(x)$ und die Dichtefunktion $\xi_c(x)$ hergeleitet. In diesem Abschnitt wird formal gezeigt, dass es sich bei der Funktion $\Xi_c(x)$ tatsächlich um eine Mutationsfunktion im weiteren Sinne handelt.

Satz 4.1: $\Xi_c(x)$ ist für beliebige $c \in \mathbb{R}$ eine Mutationsfunktion im weiteren Sinne.

Beweis 4.1: Zu zeigen ist zum einen, dass $\Xi_c(x)$ eine Wahrscheinlichkeitsverteilungsfunktion ist, indem die in Abschnitt 4.1 (Eigenschaften von Wahrscheinlichkeitsverteilungen, Seite 64) aufgeführten Eigenschaften nachgewiesen werden. Zum anderen muss gezeigt werden, dass die Eigenschaften einer Mutationsfunktion im weiteren Sinne nach Definition 4.6 des Abschnitts 4.2 (Mutationsfunktionen, Seite 70) erfüllt sind.

a. $\Xi_c(x)$ ist eine Wahrscheinlichkeitsverteilungsfunktion

1. Die Eigenschaft $\lim_{x \rightarrow -\infty} \Xi_c(x) = 0$ und $\lim_{x \rightarrow \infty} \Xi_c(x) = 1$ kann aus der Definition der Verteilungsfunktion sowie der zugehörigen Dichtefunktion abgeleitet werden. Es gilt allgemein:

$$\lim_{x \rightarrow -\infty} \Phi_X(x) = \Phi_X(-\infty) = \int_{-\infty}^{-\infty} \varphi_X(t) dt \quad (4.63)$$

Zusätzlich kann aus der Eigenschaft der Exponentialfunktion direkt abgeleitet werden, dass

$$\lim_{x \rightarrow -\infty} \varphi_X(x) = \lim_{x \rightarrow -\infty} \xi_c(x) = 0 \quad (4.64)$$

für beliebige $c \in \mathbb{R}$ gilt. Somit folgt

$$\int_{-\infty}^{-\infty} \xi_c(t) dt = \lim_{x \rightarrow -\infty} \Xi_c(x) = 0 \quad (4.65)$$

■

2. Die Eigenschaft $\lim_{x \rightarrow \infty} \Xi_c(x) = 1$ kann aus der allgemeinen Form

$$\lim_{x \rightarrow \infty} \Phi_X(x) = \Phi_X(\infty) = \int_{-\infty}^{\infty} \varphi_X(t) dt \quad (4.66)$$

abgeleitet werden. Es gilt für $c \geq 0$ unter Berücksichtigung von Gleichungen (4.50) - (4.52):

$$\begin{aligned} \int_{-\infty}^{\infty} \varphi(x) dx &= \int_{-\infty}^0 \varphi_1(x) dx + \int_0^{\infty} \varphi_2(x) dx \\ &= \int_{-\infty}^0 \frac{2}{\sqrt{\pi}(1+\sqrt{1+c})} e^{-\frac{1}{2}x^2} dx + \int_0^{\infty} \frac{2}{\sqrt{\pi}(1+\sqrt{1+c})} e^{-\frac{1}{2(1+c)}x^2} dx \\ &= \frac{1}{1+\sqrt{1+c}} + \frac{\sqrt{1+c}}{1+\sqrt{1+c}} \\ &= 1 \end{aligned} \quad (4.67)$$

Die analoge Berechnung für $c < 0$ führt zu dem gleichen Resultat. Damit gilt:

$$\int_{-\infty}^{\infty} \xi_c(t) dt = \lim_{x \rightarrow \infty} \Xi_c(x) = 1 \quad (4.68)$$

■

3. Die Eigenschaft $\lim_{\delta \rightarrow 0} \Phi_X(x + \delta) = \Phi_X(x)$, also die rechtsseitige Stetigkeit kann direkt aus der entsprechenden Eigenschaft der Exponentialfunktion abgeleitet werden. Zusätzlich ist zu zeigen, dass an der Stelle $x = 0$ die Stetigkeit der Exponentialfunktion erhalten bleibt. Für diesen Beweis wird auf den Abschnitt 4.7-4 (Stetigkeit, Seite 131) verwiesen, in dem die rechts- und linksseitige Stetigkeit nachgewiesen wird.

■

b. $\Xi_c(x)$ ist eine Mutationsfunktion

1. Die strenge Monotonie einer Verteilungsfunktion $\Phi_X(x)$ kann über die Dichtefunktion gezeigt werden. Es muss gelten:

$$\xi_c(x) > 0 \text{ für alle } x \in \mathbb{R} \quad (4.69)$$

Diese Eigenschaft folgt direkt aus der Eigenschaft der Exponentialfunk-

tion, für die gilt:

$$ae^{-bx^2} > 0, \text{ für } a, b > 0. \quad (4.70)$$

Aus der Definition der Funktion $\xi_c(x)$ folgt, dass die Bedingung $a, b > 0$ für alle $c \in \mathbb{R}$ erfüllt ist. ■

2. Als Maß für die Skalierbarkeit kann die Varianz einer Zufallsvariablen verwendet werden. Im Abschnitt 4.3-5 (Varianz einer $\Xi_c(x)$ -verteilten Zufallsvariablen, Seite 84) wird die Varianz bestimmt. Hieraus folgt unmittelbar die Skalierbarkeit. ■
3. Die Forderung nach endogenen Richtungsparametern kann durch die Verteilungsfunktion allein nicht bewertet werden, da es sich hier um eine anwendungsabhängige Eigenschaft handelt. Im Vorgriff auf das Kapitel 6 (Asymmetrische Evolutionsstrategien, Seite 167) sei an dieser Stelle angemerkt, dass die asymmetrischen Evolutionsstrategien die Richtungsparameter ausschließlich als endogene Parameter verwenden. ■
4. Für die Lage des Modus ist zu zeigen, dass er für beliebige $c \in \mathbb{R}$ einzig an der Stelle $x = 0$ existiert. Die Differenzierbarkeit wird im Abschnitt 4.7-5 (Differenzierbarkeit, Seite 133) untersucht. Hier sei nur das Resultat angegeben. Notwendige Bedingung für ein Maximum ist

$$\frac{d}{dx}\xi_c(\tilde{x}^*) = 0 \quad (4.71)$$

Diese Bedingung gilt für beliebige $c \in \mathbb{R}$ ausschließlich an der Stelle $\tilde{x}^* = 0$. Diese Eigenschaft folgt, da $\xi_c(x)$ auf einer Exponentialfunktion basiert und keine translatorische Anteile enthält, die eine Verschiebung des Optimums entlang der x -Achse bewirken. Aus der Eigenschaft der Exponentialfunktion folgt zudem

$$\frac{d^2}{dx^2}\xi_c(\tilde{x}^*) < 0 \quad (4.72)$$

für beliebige $c \in \mathbb{R}$. Daraus folgt, dass das einzige Optimum ein Maximum an der Stelle $\tilde{x}^* = 0$ ist. ■

An dieser Stelle sind die formalen Eigenschaften der Funktion $\Xi_c(x)$ bewiesen, die ihren Einsatz als Mutationsfunktion ermöglichen. In den folgenden Abschnitten werden weitere für diese Arbeit notwendige Eigenschaften der Funktion $\Xi_c(x)$ berechnet.

4.3-3 Inverse Verteilungsfunktion einer $\Xi_c(x)$ -verteilten Zufallsvariablen

Für die Entwicklung eines Zufallszahlengenerators, der eine spezielle Verteilung nachbilden kann und dem Prinzip der Transformation der Inversen folgt, ist es notwendig, die inverse Verteilungs-

funktion $\bar{\Phi}_X(y)$ einer Zufallsvariablen X zu verwenden. Die inverse Verteilungsfunktion liefert den Wert für y in

$$\bar{\Phi}_X(y) = x \quad (4.73)$$

Da die Verteilungsfunktion $\Xi_c(x)$ nach (4.61) abschnittsweise definiert ist, folgt daraus eine abschnittsweise Definition der inversen Verteilungsfunktion. Die Grenze der beiden Abschnitte liegt an der Stelle $x = 0$:

$$\Xi_c(0) = y_{Grenze} \quad (4.74)$$

bzw. unter Verwendung der inversen Verteilungsfunktion:

$$\bar{\Xi}_c(y_{Grenze}) = 0 \quad (4.75)$$

Angewandt auf die Verteilungsfunktion $\Xi_c(x)$ folgt daraus für den Fall $c \geq 0$:

$${}_{(c \geq 0)} \Xi_c(0) = \int_{-\infty}^0 \xi_c(t) dt = \frac{1}{1 + \sqrt{1+c}} \quad (4.76)$$

Analog gilt:

$${}_{(c < 0)} \Xi_c(0) = \int_{-\infty}^0 \xi_c(t) dt = \frac{\sqrt{1-c}}{1 + \sqrt{1-c}} \quad (4.77)$$

Durch (4.76) und (4.77) werden die Grenzen der abschnittswisen Definition der inversen Verteilungsfunktion festgelegt. Für den Fall $c \geq 0$ müssen folgende Gleichungen gelöst werden:

$$\frac{1 + \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right)}{1 + \sqrt{1+c}} = y \quad (4.78)$$

$$\frac{1 + \sqrt{1+c} \operatorname{erf}\left(\frac{x}{\sqrt{2+2c}}\right)}{1 + \sqrt{1+c}} = y \quad (4.79)$$

Die Umformung nach x resultiert für (4.78) in

$$x = \sqrt{2} \operatorname{inverf}(y(1 + \sqrt{1+c}) - 1) \quad (4.80)$$

für die Gleichung (4.79) gilt:

$$x = \sqrt{2(1+c)} \operatorname{inverf}\left(y + \frac{y-1}{\sqrt{1+c}}\right) \quad (4.81)$$

Die in dieser Arbeit verwendete inverse Fehlerfunktion ist definiert als die Lösung für x in

$$y = \operatorname{erf}(x) \quad (4.82)$$

Die resultierende, vollständige Definition für $(c \geq 0)$ $\bar{\Xi}_c(y)$, unter Berücksichtigung der Abschnitte nach (4.76) lautet:

$${}_{(c \geq 0)} \bar{\Xi}_c(y) = \begin{cases} \sqrt{2} \operatorname{inverf}\left(y(1 + \sqrt{1+c}) - 1\right) & \text{für } y < \frac{1}{1 + \sqrt{1+c}} \\ \sqrt{2(1+c)} \operatorname{inverf}\left(y + \frac{y-1}{\sqrt{1+c}}\right) & \text{für } y \geq \frac{1}{1 + \sqrt{1+c}} \end{cases} \quad (4.83)$$

Die analoge Herleitung führt zu:

$${}_{(c < 0)} \bar{\Xi}_c(y) = \begin{cases} \sqrt{2(1-c)} \operatorname{inverf}\left(y\left(1 + \frac{1}{\sqrt{1-c}}\right) - 1\right) & \text{für } y < \frac{\sqrt{1-c}}{1 + \sqrt{1-c}} \\ \sqrt{2} \operatorname{inverf}\left(y(1 + \sqrt{1-c}) - \sqrt{1-c}\right) & \text{für } y \geq \frac{\sqrt{1-c}}{1 + \sqrt{1-c}} \end{cases} \quad (4.84)$$

Für die inverse Verteilungsfunktion gilt dann allgemein:

$$\bar{\Xi}_c(y) = \begin{cases} {}_{(c < 0)} \bar{\Xi}_c(y) & \text{für } c < 0 \\ {}_{(c \geq 0)} \bar{\Xi}_c(y) & \text{für } c \geq 0 \end{cases} \quad (4.85)$$

Den Verlauf der inversen Verteilungsfunktion $\bar{\Xi}_c(y)$ für $c_1 = 5$ (blau), $c_2 = -10$ (rot) und $c_3 = 0$ (grün) stellt die Abbildung 4.5 dar.

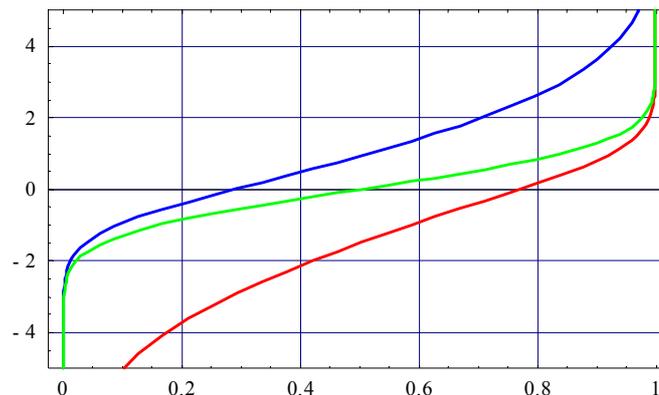


Abb. 4.5: Verlauf der inversen Verteilungsfunktion $\bar{\Xi}_c(y)$

4.3-4 Erwartungswert einer $\Xi_c(x)$ -verteilten Zufallsvariablen

Die Berechnung des Erwartungswertes einer $\Xi_c(x)$ -verteilten Zufallsvariablen X erfolgt mittels Berechnung der partiellen Erwartungswerte nach Definition 4.5. Für den Fall $c \geq 0$ gilt:

$$\begin{aligned}
 (c \geq 0) E_X &= (c \geq 0) E_X \Big|_{-\infty}^0 + (c \geq 0) E_X \Big|_0^{\infty} \\
 &= \int_{-\infty}^0 x (c \geq 0) \xi_c(x) dx + \int_0^{\infty} x (c \geq 0) \xi_c(x) dx \\
 &= -\frac{\sqrt{2}}{\sqrt{\pi}(1 + \sqrt{1+c})} + \frac{\sqrt{2}(1+c)}{\sqrt{\pi}(1 + \sqrt{1+c})} \\
 &= \frac{c\sqrt{2}}{\sqrt{\pi}(1 + \sqrt{1+c})}
 \end{aligned} \tag{4.86}$$

die analoge Herleitung für $c < 0$ resultiert in:

$$\begin{aligned}
 (c < 0) E_X &= (c < 0) E_X \Big|_{-\infty}^0 + (c < 0) E_X \Big|_0^{\infty} \\
 &= \int_{-\infty}^0 x (c < 0) \xi_c(x) dx + \int_0^{\infty} x (c < 0) \xi_c(x) dx \\
 &= \frac{\sqrt{2}(1-c)}{\sqrt{\pi}(1 + \sqrt{1-c})} - \frac{\sqrt{2}}{\sqrt{\pi}(1 + \sqrt{1-c})} \\
 &= \frac{c\sqrt{2}}{\sqrt{\pi}(1 + \sqrt{1-c})}
 \end{aligned} \tag{4.87}$$

Für E_X gilt damit allgemein:

$$E_X = \frac{c\sqrt{2}}{\sqrt{\pi}(1 + \sqrt{1+|c|})} \tag{4.88}$$

Den Verlauf des Erwartungswertes einer $\Xi_c(x)$ -verteilten Zufallsvariablen X in Abhängigkeit des Parameters c stellt Abbildung 4.6 dar.

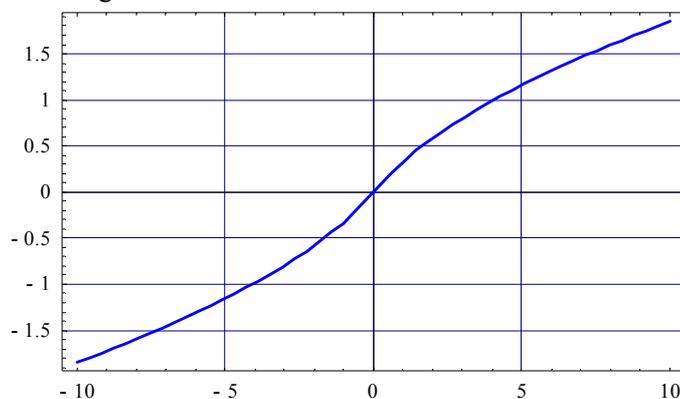


Abb. 4.6: Erwartungswert einer $\Xi_c(x)$ -verteilten Zufallsvariablen

4.3-5 Varianz einer $\Xi_c(x)$ -verteilten Zufallsvariablen

Für die Varianz einer $\Xi_c(x)$ -verteilten Zufallsvariablen X gilt nach (4.26) $V_X = m_2 - m_1^2$, mit den Momenten m_1 und m_2 . Für die Momente gilt:

$$m_2 = \int_{-\infty}^{\infty} x^2 \xi_c(x) dx \quad (4.89)$$

$$m_1^2 = \left(\int_{-\infty}^{\infty} x \xi_c(x) dx \right)^2 \quad (4.90)$$

Gleichung (4.26) kann verwendet werden, um die Varianz einer $\Xi_c(x)$ -verteilten Zufallsvariablen X zu bestimmen. Unter Verwendung von Definition 4.5 gilt hierbei:

$$V_X = E_{X^2} \Big|_{-\infty}^{\infty} - \left(E_X \Big|_{-\infty}^{\infty} \right)^2 \quad (4.91)$$

Da die Dichtefunktion $\xi_c(x)$ abschnittsweise definiert ist, ist die Anwendung von (4.91) komplex. Es gilt:

$$E_{X^2} \Big|_{-\infty}^{\infty} = E_{X^2} \Big|_{-\infty}^0 + E_{X^2} \Big|_0^{\infty} \quad (4.92)$$

$$\left(E_X \Big|_{-\infty}^{\infty} \right)^2 = \left(E_X \Big|_{-\infty}^0 \right)^2 + \left(E_X \Big|_0^{\infty} \right)^2 + 2 \cdot E_X \Big|_{-\infty}^0 \cdot E_X \Big|_0^{\infty} \quad (4.93)$$

Damit gilt für V_X :

$$V_X = E_{X^2} \Big|_{-\infty}^0 + E_{X^2} \Big|_0^{\infty} - \left[\left(E_X \Big|_{-\infty}^0 \right)^2 + \left(E_X \Big|_0^{\infty} \right)^2 + 2 \cdot E_X \Big|_{-\infty}^0 \cdot E_X \Big|_0^{\infty} \right] \quad (4.94)$$

Durch Einsetzen der abschnittswisen Definitionen für $\xi_c(x)$, resultiert daraus die folgende Varianz V_X für $c \geq 0$:

$$\begin{aligned}
(c \geq 0) V_X &= (c \geq 0) E_{X^2} \Big|_{-\infty}^0 + (c \geq 0) E_{X^2} \Big|_0^{\infty} \\
&\quad - \left[(c \geq 0) E_X \Big|_{-\infty}^0 \right]^2 + (c \geq 0) E_X \Big|_0^{\infty} \Big]^2 + 2 \cdot (c \geq 0) E_X \Big|_{-\infty}^0 \cdot (c \geq 0) E_X \Big|_0^{\infty} \\
&= \int_{-\infty}^0 x^2 (c \geq 0) \xi_c(x) dx + \int_0^{\infty} x^2 (c \geq 0) \xi_c(x) dx \\
&\quad - \left[\left(\int_{-\infty}^0 x (c \geq 0) \xi_c(x) dx \right)^2 + \left(\int_0^{\infty} x (c \geq 0) \xi_c(x) dx \right)^2 \right. \\
&\quad \left. + 2 \int_{-\infty}^0 x (c \geq 0) \xi_c(x) dx \int_0^{\infty} x (c \geq 0) \xi_c(x) dx \right] \\
&= \int_{-\infty}^0 x^2 (c \geq 0) \varphi_1(x) dx + \int_0^{\infty} x^2 (c \geq 0) \varphi_2(x) dx \\
&\quad - \left[\left(\int_{-\infty}^0 x (c \geq 0) \varphi_1(x) dx \right)^2 + \left(\int_0^{\infty} x (c \geq 0) \varphi_2(x) dx \right)^2 \right. \\
&\quad \left. + 2 \int_{-\infty}^0 x (c \geq 0) \varphi_1(x) dx \int_0^{\infty} x (c \geq 0) \varphi_2(x) dx \right] \\
&= \int_{-\infty}^0 x^2 \frac{\sqrt{2}}{\sqrt{\pi}(1+\sqrt{1+c})} e^{-\frac{1}{2}x^2} dx + \int_0^{\infty} x^2 \frac{2}{\sqrt{\pi}(1+\sqrt{1+c})} e^{-\frac{1}{2}\frac{x^2}{1+c}} dx \\
&\quad - \left[\left(\int_{-\infty}^0 x \frac{\sqrt{2}}{\sqrt{\pi}(1+\sqrt{1+c})} e^{-\frac{1}{2}x^2} dx \right)^2 + \left(\int_0^{\infty} x \frac{2}{\sqrt{\pi}(1+\sqrt{1+c})} e^{-\frac{1}{2}\frac{x^2}{1+c}} dx \right)^2 \right. \\
&\quad \left. + 2 \int_{-\infty}^0 x \frac{\sqrt{2}}{\sqrt{\pi}(1+\sqrt{1+c})} e^{-\frac{1}{2}x^2} dx \int_0^{\infty} x \frac{2}{\sqrt{\pi}(1+\sqrt{1+c})} e^{-\frac{1}{2}\frac{x^2}{1+c}} dx \right] \\
&= \frac{1}{1+\sqrt{1+c}} + \frac{(1+c)^{3/2}}{1+\sqrt{1+c}} \\
&\quad - \left(\frac{2}{(1+\sqrt{1+c})^2 \pi} + \frac{2(1+c)^2}{(1+\sqrt{1+c})^2 \pi} - \frac{4(1+c)}{(1+\sqrt{1+c})^2 \pi} \right) \\
&= 2+c-\sqrt{1+c} - \frac{2c^2}{(1+\sqrt{1+c})^2 \pi}
\end{aligned} \tag{4.95}$$

Für $c < 0$ gilt analog:

$$\begin{aligned}
({}^{c < 0})V_X &= ({}^{c < 0})E_{X^2} \Big|_{-\infty}^0 + ({}^{c < 0})E_{X^2} \Big|_0^{\infty} \\
&\quad - \left[({}^{c < 0})E_X \Big|_{-\infty}^0 \right]^2 + ({}^{c < 0})E_X \Big|_0^{\infty} \right]^2 + 2 \cdot ({}^{c < 0})E_X \Big|_{-\infty}^0 \cdot ({}^{c < 0})E_X \Big|_0^{\infty} \\
&= \int_{-\infty}^0 x^2 ({}^{c < 0})\xi_c(x) dx + \int_0^{\infty} x^2 ({}^{c < 0})\xi_c(x) dx \\
&\quad - \left[\left(\int_{-\infty}^0 x ({}^{c < 0})\xi_c(x) dx \right)^2 + \left(\int_0^{\infty} x ({}^{c < 0})\xi_c(x) dx \right)^2 \right. \\
&\quad \left. + 2 \int_{-\infty}^0 x ({}^{c < 0})\xi_c(x) dx \int_0^{\infty} x ({}^{c < 0})\xi_c(x) dx \right] \\
&= \int_{-\infty}^0 x^2 ({}^{c < 0})\Phi_1(x) dx + \int_0^{\infty} x^2 ({}^{c < 0})\Phi_2(x) dx \\
&\quad - \left[\left(\int_{-\infty}^0 x ({}^{c < 0})\Phi_1(x) dx \right)^2 + \left(\int_0^{\infty} x ({}^{c < 0})\Phi_2(x) dx \right)^2 \right. \\
&\quad \left. + 2 \int_{-\infty}^0 x ({}^{c < 0})\Phi_1(x) dx \int_0^{\infty} x ({}^{c < 0})\Phi_2(x) dx \right] \tag{4.96} \\
&= \int_{-\infty}^0 x^2 \frac{\sqrt{2}}{\sqrt{\pi}(1+\sqrt{1-c})} e^{-\frac{1}{2} \frac{x^2}{(1-c)}} dx + \int_0^{\infty} x^2 \frac{\sqrt{2}}{\sqrt{\pi}(1+\sqrt{1-c})} e^{-\frac{1}{2} x^2} dx \\
&\quad - \left[\left(\int_{-\infty}^0 x \frac{\sqrt{2}}{\sqrt{\pi}(1+\sqrt{1-c})} e^{-\frac{1}{2} \frac{x^2}{(1-c)}} dx \right)^2 + \left(\int_0^{\infty} x \frac{\sqrt{2}}{\sqrt{\pi}(1+\sqrt{1-c})} e^{-\frac{1}{2} x^2} dx \right)^2 \right. \\
&\quad \left. + 2 \int_{-\infty}^0 x \frac{\sqrt{2}}{\sqrt{\pi}(1+\sqrt{1-c})} e^{-\frac{1}{2} \frac{x^2}{(1-c)}} dx \int_0^{\infty} x \frac{\sqrt{2}}{\sqrt{\pi}(1+\sqrt{1-c})} e^{-\frac{1}{2} x^2} dx \right] \\
&= \frac{(1-c)^{3/2}}{1+\sqrt{1-c}} + \frac{1}{1+\sqrt{1-c}} \\
&\quad - \left(\frac{2(1-c)^2}{(1+\sqrt{1-c})^2 \pi} + \frac{2}{(1+\sqrt{1-c})^2 \pi} - \frac{4(c-1)}{(1+\sqrt{1-c})^2 \pi} \right) \\
&= 2-c-\sqrt{1-c} - \frac{2c^2}{(1+\sqrt{1-c})^2 \pi}
\end{aligned}$$

Für die Varianz V_X gilt damit:

$$V_X = \begin{cases} 2 - c - \sqrt{1 - c} - \frac{2c^2}{(1 + \sqrt{1 - c})^2 \pi} & \text{für } c < 0 \\ 2 + c - \sqrt{1 + c} - \frac{2c^2}{(1 + \sqrt{1 + c})^2 \pi} & \text{für } c \geq 0 \end{cases} \quad (4.97)$$

Allgemein gilt dann:

$$V_X = 2 + |c| - \sqrt{1 + |c|} - \frac{2c^2}{(1 + \sqrt{1 + |c|})^2 \pi} \quad (4.98)$$

Den Verlauf der Varianz einer $\Xi_c(x)$ -verteilten Zufallsvariablen X stellt Abbildung 4.7 dar.

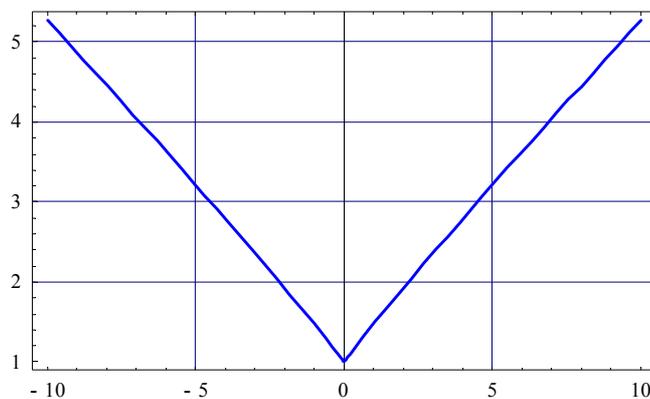


Abb. 4.7: Varianz einer $\Xi_c(x)$ -verteilten Zufallsvariablen

Den Verlauf der eng mit der Varianz verbundenen Standardabweichung $SD_X = \sqrt{V_X}$ stellt die Abbildung 4.8 dar.

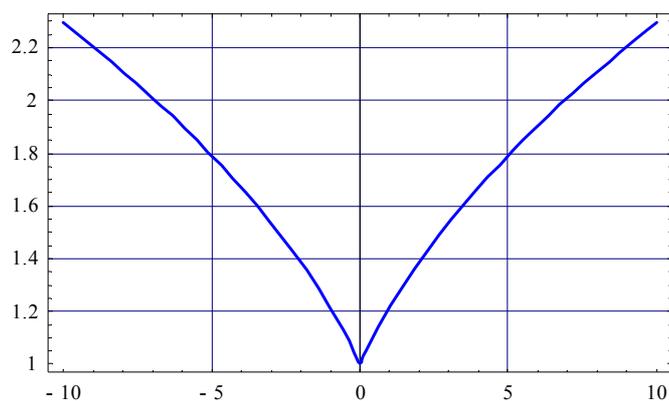


Abb. 4.8: Standardabweichung einer $\Xi_c(x)$ -verteilten Zufallsvariablen

4.3-6 Schiefe einer $\Xi_c(x)$ -verteilten Zufallsvariablen

Die Schiefe einer Verteilungsfunktion ist ein Maß für ihre Asymmetrie. Nach (4.29) gilt:

$$\eta_3 = \frac{\mu_3}{\mu_2^{3/2}} = \frac{E_{(X-E_X)^3}}{(E_{(X-E_X)^2})^{3/2}} \quad (4.99)$$

Das zentrale Moment 2. Ordnung $\mu_2 = V_X$ ist bereits im Abschnitt 4.3-5, Seite 84 bestimmt worden, sodass für $\mu_2^{3/2}$ gilt:

$$\mu_2^{3/2} = \sqrt{V_X^3} \quad (4.100)$$

μ_3 kann bestimmt werden, indem auf die Momente m_1 , m_2 und m_3 zurückgegriffen wird. Die Herleitung wird hier nur für $c \geq 0$ durchgeführt. Für den Fall $c < 0$ werden nur die Endresultate angegeben, da die Herleitung analog erfolgt. Nach (4.30) gilt:

$${}_{(c \geq 0)}\mu_3 = {}_{(c \geq 0)}E_{(X-E_X)^3} = {}_{(c \geq 0)}m_3 - 3{}_{(c \geq 0)}m_1 \cdot {}_{(c \geq 0)}m_2 + 2 \cdot {}_{(c \geq 0)}m_1^3 \quad (4.101)$$

$${}_{(c \geq 0)}m_1 = \int_{-\infty}^{\infty} x {}_{(c \geq 0)}\xi_c(x) dx = \frac{c\sqrt{2}}{\sqrt{\pi}(1 + \sqrt{1+c})} \quad (4.102)$$

$${}_{(c \geq 0)}m_2 = \int_{-\infty}^{\infty} x^2 {}_{(c \geq 0)}\xi_c(x) dx = 2 + c - \sqrt{1+c} \quad (4.103)$$

$${}_{(c \geq 0)}m_3 = \int_{-\infty}^{\infty} x^3 {}_{(c \geq 0)}\xi_c(x) dx = \frac{2c(2+c)\sqrt{2}}{\sqrt{\pi}(1 + \sqrt{1+c})} \quad (4.104)$$

Für die 3. Potenz des 1. Momentes gilt:

$${}_{(c \geq 0)}m_1^3 = \left(\int_{-\infty}^{\infty} x {}_{(c \geq 0)}\xi_c(x) dx \right)^3 \quad (4.105)$$

Diese Gleichung lässt sich aufgrund der abschnittswisen Definition der Dichtefunktion und Anwendung der Definition 4.5 umformen:

$$\begin{aligned} {}_{(c \geq 0)}m_1^3 &= \left({}_{(c \geq 0)}m_1 \Big|_{-\infty}^0 \right)^3 + 3 \left({}_{(c \geq 0)}m_1 \Big|_{-\infty}^0 \right)^2 \cdot {}_{(c \geq 0)}m_1 \Big|_0^{\infty} + \\ & 3 {}_{(c \geq 0)}m_1 \Big|_{-\infty}^0 \cdot \left({}_{(c \geq 0)}m_1 \Big|_0^{\infty} \right)^2 + \left({}_{(c \geq 0)}m_1 \Big|_0^{\infty} \right)^3 \end{aligned} \quad (4.106)$$

Es gilt:

$$\left({}_{(c \geq 0)} m_1 \Big|_{-\infty}^0 \right)^3 = -\frac{2\sqrt{2}}{\sqrt{\pi^3}(1 + \sqrt{1+c})} \quad (4.107)$$

$$\left({}_{(c \geq 0)} m_1 \Big|_{-\infty}^0 \right)^2 = \frac{2}{\pi(1 + \sqrt{1+c})} \quad (4.108)$$

$${}_{(c \geq 0)} m_1 \Big|_{-\infty}^0 = -\frac{\sqrt{2}}{\sqrt{\pi}(1 + \sqrt{1+c})} \quad (4.109)$$

$$\left({}_{(c \geq 0)} m_1 \Big|_0^{\infty} \right)^3 = \frac{2\sqrt{2}(1+c)^3}{\sqrt{\pi^3}(1 + \sqrt{1+c})} \quad (4.110)$$

$$\left({}_{(c \geq 0)} m_1 \Big|_0^{\infty} \right)^2 = \frac{2(1+c)^2}{\pi(1 + \sqrt{1+c})} \quad (4.111)$$

$${}_{(c \geq 0)} m_1 \Big|_0^{\infty} = \frac{\sqrt{2}(1+c)}{\sqrt{\pi}(1 + \sqrt{1+c})} \quad (4.112)$$

Durch Einsetzen folgt hieraus für (4.106):

$${}_{(c \geq 0)} m_1^3 \Big|_{-\infty}^{\infty} = \left(\frac{c\sqrt{2}}{\sqrt{\pi}(1 + \sqrt{1+c})} \right)^3 \quad (4.113)$$

Hieraus folgt für ${}_{(c \geq 0)} \mu_3$:

$${}_{(c \geq 0)} \mu_3 = \frac{\sqrt{2}c(4c^2 + \pi(\sqrt{1+c}(2+c) - c^2 + 2c + 2))}{(1 + \sqrt{1+c})^3 \pi^{3/2}} \quad (4.114)$$

Das Einsetzen dieser Ausdrücke in (4.99) resultiert in folgender Gleichung für das Schiefemaß ${}_{(c \geq 0)} \eta_3$:

$${}_{(c \geq 0)} \eta_3 = \frac{\sqrt{2}c(4c^2 + \pi(\sqrt{1+c}(2+c) - c^2 + 2c + 2))}{(1 + \sqrt{1+c})^3 (-4 - \sqrt{1+c}(\pi - 4) + c\pi - 2c + 2\pi)^{3/2}} \quad (4.115)$$

Die analoge Ableitung liefert

$${}_{(c < 0)} \eta_3 = -\frac{\sqrt{2}c(-4c^2 + \pi(\sqrt{1-c}(c-2) + c^2 + 2c - 2))}{(1 + \sqrt{1-c})^3 (-4 - \sqrt{1-c}(\pi - 4) - c\pi + 2c + 2\pi)^{3/2}} \quad (4.116)$$

Für η_3 gilt damit:

$$\eta_3 = \begin{cases} \frac{\sqrt{2}c(-4c^2 + \pi(\sqrt{1-c}(c-2) + c^2 + 2c - 2))}{(1 + \sqrt{1-c})^3(-4 - \sqrt{1-c}(\pi-4) - c\pi + 2c + 2\pi)^{3/2}} & \text{für } c < 0 \\ \frac{\sqrt{2}c(4c^2 + \pi(\sqrt{1+c}(2+c) - c^2 + 2c + 2))}{(1 + \sqrt{1+c})^3(-4 - \sqrt{1+c}(\pi-4) + c\pi - 2c + 2\pi)^{3/2}} & \text{für } c \geq 0 \end{cases} \quad (4.117)$$

Der Verlauf des Schiefemaßes η_3 in Abhängigkeit von c ist in Abbildung 4.9 dargestellt.

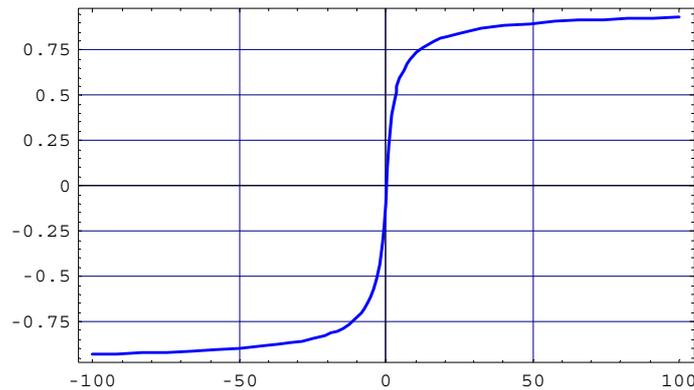


Abb. 4.9: Verlauf des Schiefemaßes η_3 in Abhängigkeit von c

Im Vergleich hierzu das Schiefemaß m_3 , das nicht auf den Erwartungswert bezogen ist und nicht bezüglich der Varianz normiert ist:

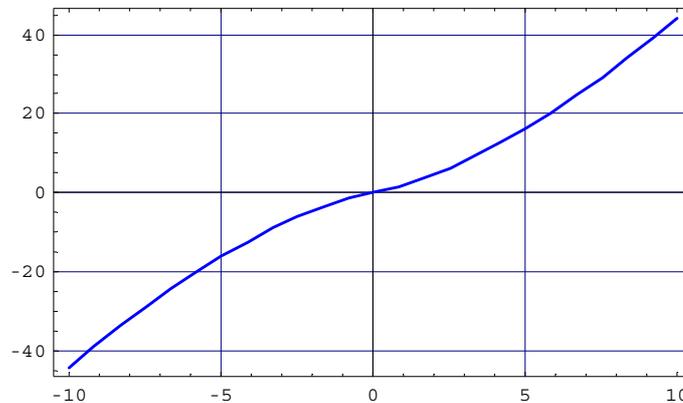


Abb. 4.10: Verlauf des Schiefemaßes m_3

4.4 Asymmetrische Verteilungsfunktion $\Xi_{c,\sigma}(x)$

Um eine Veränderung der Schrittweite, die unabhängig von der Koordinatenrichtungen ist, zu ermöglichen, muss die Funktion $\xi_c(x)$ um einen weiteren Steuerparameter σ erweitert werden [Hil96, Hil97b]. Der Parameter σ beeinflusst beide Koordinatenrichtungen gleichzeitig und ermöglicht so die richtungsunabhängige Vergrößerung und Verkleinerung der Schrittweite. Der Parameter σ stellt bei den asymmetrischen Evolutionsstrategien eine Schrittweitensteuerung dar, die von der Standardabweichung unterschieden werden muss. Bei Evolutionsstrategien, deren Mutationsfunk-

tion auf symmetrischen Verteilungen basiert, kann Schrittweite und Standardabweichung synonym verwendet werden, da es sich um den gleichen Parameter der Verteilungsfunktion handelt. Dies gilt für asymmetrische Verteilungen und asymmetrische Evolutionsstrategien nicht. Trotzdem gilt, analog zu den bekannten symmetrischen Mutationsfunktionen, dass die Schrittweite stets positiv ist: $\sigma \in \mathbb{R}_+$.

Die Grundlage für die um σ erweiterte Funktion $\xi_{c,\sigma}(x)$ bildet:

$${}_{(c \geq 0)}f(x) = e^{-\frac{1}{2\sigma(1+c)}x^2} \quad (4.118)$$

für beliebige, reelle $c \geq 0$ und $\sigma \in \mathbb{R}_+$. Auch hier soll eine abschnittsweise Definition die Asymmetrie ermöglichen:

$${}_{(c \geq 0)}f(x) = \begin{cases} e^{-\frac{1}{2\sigma}x^2} & \text{für } x < 0 \\ e^{-\frac{1}{2\sigma(1+c)}x^2} & \text{für } x \geq 0 \end{cases} \quad (4.119)$$

Die Definition für $c < 0$ ist analog, entsprechend (4.43) - (4.45). Um aus (4.119) eine Dichtefunktion zu erzeugen, muss die Gesamtfläche normiert werden, es gilt:

$$\begin{aligned} \int_{-\infty}^{\infty} f(x) dx &= \int_{-\infty}^0 {}_{(c \geq 0)}f_1(x) dx + \int_0^{\infty} {}_{(c \geq 0)}f_2(x) dx \\ &= \sqrt{\frac{\pi \sigma}{2}} + \sqrt{\frac{\pi(\sigma(1+c))}{2}} \\ &= \sqrt{\frac{\pi(\sigma(1+c))}{2}} \end{aligned} \quad (4.120)$$

Angewandt auf die Funktionen nach (4.119) folgt daraus die Dichtefunktion $\xi_{c,\sigma}(x)$:

$$\xi_{c,\sigma}(x) = \begin{cases} {}_{(c < 0)}\varphi(x) & \text{für } c < 0 \\ {}_{(c \geq 0)}\varphi(x) & \text{für } c \geq 0 \end{cases} \quad (4.121)$$

Für die hier verwendeten Funktionen ${}_{(c < 0)}\varphi(x)$ und ${}_{(c \geq 0)}\varphi(x)$ gilt:

$${}_{(c < 0)}\varphi(x) = \begin{cases} \frac{\sqrt{2}}{\sqrt{\pi\sigma(1+\sqrt{1-c})}} e^{-\frac{1}{2\sigma(1-c)}x^2} & \text{für } x < 0 \\ \frac{\sqrt{2}}{\sqrt{\pi\sigma(1+\sqrt{1-c})}} e^{-\frac{1}{2\sigma}x^2} & \text{für } x \geq 0 \end{cases} \quad (4.122)$$

$${}_{(c \geq 0)} \varphi(x) = \begin{cases} \frac{\sqrt{2}}{\sqrt{\pi\sigma}(1 + \sqrt{1+c})} e^{-\frac{1}{2}\frac{x^2}{\sigma}} & \text{für } x < 0 \\ \frac{\sqrt{2}}{\sqrt{\pi\sigma}(1 + \sqrt{1+c})} e^{-\frac{1}{2}\frac{x^2}{\sigma(1+c)}} & \text{für } x \geq 0 \end{cases} \quad (4.123)$$

Die Abbildung 4.11 zeigt den Verlauf der Dichtefunktion $\xi_{c,\sigma}(x)$ für $c = 2$, $\sigma_1 = 0.5$ (blau), $\sigma_2 = 1$ (rot) und $\sigma_3 = 5$ (grün).

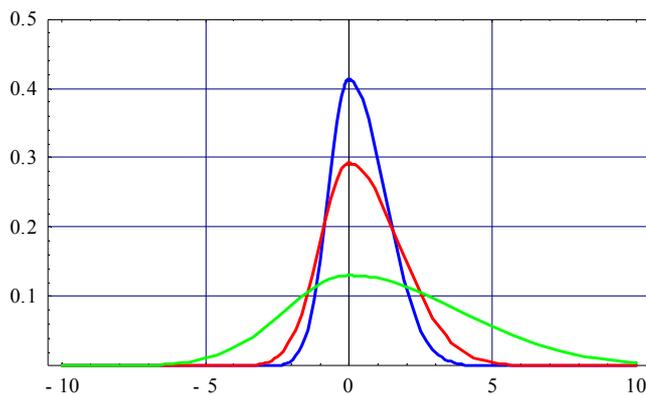


Abb. 4.11: Verlauf der Dichtefunktion $\xi_{c,\sigma}(x)$ für unterschiedliche σ

Das Diagramm 4.12 zeigt den Verlauf von $\xi_{c,\sigma}(x)$ für variierende $\sigma \in (0, 5] \subset \mathbb{R}$:

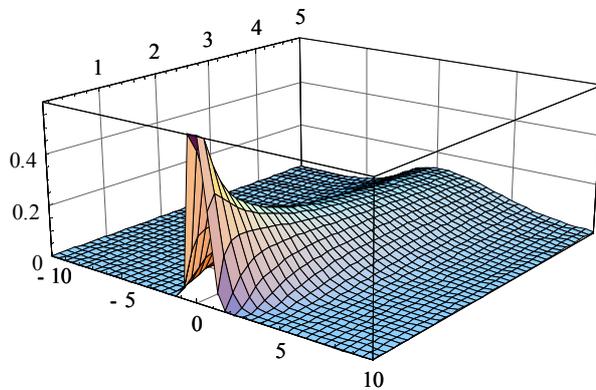


Abb. 4.12: Dichtefunktion $\xi_{c,\sigma}(x)$ für variierende σ

4.4-1 Verteilungsfunktion einer $\Xi_{c,\sigma}(x)$ -verteilten Zufallsvariablen

Die Berechnung der Verteilungsfunktion erfolgt entsprechend des Abschnittes 4.3 (Asymmetrische Verteilungsfunktion $\Xi_c(x)$, Seite 72) mittels des Ansatzes

$$\Phi_X(x) = \int_{-\infty}^x \varphi_X(t) dt \quad (4.124)$$

Angewandt auf die abschnittsweise Definition von $\xi_{c,\sigma}(x)$ folgt daraus:

$$\Xi_{c,\sigma}(x) = \begin{cases} x & \\ \int_{-\infty}^x \xi_{c,\sigma}(t) dt & \text{für } x < 0 \\ 0 & \\ \int_{-\infty}^0 \xi_{c,\sigma}(t) dt + \int_0^x \xi_{c,\sigma}(t) dt & \text{für } x \geq 0 \\ -\infty & 0 \end{cases} \quad (4.125)$$

Das Einsetzen der Definitionen für $\xi_{c,\sigma}(x)$ führt im Fall $c \geq 0$ zu folgender Verteilungsfunktion ($c \geq 0$) $\Xi_{c,\sigma}(x)$:

$${}_{(c \geq 0)} \Xi_{c,\sigma}(x) = \begin{cases} \frac{1 + \operatorname{erf}\left(\frac{x}{\sqrt{2\sigma}}\right)}{1 + \sqrt{1+c}} & \text{für } x < 0 \\ \frac{1 + \sqrt{1+c} \operatorname{erf}\left(\frac{x}{\sqrt{2\sigma(1+c)}}\right)}{1 + \sqrt{1+c}} & \text{für } x \geq 0 \end{cases} \quad (4.126)$$

für den analogen Fall gilt:

$${}_{(c < 0)} \Xi_{c,\sigma}(x) = \begin{cases} \frac{\sqrt{1-c} \left(1 + \operatorname{erf}\left(\frac{x}{\sqrt{2\sigma(1-c)}}\right)\right)}{1 + \sqrt{1-c}} & \text{für } x < 0 \\ \frac{\sqrt{1-c} + \operatorname{erf}\left(\frac{x}{\sqrt{2\sigma}}\right)}{1 + \sqrt{1-c}} & \text{für } x \geq 0 \end{cases} \quad (4.127)$$

Für die allgemeine Verteilungsfunktion $\Xi_{c,\sigma}(x)$ folgt daraus:

$$\Xi_{c,\sigma}(x) = \begin{cases} {}_{(c < 0)} \Xi_{c,\sigma}(x) & \text{für } c < 0 \\ {}_{(c \geq 0)} \Xi_{c,\sigma}(x) & \text{für } c \geq 0 \end{cases} \quad (4.128)$$

Der Verlauf von $\Xi_{c,\sigma}(x)$ wird in der Abbildung 4.13 für $\sigma_1 = 0.5$ (blau), $\sigma_2 = 1$ (rot) und $\sigma_3 = 5$ (grün) dargestellt. Für den Parameter c gilt hier: $c = 2$

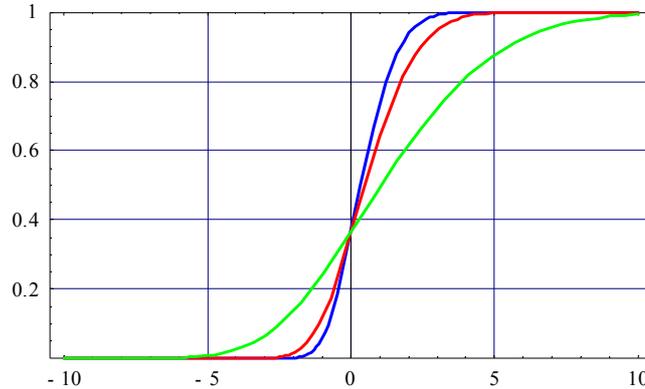


Abb. 4.13: Verlauf der Verteilungsfunktion $\Xi_{c,\sigma}(x)$ für unterschiedliche σ

4.4-2 $\Xi_{c,\sigma}(x)$ als Mutationsfunktion

An dieser Stelle kann, analog zu dem Abschnitt 4.3-2 ($\Xi_c(x)$ als Mutationsfunktion, Seite 78), gezeigt werden, dass $\Xi_{c,\sigma}(x)$ eine Mutationsfunktion im weiteren Sinne ist. Dieser Beweis ist nahezu identisch zu dem Beweis 4.1, lediglich bei den Berechnungen muss der neue Parameter σ berücksichtigt werden.

Satz 4.2: $\Xi_{c,\sigma}(x)$ ist für beliebige $c \in \mathbb{R}$ und $\sigma \in \mathbb{R}_+$ eine Mutationsfunktion im weiteren Sinne.

Beweis 4.2: Zu zeigen ist zum einen, dass $\Xi_{c,\sigma}(x)$ eine Wahrscheinlichkeitsverteilungsfunktion ist, indem die in Abschnitt 4.1 (Eigenschaften von Wahrscheinlichkeitsverteilungen, Seite 64) aufgeführten Eigenschaften nachgewiesen werden. Zum anderen muss gezeigt werden, dass der einzige Modus der Funktion $\Xi_{c,\sigma}(x)$ an der Stelle $x = 0$ existiert.

a. $\Xi_{c,\sigma}(x)$ ist eine Wahrscheinlichkeitsverteilungsfunktion

1. Die Eigenschaft $\lim_{x \rightarrow -\infty} \Xi_{c,\sigma}(x) = 0$ und $\lim_{x \rightarrow \infty} \Xi_{c,\sigma}(x) = 1$ kann aus der Definition der Verteilungsfunktion sowie der zugehörigen Dichtefunktion abgeleitet werden. Es gilt:

$$\lim_{x \rightarrow -\infty} \Phi_X(x) = \Phi_X(-\infty) = \int_{-\infty}^{-\infty} \varphi_X(t) dt \quad (4.129)$$

Zusätzlich kann aus der Eigenschaft der Exponentialfunktion direkt abgeleitet werden, dass

$$\lim_{x \rightarrow -\infty} \varphi_X(x) = \lim_{x \rightarrow -\infty} \xi_{c,\sigma}(x) = 0 \quad (4.130)$$

für beliebige $c \in \mathbb{R}$ und $\sigma \in \mathbb{R}_+$ gilt. Somit folgt

$$\int_{-\infty}^{-\infty} \xi_{c,\sigma}(t) dt = \lim_{x \rightarrow -\infty} \Xi_{c,\sigma}(x) = 0 \tag{4.131}$$



2. Die Eigenschaft $\lim_{x \rightarrow \infty} \Xi_{c,\sigma}(x) = 1$ kann aus

$$\lim_{x \rightarrow \infty} \Phi_X(x) = \Phi_X(\infty) = \int_{-\infty}^{\infty} \varphi_X(t) dt \tag{4.132}$$

abgeleitet werden. Es gilt für $c \geq 0$ unter Berücksichtigung der Gleichungen (4.121) - (4.123):

$$\begin{aligned} \int_{-\infty}^{\infty} \varphi(x) dx &= \int_{-\infty}^0 \varphi_1(x) dx + \int_0^{\infty} \varphi_2(x) dx \\ &= \int_{-\infty}^0 \frac{\sqrt{2}}{\sqrt{\pi\sigma(1+\sqrt{1+c})}} e^{-\frac{1}{2}\frac{x^2}{\sigma}} dx + \int_0^{\infty} \frac{\sqrt{2}}{\sqrt{\pi\sigma(1+\sqrt{1+c})}} e^{-\frac{1}{2}\frac{x^2}{\sigma(1+c)}} dx \tag{4.133} \\ &= \frac{1}{1+\sqrt{1+c}} + \frac{\sqrt{1+c}}{1+\sqrt{1+c}} \\ &= 1 \end{aligned}$$

Die analoge Berechnung für $c < 0$ führt zu dem gleichen Resultat. Damit gilt für beliebige c und $\sigma \in \mathbb{R}_+$:

$$\int_{-\infty}^{\infty} \xi_{c,\sigma}(t) dt = 1 = \lim_{x \rightarrow \infty} \Xi_{c,\sigma}(x) \tag{4.134}$$



3. Die Eigenschaft $\lim_{\delta \rightarrow 0} \Phi_X(x + \delta) = \Phi_X(x)$, also die rechtsseitige Stetigkeit kann direkt aus der entsprechenden Eigenschaft der Exponentialfunktion abgeleitet werden. Zusätzlich ist zu zeigen, dass an der Stelle $x = 0$ die Stetigkeit der Exponentialfunktion erhalten bleibt. Für diesen Beweis wird auf den Abschnitt 4.7-4 (Stetigkeit, Seite 131) verwiesen, in dem die rechts- und linksseitige Stetigkeit nachgewiesen wird.



b. $\Xi_{c,\sigma}(x)$ ist eine Mutationsfunktion

1. Die strenge Monotonie einer Verteilungsfunktion $\Phi_X(x)$ kann über die Dichtefunktion gezeigt werden. Es muss gelten:

$$\xi_{c,\sigma}(x) > 0 \text{ für alle } x \in \mathbb{R} \tag{4.135}$$

Diese Eigenschaft folgt direkt aus der Eigenschaft der Exponentialfunk-

tion, für die gilt:

$$ae^{-bx^2} > 0, \text{ für } a, b > 0 \quad (4.136)$$

Aus der Definition der Funktion $\xi_{c,\sigma}(x)$ folgt, dass die Bedingung $a, b > 0$ für alle $c \in \mathbb{R}$ und $\sigma \in \mathbb{R}_+$ erfüllt ist. ■

2. Als Maß für die Skalierbarkeit kann die Varianz einer Zufallsvariablen verwendet werden. Im Abschnitt 4.4-5 (Varianz einer $\Xi_{c,\sigma}(x)$ -verteilten Zufallsvariablen, Seite 99) wird die Varianz bestimmt. Hieraus folgt unmittelbar die Skalierbarkeit. ■
3. Die Forderung nach endogenen Richtungsparametern kann durch die Verteilungsfunktion allein nicht bewertet werden, da es sich hier um eine anwendungsabhängige Eigenschaft handelt. Im Vorgriff auf das Kapitel 6 (Asymmetrische Evolutionsstrategien, Seite 167) sei an dieser Stelle angemerkt, dass die asymmetrischen Evolutionsstrategien die Richtungsparameter ausschließlich als endogene Parameter verwenden. ■
4. Für die Lage des Modus ist zu zeigen, dass er für beliebige c und beliebige positive $\sigma \in \mathbb{R}_+$ einzig an der Stelle $x = 0$ existiert. Die Differenzierbarkeit wird im Abschnitt 4.7-5 (Differenzierbarkeit, Seite 133) untersucht. Hier sei nur das Resultat angegeben. Notwendige Bedingung für ein Maximum ist

$$\frac{d}{dx} \xi_{c,\sigma}(x^*) = 0 \quad (4.137)$$

Diese Bedingung gilt für beliebige c und $\sigma \in \mathbb{R}_+$ ausschließlich an der Stelle $x^* = 0$. Diese Eigenschaft folgt, da $\xi_{c,\sigma}(x)$ auf einer Exponentialfunktion basiert und keine translatorische Anteile enthält, die eine Verschiebung des Optimums entlang der x -Achse bewirken. Aus der Eigenschaft der Exponentialfunktion folgt zudem

$$\frac{d^2}{dx^2} \xi_{c,\sigma}(x^*) < 0 \quad (4.138)$$

für beliebige $c \in \mathbb{R}$ und $\sigma \in \mathbb{R}_+$. Daraus folgt, dass das einzige Optimum ein Maximum an der Stelle $x^* = 0$ ist. ■

Die Verteilungsfunktion $\Xi_{c,\sigma}(x)$ erfüllt damit die Forderungen 1 (Erreichbarkeit), 2 (Skalierbarkeit) und 4 (Stelle des Modus) der Definition 4.6 der Mutationsfunktion im weiteren Sinne. Die Forderung 3 (endogene Richtungsparameter) kann nicht von der Verteilungsfunktion realisiert werden, sondern nur von der Variante der Evolutionsstrategie, in der sie verwendet wird. Die in dieser Arbeit entwickelte asymmetrische Evolutionsstrategie erfüllt diese Forderung, sodass die Verteilungsfunktion $\Xi_{c,\sigma}(x)$ alle vier Forderungen an eine Mutationsfunktion im weiteren Sinne erfüllt.

4.4-3 Inverse Verteilungsfunktion einer $\Xi_{c,\sigma}(x)$ -verteilten Zufallsvariablen

Zur Bestimmung der inversen Verteilungsfunktion muss entsprechend dem Abschnitt 4.3-3 (Inverse Verteilungsfunktion einer $\Xi_c(x)$ -verteilten Zufallsvariablen, Seite 80), der Wert für y in

$$\bar{\Xi}_{c,\sigma}(y) = x \quad (4.139)$$

gefunden werden. Da die Verteilungsfunktion $\Xi_{c,\sigma}(x)$ nach (4.128) abschnittsweise definiert ist, folgt daraus eine abschnittsweise Definition der inversen Verteilungsfunktion. Die Grenze der beiden Abschnitte liegt an der Stelle $x = 0$:

$$\Xi_{c,\sigma}(0) = y_{Grenze} \quad (4.140)$$

bzw. unter Verwendung der inversen Verteilungsfunktion:

$$\bar{\Xi}_{c,\sigma}(y_{Grenze}) = 0 \quad (4.141)$$

Angewandt auf die Verteilungsfunktion $\Xi_{c,\sigma}(x)$ folgt daraus für den Fall $c \geq 0$:

$${}_{(c \geq 0)} \Xi_{c,\sigma}(0) = \int_{-\infty}^0 \xi_{c,\sigma}(t) dt = \frac{1}{1 + \sqrt{1+c}} \quad (4.142)$$

analog gilt:

$${}_{(c < 0)} \Xi_{c,\sigma}(0) = \int_{-\infty}^0 \xi_{c,\sigma}(t) dt = \frac{\sqrt{1-c}}{1 + \sqrt{1-c}} \quad (4.143)$$

Durch (4.142) und (4.143) werden die Grenzen der abschnittswisen Definition der inversen Verteilungsfunktion festgelegt. Unter Auslassung der Zwischenschritte resultiert die Berechnung in folgender Definition für die inverse Verteilungsfunktion:

$${}_{(c \geq 0)} \bar{\Xi}_{c,\sigma}(y) = \begin{cases} \sqrt{2\sigma} \operatorname{inverf}(y(1 + \sqrt{1+c}) - 1) & \text{für } y < \frac{1}{1 + \sqrt{1+c}} \\ \sqrt{2\sigma(1+c)} \operatorname{inverf}\left(y + \frac{y-1}{\sqrt{1+c}}\right) & \text{für } y \geq \frac{1}{1 + \sqrt{1+c}} \end{cases} \quad (4.144)$$

Die analoge Herleitung führt zu:

$${}_{(c < 0)} \bar{\Xi}_{c,\sigma}(y) = \begin{cases} \sqrt{2\sigma(1-c)} \operatorname{inverf}\left(y\left(1 + \frac{1}{\sqrt{1-c}}\right) - 1\right) & \text{für } y < \frac{\sqrt{1-c}}{1 + \sqrt{1-c}} \\ \sqrt{2\sigma} \operatorname{inverf}(y(1 + \sqrt{1-c}) - \sqrt{1-c}) & \text{für } y \geq \frac{\sqrt{1-c}}{1 + \sqrt{1-c}} \end{cases} \quad (4.145)$$

Für die inverse Verteilungsfunktion gilt dann allgemein:

$$\bar{\Xi}_{c, \sigma}(y) = \begin{cases} (c < 0) \bar{\Xi}_{c, \sigma}(y) & \text{für } c < 0 \\ (c \geq 0) \bar{\Xi}_{c, \sigma}(y) & \text{für } c \geq 0 \end{cases} \quad (4.146)$$

Den Verlauf der inversen Verteilungsfunktion $\bar{\Xi}_{c, \sigma}(y)$ für festes $c = 2$ und $\sigma_1 = 0.5$ (blau), $\sigma_2 = 1$ (rot) und $\sigma_3 = 5$ (grün) stellt die Abbildung 4.14 dar.

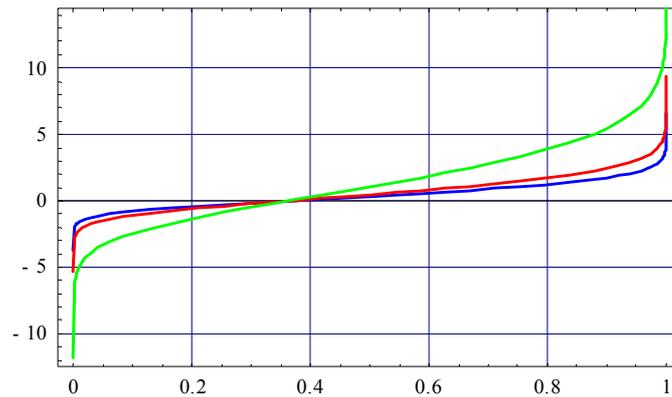


Abb. 4.14: Verlauf der inversen Verteilungsfunktion $\bar{\Xi}_{c, \sigma}(y)$ für unterschiedliche σ

4.4-4 Erwartungswert einer $\Xi_{c, \sigma}(x)$ -verteilten Zufallsvariablen

Die Berechnung des Erwartungswertes erfolgt auch bei einer Zufallsvariablen X mit Dichtefunktion $\xi_{c, \sigma}(x)$ mittels der partiellen Erwartungswerte nach Definition 4.5. Für den Fall $c \geq 0$ gilt:

$$\begin{aligned} (c \geq 0) E_X &= (c \geq 0) E_X \Big|_{-\infty}^0 + (c \geq 0) E_X \Big|_0^{\infty} \\ &= \int_{-\infty}^0 x (c \geq 0) \xi_{c, \sigma}(x) dx + \int_0^{\infty} x (c \geq 0) \xi_{c, \sigma}(x) dx \\ &= -\frac{\sqrt{2\sigma}}{\sqrt{\pi}(1 + \sqrt{1+c})} + \frac{\sqrt{2\sigma}(1+c)}{\sqrt{\pi}(1 + \sqrt{1+c})} \\ &= \frac{c\sqrt{2\sigma}}{\sqrt{\pi}(1 + \sqrt{1+c})} \end{aligned} \quad (4.147)$$

die analoge Herleitung für $c < 0$ resultiert in:

$$\begin{aligned}
(c < 0) E_X &= (c < 0) E_X \Big|_{-\infty}^0 + (c < 0) E_X \Big|_0^{\infty} \\
&= \int_{-\infty}^0 x (c < 0) \xi_{c,\sigma}(x) dx + \int_0^{\infty} x (c < 0) \xi_{c,\sigma}(x) dx \\
&= \frac{\sqrt{2\sigma}(1-c)}{\sqrt{\pi}(1+\sqrt{1-c})} - \frac{\sqrt{2\sigma}}{\sqrt{\pi}(1+\sqrt{1-c})} \\
&= \frac{c\sqrt{2\sigma}}{\sqrt{\pi}(1+\sqrt{1-c})}
\end{aligned} \tag{4.148}$$

Für E_X gilt damit allgemein:

$$E_X = \frac{c\sqrt{2\sigma}}{\sqrt{\pi}(1+\sqrt{1+|c|})} \tag{4.149}$$

Die Abbildung 4.15 zeigt den Verlauf des Erwartungswertes E_X einer $\Xi_{c,\sigma}(x)$ -verteilten Zufallsvariablen X , für festes $c = 2$, $\sigma_1 = 0.5$ (blau), $\sigma_2 = 1$ (rot) und $\sigma_3 = 5$ (grün).

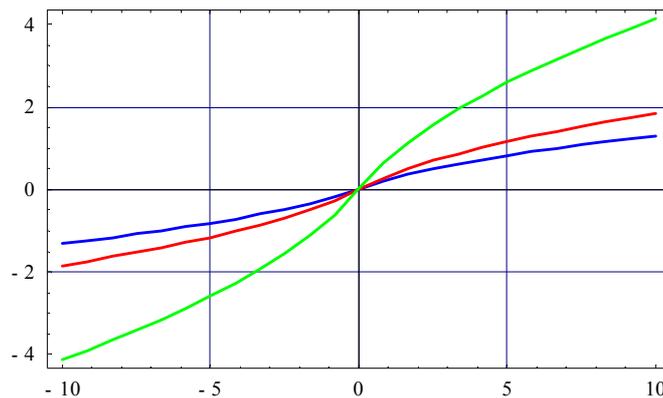


Abb. 4.15: Verlauf des Erwartungswertes einer $\Xi_{c,\sigma}(x)$ -verteilten Zufallsvariablen

4.4-5 Varianz einer $\Xi_{c,\sigma}(x)$ -verteilten Zufallsvariablen

Die Varianz V_X einer $\Xi_{c,\sigma}(x)$ -verteilten Zufallsvariable X kann entsprechend der Varianz einer $\Xi_c(x)$ -verteilten Zufallsvariablen berechnet werden. Unter Auslassung der Zwischenschritte, die denen in (4.95) und (4.96) entsprechen, gilt:

$$V_X = \begin{cases} \sigma \left((2 - \sqrt{1-c} - c) - \frac{2c^2}{\pi(1 + \sqrt{1-c})^2} \right) & \text{für } c < 0 \\ \sigma \left((2 - \sqrt{1+c} + c) - \frac{2c^2}{\pi(1 + \sqrt{1+c})^2} \right) & \text{für } c \geq 0 \end{cases} \tag{4.150}$$

Für beliebige $c \in \mathbb{R}$ und $\sigma \in \mathbb{R}_+$ gilt dann allgemein:

$$V_X = \sigma \left((2 - \sqrt{1 + |c|} + |c|) - \frac{2c^2}{\pi(1 + \sqrt{1 + |c|})^2} \right) \quad (4.151)$$

Den Verlauf der Varianz einer $\Xi_{c,\sigma}(x)$ -verteilten Zufallsvariablen X für festes $c = 2$, $\sigma_1 = 0.5$ (blau), $\sigma_2 = 1$ (rot) und $\sigma_3 = 5$ (grün) stellt Abbildung 4.16 dar.

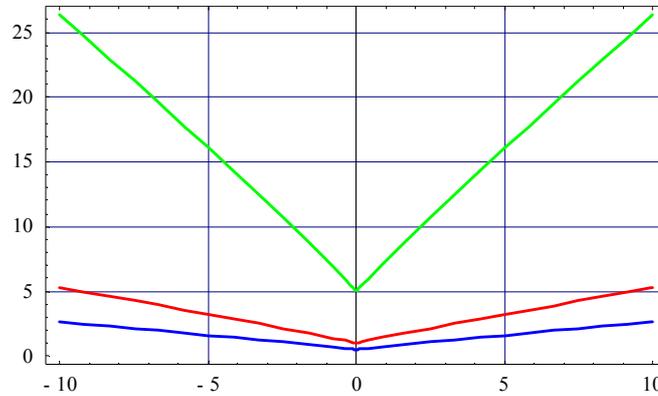


Abb. 4.16: Varianz einer $\Xi_{c,\sigma}(x)$ -verteilten Zufallsvariablen

Den Verlauf der eng mit der Varianz verbundenen Standardabweichung $SD_X = \sqrt{V_X}$ für festes $c = 2$, $\sigma_1 = 0.5$ (blau), $\sigma_2 = 1$ (rot) und $\sigma_3 = 5$ (grün) stellt die Abbildung 4.17 dar.

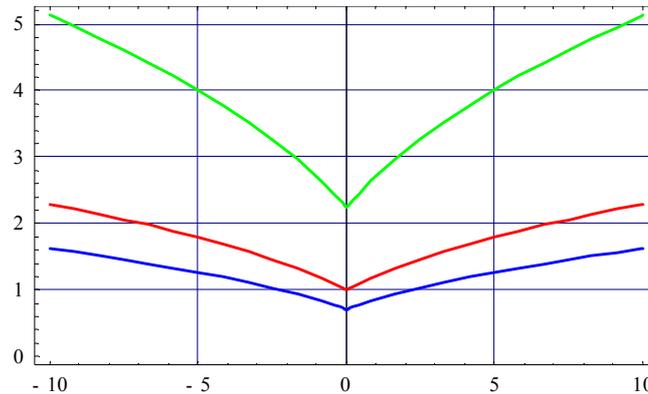


Abb. 4.17: Standardabweichung einer $\Xi_{c,\sigma}(x)$ -verteilten Zufallsvariablen

4.4-6 Schiefe einer $\Xi_{c,\sigma}(x)$ -verteilten Zufallsvariablen

Die Herleitung des Schiefemaßes η_3 wurde bereits im Abschnitt 4.3-6, Seite 88 beschrieben, daher wird hier nur das Endresultat für die Dichtefunktion $\xi_{c,\sigma}(x)$ angegeben:

$${}_{(c \geq 0)} \eta_3 = \frac{\sqrt{2}c(4c^2 + \pi(\sqrt{1+c}(2+c) - c^2 + 2c + 2))}{(1 + \sqrt{1+c})^3(-4 - \sqrt{1+c}(\pi - 4) + c\pi - 2c + 2\pi)^{3/2}} \quad (4.152)$$

Die analoge Ableitung liefert

$${}_{(c < 0)} \eta_3 = -\frac{\sqrt{2}c(-4c^2 + \pi(\sqrt{1-c}(c-2) + c^2 + 2c - 2))}{(1 + \sqrt{1-c})^3(-4 - \sqrt{1-c}(\pi - 4) - c\pi + 2c + 2\pi)^{3/2}} \quad (4.153)$$

Für η_3 gilt damit:

$$\eta_3 = \begin{cases} \frac{\sqrt{2}c(-4c^2 + \pi(\sqrt{1-c}(c-2) + c^2 + 2c - 2))}{(1 + \sqrt{1-c})^3(-4 - \sqrt{1-c}(\pi - 4) - c\pi + 2c + 2\pi)^{3/2}} & \text{für } c < 0 \\ \frac{\sqrt{2}c(4c^2 + \pi(\sqrt{1+c}(2+c) - c^2 + 2c + 2))}{(1 + \sqrt{1+c})^3(-4 - \sqrt{1+c}(\pi - 4) + c\pi - 2c + 2\pi)^{3/2}} & \text{für } c \geq 0 \end{cases} \quad (4.154)$$

Es wird deutlich, dass η_3 nicht von σ abhängt. Da die Dichtefunktion $\xi_c(x)$ mit der Maßgabe erweitert wurde, dass eine richtungsunabhängige Veränderung ermöglicht wird, ist dies jedoch nicht verwunderlich. Der Verlauf des Schiefemaßes η_3 in Abhängigkeit von c ist in Abbildung 4.18 dargestellt.

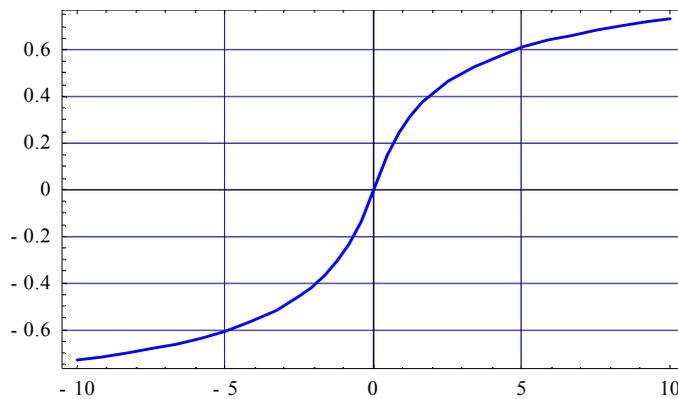


Abb. 4.18: Verlauf des Schiefemaßes η_3

Im Vergleich hierzu das Schiefemaß m_3 , das nicht auf den Erwartungswert bezogen ist und nicht bezüglich der Varianz normiert ist:

$$m_3 = \begin{cases} \frac{2c(2-c)\sqrt{2\sigma^3}}{\sqrt{\pi}(1 + \sqrt{1-c})} & \text{für } c < 0 \\ \frac{2c(2+c)\sqrt{2\sigma^3}}{\sqrt{\pi}(1 + \sqrt{1+c})} & \text{für } c \geq 0 \end{cases} \quad (4.155)$$

Der Verlauf von m_3 wird in der Abbildung 4.19 für festes $c = 2$ und $\sigma_1 = 0.5$ (blau), $\sigma_2 = 1$ (rot) und $\sigma_3 = 5$ (grün) dargestellt.

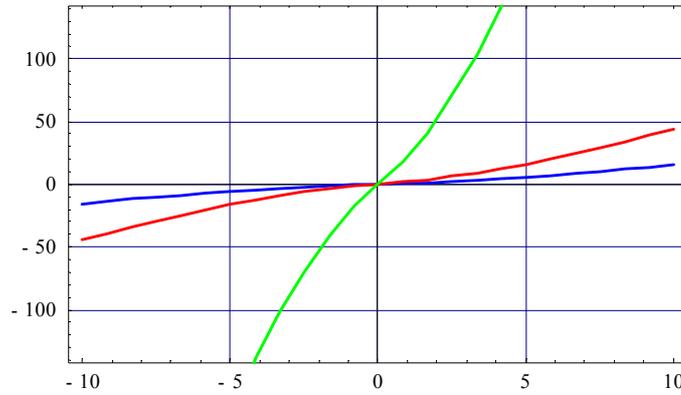


Abb. 4.19: Verlauf des Schiefemaßes m_3

Als Ergänzung sei hier auch das Schiefemaß μ_3 angegeben. Für das Schiefemaß μ_3 gilt:

$${}_{(c \geq 0)} \mu_3 = \frac{\sqrt{2}c(4c^2 + \pi(\sqrt{1+c}(2+c) - c^2 + 2c + 2))\sigma^{3/2}}{(1 + \sqrt{1+c})^3 \pi^{3/2}} \quad (4.156)$$

$${}_{(c < 0)} \mu_3 = -\frac{\sqrt{2}c(-4c^2 + \pi(\sqrt{1-c}(-2+c) - c^2 + 2c + 2))\sigma^{3/2}}{(1 + \sqrt{1-c})^3 \pi^{3/2}} \quad (4.157)$$

4.5 Asymmetrische Verteilungsfunktion $\Xi_{c,\sigma}^\gamma(x)$

Bei der späteren Verwendung der in diesem Kapitel eingeführten Verteilungsfunktionen ist eine globale Steuerung wünschenswert, mit der eine über die Parameter c und σ hinausgehende Einflussnahme auf den Verlauf der Verteilungsfunktion möglich ist [Hil96, Hil97b, Hil99b]. Insbesondere die Tatsache, dass das Schiefemaß η_3 einer $\Xi_{c,\sigma}(x)$ -verteilten Zufallsvariable nicht von dem Parameter σ abhängt, kann so kompensiert werden. Die konkrete Verwendung der Verteilungsfunktionen wird im Kapitel 6 (Asymmetrische Evolutionsstrategien, Seite 167) diskutiert, in dem die Parameter c , σ und γ den Strategieparametern einer Evolutionsstrategie gegenübergestellt werden. An dieser Stelle seien daher nur die mathematischen bzw. stochastischen Eigenschaften einer $\Xi_{c,\sigma}^\gamma(x)$ -verteilten Zufallsvariablen diskutiert.

Zur Entwicklung einer Verteilungsfunktion mit globalem Parameter $\gamma \in \mathbb{R}_+$ wird auch in diesem Abschnitt von der zugehörigen Dichtefunktion ausgegangen. Die Grundlage bildet die Dichtefunktion $\xi_{c,\sigma}(x)$ nach (4.121), die um den Parameter $\gamma \in \mathbb{R}_+$ wie folgt erweitert wird:

$$\xi_{c,\sigma}^\gamma(x) = \begin{cases} {}_{(c < 0)} \varphi(x) & \text{für } c < 0 \\ {}_{(c \geq 0)} \varphi(x) & \text{für } c \geq 0 \end{cases} \quad (4.158)$$

Für die Funktionen $_{(c < 0)}\varphi(x)$ und $_{(c \geq 0)}\varphi(x)$ gilt:

$$_{(c < 0)}\varphi(x) = \begin{cases} \frac{\sqrt{2}}{\sqrt{\pi\sigma^\gamma(1+(1-c)^\gamma)}} e^{-\frac{1}{2} \frac{x^2}{(\sigma(1-c))^\gamma}} & \text{für } x < 0 \\ \frac{\sqrt{2}}{\sqrt{\pi\sigma^\gamma(1+(1-c)^\gamma)}} e^{-\frac{1}{2} \frac{x^2}{\sigma^\gamma}} & \text{für } x \geq 0 \end{cases} \quad (4.159)$$

$$_{(c \geq 0)}\varphi(x) = \begin{cases} \frac{\sqrt{2}}{\sqrt{\pi\sigma^\gamma(1+(1+c)^\gamma)}} e^{-\frac{1}{2} \frac{x^2}{\sigma^\gamma}} & \text{für } x < 0 \\ \frac{\sqrt{2}}{\sqrt{\pi\sigma^\gamma(1+(1+c)^\gamma)}} e^{-\frac{1}{2} \frac{x^2}{(\sigma(1+c))^\gamma}} & \text{für } x \geq 0 \end{cases} \quad (4.160)$$

Der Verlauf der Dichtefunktion $\xi_{c,\sigma}^\gamma(x)$ wird in der folgenden Abbildung 4.20 für feste $c = 2$, $\sigma = 2$ und $\gamma_1 = 0.5$ (blau), $\gamma_2 = 1$ (rot) sowie $\gamma_3 = 1.5$ (grün) dargestellt.

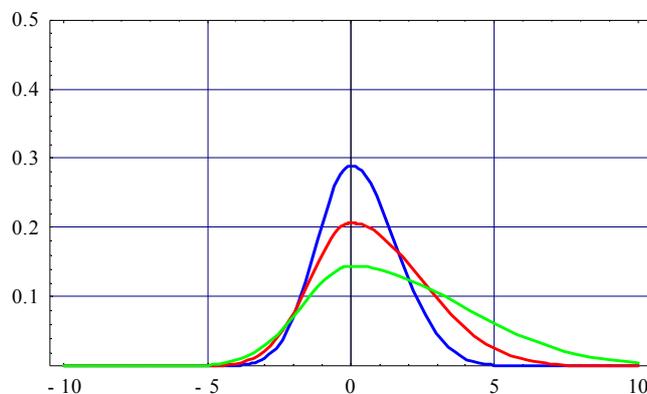


Abb. 4.20: Verlauf der Dichtefunktion $\xi_{c,\sigma}^\gamma(x)$ für unterschiedliche γ

4.5-1 Verteilungsfunktion einer $\Xi_{c,\sigma}^\gamma(x)$ -verteilten Zufallsvariablen

Die Berechnung der Verteilungsfunktion erfolgt entsprechend des Abschnittes 4.3 (Asymmetrische Verteilungsfunktion $\Xi_c(x)$, Seite 72) mittels des Ansatzes

$$\Phi_X(x) = \int_{-\infty}^x \varphi_X(t) dt \quad (4.161)$$

Angewandt auf die abschnittsweise Definition von $\xi_{c,\sigma}^\gamma(x)$ folgt daraus:

$$\Xi_{c,\sigma}^\gamma(x) = \begin{cases} x & \\ \int_{-\infty}^x \xi_{c,\sigma}^\gamma(t) dt & \text{für } x < 0 \\ 0 & \\ \int_{-\infty}^x \xi_{c,\sigma}^\gamma(t) dt + \int_0^x \xi_{c,\sigma}^\gamma(t) dt & \text{für } x \geq 0 \\ -\infty & 0 \end{cases} \quad (4.162)$$

Das Einsetzen der Definitionen für $\xi_{c,\sigma}^\gamma(x)$ führt im Fall $c \geq 0$ zu folgender Verteilungsfunktion ($c \geq 0$) $\Xi_{c,\sigma}^\gamma(x)$:

$${}_{(c \geq 0)} \Xi_{c,\sigma}^\gamma(x) = \begin{cases} \frac{1 + \operatorname{erf}\left(\frac{x}{\sqrt{2\sigma^\gamma}}\right)}{1 + \sqrt{(1+c)^\gamma}} & \text{für } x < 0 \\ \frac{1 + \sqrt{(1+c)^\gamma} \operatorname{erf}\left(\frac{x}{\sqrt{2(\sigma(1+c))^\gamma}}\right)}{1 + \sqrt{(1+c)^\gamma}} & \text{für } x \geq 0 \end{cases} \quad (4.163)$$

für den analogen Fall gilt:

$${}_{(c < 0)} \Xi_{c,\sigma}^\gamma(x) = \begin{cases} \frac{\sqrt{(1-c)^\gamma} \left(1 + \operatorname{erf}\left(\frac{x}{\sqrt{2(\sigma(1-c))^\gamma}}\right)\right)}{1 + \sqrt{(1-c)^\gamma}} & \text{für } x < 0 \\ \frac{\sqrt{(1-c)^\gamma} + \operatorname{erf}\left(\frac{x}{\sqrt{2\sigma^\gamma}}\right)}{1 + \sqrt{(1-c)^\gamma}} & \text{für } x \geq 0 \end{cases} \quad (4.164)$$

Für die allgemeine Verteilungsfunktion $\Xi_{c,\sigma}^\gamma(x)$ folgt daraus:

$$\Xi_{c,\sigma}^\gamma(x) = \begin{cases} {}_{(c < 0)} \Xi_{c,\sigma}^\gamma(x) & \text{für } c < 0 \\ {}_{(c \geq 0)} \Xi_{c,\sigma}^\gamma(x) & \text{für } c \geq 0 \end{cases} \quad (4.165)$$

Der Verlauf von $\Xi_{c,\sigma}^\gamma(x)$ wird in der Abbildung 4.13 für feste $c = 2$, $\sigma = 2$ und $\gamma_1 = 0.5$ (blau), $\gamma_2 = 1$ (rot) sowie $\gamma_3 = 1.5$ (grün) dargestellt.

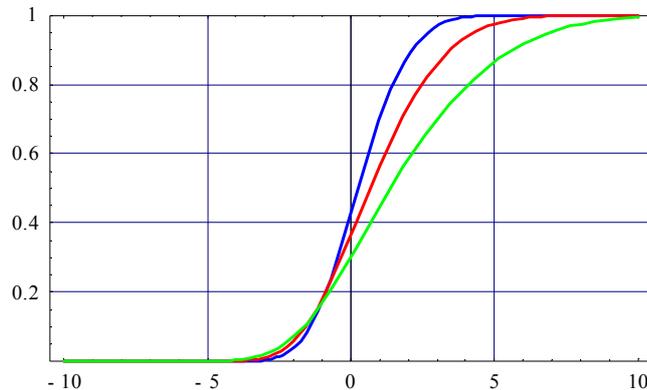


Abb. 4.21: Verlauf der Verteilungsfunktion $\Xi_{c,\sigma}^\gamma(x)$

4.5-2 $\Xi_{c,\sigma}^\gamma(x)$ als Mutationsfunktion

An dieser Stelle kann, analog zu dem Abschnitt 4.3-2 ($\Xi_c(x)$ als Mutationsfunktion, Seite 78), gezeigt werden, dass $\Xi_{c,\sigma}^\gamma(x)$ eine Mutationsfunktion im weiteren Sinne ist. Dieser Beweis ist nahezu identisch zu dem Beweis 4.2, lediglich bei den Berechnungen muss der neue Parameter γ berücksichtigt werden.

Satz 4.3: $\Xi_{c,\sigma}^\gamma(x)$ ist für beliebige $c \in \mathbb{R}$, $\sigma \in \mathbb{R}_+$ und $\gamma \in \mathbb{R}_+$ eine Mutationsfunktion im weiteren Sinne.

Beweis 4.3: Zu zeigen ist zum einen, dass $\Xi_{c,\sigma}^\gamma(x)$ eine Wahrscheinlichkeitsverteilungsfunktion ist, indem die in Abschnitt 4.1 (Eigenschaften von Wahrscheinlichkeitsverteilungen, Seite 64) aufgeführten Eigenschaften nachgewiesen werden. Zum anderen muss gezeigt werden, dass der einzige Modus der Funktion $\Xi_{c,\sigma}^\gamma(x)$ an der Stelle $x = 0$ existiert.

a. $\Xi_{c,\sigma}^\gamma(x)$ ist eine Wahrscheinlichkeitsverteilungsfunktion

- Die Eigenschaft $\lim_{x \rightarrow -\infty} \Xi_{c,\sigma}^\gamma(x) = 0$ und $\lim_{x \rightarrow \infty} \Xi_{c,\sigma}^\gamma(x) = 1$ kann aus der Definition der Verteilungsfunktion sowie der zugehörigen Dichtefunktion abgeleitet werden. Es gilt:

$$\lim_{x \rightarrow -\infty} \Phi_X(x) = \Phi_X(-\infty) = \int_{-\infty}^{-\infty} \varphi_X(t) dt \quad (4.166)$$

Zusätzlich kann aus der Eigenschaft der Exponentialfunktion direkt abgeleitet werden, dass

$$\lim_{x \rightarrow -\infty} \varphi_X(x) = \lim_{x \rightarrow -\infty} \xi_{c,\sigma}^\gamma(x) = 0 \quad (4.167)$$

für beliebige $c \in \mathbb{R}$, $\sigma \in \mathbb{R}_+$ und $\gamma \in \mathbb{R}_+$ gilt. Somit folgt

$$\int_{-\infty}^{-\infty} \xi_{c, \sigma}^{\gamma}(t) dt = 0 = \lim_{x \rightarrow -\infty} \Xi_{c, \sigma}^{\gamma}(x) \quad (4.168)$$

2. Die Eigenschaft $\lim_{x \rightarrow \infty} \Xi_{c, \sigma}^{\gamma}(x) = 1$ kann aus der allgemeinen Form

$$\lim_{x \rightarrow \infty} \Phi_X(x) = \Phi_X(\infty) = \int_{-\infty}^{\infty} \varphi_X(t) dt \quad (4.169)$$

abgeleitet werden. Es gilt für $c \geq 0$, $\sigma \in \mathbb{R}_+$ und $\gamma \in \mathbb{R}_+$:

$$\begin{aligned} \int_{-\infty}^{\infty} \varphi(x) dx &= \int_{-\infty}^0 \varphi_1(x) dx + \int_0^{\infty} \varphi_2(x) dx \\ &= \int_{-\infty}^0 \frac{\sqrt{2}}{\sqrt{\pi \sigma^{\gamma} (1 + (1+c)^{\gamma})}} e^{-\frac{1}{2} \frac{x^2}{\sigma^{\gamma}}} dx \\ &\quad + \int_0^{\infty} \frac{\sqrt{2}}{\sqrt{\pi \sigma^{\gamma} (1 + (1+c)^{\gamma})}} e^{-\frac{1}{2} \frac{x^2}{(\sigma(1+c))^{\gamma}}} dx \\ &= \frac{1}{1 + \sqrt{(1+c)^{\gamma}}} + \frac{\sqrt{(1+c)^{\gamma}}}{1 + \sqrt{(1+c)^{\gamma}}} \\ &= 1 \end{aligned} \quad (4.170)$$

Die analoge Berechnung für $c < 0$ führt zu dem gleichen Resultat. Damit gilt für beliebige c , $\sigma \in \mathbb{R}_+$ und $\gamma \in \mathbb{R}_+$:

$$\int_{-\infty}^{\infty} \xi_{c, \sigma}^{\gamma}(t) dt = \lim_{x \rightarrow \infty} \Xi_{c, \sigma}^{\gamma}(x) = 1 \quad (4.171)$$

3. Die Eigenschaft $\lim_{\delta \rightarrow 0} \Phi_X(x + \delta) = \Phi_X(x)$, also die rechtsseitige Stetigkeit kann direkt aus der entsprechenden Eigenschaft der Exponentialfunktion abgeleitet werden. Zusätzlich ist zu zeigen, dass an der Stelle $x = 0$ die Stetigkeit der Exponentialfunktion erhalten bleibt. Für diesen Beweis wird auf den Abschnitt 4.7-4 (Stetigkeit, Seite 131) verwiesen, in dem die rechts- und linksseitige Stetigkeit nachgewiesen wird.

b. $\Xi_{c, \sigma}^{\gamma}(x)$ ist eine Mutationsfunktion

1. Die strenge Monotonie einer Verteilungsfunktion $\Phi_X(x)$ kann über die Dichtefunktion gezeigt werden. Es muss gelten:

$$\xi_{c,\sigma}^\gamma(x) > 0 \text{ für alle } x \in \mathbb{R} \quad (4.172)$$

Diese Eigenschaft folgt direkt aus der Eigenschaft der Exponentialfunktion, für die gilt:

$$ae^{-bx^2} > 0, \text{ für } a, b > 0. \quad (4.173)$$

Aus der Definition der Funktion $\xi_c(x)$ folgt, dass die Bedingung $a, b > 0$ für alle $c \in \mathbb{R}$, $\sigma \in \mathbb{R}_+$ und $\gamma \in \mathbb{R}_+$ erfüllt ist. ■

2. Als Maß für die Skalierbarkeit kann die Varianz einer Zufallsvariablen verwendet werden. Im Abschnitt 4.5-5 (Varianz einer $\Xi_{c,\sigma}^\gamma(x)$ -verteilten Zufallsvariablen, Seite 111) wird die Varianz bestimmt. Hieraus folgt unmittelbar die Skalierbarkeit. ■
3. Die Forderung nach endogenen Richtungsparametern kann durch die Verteilungsfunktion allein nicht bewertet werden, da es sich hier um eine anwendungsabhängige Eigenschaft handelt. Im Vorgriff auf das Kapitel 6 (Asymmetrische Evolutionsstrategien, Seite 167) sei an dieser Stelle angemerkt, dass die asymmetrischen Evolutionsstrategien die Richtungsparameter ausschließlich als endogene Parameter verwenden. ■
4. Für die Lage des Modus ist zu zeigen, dass er für beliebige c und beliebige positive σ und γ einzig an der Stelle $x = 0$ existiert. Die Differenzierbarkeit wird im Abschnitt 4.7-5 (Differenzierbarkeit, Seite 133) untersucht. Hier sei nur das Resultat angegeben. Notwendige Bedingung für ein Maximum ist

$$\frac{d}{dx} \xi_{c,\sigma}^\gamma(x^*) = 0 \quad (4.174)$$

Diese Bedingung gilt für beliebige c ausschließlich an der Stelle $x^* = 0$. Diese Eigenschaft folgt, da $\xi_c(x)$ auf einer Exponentialfunktion basiert und keine translatorische Anteile enthält, die eine Verschiebung des Optimums entlang der x -Achse bewirken. Aus der Eigenschaft der Exponentialfunktion folgt zudem

$$\frac{d^2}{dx^2} \xi_{c,\sigma}^\gamma(x^*) < 0 \quad (4.175)$$

für beliebige $c \in \mathbb{R}$, $\sigma \in \mathbb{R}_+$ und $\gamma \in \mathbb{R}_+$. Daraus folgt, dass das einzige Optimum ein Maximum an der Stelle $x^* = 0$ ist. ■

Die Verteilungsfunktion $\Xi_{c,\sigma}^\gamma(x)$ erfüllt damit, ebenso wie die Verteilungsfunktion $\Xi_{c,\sigma}(x)$, die Forderungen 1 (Erreichbarkeit), 2 (Skalierbarkeit) und 4 (Stelle des Modus) der Definition 4.6 der Mutationsfunktion im weiteren Sinne. Die Forderung 3 (endogene Richtungsparameter) kann nicht von der Verteilungsfunktion realisiert werden, sondern nur von der Variante der Evolutionsstrategie, ■

in der sie verwendet wird. Die in dieser Arbeit entwickelte asymmetrische Evolutionsstrategie erfüllt diese Forderung, sodass die Verteilungsfunktion $\Xi_{c,\sigma}^\gamma(x)$ alle vier Forderungen an eine Mutationsfunktion im weiteren Sinne erfüllt.

4.5-3 Inverse Verteilungsfunktion einer $\Xi_{c,\sigma}^\gamma(x)$ -verteilten Zufallsvariablen

Zur Bestimmung der inversen Verteilungsfunktion muss entsprechend dem Abschnitt 4.3-3, Seite 80, der Wert für y in

$$\Xi_{c,\sigma}^\gamma(y) = x \quad (4.176)$$

gefunden werden. Da die Verteilungsfunktion $\Xi_{c,\sigma}^\gamma(x)$ nach (4.165) abschnittsweise definiert ist, folgt daraus eine abschnittsweise Definition der inversen Verteilungsfunktion. Die Grenze der beiden Abschnitte liegt an der Stelle $x = 0$:

$$\Xi_{c,\sigma}^\gamma(0) = y_{Grenze} \quad (4.177)$$

bzw. unter Verwendung der inversen Verteilungsfunktion:

$$\Xi_{c,\sigma}^\gamma(y_{Grenze}) = 0 \quad (4.178)$$

Angewandt auf die Verteilungsfunktion $\Xi_{c,\sigma}^\gamma(x)$ folgt daraus für den Fall $c \geq 0$:

$${}_{(c \geq 0)} \Xi_{c,\sigma}^\gamma(0) = \int_{-\infty}^0 \xi_{c,\sigma}^\gamma(t) dt = \frac{1}{1 + \sqrt{(1+c)^\gamma}} \quad (4.179)$$

Analog gilt:

$${}_{(c < 0)} \Xi_{c,\sigma}^\gamma(0) = \int_{-\infty}^0 \xi_{c,\sigma}^\gamma(t) dt = \frac{\sqrt{(1-c)^\gamma}}{1 + \sqrt{(1-c)^\gamma}} \quad (4.180)$$

Durch (4.179) und (4.180) werden die Grenzen der abschnittweisen Definition der inversen Verteilungsfunktion festgelegt. Unter Auslassung der Zwischenschritte resultiert die Berechnung in folgender Definition für die inverse Verteilungsfunktion:

$${}_{(c \geq 0)} \Xi_{c,\sigma}^\gamma(y) = \begin{cases} \sqrt{2\sigma^\gamma} \operatorname{inverf} \left(y(1 + \sqrt{(1+c)^\gamma}) - 1 \right) & \text{für } y < \frac{1}{1 + \sqrt{(1+c)^\gamma}} \\ \sqrt{2(\sigma(1+c))^\gamma} \operatorname{inverf} \left(y + \frac{y-1}{\sqrt{(1+c)^\gamma}} \right) & \text{für } y \geq \frac{1}{1 + \sqrt{(1+c)^\gamma}} \end{cases} \quad (4.181)$$

Die analoge Herleitung führt zu:

$${}_{(c < 0)} \bar{\Xi}_{c,\sigma}^\gamma(y) = \begin{cases} \sqrt{2(\sigma(1-c))^\gamma} \operatorname{inverf} \left(y \left(1 + \frac{1}{\sqrt{(1-c)^\gamma}} \right) - 1 \right) & \text{für } y < \frac{\sqrt{(1-c)^\gamma}}{1 + \sqrt{(1-c)^\gamma}} \\ \sqrt{2\sigma^\gamma} \operatorname{inverf} (y(1 + \sqrt{1-c}) - \sqrt{(1-c)^\gamma}) & \text{für } y \geq \frac{\sqrt{(1-c)^\gamma}}{1 + \sqrt{(1-c)^\gamma}} \end{cases} \quad (4.182)$$

Für die inverse Verteilungsfunktion gilt dann allgemein:

$$\bar{\Xi}_{c,\sigma}^\gamma(x) = \begin{cases} {}_{(c < 0)} \bar{\Xi}_{c,\sigma}^\gamma(x) & \text{für } c < 0 \\ {}_{(c \geq 0)} \bar{\Xi}_{c,\sigma}^\gamma(x) & \text{für } c \geq 0 \end{cases} \quad (4.183)$$

Den Verlauf der inversen Verteilungsfunktion $\bar{\Xi}_{c,\sigma}^\gamma(x)$ für feste $c = 2$, $\sigma = 2$ und $\gamma_1 = 0.5$ (blau), $\gamma_2 = 1$ (rot) sowie $\gamma_3 = 1.5$ (grün) ist in Abbildung 4.22 dargestellt.

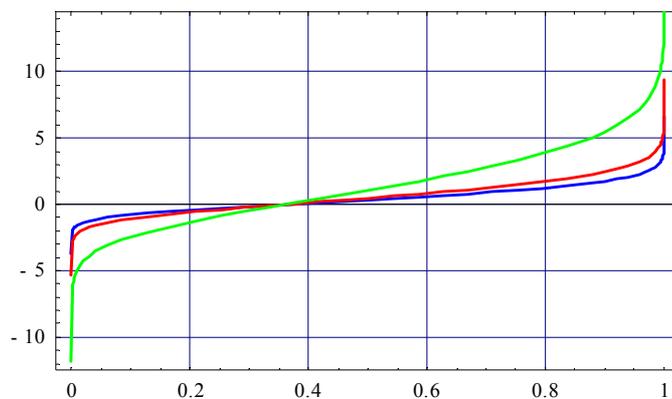


Abb. 4.22: Verlauf der inversen Verteilungsfunktion $\bar{\Xi}_{c,\sigma}^\gamma(y)$ für variierende γ

4.5-4 Erwartungswert einer $\Xi_{c,\sigma}^\gamma(x)$ -verteilten Zufallsvariablen

Die Berechnung des Erwartungswertes erfolgt auch bei einer Zufallsvariablen X mit Dichtefunktion $\xi_{c,\sigma}^\gamma(x)$ mittels der partiellen Erwartungswerte nach Definition 4.5. Für den Fall $c \geq 0$ gilt:

$$\begin{aligned}
(c \geq 0) E_X &= (c \geq 0) E_X \Big|_{-\infty}^0 + (c \geq 0) E_X \Big|_0^{\infty} \\
&= \int_{-\infty}^0 x (c \geq 0) \xi_{c, \sigma}^{\gamma}(x) dx + \int_0^{\infty} x (c \geq 0) \xi_{c, \sigma}^{\gamma}(x) dx \\
&= -\frac{\sqrt{2\sigma^{\gamma}}}{\sqrt{\pi}(1 + (1+c)^{\gamma})} + \frac{\sqrt{2\sigma^{\gamma}}(1+c)}{\sqrt{\pi}(1 + (1+c)^{\gamma})} \\
&= \frac{c\sqrt{2\sigma^{\gamma}}}{\sqrt{\pi}(1 + \sqrt{1+c})}
\end{aligned} \tag{4.184}$$

die analoge Herleitung für $c < 0$ resultiert in:

$$\begin{aligned}
(c < 0) E_X &= (c < 0) E_X \Big|_{-\infty}^0 + (c < 0) E_X \Big|_0^{\infty} \\
&= \int_{-\infty}^0 x (c < 0) \xi_{c, \sigma}^{\gamma}(x) dx + \int_0^{\infty} x (c < 0) \xi_{c, \sigma}^{\gamma}(x) dx \\
&= \frac{\sqrt{2\sigma^{\gamma}}(1-c)}{\sqrt{\pi}(1 + (1-c)^{\gamma})} - \frac{\sqrt{2\sigma^{\gamma}}}{\sqrt{\pi}(1 + (1-c)^{\gamma})} \\
&= \frac{c\sqrt{2\sigma^{\gamma}}}{\sqrt{\pi}(1 + (1-c)^{\gamma})}
\end{aligned} \tag{4.185}$$

Für $E(X)$ gilt damit allgemein:

$$E_X = \frac{c\sqrt{2\sigma^{\gamma}}}{\sqrt{\pi}(1 + (1 + |c|)^{\gamma})} \tag{4.186}$$

Die Abbildung 4.23 zeigt den Verlauf des Erwartungswertes E_X einer $\Xi_{c, \sigma}^{\gamma}(x)$ -verteilten Zufallsvariablen X , für feste $c = 2$, $\sigma = 2$ und $\gamma_1 = 0.5$ (blau), $\gamma_2 = 1$ (rot) sowie $\gamma_3 = 1.5$ (grün) dargestellt.

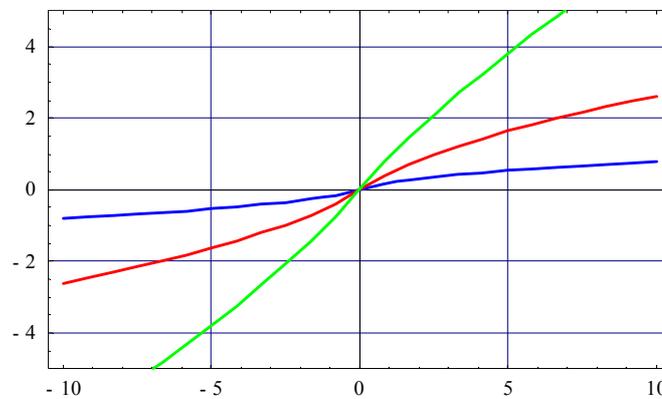


Abb. 4.23: Verlauf des Erwartungswertes einer $\Xi_{c, \sigma}^{\gamma}(x)$ -verteilten Zufallsvariablen

4.5-5 Varianz einer $\Xi_{c,\sigma}^\gamma(x)$ -verteilten Zufallsvariablen

Die Varianz V_X einer $\Xi_{c,\sigma}^\gamma(x)$ -verteilten Zufallsvariable X kann entsprechend der Varianz einer $\Xi_c(x)$ -verteilten Zufallsvariablen berechnet werden. Unter Auslassung der Zwischenschritte, die denen in (4.95) und (4.96) entsprechen, gilt:

$$V_X = \begin{cases} \sigma^\gamma \left((1 - \sqrt{(1-c)^\gamma} + (1-c)^\gamma) - \frac{2(\sqrt{(1-c)^\gamma} - 1)^2}{\pi} \right) & \text{für } c < 0 \\ \sigma^\gamma \left((1 - \sqrt{(1+c)^\gamma} + (1+c)^\gamma) - \frac{2(\sqrt{(1+c)^\gamma} - 1)^2}{\pi} \right) & \text{für } c \geq 0 \end{cases} \quad (4.187)$$

Allgemein gilt damit:

$$V_X = \sigma^\gamma \left((1 - \sqrt{(1+|c|)^\gamma} + (1+|c|)^\gamma) - \frac{2(\sqrt{(1+|c|)^\gamma} - 1)^2}{\pi} \right) \quad (4.188)$$

Den Verlauf der Varianz einer $\Xi_{c,\sigma}^\gamma(x)$ -verteilten Zufallsvariablen X , für feste $c = 2$, $\sigma = 2$ und $\gamma_1 = 0.5$ (blau), $\gamma_2 = 1$ (rot) sowie $\gamma_3 = 1.5$ (grün) stellt Abbildung 4.24 dar.

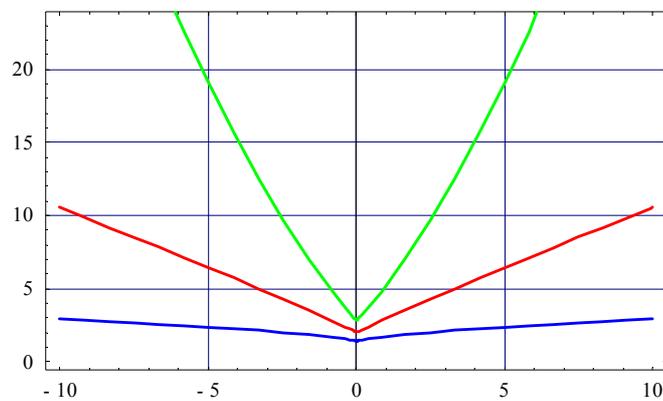


Abb. 4.24: Varianz einer $\Xi_{c,\sigma}^\gamma(x)$ -verteilten Zufallsvariablen

Den Verlauf der eng mit der Varianz verbundenen Standardabweichung $SD_X = \sqrt{V_X}$, für feste $c = 2$, $\sigma = 2$ und $\gamma_1 = 0.5$ (blau), $\gamma_2 = 1$ (rot) sowie $\gamma_3 = 1.5$ (grün) stellt die Abbildung 4.25 dar.

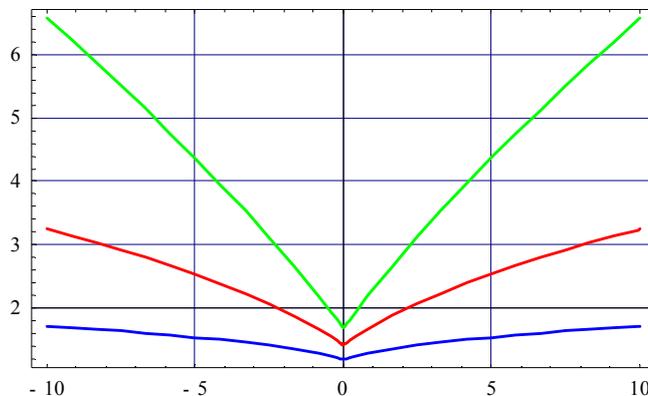


Abb. 4.25: Standardabweichung einer $\Xi_{c,\sigma}^{\gamma}(x)$ -verteilten Zufallsvariablen

4.5-6 Schiefe einer $\Xi_{c,\sigma}^{\gamma}(x)$ -verteilten Zufallsvariablen

Die Herleitung des Schiefemaßes η_3 wurde bereits im Abschnitt 4.3-6, Seite 88 beschrieben, daher wird hier nur das Endresultat für die Dichtefunktion $\xi_{c,\sigma}^{\gamma}(x)$ angegeben:

$${}_{(c \geq 0)} \eta_3 = \frac{\sqrt{2}(-1 + \sqrt{(1+c)^{\gamma}})(-4 + \sqrt{(1+c)^{\gamma}}(8-3\pi) + (1+c)^{\gamma}(-4+\pi) + \pi)}{\sqrt{(-2 - \sqrt{(1+c)^{\gamma}}(\pi-4) + (1+c)^{\gamma}(\pi-2) + \pi)^3}} \quad (4.189)$$

Die analoge Ableitung liefert

$${}_{(c < 0)} \eta_3 = -\frac{\sqrt{2}(-1 + \sqrt{(1-c)^{\gamma}})(-4 + \sqrt{(1-c)^{\gamma}}(8-3\pi) + (1-c)^{\gamma}(-4+\pi) + \pi)}{\sqrt{(-2 - \sqrt{(1-c)^{\gamma}}(\pi-4) + (1-c)^{\gamma}(\pi-2) + \pi)^3}} \quad (4.190)$$

Für η_3 gilt damit:

$$\eta_3 = \begin{cases} \frac{\sqrt{2}(-1 + \sqrt{(1-c)^{\gamma}})(-4 + \sqrt{(1-c)^{\gamma}}(8-3\pi) + (1-c)^{\gamma}(-4+\pi) + \pi)}{\sqrt{(-2 - \sqrt{(1-c)^{\gamma}}(\pi-4) + (1-c)^{\gamma}(\pi-2) + \pi)^3}} & \text{für } c < 0 \\ \frac{\sqrt{2}(-1 + \sqrt{(1+c)^{\gamma}})(-4 + \sqrt{(1+c)^{\gamma}}(8-3\pi) + (1+c)^{\gamma}(-4+\pi) + \pi)}{\sqrt{(-2 - \sqrt{(1+c)^{\gamma}}(\pi-4) + (1+c)^{\gamma}(\pi-2) + \pi)^3}} & \text{für } c \geq 0 \end{cases} \quad (4.191)$$

Auch hier wird deutlich, dass η_3 nicht von σ abhängt. Es gilt die gleiche Anmerkung wie im Fall der $\Xi_{c,\sigma}$ -Verteilung. Der Verlauf des Schiefemaßes η_3 in Abhängigkeit von c für $\sigma = 2$ und $\gamma_1 = 0.5$ (blau), $\gamma_2 = 1$ (rot) sowie $\gamma_3 = 1.5$ (grün) ist in Abbildung 4.26 dargestellt.

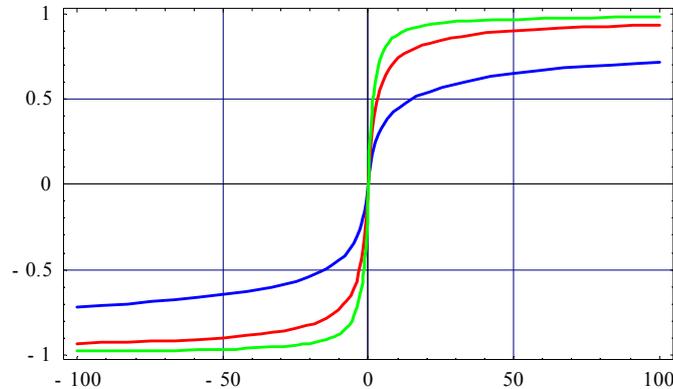


Abb. 4.26: Verlauf des Schiefemaßes η_3 für unterschiedliche γ

Im Vergleich hierzu das Schiefemaß m_3 , das nicht auf den Erwartungswert bezogen ist und nicht bezüglich der Varianz normiert ist:

$$m_3 = \begin{cases} -2(\sqrt{(1-c)^\gamma} - 1)((1-c)^\gamma + 1) \sqrt{\frac{2\sigma^{3\gamma}}{\pi}} & \text{für } c < 0 \\ 2(\sqrt{(1+c)^\gamma} - 1)((1+c)^\gamma + 1) \sqrt{\frac{2\sigma^{3\gamma}}{\pi}} & \text{für } c \geq 0 \end{cases} \quad (4.192)$$

Der Verlauf des Schiefemaßes m_3 in Abhängigkeit von c für $\sigma = 2$ und $\gamma_1 = 0.5$ (blau), $\gamma_2 = 1$ (rot) sowie $\gamma_3 = 1.5$ (grün) ist in Abbildung 4.27 dargestellt.

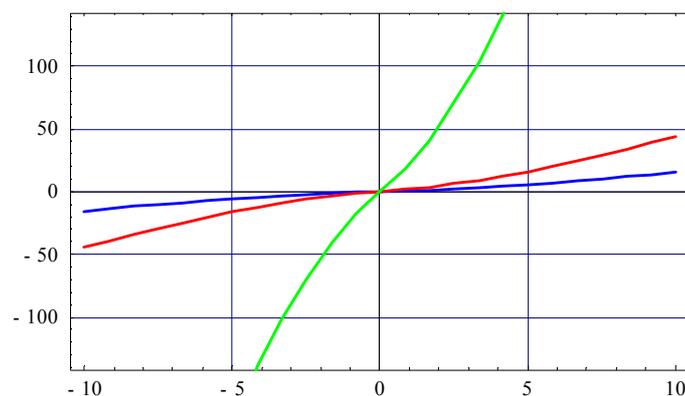


Abb. 4.27: Verlauf des Schiefemaßes m_3

Als Ergänzung sei auch hier das Schiefemaß μ_3 angegeben, für das gilt:

$$\mu_3 = \begin{cases} \frac{\sqrt{2}(-1 + \sqrt{(1-c)^\gamma})(-4 + \sqrt{(1-c)^\gamma}(8-3\pi) + (1-c)^\gamma(-4+\pi) + \pi)\sigma^{3\gamma/2}}{\sqrt{(-2 - \sqrt{(1-c)^\gamma}(\pi-4) + (1-c)^\gamma(\pi-2) + \pi)^3}} & \text{für } c < 0 \\ \frac{\sqrt{2}(-1 + \sqrt{(1+c)^\gamma})(-4 + \sqrt{(1+c)^\gamma}(8-3\pi) + (1+c)^\gamma(-4+\pi) + \pi)\sigma^{3\gamma/2}}{\sqrt{(-2 - \sqrt{(1+c)^\gamma}(\pi-4) + (1+c)^\gamma(\pi-2) + \pi)^3}} & \text{für } c \geq 0 \end{cases} \quad (4.193)$$

4.6 Anmerkungen zur Wahl der Grundfunktion

Bei der Herleitung der asymmetrischen Wahrscheinlichkeitsdichte- bzw. Verteilungsfunktionen wird in dieser Arbeit eine abschnittsweise definierte, auf e -Funktionen aufbauende Grundfunktion gewählt. Für diese Wahl gibt es Gründe, die in diesem Abschnitt erläutert werden. Zusätzlich wird in diesem Abschnitt auf alternative Grundfunktionen eingegangen sowie die Gründe angegeben, warum diese Alternativen in dieser Arbeit nicht verwendet werden.

4.6-1 Alternative Verteilungsfunktionen

Als Resultat einer Literaturrecherche über Verteilungsfunktionen erhält man eine Vielzahl nicht symmetrischer Verteilungsfunktionen. Aus dieser Vielzahl kommen jedoch nur Verteilungen in Frage, die zusätzlich zur Asymmetrie den für Evolutionsstrategien erforderlichen Definitionsbereich \mathbb{R} besitzen. Viele bekannte, asymmetrische Verteilungen fallen an dieser Stelle bereits aus der Auswahl aus. Hierzu gehören:

- Beta-Verteilung, mit Definitionsbereich $[0, 1]$,
- χ^2 -Verteilung, mit Definitionsbereich $[0, \infty)$,
- generalisierte Rayleigh-Verteilung, mit Definitionsbereich $(0, \infty)$,
- F-Verteilung, mit Definitionsbereich $[0, \infty)$,
- Gamma-Verteilung, mit Definitionsbereich $[0, \infty)$ und
- Weibull-Verteilung, mit Definitionsbereich $[0, \infty)$.

Eine der wenigen Verteilungsfunktionen, die asymmetrisch sind und den Definitionsbereich \mathbb{R} besitzen, ist die Gumbel-Verteilung [Eva93]:

$$\Phi_{\text{Gumbel}}(x) = e^{-e^{-\frac{x-a}{b}}} \quad (4.194)$$

mit dem Lageparameter $a \in \mathbb{R}$ und dem Skalierungsparameter $b \in \mathbb{R}_+$. Die zu (4.194) gehörende Dichtefunktion lautet:

$$\varphi_{\text{Gumbel}}(x) = \frac{1}{b} e^{-\frac{x-a}{b}} e^{-e^{-\frac{x-a}{b}}} \quad (4.195)$$

Abbildung 4.28 stellt den Verlauf von $\varphi_X(x)$ für $b_1 = 0.5$ (blau), $b_2 = 1$ (rot), und $b_3 = 2$ (grün) dar. Eine Veränderung des Lageparameters a bewirkt eine Translation um a auf der x -Achse.

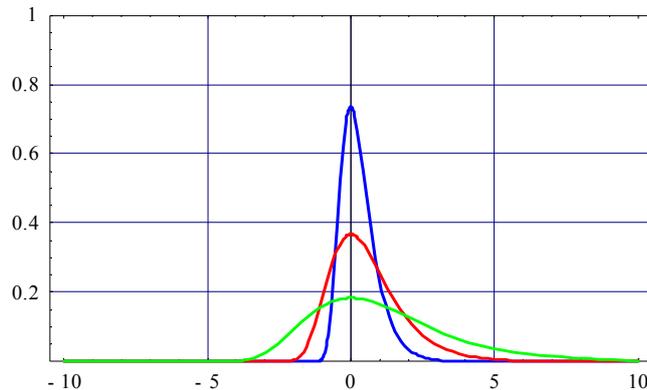


Abb. 4.28: Verlauf der Dichtefunktion einer Gumbel-verteilten Zufallsvariablen

Die Gumbel-Verteilung ist grundsätzlich rechtsschief, mit dem festen Schiefemaß $\eta_3 = 1.139547$. Für das Schiefemaß m_3 gilt:

$$m_3 = \frac{1}{2}b^3(2\gamma_{Euler}^3 + \pi^2\gamma_{Euler} + 4\zeta(3)) \quad (4.196)$$

mit der Eulerschen Konstante γ_{Euler} :

$$\gamma_{Euler} = \lim_{m \rightarrow \infty} \left(\sum_{k=1}^m \frac{1}{k} - \log(m) \right) = 0.5772157 \quad (4.197)$$

und der Riemanschen Zeta-Funktion $\zeta(x)$:

$$\zeta(x) = \lim_{m \rightarrow \infty} \sum_{k=1}^m \frac{1}{k^x} \quad (4.198)$$

Das Schiefemaß η_3 ist unabhängig vom Parameter b . Die Gumbel-Verteilung besitzt damit keine Möglichkeit, die Schiefe zu steuern. Der Parameter b ist nicht mit dem Parameter c der ξ -Verteilungen vergleichbar. Er ist eher mit dem Parameter σ der Ξ -Verteilungen vergleichbar, der eine symmetrische Änderung der Varianz bewirkt. Da die Gumbel-Verteilung grundsätzlich rechtsschief ist, muss für eine linksschiefe Variante eine Fallunterscheidung mit Transformation der Variablen x eingeführt werden. Damit lässt sie sich nicht mehr geschlossen darstellen und vereinfacht die notwendigen Berechnungen, verglichen mit den Ξ -Verteilungen, nicht. Aus diesen Gründen eignet sich die Gumbel-Verteilung nicht für den Einsatz in asymmetrischen Evolutionsstrategien.

4.6-2 Verteilungssysteme

Zusätzlich zu einzelnen, isoliert betrachteten Verteilungsfunktionen sind sog. Verteilungssysteme entwickelt worden. Ein Verteilungssystem ist eine Klasse eindimensionaler Verteilungen mit existierenden ersten vier Momenten. Sie basieren in der Regel auf einem zentralen Grundgedanken und

erlauben durch Variation einzelner Parameter oder durch Fallunterscheidung einzelner Kriterien die Erzeugung einer Vielzahl von Verteilungsfunktionen. Zu diesen Verteilungssystemen gehören das Pearson System, das Expansionssystem, die transformierten Verteilungen, das Burr System und das Toranzos System. Im folgenden wird das bekannteste dieser Systeme, das Pearson System ausführlich vorgestellt. Weiterführende Informationen über die anderen Systeme finden sich für das Expansionssystem in [Bru06, Edg07, Joh70], für die transformierten Verteilungen in [Edg16, Joh49], für das Burr System in [Bur42, Bur68] und für das Toranzos System in [Tor52].

Das Pearson System wurde von Pearson Ende des 19. Jahrhunderts entwickelt [Pea95]. Für jede Verteilungsfunktion in diesem System gehorcht die zugehörige Wahrscheinlichkeitsdichtefunktion $\varphi(x)$ der Differentialgleichung

$$\frac{d\varphi(x)}{dx} = -\varphi(x) \frac{a+x}{c_0 + c_1x + c_2x^2} \quad (4.199)$$

Die Gestalt einer Verteilungsfunktion des Pearson Systems wird somit von den Parametern a , c_0 , c_1 und c_2 bestimmt. Pearson hat auf Basis der Lösungen der Gleichung

$$c_0 + c_1x + c_2x^2 = 0 \quad (4.200)$$

zwölf Typen systematisiert und diese mit Typ I bis Typ XII bezeichnet. Häufig wird eine andere Notation verwendet:

$$\frac{d\varphi(x)}{dx} = \varphi(x) \frac{a_0 + a_1x}{C_0 + C_1x + C_2x^2} \quad (4.201)$$

für die die Ersetzungen $a = a_0/a_1$ und $c_i = C_i/a_1$, $a_1 = 1$ gelten. Der Spezialfall der Typ XII Verteilung kann dann in die allgemeinen Ableitung eingebettet werden.

Im Pearson System ist es möglich, für die gewünschte Maßzahlen der Schiefe $\gamma_1 = \eta_3$ und $\gamma_2 = \mu_4/\sigma^4 - 3$ die Koeffizienten der Gleichung (4.199) bzw. (4.201) zu bestimmen. Es gilt:

$$\begin{aligned} C_0 &= 4\gamma_2 - 3\gamma_1^2 + 12 \\ C_1 &= \gamma_1(\gamma_2 + 6) \\ C_2 &= 2\gamma_2 - 3\gamma_1^2 \\ a_0 &= -C_1 \\ a_1 &= -2(5\gamma_2 - 6\gamma_1^2 + 6) \end{aligned} \quad (4.202)$$

Der Beweis für diese Aussage findet sich in [Ras95]. Somit können die Koeffizienten der Differentialgleichung anhand γ_1 und γ_2 bestimmt werden. Andere Maßgrößen können nicht in Betracht gezogen werden. Für den Einsatz als Grundlage einer asymmetrischen Verteilungsfunktion, wie sie im Sinne einer Mutationsfunktion für asymmetrische Evolutionsstrategien verwendet wird, kommt die Bestimmung einer Verteilungsfunktion durch Lösung der entsprechenden Differentialgleichung daher nicht in Betracht. Im folgenden werden die einzelnen Typen kurz beschrieben und dann auf ihre Eignung als Grundlage einer Mutationsfunktion für die asymmetrischen Evolutionsstrategien bewertet.

Pearson Typ I Verteilungen: Eine Verteilung vom Typ I entsteht, wenn die beiden Nullstellen a_1 und a_2 der Gleichung (4.199) reellwertig und mit gegengesetztem Vorzeichen behaftet sind, d. h.:

$$a_1 < 0 < a_2 \quad (4.203)$$

Aus (4.203) folgt

$$c_0 + c_1x + c_2x^2 = -c_2(x - a_1)(a_2 - x) \quad (4.204)$$

Gleichung (4.199) kann dann wie folgt notiert werden

$$\frac{d}{dx} \log \varphi(x) = \frac{x + a}{c_2(x - a_1)(a_2 - x)} = \frac{1}{c_2(a_2 - a_1)} \left[\frac{a + a_1}{x - a_1} + \frac{a + a_2}{a_2 - x} \right] \quad (4.205)$$

Für die Dichtefunktion $\varphi(x)$ gilt damit die prinzipielle Form

$$\varphi(x) = k(x - a_1)^{m_1}(a_2 - x)^{m_2} \quad (4.206)$$

Für die Parameter m_1, m_2 gilt:

$$m_1 = \frac{a + a_1}{c_2(a_2 - a_1)} \quad (4.207)$$

$$m_2 = -\frac{a + a_2}{c_2(a_2 - a_1)} \quad (4.208)$$

Der Faktor k ist so zu wählen, dass

$$\int_{-\infty}^{\infty} \varphi(x) dx = 1 \quad (4.209)$$

gilt. Der Definitionsbereich in dieser Notation ist hierbei $a_1 < x < a_2$. Unter Auslassung von Umformungen, die in [Joh70] nachgelesen werden können, ergibt sich die von Pearson verwendete Schreibweise, die als Definitionsbereich das Intervall $[0, 1] \subset \mathbb{R}$ besitzt:

$$\varphi_I(x) = \frac{1}{B(m_1 + 1, m_2 + 1)} x^{m_1} (1 - x)^{m_2} \quad (4.210)$$

Die Gleichung (4.210) ist die allgemeine Form der Beta-Verteilung. Die Dichtefunktion der Beta-Verteilung wird mit den häufig verwendeten Parametern $p, q > 0$ wie folgt notiert:

$$\varphi_{Beta}(x) = \frac{1}{B(p, q)} x^{p-1} (1 - x)^{q-1}, \text{ für } 0 \leq x \leq 1. \quad (4.211)$$

Die Funktion $B(p, q)$ bezeichnet hier die Eulersche Beta-Funktion, für die gilt:

$$B(p, q) = \int_0^1 t^{p-1} (1-t)^{q-1} dt \quad (4.212)$$

Die folgenden Abbildungen geben einige Beispiele für typische Dichtefunktionen der Beta-Verteilung.

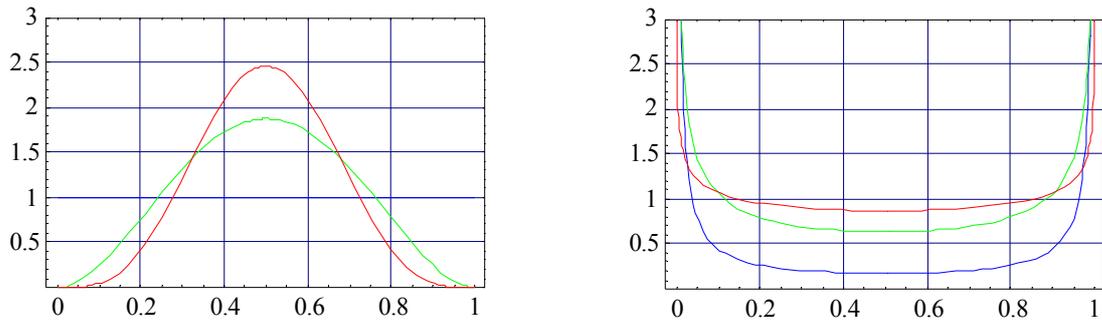


Abb. 4.29: Dichtefunktionen der Beta-Verteilung mit $p, q > 1$ und $p = q$ (links) und $p, q < 1$ (rechts)

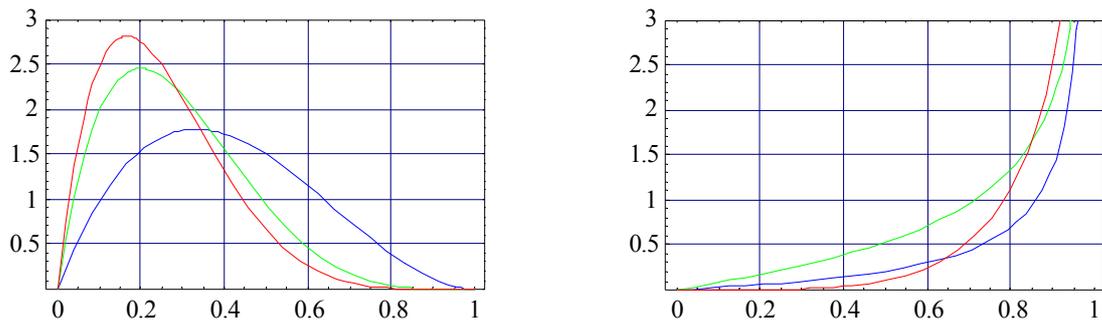


Abb. 4.30: Dichtefunktionen der Beta-Verteilung mit $p, q > 1$ (links) und $p > 1, q < 1$ (rechts)

Für $p > 1, q > 1$ und $p = q$ entstehen grundsätzlich glockenförmige Dichtefunktionen mit Maximum an der Stelle $x = 0.5$. Für $p = q = 1$ entsteht eine Gleichverteilung über dem Intervall $[0, 1] \subset \mathbb{R}$. Für $p < 1$ und $q < 1$ entstehen U-förmige Dichtefunktionen, die für $p = q$ symmetrisch um die Stelle $x = 0.5$ sind. Für den Fall $p > 1, q > 1$ entstehen asymmetrisch verzerrte glockenförmige Kurve. Für das Diagramm in Abbildung 4.30 rechts gilt $p < q$. Für $p > q$ entstehen Kurven, bei denen das Maximum in Richtung $x = 1$ verschoben ist. Für den Fall, dass einer der beiden Parameter kleiner, der andere größer als 1 ist, entstehen steigende bzw. fallende Kurven.

Anhand der allgemeinen Definition in (4.207) werden in Abhängigkeit der Parameter m_1 und m_2 weitere Typen unterschieden. Alle diese Typen sind Varianten der Beta-Verteilung und in dem obigen Abschnitt bereits erläutert. Bei diesen Varianten handelt es sich um Typ II ($m_1 = m_2$), Typ VIII ($m_1 < m_2, m_2 \neq 0$), Typ IX ($m_1 < m_2, m_1 = 0$) und Typ XII ($m_1 < 0 < m_2 < 1$). Sie werden im folgenden näher beschrieben.

Pearson Typ II Verteilungen: Eine Verteilung vom Typ II entsteht, wenn für die Parameter m_1 und m_2 nach (4.207) und (4.208) die Identität gilt, d. h.:

$$m_1 = m_2 \quad (4.213)$$

Für die Gleichungen (4.210) und (4.210) gilt dann:

$$\varphi_{II}(x) = \frac{1}{B(m+1, m+1)} x^m (1-x)^m \quad (4.214)$$

$$\varphi_{II}(x) = \frac{1}{B(p, p)} x^{p-1} (1-x)^{p-1} \quad (4.215)$$

Beispiele für Typ II Verteilungen sind in der folgenden Abbildung 4.31 dargestellt.

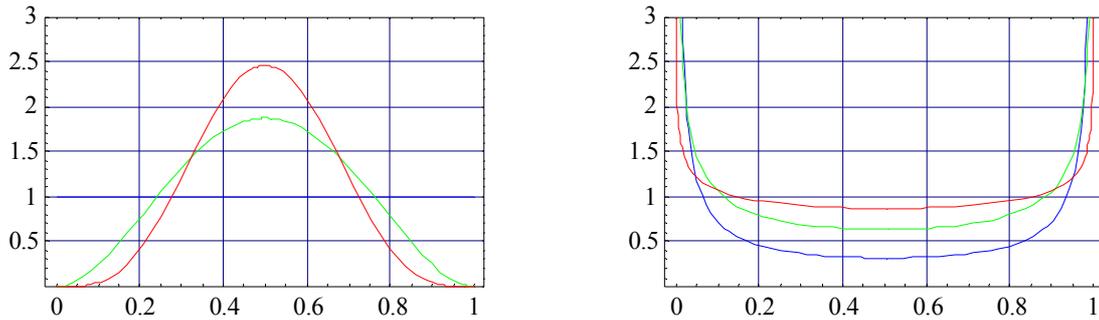


Abb. 4.31: Dichtefunktionen der Typ II Verteilung mit $p, q > 1$ (links) und $p, q < 1$ (rechts)

Pearson Typ III Verteilungen: Eine Verteilung vom Typ III entsteht, wenn für die Parameter der Gleichung (4.199) $c_2 = 0$ und $c_1 \neq 0$ gilt. In diesem Fall gilt in Analogie zu (4.205):

$$\begin{aligned} \frac{d}{dx} \log \varphi(x) &= -\frac{a+x}{c_0 + c_1 x} \\ &= -\frac{1}{c_1} - \frac{a - c_0/c_1}{c_0 + c_1 x} \end{aligned} \quad (4.216)$$

Für die Dichtefunktion $\varphi(x)$ gilt damit die prinzipielle Form

$$\varphi_{III}(x) = k(c_0 + c_1 x)^m e^{-x/c_1} \quad (4.217)$$

mit

$$m = \frac{1}{c_1} \left(\frac{c_0}{c_1} - a \right) \quad (4.218)$$

Der Faktor k ist so zu wählen, dass

$$\int_{-\infty}^{\infty} \varphi_{III}(x) dx = 1 \quad (4.219)$$

gilt. Der Definitionsbereich in dieser Notation ist für den Fall $c_1 > 0$ $x > -c_0/c_1$ und für den Fall $c_1 < 0$ $x < -c_0/c_1$. Die Form nach Gleichung (4.217) ist die allgemeine Form der Gamma-Verteilung. Die Gamma-Verteilung wird häufig in einer anderen Form definiert, bei der der Definitionsbereich verschoben und der Skalierungsfaktor k explizit angegeben wird:

$$\varphi_{Gamma}(x) = \frac{1}{\Gamma(\lambda)} x^{\lambda-1} e^{-x} \quad (4.220)$$

mit $\lambda > 0$ und $0 \leq x < \infty$. Die Abbildung 4.32 gibt einige Beispiele für den typischen Verlauf der Dichtefunktion unterschiedlicher Gamma-Verteilungen:

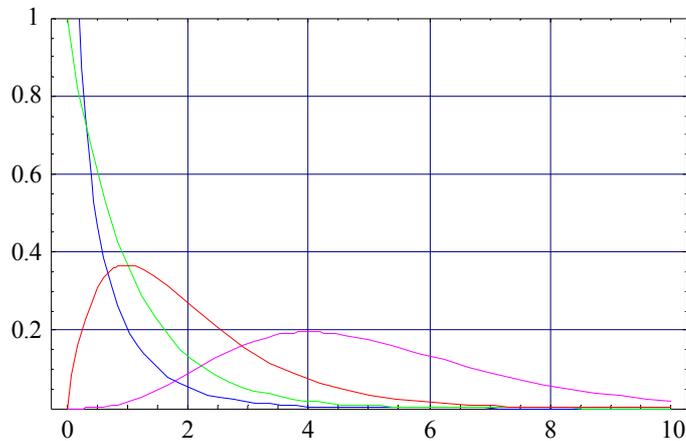


Abb. 4.32: Dichtefunktionen der Gamma-Verteilungen mit variierendem Parameter λ

Pearson Typ IV Verteilungen: Eine Verteilung vom Typ IV entsteht, wenn die Nullstellen der Gleichung

$$c_0 + c_1x + c_2x^2 = 0 \quad (4.221)$$

imaginär sind. In diesem Fall wird folgende Ersetzung durchgeführt:

$$c_0 + c_1x + c_2x^2 = C_0 + c_2(x + C_1)^2 \quad (4.222)$$

$$C_0 = c_0 - \frac{1}{4}c_1^2 \frac{1}{c_2} \quad (4.223)$$

$$C_1 = \frac{1}{2} \frac{c_1}{c_2} \quad (4.224)$$

Es gilt in Analogie zu (4.205):

$$\frac{d}{dx} \log \varphi(x) = \frac{-(x + C_1) - (a - C_1)}{C_0 + c_2(x + C_1)^2} \quad (4.225)$$

Für die Dichtefunktion $\varphi(x)$ gilt damit die prinzipielle Form

$$\varphi(x) = k [C_0 + c_2(x + C_1)^2]^{-\frac{1}{2c_2}} e^{\left[\frac{a - C_1}{\sqrt{c_2} C_0} \arctan \left(\frac{x + C_1}{\sqrt{C_0/c_2}} \right) \right]} \quad (4.226)$$

Dargestellt in der von Pearson verwendeten Notation gilt:

$$\varphi_{IV}(x) = k \left(1 + \frac{x^2}{a^2} \right)^{-m} e^{-v \cdot \arctan(x/a)} \quad (4.227)$$

In der Statistik existiert keine bekannte Verteilungsfunktion, die dem Pearson Typ IV entspricht. Aus diesem Grund werden hier nur allgemein einige mögliche Verläufe von Typ IV Verteilungen dargestellt. Es handelt sich hierbei nicht um Dichtefunktionen, da der Faktor k in diesen Beispielen nicht bestimmt wurde.

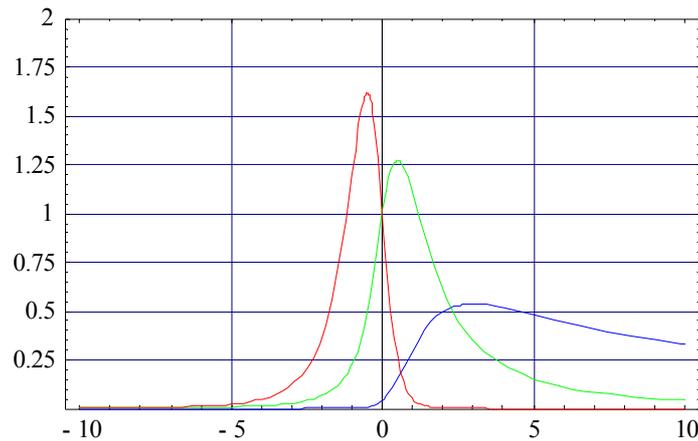


Abb. 4.33: Dichtefunktionen der Typ IV Verteilungen mit variierendem Parametern

Pearson Typ V Verteilungen: Eine Verteilung vom Typ V entsteht, wenn die Gleichung

$$c_0 + c_1x + c_2x^2 = 0 \quad (4.228)$$

ein perfektes Quadrat bildet, d. h.: $c_1^2 = 4c_0c_2$ gilt. Es gilt in Analogie zu (4.205):

$$\begin{aligned} \frac{d}{dx} \log \varphi(x) &= -\frac{x+a}{c_2(x+c_1)^2} \\ &= -\frac{1}{c_2(x+c_1)} - \frac{a-c_1}{c_2(x+c_1)^2} \end{aligned} \quad (4.229)$$

Für die Dichtefunktion $\varphi(x)$ gilt damit die prinzipielle Form

$$\varphi(x) = k(x+c_1)^{-\frac{1}{c_2}} e^{\left[\frac{a-c_1}{c_2(x+c_1)} \right]} \quad (4.230)$$

In dieser Form dargestellt gilt der Definitionsbereich $x > -c_1$, falls $(a-c_1)/c_2 < 0$ und $x > -c_1$, falls $(a-c_1)/c_2 > 0$. Durch Transformation des Definitionsbereichs erhält man die häufig verwendete Darstellung der Typ V Verteilungen:

$$\varphi_V(x) = \frac{x^{-(\alpha+1)} e^{-\beta/x}}{\beta^{-\alpha} \Gamma(\alpha)}, \text{ für } x > 0 \quad (4.231)$$

Die Abbildung 4.34 gibt einige Beispiele für den typischen Verlauf der Dichtefunktion unterschiedlicher Typ V Verteilungen:

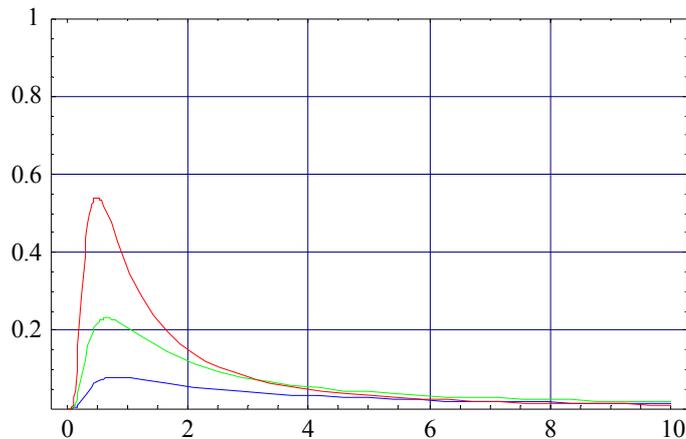


Abb. 4.34: Dichtefunktionen der Typ V Verteilungen mit variierenden Parametern

Pearson Typ VI Verteilungen: Eine Verteilung vom Typ VI entsteht, wenn die Gleichung

$$c_0 + c_1x + c_2x^2 = 0 \quad (4.232)$$

reelle Nullstellen mit gleichem Vorzeichen besitzt. Die Berechnung kann analog zu (4.205) durchgeführt werden. Für die Dichtefunktion $\varphi(x)$ gilt damit die prinzipielle Form

$$\varphi(x) = k(x - a_1)^{m_1}(x - a_2)^{m_2} \quad (4.233)$$

In dieser Form dargestellt gilt der Definitionsbereich $x > a_2$ bzw. $m_2 < -1$ und $m_1 + m_2 < 0$. Durch Transformation des Definitionsbereichs erhält man die häufig verwendete Darstellung der Typ VI Verteilungen als Beta-Verteilung zweiter Art:

$$\varphi_{VI}(x) = \frac{(x/\beta)^{\alpha_1 - 1}}{\beta B(\alpha_1, \alpha_2)(1 + x/\beta)^{\alpha_1 + \alpha_2}}, \text{ für } x > 0 \quad (4.234)$$

Die Abbildung 4.35 gibt einige Beispiele für den typischen Verlauf der Dichtefunktion unterschiedlicher Typ VI Verteilungen:

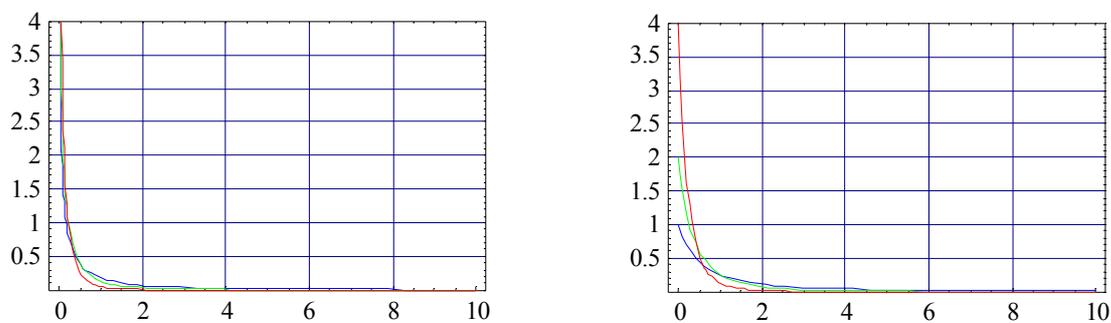


Abb. 4.35: Dichtefunktionen der Typ VI Verteilungen mit variierenden Parametern

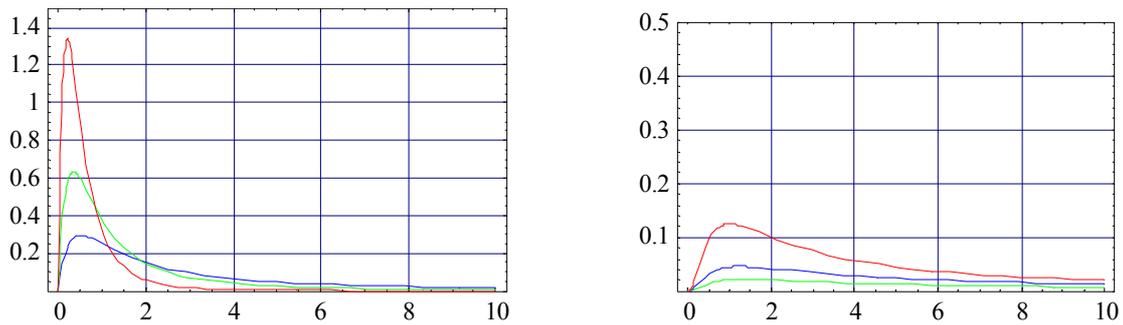


Abb. 4.36: Dichtefunktionen der Typ VI Verteilungen mit variierenden Parametern

Pearson Typ VII Verteilungen: Eine Verteilung vom Typ V entsteht, wenn für die Gleichung (4.199) gilt:

$$a = 0 \quad (4.235)$$

$$c_0 > 0 \quad (4.236)$$

$$c_1 = 0 \quad (4.237)$$

$$c_2 > 0 \quad (4.238)$$

Es gilt in Analogie zu (4.205):

$$\frac{d}{dx} \log \varphi(x) = -\frac{x}{c_0 + c_2 x^2} \quad (4.239)$$

Für die Dichtefunktion $\varphi(x)$ gilt damit die prinzipielle Form

$$\varphi(x) = k(c_0 + c_2 x^2)^{-\frac{1}{2c_2}} \quad (4.240)$$

Ein Vertreter der Typ VII Verteilungen ist die Student's t-Verteilung, für deren Dichtefunktion gilt:

$$\varphi_{\text{Student}}(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\Gamma\left(\frac{\nu}{2}\right)\left(1+\frac{x^2}{\nu}\right)^{(\nu+1)/2}} \quad (4.241)$$

Die Abbildung 4.34 gibt einige Beispiele für den typischen Verlauf der Dichtefunktion unterschiedlicher Student's t- Verteilungen:

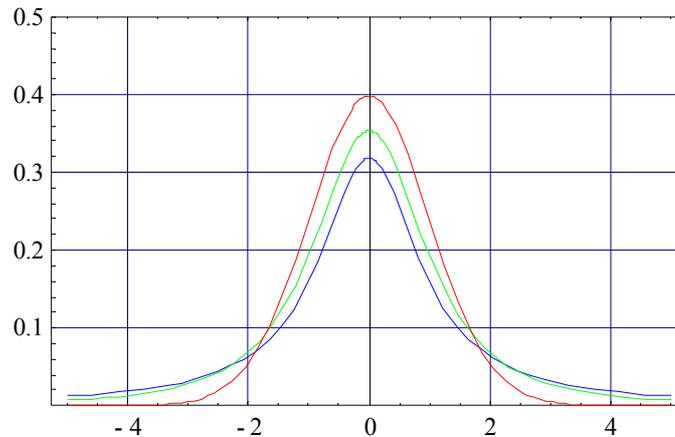


Abb. 4.37: Dichtefunktionen der Student's t-Verteilung mit variierenden Parametern

Pearson Typ VIII Verteilungen: Diese Verteilungen sind Untertypen der Beta-Verteilung. Sie entstehen, wenn für die allgemeine Form nach (4.206)

$$\varphi(x) = k(x - a_1)^{m_1} (a_2 - x)^{m_2} \quad (4.242)$$

$$m_1 < m_2 \text{ und } m_2 \neq 0 \quad (4.243)$$

gilt. Die folgende Abbildung zeigt einige Beispiele für diesen Fall.

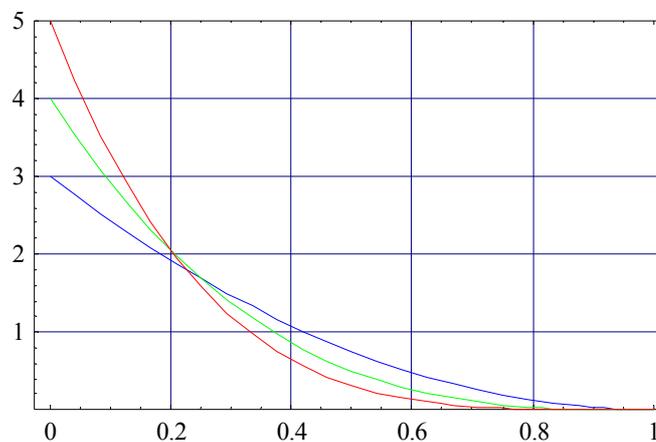


Abb. 4.38: Dichtefunktionen der Typ VIII Verteilungen mit variierenden Parametern

Pearson Typ IX Verteilungen: Diese Verteilungen sind ebenfalls Untertypen der Beta-Verteilung. Sie entstehen, wenn für die allgemeine Form nach (4.206)

$$\varphi(x) = k(x - a_1)^{m_1} (a_2 - x)^{m_2} \quad (4.244)$$

$$m_1 < m_2 \text{ und } m_1 = 0 \quad (4.245)$$

gilt. Die folgende Abbildung zeigt einige Beispiele für diesen Fall.

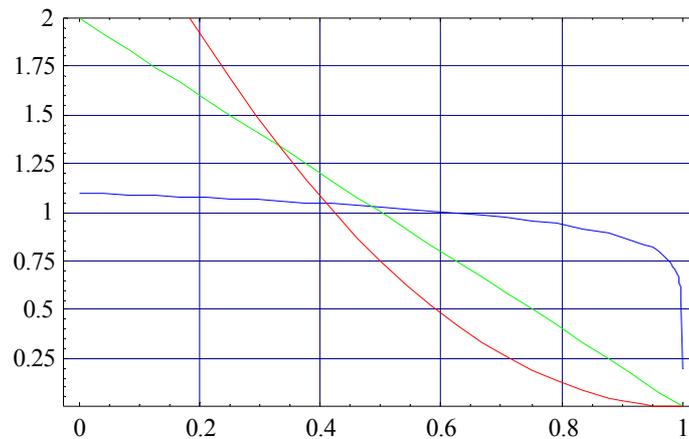


Abb. 4.39: Dichtefunktionen der Typ IX Verteilungen mit variierenden Parametern

Pearson Typ X Verteilungen: Diese Verteilungen entspricht der Exponentialverteilung mit Dichtefunktion

$$\varphi_{\text{Exponential}}(x) = \frac{1}{\beta} e^{-\frac{x}{\beta}} \quad (4.246)$$

Die Exponentialverteilung ist ein Spezialfall der Gamma-Verteilung nach (4.220) für $\lambda = 1$.

Pearson Typ XI Verteilungen: Eine Verteilung vom Typ XI ist identisch zur Verteilung von Typ VI, wenn in der allgemeinen Form

$$\varphi(x) = k(x - a_1)^{m_1} (x - a_2)^{m_2} \quad (4.247)$$

der Parameter $m_2 = 0$ gesetzt wird. Die Abbildung 4.40 gibt einige Beispiele für den typischen Verlauf der Dichtefunktion unterschiedlicher Typ XI Verteilungen:

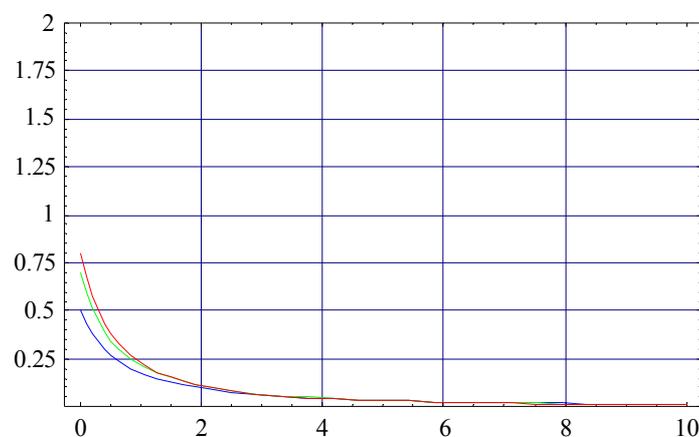


Abb. 4.40: Dichtefunktionen der Typ XI Verteilungen mit variierenden Parametern

Pearson Typ XII Verteilungen: Diese Verteilungen sind ebenfalls Untertypen der Beta-Verteilung. Sie entstehen, wenn für die allgemeine Form nach (4.206)

$$\varphi(x) = k(x - a_1)^{m_1} (a_2 - x)^{m_2} \quad (4.248)$$

$$m_1 < 0 < m_2 < 1 \quad (4.249)$$

gilt. Die folgende Abbildung zeigt einige Beispiele für diesen Fall.

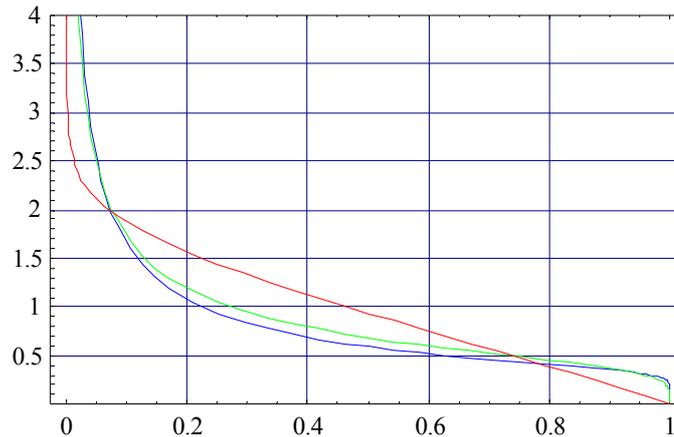


Abb. 4.41: Dichtefunktionen der Typ XII Verteilungen mit variierenden Parametern

Für die unterschiedlichen Pearson Verteilungen gelten die gleichen Anmerkungen, die bereits im Abschnitt 4.6-1 (Alternative Verteilungsfunktionen, Seite 114) zu den dort diskutierten Verteilungen gemacht wurden. Insbesondere muss die Komplexität der mathematischen Beschreibung einer Verteilungsfunktion sowie der zugehörige Definitionsbereich bewertet werden. Unter den Verteilungen des Pearson Systems sind lediglich die Typen IV und VII, die über gesamt \mathbb{R} definiert sind. Bei allen anderen Typen ist der Definitionsbereich durch Randbedingungen auf Teilintervalle der reellen Zahlen eingeschränkt. Zu dem Typ VII gehört die Student's t-Verteilung, die im Abschnitt 4.6-1 (Alternative Verteilungsfunktionen, Seite 114) bereits diskutiert und nur für eingeschränkt geeignet bewertet wurde. Aus dem Pearson System bleibt damit lediglich die Typ IV Verteilung. Die Typ IV Verteilungen entstehen, wenn die Nullstellen des bestimmenden Polynoms (4.200) imaginär sind. Es sind dann aufwendige Ersetzungen durchzuführen, die in folgender allgemeinen Form der Dichtefunktion resultieren:

$$\varphi(x) = k[C_0 + c_2(x + C_1)^2]^{-\frac{1}{2c_2}} e^{\left[-\frac{a-C_1}{\sqrt{c_2}C_0} \arctan\left(\frac{x+C_1}{\sqrt{C_0/c_2}}\right)\right]} \quad (4.250)$$

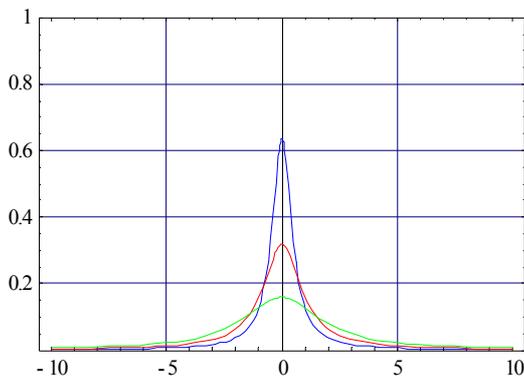
In der Dichtefunktion nach (4.250) muss der Korrekturfaktor k bestimmt werden, um aus der allgemeinen Form eine gültige Dichtefunktion zu erzeugen. Zusätzlich muss die Lage des Modus aufwendig korrigiert werden, da der Modus bei der allgemeinen Form nach (4.250) nicht an der Stelle $x = 0$ fixiert ist. Von der Komplexität der mathematischen Beschreibung ist eine solche Dichtefunktion ein vielfaches komplexer, als die in dieser Arbeit entwickelten Ξ -Verteilungen. Aus diesem Grund wird auf eine weitere Betrachtung des Pearson Systems im Allgemeinen und eine Betrachtung der Typ IV Verteilungen im Speziellen in dieser Arbeit verzichtet.

4.6-3 Alternative Grundfunktionen

Da es in der Literatur keine Verteilungsfunktion gibt, die den Ansprüchen einer asymmetrischen Evolutionsstrategie gerecht wird, muss eine solche neu entwickelt werden. In dieser Arbeit wurde als Grundlage eine modifizierte e -Funktion gewählt. Diese Wahl ist an die Darstellung der Normalverteilung angelehnt. Die Gründe hierfür werden in dem folgenden Abschnitt 4.7 (Spezielle Eigenschaften der Ξ -Verteilungen, Seite 129) dargelegt. Alternativ kann die Grundlage einer asymmetrischen Verteilungsfunktion an andere Verteilungen angelehnt sein. Da die Definition abschnittsweise erfolgt, für die positive bzw. negative Halbachse getrennt, muss die Grundfunktion zumindest für eine der beiden Halbachsen definiert sein und an der Stelle $x = 0$ den Modus besitzen. Im folgenden werden einige Verteilungsfunktionen, die diese Eigenschaften erfüllen, diskutiert.

- Cauchy-Verteilung

Die Cauchy-Verteilung verfügt über den Lageparameter $a \in \mathbb{R}$ und den Skalierungsparameter $b \in \mathbb{R}_+$. Die Momente sind nicht berechenbar. Der Modus sowie das Symmetriezentrum befinden sich an der Stelle $x = a$. Der Verlauf der Cauchy-Verteilung für $a = 0$, $b_1 = 0.5$, $b_2 = 1$ und $b_3 = 2$ ist in der Abbildung 4.42 anhand der Dichtefunktion (4.251) dargestellt.



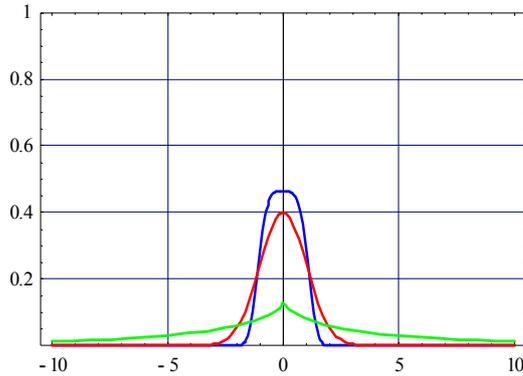
$$\Phi_{\text{Cauchy}}(x) = \frac{1}{\pi b \left(1 + \left(\frac{x-a}{b} \right)^2 \right)} \quad (4.251)$$

Abb. 4.42: Dichtefunktion der Cauchy-Verteilung

- Fehler-Verteilung

Die Fehler-Verteilung verfügt über den Lageparameter $a \in \mathbb{R}$, den Skalierungsparameter $b \in \mathbb{R}_+$ sowie über den Formparameter $c \in \mathbb{R}_+$. Der Formparameter c steuert die Kurtosis der Fehler-Verteilung, für $c < 1$ ist die Verteilung leptokurtisch, für $c = 1$ ist sie mesokurtisch und für $c > 1$ platykurtisch. Für $a = 0$ und $b = c = 1$ entspricht die Fehlerverteilung der $N(0, 1)$ -Normalverteilung, für $c \leftarrow 0$ degeneriert die Fehlerverteilung zu einer Rechteckverteilung im Intervall $[a - b, a + b]$. Der Modus befindet sich an der Stelle $x = a$, die Momente lassen sich mittels der Γ -Funktion berechnen. Der Verlauf der Fehler-Verteilung für $a = 0$,

$b = 1$, $c_1 = 0.5$, $c_2 = 1$ und $c_3 = 3$ ist in der Abbildung 4.43 dargestellt. Es gilt die Dichtefunktion nach (4.252).



$$\varphi_{\text{Fehler}}(x) = \frac{e^{-\frac{c\sqrt{|x-a|}}{2b}}}{\sqrt{b}2^{c/2+1}\Gamma(1+c/2)} \quad (4.252)$$

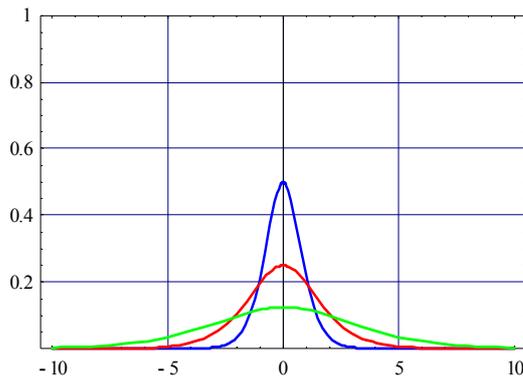
Abb. 4.43: Dichtefunktion der Fehler-Verteilung

- Logistische Verteilung

Die logistische Verteilung verfügt über den Lageparameter $a \in \mathbb{R}$ und den Skalierungsparameter $b \in \mathbb{R}_+$. Der Modus befindet sich an der Stelle $x = a$, alle Momente existieren. Die logistische Verteilung ist die Grenzverteilung für die gewichtete Summe n unabhängiger Gumbel-verteilter Zufallsvariablen X_i , mit den Parametern $a = 0$ und $b = 1$

$$\lim_{n \rightarrow \infty} \left(\sum_{k=1}^n \frac{X_i}{i} \right) \quad (4.253)$$

Der Verlauf der logistischen Verteilung für $a = 0$, $b_1 = 0.5$, $b_2 = 1$ und $b_3 = 2$ ist in der Abbildung 4.44 dargestellt. Es gilt die Dichtefunktion nach (4.254).



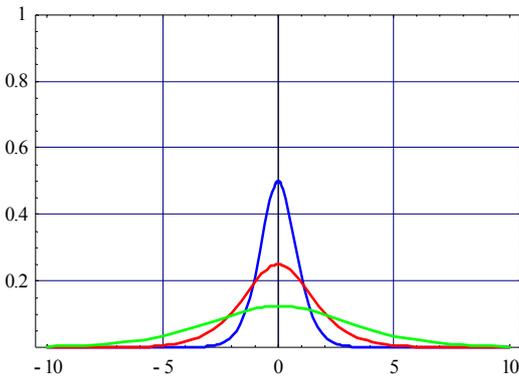
$$\varphi_{\text{Logistisch}}(x) = \frac{\left(\operatorname{sech}\left(\frac{x-a}{2b}\right) \right)^2}{4b} \quad (4.254)$$

Abb. 4.44: Dichtefunktion der logistischen Verteilung

- Student's t -Verteilung

Die Student's t -Verteilung verfügt ausschließlich über den Skalierungsparameter $v \in \mathbb{R}_+$. Der Modus befindet sich fest an der Stelle $x = 0$, alle Momente existieren. Die Student's t -Verteilung ist eng mit der χ^2 -Verteilung und der Normalverteilung verbunden. Der Verlauf der Stu-

dent's t -Verteilung für $\nu_1 = 1$, $\nu_2 = 5$ und $\nu_3 = 100$ ist in der Abbildung 4.45 dargestellt. Es gilt die Dichtefunktion nach (4.255).



$$\varphi_{\text{Student}}(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\Gamma\left(\frac{\nu}{2}\right)\left(1+\frac{x^2}{\nu}\right)^{(\nu+1)/2}} \quad (4.255)$$

Abb. 4.45: Dichtefunktion der Student's t -Verteilung

Es existieren weitere Verteilungsfunktionen, die einen ähnlichen Verlauf und ähnliche Eigenschaften aufweisen wie die dargestellten vier Beispiele. Oft ist es jedoch so, dass bereits die symmetrische Variante eine aufwendige Berechnung der Verteilungsfunktion erfordert. Zur Berechnung der Student's t -Verteilung muss bereits auf eine Fallunterscheidung bezüglich des Skalierungsparameters ν zurückgegriffen werden. Eine weitere Fallunterscheidung zur Realisierung der Asymmetrie kompliziert nur alle weiteren Berechnungen.

Bei der Analyse der obigen vier Beispiele fällt auf, dass mit Ausnahme der Fehler-Verteilung alle Dichtefunktionen einen Verlauf besitzen, der dem Verlauf der Normalverteilung ähnlich ist. Bereits in 3.1 (Variation der Mutationsfunktion, Seite 35) wurde jedoch gezeigt, dass für Evolutionsstrategien eine der normalverteilten Mutation ähnelnde Mutation keine grundlegenden Änderungen des Konvergenzverhaltens hervorruft. Die Fehler-Verteilung bringt den Nachteil mit sich, dass sie bereits für moderate $c < 0.2$ zu einer nahezu $(-1, 1)$ -verteilten Rechteckverteilung degeneriert. Dieses Verhalten ist für Evolutionsstrategien unbrauchbar und müsste durch Reparaturalgorithmen korrigiert werden.

Betrachtet man die diskutierten alternativen Verteilungen mit ihren Eigenschaften, ihrer Berechnungs- und Herleitungskomplexität sowie ihren Vor- und Nachteilen, stellt die Wahl einer e -Funktion als Grundlage für eine asymmetrische Verteilungsfunktion eine elegante Lösung dar. Für alle Verteilungen der Ξ -Familie sind die Herleitungen von Dichte- und Verteilungsfunktionen mit geringem Aufwand möglich. Alle Momente existieren und die inversen Verteilungsfunktionen sind algorithmisch realisierbar (hierzu siehe auch Kapitel 5 (Implementierung, Seite 149)). Zudem ist es möglich, die Schiefe unabhängig von einer symmetrischen Skalierung einzustellen.

4.7 Spezielle Eigenschaften der Ξ -Verteilungen

Für die Verteilungsfunktionen $\Xi_c(x)$, $\Xi_{c,\sigma}(x)$ und $\Xi_{c,\sigma}^\gamma(x)$ gelten einige zusätzliche Eigenschaften, die an dieser Stelle erläutert werden.

4.7-1 Symmetrie

Alle Varianten der Ξ -Verteilungen sind für $c = 0$ symmetrisch. Im symmetrischen Fall gilt für alle Varianten:

$$E_X = 0 \text{ und} \quad (4.256)$$

$$\mu_3 = m_3 = \eta_3 = 0 \quad (4.257)$$

Für die Varianzen V_X gilt:

$$V_X = 1, \text{ falls } X \sim \Xi_c(x), \text{ mit } c = 0 \quad (4.258)$$

$$V_X = \sigma, \text{ falls } X \sim \Xi_{c,\sigma}(x), \text{ mit } c = 0 \quad (4.259)$$

$$V_X = \sigma^\gamma, \text{ falls } X \sim \Xi_{c,\sigma}^\gamma(x), \text{ mit } c = 0 \quad (4.260)$$

Die Verteilungsfunktionen Ξ_X reduzieren sich für $c = 0$ ebenfalls, es gilt:

$$\Xi_X(x) = \frac{1 + \operatorname{erf}(x/\sqrt{2})}{2}, \text{ falls } X \sim \Xi_c(x) \quad (4.261)$$

$$\Xi_X(x) = \frac{1 + \operatorname{erf}(x/\sqrt{2\sigma})}{2}, \text{ falls } X \sim \Xi_{c,\sigma}(x) \quad (4.262)$$

$$\Xi_X(x) = \frac{1 + \operatorname{erf}(x/(\sqrt{2\sigma^\gamma}))}{2}, \text{ falls } X \sim \Xi_{c,\sigma}^\gamma(x) \quad (4.263)$$

Für die inversen Verteilungsfunktionen $\bar{\Xi}_X$ gilt im Fall $c = 0$:

$$\bar{\Xi}_X(x) = \sqrt{2} \operatorname{inverf}(2y - 1), \text{ falls } X \sim \Xi_c(x) \quad (4.264)$$

$$\bar{\Xi}_X(x) = \sqrt{2\sigma} \operatorname{inverf}(2y - 1), \text{ falls } X \sim \Xi_{c,\sigma}(x) \quad (4.265)$$

$$\bar{\Xi}_X(x) = \sqrt{2\sigma^\gamma} \operatorname{inverf}(2y - 1), \text{ falls } X \sim \Xi_{c,\sigma}^\gamma(x) \quad (4.266)$$

4.7-2 Vollständige Äquivalenz zur Normalverteilung

Die $\Xi_{c,\sigma}^\gamma(x)$ -Verteilung besitzt für eine spezielle Belegung der Parameter c und γ eine interessante Eigenschaft. Für $c = 0$ und $\gamma = 2$ entspricht sie einer zentralen Normalverteilung mit der Standardabweichung σ , d. h.

$$\xi_{c,\sigma}^\gamma(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{x^2}{2\sigma^2}}, \text{ für } c = 0 \text{ und } \gamma = 2 \quad (4.267)$$

Im Vorgriff auf das Kapitel 6 (Asymmetrische Evolutionsstrategien, Seite 167) sei an dieser Stelle bereits auf ein Resultat hingewiesen, das aus (4.267) entsteht: Alle Varianten der Evolutionsstrategien, die unkorrelierte Zufallsvariablen verwenden¹, lassen sich durch Einsatz der $\Xi_{c,\sigma}^\gamma(x)$ -Verteilung mit den Parametern $c = 0$ und $\gamma = 2$ nachbilden.

4.7-3 Fallweise Äquivalenz zur Normalverteilung

Im Abschnitt 6.2-1 (Klassische Beweisstrategien, Seite 173) wird zusätzlich zur Äquivalenz von $\Xi_{c,\sigma}^\gamma(x)$ -Verteilung und Normalverteilung die Äquivalenz beider Verteilungen bezüglich der Fallunterscheidung $x < 0$ und $x \geq 0$ benötigt. Die Bedingungen, unter denen die $\Xi_{c,\sigma}^\gamma(x)$ -Verteilung abschnittsweise äquivalent zu Abschnitten der Normalverteilung ist, sind im folgenden angegeben. Für den Fall $x \geq 0$ muss gelten:

$$\frac{\sqrt{2}}{\sqrt{\pi\sigma^\gamma(1+(1+c)^\gamma)}} e^{-\frac{1}{2(\sigma(1+c))^\gamma}x^2} = \frac{1}{\sqrt{2\pi\sigma'}} e^{-\frac{x^2}{2\sigma'^2}} \quad (4.268)$$

Da $\gamma = 2$ gelten muss, lässt sich $\sigma' = \sigma(1+c)$ substituieren. Daraus folgt durch Vereinfachen der Terme folgender Ausdruck, der nur für $c = 0$ gültig ist:

$$\frac{2}{2+c} = \frac{1}{1+c} \quad (4.269)$$

Für den Fall $x < 0$ muss gelten:

$$\frac{\sqrt{2}}{\sqrt{\pi\sigma^\gamma(1+(1+c)^\gamma)}} e^{-\frac{1}{2\sigma^\gamma}x^2} = \frac{1}{\sqrt{2\pi\sigma'}} e^{-\frac{x^2}{2\sigma'^2}} \quad (4.270)$$

Hier lässt sich $\sigma' = \sigma$ substituieren. Durch Vereinfachen folgt die Gleichung

$$\frac{2}{2+c} = \frac{1}{1+c} \quad (4.271)$$

die ebenfalls nur für $c = 0$ gültig ist.

Aus den Gleichungen (4.269) und (4.271) folgt, dass die Äquivalenz zur Normalverteilung, bei der beide Fälle getrennt betrachtet werden, ebenfalls nur für die triviale Kombination $\gamma = 2$ und $c = 0$ gegeben ist.

4.7-4 Stetigkeit

Zu den wichtigen Eigenschaften einer Funktion gehört die Stetigkeit. Die Stetigkeit wird in diesem Abschnitt für die Dichtefunktionen der Ξ -Verteilungen gezeigt.

1. Dies sind mit Ausnahme der Evolutionsstrategien mit korrelierter Selbstadaptation alle weiteren Varianten.

Satz 4.4: Für alle in dieser Arbeit diskutierten Varianten der Ξ -Verteilungen sind für beliebige $c \in \mathbb{R}$, $\sigma \in \mathbb{R}_+$ und $\gamma \in \mathbb{R}_+$ die zugehörigen Dichtefunktionen stetig.

Beweis 4.4: Um diese Eigenschaft zu beweisen, ist für beide Abschnittsfunktionen die Stetigkeit sowie für den Abschnittswechsel an der Stelle $x_0 = 0$ die rechtsseitige bzw. linksseitige Stetigkeit, nachzuweisen. Alle Abschnittsdefinitionen lassen sich auf die Form

$$f(x) = ae^{-bx^2} \quad (4.272)$$

reduzieren. Für die Ersetzungen a und b gilt

$$\frac{\sqrt{2}}{\sqrt{\pi\sigma^\gamma(1+\sqrt{(1+c)^\gamma})}} = a \quad (4.273)$$

sowie

$$\frac{1}{2\sigma^\gamma} = b, \text{ für } c < 0 \quad (4.274)$$

und

$$\frac{1}{2(\sigma(1+c))^\gamma} = b, \text{ für } c \geq 0 \quad (4.275)$$

Die weiteren Varianten folgen aus (4.273) - (4.275) durch Ersetzung von $\gamma = 1$ und $\sigma = 1$. Da die Form (4.272) gilt, lässt sich somit jede Abschnittsdefinition auf eine quadratische Exponentialfunktion reduzieren. Quadratische Exponentialfunktionen der Form (4.272) sind wiederum stetig, siehe z. B. [Bro89].

Zu zeigen ist damit nur noch die Stetigkeit an der Stelle $x_0 = 0$. Eine Funktion f ist an der Stelle x_0 genau dann stetig, wenn der rechtsseitige Grenzwert und der linksseitige Grenzwert an der Stelle x_0 mit dem Funktionswert $f(x_0)$ übereinstimmen. Der Beweis wird an dieser Stelle für $c \geq 0$ durchgeführt, der Fall $c < 0$ folgt diesem Beweis. Für den linksseitigen Grenzwert gilt:

$$\lim_{x \rightarrow x_0 - 0} \left[\frac{\sqrt{2}}{\sqrt{\pi\sigma^\gamma(1+\sqrt{(1+c)^\gamma})}} e^{-\frac{x^2}{2\sigma^\gamma}} \right] = \frac{\sqrt{2}}{\sqrt{\pi\sigma^\gamma(1+\sqrt{(1+c)^\gamma})}} \quad (4.276)$$

Für den rechtsseitigen Grenzwert gilt:

$$\lim_{x \rightarrow x_0 + 0} \left[\frac{\sqrt{2}}{\sqrt{\pi\sigma^\gamma(1+\sqrt{(1+c)^\gamma})}} e^{-\frac{x^2}{2(\sigma(1+c))^\gamma}} \right] = \frac{\sqrt{2}}{\sqrt{\pi\sigma^\gamma(1+\sqrt{(1+c)^\gamma})}} \quad (4.277)$$

Beide Gleichungen folgen unmittelbar aus

$$e^{(x+\Delta x)} = e^x \cdot e^{\Delta x} \quad (4.278)$$

Der Funktionswert an der Stelle x_0 wird durch (4.160) bestimmt, es gilt:

$$f(x_0) = f(0) = \frac{\sqrt{2}}{\sqrt{\pi\sigma^\gamma(1+\sqrt{(1+c)^\gamma})}} \quad (4.279)$$

Somit ist der rechtsseitige Grenzwert und der linksseitige Grenzwert an der Stelle x_0 mit dem Funktionswert $f(x_0)$ identisch und die Dichtefunktion $\xi_{c,\sigma}^\gamma(x)$ stetig. ■

4.7-5 Differenzierbarkeit

Eine weitere wichtige Eigenschaft einer Funktion ist die Differenzierbarkeit. In diesem Abschnitt wird die Differenzierbarkeit für die Ξ -Verteilungen gezeigt.

Satz 4.5: Die Dichtefunktionen der Ξ -Verteilungen sind für beliebige $\sigma \in \mathbb{R}_+$ und $\gamma \in \mathbb{R}_+$ (sofern vorhanden) und alle $c \in \mathbb{R}$ stetig differenzierbar.

Beweis 4.5: Diese Tatsache folgt unmittelbar aus den Eigenschaften der Exponentialfunktion. Für diesen Abschnitt gelten die Ersetzungen nach (4.273) - (4.275). Es gilt dann:

$$\frac{d}{dx} a e^{-bx^2} = -2ab e^{-bx^2} x \quad (4.280)$$

Für den Fall $c \geq 0$ folgt daraus:

$$\frac{d}{dx} \xi_{c,\sigma}^\gamma(x) = \begin{cases} \frac{\sqrt{2}}{\sqrt{\pi\sigma^{3\gamma}(1+\sqrt{(1+c)^\gamma})}} x e^{-\frac{1}{2}\frac{x^2}{\sigma^\gamma}} & \text{für } x < 0 \\ \frac{\sqrt{2}}{\sqrt{\pi\sigma^{3\gamma}[(1+c)^\gamma + \sqrt{(1+c)^{3\gamma}}]}} x e^{-\frac{x^2}{2(\sigma(1+c))^\gamma}} & \text{für } x \geq 0 \end{cases} \quad (4.281)$$

Für den Spezialfall $c = 0$ sind die Dichtefunktionen beliebig oft stetig differenzierbar. Für den allgemeinen Fall gilt dies nicht. Der Beweis folgt unmittelbar aus der Tatsache, dass für $c \neq 0$ die rechtsseitige bzw. linksseitige zweite Ableitung an der Stelle $x_0 = 0$ nicht übereinstimmen, d. h.

$$\lim_{x \rightarrow x_0 - 0} \left(\frac{\frac{d}{dx} \xi_{c,\sigma}^\gamma(x) - \frac{d}{dx} \xi_{c,\sigma}^\gamma(x_0)}{x - x_0} \right) \neq \lim_{x \rightarrow x_0 + 0} \left(\frac{\frac{d}{dx} \xi_{c,\sigma}^\gamma(x) - \frac{d}{dx} \xi_{c,\sigma}^\gamma(x_0)}{x - x_0} \right) \quad (4.282)$$

4.7-6 Summe von Ξ -verteilten Zufallsvariablen

Für eine Vielzahl von Anwendungen ist das additive Verhalten zur Anwendung kommender Zufallsvariablen von Interesse. Die Summe von Zufallsvariablen ist für eine Vielzahl von Verteilungen untersucht. Es existieren Grenzverteilungen für die Summe unendlich vieler Zufallsvariablen [Gne68]. Für die Summe Ξ -verteilter Zufallsvariablen existieren in der Literatur keine Angaben bzw. keine Grenzverteilungen. Dieser Abschnitt beschäftigt sich mit der Summe von zwei Zufallsvariablen, die Ξ -verteilt sind. Der darauf folgende Abschnitt gibt einige wichtige Eigenschaften der Summe. Aussagen über Grenzverteilungen ist das Thema des abschließenden Abschnitts.

Die Dichtefunktion der Summe von zwei kontinuierlichen und unabhängigen Zufallsvariablen X und Y , die durch ihre Dichtefunktionen $\varphi_X(x)$ und $\varphi_Y(y)$ dargestellt werden, kann durch die Faltung $(\varphi_X * \varphi_Y)(t)$ bestimmt werden. Es gilt nach [Kie98]:

$$(\varphi_X * \varphi_Y)(t) = \int_{-\infty}^{\infty} \varphi_X(t - \tau) \varphi_Y(\tau) d\tau \quad (4.283)$$

$$(\varphi_X * \varphi_Y)(t) = \int_{-\infty}^{\infty} \varphi_Y(t - \tau) \varphi_X(\tau) d\tau \quad (4.284)$$

Für Zufallsvariablen, deren Dichtefunktion als geschlossener Ausdruck darstellbar ist, gestaltet sich die Berechnung der Faltung als wenig aufwendig. Für die Summe zweier normalverteilter Zufallsvariablen X und Y mit Erwartungswert E_X und E_Y sowie den Varianzen V_X und V_Y , gilt beispielsweise, dass die Dichtefunktion der Summe $Z = X + Y$ einer normalverteilten Zufallsvariable entspricht, mit dem Erwartungswert $E_Z = E_X + E_Y$ und der Varianz $V_Z = V_X + V_Y$. Die gesamte Herleitung ist trivial und erfordert nur wenige Rechenschritte.¹

Da die Dichtefunktionen der Ξ -Verteilungen abschnittsweise definiert sind, gestaltet sich die Berechnung der Faltung hier aufwendiger. Alle Berechnungen werden an dieser Stelle für die Faltung von zwei rechtsschiefen Ξ -verteilten Zufallsvariablen $X_1 \sim \xi_{c_1, \sigma_1}^\gamma$ und $X_2 \sim \xi_{c_2, \sigma_2}^\gamma$, mit $c_1, c_2 > 0$ durchgeführt. Die weiteren Fälle können analog berechnet werden. Die triviale Lösung, bei der lediglich die entsprechenden Funktionsdefinitionen in das Faltungsintegral eingesetzt werden, lautet:

1. Die Additivität der Erwartungswerte und Varianzen, die hier nur exemplarisch für die Normalverteilung genannt wurde, ist allgemeingültig, wenn Summen **unabhängiger** Zufallsvariablen betrachtet werden. Somit gilt die Additivität auch für Zufallsvariablen, die den Ξ -Verteilungen unterliegen.

$$\begin{aligned}
(\xi_{c_1, \sigma_1}^\gamma * \xi_{c_2, \sigma_2}^\gamma)(t) &= \int_{-\infty}^{\infty} \xi_{c_1, \sigma_1}^\gamma(t-\tau) \cdot \xi_{c_2, \sigma_2}^\gamma(\tau) d\tau \\
&= \begin{cases} \int_{-\infty}^{\infty} d_1 e^{-\frac{1(t-\tau)^2}{2\sigma_1^\gamma}} \cdot d_2 e^{-\frac{1\tau^2}{2\sigma_2^\gamma}} d\tau & \text{für } (t-\tau < 0) \wedge (t < 0) \\ \int_{-\infty}^{\infty} d_1 e^{-\frac{1(t-\tau)^2}{2\sigma_1^\gamma}} \cdot d_2 e^{-\frac{1\tau^2}{2(\sigma_2(1+c_2))^\gamma}} d\tau & \text{für } (t-\tau < 0) \wedge (t \geq 0) \\ \int_{-\infty}^{\infty} d_1 e^{-\frac{1(t-\tau)^2}{2(\sigma_1(1+c_1))^\gamma}} \cdot d_2 e^{-\frac{1\tau^2}{2\sigma_2^\gamma}} d\tau & \text{für } (t-\tau \geq 0) \wedge (t < 0) \\ \int_{-\infty}^{\infty} d_1 e^{-\frac{1(t-\tau)^2}{2(\sigma_1(1+c_1))^\gamma}} \cdot d_2 e^{-\frac{1\tau^2}{2(\sigma_2(1+c_2))^\gamma}} d\tau & \text{für } (t-\tau \geq 0) \wedge (t \geq 0) \end{cases} \quad (4.285)
\end{aligned}$$

Für die Variablen d_1 und d_2 gilt:

$$d_1 = \frac{\sqrt{2}}{\sqrt{\pi\sigma_1^\gamma(1+(1+c_1)^\gamma)}} \quad (4.286)$$

$$d_2 = \frac{\sqrt{2}}{\sqrt{\pi\sigma_2^\gamma(1+(1+c_2)^\gamma)}} \quad (4.287)$$

Diese Lösung lässt sich nicht weiter zusammenfassen, da die Integrationsvariable τ Bestandteil der Fallunterscheidung ist und somit keine geschlossene Form bezüglich der Faltungsvariablen t möglich ist. Ein aufwendigerer Ansatz, der auf der Distributivität der Faltungsoperation aufbaut, ermöglicht jedoch eine Darstellung des Faltungsintegrals, die lediglich die Fallunterscheidung bezüglich t notwendig macht [Kie98, Opp89]. Als Resultat entstehen so acht Faltungsintegrale, vier Faltungsintegrale für $t < 0$ sowie weitere vier Faltungsintegrale für $t \geq 0$. Die Distributivität der Faltungsoperation ermöglicht die Aufsummierung der jeweils vier Faltungsintegrale und ermöglicht so die bereits erwähnte Darstellung, die lediglich von t abhängt. Um die Ableitung übersichtlich zu halten, werden folgende verkürzten Schreibweisen und Farbdarstellungen eingeführt:

$$\xi_1^-(t) = \xi_{c_1, \sigma_1}^\gamma(t), \text{ für } t < 0 \quad (4.288)$$

$$\xi_1^+(t) = \xi_{c_1, \sigma_1}^\gamma(t), \text{ für } t \geq 0 \quad (4.289)$$

$$\xi_2^-(t) = \xi_{c_2, \sigma_2}^\gamma(t), \text{ für } t < 0 \quad (4.290)$$

$$\xi_2^+(t) = \xi_{c_2, \sigma_2}^\gamma(t), \text{ für } t \geq 0 \tag{4.291}$$

Die Teilfunktion $\xi_1^-(t)$ wird in den folgenden Diagrammen schwarz, die Teilfunktion $\xi_1^+(t)$ grün, die Teilfunktion $\xi_2^-(t)$ rot und die Teilfunktion $\xi_2^+(t)$ blau dargestellt. Die grafische Darstellung der Gleichungen (4.288) - (4.291) ist in Abbildung 4.46 dargestellt. Für diese Beispielrechnung sind die Parameter wie folgt gewählt: $c_1 = 1, \sigma_1 = 1, c_2 = 2, \sigma_2 = 1$ sowie $\gamma = 2$.

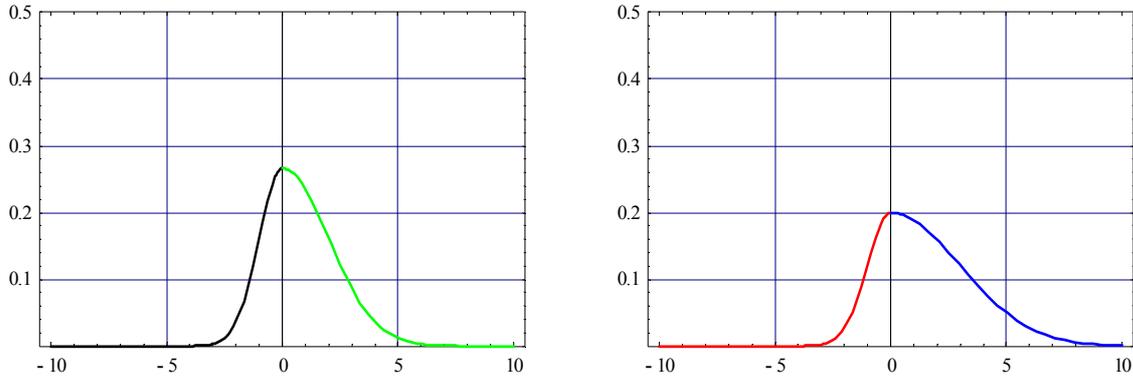
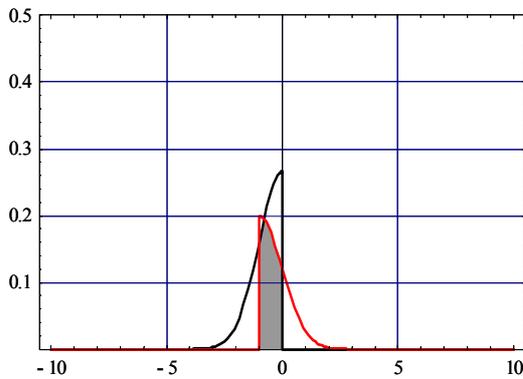
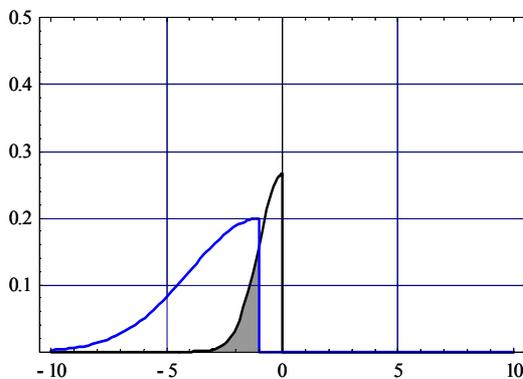


Abb. 4.46: Darstellung der zu faltenden Funktionen $\xi_{c_1, \sigma_1}^\gamma(t)$ und $\xi_{c_2, \sigma_2}^\gamma(t)$

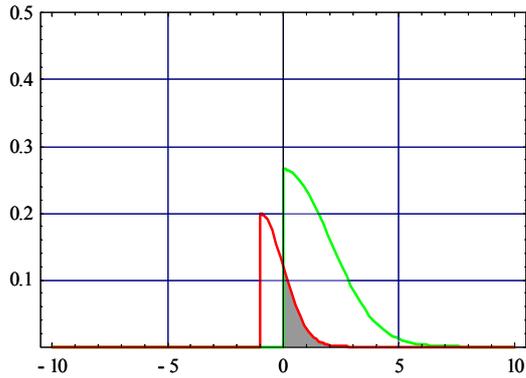
Für den Fall $t < 0$ entstehen die folgenden vier Faltungsintegrale (4.292) - (4.295). Der fixierte Term ist jeweils der Term mit Argument τ , der variierte Term hat das Argument $t - \tau$. Die schraffierte Fläche repräsentiert hier das Resultat der entsprechenden Faltung für ein konkretes t .



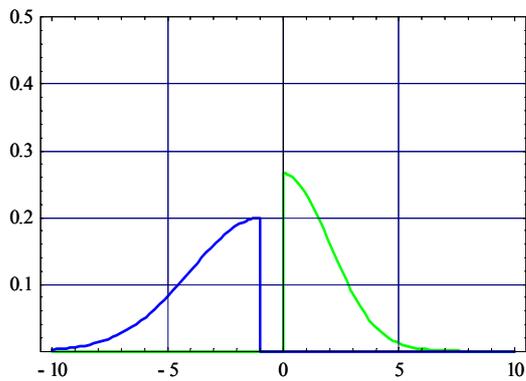
$$(\xi_1^- * \xi_2^-)(t) = \int_t^0 \xi_1^-(\tau) \cdot \xi_2^-(t - \tau) d\tau \tag{4.292}$$



$$(\xi_1^- * \xi_2^+)(t) = \int_{-\infty}^t \xi_1^-(\tau) \cdot \xi_2^+(t - \tau) d\tau \tag{4.293}$$

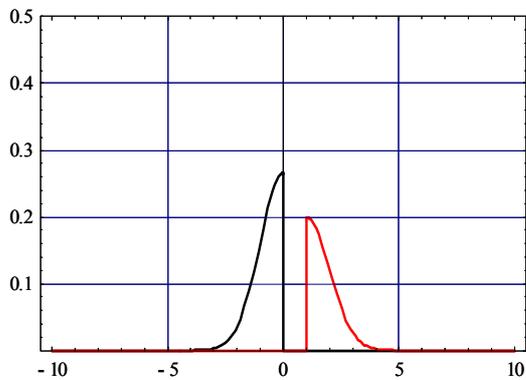


$$(\xi_1^+ * \xi_2^-)(t) = \int_0^{\infty} \xi_1^+(\tau) \cdot \xi_2^-(t-\tau) d\tau \quad (4.294)$$

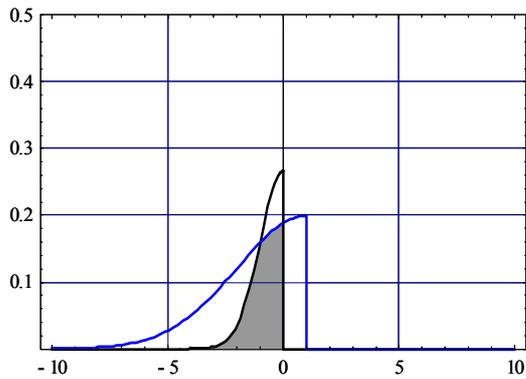


$$(\xi_1^+ * \xi_2^+)(t) = \int_{-\infty}^t \xi_1^+(\tau) \cdot \xi_2^+(t-\tau) d\tau \quad (4.295)$$

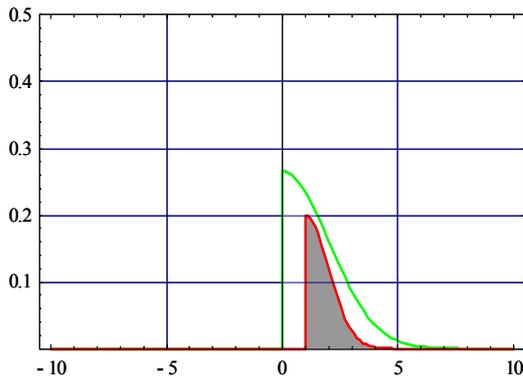
Für den Fall $t \geq 0$ lauten die entsprechenden Faltungsintegrale wie folgt:



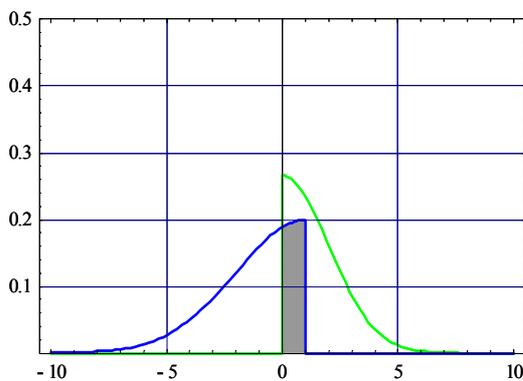
$$(\xi_1^- * \xi_2^-)(t) = \int_t^{\infty} \xi_1^-(\tau) \cdot \xi_2^-(t-\tau) d\tau \quad (4.296)$$



$$(\xi_1^- * \xi_2^+)(t) = \int_{-\infty}^0 \xi_1^-(\tau) \cdot \xi_2^+(t-\tau) d\tau \quad (4.297)$$



$$(\xi_1^+ * \xi_2^-)(t) = \int_t^{\infty} \xi_1^+(\tau) \cdot \xi_2^-(t-\tau) d\tau \quad (4.298)$$



$$(\xi_1^+ * \xi_2^+)(t) = \int_0^t \xi_1^+(\tau) \cdot \xi_2^+(t-\tau) d\tau \quad (4.299)$$

Die Terme (4.292) - (4.299) gelten für beliebige c_i , σ_i und γ , die zugehörigen Abbildungen gelten hier selbstverständlich nur für die o. a. Belegungen der Parameter. Für konkrete Berechnungen müssen folgende Fälle unterschieden werden: $(c_1 < 0) \wedge (c_2 < 0)$, $(c_1 \geq 0) \wedge (c_2 \geq 0)$ ¹ sowie $(c_1 < 0) \wedge (c_2 \geq 0)$ und $(c_1 \geq 0) \wedge (c_2 < 0)$. Die beiden zuletzt genannten Fälle lassen sich aufgrund der Kommutativität der Faltungsoperation auf einen Fall reduzieren, sodass insgesamt drei Fälle getrennt zu betrachten sind. Da die notwendigen Rechenschritte für alle drei Fälle nahezu identisch sind, wird hier nur der Fall $(c_1 \geq 0) \wedge (c_2 \geq 0)$ explizit berechnet. Die anderen Fälle folgen analog.

Zur Berechnung der Faltung müssen die allgemeinen Definitionen $\xi_{1,2}^{+,\gamma}(t)$ durch die entsprechenden Definitionen nach (4.50), (4.121) oder (4.158) ersetzt werden. Die hier durchgeführte Berechnung verwendet die Definition nach (4.158), d.h. die Berechnung der Faltung von zwei $\Xi_{c,\sigma}^\gamma(x)$ -verteilten Zufallsvariablen. Durch $\gamma = 1$ gilt diese Berechnung auch für $\Xi_{c,\sigma}(x)$ -verteilte Zufallsvariablen. Gilt zusätzlich $\sigma = 1$, ist die Ableitung auch für $\Xi_c(x)$ -verteilte Zufallsvariablen gültig. Im folgenden werden die entsprechenden Definitionen in die Gleichungen (4.292) - (4.299) eingesetzt. Es gilt dann für $t < 0$:

1. Der Fall $c = 0$ ist aufgrund der Definition der ξ -Verteilungen in dem Fall $c \geq 0$ enthalten.

$$\begin{aligned}
(\xi_1^- * \xi_2^-)(t) &= \int_0^t \xi_1^-(\tau) \cdot \xi_2^-(t-\tau) d\tau \\
&= \frac{e^{-\frac{t^2}{2(\sigma_1^\gamma + \sigma_2^\gamma)}} \sqrt{2} \left(\operatorname{erf} \left(\frac{t}{(\sqrt{2}\sigma_1^\gamma k)} \right) + \operatorname{erf} \left(\frac{t}{(\sqrt{2}\sigma_2^\gamma k)} \right) \right)}{\sqrt{\pi\sigma_1^\gamma \sigma_2^\gamma k_1 k_2 k}}
\end{aligned} \tag{4.300}$$

$$\begin{aligned}
(\xi_1^- * \xi_2^+)(t) &= \int_{-\infty}^t \xi_1^-(\tau) \cdot \xi_2^+(t-\tau) d\tau \\
&= \frac{e^{-\frac{t^2}{2(\sigma_1^\gamma + K_2^\gamma)}} \sqrt{2} \left(1 + \operatorname{erf} \left(\frac{t}{(\sqrt{2}\sigma_1^\gamma \sqrt{\sigma_1^{-\gamma} + K_2^{-\gamma}})} \right) \right)}{\sqrt{\pi\sigma_1^\gamma \sigma_2^\gamma k_1 k_2} \sqrt{\sigma_1^{-\gamma} + K_2^{-\gamma}}}
\end{aligned} \tag{4.301}$$

$$\begin{aligned}
(\xi_1^+ * \xi_2^-)(t) &= \int_0^\infty \xi_1^+(\tau) \cdot \xi_2^-(t-\tau) d\tau \\
&= \frac{e^{-\frac{t^2}{2(K_1^\gamma + \sigma_2^\gamma)}} \sqrt{2} \left(1 + \operatorname{erf} \left(\frac{t}{(\sqrt{2}\sigma_2^\gamma \sqrt{K_1^{-\gamma} + \sigma_2^{-\gamma}})} \right) \right)}{\sqrt{\pi\sigma_1^\gamma \sigma_2^\gamma k_1 k_2} \sqrt{K_1^{-\gamma} + \sigma_2^{-\gamma}}}
\end{aligned} \tag{4.302}$$

$$\begin{aligned}
(\xi_1^+ * \xi_2^+)(t) &= \int_{-\infty}^0 \xi_1^+(\tau) \cdot \xi_2^+(t-\tau) d\tau \\
&= 0
\end{aligned} \tag{4.303}$$

Für den Fall $t \geq 0$ gilt:

$$\begin{aligned}
(\xi_1^- * \xi_2^-)(t) &= \int_0^\infty \xi_1^-(\tau) \cdot \xi_2^-(t-\tau) d\tau \\
&= 0
\end{aligned} \tag{4.304}$$

$$\begin{aligned}
(\xi_1^- * \xi_2^+)(t) &= \int_{-\infty}^0 \xi_1^-(\tau) \cdot \xi_2^+(t-\tau) d\tau \\
&= \frac{e^{-\frac{t^2}{2(\sigma_1^\gamma + K_2^\gamma)}} \sqrt{2} \left(1 - \operatorname{erf} \left(\frac{t}{(\sqrt{2} K_2^\gamma \sqrt{\sigma_1^{-\gamma} + K_2^{-\gamma}})} \right) \right)}{\sqrt{\pi \sigma_1^\gamma \sigma_2^\gamma k_1 k_2} \sqrt{\sigma_1^{-\gamma} + K_2^{-\gamma}}}
\end{aligned} \tag{4.305}$$

$$\begin{aligned}
(\xi_1^+ * \xi_2^-)(t) &= \int_t^{\infty} \xi_1^+(\tau) \cdot \xi_2^-(t-\tau) d\tau \\
&= \frac{e^{-\frac{t^2}{2(K_1^\gamma + \sigma_2^\gamma)}} \sqrt{2} \left(1 - \operatorname{erf} \left(\frac{t}{(\sqrt{2} K_1^\gamma \sqrt{K_1^{-\gamma} + \sigma_2^{-\gamma}})} \right) \right)}{\sqrt{\pi \sigma_1^\gamma \sigma_2^\gamma k_1 k_2} \sqrt{K_1^{-\gamma} + \sigma_2^{-\gamma}}}
\end{aligned} \tag{4.306}$$

$$\begin{aligned}
(\xi_1^+ * \xi_2^+)(t) &= \int_0^t \xi_1^+(\tau) \cdot \xi_2^+(t-\tau) d\tau \\
&= \frac{e^{-\frac{t^2}{2(K_1^\gamma + K_2^\gamma)}} \sqrt{2} \left(\operatorname{erf} \left(\frac{t}{(\sqrt{2} K_1^\gamma K)} \right) + \operatorname{erf} \left(\frac{t}{(\sqrt{2} K_2^\gamma K)} \right) \right)}{\sqrt{\pi \sigma_1^\gamma \sigma_2^\gamma k_1 k_2} \sqrt{K_1^{-\gamma} + K_2^{-\gamma}}}
\end{aligned} \tag{4.307}$$

Für die neu eingeführten Variablen k , k_1 , k_2 , K , K_1 und K_2 gilt:

$$k = \sqrt{\sigma_1^{-\gamma} + \sigma_2^{-\gamma}} \tag{4.308}$$

$$k_1 = 1 + \sqrt{(1 + c_1)^\gamma} \tag{4.309}$$

$$k_2 = 1 + \sqrt{(1 + c_2)^\gamma} \tag{4.310}$$

$$K = \sqrt{((1 + c_1)\sigma_1)^{-\gamma} + ((1 + c_2)\sigma_2)^{-\gamma}} \tag{4.311}$$

$$K_1 = (1 + c_1)\sigma_1 \tag{4.312}$$

$$K_2 = (1 + c_2)\sigma_2 \tag{4.313}$$

Aufgrund der Distributivität lassen sich diese Terme additiv zusammenfassen. Es gilt dann für die Faltung¹:

$$(\xi_{c_1, \sigma_1}^\gamma * \xi_{c_2, \sigma_2}^\gamma)(t) = \int_{-\infty}^{\infty} \xi_{c_1, \sigma_1}^\gamma(t-\tau) \cdot \xi_{c_2, \sigma_2}^\gamma(\tau) d\tau$$

$$= \begin{cases} \int_{-\infty}^t \xi_1^-(\tau) \cdot \xi_2^+(t-\tau) d\tau + \int_0^t \xi_1^-(\tau) \cdot \xi_2^-(t-\tau) d\tau + \int_t^{\infty} \xi_1^+(\tau) \cdot \xi_2^-(t-\tau) d\tau & \text{für } t < 0 \\ 0 & \\ \int_{-\infty}^0 \xi_1^-(\tau) \cdot \xi_2^+(t-\tau) d\tau + \int_0^t \xi_1^+(\tau) \cdot \xi_2^+(t-\tau) d\tau + \int_t^{\infty} \xi_1^+(\tau) \cdot \xi_2^-(t-\tau) d\tau & \text{für } t \geq 0 \end{cases} \quad (4.314)$$

Durch (4.314) ist es nun möglich, die Summe von zwei Ξ -verteilten Zufallsvariablen zu bestimmen.

4.7-7 Eigenschaften der Summe

An dieser Stelle sind einige Eigenschaften bezüglich der Faltung kommentiert.

1. Symmetrieerhaltung

Die Dichtefunktion der Summe von zwei symmetrisch Ξ -verteilten Zufallsvariablen, d. h. $c_1 = 0$ und $c_2 = 0$, bleibt symmetrisch. Es gilt dann:

$$(\xi_{0, \sigma_1}^\gamma * \xi_{0, \sigma_2}^\gamma)(t) = \frac{1}{\sqrt{2\pi} \sqrt{\sigma_1^\gamma + \sigma_2^\gamma}} e^{-\frac{t^2}{2(\sigma_1^\gamma + \sigma_2^\gamma)}} \quad (4.315)$$

Die folgende Abbildung 4.49 gibt für diesen Fall ein Beispiel:

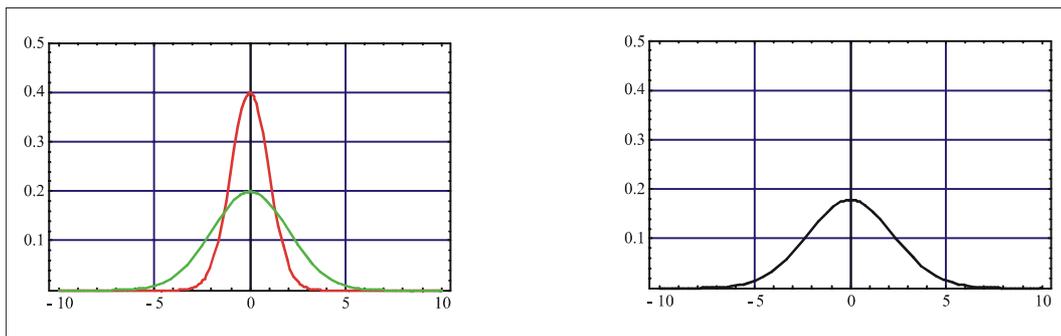


Abb. 4.47: Dichtefunktion zweier symmetrischer Zufallsvariablen (links) sowie die Dichtefunktion der Summe (rechts)

1. Da die Nenner der berechneten Terme unterschiedlich sind, existiert keine wirkliche Vereinfachung der Terme in der Summe. Aus diesem Grund wird an dieser Stelle keine weitere Termersetzung durchgeführt. Für konkrete Berechnungen sind die Faltungsintegrale durch ihre entsprechenden Terme zu ersetzen.

2. Symmetrieerzeugung

Die Dichtefunktion der Summe von zwei asymmetrisch Ξ -verteilten Zufallsvariablen, d. h. $c_1 \neq 0$ und $c_2 \neq 0$, ist symmetrisch, falls $c_1 = -c_2$ gilt. Die folgende Abbildung 4.49 gibt für diesen Fall ein Beispiel:

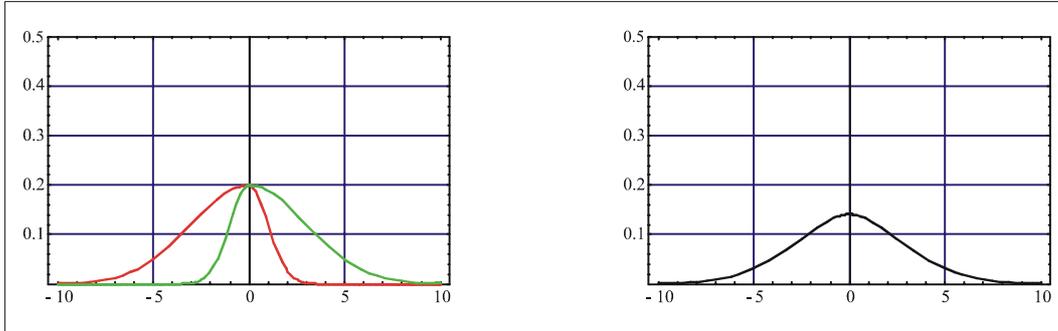


Abb. 4.48: Dichtefunktion einer links- und einer rechtsschiefen Zufallsvariablen mit identischem Betrag des Schiefemaßes (links) sowie die Dichtefunktion der Summe (rechts)

3. Strukturhaltung

Die Dichtefunktion der Summe von zwei symmetrisch Ξ -verteilten Zufallsvariablen, d. h. $c_1 = 0$ und $c_2 = 0$, mit globalem Parameter $\gamma = 2$, ist wiederum eine Ξ -verteilte Zufallsvariable. Es gilt:

$$(\xi_{0, \sigma_1}^2 * \xi_{0, \sigma_2}^2)(t) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{t^2}{2\sigma^2}} \quad (4.316)$$

Für σ gilt hierbei:

$$\sigma = \sqrt{\sigma_1^\gamma + \sigma_2^\gamma} \quad (4.317)$$

Ξ -verteilte Zufallsvariablen, für die $c = 0$ und $\gamma = 0$ gilt sind nach (4.267) normalverteilt mit den Parametern $\mu = 0$ und Varianz $V = \sigma^2$. Die Eigenschaft der Strukturhaltung folgt damit direkt aus den Faltungseigenschaften der Normalverteilung. Es ist jedoch zu bemerken, dass die Ξ -Verteilungen bereits für $c_1 = 0$, $c_2 = 0$ und $\gamma \neq 2$ nicht strukturhaltend sind.

Für den allgemeinen Fall gilt, dass die Summe von Ξ -verteilten Zufallsvariablen nicht Ξ -verteilt ist. Die folgenden drei Abbildungen geben Beispiele für diesen allgemeinen Fall. Abbildung 4.49 beschreibt den Fall, dass die Dichtefunktion zweier rechtsschiefer Zufallsvariablen ($c_1 > 0$, $c_2 > 0$) addiert werden. Die Abbildung 4.50 beschreibt den entsprechenden Fall für zwei linksschiefe

Zufallsvariablen ($c_1 < 0$, $c_2 < 0$), die Abbildung 4.51 den Fall einer rechtsschiefen sowie einer linksschiefen Zufallsvariablen.

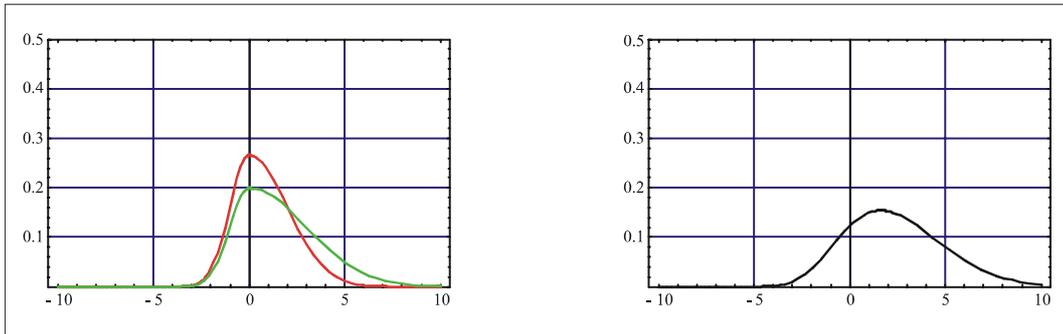


Abb. 4.49: Dichtefunktion zweier rechtsschiefer Zufallsvariablen (links) sowie die Dichtefunktion der Summe (rechts)

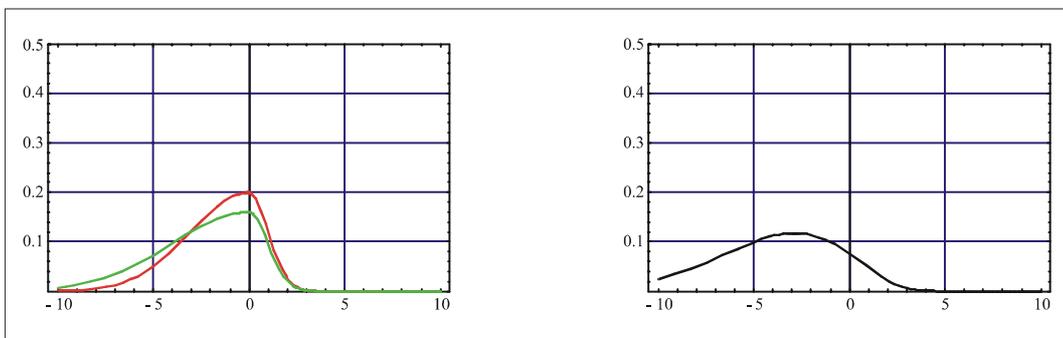


Abb. 4.50: Dichtefunktion zweier linksschiefer Zufallsvariablen (links) sowie die Dichtefunktion der Summe (rechts)

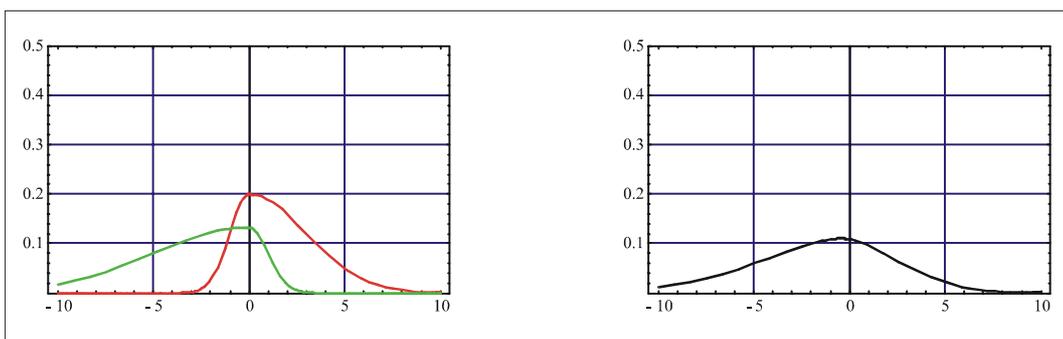


Abb. 4.51: Dichtefunktion zweier rechtsschiefer Zufallsvariablen (links) sowie die Dichtefunktion der Summe (rechts)

Es kann vermutet werden, dass eine Erweiterung der Ξ -Verteilungen, die in dieser Arbeit grundsätzlich zentral sind, auf nicht zentrale Verteilungen Zusammenhänge aufzeigt, in denen die Strukturhaltung ausgeprägter ist. Nichtzentrale Verteilungen sind jedoch nicht Gegenstand dieser Arbeit, da sie nicht der Definition der Mutationsfunktionen im weiteren Sinne (siehe 4.2 (Mutationsfunktionen, Seite 70)) entsprechen.

4.7-8 Zentraler Grenzwertsatz

Grenzwertsätze ermöglichen Aussagen über die Eigenschaften von Folgen von Zufallsvariablen. Von besonderer Bedeutung sind hier Grenzwertsätze, die Aussagen über die Summe von Zufallsvariablen geben, bei denen die Anzahl der Summanden gegen unendlich strebt. Ziel dieser Grenzwertsätze ist die Bestimmung der Grenzverteilungen, die durch solche Summen entstehen bzw. die Charakterisierung von Zufallsvariablen, für die bestimmte Formen von Grenzverteilungen existieren. Grundlegende Grenzwertsätze sind von Bernoulli und de Moivre zu Beginn des 18. Jahrhunderts geschaffen worden [Gne60, Gne68]. Methoden zum Beweis der Grenzwertsätze, die sich auf Summen beliebig verteilter, unabhängiger Zufallsvariablen beziehen, sind in der zweiten Hälfte des 19. Jahrhunderts von Tschebyschew entwickelt worden. Tschebyschew hat sich im besonderen um die Bedingungen bemüht, unter denen das Gesetz der großen Zahlen sowie der zentrale Grenzwertsatz für Summen unabhängiger Zufallsvariablen mit endlichem Erwartungswert E_X und endlicher Varianz V_X gelten. Die Theorie, die hierzu aufgestellt wurde, wird in [Gne60, Gne68, Zub72] beschrieben. Der zentrale Grenzwertsatz nach Lindeberg-Lévy lautet:

Satz 4.6: Es seien X_1, \dots, X_n stochastisch unabhängige und identisch verteilte Zufallsvariablen mit positiver Varianz $V = V_{X_1}$ und Erwartungswert $\mu = E_{X_1}$. Mit $X = X_1 + X_2 + \dots + X_n$ gilt:

$$\lim_{n \rightarrow \infty} P\left(a \leq \frac{X - n\mu}{V\sqrt{n}} \leq b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{t^2}{2}} dt \quad (4.318)$$

$$\lim_{n \rightarrow \infty} P\left(\frac{X - n\mu}{V\sqrt{n}} \leq b\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^b e^{-\frac{t^2}{2}} dt \quad (4.319)$$

Der Beweis für den zentralen Grenzwertsatz findet sich in [Gne60, Gne68]. Neben dieser Variante des zentralen Grenzwertsatzes gibt es weitere, die die Konvergenz gegen die Poisson-Verteilung zeigen, falls die Zufallsvariablen die hierfür notwendige Verteilung besitzen. Bekannt ist ebenfalls der zentrale Grenzwertsatz nach Moivre und Laplace, der Aussagen über die Grenzverteilung der Summe binomialverteilter Zufallsvariablen macht. Der folgende Satz von Chintschin gibt die Bedingung für den hier benötigten Fall, dass die Verteilungsfunktion der Summe unabhängiger Zufallsvariablen gegen die Normalverteilung konvergiert.

Satz 4.7: Wenn die Verteilungen der Summen

$$X_n = X_{n1} + X_{n2} + \dots + X_{nk_n} \quad (4.320)$$

infinitesimaler, in jeder Zeile unabhängiger Zufallsgrößen X_{nk} , mit $1 \leq k \leq k_n$ gegen eine Grenzverteilung konvergieren, so ist diese dann, und nur dann, eine Normalverteilung, wenn für beliebige $\varepsilon > 0$ gilt:

$$\lim_{n \rightarrow \infty} \left(\sum_{k=1}^{k_n} \int_{|x| \geq \varepsilon} d\Phi_{X_{nk}}(x) \right) = 0 \quad (4.321)$$

Der Beweis für diesen Satz findet sich z. B. in der Arbeit von Gnedenko und Kolmogorov [Gne60]. Die Übertragung des zentralen Grenzwertsatzes auf die asymmetrisch Ξ -verteilten Zufallsvariablen durch Anwendung des Satzes 4.7 ist sehr aufwendig. Zum einen muss gezeigt werden, dass die X_n unbegrenzt teilbar sind, zum anderen muss das Integral über dem Bereich $|x| \geq \varepsilon$ aufgelöst werden. Ein Problemfall vieler Grenzwertsätze ist zudem die Forderung, dass die Summe unabhängiger und identisch verteilter Zufallsvariablen betrachtet wird. Die Forderung nach Unabhängigkeit der Zufallsvariablen ist bei der Bestimmung der Grenzverteilung Ξ -verteilter Zufallsvariablen unproblematisch, da für den Einsatz in asymmetrischen Evolutionsstrategien die Unabhängigkeit der Zufallsvariablen ein zentrales Charakteristikum ist. Problematisch ist jedoch die Forderung nach identisch verteilten Zufallsvariablen. Bezogen auf das Themengebiet dieser Arbeit würde dies bedeuten, dass Schrittweiten und Schiefemaße für alle Dimensionen der zu optimierenden Funktion identisch gewählt werden müssen. Dies würde die asymmetrischen Evolutionsstrategien trivialisieren und die Betrachtung der Grenzverteilung nutzlos machen.

Der Satz von Ljapunow ist in einer Form bewiesen worden, die von unabhängigen Zufallsvariablen ausgeht, deren Forderung an die Momente jedoch weit schwächer ist und damit für die Grenzverteilung der Summe Ξ -verteilter Zufallsvariablen anwendbar ist. Der Satz lautet wie folgt:

Satz 4.8: Es seien X_1, \dots, X_n stochastisch unabhängige Zufallsvariablen mit Erwartungswerten E_{X_i} , Varianzen $V_{X_i} \neq 0$ und endlichem dritten Moment $\mu_3(X_i)$. Es gelte zusätzlich

$$V_{\Sigma} = \sqrt{V_{X_1} + \dots + V_{X_n}} \text{ und} \quad (4.322)$$

$$\mu_{3\Sigma} = \sqrt[3]{\mu_3(X_1) + \dots + \mu_3(X_n)}. \quad (4.323)$$

Wenn der Grenzwert

$$\lim_{n \rightarrow \infty} \frac{\mu_{3\Sigma}}{V_{\Sigma}} = 0 \quad (4.324)$$

existiert, dann gilt für die Summe

$$X = \frac{X_1 + \dots + X_n - E_{X_1} - \dots - E_{X_n}}{V_{\Sigma}} \quad (4.325)$$

dass sie für $n \rightarrow \infty$ gegen die Normalverteilung mit Verteilungsfunktion

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^b e^{-\frac{t^2}{2}} dt \quad (4.326)$$

konvergiert.

Der Beweis dieses Satzes findet sich in [Zub72]. Da dieser Satz lediglich fordert, dass die Varianzen nicht verschwinden dürfen und die dritten Momente endlich sind, kann er für die Grenzverteilung Ξ -verteilter Zufallsvariablen eingesetzt werden. Dies resultiert in dem folgenden Satz:

Satz 4.9: Es seien X_1, \dots, X_n stochastisch unabhängige Zufallsvariablen. Jede Zufallsvariable X_i sei $\Xi_{c_i, \sigma_i}^\gamma$ -verteilt. Es gelte zusätzlich

$$V_\Sigma = \sqrt{V_{X_1} + \dots + V_{X_n}}. \quad (4.327)$$

Für die Summe X der Zufallsvariablen X_i

$$X = \frac{X_1 + \dots + X_n - E_{X_1} - \dots - E_{X_n}}{V_\Sigma} \quad (4.328)$$

gilt, dass sie für $n \rightarrow \infty$ gegen die Normalverteilung mit Verteilungsfunktion

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt \quad (4.329)$$

konvergiert.

Beweis 4.9: Nach Satz 4.8 ist die Summe stochastisch unabhängiger Zufallsvariablen $X = (X_1 + \dots + X_n - E_{X_1} - \dots - E_{X_n})/V_\Sigma$ normalverteilt, wenn

- a. die Varianzen nicht verschwinden,
- b. die dritten Momente endlich sind und
- c. der Grenzwert $\lim_{n \rightarrow \infty} (\mu_{3\Sigma}/V_\Sigma) = 0$, existiert.

Die Gültigkeit dieser drei Forderungen wird nun gezeigt:

- a. Für die Varianzen V_{X_i} gilt:

$$V_{X_i} = \sigma_i^\gamma \left((1 - \sqrt{(1 + |c_i|)^\gamma} + (1 + |c_i|)^\gamma) - \frac{2(\sqrt{(1 + |c_i|)^\gamma} - 1)^2}{\pi} \right) \quad (4.330)$$

Die Varianzen können nur dann verschwinden, wenn einer der beiden Faktoren des Produktes gleich 0 ist. σ_i^γ ist, nach Definition, immer größer als 0, da $0 < \sigma_i < \infty$ gilt. Der zweite Faktor steigt mit wachsendem c_i . Den kleinsten Wert nimmt er mit 1 für $c_i = 0$ an. Daraus folgt, dass die Varianzen nicht verschwinden, da keiner der beiden Faktoren gleich 0 gesetzt werden kann. ■

- b. Für die dritten Momente $\mu_3(X_i)$ gilt:

$$\begin{aligned} \mu_3(X_i) = & \frac{\sqrt{2}(-1 + \sqrt{(1 - c_i)^\gamma})\sqrt{\sigma_i^{3\gamma}}}{\sqrt{\pi^3}} \\ & \cdot (-4 + \sqrt{(1 - c_i)^\gamma}(8 - 3\pi) + (1 - c_i)^\gamma(-4 + \pi) + \pi) \end{aligned} \quad (4.331)$$

Laut Definition gilt $-\infty < c_i < \infty$, $0 < \sigma_i < \infty$ und $0 < \gamma < \infty$. Der Nenner des Quotienten in (4.331) ist konstant und von c_i und γ unabhängig. Alle Operationen, die zur Berechnung von $\mu_3(X_i)$ verwendet werden, haben endlichen Wertebereich bei endlichen Argumenten. Daher folgt, dass die dritten Momente endlich sind. ■

- c. Unter der Vereinfachung, dass $c_i = c$ und $\sigma_i = \sigma$ gilt sowie einer Neuordnung der Terme folgt für den Quotienten $\mu_{3\Sigma}/V_\Sigma$:

$$\frac{\mu_{3\Sigma}}{V_\Sigma} = -\frac{1}{\sqrt[6]{2n}} \cdot \sqrt[3]{(\sqrt{(1+c)^\gamma} - 1)} \cdot \frac{\sqrt[3]{(\sqrt{(1+c)^\gamma}(8-3\pi) + (1+c)^\gamma(\pi-4) + \pi-4)}}{\sqrt{(1+c)^\gamma(\pi-2) + \pi - \sqrt{(1+c)^\gamma(\pi-4)} - 2}} \quad (4.332)$$

Für die Berechnung des Grenzwertes für $n \rightarrow \infty$ ist lediglich der erste Faktor von Bedeutung. Da die Varianzen nicht verschwinden, ist der Quotient $\mu_{3\Sigma}/V_\Sigma$ unabhängig von σ . Der dritte Faktor enthält nur Parameter, die von n unabhängig sind. Für γ gilt nach Definition: $0 < \gamma < \infty$. Daher gilt, dass für die Grenzwertbetrachtung alle Faktoren, mit Ausnahme des ersten, Konstanten $< \infty$ sind und der Grenzwert lediglich durch den ersten Faktor bestimmt wird. Trivialerweise folgt

$$\lim_{n \rightarrow \infty} \left(\frac{1}{\sqrt[6]{2n}} \right) = 0 \quad (4.333)$$

und somit unter der Vereinfachung $c_i = c$ und $\sigma_i = \sigma$:

$$\lim_{n \rightarrow \infty} (\mu_{3\Sigma}/V_\Sigma) = 0. \quad (4.334)$$

Für den Fall, dass nicht mehr $c_i = c$ und $\sigma_i = \sigma$ gilt, ist die Vereinfachung des Produktes nach (4.332) nicht mehr möglich. Trotzdem lässt sich eine Neuordnung der Faktoren in einen von n abhängigen und einen von n unabhängigen Teil durchführen. Auch hier gilt dann mit der gleichen Argumentation

$$\lim_{n \rightarrow \infty} \left(\frac{1}{\sqrt[6]{2n}} \right) = 0 \quad (4.335)$$

und daraus dann

$$\lim_{n \rightarrow \infty} (\mu_{3\Sigma}/V_\Sigma) = 0 \quad (4.336)$$

für beliebige σ_i , c_i und γ , mit $0 < \sigma_i < \infty$, $-\infty < c_i < \infty$ und $0 < \gamma < \infty$. ■

4.8 Zusammenfassung

Die Wahrscheinlichkeitsverteilungen Ξ_c , $\Xi_{c,\sigma}$ und $\Xi_{c,\sigma}^\gamma$ stellen Verteilungen dar, deren Eigenschaften und Verlauf durch die Parameter c , σ und γ in einem weiten Bereich beeinflusst werden können. So ist es möglich, durch geeignete Parametrisierung einer $\Xi_{c,\sigma}^\gamma$ -Verteilung die Normalverteilung zu erzeugen. Wird der Parameter c konstant auf 0 gesetzt, werden nur symmetrische Verteilung erzeugt, die einen ähnlichen Verlauf der Verteilungsfunktion besitzen, wie die Normalverteilung. Durch Wahl eines positiven Parameters c kann eine rechtsschiefe Verteilung erzeugt werden, für negative c ist die Verteilung linksschief. Für die Wahrscheinlichkeitsverteilungen Ξ_c , $\Xi_{c,\sigma}$ und $\Xi_{c,\sigma}^\gamma$ können alle Momente berechnet werden sowie die zugehörigen inversen Verteilungsfunktionen. Das asymptotische Verhalten der Summe von Ξ_c , $\Xi_{c,\sigma}$ oder $\Xi_{c,\sigma}^\gamma$ -verteilten Zufallszahlen entspricht dem Verhalten der Summe normalverteilter Zufallszahlen. Zusätzlich zu diesen Eigenschaften sind die Wahrscheinlichkeitsverteilungen Ξ_c , $\Xi_{c,\sigma}$ und $\Xi_{c,\sigma}^\gamma$ für beliebige $c \in \mathbb{R}$, $\sigma \in \mathbb{R}_+$ und $\gamma \in \mathbb{R}_+$ stetig und mindestens einmal stetig differenzierbar. Für $c = 0$ sind sie beliebig oft stetig differenzierbar.

Kapitel 5 Implementierung

*Es gibt nichts Nebensächlicheres als die Zeit.
(Rainer Maria Rilke)*

Die Realisierung der in dieser Arbeit vorgestellten asymmetrischen Evolutionsstrategie erfordert keine speziellen Implementierungstechniken in Bezug auf Datenstrukturen und Algorithmen. Grundsätzlich ähnelt sie der Realisierung einer (μ, λ) -ES mit n Schrittweiten¹, die um n weitere Strategieparameter, die Asymmetriemaße, erweitert wurde. Die Anzahl der Strategieparameter ist damit linear von der Dimensionalität des Optimierproblems abhängig. Diese lineare Abhängigkeit resultiert in einer linearen Zunahme der Speicheranforderung in einem Rechnersystem, die heutige Rechner vor kein Problem stellt.

Auch wenn die Implementierung der asymmetrischen Evolutionsstrategie keine speziellen Techniken erfordert, gibt es im Zusammenhang mit der Erzeugung asymmetrisch verteilter Zufallszahlen Implementierungsdetails, die nicht zum Standardrepertoire einer Entwicklungsumgebung gehören. Insbesondere sind dies Methoden zur Generierung (fast) beliebig verteilter Zufallszahlen und die damit verbundenen Approximationen der Gaußschen Fehlerfunktion sowie der inversen Gaußschen Fehlerfunktion. Die folgenden Abschnitte behandeln diese Themengebiete.

5.1 Eigenschaften von Zufallszahlengeneratoren

Für die Funktionsweise einer Evolutionsstrategie ist die Generierung von Zufallszahlen von zentraler Bedeutung. Zufallszahlen werden benötigt, um einzelne Individuen aus der Menge der Eltern oder Nachkommen auszuwählen, oder um eine neue Position eines Individuums im Suchraum zu bestimmen. Je nach Variante der Evolutionsstrategie werden für diese Aufgaben Zufallsvariablen mit unterschiedlichen Eigenschaften benötigt. So wird für die Auswahl eines einzelnen Individuums in der Regel eine gleichverteilte Zufallsvariable, für die Veränderung der Position eines Individuums meist eine normalverteilte Zufallsvariable verwendet. Auch in diesem Kapitel wird zwischen einer Zufallsvariablen, d. h. einer Variablen, die in Abhängigkeit vom Zufall unterschiedliche Werte annimmt, und einer Zufallszahl, d. h. einer konkreten Realisierung einer Zufallsvariablen, unterschieden.

Eine der wichtigsten Eigenschaften einer Zufallsvariablen ist die zu Grunde liegende Verteilung, die bestimmt, welche Zahlen in welcher Häufigkeit von einem zugehörigen Zufallszahlengenerator erzeugt werden. Auf diese Eigenschaft wird in diesem Abschnitt der Schwerpunkt gelegt. Weitere

1. n entspricht der Dimensionalität des Optimierproblems

Eigenschaften, die in Arbeiten über Evolutionsstrategien häufig nur eine Nebenrolle spielen, sind die Genauigkeit, die Effizienz und die Robustheit eines Generators. Generatoren für die bei klassischen Evolutionsstrategien benötigten Zufallsvariablen sind im Laufe der Jahre immer wieder verbessert worden, sodass die oben genannten Eigenschaften alle erfüllt sind. Speziell bei der Generierung asymmetrisch verteilter Zufallszahlen, die in dieser Arbeit eine zentrale Bedeutung besitzen, kann jedoch nicht auf bereits existierende Generatoren zurückgegriffen werden, die bereits alle Anforderungen erfüllen. Daher wird im folgenden auf die Eigenschaften Genauigkeit, Effizienz und Robustheit eingegangen.

5.1-1 Genauigkeit

Unter Genauigkeit eines Zufallszahlengenerators wird die Eigenschaft verstanden, zufällige Folgen von Zahlen zu erzeugen, die exakt der gewünschten Verteilung gehorchen. Gleichzeitig muss jedoch der begrenzten Genauigkeit bei der Darstellung reeller Zahlen Rechnung getragen werden, die mit steigender Wortlänge der Prozessoren zunimmt, jedoch auf Grund des endlichen Speicherplatzes eines Rechners nicht beliebig ist. Für die Generierung von Zufallszahlen bieten heutige Prozessoren jedoch eine ausreichende Wortlänge, sodass hier keine technisch bedingten Probleme auftreten. Nahezu alle Verfahren, die beliebig verteilte Zufallszahlen erzeugen, benötigen während der Generierung eine kontinuierliche, gleichverteilte Zufallsvariable. Für eine beliebige, kontinuierliche, gleichverteilte Zufallsvariable $G(a, b)$ gilt:

$$\varphi_{G(a,b)}(x) = \begin{cases} \frac{1}{b-a} & \text{für } a \leq x \leq b \\ 0 & \text{sonst} \end{cases} \quad (5.1)$$

$$\Phi_{G(a,b)}(x) = \begin{cases} 0 & \text{für } x < a \\ \frac{x-a}{b-a} & \text{für } a \leq x \leq b \\ 0 & \text{für } x > b \end{cases} \quad (5.2)$$

Eine $G(a, b)$ -verteilte Zufallsvariable erzeugt damit Zufallszahlen in dem Wertebereich $[a, b] \subset \mathbb{R}$. Alle Zahlen in diesem Intervall werden mit gleicher Wahrscheinlichkeit erzeugt. Die Genauigkeit der $G(a, b)$ -verteilten Zufallsvariable beeinflusst direkt die Genauigkeit der daraus abgeleiteten Zufallsvariable. Da genaue und effiziente Algorithmen zur Erzeugung $G(a, b)$ -verteilter Zufallszahlen existieren, sollten diese auch verwendet werden. Law und Kelton geben in [Law91] eine Übersicht über diese Generatoren. In [Hel98] erläutert Hellekalek Probleme, die mit den in vielen Entwicklungsumgebungen bereits integrierten Generatoren auftreten. In dieser Arbeit wird ein PMMLC¹ Generator verwendet, der mittels der Vorschrift

$$Z_i = (630360019 \cdot Z_{i-1}) \bmod 2147483647 \quad (5.3)$$

alle Zahlen aus dem Intervall $[0, 2^{31} - 1] \subset \mathbb{N}$ während eines Zyklus erzeugt. Die Ausschöpfung, d. h. die Sequenzlänge dieses Generators ist damit optimal. Empirische Untersuchungen dieses

1. **prime modulus multiplicative linear congruential**

Generators mittels eines χ^2 -Tests haben zudem gezeigt, dass die erzeugte Zahlenfolge einer perfekten Gleichverteilung entspricht [Law91]. Die Originalarbeit zu diesem Generator stammt von Marsse und Roberts aus dem Jahr 1983. Anpassungen an Pascal und C finden sich in der Arbeit von Law und Kelton [Mar83, Law91].

Durch Einsatz eines genauen $G(0, 1)$ -Generators besteht nun die Möglichkeit, aus einer gleichverteilten Zufallsvariable eine möglichst genaue, beliebig verteilte Zufallsvariable zu erzeugen. Ist die Verteilungsfunktion Φ_X einer Zufallsvariable X bekannt, können statistische Tests benutzt werden, um die Abweichung einer durch einen Generator erzeugten Zufallszahl \bar{X} von der idealen, Φ -verteilten Zufallsvariable X zu bestimmen und zu bewerten. Da in dieser Arbeit sowohl die Verteilungsfunktion Φ als auch die Dichtefunktion φ bekannt sind, lassen sich diese Tests, zu denen der χ^2 -Test und der Kolmogorov-Smirnov-Test gehören, anwenden.

χ^2 -Test: Die Grundidee des χ^2 -Test ist, die bekannte Dichtefunktion φ_X mit einem Histogramm der erzeugten Zufallszahl \bar{X} zu vergleichen. Der erste Schritt teilt den Definitionsbereich von \bar{X} in k disjunkte Intervalle ein:

$$[a_0, a_1), [a_1, a_2), \dots, [a_{k-1}, a_k) \quad (5.4)$$

Die Fälle $a_0 = -\infty$ und $a_k = \infty$ sind erlaubt. Bei den in dieser Arbeit betrachteten Zufallsvariablen gilt damit:

$$[-\infty, a_1), [a_1, a_2), \dots, [a_{k-1}, \infty) \quad (5.5)$$

Danach werden n Zufallszahlen \bar{x}_i erzeugt, die jeweils dem Intervall j zugeordnet werden, sodass $a_{j-1} \leq \bar{x}_i < a_j$ gilt. Die Anzahl der den Intervallen zugeordneten \bar{x}_i wird mit N_j bezeichnet. Im nächsten Schritt wird bestimmt, wie hoch der Anteil p_j eines jeden Intervalls an den erzeugten Zufallszahlen laut der vorgegebenen Dichtefunktion φ sein soll:

$$p_j = \int_{a_{j-1}}^{a_j} \varphi_X(x) dx \quad (5.6)$$

Mit diesen Vorgaben kann der Wert χ^2 berechnet werden. Es gilt:

$$\chi^2 = \sum_{j=1}^k \frac{(N_j - np_j)^2}{np_j} \quad (5.7)$$

Die Hypothese H_0 für diesen Test lautet: Die generierten \bar{x}_i sind unabhängig und identisch verteilt nach Φ_X . Da alle Parameter der Verteilungsfunktion Φ_X bekannt sind, schlagen Law und Kelton folgende Bewertung für χ^2 vor. Falls H_0 wahr ist, konvergiert χ^2 für $n \rightarrow \infty$ gegen eine χ^2 -Verteilung mit $k-1$ Freiheitsgraden. Daraus folgt, dass die Hypothese H_0 mit der Signifikanz α verworfen wird, falls $\chi^2 > \chi_{k-1, 1-\alpha}^2$ gilt [Law91].

Kolmogorov-Smirnov-Test: Ein Nachteil des χ^2 -Tests ist die Notwendigkeit, den Definitionsbereich der Zufallszahl in Intervalle einzuteilen. Hier schafft der Kolmogorov-Smirnov-Test Abhilfe, da es bei diesem Test nicht notwendig ist, Intervalle zu spezifizieren.

Auch bei dem Kolmogorov-Smirnov-Test wird eine empirische Größe, in diesem Fall eine von einem Generator erzeugte Folge von Zufallszahlen, mit der gewünschten Größe verglichen. Anders als bei dem χ^2 -Test wird bei dem Kolmogorov-Smirnov-Test jedoch die Verteilungsfunktion und nicht die Dichtefunktion verwendet. Aus der erzeugten Folge von n Zufallszahlen \bar{x}_i wird durch die Vorschrift

$$\bar{\Phi}_n(\bar{x}) = \frac{\|\bar{x}_i \leq x\|}{n} \quad (5.8)$$

eine empirische Verteilungsfunktion erzeugt. Der charakteristische Wert D_n des Kolmogorov-Smirnov-Test wird durch

$$D_n = \sup_x (|\bar{\Phi}_n(\bar{x}) - \Phi_n(x)|) \quad (5.9)$$

berechnet. Für den Fall, dass alle Parameter der Verteilungsfunktion Φ_n bekannt sind, dies ist in dieser Arbeit der Fall, gilt folgende Gleichung für die Ablehnung der Hypothese H_0 :

$$\left(\sqrt{n} + 0.12 + \frac{0.11}{\sqrt{n}}\right) D_n > c_{1-\alpha} \quad (5.10)$$

Die Werte $c_{1-\alpha}$ für unterschiedliche Signifikanzen α können der folgenden Tabelle 5-1 entnommen werden:

$1 - \alpha$	0.850	0.900	0.950	0.975	0.990
$c_{1-\alpha}$	1.138	1.224	1.358	1.480	1.628

Tab. 5-1: Kritische Werte $c_{1-\alpha}$ für unterschiedliche Signifikanzen

Der Wert D_n , der durch den Kolmogorov-Smirnov-Test berechnet und bewertet wird, ist die maximale Abweichung zwischen der empirischen, hier also der generierten Zufallszahl, und der idealen Verteilungsfunktion. Daraus folgt, dass je größer der Wert D_n ist, der Zufallszahlengenerator um so stärker von der gewünschten Verteilungsfunktion abweicht. Im Gegensatz zum χ^2 -Test muss der Definitionsbereich hier nicht in Intervalle aufgeteilt werden. Es entfällt somit die subjektiv geprägte und evtl. fehlerträchtige Einteilung in n Intervalle. Zudem ist der Kolmogorov-Smirnov-Test für beliebig lange, oder auch beliebig kurze, Folgen von Zufallszahlen geeignet, während die Resultate des χ^2 -Tests nur asymptotisch für $n \rightarrow \infty$ gelten. Es muss beim Kolmogorov-Smirnov-Test jedoch berücksichtigt werden, dass die Bestimmung des Supremums eine sortierte Liste der erzeugten Zufallszahlen erfordert. Der Zeitaufwand zur Sortierung beträgt $O(n \ln(n))$, d. h. bei dem Test von langen Folgen kann die Rechenzeit problematisch werden.

5.1-2 Effizienz

Ein Generator zur Erzeugung beliebig verteilter Zufallszahlen sollte effizient mit den Ressourcen Speicher und Rechenzeit umgehen. Die Ressource Speicher sollte bei modernen Rechnern kein Pro-

blem darstellen. Spezielle Algorithmen jedoch, die große Wertetabellen benötigen, können den vorhandenen Prozessorcaché überfordern, sodass trotz großzügig bemessenem Hauptspeicher ein negativer Einfluss auf die Rechenzeit ausgeübt wird. Abhilfe kann hier durch Funktionaldarstellung bzw. Approximation der Tabellenwerte erreicht werden.

Die Ressource Rechenzeit muss kritischer betrachtet werden, da bereits die Generierung einer gleichverteilten Zufallszahl mit Rechenaufwand verbunden ist. Die Transformation der gleichverteilten Zufallsvariable in einer Zufallsvariable mit der gewünschten Verteilung kann weitere, rechenintensive Verarbeitungsschritte erfordern. Für die Realisierung evolutionärer Algorithmen durch VLSI-Schaltungen existieren Zufallszahlengeneratoren, die als elektronische Schaltung aufgebaut sind und damit sehr schnell arbeiten können. Solche Ansätze können allgemein genutzt werden, um Zufallszahlen als Hardware-Lösung zu erzeugen [Alp91, Ang95, Bla96].

5.1-3 Robustheit

Die Robustheit eines Zufallszahlengenerators bezieht sich auf das Verhalten über dem gesamten Definitionsbereich. Nach Möglichkeit sollten sich Effizienz und Genauigkeit über dem Definitionsbereich gleich verhalten. Abschnittsweise definierte Generatoren, die z. B. nach der in Abschnitt 5.2-2 (Kompositionsmethode) beschriebenen Methode arbeiten, müssen diese Eigenschaft nicht immer erfüllen, da sie intern mit unterschiedlichen Generatoren arbeiten können. Diese Generatoren können sich sowohl in der Genauigkeit als auch in der Effizienz deutlich unterscheiden.

5.1-4 Reproduzierbarkeit

An viele Zufallszahlengeneratoren wird die Anforderung gestellt, dass sie reproduzierbare Folgen von Zufallszahlen erzeugen. Bei Simulationen kann so die Auswirkung von Parameteränderungen untersucht werden, ohne dass der Einfluss von unterschiedlichen Zufallszahlen die Ergebnisse zusätzlich verändert. Bei nahezu allen mathematisch orientierten Zufallszahlengeneratoren können initiale Werte durch den Benutzer definiert werden. Mit identischen Initialwerten werden dann auch identische Folgen von Zufallszahlen erzeugt. Um unterschiedliche Folgen von Zufallszahlen zu erzeugen, werden diese Initialwerte in der Regel durch häufig wechselnde Werte initialisiert. Die letzten Stellen der internen Zeitfunktion, die bei modernen Rechnersystemen im Mikrosekundenbereich aktualisiert wird, eignen sich für solche Aufgaben.

Generatoren für Zufallszahlen, die sich physikalische Eigenschaften eines Bauteils zu Nutze machen, erzeugen in nahezu allen Fällen keine reproduzierbaren Folgen von Zufallszahlen. Typische Vertreter dieser Klasse von Generatoren sind solche, die das thermische Rauschen eines Transistors zur Bestimmung der nächsten Zufallszahl ausnutzen. Auf dieser Art generierte Folgen von Zufallszahlen sind, nach heutigem Stand der Forschung nicht deterministisch und nicht reproduzierbar. Dieses Verhalten muss auch in Betracht gezogen werden, wenn Generatoren verwendet werden, wie die in Abschnitt 5.1-2 (Effizienz, Seite 152) beschriebenen VLSI-Realisierungen.

5.2 Typen von Zufallszahlengeneratoren

Die Erzeugung gleichverteilter Zufallszahlen ist in modernen Entwicklungsumgebungen eine fest integrierte Funktionalität. Durch den Einsatz von linearen Kongruenzgeneratoren bzw. deren Weiterentwicklungen und die Wahl geeigneter Parameter ist es möglich, effizient genaue, gleichverteilte Zufallszahlen zu erzeugen. Der im Abschnitt 5.1-1, Seite 150 beschriebene PMMLC Generator ist ein solcher genauer und effizienter Generator.

Auch wenn Funktionen zur Erzeugung gleichverteilter Zufallszahlen in fast allen Entwicklungsumgebungen existieren, ist die Erzeugung von Zufallszahlen mit anderen Verteilungen in der Regel nur durch Eigenentwicklung bzw. Verwendung zusätzlicher statistischer Bibliotheken möglich. Grundsätzlich ist die Verwendung statistischer Bibliotheken vorzuziehen, da diese ausführlich getestet und bewertet sind. Existiert keine Bibliotheksfunktion für einen Zufallszahlengenerator mit gewünschter Verteilung, für den Fall einer $A_\gamma(c, \sigma)$ -verteilten Zufallszahl ist dies der Fall, muss der Generator selbst entwickelt werden. Es existieren mehrere Methoden, aus einer gleichverteilten Zufallszahl eine beliebig verteilte Zufallszahl zu erzeugen: die Transformation der Inversen, die Kompositionsmethode, die Konvolution oder die Akzeptanz/Ablehnungs-Methode sind hier als die wichtigsten zu nennen.

5.2-1 Transformation der Inversen

Das Ziel der Transformation der Inversen ist es, bei bekannter Verteilungsfunktion Φ der gewünschten Zufallszahl, aus einer $G(0, 1)$ -verteilten Zufallszahl eine Φ -verteilte Zufallszahl zu erzeugen. Hierzu wird die inverse Verteilungsfunktion $\bar{\Phi}$ benötigt. Der Algorithmus zur Erzeugung einer Φ -verteilten Zufallszahl lautet:

- 1 Funktion randtransinv: Eingabe = $\bar{\Phi}(x)$, Ausgabe = Φ -verteilte Zufallszahl
- 2 Erzeuge $G \sim G(0, 1)$
- 3 **Rückgabe** $\leftarrow \bar{\Phi}(G)$

Alg. 5-1: Transformation der Inversen

Die Zufallszahl X , die durch diesen Algorithmus erzeugt wird, ist Φ -verteilt. Die Funktion $X = \bar{\Phi}(G)$ ist immer definiert, da für den Wertebereich von $G: G \in [0, 1] \subset \mathbb{R}$ gilt und für den Definitionsbereich von $\bar{\Phi}: \bar{\Phi}(x): (x \in [0, 1] \subset \mathbb{R})$ gilt. Der Algorithmus 5-1 basiert auf dem Beweis, dass für beliebige $x \in \mathbb{R}$, $P(X \leq x) = \Phi(x)$ gilt:

$$\begin{aligned} P(X \leq x) &= P(\bar{\Phi}(G) \leq x) \\ &= P(G \leq \Phi(x)) \\ &= \Phi(x) \end{aligned} \tag{5.11}$$

Diese vereinfachte Form des Beweises gilt für alle streng monotonen Verteilungsfunktionen und damit insbesondere für Ξ_γ . Für monotone Verteilungsfunktionen oder diskrete Verteilungen muss über das Infimum

$$\inf_{0 < g < 1} (x: \Phi(x)=g) \tag{5.12}$$

argumentiert werden. Der Beweis für diesen Fall findet sich in [Dev86].

Wird die Erzeugung einer $G(0, 1)$ -verteilten Zufallszahl als elementare Operation betrachtet, hängt die Effizienz einer durch die Transformation der Inversen erzeugten Zufallszahl ausschließlich von der Effizienz der Erzeugung der inversen Verteilungsfunktion ab. Besteht diese aus elementaren Operationen, ist die Methode der Transformation der Inversen ein genaues und effizientes Werkzeug zur Erzeugung beliebig verteilter Zufallszahlen. In dieser Tatsache liegt aber auch ein großer Nachteil dieser Methode: nicht alle inversen Verteilungsfunktionen lassen sich durch einen geschlossenen

Ausdruck angeben. Dies gilt z. B. für die inverse Verteilungsfunktion der Normalverteilung. Lässt sich die inverse Verteilungsfunktion approximieren, kann diese Methoden trotzdem angewandt werden, die Effizienz wird jedoch sinken.

5.2-2 Kompositionsmethode

Die Kompositionsmethode kann angewandt werden, wenn sich die Verteilungsfunktion Φ als konvexe Summe anderer Verteilungsfunktionen Φ_i darstellen lässt und die Erzeugung von Φ_i -verteilten Zufallszahlen einfacher ist als die der Φ -verteilten Zufallszahl. Für Φ muss gelten:

$$\Phi(x) = \sum_{i=1}^{\infty} p_i \Phi_i(x) \quad (5.13)$$

$$p_i \geq 0, \text{ und } \sum_{j=1}^{\infty} p_j = 1 \quad (5.14)$$

Für praktische Anwendungen gilt hierbei, dass die Anzahl der Summanden begrenzt ist, d. h. ab einem Index i_k gilt:

$$p_i = 0 \text{ für } i \geq i_k \quad (5.15)$$

Die p_i können bei dieser Methode als Vektor von Einzelwahrscheinlichkeiten interpretiert werden. Angewendet wird die Kompositionsmethode nur für eine endliche Anzahl Summanden n . Der Algorithmus für die Kompositionsmethode lautet dann:

- 1 Funktion randkomp: Eingabe = $p_1, \dots, p_n, \Phi_1, \dots, \Phi_n$ Ausgabe = Φ -verteilte Zufallszahl
- 2 Erzeuge G , sodass $P(G=g) = p_g$ für $g = \{1, \dots, n\}$ gilt
- 3 Generiere zufällig eine Zahl X , die Φ_G -verteilt ist
- 4 **Rückgabe** $\leftarrow X$

Alg. 5-2: Kompositionsmethode

Für die Umsetzung bietet es sich an, eine Indexfunktion $I_i(x)$ einzuführen, für die gilt:

$$I_i(x) = \begin{cases} 1 & \text{falls } x \in \left[\sum_{j=0}^{i-1} p_j, \sum_{j=0}^i p_j \right) \\ 0 & \text{sonst} \end{cases} \quad (5.16)$$

Für p_0 gilt $p_0 = 0$.

Wenn die Zufallszahl G gleichverteilt erzeugt wird, kann durch $I_i(G)$ bestimmt werden, welche Verteilung Φ_G für die Generierung der Zufallszahl X verwendet werden soll. Eine anschauliche Deutung der Kompositionsmethode besteht darin, den Definitionsbereich von Φ in n Intervalle

aufzuteilen, die Wahrscheinlichkeit zu bestimmen, wie häufig eine Φ -verteilte Zufallszahl in einem Intervall i liegt, und dann eine Φ_i -verteilte Zufallszahl zu erzeugen. Die Kompositionsmethode wird in der Regel verwendet, wenn es möglich ist, die Verteilungsfunktion Φ durch Komposition von Gleich- oder Triangulärverteilungen bzw. durch ähnlich simple Verteilungen darzustellen. Der Beweis für die Gültigkeit der Kompositionsmethode findet sich z. B. in [Dev86, Ros89].

5.2-3 Konvolutionsmethode

Im Gegensatz zur Kompositionsmethode, bei der Voraussetzung ist, dass sich eine beliebige Verteilung als Summe anderer, einfacher zu bestimmender Verteilungsfunktionen darstellen lässt, ist die Voraussetzung bei der Konvolutionsmethode, dass sich eine beliebige Zufallsvariable X als Summe anderer, identisch verteilter Zufallsvariablen Y_i mit Verteilungsfunktion Φ_Y darstellen lässt. Die Idee hierbei ist auch wieder, dass die Zufallsvariablen Y_i in der Summe einfacher zu erzeugen sind als die Zufallsvariable X :

$$X = \sum_{i=1}^n Y_i \quad (5.17)$$

Der Algorithmus lautet dann:

- 1 Funktion `randkonv`: Eingabe = Φ_Y , Ausgabe = Φ -verteilte Zufallszahl
- 2 Erzeuge n identisch verteilte Zufallszahlen Y_1, \dots, Y_n mit Verteilungsfunktion Φ_Y
- 3 Bestimme $X = \sum_{i=1}^n Y_i$
- 4 **Rückgabe** $\leftarrow X$

Alg. 5-3: Konvolutionsmethode

Die Konvolutionsmethode nach (5.17) ist gültig, da folgende Gleichung gilt:

$$\begin{aligned} P(X \leq x) &= P\left(\sum_{i=1}^n Y_i \leq x\right) \\ &= \Phi(x) \end{aligned} \quad (5.18)$$

Der Beweis für die Gültigkeit der Gleichung (5.18) findet sich in [Law91].

5.2-4 Akzeptanz/Ablehnungs-Methode

Die vorgestellten Methoden zur Generierung von Zufallszahlen aus den Abschnitten 5.2-1 (Transformation der Inversen), 5.2-2 (Kompositionsmethode) und 5.2-3 (Konvolutionsmethode) zählen zu den direkten Methoden, da sie in einem Arbeitsschritt eine Zufallszahl erzeugen. Der Nachteil dieser Methoden ist, dass nicht für beliebige Verteilungen Zufallszahlen erzeugt werden können, da bestimmte Grundvoraussetzungen erfüllt sein müssen:

1. Transformation der Inversen: inverse Verteilungsfunktion $\bar{\Phi}$ muss bekannt sein,
2. Kompositionsmethode: die Verteilung Φ muss als Summe von Einzelverteilungen Φ_i darstellbar sein,
3. Konvolutionsmethode: Die Zufallszahl X muss als Summe von identisch verteilten Zufallszahlen Y_i darstellbar sein.

Die in diesem Abschnitt vorgestellte Akzeptanz/Ablehnungs-Methode stellt keine starken Voraussetzungen, ist jedoch auch ineffizienter als die drei direkten Methoden.

Voraussetzung für die Akzeptanz/Ablehnungs-Methode ist es, eine Funktion $t(x)$ zu spezifizieren, die die gewünschte Dichtefunktion $\varphi(x)$ majorisiert, d. h.: $t(x) \geq \varphi(x)$, für alle x . Bei der Funktion $t(x)$ wird es sich in der Regel nicht um eine Dichtefunktion handeln, da häufig

$$\int_{-\infty}^{\infty} t(x) dx \geq 1 \tag{5.19}$$

gelten wird. Wird die Funktion $r(x)$ wie folgt definiert:

$$r(x) = \frac{t(x)}{\int_{-\infty}^{\infty} t(x) dx} \tag{5.20}$$

mit der Annahme, dass $\int_{-\infty}^{\infty} t(x) dx < \infty$, handelt es sich bei $r(x)$ um eine Dichtefunktion. Der Algorithmus zur Erzeugung einer Zufallszahl nach der Akzeptanz/Ablehnungs-Methode lautet dann:

- 1 Funktion randakz: Eingabe = $r(x)$, Ausgabe = Φ -verteilte Zufallszahl
- 2 Erzeuge $Y \sim r(x)$ mit Dichtefunktion $r(x)$
- 3 Erzeuge $G \sim G(0, 1)$
- 4 **Falls** $G \leq f(Y)/t(Y)$: **Rückgabe** $\leftarrow Y$
- 5 Starte neu bei Schritt 2

Alg. 5-4: Akzeptanz/Ablehnungs-Methode

Wichtig ist es bei dieser Methode, eine Funktion $t(x)$ zu finden, sodass für die resultierende Funktion $r(x)$ mit einfachen Mitteln eine Zufallszahl bestimmt werden kann (Schritt 2 in Alg. 5-4). Die Akzeptanz/Ablehnungs-Methode ist ein iterativer Prozess, bei dem die Effizienz stark von den Funktionen $t(x)$ und $r(x)$ abhängt. Law und Kelton geben in [Law91] einen Beweis, dass die Wahrscheinlichkeit der Akzeptanz in Schritt 3 zu

$$1 / \int_{-\infty}^{\infty} t(x) dx \tag{5.21}$$

berechnet werden kann. Daraus folgt, dass $\int_{-\infty}^{\infty} t(x) dx$ möglichst klein sein sollte. Dies wiederum bedingt, dass die majorisierende Funktion $t(x)$ möglichst genau die gewünschte Dichtefunktion $\varphi(x)$ approximieren sollte.

5.3 Approximation der inversen Gaußschen Fehlerfunktion

Der $A_\gamma(c, \sigma)$ -Generator zur Erzeugung von Zufallszahlen, die Ξ -verteilt sind, setzt die Transformation der Inversen ein. Er benötigt daher numerische Werte für die inverse Verteilungsfunktion $\bar{\Xi}$. Zur Berechnung von $\bar{\Xi}(y)$ wird die inverse Fehlerfunktion $\overline{erf}(z)$, d. h. die Lösung für z in $s = erf(z)$, verwendet, für die es keinen geschlossenen Ausdruck gibt, und die daher nur approximativ bestimmt werden kann. Bevor die Approximation erläutert wird, seien einige Eigenschaften normalverteilter Zufallszahlen, der Normalverteilung an sich und der Fehlerfunktion $erf(x)$ wiederholt.

Eine Zufallszahl heißt normalverteilt, wenn sie eine Dichtefunktion folgender Gestalt hat:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \frac{(x-a)^2}{\sigma^2}} \quad (5.22)$$

a und σ sind die Lageparameter der Normalverteilung. In a liegt das Maximum der Normalverteilung, σ gibt den Abstand von der Stelle des Maximums zu den Wendepunkten an. Ist $a = 0$ und $\sigma = 1$, handelt es sich um die normierte und zentrale Normalverteilung. Die zu $\varphi(x)$ gehörende Verteilungsfunktion lautet:

$$\Phi(x) = \int_{-\infty}^x \varphi(t) dt \quad (5.23)$$

$\Phi(x)$ lässt sich nicht geschlossen darstellen. Aufgrund der Symmetrie der Dichtefunktion $\varphi(x)$ gilt für negative Argumente: $\Phi(-x) = 1 - \Phi(x)$. Zur Bestimmung von $\Phi(x)$ wird häufig das Wahrscheinlichkeitsintegral $\Phi_0(x)$ der normierten und zentralen Normalverteilung verwendet, für das gilt:

$$\Phi_0(x) = \int_0^x \varphi(t) dt = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{1}{2}t^2} dt \quad (5.24)$$

Mit $\Phi_0(x)$, das für $x \geq 0$ definiert ist und der Symmetriebedingung, lassen sich die Werte für $\Phi(x)$ der normierten und zentralen Normalverteilung durch

$$\Phi(x) = \Phi_0(x) + \frac{1}{2} \quad (5.25)$$

bestimmen. Das Wahrscheinlichkeitsintegral $\Phi_0(x)$ wird in der Literatur auch häufig als Gaußsches Fehlerintegral, Gaußsche Fehlerfunktion oder Fehlerfunktion bezeichnet. Diese Bezeichnung kollidiert mit der Definition der Gaußschen Fehlerfunktion als:

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \quad (5.26)$$

die in dieser Arbeit verwendet wird. Zwischen der Fehlerfunktion $erf(x)$ und dem Wahrscheinlichkeitsintegral $\Phi_0(x)$ bestehen folgende Beziehungen:

$$\Phi_0(x) = \frac{1}{2} \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right) \quad (5.27)$$

und

$$\operatorname{erf}(x) = 2\Phi_0(\sqrt{2}x) \quad (5.28)$$

Das in der Literatur gelegentlich verwendete Wahrscheinlichkeitsintegral wird in dieser Arbeit selten verwendet. Bei der Definition der inversen Verteilungsfunktionen wird die Fehlerfunktion $\operatorname{erf}(x)$ jedoch oft benötigt.

Zur graphischen Veranschaulichung der beiden Funktionen $\Phi_0(x)$ und $\operatorname{erf}(x)$ ist folgendes Diagramm gegeben:

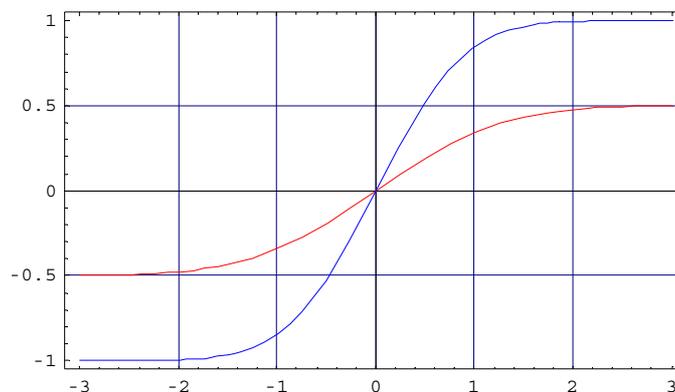


Abb. 5.1: Verlauf der Funktionen $\Phi_0(x)$ (rot) und $\operatorname{erf}(x)$ (blau)

Die Fehlerfunktion $\operatorname{erf}(x)$ ist über dem Intervall $x \in [-\infty, \infty]$ definiert. Es gilt für den Wertebereich das Intervall $[-1, 1] \subset \mathbb{R}$. Daraus folgt, dass $\overline{\operatorname{erf}}(z)$ auf dem Intervall $[-1, 1]$ definiert ist mit $\overline{\operatorname{erf}}(-1) = -\infty$ und $\overline{\operatorname{erf}}(1) = \infty$. Für die Approximation der Normalverteilung und damit nach (5.25) und (5.28) auch der Fehlerfunktion, gibt es in der Literatur genaue und effiziente Algorithmen. Zu nennen ist hier die polynomielle Approximation nach Abramowitz und Stegun [Bey88] mit einer maximalen Abweichung $|\varepsilon(x)| < 7.5 \cdot 10^{-8}$.

Für die Approximation der inversen Normalverteilung $\overline{\Phi}(x)$ gibt es wenige Algorithmen, die mittels einer polynomiellen Berechnung die Approximation durchführen. Iterative Algorithmen existieren, benötigen jedoch mehr Rechenzeit [Cun69]. Eine polynomielle Approximation, mit einer maximalen Abweichung $|\varepsilon(x)| < 1.6 \cdot 10^{-6}$, ist von Hill und Davis entwickelt worden [Hil73]. Sie wird in den Algorithmen 5-5 und 5-6 vorgestellt. Weitere Verfahren finden sich in der Arbeit von Odeh und Evans [Ode74] sowie in der Standardbibliothek IMSL als Algorithmus ANORIN [IMS87].

- 1 Funktion gausin: Eingabe= z , Ausgabe $\overline{\Phi}(z)$
- 2 **Falls** $z \leq 0$ oder $z \geq 1$: **Rückgabe** \leftarrow Fehler: ungültige Eingabe
- 3 **Falls** $z = 1/2$: **Rückgabe** $\leftarrow 0$
- 4 $x \leftarrow z$
- 5 **Falls** $z > 1/2$: $x \leftarrow 1 - z$
- 6 $x \leftarrow \sqrt{-2 \log(x)}$

$$7 \quad x \leftarrow x - \frac{1271.06 + 450.636x^2 + 7.45551x^2}{500.756 + 750.365x + 110.421x^2 + x^3}$$

$$8 \quad \mathbf{Falls} \quad z < 1/2 : \quad x \leftarrow -x$$

$$9 \quad \mathbf{Rückgabe} \leftarrow x + 2.50662827 \cdot (z - \mathit{freq}(x)) \cdot e^{1/2 \cdot x^2}$$

Alg. 5-5: Approximation der inversen, zentralen und normierten Normalverteilung

In diesem Verfahren wird von der Hilfsfunktion freq Gebrauch gemacht, die wie folgt definiert ist:

$$1 \quad \text{Funktion } \mathit{freq}: \text{Eingabe}=x, \text{ Ausgabe } \mathit{freq}(x)$$

$$2 \quad v \leftarrow |x|$$

$$3 \quad \mathbf{Falls} \quad v < 1/2 : \quad h \leftarrow \frac{3.6767877v - 0.097970465v^3}{3.2584593 + v^2}$$

$$hc \leftarrow 1 - h$$

$$\mathbf{Falls} \quad x > 0 : \quad \mathbf{Rückgabe} \leftarrow \frac{1}{2}h + \frac{1}{2}$$

$$\mathbf{Sonst: Rückgabe} \leftarrow \frac{1}{2}hc$$

$$4 \quad \mathbf{Falls} \quad v < 4.0 : \quad ap \leftarrow 7.37389 + 6.86502v + 3.0318v^2$$

$$\quad \quad \quad + 0.56317v^3 + 0.0000431878v^4$$

$$aq \leftarrow 7.37396 + 15.1849v + 12.7955v^2$$

$$\quad \quad \quad + 5.35422v^3 + v^4$$

$$h \leftarrow 1 - e^{-v^2 \frac{ap}{aq}}$$

$$hc \leftarrow 1 - h$$

$$\mathbf{Falls} \quad x > 0 : \quad \mathbf{Rückgabe} \leftarrow \frac{1}{2}h + \frac{1}{2}$$

$$\mathbf{Sonst: Rückgabe} \leftarrow \frac{1}{2}hc$$

$$5 \quad h \leftarrow 1 - e^{-v^2 \frac{-0.096821 + 0.439821v^2 + 0.248761v^4}{v^3 + 0.440917v^5}}$$

$$6 \quad hc \leftarrow 1 - h$$

$$7 \quad \mathbf{Falls} \quad x > 0 : \quad \mathbf{Rückgabe} \leftarrow \frac{1}{2}h + \frac{1}{2}$$

$$\mathbf{Sonst: Rückgabe} \leftarrow \frac{1}{2}hc$$

Alg. 5-6: Hilfsfunktion $\mathit{freq}(x)$

Ausgehend von der Approximation der inversen Normalverteilung lässt sich die Approximation der inversen Fehlerfunktion dann durch

$$\overline{erf}(x) = \frac{1}{\sqrt{2}} \overline{\Phi}\left(\frac{x+1}{2}\right) \quad (5.29)$$

erreichen [Hil99b].

5.4 Der Zufallszahlengenerator $A_\gamma(c, \sigma)$

Ein Zufallszahlengenerator $A_\gamma(c, \sigma)$, der effizient Ξ_γ -verteilte Zufallszahlen erzeugt, existiert nur bedingt. Es existieren in mathematischen und statistischen Funktionsbibliotheken Prozeduren, die bei gegebener Verteilungsfunktion Φ Zufallszahlen erzeugen, die Φ -verteilt sind. Zwei Beispiele für solche Prozeduren sind FUNLXP und FUNLUX, zwei Realisierungen der Prozedur V152 der `mathlib`. Bei diesen Prozeduren handelt es sich um Realisierungen der Akzeptanz/Ablehnungs-Methode, mit den bereits beschriebenen Vor- und Nachteilen. In einer Initialisierungsphase wird die majorisierende Funktion $t(x)$ durch Approximation von Φ aus Dreieck- und Gaußfunktionen gebildet. Die Anzahl der Stützstellen beträgt fest 200. Die majorisierende Funktion $t(x)$ muss bei jeder Änderung der Parameter, die den Verlauf der Verteilungsfunktion ändern, neu approximiert werden. Für die asymmetrische Verteilungsfunktion Ξ bedeutet dies, dass bei jeder Änderung von c , σ oder γ der FUNLXP- oder FUNLUX-Generator initialisiert werden muss. In der späteren Anwendung wird eine Änderung der Parameter c und σ jedoch mit nahezu jedem Aufruf des Zufallszahlengenerators einhergehen. Damit wird die Anwendung dieser allgemeinen Generatoren in Bezug auf die Geschwindigkeit ineffizient. Ziel muss es daher sein, einen Zufallszahlengenerator zu entwickeln, der auch für häufig wechselnde Parameter ein effizientes Laufzeitverhalten zeigt.

5.4-1 Realisierung

Eine effiziente Realisierung eines $A_\gamma(c, \sigma)$ lässt sich durch Einsatz der Transformation der Inversen in Kombination mit der Kompositionsmethode erreichen. Am Beispiel $c \geq 0$ wird dieses Verfahren beschrieben. Da $\Xi_{c, \sigma}^\gamma$ durch $n = 2$ abschnittsweise definierte Funktionen zusammengesetzt ist, kann die Kompositionsmethode auf diesen beiden Abschnitten arbeiten. Der Wechsel der Abschnitte befindet sich an der Stelle $x = 0$, mit

$${}_{(c \geq 0)} \Xi_{c, \sigma}^\gamma(0) = \frac{1}{1 + (1 + c)^{\gamma/2}} \quad (5.30)$$

nach (4.179). Der Definitionsbereich von $\Xi_{c, \sigma}^\gamma$ ist \mathbb{R} . Aus (5.30) folgt, dass für p_1 und p_2 gilt:

$$\begin{aligned} p_1 &= \frac{1}{1 + (1 + c)^{\gamma/2}} \\ p_2 &= 1 - \frac{1}{1 + (1 + c)^{\gamma/2}} \end{aligned} \quad (5.31)$$

Die Abbildung 5.2 stellt diesen Sachverhalt graphisch dar. Der rot eingefärbte Bereich von $-\infty \leq x < 0$ hat die kumulierte Wahrscheinlichkeit p_1 , der blau eingefärbte Bereich, von $0 \leq x \leq \infty$, die kumulierte Wahrscheinlichkeit p_2 :

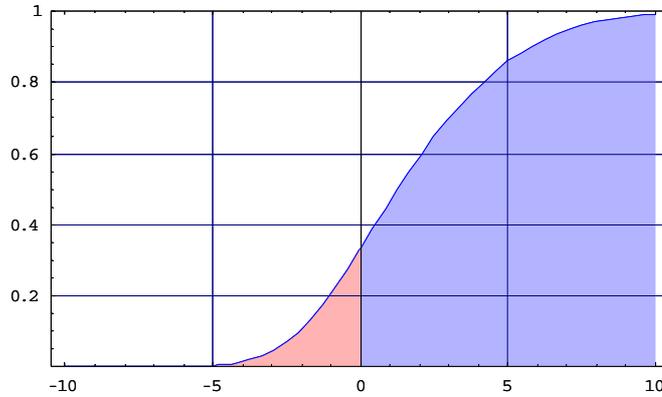


Abb. 5.2: Verlauf der Verteilungsfunktion $\Xi_{c,\sigma}^\gamma(x)$

Für die Komposition von $\Xi_{c,\sigma}^\gamma$ gilt dann:

$$\begin{aligned} \Xi_{c,\sigma}^\gamma(x) &= \sum_{j=1}^n p_j \Phi_j(x) \\ &= \frac{1}{1 + (1+c)^{\gamma/2}} \Phi_1(x) + \left(1 - \frac{1}{1 + (1+c)^{\gamma/2}}\right) \Phi_2(x) \end{aligned} \quad (5.32)$$

bzw. mit Sicht auf eine spätere Implementierung:

$$\Xi_{c,\sigma}^\gamma(x) = I_1(x) \Phi_1(x) + I_2(x) \Phi_2(x), \text{ mit} \quad (5.33)$$

$$I_1(x) = \begin{cases} 1 & \text{falls } x \in \left[0, \frac{1}{1 + (1+c)^{\gamma/2}}\right) \\ 0 & \text{sonst} \end{cases} \quad (5.34)$$

$$I_2(x) = \begin{cases} 1 & \text{falls } x \in \left[\frac{1}{1 + (1+c)^{\gamma/2}}, 1\right) \\ 0 & \text{sonst} \end{cases} \quad (5.35)$$

Für die Verteilungsfunktionen $\Phi_1(x)$ und $\Phi_2(x)$ gilt nach (4.163):

$$\Phi_1(x) = \frac{1 + \operatorname{erf}\left(\frac{x}{\sqrt{2\sigma^\gamma}}\right)}{1 + \sqrt{(1+c)^\gamma}} \quad (5.36)$$

$$\Phi_2(x) = \frac{1 + \sqrt{(1+c)^\gamma} \operatorname{erf}\left(\frac{x}{\sqrt{2(\sigma(1+c))^\gamma}}\right)}{1 + \sqrt{(1+c)^\gamma}} \quad (5.37)$$

Die Kompositionsmethode vereinfacht die Erzeugung von Zufallszahlen jedoch nur, wenn die Φ_i einfacher zu generieren sind als die ursprüngliche Verteilung Φ . Weder Φ_1 noch Φ_2 lassen sich auf bekannte Verteilungen zurückführen. Auch wenn die Verteilungsfunktionen von Φ_1 und Φ_2 an eine Normalverteilung angelehnt sind, lassen sie sich nicht in eine $N(a, \sigma)$ -Verteilung überführen, um bekannte Verfahren zur Erzeugung von normalverteilten Zufallszahlen verwenden zu können. An dieser Stelle muss daher die Transformation der Inversen verwendet werden, um die Φ_i -verteilten Zufallszahlen zu erzeugen.

Im Abschnitt 4.5-3 (Inverse Verteilungsfunktion einer $\Xi_{c,\sigma}^\gamma(x)$ -verteilten Zufallsvariablen, Seite 108) wurden bereits die inversen Verteilungsfunktionen $\bar{\Phi}_1$ und $\bar{\Phi}_2$ bestimmt. Es gilt:

$$\bar{\Phi}_1(y) = \sqrt{2\sigma^\gamma} \operatorname{inverf}(y(1 + \sqrt{(1+c)^\gamma}) - 1) \quad (5.38)$$

$$\bar{\Phi}_2(y) = \sqrt{2(\sigma(1+c))^\gamma} \operatorname{inverf}\left(y + \frac{y-1}{\sqrt{(1+c)^\gamma}}\right). \quad (5.39)$$

Auch hier wird der Verlauf der inversen Verteilungsfunktion graphisch dargestellt:

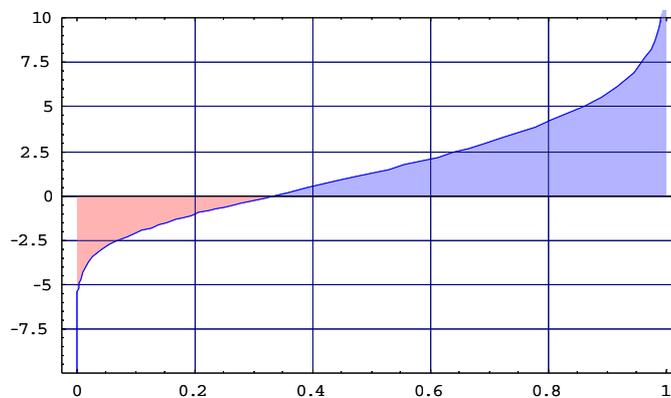


Abb. 5.3: Verlauf der inversen Verteilungsfunktionen $\bar{\Phi}_1(y)$ und $\bar{\Phi}_2(y)$

Mit diesen inversen Verteilungsfunktionen ist es nun möglich, durch Einsatz der Transformation der Inversen, Φ_1 und Φ_2 verteilte Zufallszahlen zu erzeugen. Diese können in (5.33) verwendet werden, um $\Xi_{c,\sigma}^\gamma(x)$ -verteilte Zufallszahlen zu erzeugen. Der Algorithmus sieht wie folgt aus [Hil99b]:

- 1 Funktion asym: Eingabe = γ, c, σ , Ausgabe $\Xi_{c,\sigma}^\gamma(x)$ - verteilte Zufallszahl x
- 2 Erzeuge $G \sim G(0, 1)$
- 3 **Falls** $G < \frac{1}{1 + (1+c)^{\gamma/2}}$: $x \leftarrow \sqrt{2\sigma^\gamma} \operatorname{erf}\left((1+c)^{\gamma/2}G + G - 1\right)$

Sonst: $x \leftarrow \sqrt{2((1+c)\sigma)^\gamma} \cdot (\overline{\text{erf}}(-(1+c)^{-\gamma/2} + (1+c)^{-\gamma/2}G + G))$
4 Rückgabe $\leftarrow x$

Alg. 5-7: Erzeugung einer $\Xi_{c,\sigma}^\gamma(x)$ -verteilten Zufallszahl

5.4-2 Komplexität

Da der Mutationsoperator auf jede Koordinate jedes Individuums angewandt wird, ist eine möglichst schnelle Realisierung dieses Operators von Bedeutung. Der Mutationsoperator besteht ausschließlich aus der Generierung einer $\Xi_{c,\sigma}^\gamma(x)$ -verteilten Zufallszahl sowie einer (in Bezug auf Rechenaufwand) vernachlässigbaren Verknüpfung der Zufallszahl mit dem Ursprungswert der Koordinate. Die Komplexität einer Funktion lässt sich in Bezug auf den Speicherverbrauch und den Verbrauch an Elementaroperationen messen. Der Speicherverbrauch des Generators $A_\gamma(c, \sigma)$ ist im aufwendigsten Fall 6, da 6 reelle Hilfsvariablen zur Speicherung von Zwischenwerten benötigt werden. Die folgende Tabelle gibt die Anzahl der benötigten Operationen für die einzelnen Algorithmenteile an. Unterschieden werden hier Vergleiche, Additionen, Multiplikationen und Logarithmierungen. Da die Anzahl der benötigten Operationen von der Belegung der Parameter c , σ und γ abhängt, beziehen sich alle Angaben auf den Worst-Case (schlechtesten Fall).

	Alg. 5-5 ($\overline{\Phi}(z)$)	Alg. 5-6 ($\text{freq}(x)$)	Gl. (5.29)	Alg. 5-7 ($A_\gamma(c, \sigma)$)	Σ
Vergleiche	4	5	-	1	10
Additionen	9	21	1	5	36
Multiplikationen	9	24	2	7	42
Logarithmierung	3	2	1	3	9

Tab. 5-2: Anzahl der benötigten Operationen zur Bestimmung einer $\Xi_{c,\sigma}^\gamma(x)$ -verteilten Zufallszahl

Zusätzlich zu den hier aufgeführten Operationen muss der Aufwand zur Bestimmung einer $G(0, 1)$ -verteilten Zufallszahl berücksichtigt werden. Der Aufwand zur Bestimmung einer $\Xi_{c,\sigma}^\gamma(x)$ -verteilten Zufallszahl ist damit deutlich höher als der Aufwand zur Bestimmung einer normalverteilten Zufallszahl nach Box und Muller [Box58]. In praktischen Anwendungen ist der Rechenaufwand jedoch zu vernachlässigen, da die Bestimmung des Fitnesswertes eines Individuums in der Regel um Größenordnungen aufwendiger ist als die Durchführung der Mutationsfunktion.

5.5 Zusammenfassung

In den Abschnitten 5.1-1 (Genauigkeit, Seite 150), 5.1-2 (Effizienz, Seite 152), 5.1-3 (Robustheit, Seite 153) und 5.1-4 (Reproduzierbarkeit, Seite 153) sind wichtige Eigenschaften erläutert worden, über die ein Zufallszahlengenerator verfügen sollte. In diesem Abschnitt wird der Zufallszahlengenerator $A_\gamma(c, \sigma)$ hinsichtlich dieser Eigenschaften bewertet.

1. Genauigkeit

Der $A_\gamma(c, \sigma)$ Generator erfüllt die Anforderungen an die Genauigkeit. Mit einer Signifikanz von $\alpha = 0.05$ kann weder die H_0 -Hypothese des χ^2 -Tests, noch die des Kolmogorov-Smirnov-Tests abgelehnt werden.

2. Effizienz

In dem vorhergehenden Abschnitt ist der Ressourcenverbrauch des $A_\gamma(c, \sigma)$ Generators bestimmt worden. Sowohl in Bezug auf Speicherverbrauch als auch in Bezug auf Verbrauch an Rechenzeit ist der Generator für die Anwendung als Mutationsoperator bei asymmetrischen Evolutionsstrategien als effizient einzustufen.

3. Robustheit

Der $A_\gamma(c, \sigma)$ Generator verwendet ausschließlich ein internes Verfahren zur Erzeugung der Zufallszahlen. Dieses Verfahren basiert auf einer linearen, d. h. nicht iterativen oder rekursiven Abfolge von Operationen. Das Verhalten des Generators ist für alle Wertebereiche der erzeugten Zahlen nahezu identisch. Dies äußert sich auch im Vergleich von Best-Case, Average-Case und Worst-Case Szenarien, bei denen der Verbrauch an Rechenzeit nahezu konstant ist.

4. Reproduzierbarkeit

Der Generator $A_\gamma(c, \sigma)$ verwendet lediglich im Algorithmus 5-7 eine stochastische Variable. Zur Bestimmung einer einzelnen Zufallszahl wird immer genau einmal diese stochastische Variable realisiert. Die Folge, die der Generator $A_\gamma(c, \sigma)$ erzeugt, wird daher ausschließlich durch die Initialisierung des Zufallszahlengenerators zur Erzeugung dieser stochastischen Variable bestimmt. Da der hier verwendete PMMLC-Generator die Anforderung an Reproduzierbarkeit erfüllt, erzeugt auch der Generator $A_\gamma(c, \sigma)$ reproduzierbare Folgen von Zufallszahlen.

Der Zufallszahlengenerator $A_\gamma(c, \sigma)$, dessen Herleitung und Implementierung in diesem Kapitel beschrieben wurde, ist somit ein effizienter und robuster Generator, der alle Anforderungen an die gewünschte Genauigkeit erfüllt. Zudem ist er in der Lage, reproduzierbare Folgen von Zufallszahlen zu erzeugen, die eine exakte Wiederholung von Experimenten erlauben. Durch eine pseudo-zufällige Initialisierung, z. B. durch Einsatz einer rechner-internen Uhr, ist es natürlich möglich, nicht reproduzierbare Folgen von Zufallszahlen zu erhalten.

Kapitel 6 Asymmetrische Evolutionstrategien

*Worte sind Zaubersprüche und
Namen sind es erst recht.
(Rainer Maria Rilke)*

Die asymmetrische Evolutionstrategie (aES) ist eine Erweiterung der (μ, λ) -Evolutionstrategie mit linearer Adaptation der Schrittweiten. Im Gegensatz zur (μ, λ) -Evolutionstrategie wird zur Mutation keine normalverteilte Zufallszahl, sondern eine asymmetrisch verteilte Zufallsvariable verwendet. Auf eine Adaptation von Lagewinkeln wird zugunsten einer Adaptation der Schiefe- maße verzichtet. Damit beträgt die Anzahl der zu adaptierenden Strategieparameter $2n$, bei gegebener Dimensionalität des Optimierproblems n . Im folgenden wird die aES ausführlich beschrieben.

6.1 Asymmetrische Evolutionstrategie

Eine aES ist, wie die klassischen Varianten der Evolutionstrategien, ein populationsbasiertes Opti- miervorgehen. In einer Generation befinden sich μ Elternindividuen, aus denen durch Rekombina- tion λ Nachkommen erzeugt werden. Diese Nachkommen werden mutiert, wobei diese Mutation auf einer asymmetrischen Verteilungsfunktion basiert. Die asymmetrische Verteilung wird durch eine Schrittweite σ und ein Schiefemaß c gesteuert. Die Schrittweite σ ermöglicht es, die asym- metrische Verteilungsfunktion ohne Bevorzugung einer Koordinatenrichtung zu verändern. Das Schiefemaß c steuert das Ausmaß der Bevorzugung einer Koordinatenrichtung. Für jede Dimension des Optimierproblems existiert eine Schrittweite σ_i und ein Schiefemaß c_i . Die Strategieparameter einer Koordinatenachse sind unabhängig von allen weiteren Koordinatenachsen. Sowohl Schritt- weite als auch Schiefemaß unterliegen einer Selbstadaptation.

6.1-1 Grundlegender Algorithmus

Die informelle Beschreibung einer aES des einführenden Abschnitts wird nun formalisiert. Eine aES ist ein 14-Tupel

$$(\mu, \lambda)\text{-aES} = (P^{(0)}, \mu, \lambda, \text{rek}, \text{mut}, \text{sel}, \tau, \tau', \tau_c, \tau_c', \gamma, f, g, \kappa), \text{ mit} \quad (6.1)$$

$$\begin{aligned} P^{(0)} &= (\mathbf{I}_1^{(0)}, \dots, \mathbf{I}_\mu^{(0)}) \in I^\mu, \text{ mit } I = \mathbb{R}^n \times \mathbb{R}_+^n \times \mathbb{R}^n && \text{Startpopulation} \\ \mu &: \mu > 1 && \text{Anzahl der Eltern} \end{aligned}$$

λ	: $\lambda > \mu$	Anzahl der Nachkommen
rek	: $I^\mu \rightarrow I$	Rekombinationsoperator
mut	: $I \rightarrow I$	Mutationsoperator
sel	: $I^\lambda \rightarrow I^\mu$	Selektionsoperator
τ, τ'	: $\tau, \tau' \in \mathbb{R}_+$	Steuerparameter Schrittweiten
τ_c, τ_c'	: $\tau_c, \tau_c' \in \mathbb{R}_+$	Steuerparameter Schiefemaße
γ	: $\gamma \in \mathbb{R}_+$	globaler Steuerparameter
f	: $\mathbb{R}^n \rightarrow \mathbb{R}$	Zielfunktion
g_j	: $\mathbb{R}^n \rightarrow \mathbb{R}$, mit $j \in \{1, \dots, q\}$	Nebenbedingungen
κ	: $I^\mu \rightarrow \{0, 1\}$	Abbruchkriterium

Den Ablauf einer aES zeigt der Algorithmus 6-1:

```

1   t:=0;
2   initialisiere  $P^{(t)} := \{\mathbf{I}_1^{(t)}, \dots, \mathbf{I}_\mu^{(t)}\}$  ;
3   solange  $\kappa(P^{(t)}) \neq 1$ 
4        $\forall(k \in \{1, \dots, \lambda\}) : \mathbf{I}_k^{(t)} := rek(P^{(t)})$  ;           //Rekombination
5        $\forall(k \in \{1, \dots, \lambda\}) : \mathbf{I}_k^{(t)} := mut(\mathbf{I}_k^{(t)})$  ;           //Mutation
6        $P^{n(t)} := \{\mathbf{I}_1^{n(t)}, \dots, \mathbf{I}_\lambda^{n(t)}\}$  ;
7        $\Psi(P^{n(t)}) = (\Psi(\mathbf{I}_1^{n(t)}), \dots, \Psi(\mathbf{I}_\lambda^{n(t)}))$  ;           //Bewertung
8        $P^{(t+1)} := sel(P^{n(t)})$  ;           //Selektion
9       t:= t + 1 ;
10  ende
```

Alg. 6-1: Ablauf einer (μ, λ) -asymmetrischen Evolutionsstrategie

6.1-2 Repräsentation

Bei einer asymmetrischen Evolutionsstrategie besteht ein Individuum aus drei Komponenten, der Objektkomponente $\mathbf{x}_i^{(t)}$ sowie den beiden Strategiekomponenten Schrittweite $\sigma_i^{(t)}$ und Schiefemaß $\mathbf{c}_i^{(t)}$:

$$\mathbf{I}_i^{(t)} = (\mathbf{x}_i^{(t)}, \sigma_i^{(t)}, \mathbf{c}_i^{(t)}) \quad (6.2)$$

Die Schrittweite wird auf positive reelle Zahlen eingeschränkt. Die Elemente $c_{i,j}^{(t)}$ des Schiefemaßes $\mathbf{c}_i^{(t)}$ erlauben für $c_{i,j}^{(t)} < 0$ eine linksschiefe, für $c_{i,j}^{(t)} = 0$ eine symmetrische und für $c_{i,j}^{(t)} > 0$ eine rechtsschiefe Mutation.

6.1-3 Rekombination

Zur Erzeugung der Population $P^{(t+1)}$ werden die Eltern im ersten Schritt rekombiniert (Zeile 4):

$$\forall (k \in \{1, \dots, \lambda\}) : \mathbf{I}_k^{(t)} := \text{rek}(P^{(t)}) \quad (6.3)$$

Die Komponenten \mathbf{x} , $\boldsymbol{\sigma}$ und \mathbf{c} werden gemäß der Empfehlung von Bäck und Hoffmeister unterschiedlich rekombiniert [Hof92]. Für die Objektvariable \mathbf{x} wird eine globale Rekombination, für die Strategieparameter $\boldsymbol{\sigma}$ und \mathbf{c} eine intermediäre Rekombination verwendet. Die Rekombination wird wie in Abschnitt 2.4-3 (Rekombination, Seite 29) beschrieben durchgeführt.

6.1-4 Mutation

In dem folgenden Schritt werden die Komponenten der λ Individuen $\mathbf{I}_i^{(t)}$ mutiert (Zeile 5):

$$\forall (k \in \{1, \dots, \lambda\}) : \mathbf{I}_k^{(t)} := \text{mut}(\mathbf{I}_k^{(t)}) \quad (6.4)$$

Die Mutation erfolgt in zwei Schritten: Im ersten Schritt werden die Strategievariablen mutiert, im zweiten Schritt wird mit diesen neuen Strategievariablen die Objektvariable mutiert:

$$\boldsymbol{\sigma}_i^{(t)} = \text{mut}(\boldsymbol{\sigma}_i^{(t)}) \quad (6.5)$$

$$\mathbf{c}_i^{(t)} = \text{mut}(\mathbf{c}_i^{(t)}) \quad (6.6)$$

$$\mathbf{x}_i^{(t)} = \text{mut}(\mathbf{x}_i^{(t)}) \quad (6.7)$$

Durch dieses Vorgehen ist es möglich, die Qualität der Strategievariablenmutation indirekt über die Fitness des erzeugten Individuums zu bestimmen. Die Mutation erfolgt Komponentenweise mit den Steuerfaktoren τ , τ' und τ_c , τ_c' für die Mutation der Strategievariablen und dem globalen Steuerparameter γ . Für die Schrittweite gilt:

$$\sigma_{i,j}^{(t)} = \sigma_{i,j}^{(t)} \cdot z \cdot z_i \quad (6.8)$$

mit der Zufallsvariablen

$$z \sim LN(0, \tau') \quad (6.9)$$

die einmal generiert für die gesamte Generation gilt und der Zufallsvariablen

$$z_i \sim LN(0, \tau), \text{ mit } 1 \leq i \leq n \quad (6.10)$$

die für jede Koordinate neu realisiert wird. Für die Steuerparameter τ und τ' werden die Empfehlungen von Schwefel, die aus dem Kugelmodell abgeleitet sind, übernommen. Es gilt:

$$\tau = 1 / \sqrt{2 \sqrt{n}} \quad (6.11)$$

$$\tau' = 1 / \sqrt{2n} \quad (6.12)$$

Für die Schiefemaße gilt folgender Mutationsschritt:

$$c''_{i,j} = c'_{i,j} + z^* + z_i^* \quad (6.13)$$

mit der Zufallsvariablen

$$z^* \sim N(0, \tau_c') \quad (6.14)$$

die einmal generiert für die gesamte Generation gilt und der Zufallsvariablen

$$z_i^* \sim N(0, \tau_c), \text{ mit } 1 \leq i \leq n \quad (6.15)$$

die für jede Koordinate neu realisiert wird. Die Steuerparameter τ_c und τ_c' werden identisch zu den Parametern τ und τ' berechnet. Die Mutation der Schrittweite erfolgt nicht durch Addition, sondern durch Multiplikation mit der logarithmischen Normalverteilung, um negative Schrittweiten zu verhindern. Die Mutation der Schiefemaße erfolgt durch Addition, da für die Schiefemaße keine Einschränkungen existieren.

Mit den so mutierten Strategievariablen ist eine Mutation der Objektvariablen möglich

$$x''_{i,j} = x'_{i,j} + A_\gamma(c''_{i,j}, \sigma''_{i,j}) \quad (6.16)$$

die den im Kapitel 5 (Implementierung, Seite 149) beschriebenen Zufallszahlengenerator $A_\gamma(c, \sigma)$ verwendet. Nach der Bewertung der Individuen (Zeile 7) wird die Selektion durchgeführt (Zeile 8). Es handelt sich bei der Selektion der asymmetrischen Evolutionsstrategie um eine Komma-Selektion, die lediglich aus den Nachkommen die Eltern der folgenden Generation auswählt.

6.1-5 Selbstadaptation der Strategieparameter

Da eine für die Optimierung geeignete Belegung der Schrittweiten und Schiefemaße sich während des Laufs einer aES ändert, und daher im Voraus nicht getroffen werden kann, unterliegen die beiden Strategiekomponenten Schrittweite und Schiefemaß einer Selbstadaptation. Die Strategiekomponenten werden vor der Objektkomponente mutiert. Diese mutierten Strategiekomponenten bilden die Grundlage für die Mutation der Objektkomponente. Die Bewertung der Strategiekomponente erfolgt indirekt über die Güte der Objektkomponente.

Dieses Verfahren unterscheidet sich nicht von den Adaptationsmechanismen, die von Rechenberg und Schwefel in den ursprünglichen Arbeiten vorgeschlagen wurden. Es findet somit keine ent-stochastisierte Mutation nach Ostermeier und Hansen statt. Die Mutation der Schrittweiten findet identisch zur Adaptation der Schrittweiten statt, die von Schwefel vorgeschlagen wurde und in Abschnitt 2.4 ((μ, λ) -Evolutionstrategie, Seite 27) beschrieben wurde. Die Belegung der Parameter τ_c und τ_c' ist ebenfalls identisch zu der Belegung von Schwefel. Die Mutation der Schiefemaße erfolgt analog zur Mutation der Schrittweiten. Die Initialisierung der Strategiekomponenten der Startpopulation erfolgt wie nachstehend beschrieben:

$$\sigma''_{i,j} = 1, \text{ mit } 1 \leq i \leq \mu \text{ und } 1 \leq j \leq n \quad (6.17)$$

$$c_{i,j}^{(t)} = 0, \text{ mit } 1 \leq i \leq \mu \text{ und } 1 \leq j \leq n \quad (6.18)$$

Da die asymmetrische Evolutionsstrategie n Schrittweiten und n Schiefemaße verwendet, beträgt die Gesamtzahl der Strategievariablen $2n$.

Die folgenden Abbildungen zeigen beispielhaft anhand von 2-dimensionalen Beispielen der Dichtefunktion (hier Mutationsraum genannt), welche Formen die Verteilung für spezielle Werte von σ_1 , σ_2 , c_1 und c_2 annehmen kann. In diesen Beispielen ist der globale Parameter γ konstant auf den Wert 2 gesetzt.

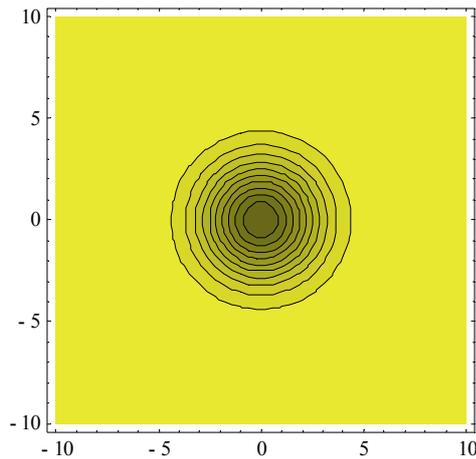


Abb. 6.1: Mutationsraum einer aES

$$\begin{aligned} \sigma_1 &= 2 \\ \sigma_2 &= 2 \\ c_1 &= 0 \\ c_2 &= 0 \end{aligned} \quad (6.19)$$

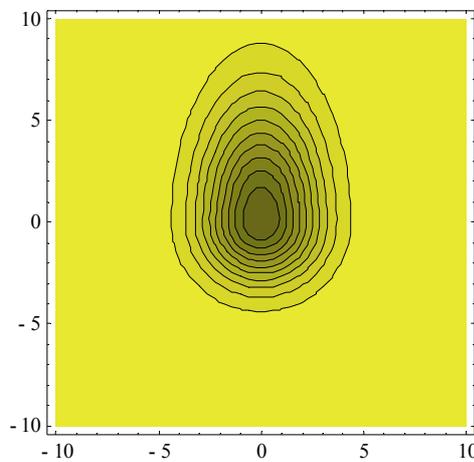


Abb. 6.2: Mutationsraum einer aES

$$\begin{aligned} \sigma_1 &= 2 \\ \sigma_2 &= 2 \\ c_1 &= 0 \\ c_2 &= 1 \end{aligned} \quad (6.20)$$

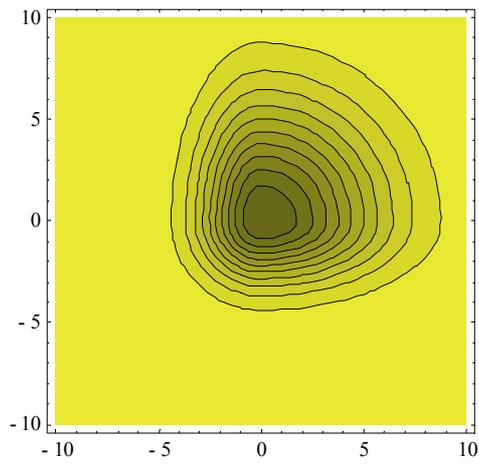


Abb. 6.3: Mutationsraum einer aES

$$\begin{aligned} \sigma_1 &= 2 \\ \sigma_2 &= 2 \\ c_1 &= 1 \\ c_2 &= 1 \end{aligned} \quad (6.21)$$

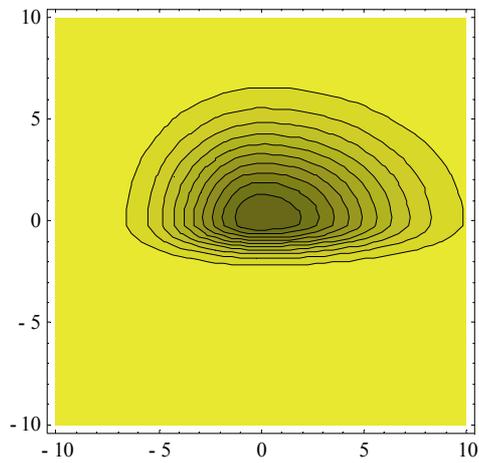


Abb. 6.4: Mutationsraum einer aES

$$\begin{aligned} \sigma_1 &= 3 \\ \sigma_2 &= 1 \\ c_1 &= 1/2 \\ c_2 &= 2 \end{aligned} \quad (6.22)$$

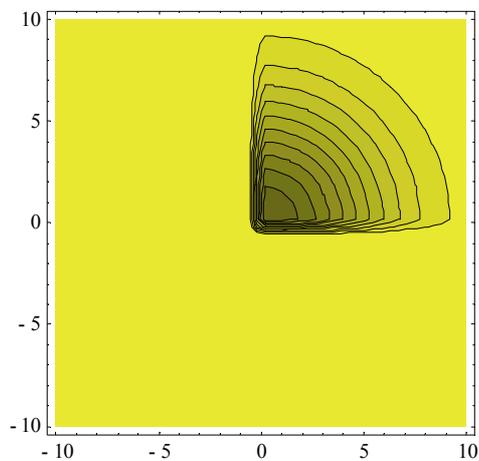


Abb. 6.5: Mutationsraum einer aES

$$\begin{aligned} \sigma_1 &= 1/5 \\ \sigma_2 &= 1/5 \\ c_1 &= 20 \\ c_2 &= 20 \end{aligned} \quad (6.23)$$

Der Mutationsraum kann die entsprechenden Mutationsräume der Evolutionsstrategie mit Adaptation einer Schrittweite sowie die Mutationsräume der Evolutionsstrategie mit Adaptation von n Schrittweiten nachbilden. Eine Rotation der Mutationsräume, wie sie bei der Evolutionsstrategie mit Adaptation von Schrittweiten und Lagewinkeln möglich ist, kann durch die asymmetrische Evolutionsstrategie nicht nachgebildet werden. Dafür ist die asymmetrische Evolutionsstrategie in der Lage, Mutationsräume zu erzeugen, die in einer starken Bevorzugung einzelner Koordinatenrichtungen resultieren. Beispiele hierfür sind in den Abbildungen 6.2 - 6.5 gegeben.

6.2 Eigenschaften der asymmetrischen Evolutionsstrategien

In diesem Abschnitt werden theoretische Aspekte der asymmetrischen Evolutionsstrategien betrachtet. Es wird untersucht, ob die klassischen Ansätze von Rechenberg und Schwefel zur Bestimmung von Konvergenzgeschwindigkeit und optimalen Schrittweiten, die für das Kugelmodell und das Korridormodell durchgeführt wurden, auf die asymmetrischen Evolutionsstrategien übertragbar sind. Danach werden spezielle Aspekte der asymmetrischen Evolutionsstrategien aufgezeigt, die einen von symmetrischen Evolutionsstrategien abweichenden, zusätzlichen Konvergenzmechanismus aufzeigen.

6.2-1 Klassische Beweisstrategien

Eine theoretische Untersuchung der Eigenschaften einer asymmetrischen Evolutionsstrategie erweist sich als weitaus aufwendiger als die einer symmetrischen Evolutionsstrategie, da viele Vereinfachungen und Sätze nur für normalverteilte Mutationen gelten. Beispiele hierfür sind:

1. Identisch verteilte Zufallszahlen

Viele Vereinfachungen setzen die Schrittweiten der einzelnen Dimensionen gleich, d. h.

$$\sigma_i = \sigma \quad (6.24)$$

Diese Vereinfachung kann für asymmetrische Evolutionsstrategien ebenfalls verwendet werden. Die in Konsequenz daraus folgende Vereinfachung, dass alle Schiefemaße identisch angenommen werden, d. h.

$$c_i = c \quad (6.25)$$

ist jedoch nicht zulässig, da sie den Hauptmechanismus, der eine asymmetrische Evolutionsstrategie charakterisiert, ausschaltet.

2. Identische Eltern

Aussagen über die Fortschrittsrate einer (μ, λ) -ES gelten häufig unter der Annahme, dass alle Elternindividuen einer Generation identisch sind. Aus o. a. Gründen schließt eine solche Vereinfachung die Betrachtung von Asymmetriemaßen aus.

3. Reduktion von $\sum_{i=1}^n z_i^2$ auf χ^2 -Verteilung

Diese Vereinfachung ist ein elementarer Schritt für die geschlossene Darstellung der Länge des Zufallsvektors nach (2.32). Eine entsprechende Vereinfachung für $\Xi_{c, \sigma}^{\gamma}$ -verteilte Zufallsvariablen existiert nicht.

4. Fehlende Rückführbarkeit

Ein Mutationsoperator, der auf $\Xi_{c, \sigma}^\gamma$ -verteilten Zufallsvariablen basiert, kann nicht auf einen Mutationsoperator, der auf normalverteilten Zufallsvariablen basiert, zurückgeführt werden. Im Abschnitt 4.7-2 (Vollständige Äquivalenz zur Normalverteilung, Seite 130) wurde gezeigt, dass diese Rückführung nur für $\gamma = 2$ und die triviale Belegung $c = 0$ gilt. Auch die abschnittsweise Rückführung ist, wie in Abschnitt 4.7-3 (Fallweise Äquivalenz zur Normalverteilung, Seite 131) gezeigt, nur für die triviale Belegung $\gamma = 2$ und $c = 0$ möglich.

Ein weitere Herausforderung besteht in der Tatsache, dass die Dichte- bzw. Verteilungsfunktion von $\Xi_{c, \sigma}^\gamma$ -verteilten Zufallsvariablen abschnittsweise definiert ist. Die Herleitung entsprechender Sätze und Vereinfachungen gestaltet sich daher weit aufwendiger und ist für den allgemeinen Fall nicht geschlossen darstellbar.

6.2-2 Asymmetrischer Konvergenzmechanismus

Grundsätzlich stellt sich hier die Frage, ob die theoretischen Resultate bezüglich der Konvergenzgeschwindigkeit oder der optimalen Schrittweite auf die asymmetrischen Evolutionsstrategien vollständig übertragbar sind. Im Gegensatz zu allen anderen in dieser Arbeit vorgestellten Varianten der Evolutionsstrategien gilt für die aES, dass die verwendeten Zufallsvariablen z_i im allgemeinen nicht zentriert sind, d.h.:

$$E(z_i) \neq 0 \quad (6.26)$$

gilt fast immer. Anders als bei den symmetrischen Varianten der ES ist es bei der aES damit möglich, die Strategiekomponenten so zu wählen, dass der Erwartungswert mit dem Optimum zusammenfällt. Die Wahl der optimalen Strategiekomponenten erfolgt dabei koordinatenweise. Sei $f(\mathbf{x})$ die zu minimierende n -dimensionale Zielfunktion und die Stelle

$$\mathbf{x}^* = (x_1^*, \dots, x_n^*), \text{ mit } 1 \leq i \leq n \quad (6.27)$$

die Position des Optimums. Sei $\mathbf{I}_i^{(t)}$ ein beliebiges Individuum der Generation t , mit n Schrittweiten $\sigma_{i,j}^{(t)} > 0$, $1 \leq j \leq n$ und n Schiefemaßen $c_{i,k}^{(t)}$, $1 \leq k \leq n$. Dann existiert für jede Koordinate x_l , $1 \leq l \leq n$ der Objektkomponente \mathbf{x} ein Paar σ_l^* und c_l^* sowie eine reelle Zahl $\gamma > 0$, sodass eine $\xi_{c_l^*, \sigma_l^*}^\gamma$ -verteilte Zufallszahl z_l^* existiert, mit

$$E(z_l^*) = x_l^* \quad (6.28)$$

Für den Erwartungswert einer $\xi_{c, \sigma}^\gamma$ -verteilten Zufallszahl Z gilt nach (4.186):

$$E(Z) = \frac{c\sqrt{2\sigma^\gamma}}{\sqrt{\pi}(1 + (1 + |c|)^\gamma)} \quad (6.29)$$

Mit $E(Z) = x_l^*$ gilt für die Koordinaten der Strategiekomponenten:

$$c_l^* = \begin{cases} 1 - 4^{-1/\gamma} (2 - x_l^* \sqrt{2\pi} (\sigma_l^{*- \gamma/2}))^{2/\gamma} & \text{für } x_l^* < 0 \\ -1 + 4^{-1/\gamma} (2 + x_l^* \sqrt{2\pi} (\sigma_l^{*- \gamma/2}))^{2/\gamma} & \text{für } x_l^* \geq 0 \end{cases} \quad (6.30)$$

$$\sigma_l^* = \begin{cases} \left(\frac{x_l^*}{\sqrt{(1 - c_l^*)^\gamma - 1}} \right)^{2/\gamma} \left(\frac{\pi}{2} \right)^{1/\gamma} & \text{für } x_l^* < 0 \wedge c_l^* < 0 \\ \left(\frac{x_l^*}{\sqrt{(1 + c_l^*)^\gamma - 1}} \right)^{2/\gamma} \left(\frac{\pi}{2} \right)^{1/\gamma} & \text{für } x_l^* > 0 \wedge c_l^* > 0 \\ \text{beliebig} & \text{für } x_l^* = 0 \wedge c_l^* = 0 \\ \text{nicht definiert} & \text{sonst} \end{cases} \quad (6.31)$$

Die Gleichung (6.30) ist für beliebige, zulässige σ_l^* definiert. Die Gleichung (6.31) ist nur für bestimmte Kombinationen von x_l^* und c_l^* definiert. Diese Definitionslücken in (6.31) spiegeln die Eigenheit der aES wieder: Nur für Schiefeße $c_l^* > 0$ ist ein positiver Erwartungswert bzw. für $c_l^* < 0$ ein negativer Erwartungswert der $\xi_{c_l^*, \sigma_l^*}^\gamma$ -verteilten Zufallszahlen möglich.

Für die Verteilungen der Abschnitte 4.3 (Asymmetrische Verteilungsfunktion $\Xi_c(x)$, Seite 72) und 4.4 (Asymmetrische Verteilungsfunktion $\Xi_{c,\sigma}(x)$, Seite 90) reduzieren sich die Gleichungen (6.30) und (6.31) wie folgt:

$$c_l^* = \begin{cases} 1 - \frac{1}{4} (2 - x_l^* \sqrt{2\pi})^2 & \text{für } x_l^* < 0 \\ -1 + \frac{1}{4} (2 + x_l^* \sqrt{2\pi})^2 & \text{für } x_l^* \geq 0 \end{cases} \quad (6.32)$$

$$c_l^* = \begin{cases} 1 - \frac{1}{4} \left(2 - x_l^* \frac{\sqrt{2\pi}}{\sqrt{\sigma_l^*}} \right)^2 & \text{für } x_l^* < 0 \\ -1 + \frac{1}{4} \left(2 + x_l^* \frac{\sqrt{2\pi}}{\sqrt{\sigma_l^*}} \right)^2 & \text{für } x_l^* \geq 0 \end{cases} \quad (6.33)$$

$$\sigma_i^* = \begin{cases} \frac{\pi(x_i^*)^2}{2(\sqrt{1-c_i^*}-1)^2} & \text{für } x_i^* < 0 \wedge c_i^* < 0 \\ \frac{\pi(x_i^*)^2}{2(\sqrt{1+c_i^*}-1)^2} & \text{für } x_i^* > 0 \wedge c_i^* > 0 \\ \text{beliebig} & \text{für } x_i^* = 0 \wedge c_i^* = 0 \\ \text{nicht definiert} & \text{sonst} \end{cases} \quad (6.34)$$

Die in Abschnitt 4.3 (Asymmetrische Verteilungsfunktion $\Xi_c(x)$, Seite 72) eingeführte Verteilung besitzt keine Schrittweitensteuerung. Daher existiert lediglich die in (6.32) dargestellte Anpassung von c_i^* .

In der folgenden Abbildung ist am Beispiel eines 3-dimensionalen Koordinatensystems die gezielte Steuerung der Schiefemaße zur Anpassung des Erwartungswertes graphisch dargestellt. Ein Elternindividuum (in der Graphik rot dargestellt) befindet sich an der Stelle $(x_1, x_2, x_3)^T = (-1, -4, 2)^T$. Das Optimum befindet sich an der Stelle $(x_1^*, x_2^*, x_3^*)^T = (0, 0, 0)^T$. Für dieses Beispiel gilt zudem, dass alle Schrittweiten σ_i mit $1 \leq i \leq 3$ mit 1.0 belegt sind und $\gamma = 2.0$ gilt.

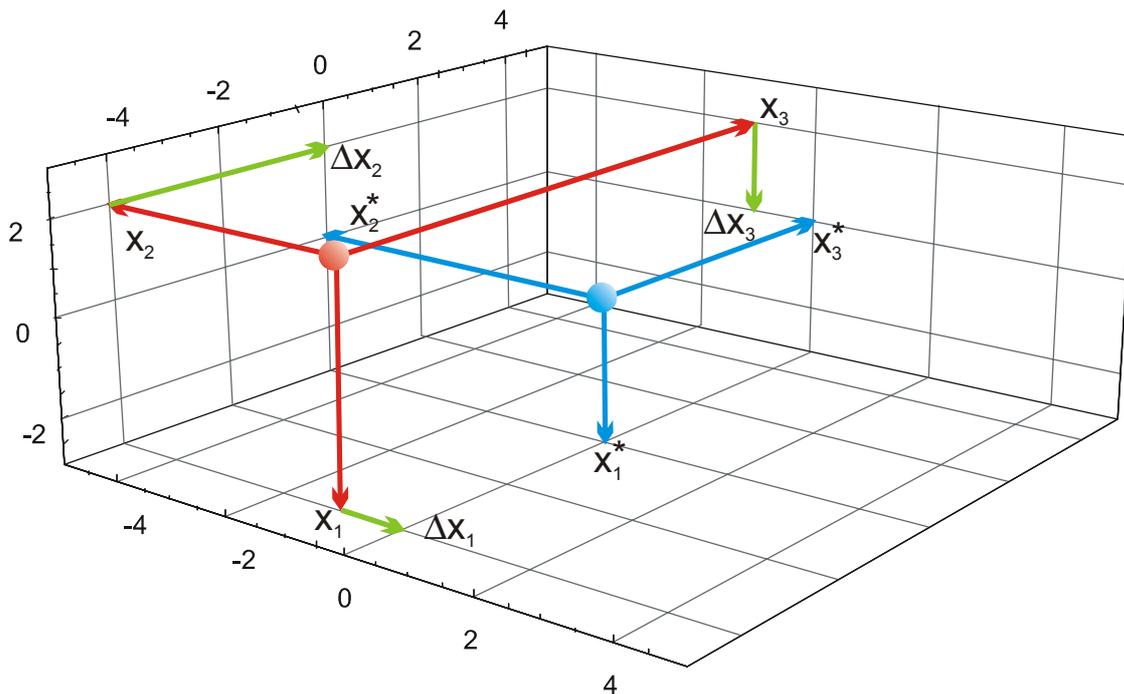


Abb. 6.6: Definition der Variablen zur Steuerung der Schiefemaße

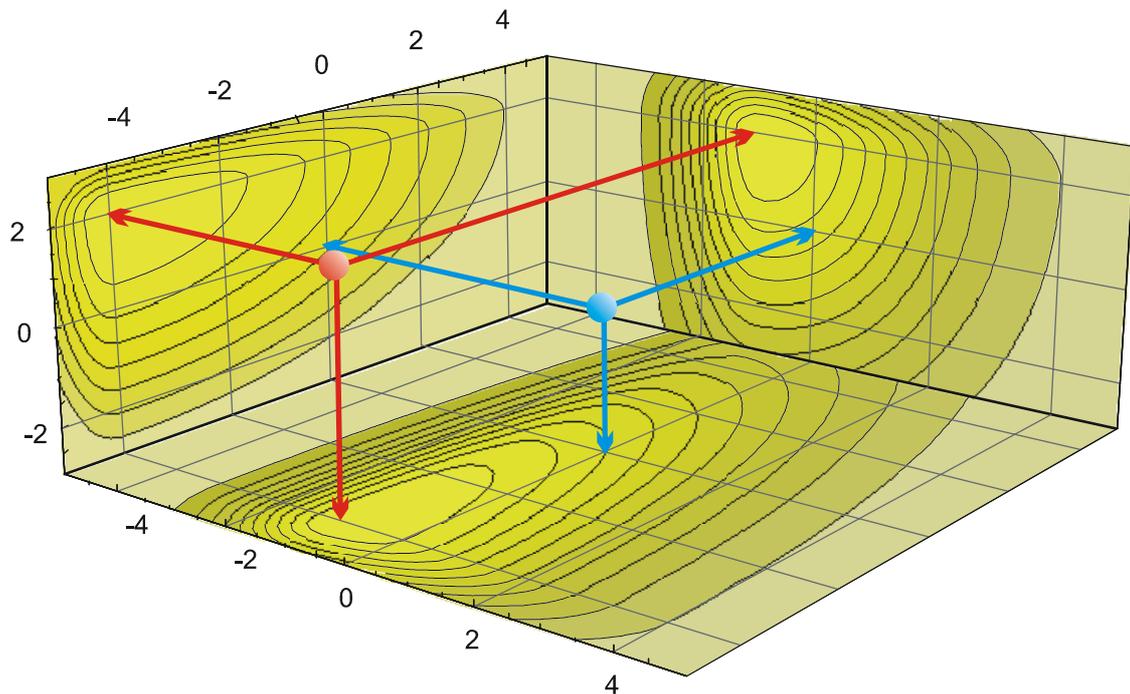


Abb. 6.7: Position von Elter und Nachkomme

6.2-3 Reversibilität

Der Einsatz gerichteter Mutation, wie sie in dieser Arbeit am Beispiel der asymmetrischen Ξ -Verteilungen vorgestellt wird, birgt die Gefahr, dass eine Richtung adaptiert wird, die im Laufe der Optimierung ungünstig wird. Durch die Selektion werden auch in diesem Fall nur die Individuen mit der höchsten Fitness ausgewählt. Bei einer ungeeigneten Richtungsadaptation sollten dies die Individuen sein, die in andere, geeignetere Richtungen orientiert sind. Für den Fall das ausreichend unterschiedliche Individuen erzeugt werden, kann diese Umorientierung durch die Selektion erreicht werden. Ist die Diversität, speziell in Bezug auf die Orientierung der Individuen, jedoch gering, können nur relativ wenige geeignete Individuen selektiert werden. Dies kann in einer relativ langen Umorientierungsphase resultieren. Ausschlaggebend für dieses Phänomen ist eine Eigenschaft der, für die Mutation verwendeten Verteilungsfunktion: Das Verhältnis der Wahrscheinlichkeit p_n negative Zufallszahlen zu erzeugen zu der Wahrscheinlichkeit p_p positive Zufallszahlen zu erzeugen¹. Diese Eigenschaft wird in dieser Arbeit Reversibilität genannt und im folgenden definiert und an Beispielen erläutert.

Definition 6.1: Reversibilität

Für eine beliebige, stetige und über \mathbb{R} definierte Zufallsvariable X mit Dichtefunktion $\varphi(x)$ wird die Reversibilität ρ wie folgt definiert:

1. Diese Aussage gilt für den Fall eines positiven Schiefemaßes der Ξ -Verteilung. Alle Aussagen gelten analog für den Fall eines negativen Schiefemaßes.

$$\rho = \frac{P_n}{P_p} = \frac{\int_{-\infty}^0 \varphi(x) dx}{\int_0^{\infty} \varphi(x) dx} \quad (6.35)$$

Die Reversibilität ρ nimmt Werte aus dem reellwertigen Intervall $[0, 1]$ an. Für jede zentrale Normalverteilung gilt, dass für die Reversibilität $\rho = 1$ gilt, da die Wahrscheinlichkeit für die Erzeugung positiver Zufallszahlen identisch zur Wahrscheinlichkeit der Erzeugung negativer Zufallszahlen ist. Es gilt für die zentrale Normalverteilung mit Dichtefunktion $\varphi(x)$:

$$\int_{-\infty}^0 \varphi(x) dx = \int_0^{\infty} \varphi(x) dx = \frac{1}{2} \quad (6.36)$$

Für die nicht-zentrale Normalverteilung mit Lageparameter a und Dichtefunktion

$$\varphi_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-a)^2} \quad (6.37)$$

gilt für die Reversibilität ρ :

$$\rho = \frac{1 - \operatorname{erf}(a/\sqrt{2})}{1 + \operatorname{erf}(a/\sqrt{2})} \quad (6.38)$$

Eine nicht-zentrale Normalverteilung wird von Weicker eingesetzt, um die Verfolgung eines dynamischen Optimums zu ermöglichen. Das Verfahren wird in dem Abschnitt 3.8 (Adaptation einer nicht-zentralen Normalverteilung, Seite 54) erläutert. An dieser Stelle wird der Verlauf der Reversibilität für die adaptierten Lageparameter dargestellt (Abbildung 6.8). Der von Weicker verwendete Lageparameter der nicht-zentralen Normalverteilung ist dabei identisch mit dem Erwartungswert der Verteilung.

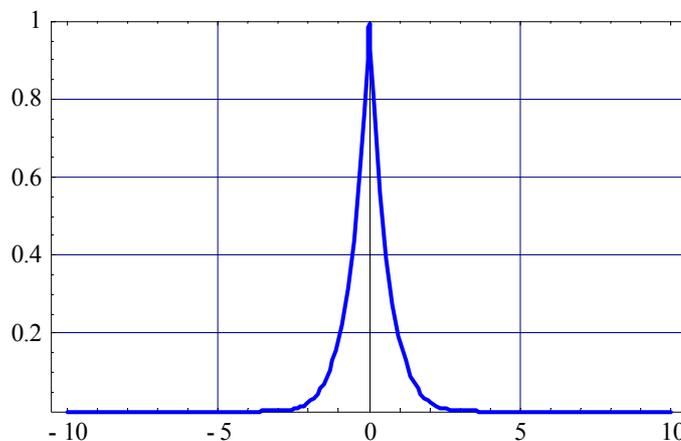


Abb. 6.8: Reversibilität der nicht-zentralen Normalverteilung mit Lageparameter a

Es ist deutlich zu erkennen, dass bereits eine moderate Richtungsadaptation eine geringe Reversibilität bedingt. Die Bedeutung dieser Aussage wird nach der Berechnung der Reversibilität der Ξ -Verteilungen diskutiert.

Um eine Vergleichbarkeit der Reversibilität zwischen Normalverteilung und Ξ -Verteilung zu erreichen, muss für die Ξ -Verteilung in dem ersten Schritt das zu einem Erwartungswert a gehörende Schiefemaß c bestimmt werden. Die Parameter σ und γ werden hier als fixierte Größen betrachtet. Nach (6.30) existiert für einen beliebigen, positiven Erwartungswert a das zugehörige Schiefemaß c als

$$c = 4^{-1/\gamma}(2 + a\sqrt{2\pi}\sigma^{-\gamma/2})^{2/\gamma} - 1 \quad (6.39)$$

Analog zu (6.36) müssen nun die Einzelintegrale bestimmt werden. Es gilt:

$$\int_{-\infty}^0 \varphi(x) dx = \frac{1}{1 + (1 + c)^{\gamma/2}} \quad (6.40)$$

$$\int_0^{\infty} \varphi(x) dx = \frac{(1 + c)^{\gamma/2}}{1 + (1 + c)^{\gamma/2}} \quad (6.41)$$

Das Ersetzen von c durch den Erwartungswert a führt zu den Gleichungen

$$\int_{-\infty}^0 \varphi(x) dx = \frac{1}{1 + (4^{-1/\gamma}(2 + a\sqrt{2\pi}\sigma^{-\gamma/2}))} \quad (6.42)$$

$$\begin{aligned} \int_0^{\infty} \varphi(x) dx &= \frac{4^{-1/\gamma}(2 + a\sqrt{2\pi}\sigma^{-\gamma/2})}{1 + (4^{-1/\gamma}(2 + a\sqrt{2\pi}\sigma^{-\gamma/2}))} \\ &= 1 - \frac{1}{1 + (4^{-1/\gamma}(2 + a\sqrt{2\pi}\sigma^{-\gamma/2}))} \end{aligned} \quad (6.43)$$

Für die Reversibilität ρ gilt damit:

$$\begin{aligned} \rho &= \frac{\int_{-\infty}^0 \varphi(x) dx}{\int_0^{\infty} \varphi(x) dx} \\ &= \frac{1}{1 + (4^{-1/\gamma}(2 + a\sqrt{2\pi}\sigma^{-\gamma/2}))} \\ &= \frac{1}{1 - \frac{1}{1 + (4^{-1/\gamma}(2 + a\sqrt{2\pi}\sigma^{-\gamma/2}))}} \\ &= \frac{1}{1 + a\sqrt{2\pi}\sigma^{-\gamma/2}} \end{aligned} \quad (6.44)$$

Die Abbildung 6.9 zeigt den Verlauf der Reversibilität für konstantes $\sigma = 1$ und $\gamma = 2$. Abgetragen ist die Reversibilität ρ in Abhängigkeit des Erwartungswertes der zugehörigen Ξ -Verteilung.

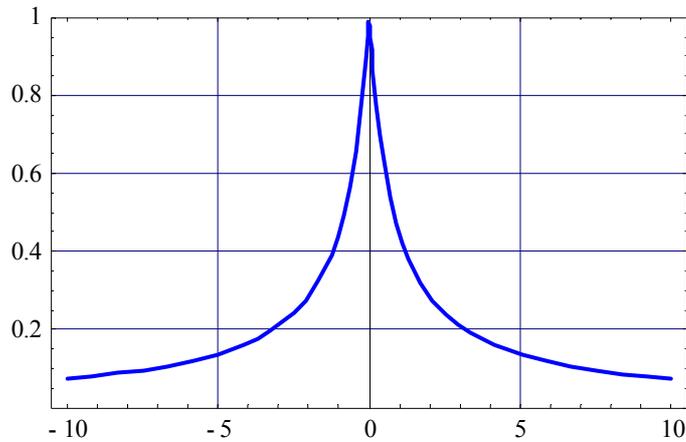


Abb. 6.9: Reversibilität der Ξ -Verteilung mit in Abhängigkeit des Erwartungswertes

Sollte sich eine adaptierte Richtung im Laufe der Optimierung als ungeeignet erweisen, müsste eine schnelle Neadaptation durch eine geeignete Wahl und Parametrisierung der Operatoren möglich sein. Analog zu dem Evolutionsfenster, das ein Intervall für die günstige Wahl der Schrittweite definiert, kann ein Reversibilitätsfenster definiert werden. Ausgangspunkt für die Definition dieses Fensters ist das Verhältnis der Anzahl selektierter Nachkommen zu der Anzahl der nicht selektierten Nachkommen $\mu/(\lambda - \mu)$. Richtungsadaptationen, bei denen das Verhältnis p_n/p_p kleiner als $\mu/(\lambda - \mu)$ ist, gelten hierbei als ungünstig, da die Selektion mehr geeignete Individuen selektieren muss, als durch die richtungs-adaptierte Mutation erzeugt werden. Die Abbildung 6.10 stellt das Reversibilitätsfenster für 15 Eltern und 100 Nachkommen für die Adaptation einer nicht-zentralen Normalverteilung sowie für die Verwendung einer Ξ -Verteilung (mit Schrittweite $\sigma = 1$ und globalem Skalierungsparameter $\gamma = 2$) dar. Bei der Adaptation einer nicht-zentralen Normalverteilung wird bereits für eine Richtungsadaptation, die sich als Erwartungswert 1.03 auszeichnet das Reversibilitätsfenster verlassen. Bei der Verwendung einer Ξ -Verteilung befindet sich die Mutation bis zu einem Erwartungswert von 3.73 innerhalb des Reversibilitätsfensters. Dieser Erwartungswert entspricht einem Schiefemaß von $c = 31.28$.

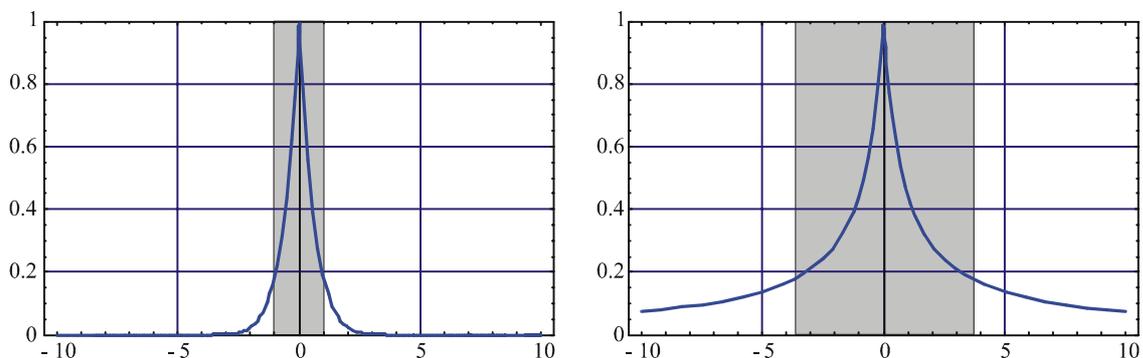


Abb. 6.10: Reversibilitätsfenster bei 15 Eltern und 100 Nachkommen

Die folgenden Abbildungen stellen die Grenze des Reversibilitätsfenster in Abhängigkeit des Verhältnis von selektierten zu nicht selektierten Nachkommen dar.

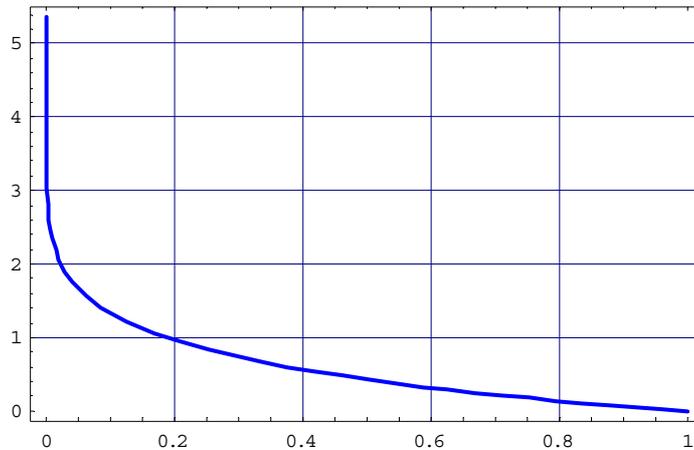


Abb. 6.11: Grenze des Reversibilitätsfenster für die Adaptation einer nicht-zentralen Normalverteilung

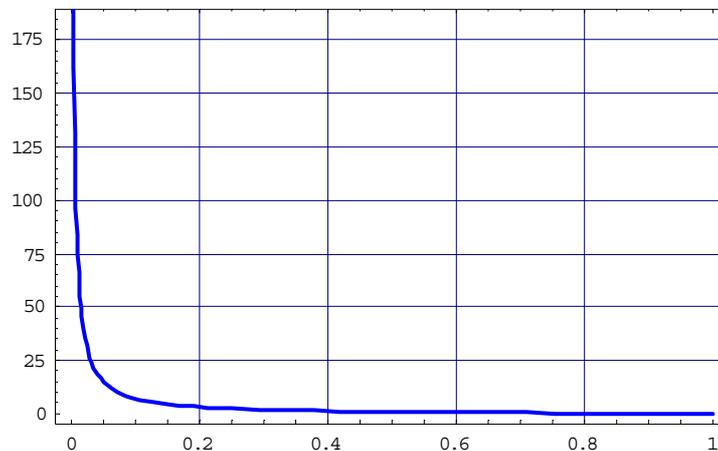


Abb. 6.12: Grenze des Reversibilitätsfenster für die Ξ -Verteilung

Die höhere Reversibilität der Ξ -Verteilung und damit die größere Fähigkeit zur Umorientierung ist deutlich erkennbar.

6.3 Empfehlungen zur Parametrisierung

Die asymmetrische Evolutionsstrategie wurde mit der Maßgabe entwickelt, dass sie über keine externen Parameter verfügen soll. Der globale Steuerparameter γ scheint diese Maßgabe zu verletzen. Daher wird an dieser Stelle die Bedeutung des Parameters γ beschrieben.

1. Steuerung der initialen Schrittweiten

Mit dem Parameter γ ist es möglich, bei konstanter Initialisierung der Schrittweiten und Schiefemaße die initiale Gestalt des Mutationsraumes zu beeinflussen. Die Abbildung 6.13 gibt hierfür ein Beispiel. In dem linken Diagramm wurde $\gamma = 1$, in dem rechten Diagramm

$\gamma = 4$ gewählt.

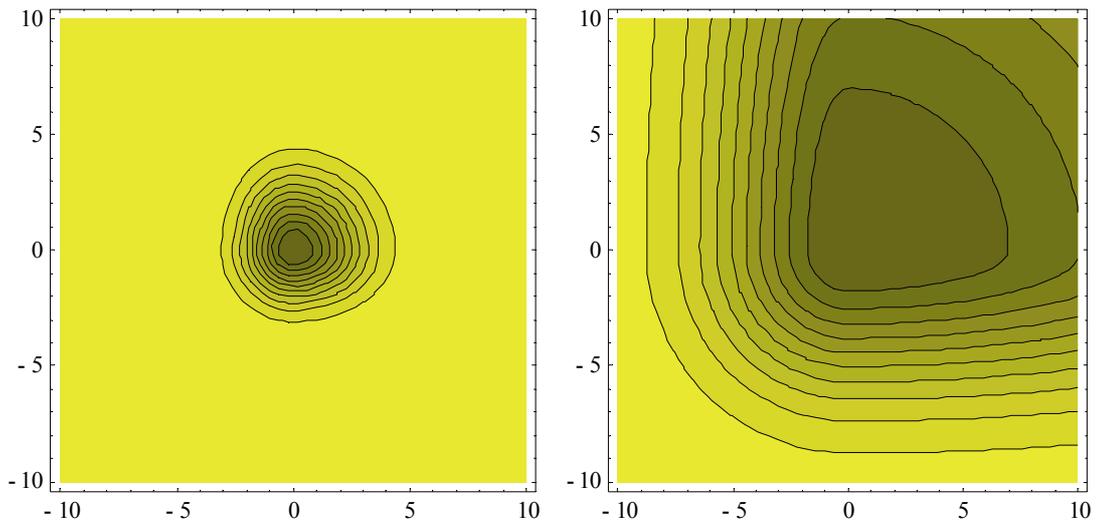


Abb. 6.13: Mutationsraum für unterschiedliche Belegungen des globalen Parameters γ

Durch die Selbstadaptation werden im Laufe der Optimierung die Schrittweiten und Schiefe-
maße adaptiert, sodass der Einfluss des Parameters γ verschwinden wird. Trotzdem kann in
der initialen Phase der Optimierung ohne Anpassung der Initialisierung zwischen unter-
schiedlichen Arbeitscharakteristika des Mutationsoperators gewählt werden.

2. Äquivalenz zur Normalverteilung

Der Steuerparameter γ wird benötigt, um die Normalverteilung als Spezialfall der $\Xi_{c,\sigma}^\gamma$ -Ver-
teilung betrachten zu können. Es gilt:

$$N(0, \sigma) = \Xi_{0,\sigma}^2 \quad (6.45)$$

3. Robustheit für $\gamma = 2$

Für alle experimentellen Versuche mit der asymmetrischen Evolutionsstrategie wurde der
Parameter $\gamma = 2$ gesetzt. Die asymmetrische Evolutionsstrategie hat bei keinem der unter-
suchten Problemen versagt.

4. Reversibilität

Mit steigendem Parameter γ steigt die Reversibilität der Mutation. Daraus folgt, dass die
Eigenschaft der Neoadaptation einer Koordinatenrichtung ebenfalls steigt. Gleichzeitig
nimmt die Ausprägung der Asymmetrie, d. h. die Fähigkeit eine Koordinatenrichtung zu
adaptieren ab. Auch hier stellt die Belegung $\gamma = 2$ einen, experimentell ermittelten, guten
Kompromiss dar.

Im Vorgriff auf die experimentellen Resultate, die im folgenden Kapitel dargestellt werden, sei an
dieser Stelle bereits auf die Tatsache hingewiesen, dass die Selbstadaptation so gut funktioniert, dass
auch bei einer absichtlichen Fehlinitialisierung die asymmetrische Evolutionsstrategie für die in die-
ser Arbeit untersuchten Testfunktionen immer konvergierte.

6.4 Zusammenfassung

Eine asymmetrische Evolutionsstrategie unterscheidet sich in ihrem generellen Aufbau nur unwesentlich von einer (μ, λ) -Evolutionsstrategie mit Adaptation von n Schrittweiten. Aus einer solchen Variante kann durch Ergänzung einer weiteren Strategiekomponente zur Verarbeitung der Schiefemaße und Austausch des Mutationsoperators eine asymmetrische Evolutionsstrategie entstehen. Diese Nähe der beiden Varianten zeigt sich auch in der Tatsache, dass jede asymmetrische Evolutionsstrategie in der Lage ist, eine (μ, λ) -Evolutionsstrategie mit Adaptation von n Schrittweiten zu simulieren. Es sollte daher ohne großen Aufwand möglich sein, bestehende Programmsysteme durch den Einsatz einer asymmetrischen Evolutionsstrategie aufzuwerten. Für das Anwendungsbeispiel Schraubenrotoren, das in Abschnitt 7.6 (Optimierung von Schraubenrotorprofilen, Seite 229) beschrieben ist, wurde eine solche Erweiterung auf asymmetrische Evolutionsstrategien bereits durchgeführt. Der Aufwand zur Anpassung des bestehenden Programmsystems betrug in diesem Fall weniger als fünf Stunden.

Trotz der Nähe der (μ, λ) -Evolutionsstrategie mit Adaptation von n Schrittweiten und der asymmetrischen Evolutionsstrategie zeigt die asymmetrische Evolutionsstrategie beeindruckende Vorteile, die in dem folgenden Kapitel am Beispiel von Testfunktionen und realen Anwendungsproblemen gezeigt werden.

Kapitel 7 Experimentelle Resultate

*Ob etwas Leben werden kann, das hängt nicht von
den grossen Ideen ab, sondern davon, ob man sich
aus ihnen ein Handwerk schafft, ein Tägliches,
etwas, was bei einem aushält bis ans Ende.
(Rainer Maria Rilke)*

Anhand der Struktur und der verwendeten Methoden bzw. Operatoren, lässt sich in der Regel nicht auf die Leistungsfähigkeit eines Optimieralgorithmus zurückschließen. Die Leistungsfähigkeit und die Einsatzmöglichkeiten eines solchen Algorithmus kann zum einen durch sein Verhalten bei realen Anwendungen, zum anderen aber auch durch den Einsatz von allgemein anerkannten Problemen bewertet werden. Soll ein Optimieralgorithmus zur Lösung eines realen Problems herangezogen werden, wird eine Vorauswahl getroffen, bei der bestimmte Algorithmustypen im Vorfeld bevorzugt oder abgelehnt werden. Daher ist in dem Bereich der realen Problemstellungen eine vergleichende Analyse, in der das Verhalten unterschiedlicher Optimieralgorithmen untersucht wird, nur unvollständig möglich. Um dennoch die Eigenschaften von Optimieralgorithmen vergleichen zu können, wird häufig auf sogenannte Benchmark-Probleme zurückgegriffen. In diesem Kapitel werden die Eigenschaften einer asymmetrischen Evolutionsstrategie sowohl anhand von Benchmark-Problemen als auch anhand von realen Problemstellungen aufgezeigt. Die Auswahl der Benchmark-Probleme orientiert sich an den Standardproblemen, die häufig zur Analyse von evolutionären Algorithmen eingesetzt werden. Die realen Probleme stammen zum einen aus der Konstruktion von metall-keramischen Verbundwerkstoffen und zum anderen aus der Qualitätssicherung von Punktschweißverbindungen.

7.1 Varianten des Kugelmodells

Für alle Experimente in diesem Abschnitt gelten folgende Rahmenbedingungen:

- Initialisierung der Startpopulation gleichverteilt nach $G[-50, 50]$,
- Initialisierung der Schrittweite(n) mit 1.0,
- Initialisierung der Lagewinkel mit 0.0,
- Initialisierung der Schiefemaße gleichverteilt nach $G[-0.5, 0.5]$,
- Abbruchkriterium $f(\mathbf{x}) < 10^{-6}$ oder 1000 bzw. 10000 Generationen,

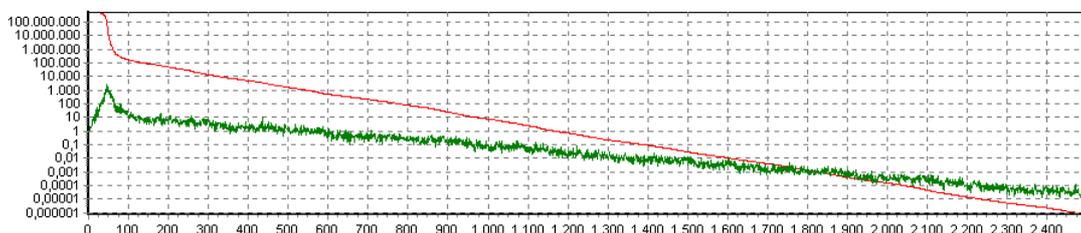
- Anzahl der Elterindividuen $\mu = 15$, Anzahl der Nachkommen $\lambda = 100$,
- korrelierte Mutation mit der maximalen Anzahl an Lagewinkeln.

Zusätzlich zu diesen Randbedingungen, die sich an Bedingungen orientieren, die in dieser Form auch für Testreihen aus der Literatur gelten, wird eine spezielle Initialisierung der Startpopulation untersucht. Bei dieser speziellen Initialisierung der Startpopulation werden die Individuen gezielt in einem Bereich des Suchraums erzeugt, der vom Optimum entfernt ist. Diese Fehlinitialisierung soll Bedingungen erzeugen, die sich an realistischen Problemen orientieren, bei denen das Optimum nicht bekannt ist und somit eine ideale Initialisierung um das Optimum nicht möglich ist. Für die Kugelmodellvarianten dieses Abschnittes wird die Startpopulation bei der Fehlinitialisierung gleichverteilt nach $G[9050, 1050]$ erzeugt.

Die Resultate der Experimente sind in Form eines Balkendiagramms aufgetragen. Die x_1 -Achse stellt die unterschiedlichen Varianten der Selbstadaptation dar. Die Adaptation einer Schrittweite wird mit einfach bezeichnet und blau dargestellt. Die Adaptation von n Schrittweiten wird mit linear bezeichnet und grün dargestellt. Die korrelierte Adaptation aller Schrittweiten und Lagewinkel wird mit korreliert bezeichnet und rot dargestellt. Die asymmetrische Adaptation von n Schrittweiten und n Schiefemaßen wird als asymmetrisch bezeichnet und in gelb dargestellt. Auf der x_2 -Achse sind die Generationen abgetragen, beginnend bei 0. Der letzte Balken der x_2 -Achse steht für das Erreichen der maximalen Generationenzahl. Hier werden alle Testläufe zusammengefasst, die nach der maximalen Anzahl an Generationen nicht das Abbruchkriterium Q_{stop} erreicht haben. Auf der z -Achse sind in Prozent die Anzahl der Testläufe angegeben, bei denen das Endkriterium Q_{stop} vor Erreichen der maximalen Generationenzahl erreicht wurde.

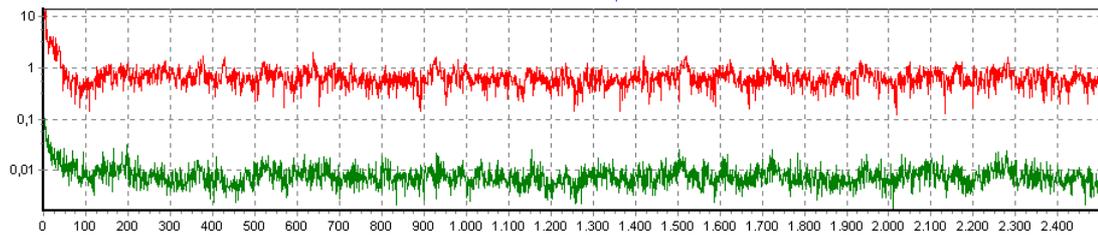
Bei einigen Experimenten wird sehr häufig die maximale Anzahl an Generationen erreicht und das Experiment abgebrochen. Dies kann bedeuten, dass die maximale Generationenzahl zu gering gewählt wurde und das Optimum mit vergrößerter Generationenzahl gefunden worden wäre. Es kann aber auch bedeuten, dass die Evolutionsstrategie stagniert, ohne das Optimum gefunden zu haben oder dass die Evolutionsstrategie divergiert. Eine genaue Charakterisierung ist nicht möglich, da es bei stochastischen Optimierverfahren möglich ist, dass in der Generation nach Erreichen des Abbruchkriteriums das Optimum gefunden worden wäre. Trotzdem lassen sich grobe Aussagen über das Konvergenzverhalten machen. Es wird in dieser Arbeit zwischen drei Fällen unterschieden. Im folgenden werden für diese drei Fälle typische Verläufe der Fitness angegeben.

1. Folgender Verlauf des Fitnesswertes (rot) wird bei der Diskussion der Resultate als **langsam konvergent** bezeichnet. Bei vergrößerter maximaler Generationenzahl ist ein Erreichen des Optimums wahrscheinlich.

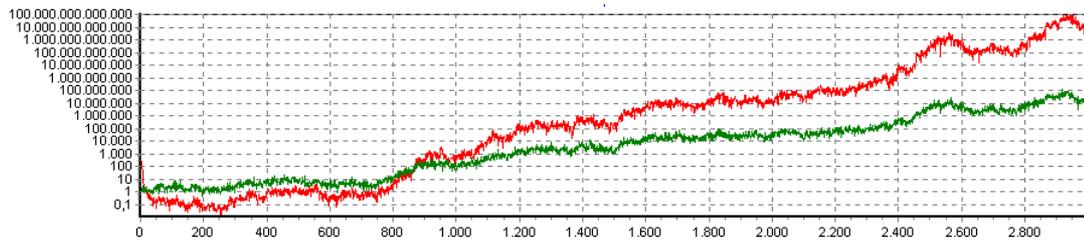


2. Folgender Verlauf wird als **stagnierend** bezeichnet. Ein Erreichen des Optimums ist nicht ausgeschlossen, jedoch unwahrscheinlich. Ein stagnierender Verlauf ist in dieser Arbeit zusätzlich abgesichert, indem isolierte Experimente durchgeführt wurden, bei denen die

maximale Generationenzahl um den Faktor 10 vergrößert wurde.



3. Folgender Verlauf wird als **divergent** bezeichnet. Ein Erreichen des Optimums kann aber auch hier nicht ausgeschlossen werden.



7.1-1 f_1 - Kugelmodell

Das Kugelmodell ist eine n -dimensionale quadratische Funktion, die kontinuierlich, streng konvex und unimodal ist. Sie ist definiert durch:

$$f_1(\mathbf{x}) = \sum_{i=1}^n x_i^2, \text{ mit } \mathbf{x} \in \mathbb{R}^n \tag{7.1}$$

Das Optimum von f_1 befindet sich an der Stelle $\mathbf{x}^* = (0, \dots, 0)^T$, mit $f(\mathbf{x}^*) = 0$. Die Abbildung 7.1 zeigt den Verlauf von f_1 , für $n = 2$.

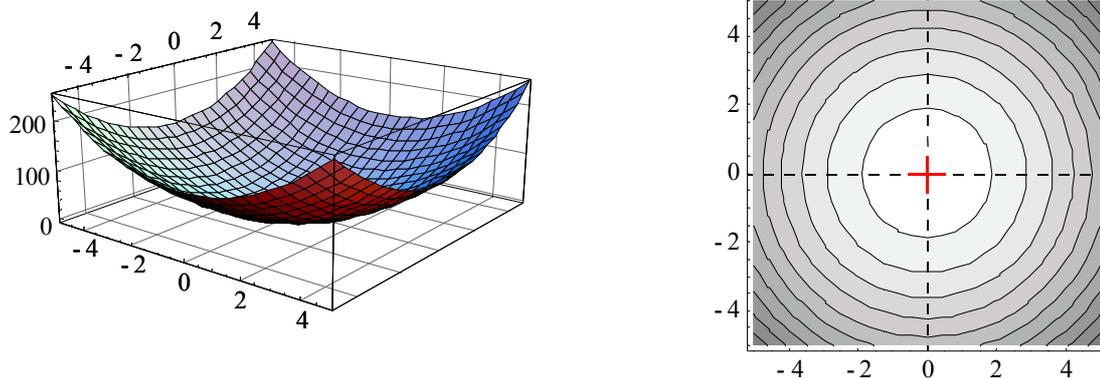


Abb. 7.1: Verlauf des Kugelmodells f_1 für $n=2$ Dimensionen

Die folgende Abbildung 7.2 zeigt die Ergebnisse der Testläufe. Die maximale Generationenzahl betrug in dieser Testreihe 200, zu jeder Adaptationsvariante wurden 100 Testläufe durchgeführt.

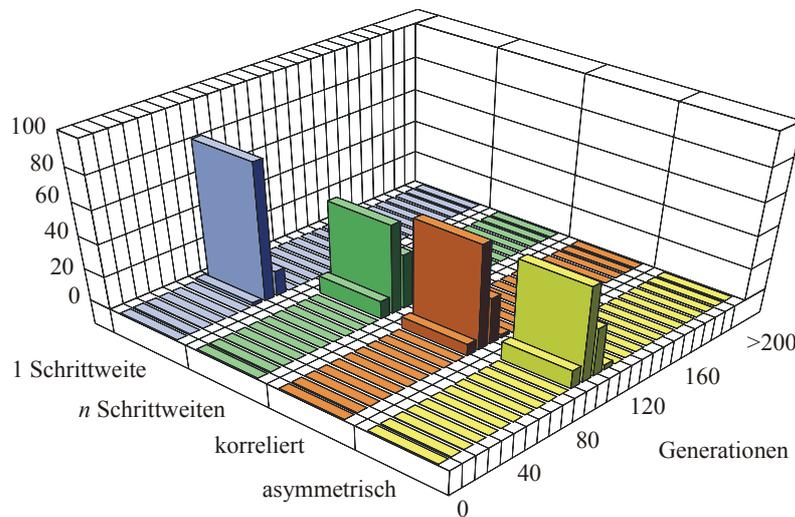


Abb. 7.2: Experimentelle Resultate zur Testfunktion f_1

In der Literatur werden häufig zwei weitere Belegungen von Startvektor, Strategieparametern und Abbruchkriterium gewählt. Hoffmeister und Bäck verweisen in [Hof92] auf die Wahl von DeJong und Schwefel, mit

$$-5.12 \leq x_i \leq 5.12 \quad (7.2)$$

Hansen schränkt den Definitionsbereich von f_1 nicht ein, sondern gibt Vorgaben für Startvektor, Schrittweite und Abbruchkriterium [Han98]:

$$\mathbf{x}^{(0)} = (1, \dots, 1)^T \quad (7.3)$$

$$\boldsymbol{\sigma}^{(0)} = (1, \dots, 1)^T \quad (7.4)$$

$$Q_{stop} = 10^{-10} \quad (7.5)$$

Die experimentellen Resultate zu diesen beiden Testmodellen finden sich im Abschnitt 7.2-2 (B2 - Kugelmodell, Seite 198) bzw. in [Hof92]. Für den Fall der Fehlinitialisierung lauten die Resultate des Experiments wie in Abbildung 7.3 dargestellt.

Die maximale Generationenzahl wurde hier auf 500 erhöht. Da das Kugelmodell eine einfache innere Struktur aufweist, ist die Adaptation einer Schrittweite ausreichend. Dies erklärt das gute Konvergenzverhalten der Evolutionsstrategie mit Adaptation einer Schrittweite. Die anderen Varianten verhalten sich untereinander ähnlich, bei leicht schlechterem Konvergenzverhalten. Im Fall der Fehlinitialisierung zeigt die asymmetrische Evolutionsstrategie bereits das beste Verhalten der komplexen Adaptationsvarianten.

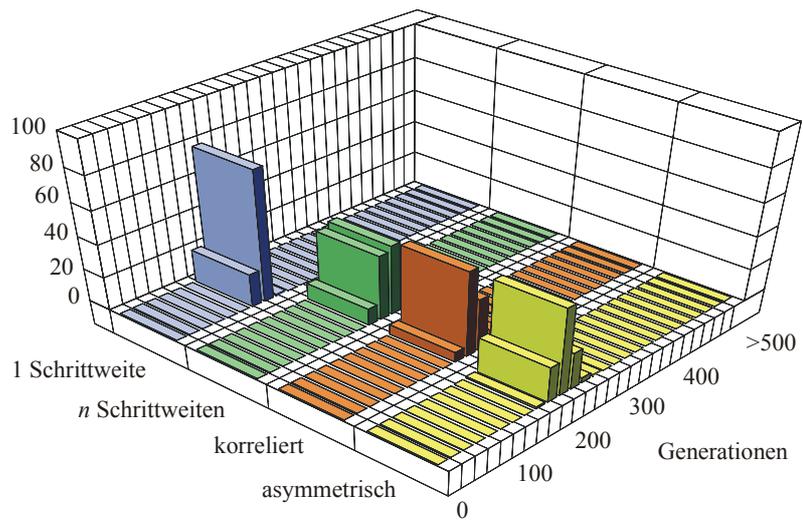


Abb. 7.3: Experimentelle Resultate zur Testfunktion f_1 mit Fehlinitialisierung

7.1-2 f_2 - alternierendes Kugelmodell

Das alternierende Kugelmodell entspricht dem Kugelmodell f_1 mit der Änderung, dass die geraden Koordinaten mit dem Faktor 0.01 skaliert werden. Damit gilt:

$$f_2(x) = \sum_{i=1}^n a_i x_i^2, \text{ mit } \mathbf{x} \in \mathbb{R}^n \text{ und } \mathbf{a} = (1, 0.01, \dots, 1, 0.01)^T \quad (7.6)$$

Das Optimum von f_2 befindet sich an der Stelle $\mathbf{x}^* = (0, \dots, 0)^T$, mit $f_2(\mathbf{x}^*) = 0$. Die Abbildung 7.4 zeigt den Verlauf von f_2 . Die Summe der ungeraden Koordinaten wurde in dieser Abbildung auf die x_1 -Achse projiziert, die Summe der geraden Koordinaten auf die x_2 -Achse.

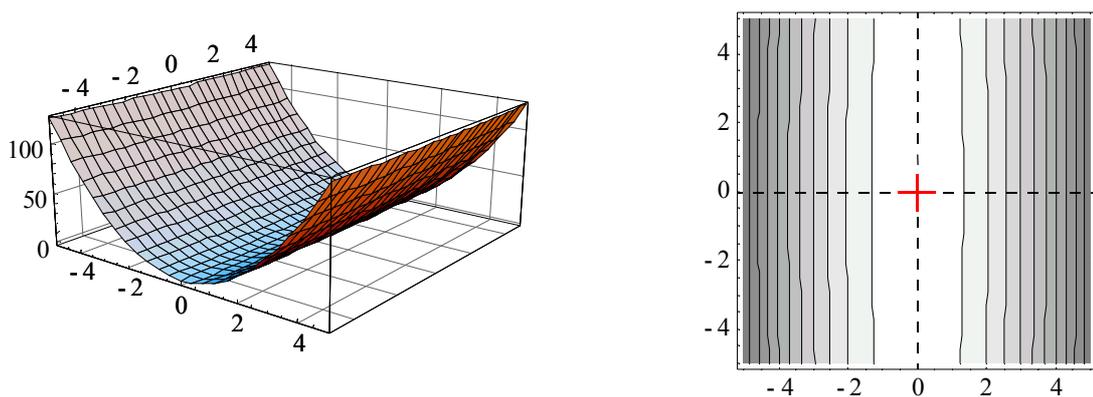


Abb. 7.4: Verlauf des Kugelmodells f_2 mit projizierten Koordinaten

Die folgende Abbildung 7.5 zeigt die Ergebnisse der Testläufe. Die maximale Generationenzahl betrug in dieser Testreihe 1000, zu jeder Adaptationsvariante wurden 100 Testläufe durchgeführt.

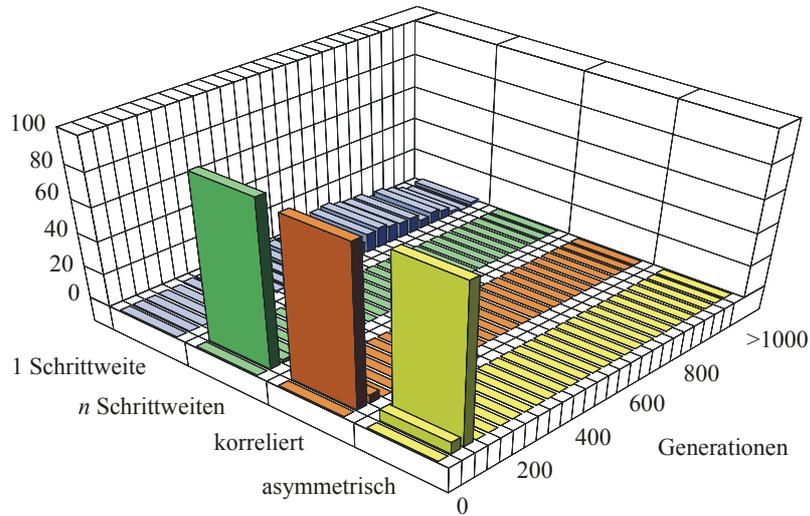


Abb. 7.5: Experimentelle Resultate zur Testfunktion f_2

Für den Fall, dass die Objektkomponente fehlinitialisiert wurde gelten folgende Resultate:

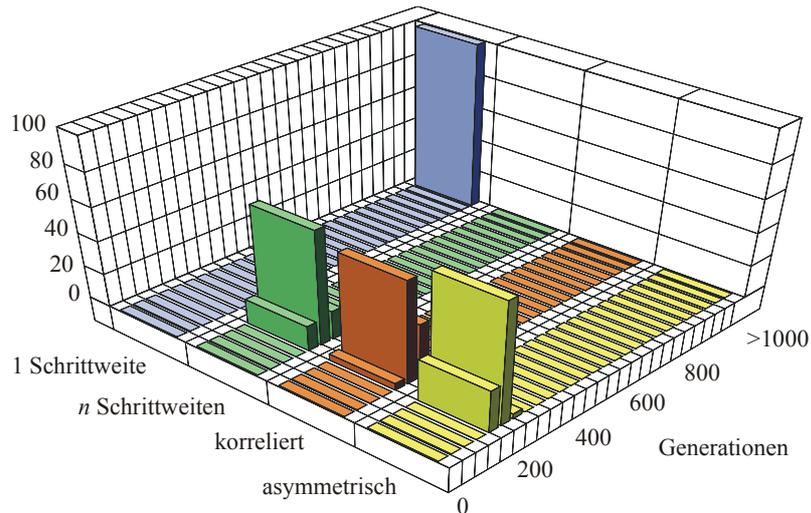


Abb. 7.6: Experimentelle Resultate zur Testfunktion f_2 mit Fehlinitialisierung

Die Evolutionsstrategie mit Adaptation einer Schrittweite ist langsam konvergent. Es werden jedoch mehr als 3000 Generationen benötigt, um das Optimum zu erreichen.

7.1-3 f_3 - steigendes Kugelmodell

Das steigende Kugelmodell entspricht dem Kugelmodell f_1 mit der Änderung, dass die Koordinaten mit dem Faktor 10^{i-1} skaliert werden. Damit gilt:

$$f_3(\mathbf{x}) = \sum_{i=1}^n 10^{i-1} x_i^2, \text{ mit } \mathbf{x} \in \mathbb{R}^n \quad (7.7)$$

Das Optimum von f_3 befindet sich an der Stelle $\mathbf{x}^* = (0, \dots, 0)^T$, mit $f_3(\mathbf{x}^*) = 0$. Die Abbildung 7.7 zeigt den Verlauf von f_3 . Die Summe der ungeraden Koordinaten wurde in dieser Bildung auf die x_1 -Achse projiziert, die Summe der geraden Koordinaten auf die x_2 -Achse.

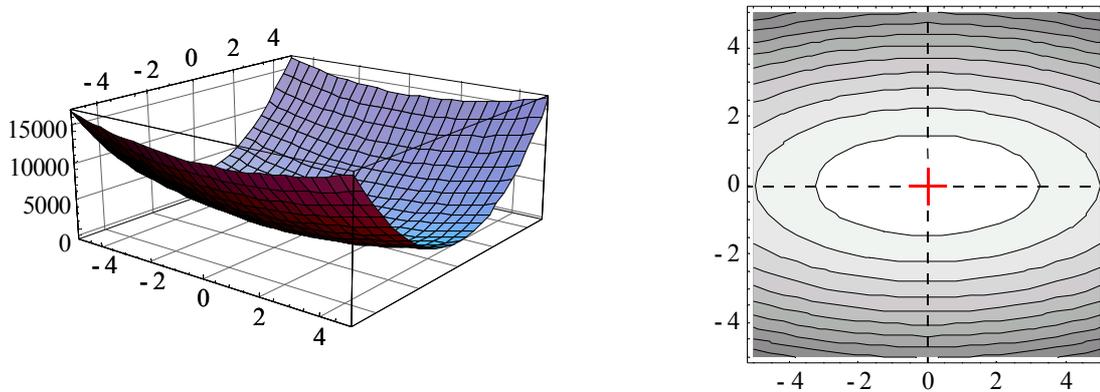


Abb. 7.7: Verlauf des Kugelmodells f_3 mit projizierten Koordinaten

Die folgende Abbildung 7.8 zeigt die Ergebnisse der Testläufe. Die maximale Generationenzahl betrug in dieser Testreihe 1000, zu jeder Adaptationsvariante wurden 100 Testläufe durchgeführt.

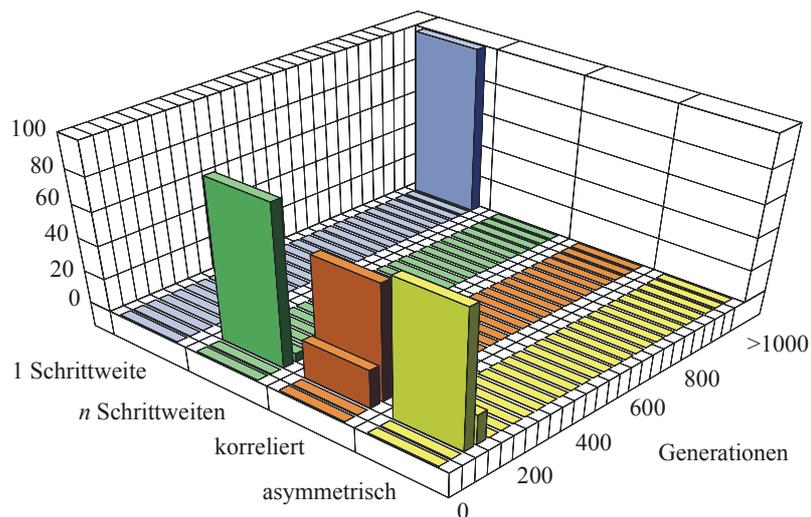


Abb. 7.8: Experimentelle Resultate zur Testfunktion f_3

Die Evolutionsstrategie mit Adaptation einer Schrittweite ist langsam konvergent. Es werden jedoch mehr als 10^5 Generationen benötigt, um das Optimum zu erreichen.

Die Resultate der Experimente bei Fehlinitialisierung sind in der folgenden Abbildung 7.9 dargestellt. Auch hier ist die Variante mit Adaptation einer Schrittweite langsam konvergent. Es werden jedoch mehr als 10^8 Generationen benötigt, um das Optimum zu erreichen.

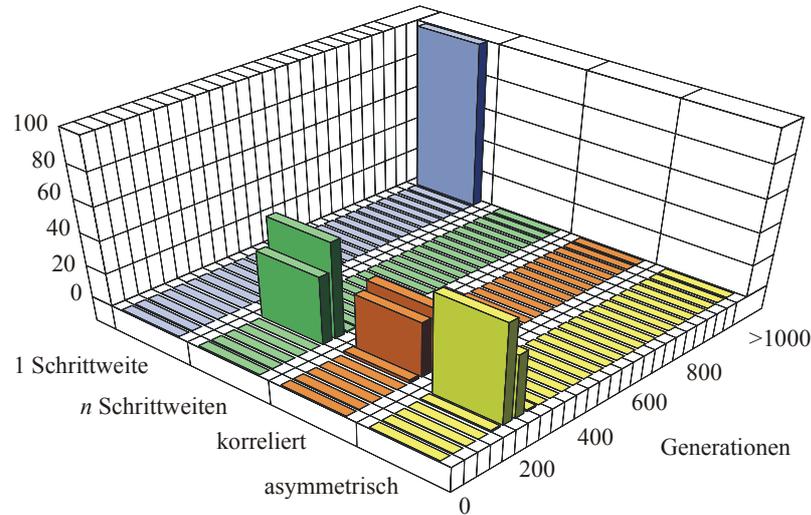


Abb. 7.9: Experimentelle Resultate zur Testfunktion f_3 mit Fehlinitialisierung

7.1-4 f_4 - Doppelsumme

Das f_4 -Doppelsummenmodell bildet die Grundform für die folgenden Varianten der Doppelsumme f_5 und f_6 . In der Literatur wird die Doppelsumme auch als Schwefel's Problem bezeichnet. Die Grundfunktion f_4 lautet:

$$f_4(\mathbf{x}) = \sum_{i=1}^n \left(\sum_{j=1}^i x_j \right)^2, \text{ mit } \mathbf{x} \in \mathbb{R}^n \quad (7.8)$$

Das Optimum von f_4 befindet sich an der Stelle $\mathbf{x}^* = (0, \dots, 0)^T$, mit $f_4(\mathbf{x}^*) = 0$. Bedingt durch die Doppelsumme werden jedoch die niederwertigen Koordinaten betont, sodass bei der Doppelsumme f_4 eine lineare Skalierung der Koordinaten erreicht wird. Die Abbildung 7.10 zeigt den Verlauf von f_4 . Die Summe der ungeraden Koordinaten wurde in dieser Abbildung auf die x_1 -Achse projiziert, die Summe der geraden Koordinaten auf die x_2 -Achse.

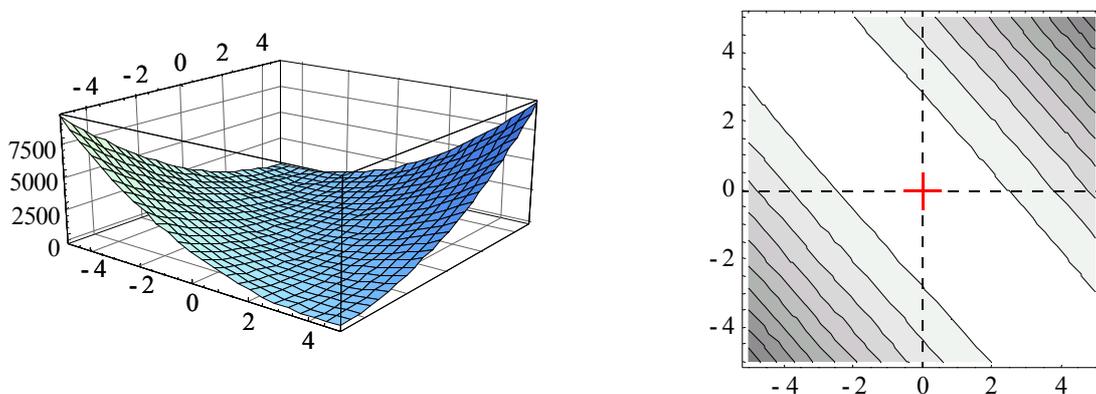


Abb. 7.10: Verlauf des Kugelmodells f_4 mit projizierten Koordinaten

Die folgende Abbildung 7.11 zeigt die Ergebnisse der Testläufe. Die maximale Generationenzahl betrug in dieser Testreihe 500, zu jeder Adaptationsvariante wurden 100 Testläufe durchgeführt.

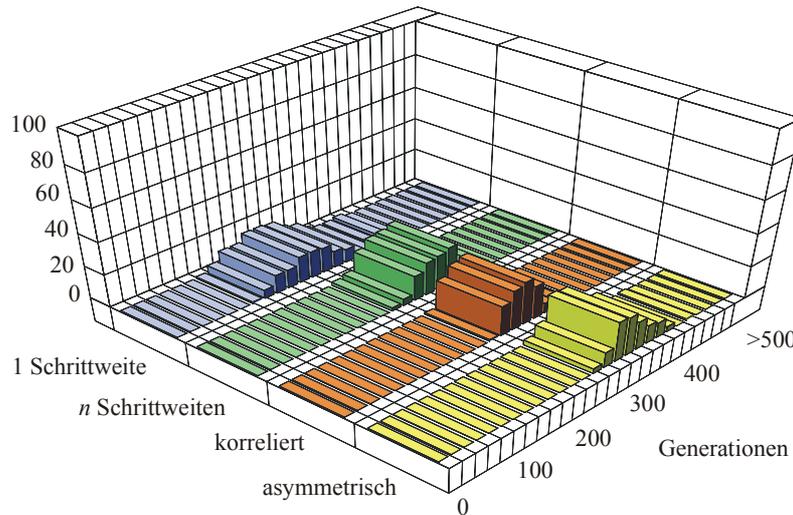


Abb. 7.11: Experimentelle Resultate zur Testfunktion f_4

Die Resultate der Experimente bei Fehlinitialisierung sind in der folgenden Abbildung 7.12 dargestellt.

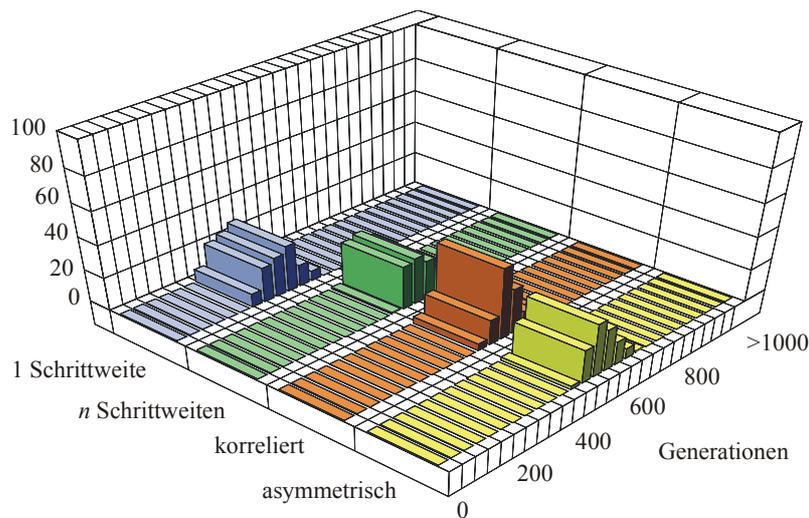


Abb. 7.12: Experimentelle Resultate zur Testfunktion f_4 mit Fehlinitialisierung

7.1-5 f_5 - alternierende Doppelsumme

Die alternierende Doppelsumme entspricht der Doppelsumme f_3 mit der Änderung, dass die geraden Koordinaten mit dem Faktor 0.01 skaliert werden. Damit gilt:

$$f_5(\mathbf{x}) = \sum_{i=1}^n \left(\sum_{j=1}^i a_j x_j \right)^2, \text{ mit } \mathbf{x} \in \mathbb{R}^n \text{ und } \mathbf{a} = (1, 0.01, \dots, 1, 0.01)^T \quad (7.9)$$

Das Optimum von f_5 befindet sich an der Stelle $\mathbf{x}^* = (0, \dots, 0)^T$, mit $f_5(\mathbf{x}^*) = 0$. Die Abbildung

7.13 zeigt den Verlauf von f_5 . Die Summe der ungeraden Koordinaten wurde in dieser Abbildung auf die x_1 -Achse projiziert, die Summe der geraden Koordinaten auf die x_2 -Achse.

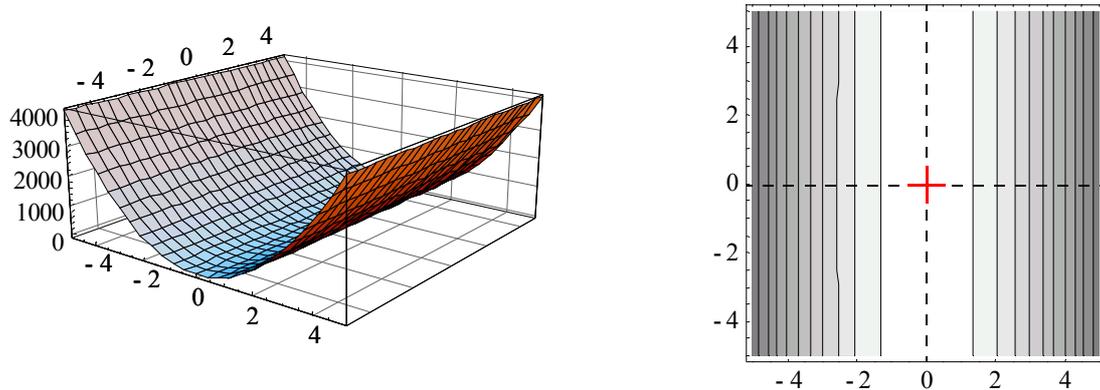


Abb. 7.13: Verlauf des Kugelmodells f_5 mit projizierten Koordinaten

Die folgende Abbildung 7.14 zeigt die Ergebnisse der Testläufe. Die maximale Generationenzahl betrug in dieser Testreihe 1000, zu jeder Adaptationsvariante wurden 100 Testläufe durchgeführt.

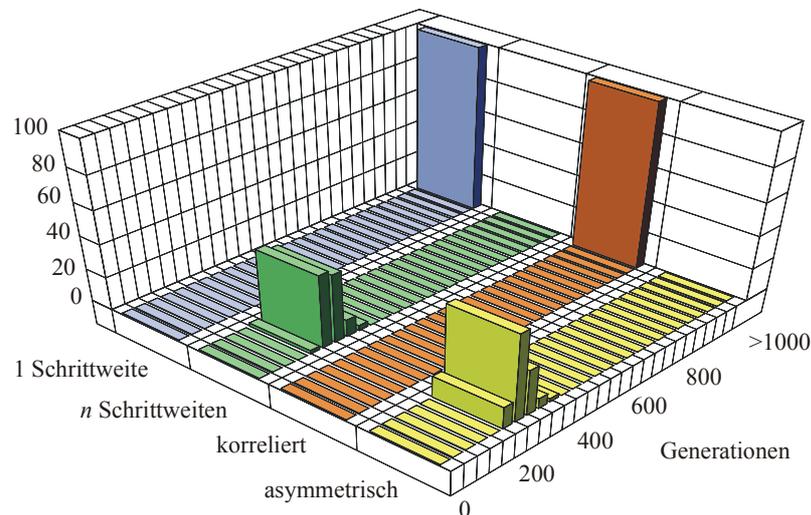


Abb. 7.14: Experimentelle Resultate zur Testfunktion f_5

Die Evolutionsstrategie mit Adaptation einer Schrittweite ist langsam konvergent. Es werden jedoch mehr als 10^6 Generationen benötigt, um das Optimum zu erreichen. Die Evolutionsstrategie mit korrelierender Adaptation, also der Adaptation von Schrittweiten und Lagewinkeln, ist divergent. Bereits nach 5000 Generationen werden Zielfunktionswerte $> 10^{100}$ erreicht.

Die Resultate der Experimente bei Fehlinitialisierung sind in der folgenden Abbildung 7.15 dargestellt. Auch hier ist die Variante mit Adaptation einer Schrittweite langsam konvergent, die Variante mit korrelierender Adaptation divergent.

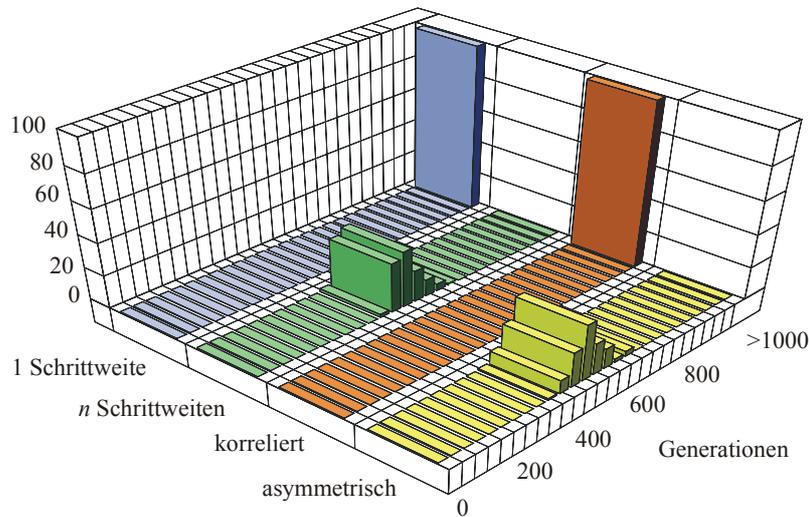


Abb. 7.15: Experimentelle Resultate zur Testfunktion f_5 mit Fehlinitialisierung

7.1-6 f_6 - steigende Doppelsumme

Die steigende Doppelsumme entspricht der Doppelsumme f_3 mit der Änderung, dass die Koordinaten mit dem Faktor 10^{j-1} skaliert werden. Damit gilt:

$$f_6(\mathbf{x}) = \sum_{i=1}^n \left(\sum_{j=1}^i 10^{j-1} x_j \right)^2, \text{ mit } \mathbf{x} \in \mathbb{R}^n \tag{7.10}$$

Das Optimum von f_6 befindet sich an der Stelle $\mathbf{x}^* = (0, \dots, 0)^T$, mit $f_6(\mathbf{x}^*) = 0$. Die Abbildung 7.16 zeigt den Verlauf von f_6 . Die Summe der ungeraden Koordinaten wurde in dieser Bildung auf die x_1 -Achse projiziert, die Summe der geraden Koordinaten auf die x_2 -Achse.

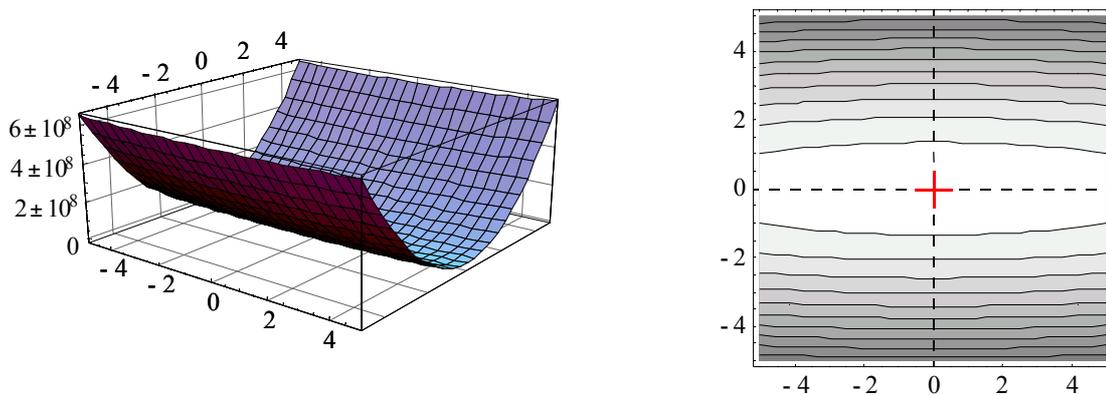


Abb. 7.16: Verlauf des Kugelmodells f_6 mit projizierten Koordinaten

Die folgende Abbildung 7.17 zeigt die Ergebnisse der Testläufe. Die maximale Generationenzahl betrug in dieser Testreihe 1000, zu jeder Adaptationsvariante wurden 100 Testläufe durchgeführt.

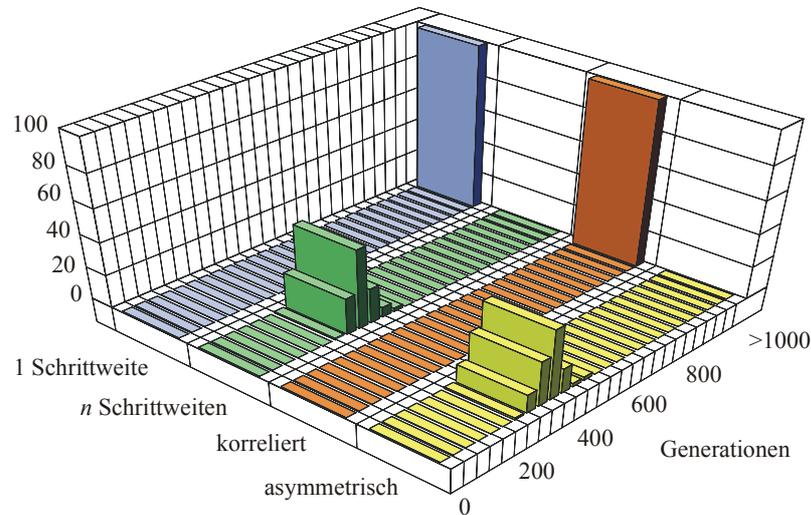


Abb. 7.17: Experimentelle Resultate zur Testfunktion f_6

Für den Fall der Fehlinitialisierung der Objektkomponenten sind die experimentellen Resultate in der Abbildung 7.18 dargestellt.

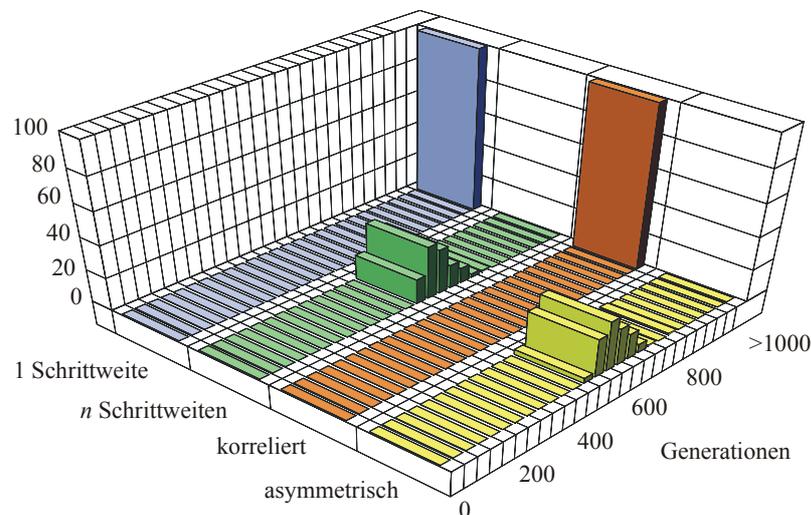


Abb. 7.18: Experimentelle Resultate zur Testfunktion f_6 mit Fehlinitialisierung

7.1-7 Bewertung der Resultate

Die Evolutionsstrategie mit Adaptation einer Schrittweite kann das Optimum der Testfunktion f_1 gut erreichen, da diese Funktion aufgrund ihrer Symmetrie zum Ursprung nur eine adaptierte Schrittweite benötigt. Auch das Auffinden des Optimums der Testfunktion f_4 kann von einer Evolutionsstrategie mit Adaptation einer Schrittweite gut durchgeführt werden, da bei der Dimensionalität 10 die Skalierung der einzelnen Koordinatenachsen moderat ausfällt. Für diese beiden Testfunktionen kann die Evolutionsstrategie mit Adaptation einer Schrittweite auch im Fall der Fehlinitialisie-

rung das Optimum mit der geringsten Zahl an Funktionsauswertungen finden. Für alle weiteren Testfunktionen dieses Abschnittes versagt die Evolutionsstrategie mit Adaptation einer Schrittweite. Die Evolutionsstrategie mit Adaptation von n Schrittweiten kann bei allen Testfunktionen das Optimum bestimmen. Die Fehlinitialisierung stellt für diese Variante keine Schwierigkeit dar. Die Evolutionsstrategie mit Adaptation von Schrittweiten und Lagewinkeln kann das Optimum der Testfunktionen f_1 bis f_4 zuverlässig finden. Bei den Testfunktionen f_5 und f_6 , sowohl bei den standard, als auch bei den fehlinitialisierten Testfällen, versagt sie. Bei diesen Fällen ist zudem starke Divergenz festzustellen. Die asymmetrische Evolutionsstrategie kann bei allen Testfunktionen das Optimum finden. In den Fällen mit Fehlinitialisierung benötigt sie in der Regel die geringste Anzahl an Funktionsauswertungen. Ausnahmen sind hier nur die Funktionen f_1 und f_4 , bei denen die Evolutionsstrategie mit Adaptation einer Schrittweite aus bereits genannten Gründen die besten Werte erreicht.

7.2 Problemklassen nach Hansen und Ostermeier

Die Problemklassen nach Ostermeier und Hansen stammen aus der Dissertation von Hansen. Sie sind als Testfunktionen für die verallgemeinerte, individuelle Schrittweitenregulierung verwendet worden [Han98]. Für diese Funktionen gilt, falls nicht anders beschrieben:

- Initialisierung der Startpopulation entsprechend der Vorgaben nach [Han98].
- Initialisierung der Schrittweite(n) mit 1.0 .
- Initialisierung der Lagewinkel mit 0.0 .
Die kumulative Schrittweitenregelung sowie die Kovarianzmatrix-Adaptation nach Ostermeier und Hansen verwenden keine Selbstadaptation der Lagewinkel, wie sie von Schwefel vorgeschlagen wird. Daher ist für diese Verfahren keine Initialisierung notwendig. Da die Resultate trotzdem mit der korrelierten Selbstadaptation verglichen werden, ist hier die Initialisierung der Lagewinkel ausschließlich für die Variante nach Schwefel von Bedeutung.
- korrelierte Mutation mit der maximalen Anzahl an Lagewinkeln.
- Initialisierung der Schiefemaße gleichverteilt nach $G[-0.5, 0.5]$.
- Abbruchkriterium $Q_{stop} = 10^{-10}$ oder Erreichen der maximalen Anzahl Generationen.
- Anzahl der Elterindividuen $\mu = 15$, Anzahl der Nachkommen $\lambda = 100$.

7.2-1 B_1 - Ebenenmodell

Die Testfunktion B_1 ist n -dimensional. Es wird jedoch nur eine Koordinate zur Berechnung des Funktionswertes verwendet. Es gilt:

$$B_1(\mathbf{x}) = -x_1, \text{ mit } \mathbf{x} \in \mathbb{R}^n \quad (7.11)$$

Das Optimum von B_1 befindet sich an den Stellen $\mathbf{x}^* = (\infty, x_2, \dots, x_n)^T$, mit $B_1(\mathbf{x}^*) = -\infty$, die Werte für x_2 bis x_n sind beliebig. Als Abbruchkriterium dient das Erreichen des Zielfunktionswertes -10^{10} . Die folgende Abbildung 7.19 zeigt die Ergebnisse der Testläufe. Die maximale Generationenzahl betrug in dieser Testreihe 100.

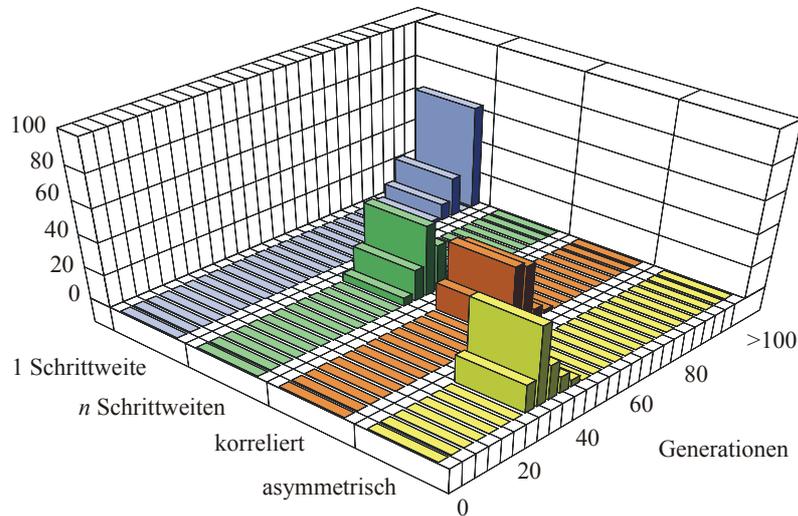


Abb. 7.19: Experimentelle Resultate zur Testfunktion B_1

Der hohe Anteil der Experimente zur Evolutionsstrategie mit Adaptation einer Schrittweite, bei der die maximale Anzahl an Generationen überschritten wurde, liegt in der langsamen Annäherung an das Optimum für diese Variante begründet. Bei allen Experimenten zu der Testfunktion B_1 wurde das Optimum nach max. 300 Generationen gefunden.

7.2-2 B_2 - Kugelmodell

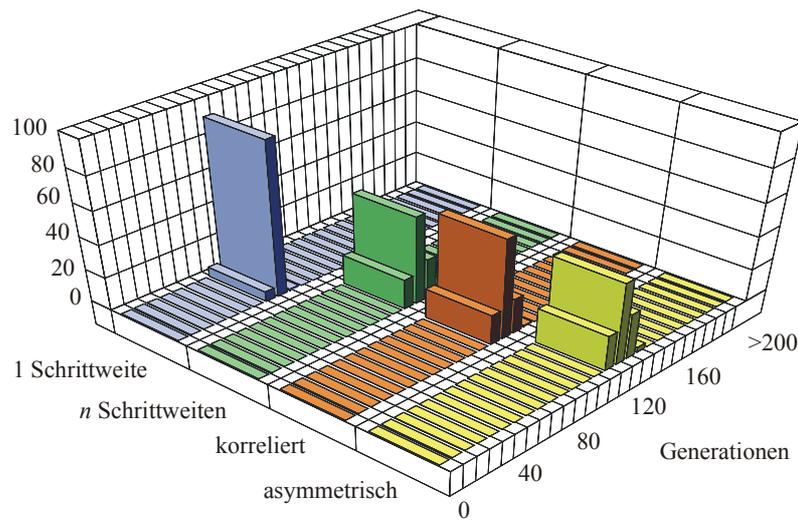
Das Kugelmodell B_2 ist identisch zu dem Kugelmodell f_1 :

$$f_1(\mathbf{x}) = \sum_{i=1}^n x_i^2, \text{ mit } \mathbf{x} \in \mathbb{R}^n \quad (7.12)$$

Unterschiede zum f_1 -Modell liegen in der Initialisierung der Startpopulation und der Schrittweiten sowie im Abbruchkriterium. Schrittweite und Abbruchkriterium folgen den allgemeinen Angaben für die Testfunktionen nach [Han98], für die Initialisierung der Startpopulation gilt:

$$\mathbf{x}^{(0)} = (1, \dots, 1)^T \quad (7.13)$$

Die folgende Abbildung 7.20 zeigt die Ergebnisse der Testläufe. Die maximale Generationenzahl betrug in dieser Testreihe 200, zu jeder Adaptationsvariante wurden 100 Testläufe durchgeführt.


 Abb. 7.20: Experimentelle Resultate zur Testfunktion B_2

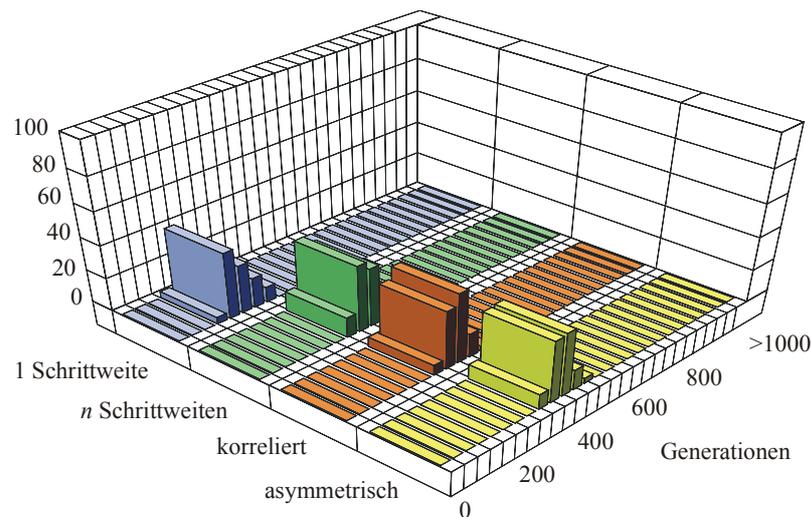
Die Resultate zur Testfunktion B_2 stimmen mit denen der Testfunktion f_1 überein. Leichte Abweichungen entstehen nur, da Abbruchkriterium und Initialisierung in den beiden Fällen unterschiedlich sind.

7.2-3 B_4 - Doppelsumme

Die Testfunktion B_4 ist identisch mit der Testfunktion f_4 :

$$B_4(\mathbf{x}) = \sum_{i=1}^n \left(\sum_{j=1}^i x_j \right)^2, \text{ mit } \mathbf{x} \in \mathbb{R}^n \quad (7.14)$$

Das Optimum von B_4 befindet sich an der Stelle $\mathbf{x}^* = (0, \dots, 0)^T$, mit $B_4(\mathbf{x}^*) = 0$. Die Unterschiede zwischen B_4 und f_4 liegen in der Initialisierung der Startpopulation und der Schrittweiten sowie im Abbruchkriterium. Diese Angaben entsprechen den allgemeinen Daten aus Abschnitt 7.2 (Problemklassen nach Hansen und Ostermeier, Seite 197). Die Abbildung 7.21 zeigt die Resultate der Testläufe für die (15, 100)-Evolutionstrategien mit den unterschiedlichen Adaptationsvarianten.

Abb. 7.21: Experimentelle Resultate zur Testfunktion B_4

Bei der Testfunktion B_4 ist das Verhalten nahezu identisch zur Testfunktion f_4 , da lediglich ein anderes Abbruchkriterium sowie eine andere Initialisierung vorgenommen wurde.

7.2-4 B_5 - Rosenbrock-Funktion

Die Definition von B_5 lautet:

$$B_5(\mathbf{x}) = \sum_{i=1}^{n-1} (100(x_i^2 - x_{i+1})^2 + (x_i - 1)^2), \text{ mit } \mathbf{x} \in \mathbb{R}^n \quad (7.15)$$

Das Optimum befindet sich an der Stelle $\mathbf{x}^* = (1, \dots, 1)^T$, mit $B_5(\mathbf{x}^*) = 0$. Für höhere Dimensionen besitzt die Rosenbrock-Funktion ein lokales Minimum an der Stelle $\mathbf{x}^* = (-1, \dots, 1)^T$. Abbildung 7.22 zeigt das Diagramm dieser Funktion B_5 sowie das zugehörige Diagramm der Höhenlinien. Zur Verdeutlichung der Topologie ist die Summe der ungeraden Koordinaten der Testfunktion auf die x_1 -Achse des Diagramms, die Summe der geraden Koordinaten auf die x_2 -Achse des Diagramms aufgetragen. Das Diagramm zeigt damit die Funktion

$$B_5(x_1, x_2) = 5(x_1 - 1)^2 + 500(x_1^2 - x_2)^2 + 4(x_2 - 1)^2 + 400(x_2^2 - x_1)^2 \quad (7.16)$$

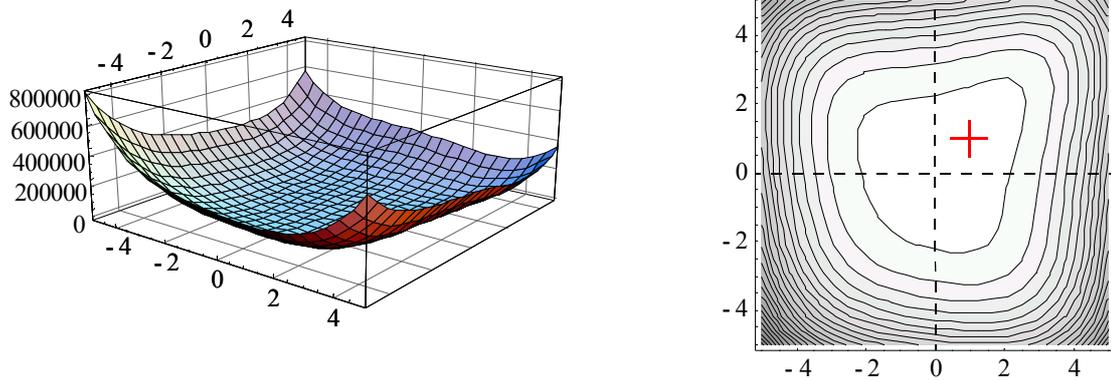


Abb. 7.22: Verlauf der Rosenbrock-Funktion B_5

Die Abbildung 7.23 zeigt die Resultate der Testläufe für die (15, 100)-Evolutionstrategien mit den unterschiedlichen Adaptationsvarianten.

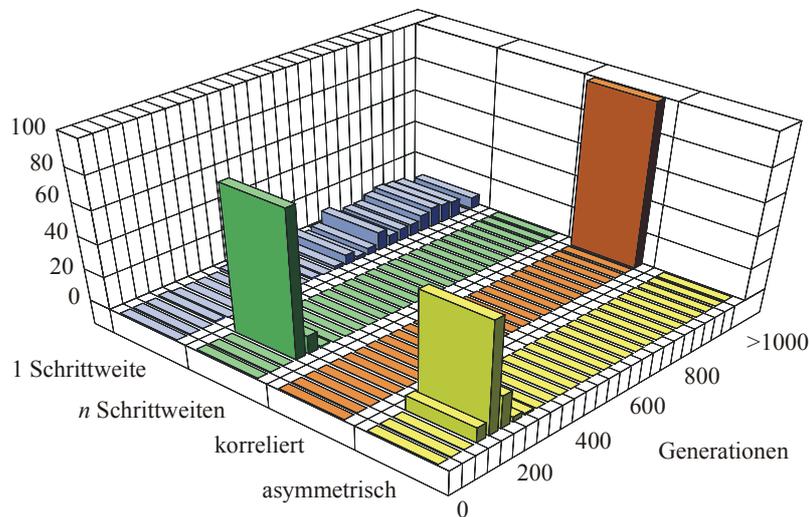


Abb. 7.23: Experimentelle Resultate zur Testfunktion B_5

Das Optimum der Rosenbrock-Funktion B_5 wird von der Evolutionstrategie mit Adaptation einer Schrittweite nur sehr langsam gefunden. Die Anzahl der benötigten Generationen ist ca. 400% höher als bei der Evolutionstrategie mit Adaptation von n Schrittweiten und der asymmetrischen Evolutionstrategie. Diese beiden Varianten finden das Optimum zuverlässig. Die Evolutionstrategie mit Adaptation von Schrittweiten und Lagewinkeln stagniert bei einem Zielfunktionswert um 1.0.

7.2-5 B_6 - Ellipse

Die Testfunktion B_6 besitzt eine Skalierung der Koordinatenachsen, bei der benachbarte Koordinaten um den konstanten Faktor $1000^{1/(n-1)}$ skaliert sind. Die Testfunktion B_6 ist den linearen Kugelmodellen ähnlich. Die Definition von B_6 lautet:

$$B_6(\mathbf{x}) = \sum_{i=1}^n \left(1000^{\frac{i-1}{n-1}} x_i \right)^2, \text{ mit } \mathbf{x} \in \mathbb{R}^n \tag{7.17}$$

Abbildung 7.24 zeigt das Diagramm dieser Funktion B_6 sowie das zugehörige Diagramm der Höhenlinien. Zur Verdeutlichung der Topologie ist die Summe der ungeraden Koordinaten der Testfunktion auf die x_1 -Achse des Diagramms, die Summe der geraden Koordinaten auf die x_2 -Achse des Diagramms aufgetragen. Das Diagramm zeigt damit die Funktion

$$B_6(x_1, x_2) = 10001x_1^2 + 100010 \cdot 10^{1/3}x_1^2 + 100 \cdot 10^{2/3}x_1^2 + 1000100x_2^2 + 100 \cdot 10^{1/3}x_2^2 + 10001 \cdot 10^{2/3}x_2^2 \quad (7.18)$$

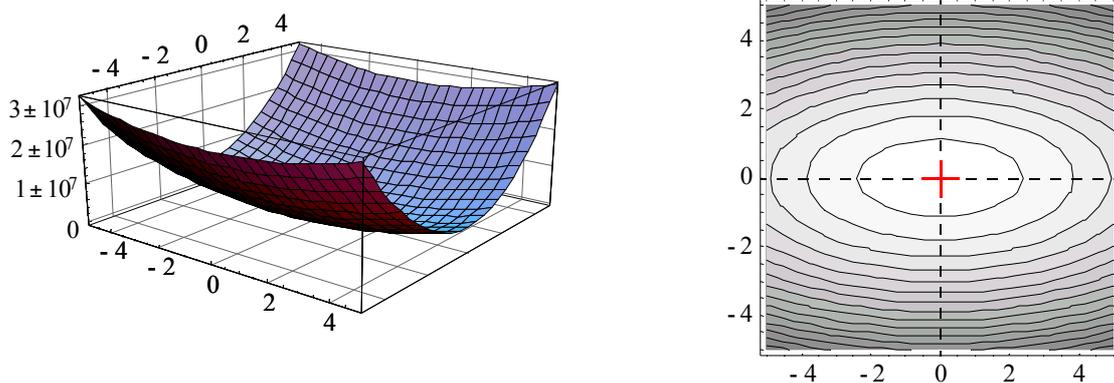


Abb. 7.24: Verlauf der Ellipse B_6

Die folgende Abbildung 7.25 zeigt die Ergebnisse der Testläufe. Die maximale Generationenzahl betrug in dieser Testreihe 500, zu jeder Adaptationsvariante wurden 100 Testläufe durchgeführt.

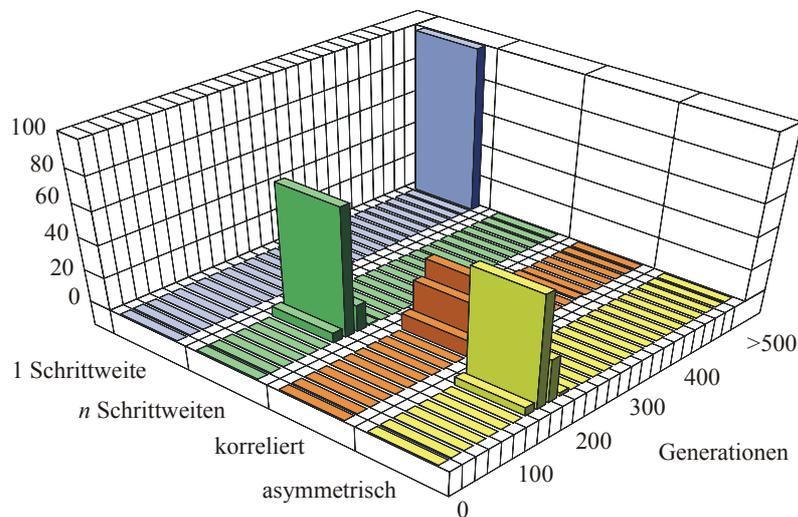


Abb. 7.25: Experimentelle Resultate zur Testfunktion B_6

Mit Ausnahme der Evolutionsstrategie mit Adaptation einer Schrittweite können alle Varianten das Optimum der Testfunktion B_6 zuverlässig finden. Die Evolutionsstrategie mit Adaptation einer Schrittweite konvergiert bei dieser Testfunktion langsam. Die Annäherung an das Optimum erfolgt mit linear sinkendem Zielfunktionswert. Allerdings wurde nach 10^6 Generationen erst ein Zielfunktionswert von 10^{-3} erreicht.

7.2-6 B_7 - Zigarre

Die Testfunktion B_7 besitzt eine “lange” und $n - 1$ “kurze” Koordinatenachsen. Die Höhenlinien der Testfunktion gleichen damit einer Zigarre. Die Definition von B_7 lautet:

$$B_7(\mathbf{x}) = x_1^2 + 10^6 \sum_{i=2}^n x_i^2, \text{ mit } \mathbf{x} \in \mathbb{R}^n \tag{7.19}$$

Abbildung 7.26 zeigt das Diagramm dieser Funktion B_7 sowie das zugehörige Diagramm der Höhenlinien. Zur Verdeutlichung der Topologie ist die x_1 -Koordinate der Testfunktion mit dem Skalierungsfaktor 1000 auf die x_1 -Achse des Diagramms und die Summe der $n - 1$ weiteren Koordinaten auf die x_2 -Achse des Diagramms aufgetragen. Die Gleichung der Abbildung 7.26 lautet damit:

$$B_6(x_1, x_2) = 10^3 x_1^2 + 9 \cdot 10^6 x_2^2 \tag{7.20}$$

Die Abbildung 7.27 zeigt den unskalierten Verlauf der Testfunktion B_7 mit der Gleichung

$$B_6(x_1, x_2) = x_1^2 + 9 \cdot 10^6 x_2^2 \tag{7.21}$$

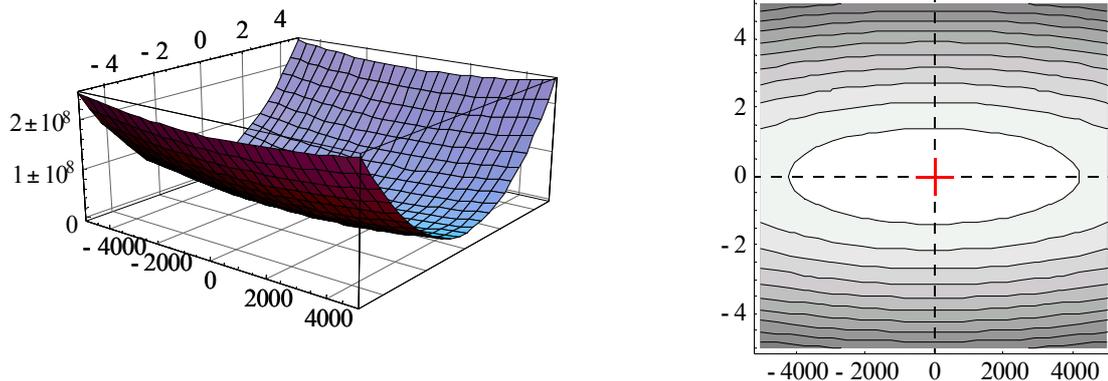


Abb. 7.26: Verlauf der Zigarre B_7 mit Skalierung

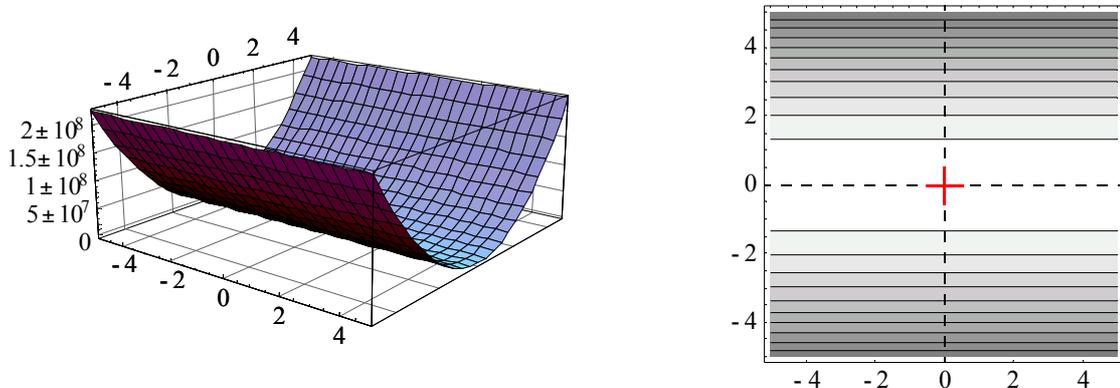


Abb. 7.27: Verlauf der Zigarre B_7 ohne Skalierung

Die folgende Abbildung 7.28 zeigt die Ergebnisse der Testläufe. Die maximale Generationenzahl betrug in dieser Testreihe 500, zu jeder Adaptationsvariante wurden 100 Testläufe durchgeführt.

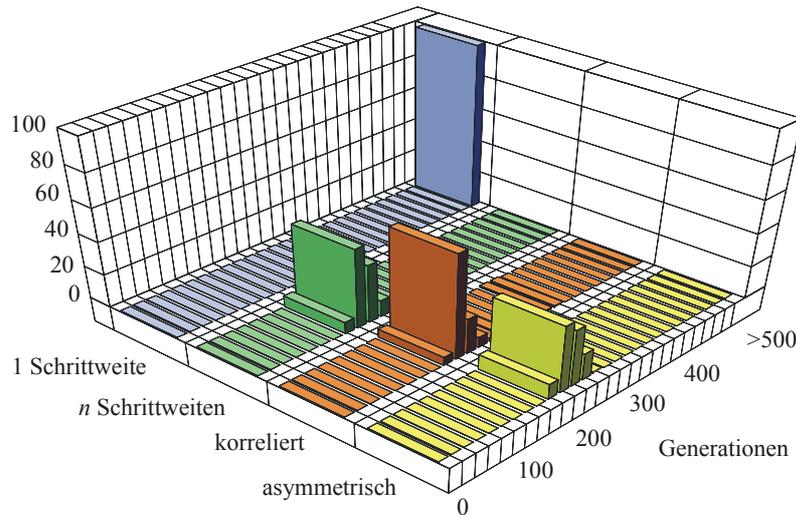


Abb. 7.28: Experimentelle Resultate zur Testfunktion B_7

Auch bei der Testfunktion B_7 können mit Ausnahme der Evolutionsstrategie mit Adaptation einer Schrittweite alle Varianten das Optimum zuverlässig finden. Die Evolutionsstrategie mit Adaptation einer Schrittweite erreicht nach ca. 100 Generationen einen Zielfunktionswert von 0.1. Der weitere Fortschritt, bezogen auf den Zielfunktionswert, liegt danach jedoch nur noch bei 10^{-4} pro 1000 Generationen. Alle Läufe dieser Variante wurden nach 10^6 Generationen endgültig abgebrochen.

7.2-7 B_8 -Tablette

Die Testfunktion B_8 besitzt eine "kurze" und $n - 1$ "lange" Koordinatenachsen und stellt damit den Gegensatz zu der Testfunktion B_7 dar. Die Höhenlinien der Testfunktion gleichen somit einer tablettenförmigen Ellipse. Die Definition von B_8 lautet:

$$B_8(\mathbf{x}) = 10^6 x_1^2 + \sum_{i=2}^n x_i^2, \text{ mit } \mathbf{x} \in \mathbb{R}^n \quad (7.22)$$

Abbildung 7.32 zeigt das Diagramm dieser Funktion B_8 sowie das zugehörige Diagramm der Höhenlinien. Zur Verdeutlichung der Topologie ist die x_1 -Koordinate der Testfunktion auf die x_1 -Achse des Diagramms und die Summe der $n - 1$ weiteren Koordinaten mit dem Skalierungsfaktor 100 auf die x_2 -Achse des Diagramms aufgetragen. Die Gleichung der Abbildung lautet daher:

$$B_8(x_1, x_2) = 10^6 x_1^2 + 9x_2^2 \quad (7.23)$$

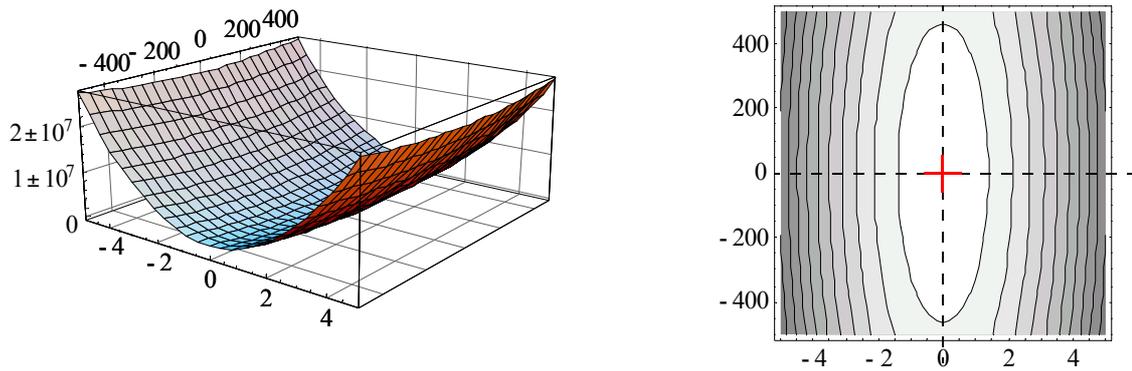


Abb. 7.29: Verlauf der Tablette B_8 , Skalierung der y-Achse

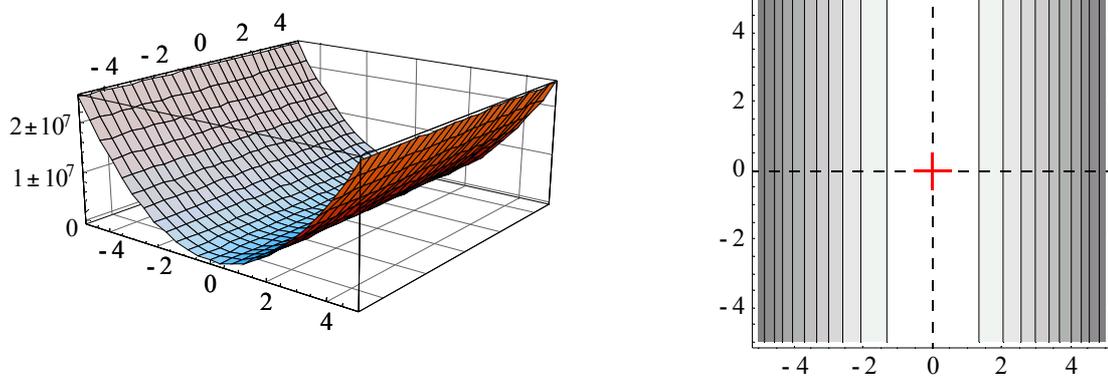


Abb. 7.30: Verlauf der Tablette B_8

Die folgende Abbildung 7.31 zeigt die Ergebnisse der Testläufe. Die maximale Generationenzahl betrug in dieser Testreihe 200, zu jeder Adaptationsvariante wurden 100 Testläufe durchgeführt.

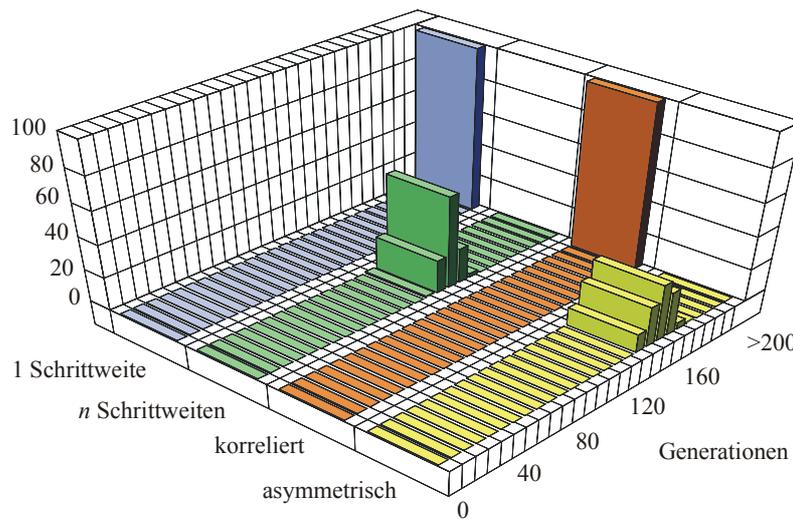


Abb. 7.31: Experimentelle Resultate zur Testfunktion B_8

Das Optimum der Testfunktion B_8 kann zuverlässig von der Evolutionsstrategie mit Adaptation von n Schrittweiten und der asymmetrischen Evolutionsstrategie gefunden werden. Bei der Evolutionsstrategie mit Adaptation einer Schrittweite findet eine Stagnation bei einem Zielfunktionswert zwischen 10 und 100 statt. Die Evolutionsstrategie mit Adaptation von Schrittweiten und Lagewinkeln divergiert. Dieses Verhalten wurde bis zu einer maximalen Generationenzahl von 10000 bei einem Zielfunktionswert von 10^{120} oder mehr beobachtet.

7.2-8 B_9 - Summe verschiedener Potenzen

Die Definition von B_9 lautet:

$$B_9(\mathbf{x}) = \sum_{i=1}^n |x_i|^{2+10\frac{i-1}{n-1}}, \text{ mit } \mathbf{x} \in \mathbb{R}^n \quad (7.24)$$

Zur Verdeutlichung der Topologie ist die Summe der ungeraden Koordinaten der Testfunktion auf die x_1 -Achse des Diagramms, die Summe der geraden Koordinaten auf die x_2 -Achse des Diagramms aufgetragen. Die Gleichung der Abbildung 7.32 lautet damit:

$$B_9(x_1, x_2) = |x_1|^2 + |x_1|^{38/9} + |x_1|^{58/9} + |x_1|^{26/3} + |x_1|^{98/9} + |x_2|^{12} + |x_2|^{88/9} + |x_2|^{68/9} + |x_2|^{16/3} + |x_2|^{28/9} \quad (7.25)$$

Abbildung 7.32 zeigt das Diagramm dieser Funktion.

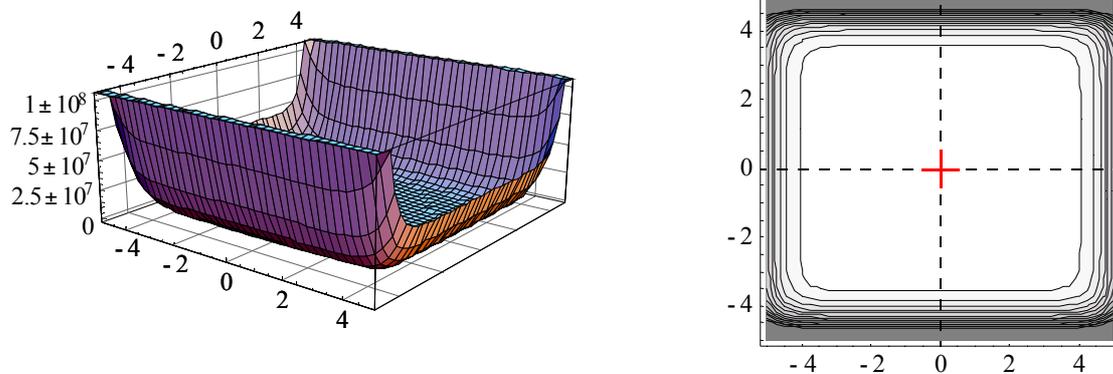


Abb. 7.32: Verlauf der Summe verschiedener Potenzen B_9

Die folgende Abbildung 7.33 zeigt die Ergebnisse der Testläufe. Die maximale Generationenzahl betrug in dieser Testreihe 200, zu jeder Adaptationsvariante wurden 100 Testläufe durchgeführt.

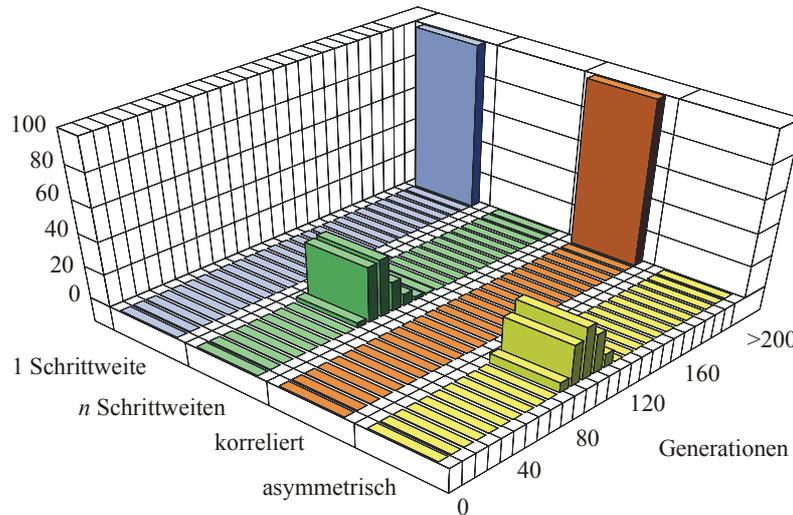


Abb. 7.33: Experimentelle Resultate zur Testfunktion B_9

Die Evolutionsstrategie mit Adaptation einer Schrittweite erreicht nach ca. 500 Generationen Zielfunktionswerte zwischen 10^{-7} und 10^{-9} . Danach ist eine langsame Konvergenz zu beobachten, bei der das Abbruchkriterium nach ca. 50000 Generationen erreicht wird. Die Evolutionsstrategie mit Adaptation von Schrittweiten und Lagewinkeln erreicht nach ca. 200 Generationen einen Zielfunktionswert von 10^{-8} . Danach stagniert diese Variante mit Zielfunktionswerten zwischen 10^{-7} und 10^{-9} . Das Optimum der Testfunktion B_9 kann zuverlässig von der Evolutionsstrategie mit Adaptation von n Schrittweiten und der asymmetrischen Evolutionsstrategie gefunden werden.

7.2-9 B_{10} - Parabelgrat

Die Definition von B_{10} lautet:

$$B_{10}(\mathbf{x}) = -x_1 + \sum_{i=2}^n x_i^2, \text{ mit } \mathbf{x} \in \mathbb{R}^n \tag{7.26}$$

Es gilt das Abbruchkriterium $Q_{stop} = -10^5$.

Zur Verdeutlichung der Topologie ist die Summe der ungeraden Koordinaten der Testfunktion B_{10} auf die x_1 -Achse des Diagramms, die Summe der geraden Koordinaten auf die x_2 -Achse des Diagramms aufgetragen. Die Gleichung der Abbildung 7.34 lautet damit

$$B_{10}(x_1, x_2) = -x_1 + 9x_2^2 \tag{7.27}$$

Abbildung 7.34 zeigt das Diagramm dieser Funktion.

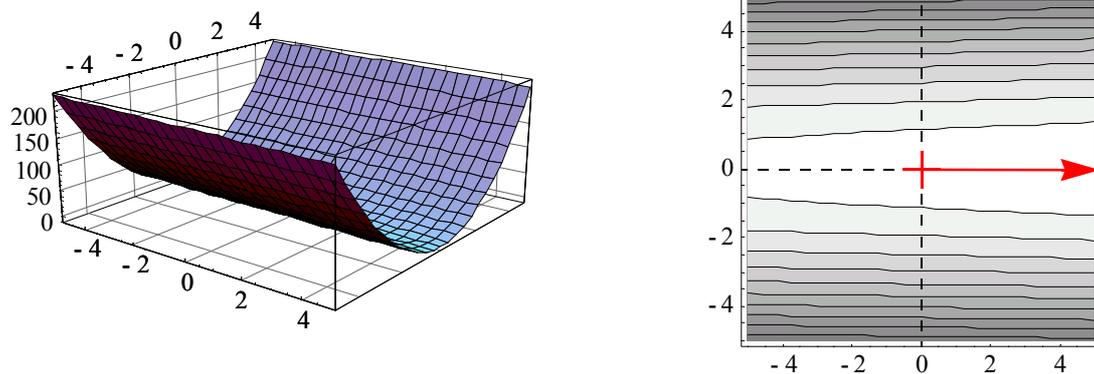


Abb. 7.34: Verlauf des Parabelgrates B_{10}

Die folgende Abbildung 7.35 zeigt die Ergebnisse der Testläufe. Die maximale Generationenzahl betrug in dieser Testreihe 100, zu jeder Adaptationsvariante wurden 100 Testläufe durchgeführt.

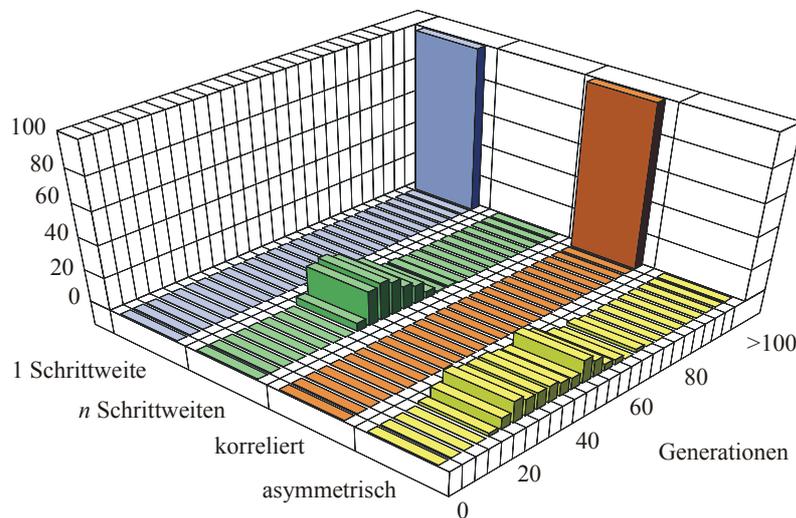


Abb. 7.35: Experimentelle Resultate zur Testfunktion B_{10}

Die Evolutionsstrategie mit Adaptation einer Schrittweite konvergiert langsam. Nach 10000 Generationen ist ein Zielfunktionswert von ca. -10^3 erreicht. Die Evolutionsstrategie mit Adaptation von Schrittweiten und Lagewinkeln erreicht nach ca. 200 Generationen einen Zielfunktionswert von -10^3 . Danach stagniert diese Variante mit Zielfunktionswerten zwischen -10^3 und -10^4 . Das Optimum der Testfunktion B_{10} kann zuverlässig von der Evolutionsstrategie mit Adaptation von n Schrittweiten und der asymmetrischen Evolutionsstrategie gefunden werden.

7.2-10 B_{11} - Spitzer Grat

Die Testfunktion B_{11} ist eine Variante des Kugelmodells und ähnlich der Testfunktion B_{10} , dem Parabelgrat. Auch bei der Testfunktion B_{11} wird eine Koordinate (o. B.d A. die erste Koordinate) linear in den Funktionswert einbezogen, die Koordinaten 2 bis n werden mit dem 100-fachen ihres geometrischen Mittelwertes einbezogen. Die Definition von B_{11} lautet:

$$B_{11}(\mathbf{x}) = -x_1 + 100\sqrt{\sum_{i=2}^n x_i^2}, \text{ mit } \mathbf{x} \in \mathbb{R}^n \tag{7.28}$$

Analog zur Testfunktion B_{10} gilt auch hier das Abbruchkriterium $Q_{stop} = -10^5$. Zur Verdeutlichung der Topologie ist die Summe der ungeraden Koordinaten der Testfunktion B_{11} auf die x_1 -Achse des Diagramms, die Summe der geraden Koordinaten auf die x_2 -Achse des Diagramms aufgetragen. Die Gleichung der Abbildung 7.36 lautet damit

$$B_{11}(x_1, x_2) = -x_1 + 300\sqrt{x_2^2} \tag{7.29}$$

Abbildung 7.36 zeigt das Diagramm dieser Funktion.

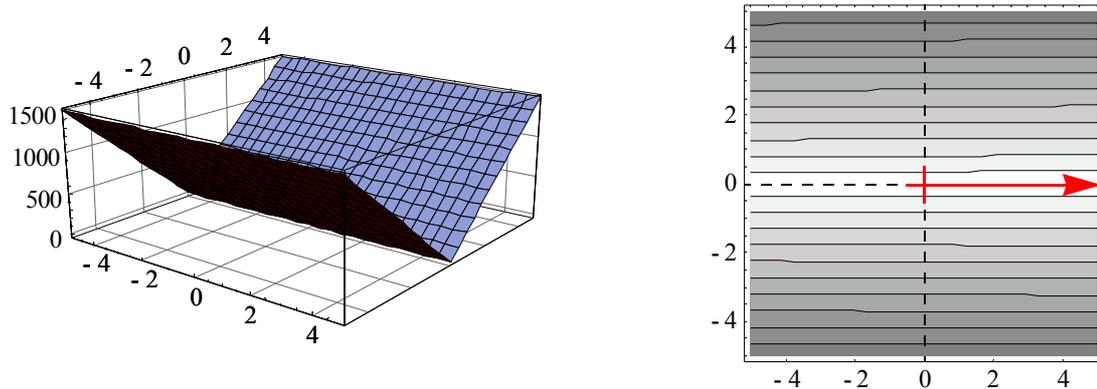


Abb. 7.36: Verlauf des spitzen Grates B_{11}

Das Optimum von B_{11} befindet sich an den Stellen $\mathbf{x}^* = (\infty, 0, \dots, 0)^T$, mit $B_{11}(\mathbf{x}^*) = -\infty$. Die folgende Abbildung 7.37 zeigt die Ergebnisse der Testläufe. Die maximale Generationenzahl betrug in dieser Testreihe 300, zu jeder Adaptationsvariante wurden 100 Testläufe durchgeführt.

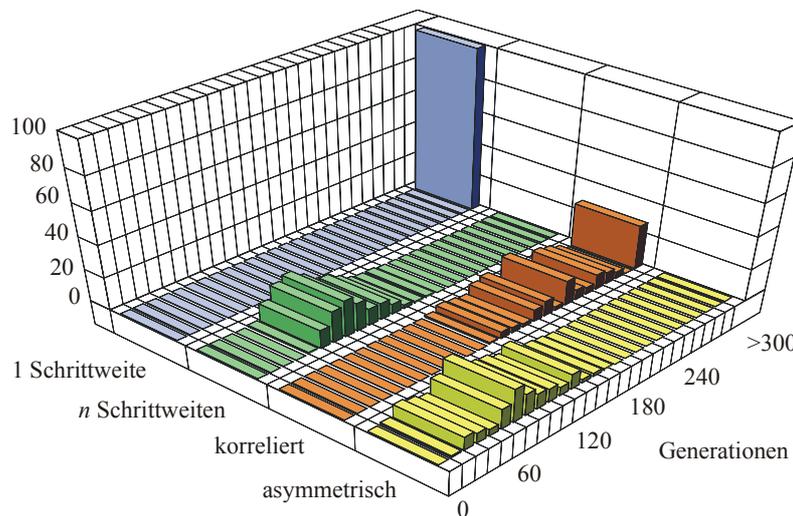


Abb. 7.37: Experimentelle Resultate zur Testfunktion B_{11}

Die Evolutionsstrategie mit Adaptation einer Schrittweite konvergiert nicht, da die Schrittweite bereits nach wenigen 100 Generationen kleiner als 10^{-20} ist. Alle anderen Varianten konvergieren zuverlässig. Lediglich die Evolutionsstrategie mit Adaptation von Schrittweiten und Lagewinkeln

benötigt mehr Generationen als in der Abbildung dargestellt wird. Bei allen Läufen dieser Variante wurde das Optimum innerhalb von 500 Generationen gefunden.

7.2-11 Bewertung der Resultate

Auch die Problemklassen nach Hansen und Ostermeier zeigen das aus dem vorhergehenden Abschnitt bekannte Verhalten der unterschiedlichen Varianten der Evolutionsstrategien. Die Evolutionsstrategie mit Adaptation einer Schrittweite kann das Optimum nur bei relativ einfachen Testfunktionen finden. Die Evolutionsstrategie mit Adaptation von Schrittweiten und Lagewinkeln divergiert häufig oder endet in einer Stagnationsphase. Die Evolutionsstrategie mit Adaptation von n Schrittweiten und die asymmetrische Evolutionsstrategie zeigen ähnliches Verhalten. Bei Testfunktionen mit ausgeprägter Richtungscharakteristik, z. B. die Funktionen B_{10} und B_{11} , benötigt die asymmetrische Evolutionsstrategie jedoch in der Regel weniger Generationen, um das Optimum zu finden.

7.3 Hoch-dimensionale Probleme

Es gibt Optimierverfahren, die für Testfunktionen mit geringer Dimensionalität ($n < 30$) gute Ergebnisse liefern, jedoch bei den gleichen Funktionen mit höherer Dimensionalität versagen. Der folgende Abschnitt stellt das Verhalten der asymmetrischen Evolutionsstrategie bei hohen Dimensionen dar. Diese Untersuchung findet am Beispiel des Ebenenmodells, des Kugelmodells und der Doppelsumme statt.

Bei allen Experimenten in diesem Abschnitt wurden im Vorfeld Untersuchungen über das optimale Verhältnis von Anzahl Eltern zu Anzahl Nachkommen durchgeführt. Die Resultate sind in dem Anhang E (Vollständige Übersicht über die Experimente, Seite 277) dargestellt. Somit gelten die Resultate in diesem Abschnitt nicht mehr für eine (15, 100)-ES bzw. eine (15, 100)-aES, sondern für die aus diesen Untersuchungen abgeleiteten optimalen Belegungen von μ und λ . Aus Gründen des Rechenaufwandes sind alle Experimente fünf mal durchgeführt worden. Die Diagramme stellen den Mittelwert dieser fünf Experimente dar, bezogen auf die Anzahl der Zielfunktionsauswertungen. Die Generationen sind in diesen Diagrammen nicht abgetragen, da sonst Experimente mit hoher Nachkommenzahl einen nicht gerechtfertigten, scheinbaren Vorteil hätten. Zusätzlich zu der Mittelung der Testläufe sind die Resultate durch eine Ausgleichskurve angenähert, um den Einfluss von statistischen Ausreißern zu vermindern.

7.3-1 B_1 - Ebenenmodell

Das Ebenenmodell entspricht der Testfunktion B_1 aus 7.2-1 (B1 - Ebenenmodell, Seite 197):

$$B_1(\mathbf{x}) = -x_1, \text{ mit } \mathbf{x} \in \mathbb{R}^n \quad (7.30)$$

Das Optimum von B_1 befindet sich an den Stellen $\mathbf{x}^* = (\infty, x_2, \dots, x_n)^T$, mit $B_1(\mathbf{x}^*) = -\infty$, die Werte für x_2 bis x_n sind beliebig. Als Abbruchkriterium dient das Erreichen des Zielfunktionswertes -10^{10} .

Die Abbildung 7.38 stellt den Verlauf der erforderlichen Zielfunktionsauswertungen in Abhängigkeit von der Dimensionalität dar. Abgetragen sind die Resultate für die Evolutionsstrategie mit Adaptation von n Schrittweiten sowie die Resultate der asymmetrischen Evolutionsstrategie.

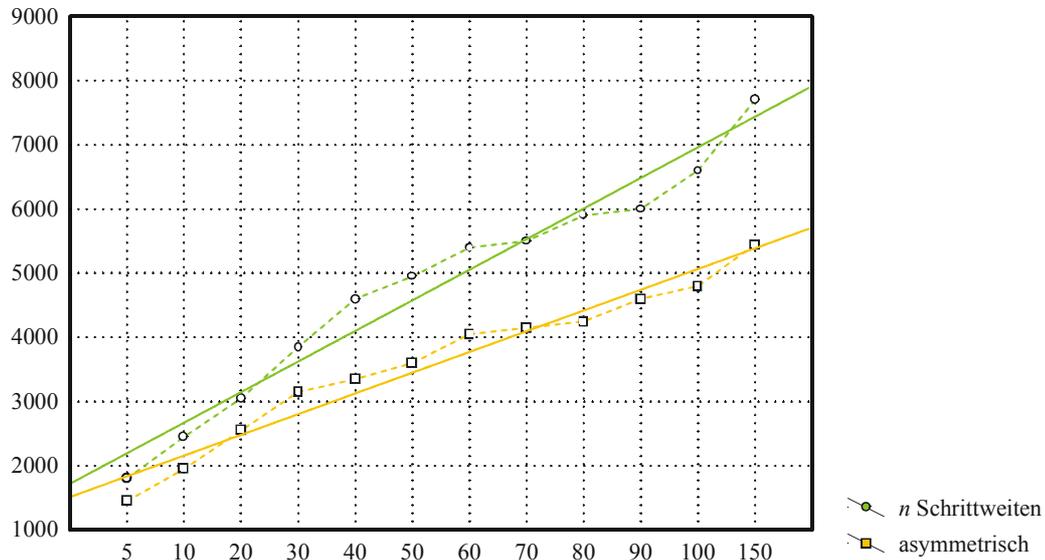


Abb. 7.38: Ebenenmodell für hohe Dimensionen

Die asymmetrische Evolutionsstrategie benötigt in allen Fällen weniger Zielfunktionsauswertungen als die Evolutionsstrategie mit Adaptation von n Schrittweiten. Aus den Diagrammen des Abschnitts E.1 (B1 - Ebenenmodell in hohen Dimensionen, Seite 277) wird zudem deutlich, dass die asymmetrische Evolutionsstrategie ein nahezu identisches Verhalten für unterschiedliche Anzahl von Eltern und Nachkommen zeigt. Die Evolutionsstrategie mit Adaptation von n Schrittweiten reagiert sensibler auf das Abweichen der optimalen Belegung.

7.3-2 f_1 - Kugelmodell

Das Kugelmodell ist eine n -dimensionale quadratische Funktion, die kontinuierlich, streng konvex und unimodal ist. Sie ist definiert durch:

$$f_1(\mathbf{x}) = \sum_{i=1}^n x_i^2, \text{ mit } \mathbf{x} \in \mathbb{R}^n \tag{7.31}$$

Das Optimum von f_1 befindet sich an der Stelle $\mathbf{x}^* = (0, \dots, 0)^T$, mit $f(\mathbf{x}^*) = 0$.

Die Abbildung 7.39 stellt den Verlauf der erforderlichen Zielfunktionsauswertungen in Abhängigkeit von der Dimensionalität dar. Abgetragen sind die Resultate für die Evolutionsstrategie mit Adaptation von n Schrittweiten sowie die Resultate der asymmetrischen Evolutionsstrategie.

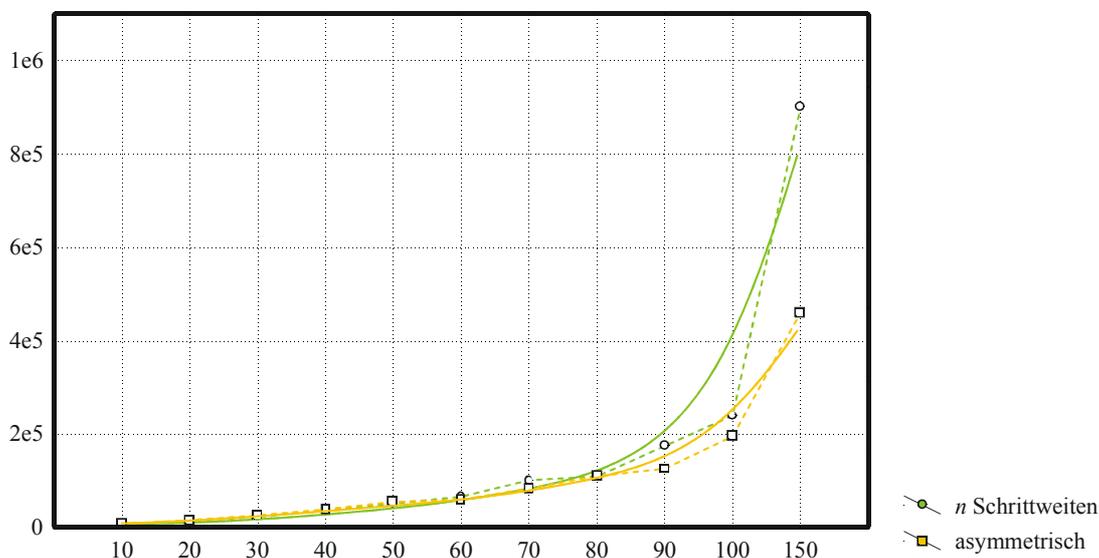


Abb. 7.39: Kugelmodell für hohe Dimensionen

Die asymmetrische Evolutionsstrategie benötigt mit steigender Dimensionalität weniger Zielfunktionsauswertungen als die Evolutionsstrategie mit Adaptation von n Schrittweiten. Grundsätzliche Unterschiede bei der Untersuchung des optimalen Verhältnisses von Eltern zu Nachkommen gibt es jedoch nicht, wie der Abschnitt E.2 (f_1 - Kugelmodell in hohen Dimensionen, Seite 281) deutlich macht.

7.3-3 f_4 - Doppelsumme

Die Testfunktion f_4 ist ausführlich im Abschnitt 7.1-4 (f_4 - Doppelsumme, Seite 192) erläutert. Die Grundfunktion f_4 lautet:

$$f_4(\mathbf{x}) = \sum_{i=1}^n \left(\sum_{j=1}^i x_j \right)^2, \text{ mit } \mathbf{x} \in \mathbb{R}^n \quad (\text{G.32})$$

Das Optimum von f_4 befindet sich an der Stelle $\mathbf{x}^* = (0, \dots, 0)^T$, mit $f_4(\mathbf{x}^*) = 0$. Die Abbildung 7.40 stellt den Verlauf der erforderlichen Zielfunktionsauswertungen in Abhängigkeit von der Dimensionalität dar. Abgetragen sind die Resultate für die Evolutionsstrategie mit Adaptation von

n Schrittweiten sowie die Resultate der asymmetrischen Evolutionsstrategie.

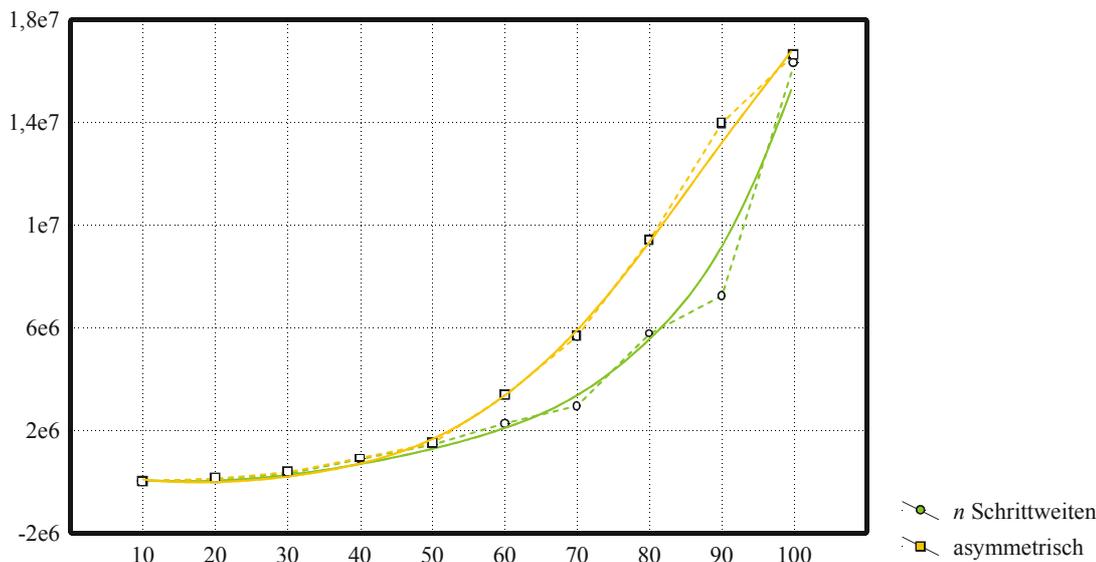


Abb. 7.40: Doppelsumme für hohe Dimensionen

Die asymmetrische Evolutionsstrategie benötigt in allen Fällen mehr Zielfunktionsauswertungen als die Evolutionsstrategie mit Adaptation von n Schrittweiten. Grundsätzliche Unterschiede bei der Untersuchung des optimalen Verhältnisses von Eltern zu Nachkommen gibt es jedoch nicht, wie der Abschnitt E.3 (f_4 - Doppelsumme in hohen Dimensionen, Seite 285) deutlich macht.

7.4 Optimierung von unscharfen Mengen zur Beschreibung von Farben

Der Einsatz von unscharfer Logik zur Beschreibung von Farben und insbesondere zur Beschreibung von Farbschattierungen, ist ein weltweit einmaliger Ansatz, dessen Idee 1998 erstmalig vorgestellt wurde [Reu98]. Die vollständige Beschreibung des Verfahrens findet sich in [Hil00a]. Eine Einbettung des Verfahrens in den Gesamtkomplex *Intelligente Systeme* findet sich in [Reu99]. Im folgenden Abschnitt wird kurz das Verfahren an sich erläutert und daran anschließend die Umwandlung von 2-dimensionalen Häufigkeitsverteilungen in unscharfe Mengen als Optimierproblem beschrieben. Zur Lösung dieses Optimierproblems wird eine asymmetrische Evolutionsstrategie eingesetzt, deren Resultate abschließend diskutiert werden.

7.4-1 Problembeschreibung

Die Idee, Farben durch unscharfe Mengen zu beschreiben und durch ein unscharfes Inferenzsystem auszuwerten, ist die Folge einer Problemstellung aus dem Bereich der industriellen Qualitätssicherung von Widerstandspunktschweißverbindungen. Das Widerstandspunktschweißen ist ein Schweißprozeß, der sich den elektrischen Widerstand von metallischen Blechen zunutze macht, um eine Verbindung zu erzeugen. Mittels Kupfer- oder Wolframelektroden wird ein hoher Strom durch die Bleche geleitet, der zu einer lokalen Erwärmung und Verflüssigung des Metalls führt. Die Elektroden pressen die erwärmten Bleche gegeneinander, sodass sie beim Abkühlen eine feste Verbindung eingehen. Das Widerstandspunktschweißen ist ein wirtschaftlicher Prozeß, da er schnell ist und außer elektrischem Strom keine weiteren Betriebsmittel benötigt. Es wird hauptsächlich im Automobilbau, der Flugzeugindustrie und in dem Maschinenbau angewendet. Der gesamte

Schweißprozeß ist stark automatisierbar. In der Regel werden nur Hilfskräfte oder Roboter benötigt, die das Material bereitstellen oder es in die Schweißmaschine einlegen. Dieser hohe Grad an Automatisierung hat einerseits zu dem wirtschaftlichen Erfolg des Widerstandspunktschweißverfahrens geführt, andererseits aber auch aufwendige Verfahren zur Qualitätssicherung notwendig gemacht. In der Regel müssen Verfahren zur Überprüfung von Punktschweißverbindungen auf zerstörende Prüfmethoden zurückgreifen. Zu diesen Methoden gehören u. a. Schälzug-, Torsions- und Ausknöpfverfahren. Mikro- und makroskopische Überprüfungen erfordern ebenfalls eine Auftrennung und damit eine Zerstörung der Schweißverbindung. Zerstörungsfreie Prüfmethoden, wie z. B. Gamma- oder Röntgenuntersuchungen, benötigen entweder große und kostenintensive Messeinrichtungen oder sind sehr empfindlich und können nicht in direkter Nähe zur Fertigungseinrichtung eingesetzt werden.

7.4-2 Qualitätsmerkmale

Die Qualitätskriterien einer Widerstandspunktschweißverbindung können in zwei Hauptbereiche unterteilt werden: physikalische und optische Kriterien. Physikalische Kriterien, wie z. B. Härte und Zugfestigkeit, beschreiben physikalische Eigenschaften einer Punktschweißverbindung und werden in dieser Arbeit nicht weiter untersucht. Schwerpunkt dieser Arbeit ist die Untersuchung optischer Qualitätskriterien, z. B. Verfärbung des Schweißpunktes oder der Wärmeeinflusszone sowie die Untersuchung von Qualitätskriterien, die aus einer optischen Vermessung abgeleitet werden können. Beispiele für diese Kriterien sind Form und Abmessung der charakteristischen Bereiche einer Punktschweißverbindung, wie sie in Abbildung 7.41 dargestellt sind.

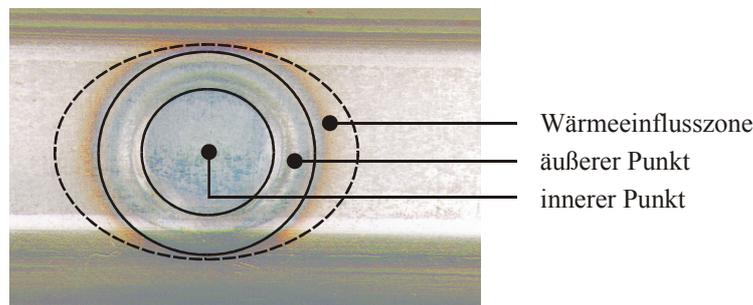


Abb. 7.41: Informationstragende Bereiche einer Punktschweißverbindung

Wissensakquisitionen mit Schweißern und Schweißfachingenieuren haben zu folgender Liste von optischen Qualitätskriterien geführt, die sich für eine wissensbasierte Verarbeitung eignen:

- Größe, Form und Farbe des inneren Punktes
- Größe, Form und Farbe des äußeren Punktes
- Größe, Form und Farbe der Wärmeeinflusszone

Die folgende Abbildung 7.42 stellt einen Schweißpunkt mit guter und einen Schweißpunkt mit schlechter Qualität dar. Die Ausprägung der einzelnen charakteristischen Bereiche ist deutlich zu erkennen. Schweißfachingenieure und erfahrene Einrichter eines Schweißautomaten können anhand der Färbung der Schweißpunkte einen ersten Rückschluß auf die Qualität der Schweißverbindung ableiten.

Problematisch ist jedoch, dass der Begriff *Farbe* in hohem Maße subjektiv geprägt ist. Jeder Mensch hat leicht unterschiedliche Vorstellungen von einer bestimmten Farbe bzw. würde eine vorgegebene Farbe in der Regel mit anderen Worten beschreiben als ein Kollege, siehe auch [Bac98, Har98,

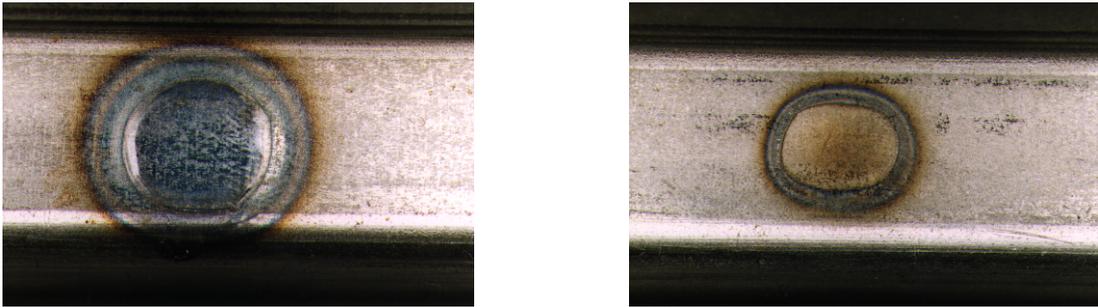


Abb. 7.42: Schweißpunkte mit guter (links) und mangelhafter (rechts) Qualität

Sto98]. Lösungsansätze, die diese problem inherente Unschärfe nicht berücksichtigen, sind daher auf die hier beschriebene Problemstellung nicht anwendbar. Gleichzeitig haben sich technisch orientierte Farbmodelle als nicht aussagekräftig herausgestellt. Abhilfe schafft hier die unscharfe Beschreibung von Farben, basierend auf 2-dimensionalen Untermodellen des HSI-Farbmodells [Reu98, Hil00a].

7.4-3 Unscharfe Farbverarbeitung

Für Details bezüglich der unscharfen Farbverarbeitung, insbesondere der regelbasierten Auswertung, sei auf [Cox98, Hil00a, Hil98b, Zad92] verwiesen. Für das Verständnis ist an dieser Stelle lediglich der Aufbau einer unscharfen Menge zur Farbbeschreibung von Bedeutung. Dieser Aufbau wird am Beispiel der Farbe blau in Abbildung 7.43 erläutert. Eine farbcharakterisierende, unscharfe

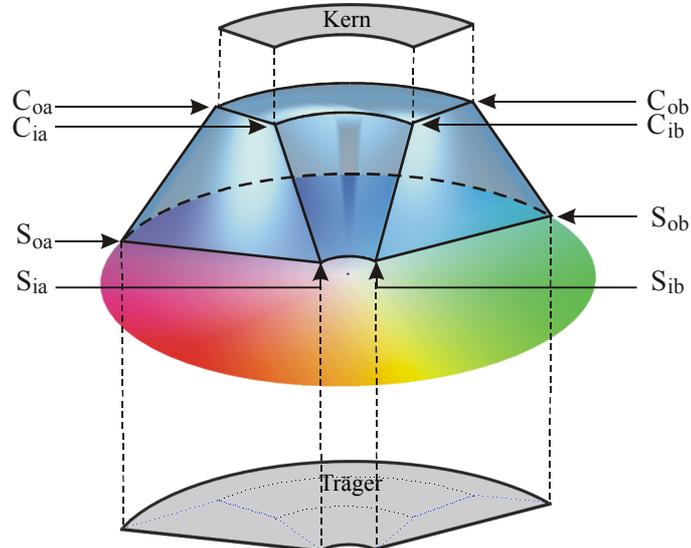


Abb. 7.43: Definition der unscharfen Menge blau

Menge ist über einem 2-dimensionalen Unterraum des HSI-Farbraumes definiert. Sie enthält einen Kernbereich mit Zugehörigkeitswert 1.0, in der Abbildung als oberes graues Feld markiert. Der Träger ist in der Abbildung als unteres, graues Feld markiert. Die graduelle Zugehörigkeit einer konkreten Farbe zu der unscharfen Menge blau wird durch die Flanken von Träger zu Kern erreicht. Gerade diese Flanken sind die Besonderheit der unscharfen Beschreibung von Farben, da hier subjektive Heuristiken in die Farbbeschreibung einfließen können. Die geometrische Form einer

unscharfen Farbbeschreibung entspricht damit einem kartesischen Pyramidenstumpf, der in ein 3-dimensionales Zylinderkoordinatensystem projiziert wurde.

7.4-4 Transformation der Farbmodelle

Das HSI-Farbmodell, als Ausgabeformat von elektronischen Farbkameras, ist zur Zeit nicht weit verbreitet. Lediglich Kameras für Spezialanwendungen verfügen über dieses Format, alle anderen liefern ihre Signale in Kompositform oder als RGB-Signal. Das RGB-Signal kann in das HSI-Format transformiert werden. Hierzu werden folgende Operatoren und Vektoren verwendet:

$$\|\mathbf{v}\| = \frac{\mathbf{v}}{|\mathbf{v}|} = \begin{pmatrix} v_1 / \sqrt{v_1^2 + v_2^2 + v_3^2} \\ v_2 / \sqrt{v_1^2 + v_2^2 + v_3^2} \\ v_3 / \sqrt{v_1^2 + v_2^2 + v_3^2} \end{pmatrix} \quad (7.33)$$

$$\mathbf{u} \cdot \mathbf{v} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = u_1 v_1 + u_2 v_2 + u_3 v_3 \quad (7.34)$$

$$\mathbf{u} \times \mathbf{v} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} \times \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} u_2 v_3 - u_3 v_2 \\ u_3 v_1 - u_1 v_3 \\ u_1 v_2 - u_2 v_1 \end{pmatrix} \quad (7.35)$$

$$\mathbf{I}_u = \left\| \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\| = \begin{pmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{pmatrix} \quad (7.36)$$

$$\mathbf{R} = \begin{pmatrix} 2 \\ -1 \\ -1 \end{pmatrix} \quad (7.37)$$

$$\mathbf{G}_u = \|\mathbf{I}_u \times \mathbf{R}\| = \begin{pmatrix} 0 \\ 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix} \quad (7.38)$$

$$\mathbf{R}_u = \|\mathbf{R}\| = \begin{pmatrix} \sqrt{6}/3 \\ -\sqrt{6}/6 \\ -\sqrt{6}/6 \end{pmatrix} \quad (7.39)$$

Eine Farbe \mathbf{C}_{RGB} , die im RGB-Farbmodell als $\mathbf{C}_{RGB} = (R, G, B)^T$ gegeben ist, kann in die identische Farbe (bezogen auf die visuelle Farberscheinung) \mathbf{C}_{HSI} transformiert werden, die im HSI-Farbmodell als $\mathbf{C}_{HSI} = (H, S, I)^T$ gegeben ist. Dazu sind folgende Berechnungsschritte notwendig:

$$H(\mathbf{C}_{RGB}) = \text{atant}(\mathbf{C}_{RGB} \cdot \mathbf{G}_w \cdot \mathbf{C}_{RGB} \cdot \mathbf{R}_u) \quad (7.40)$$

$$S(\mathbf{C}_{RGB}) = 1 - \frac{\min(\mathbf{C}_{RGB})}{I(\mathbf{C}_{RGB})} \quad (7.41)$$

$$I(\mathbf{C}_{RGB}) = \frac{1}{\sqrt{3}}(\mathbf{C}_{RGB} \cdot \mathbf{I}_u) \quad (7.42)$$

Die Funktion $\text{atant}(x, y)$ ist wie folgt definiert:

$$\text{atant}(x, y) = \begin{cases} \text{atanx}(x, y) & , \text{ falls } \text{atanx}(x, y) \geq 0 \\ \text{atanx}(x, y) + 2\pi & , \text{ sonst} \end{cases} \quad (7.43)$$

$$\text{atanx}(x, y) = \begin{cases} \arctan(y/x) & , \text{ falls } x > 0 \\ \arctan(y/x) + \pi & , \text{ falls } x < 0, y \geq 0 \\ \arctan(y/x) - \pi & , \text{ falls } x < 0, y < 0 \end{cases} \quad (7.44)$$

Die undefinierten Funktionswerte für $x = 0$ sind wie folgt definiert:

$$\text{atanx}(0, y) = \begin{cases} 0 & , \text{ falls } y = 0 \\ \pi/2 & , \text{ falls } y > 0 \\ -\pi/2 & , \text{ sonst} \end{cases} \quad (7.45)$$

7.4-5 Generierung unscharfer Farbbeschreibungen

Die unscharfen Farbbeschreibungen können mittels einer Benutzerschnittstelle direkt definiert werden. Hierfür ist keinerlei Optimierung notwendig. Es ist jedoch auch möglich, die relevanten Farben eines bestimmten Bereiches als relative Häufigkeitsverteilung darzustellen und darauf aufbauend

eine unscharfe Menge automatisch zu generieren bzw. zu optimieren. Eine typische Häufigkeitsverteilung ist in der Abbildung 7.44 dargestellt.

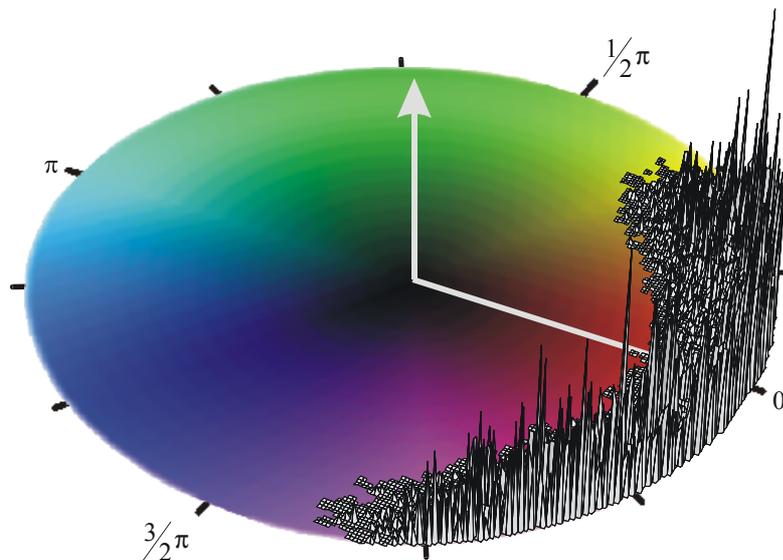


Abb. 7.44: Häufigkeitsverteilung der Farben eines informationstragenden Bereiches

Die allgemeine Struktur der unscharfen Mengen zur Farbbeschreibung lässt sich in dieser Häufigkeitsverteilung wiederfinden. Es existiert ein Bereich aus häufig gemessenen Farben sowie die von diesem Bereich ausgehenden fallenden Flanken, die eine verminderte Häufigkeit der entsprechenden Farben darstellen. Zusätzlich existiert ein großer Abschnitt, der Farben repräsentiert, die in dem informationstragenden Bereich nicht vorkommen. Diese Abschnitte sollen nun durch Einsatz einer asymmetrischen Evolutionsstrategie in eine unscharfe Farbbeschreibung umgewandelt werden.

7.4-6 Repräsentation

Die einfachste Definition einer unscharfen Farbbeschreibung besteht aus den acht Eckpunkten des projizierten Pyramidenstumpfs. Unscharfe Farbbeschreibungen sind stets normal¹, d. h. die z -Koordinate kann vernachlässigt werden. Somit ergeben sich für die einfache Darstellung acht Zylinderkoordinatenpaare, die je aus einem Winkel und einem Radius bestehen. Die Bezeichnung der Koordinatenpaare kann der Abbildung 7.43 entnommen werden. In dieser Notation bezeichnen die Buchstaben **C** und **S** Koordinatenpaare, die entweder den Kern (engl. Core) oder den Träger (engl. Support) bezeichnen. Die Indizes **i** und **o** geben die Position innen (engl. inner) und außen (engl. outer) an. Die Indizes **a** und **b** dienen lediglich einer fortlaufenden Kennzeichnung. Eine unscharfe Menge F , die eine Farbe definiert, kann daher wie folgt beschrieben werden:

$$F = (\mathbf{C}_{oa}, \mathbf{C}_{ob}, \mathbf{C}_{ia}, \mathbf{C}_{ib}, \mathbf{S}_{oa}, \mathbf{S}_{ob}, \mathbf{S}_{ia}, \mathbf{S}_{ib}) \quad (7.46)$$

Jedes Koordinatenpaar besteht wiederum aus zwei reellen Zahlen, die die Position bestimmen. Als

1. Bei normalen, unscharfen Mengen gilt, dass die Höhe (definiert als das Supremum der Zugehörigkeitsgrade) gleich eins ist. Für die hier verwendeten unscharfen Farbbeschreibungen gilt, dass die z -Koordinaten des Trägers gleich null, die z -Koordinaten des Kerns gleich eins sind.

Beispiel sei hier das Koordinatenpaar \mathbf{C}_{oa} angegeben. Es gilt: $\mathbf{C}_{oa} = (\alpha_{\mathbf{C}_{oa}}, r_{\mathbf{C}_{oa}})$. Die Objektkomponente besteht damit aus 16 reellen Zahlen, d. h. die Dimensionalität dieses Optimierproblems ist 16.

Zur Optimierung der unscharfen Mengen werden zwei verschiedene Evolutionsstrategien eingesetzt, eine Evolutionsstrategie mit Adaptation von n Schrittweiten sowie die asymmetrische Evolutionsstrategie. Die Repräsentation eines Individuums für die erste Variante sieht somit wie folgt aus:

$$I = (\mathbf{x}, \boldsymbol{\sigma}), \text{ mit} \quad (7.47)$$

$$\mathbf{x} = (\alpha_{\mathbf{C}_{oa}}, r_{\mathbf{C}_{oa}}, \alpha_{\mathbf{C}_{ob}}, r_{\mathbf{C}_{ob}}, \dots, \alpha_{\mathbf{S}_{ib}}, r_{\mathbf{S}_{ib}}) \text{ und} \quad (7.48)$$

$$\boldsymbol{\sigma} = (\sigma_{\mathbf{C}_{oa}}, \sigma_{\mathbf{C}_{oa}}, \sigma_{\mathbf{C}_{ob}}, \sigma_{\mathbf{C}_{ob}}, \dots, \sigma_{\mathbf{S}_{ib}}, \sigma_{\mathbf{S}_{ib}}) \quad (7.49)$$

Für die asymmetrische Evolutionsstrategie gilt die folgende Repräsentation:

$$I = (\mathbf{x}, \boldsymbol{\sigma}, \mathbf{c}), \text{ mit} \quad (7.50)$$

$$\mathbf{x} = (\alpha_{\mathbf{C}_{oa}}, r_{\mathbf{C}_{oa}}, \alpha_{\mathbf{C}_{ob}}, r_{\mathbf{C}_{ob}}, \dots, \alpha_{\mathbf{S}_{ib}}, r_{\mathbf{S}_{ib}}) \quad (7.51)$$

$$\boldsymbol{\sigma} = (\sigma_{\mathbf{C}_{oa}}, \sigma_{\mathbf{C}_{oa}}, \sigma_{\mathbf{C}_{ob}}, \sigma_{\mathbf{C}_{ob}}, \dots, \sigma_{\mathbf{S}_{ib}}, \sigma_{\mathbf{S}_{ib}}) \quad (7.52)$$

$$\mathbf{c} = (c_{\mathbf{C}_{oa}}, c_{\mathbf{C}_{oa}}, c_{\mathbf{C}_{ob}}, c_{\mathbf{C}_{ob}}, \dots, c_{\mathbf{S}_{ib}}, c_{\mathbf{S}_{ib}}) \quad (7.53)$$

7.4-7 Fitnessfunktion

Die Qualität eines Individuums wird durch die Fitnessfunktion bestimmt. In diesem Anwendungsbeispiel wird hierzu die Summe der Fehlerquadrate zwischen der Häufigkeitsverteilung und der durch die Eckkoordinatenpaare bestimmten unscharfen Menge bestimmt. Im Idealfall ist diese Summe 0. Dieser Wert kann in der Regel nicht erreicht werden, da die hier gewählte Form der zylindrischen Pyramidenstümpfe nur eine grobe Annäherung an beliebige Häufigkeitsverteilungen bietet. Trotzdem ist so das (theoretisch) erreichbare Optimum bekannt. Die Fitnessfunktion kann wie folgt beschrieben werden:

$$\Psi(\mathbf{I}) = \Psi(\mathbf{x}) = \sum_{i_{\alpha} = \alpha_{\mathbf{S}_{oa}}}^{\alpha_{\mathbf{S}_{ob}}} \left(\sum_{i_r = r_{\mathbf{S}_{oa}}}^{r_{\mathbf{S}_{ia}}} (H_F(i_{\alpha}, i_r) - \mu_F(i_{\alpha}, i_r))^2 \right) \quad (7.54)$$

$H_F(i_{\alpha}, i_r)$ stellt hierbei die Häufigkeit der Farbe an der Stelle (i_{α}, i_r) dar, $\mu_F(i_{\alpha}, i_r)$ den Wert der entsprechenden Zugehörigkeitsfunktion. Die Berechnung des Wertes $\mu_F(\alpha, r)$ basiert auf einer Fallunterscheidung, die alle sechs möglichen Seitenflächen des zylindrischen Pyramidenstumpfes einschließt. Die vollständige Berechnung findet sich in [Hil00a]. Für die Komponenten der Objektkoordinate gilt die Bezeichnung nach (7.48) und (7.51).

Alle anderen Parameter der verwendeten Evolutionsstrategien werden entsprechend den Empfehlungen aus den Kapiteln 2 (Symmetrische Evolutionsstrategien, Seite 13) und 6 (Asymmetrische Evolutionsstrategien, Seite 167) belegt.

7.4-8 Resultate

Die folgenden Diagramme zeigen die Resultate, die bei zwei unterschiedlichen Testfällen durch die beiden Varianten der Evolutionsstrategien erreicht wurden. Die beiden Testfälle unterscheiden sich in der Initialisierung. In dem ersten Fall wurden die Objektkomponenten der Individuen mit Werten belegt, die sich im Träger der Häufigkeitsverteilung befinden, d.h.

$$H_F(\alpha, r) \neq 0 \quad (7.55)$$

gilt für alle Koordinatenpaare der Objektkomponenten der Startpopulation. Bei dem zweiten Testfall wurden die Objektkomponenten bewusst in den Co-Träger der Häufigkeitsverteilung gelegt, d. h.

$$H_F(\alpha, r) = 0 \quad (7.56)$$

gilt für alle Koordinatenpaare der Objektkomponenten der Startpopulation. Diese Belegung simuliert eine Fehlinitialisierung entsprechend den fehlinitialisierten Experimenten aus den Abschnitten 7.1-1 (f_1 - Kugelmodell, Seite 187) bis 7.1-6 (f_6 - steigende Doppelsumme, Seite 195). Die Resultate zu diesen Testfällen sind in der Abbildung 7.45 für die Evolutionsstrategie mit Adaptation von n Schrittweiten und für die asymmetrische Evolutionsstrategie dargestellt.

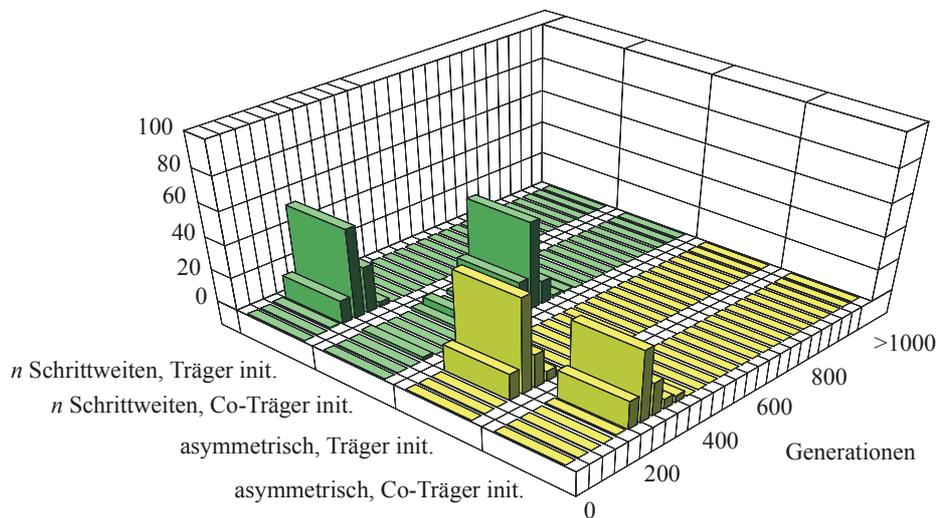


Abb. 7.45: Experimentelle Resultate der unscharfen Farbverarbeitung

Die vier Testfälle wurden jeweils 100 mal durchgeführt. Es gilt folgendes Abbruchkriterium:

$$Q_{stop} = \frac{180}{\pi}(\alpha_{C_{ob}} - \alpha_{C_{oa}})(r_{S_{oa}} - r_{S_{ia}})100q_m^2 \quad (7.57)$$

Bei der Festlegung des Abbruchkriteriums wurde folgende Annahme zugrunde gelegt: Der H -Anteil wird in einer Auflösung von 1° gemessen, der S bzw. I -Anteil in einer Auflösung von 0.01 . Dadurch gibt es $(180/\pi)(\alpha_{C_{ob}} - \alpha_{C_{oa}})(r_{S_{oa}} - r_{S_{ia}})100$ Werte in der Häufigkeitsverteilung. Mit dem Faktor q_m kann nun die tolerierbare mittlere Differenz von Häufigkeitsverteilung und unscharfer Menge eingestellt werden. In allen Testfällen wurde der Wert q_m auf 0.2 gesetzt.

7.4-9 Bewertung der Resultate

Dieser Anwendungsfall demonstriert den Vorteil der asymmetrischen Evolutionsstrategien gegenüber der Evolutionsstrategie mit Adaptation von n Schrittweiten für den Fall einer ungünstigen Initialisierung. Letztgenannte Variante ist zwar in der Lage, die unscharfen Mengen in erforderlicher Genauigkeit anzunähern, benötigt jedoch im Falle einer ungünstigen Initialisierung deutlich mehr Generationen als die asymmetrische Evolutionsstrategie.

7.5 Unscharfe Inferenzsysteme als Funktionsapproximatoren

Unscharfe Inferenzsysteme können eingesetzt werden, um beliebige Funktionen zu approximieren. Das hier vorgestellte Beispiel stammt aus einem Anwendungsbereich, der im Rahmen des Sonderforschungsbereichs 316 – Herstellung, Be- und Verarbeitung sowie Prüfung von metallischen und metall-keramischen Verbundwerkstoffen – der DFG an der Universität Dortmund untersucht wurde. Bei dieser Untersuchung wurde der Einsatz von unscharfen Inferenzsystemen zur Modellierung und Verarbeitung von domänenspezifischem Wissen und Fakten überprüft. Als Resultat dieser Untersuchung wurde festgestellt, dass sich unscharfe Inferenzsysteme eignen, wenn ihr Einsatz durch Werkzeuge unterstützt wird, die eine automatische Transformation von numerischen Zusammenhängen in linguistisch repräsentierte Zusammenhänge ermöglichen. Dann ist eine gleichzeitige Verarbeitung von exaktem und unscharfem Wissen möglich [Fat95].

7.5-1 Problembeschreibung

Zu den Werkzeugen, die in diesem Sonderforschungsbereich entwickelt wurden, gehört ein Transformationssystem, das Funktionen, die entweder als Gleichungen, Tabellen oder Diagramme gege-

ben sind, in unscharfe Inferenzsysteme umformt. Ein Beispiel für eine solche Umformung ist in der folgenden Abbildung gegeben:

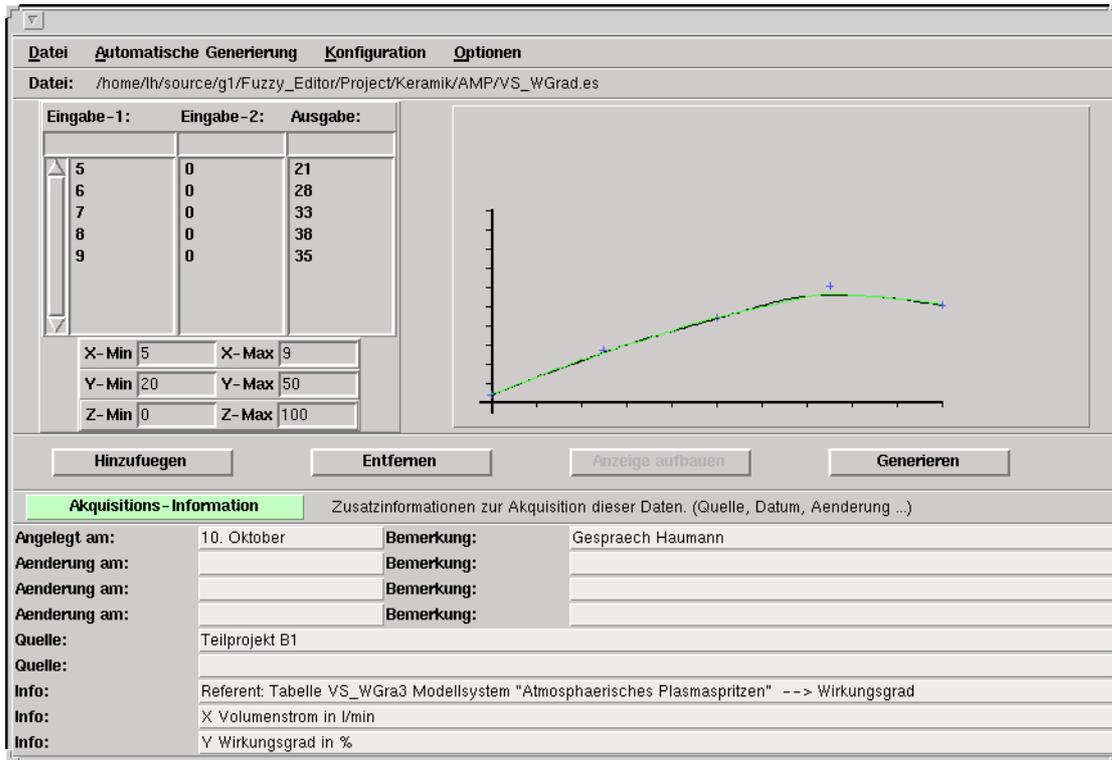


Abb. 7.46: Transformationssystem für unscharfe Inferenzsysteme

Im linken, oberen Bereich ist die Tabelle der akquirierten Informationen zu finden, im rechten, oberen Bereich das resultierende Diagramm. Der gesamte untere Bereich enthält ausschließlich Akquisitionsinformationen, die hier nicht relevant sind. Das Problem bestand in der Transformation dieses Diagramms in ein unscharfes Inferenzsystem. Ein unscharfes Inferenzsystem besteht aus fünf Teilkomponenten, die in Abbildung 7.47 dargestellt sind.

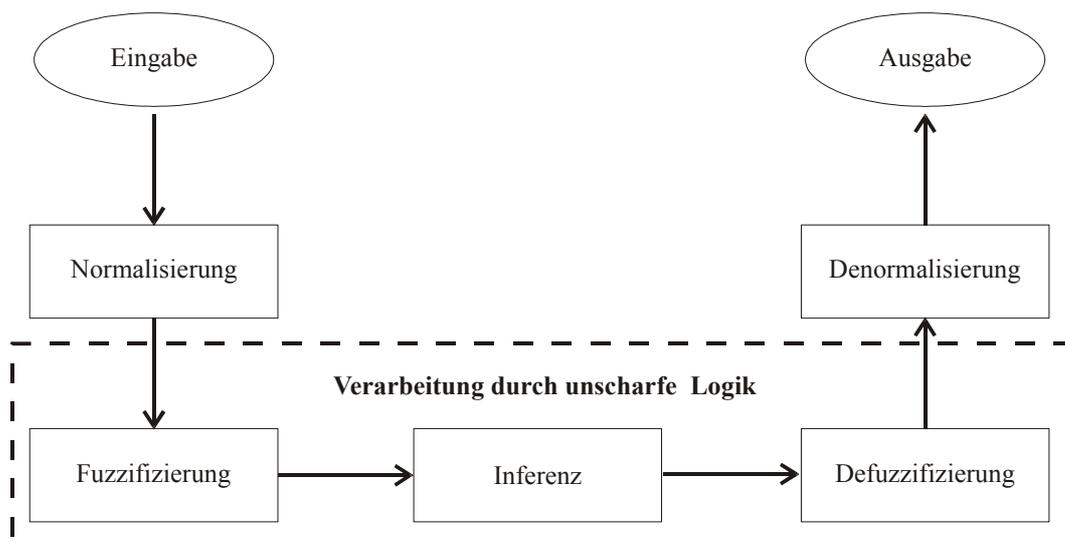


Abb. 7.47: Struktur eines unscharfen regelbasierten Systems

Die Bedeutung der einzelnen Komponenten ist wie folgt:

- Normalisierung: Dieses Modul soll Eingangswerte, die aus einem beliebigen Wertebereich stammen können, auf einen einheitlichen Wertebereich transformieren. Der Einsatz eines solchen Moduls ist optional. In der Regel wird es nur bei Systemen benötigt, die einen Teil des Wissens durch automatische Werkzeuge erstellt haben und hierbei von einem vordefinierten Wertebereich ausgehen.
- Fuzzifizierung: Anhand der Zugehörigkeitsfunktionen der Eingangsvariablen muß die Zugehörigkeit des Eingabewertes in Bezug auf die linguistischen Terme ermittelt werden.
- Inferenz: Der Inferenzmechanismus ist das eigentliche Rechenwerk eines Expertensystems. Mittels ihm kann aus bekanntem Wissen und Fakten neues Wissen abgeleitet werden. Eng verknüpft mit dem Inferenzmechanismus ist die Regelbasis, die das bekannte Wissen in Form von z. B. Produktionensystemen enthält.
- Defuzzifizierung: Neues Wissen, das mittels des Inferenzmechanismus abgeleitet worden ist, wird durch dieses Modul weiterverarbeitet, um als Endresultat die Belegung der Ausgangsvariablen zu liefern. Grundlage hierfür ist z. B. die Schwerpunkt- oder die Maximummethode.
- Denormalisierung: Falls durch ein Normalisierungsmodul der Definitionsbereich der Eingabe verändert wurde, muß eine Rücktransformation der ermittelten Werte in diesem Modul vorgenommen werden.

Im Gegensatz zu traditionellen Expertensystemen kommen bei unscharfen Expertensystemen fast nur vorwärtsverkettende Inferenzmaschinen zum Einsatz. Grund hierfür ist die Tatsache, daß das Ergebnis eines einzelnen Inferenzschrittes auf unendlich viele Belegungen der Eingangswerte zurückgeführt werden kann. Anders als bei boolescher Logik ist es damit nicht möglich, den Suchraum über den Eingangswerten sinnvoll auf ein berechenbares Maß zu reduzieren. Es gibt jedoch auch hybride Systeme, bei denen scharfe und unscharfe Logik kombiniert wird. Bei diesen Systemen ist unter bestimmten Bedingungen eine rückwärtsverkettenden Inferenz möglich [Die94].

Das Ein/Ausgabeverhalten eines unscharfen Inferenzsystems wird durch die Regelbasis und durch die Zugehörigkeitsfunktionen der Ein- bzw. Ausgabevariablen bestimmt. Abbildung 7.48 zeigt eine graphische Darstellung eines solchen Inferenzsystems für den eindimensionalen Fall. Es besteht aus

zwei linguistischen Variablen Eingang und Ausgang sowie pro Variablen fünf linguistischen Termen. Die Auswertung eines solchen Inferenzsystems resultiert dann in einer Kurve.

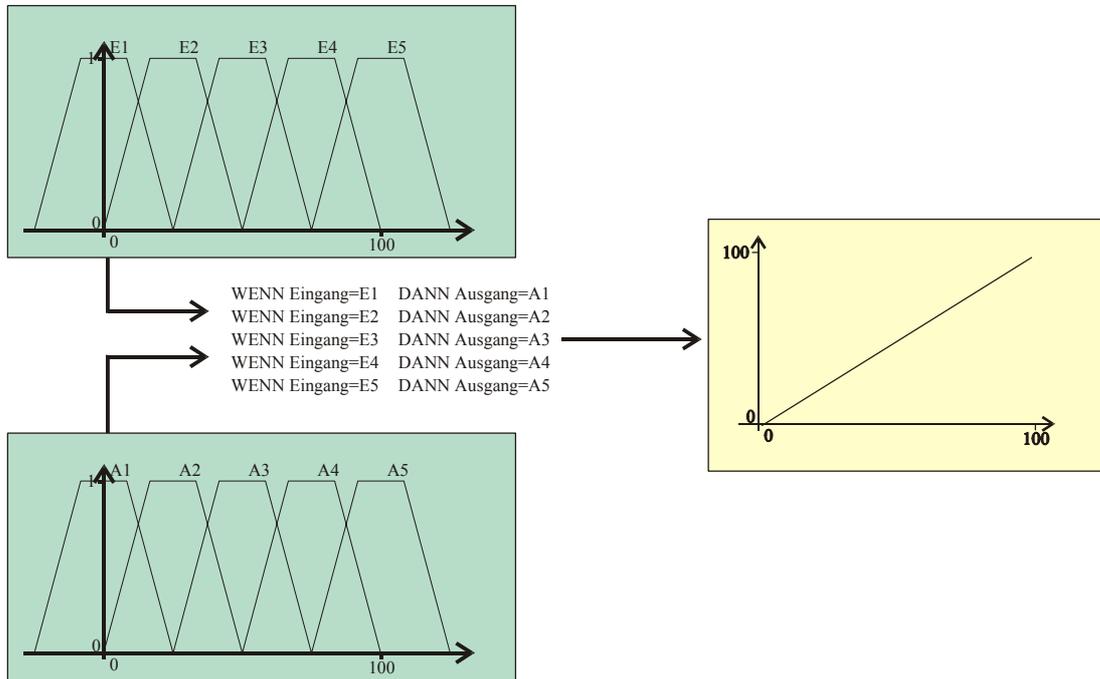


Abb. 7.48: Unscharfes Inferenzsystem mit einer Eingangs- und einer Ausgangsvariablen

7.5-2 Repräsentation

Ein unscharfes Inferenzsystem besteht aus einer Regelbasis und der Beschreibung der unscharfen Mengen. Die Kodierung der unscharfen Mengen ist einfach: Für jede Menge werden die Koordinaten der vier Eckpunkte gespeichert, bei normalen unscharfen Mengen reicht es aus, die x -Koordinaten zu speichern. Die unscharfe Menge E1 kann dann durch (x_1, x_2, x_3, x_4) repräsentiert werden mit x_1 bis x_4 nach Abbildung 7.49.

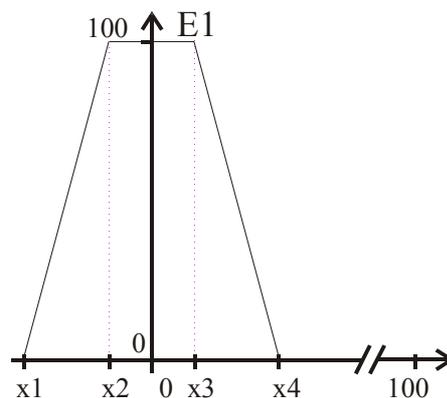


Abb. 7.49: Kodierung einer unscharfen Menge

Ein Vektor, der alle unscharfen Mengen eines unscharfen Inferenzsystems kodiert, hat dann die Form:

$$\left(\underbrace{x_1 x_2 x_3 x_4}_{E1} \cdots \underbrace{x_1 x_2 x_3 x_4}_{E5} \underbrace{x_1 x_2 x_3 x_4}_{A1} \cdots \underbrace{x_1 x_2 x_3 x_4}_{A5} \right) \quad (7.58)$$

Die Anzahl n der Komponenten in der Objektvariablen lässt sich wie folgt bestimmen:

$$n = 4 \cdot n_{LT} \cdot n_{LV} \quad (7.59)$$

Die Variablen n_{LT} bezeichnet die Anzahl der linguistischen Terme und n_{LV} die Anzahl der linguistischen Variablen. Für das oben beschriebenen Beispiel gilt damit $n = 40$, d. h. es handelt sich um ein 40-dimensionales Optimierproblem. Die Dimensionalität kann selbst bei einfachen Inferenzsystemen stark steigen. So resultiert die Optimierung des Ein-/Ausgabeverhaltens eines Inferenzsystems mit 3 Eingangs- und einer Ausgangsvariablen, jeweils mit einer Granularität von 7 linguistischen Termen einem 112-dimensionalen Optimierproblem. Für verkettete Inferenzsysteme, bei denen die Ausgangsvariablen der Regelgruppen die Eingangsvariablen nachfolgender Regelgruppen sein können, kann die Dimensionalität auf weit über 1000 steigen.

Nachdem die Kodierung der unscharfen Mengen erläutert ist, muß nun die Kodierung der Regeln vorgenommen werden. Hierzu gibt es eine elegante Lösung: Verzicht auf explizite Kodierung der Regeln, statt dessen eine implizite, im Inferenzmechanismus festgelegte Kodierung. Dieses bedeutet, daß die folgende Zuordnung von Eingangstermen zu Ausgangstermen vorgegeben ist:

$$\begin{aligned} E1 &\rightarrow A1 \\ E2 &\rightarrow A2 \\ E3 &\rightarrow A3 \\ E4 &\rightarrow A4 \\ E5 &\rightarrow A5 \end{aligned} \quad (7.60)$$

Diese Zuordnung der Terme ist fixiert. Im Laufe einer Optimierung kann sich jedoch die Semantik, d. h. die Anordnung der linguistischen Terme, innerhalb des Universums ändern. Aus diesem Grund werden während einer Optimierung die linguistischen Terme nicht mit Bezeichnern wie klein oder groß gekennzeichnet, sondern nur durch Nummern. Erst nach der Optimierung werden, mit Methoden aus dem Bereich der Ordnungsrelationen auf unscharfen Mengen, die Terme mit linguistischen Bezeichnern belegt. Als Ordnungsrelationen bzw. Anordnungsverfahren bieten sich die Methoden von Adamo [Ada80], Baas und Kwakernaak [Baa77], Bortolan und Degani [Bor85], Kerre [Ker82] sowie die Verfahren nach Yager [Yag78, Yag81] an.

7.5-3 Fitnessfunktion

In diesem Anwendungsbeispiel wird zur Bestimmung der Fitnessfunktion die Summe der Fehlerquadrate zwischen der gewünschten Kennlinie und der durch das unscharfe Inferenzsystem erzeugten Kennlinie bestimmt. Im Idealfall ist diese Summe 0. Analog zu dem Anwendungsbeispiel aus dem Abschnitt 7.4 (Optimierung von unscharfen Mengen zur Beschreibung von Farben, Seite 213) kann dieser Wert in der Regel nicht erreicht werden. Unscharfe Inferenzsysteme haben einen inter-

polierenden Charakter und können nur eine durch die Anzahl der linguistischen Terme begrenzte Anzahl von Knickstellen enthalten. Die Fitnessfunktion kann wie folgt beschrieben werden:

$$\Psi(\mathbf{I}) = \Psi(\mathbf{x}) = \sum_{i=0}^{100} (F_S(i) - F_I(i))^2 \quad (7.61)$$

Die Funktion $F_S(i)$ bezeichnet hier den Sollwert, d. h. die gewünschte Kennlinie. Die Funktion $F_I(i)$ stellt den Istwert, also die Kennlinie des aktuellen Individuums, dar.

7.5-4 Resultate

Als Testfunktionen wurden zwei Gruppen von Kennlinien ausgesucht, die im folgenden mit Linear bzw. Linear-invers und Sprung bzw. Sprung-invers gekennzeichnet sind. Die Testbeispiele Linear und Sprung sind dabei der Initialisierung der Objektkomponenten ähnlich, da sie dem generellen Verlauf der Initialisierung entsprechen, d. h. bei kleinen Eingabewerten ist die Ausgabe klein, mit steigenden Eingabewerten steigen auch die Ausgabewerte bzw. bleiben gleich. Die Initialisierung der Objektkomponenten ist in der Abbildung 7.50 dargestellt.

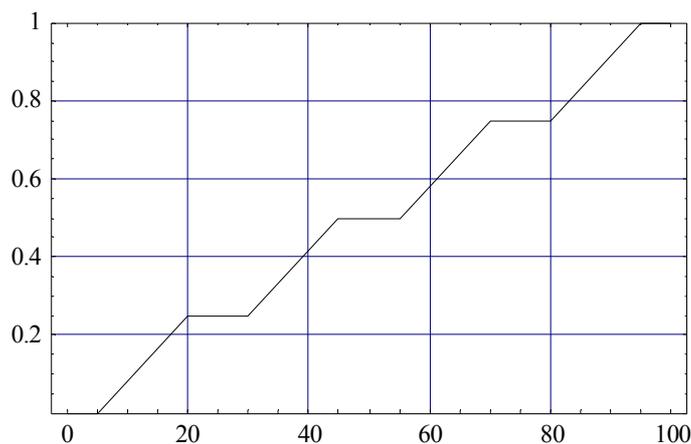


Abb. 7.50: Initialisierung der Objektkomponente

Die Testbeispiele Linear-invers und Sprung-invers erfordern im Gegensatz zu den erst genannten Beispielen eine aufwendige Veränderung der initialen unscharfen Mengen. Da der Kennlinienverlauf der initialen Generation nahezu umgekehrt werden muss, müssen entweder die unscharfen Mengen der Eingangsvariablen oder die unscharfen Mengen der Ausgangsvariablen symmetrisch zu der mittleren unscharfen Menge ihre Position tauschen. Dieser Tausch darf jedoch nur bei einer der beiden Variablen geschehen, da ein gleichzeitiger Tausch der Eingangs- und Ausgangsterme keine Umkehr des Kennlinienverlaufs bewirkt.

Die folgenden beiden Diagramme zeigen die Sollkennlinie, die durch ein unscharfes Inferenzsystem dargestellt werden sollen. Die blau dargestellte Kennlinie entspricht den Testfällen Linear bzw. Sprung, die rot dargestellte Kennlinie den Testfällen der inversen Kennlinien.

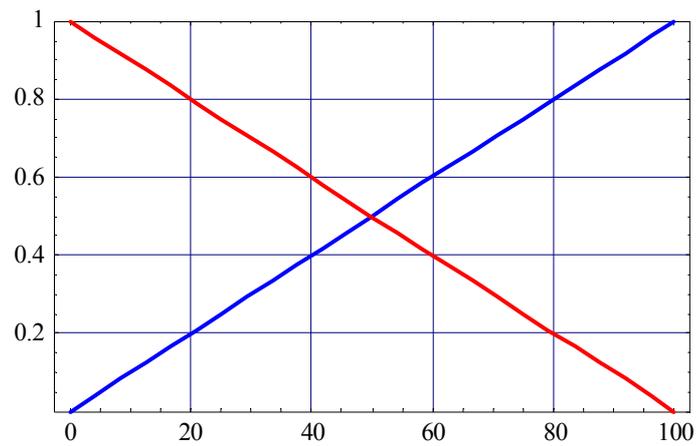


Abb. 7.51: Sollkennlinien Linear und Linear-invers

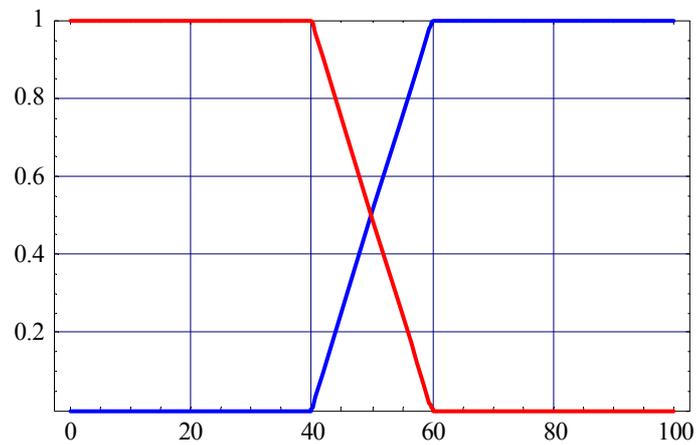


Abb. 7.52: Sollkennlinien Sprung und Sprung-invers

Für die Beispiele Linear und Linear-invers wurden folgende Parameter für die beiden Varianten der Evolutionsstrategie gewählt: $\mu = 15$, $\lambda = 100$ und $Q_{stop} = 1.0$. Die Abbildung 7.53 stellt die Resultate in der üblichen Darstellung dar.

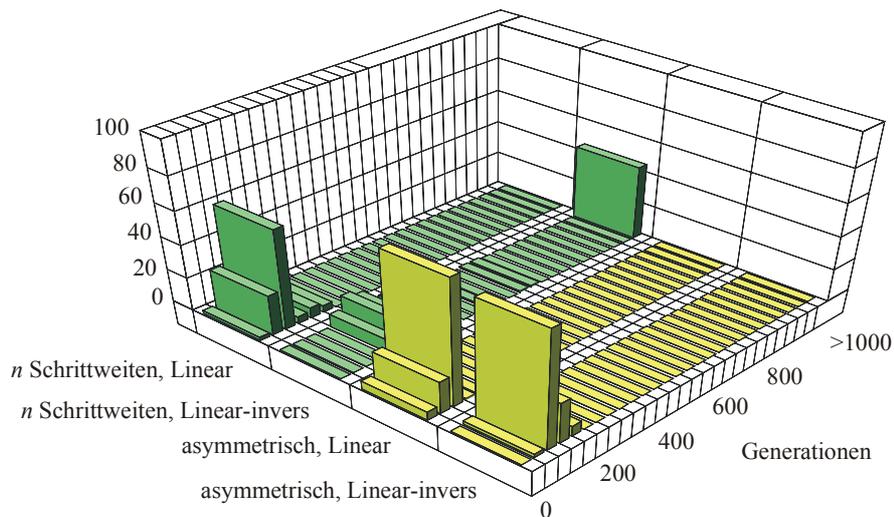


Abb. 7.53: Experimentelle Resultate der Kennlinien Linear und Linear-invers

Die Abbildung 7.54 zeigt die Resultate zu den Kennlinien Sprung und Sprung-invers:

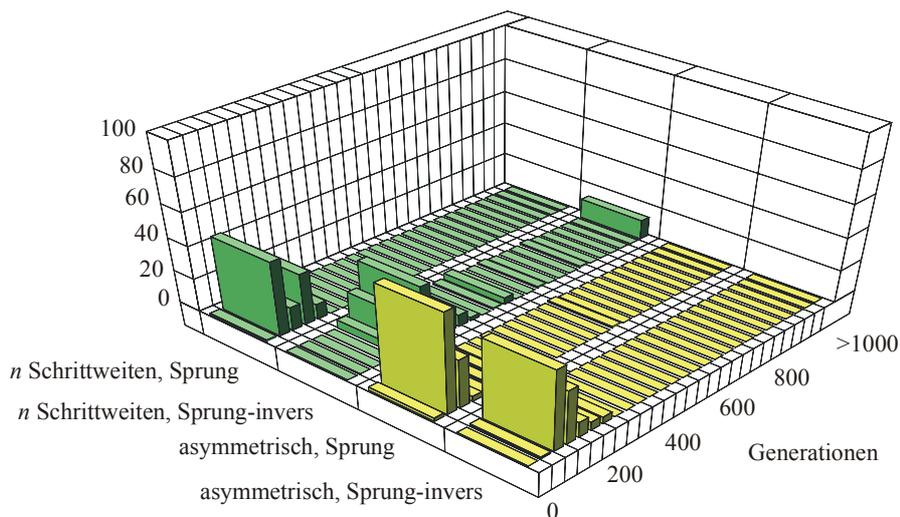


Abb. 7.54: Experimentelle Resultate der Kennlinien Sprung und Sprung-invers

7.5-5 Bewertung der Resultate

Das Beispiel unscharfe Inferenzsysteme zeigt sehr deutlich, wie robust die asymmetrische Evolutionsstrategie gegenüber nicht idealen Initialisierungen ist. Bei einer Initialisierung, deren Kennlinie dem gewünschten Verlauf ähnlich ist, zeigen sowohl die Evolutionsstrategie mit Adaptation von n Schrittweiten als auch die asymmetrische Evolutionsstrategie ein nahezu identisches Verhalten. Deutliche Unterschiede zeigen sich jedoch bei den inversen Testbeispielen. Während die asymmetrische Evolutionsstrategie unwesentlich mehr Generationen benötigt, um die Sollkennlinie zu erreichen, besteht bei der Evolutionsstrategie mit Adaptation von n Schrittweiten sogar die Gefahr, dass sie in dem gegebenen Generationsrahmen die Sollkennlinie nicht mit der erforderlichen Genauigkeit annähern kann. Grundsätzlich braucht sie bei den inversen Testbeispielen deutlich mehr Generatio-

nen, um das Optimum zu erreichen.

Der hier geschilderte Fall der inversen Testbeispiele ist im Bereich der Optimierung unscharfer Inferenzsysteme sehr häufig anzutreffen. Da diese Systeme, besonders im Fall verketteter Inferenzsysteme, äußerst komplex sind, kann eine möglichst ideale Initialisierung selten gegeben werden. Sollte die Anzahl der Eingangsvariablen eines unscharfen Inferenzsystems größer als drei sein, entfallen nahezu alle Visualisierungsmöglichkeiten. Daher ist auch eine Überprüfung der Initialisierung durch einen Anwender bzw. die Vorgabe idealer Initialisierungen kaum möglich.

7.6 Optimierung von Schraubenrotorprofilen

Der letzte Abschnitt dieses Kapitels beschreibt den Einsatz asymmetrischer Evolutionsstrategien zur Optimierung von Schraubenrotorprofilen. Die Bearbeitung dieses Themas findet im Rahmen des DFG-Projektes Ka 663/23-1, *Methode zur Auslegung trockenlaufender Rotationsverdrängermaschinen* statt. Dieses Forschungsprojekt ist das erste Projekt, bei dem die asymmetrische Evolutionsstrategie externen Arbeitsgruppen als Optimierverfahren zur Verfügung gestellt wurde. Hier lässt sich daher auch der Arbeitsaufwand abschätzen, der mit der Integration der asymmetrischen Evolutionsstrategie in ein existierendes Programmsystem verbunden ist.

7.6-1 Problembeschreibung

Das Ziel des Forschungsprojektes ist die Entwicklung einer Methode, mit der Schraubenprofile optimal an ihr Einsatzgebiet angepasst werden können. Bei Maschinen, in die Schraubenrotoren eingebaut werden, handelt es sich z. B. um Schraubenverdichter. Schraubenverdichter werden in der Industrie zum Erzeugen von Druckunterschieden für unterschiedliche Medien, den sogenannten Arbeitsfluiden, eingesetzt. Ein großer Bereich ist dabei die Erzeugung von Druckluft zum Antrieb von Maschinen. Ein anderer Bereich ist das Fördern unterschiedlicher Güter. Dabei wird das Fördergut mit Luft vermischt und durch Schläuche oder Rohre transportiert. Die Palette möglicher Fördergüter reicht hierbei von Baustoffen wie Zement und Kalk, bis hin zu Lebensmitteln wie Mehl oder Zucker. Zum Teil werden bereits LKW mit Verdichtern ausgerüstet, die eine Be- bzw. Entladeleistung von einer Tonne pro Minute erreichen.

Schraubenverdichter gehören zur Gruppe der mehrwelligen Verdrängermaschinen. Das gemeinsame Prinzip bei allen Typen von Verdrängermaschinen ist ein gekapselter Arbeitsraum, in dem das

Arbeitsfluid komprimiert wird. Abbildung 7.55 zeigt ein Beispiel für einen solchen Schraubenverdichter [Lel88]. Einer der Schraubenrotoren ist hier farblich hervorgehoben.

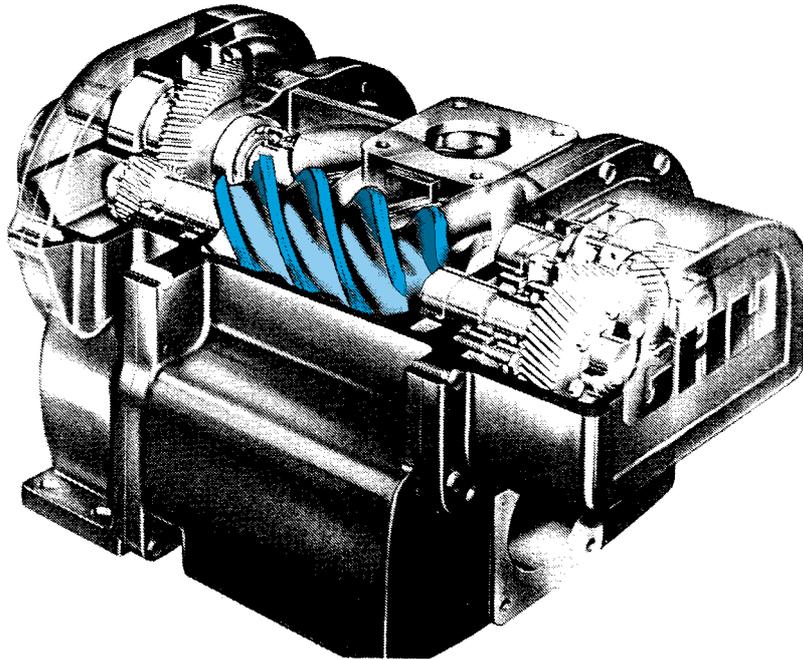


Abb. 7.55: Schraubenverdichter

Deutlich ist in der Mitte das Schraubenrotorenpaar zu sehen. Dieses Rotorenpaar besteht aus zwei schrägverzahnten Zahnrädern, die ineinander greifen. Abbildung 7.56 zeigt diese Verzahnung schematisch. Das Gehäuse des Verdichters muß das Rotorenpaar eng umschließen, um die zwischen Rotoren und Gehäuse befindlichen Arbeitsräume abzudichten [Dre90].

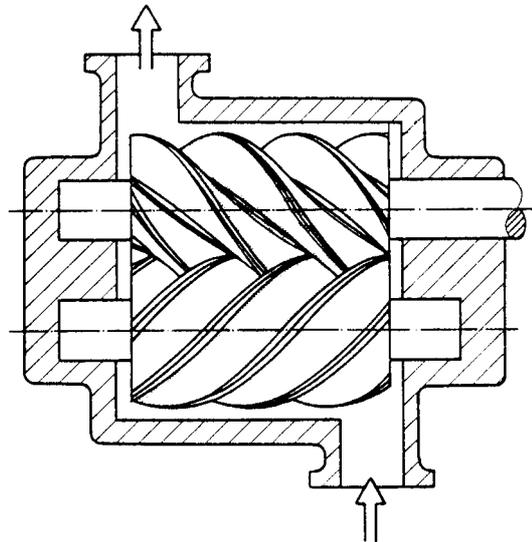


Abb. 7.56: Verzahnung von Haupt- und Nebenläufer

7.6-2 Repräsentation

Zur Repräsentation eines Schraubenrotorpaars muss lediglich der Hauptrotor in Form eines Individuums kodiert werden. Der zugehörige Nebenrotor kann passend berechnet werden. Die Kodierung

verwendet b -Spline Kurven, eine vollständige Beschreibung findet sich in [Kau00]. Ein Beispiel eines Schraubenrotorpaars ist in der folgenden Abbildung gegeben.

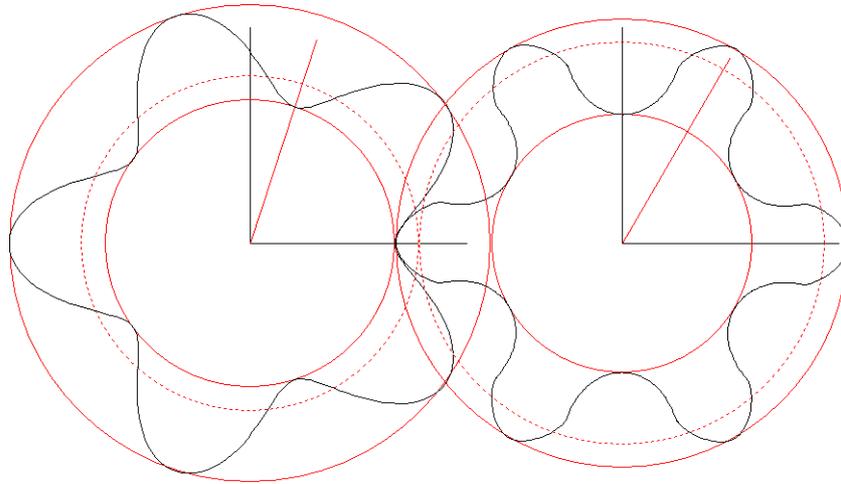


Abb. 7.57: Schraubenrotorpaar

7.6-3 Resultate

Der Einsatz einer asymmetrischen Evolutionsstrategie, die eine symmetrische Evolutionsstrategie mit Adaptation von n Schrittweiten ablöst, hat auf die Anzahl der benötigten Generationen beeindruckende Auswirkungen. Für das Optimierkriterium *Maximierung der Schöpffläche* konnten bereits nach 800 Generationen Werte erreicht werden, die bei Einsatz der symmetrischen Evolutionsstrategie erst nach 20.000 Generationen vorlagen. Da die Berechnung der Gegenrotoren sowie die Berechnung der Fitnessfunktion rechenintensiv sind, bedeutet dies eine Reduktion der Laufzeiten von mehreren Tagen auf wenige Stunden. Für weitere Kriterien, wie *Minimierung der polaren Flächenträgheitsmomente* oder *Minimierung der Schöpffläche* konnten ähnlich reduzierte Laufzeiten ermittelt werden. Eine vollständige Übersicht über die Resultate findet sich in [Kau00].

7.7 Zusammenfassung

Die asymmetrische Evolutionsstrategie zeigt sowohl für Testfunktionen, als auch für reale Anwendungen ein hervorragendes Verhalten. Bei den Testfunktionen benötigt sie, mit Ausnahme des einfachen Kugelmodells und der einfachen Doppelsumme, entweder die geringste Anzahl an Generationen oder unwesentlich mehr als die beste andere Variante. Für alle Fälle der Fehlinitialisierung findet sie das Optimum in der kürzesten Zeit. Bei den realen Anwendungen führt der Einsatz einer asymmetrischen Evolutionsstrategie zu einem zuverlässigen und robustem Laufzeitverhalten. Dadurch, dass die asymmetrische Evolutionsstrategie keine Probleme mit einer ungünstigen Initialisierung der Startpopulation hat, kann sie ohne Anpassung auf verschiedenartige Problemstellung angewandt werden. Da sie zudem über keine externen Steuerparameter verfügt kann sie als universelles Optimierverfahren angesehen werden, das allgemein eingesetzt werden kann.

Kapitel 8 Zusammenfassung & Ausblick

*Doch der Gedanke an diesen einsamen Computer, vierzig Meter unter der Erdoberfläche in einer texanischen Wüste vergraben, ausgestattet mit der Erfahrung eines Korporals und der mathematischen Begabung eines Einstein und Kepler zusammen, Jahr für Jahr über Strindbergs Mächten brütend und immer wieder das Drama nachvollziehend, das sich in den neunziger Jahren des vorigen Jahrhunderts im Hotel Orfila in Paris abspielte, das ganze Drama, in all seinen Varianten, mit allen Alternativdramen und möglichen Verwicklungen – dieser Gedanke verursacht mir manchmal einen Anflug von schlechtem Gewissen.
(Lars Gustafsson, letzter Satz der Tennisspieler)*

8.1 Zusammenfassung

Die Entwicklung und Realisierung der asymmetrischen Evolutionsstrategie stellt eine weitreichende Weiterentwicklung der Evolutionsstrategien und der evolutionären Algorithmen dar. In dieser Arbeit ist zum ersten Mal die Möglichkeit einer Richtungsadaptation sowie die Auswirkung dieser neuen Adaptationsart systematisch untersucht worden. Alle experimentellen Resultate, sowohl für Testfunktionen als auch für reale Anwendungen haben gezeigt, dass die Kombination von ungerichteter Mutation und anschließender Selektion (wie sie bei den symmetrischen Varianten durchgeführt wird) einer gerichteten Mutation mit anschließender Selektion unterlegen ist. Die allgemeine Forderung, dass eine Adaptation ausschließlich durch die Selektion erfolgen muss, verliert damit ihre Gültigkeit.

Da die für die asymmetrischen Evolutionsstrategien notwendigen Verteilungsfunktionen und Zufallszahlengeneratoren nicht existieren, ist ein Großteil der vorliegenden Arbeit diesem Thema gewidmet. In der Reihenfolge ihrer steigenden Komplexität werden in dieser Arbeit drei asymmetrische Verteilungsfunktionen entwickelt:

1. Ξ_c -Verteilung mit dem Parameter c zur Steuerung der Asymmetrie
2. $\Xi_{c,\sigma}$ -Verteilung, mit dem zusätzlichen Parameter σ zur richtungsunabhängigen Steuerung der Schrittweite
3. $\Xi_{c,\sigma}^\gamma$ -Verteilung mit einem weiteren Parameter γ zur Steuerung des initialen Verhaltens einer

asymmetrischen Evolutionsstrategie

Für alle Varianten der Ξ -Verteilungen werden in dieser Arbeit grundlegende stochastische Eigenschaften gezeigt. Hierzu zählen die Bestimmung der Verteilungs- und Dichtefunktion sowie die Herleitung der wichtigen Momente, wie Erwartungswert, Varianz und Schiefe. Da für die asymmetrischen Evolutionsstrategien Ξ -verteilte Zufallszahlen benötigt werden, wird für jede Verteilungsvariante die inverse Verteilungsfunktion bestimmt. Die Kenntnis der inversen Verteilungsfunktion ist zwingend erforderlich, da der in dieser Arbeit entwickelte Zufallszahlengenerator die Transformation der Inversen einsetzt, um Ξ -verteilte Zufallszahlen zu generieren. Um die Herleitung und Beschreibung der Ξ -Verteilungen möglichst allgemein abzuschließen, werden weitere Eigenschaften gezeigt. Hierzu zählen:

- Stetigkeit
Alle Varianten der Ξ -Verteilungen sind stetig.
- Differenzierbarkeit
Alle Varianten der Ξ -Verteilungen sind grundsätzlich einmal stetig differenzierbar. Für bestimmte Belegungen der Parameter sind sie beliebig oft stetig differenzierbar.
- Verteilung der Summe
In dieser Arbeit wird die Verteilungsfunktion der Summe von zwei Ξ -verteilten Zufallszahlen hergeleitet. Zusätzlich werden Eigenschaften der Summe gezeigt, zu denen Symmetrienerhaltung, Symmetrieerzeugung und die Strukturhaltung im Sinne der stabilen Verteilungsfunktionen gehört.
- Grenzverteilung
Als Grenzverteilung der normierte Summe Ξ -verteilter Zufallszahlen wird in dieser Arbeit die Normalverteilung bestimmt sowie ihre Gültigkeit formal bewiesen.

In einer ausführlichen Abhandlung wird in dieser Arbeit gezeigt, warum sich existierende Verteilungen nicht als Alternative zu den Ξ -Verteilungen eignen. Zu den in diesem Rahmen untersuchten Verteilungen gehören die Cauchy-Verteilung, die Fehler-Verteilung, die logistische Verteilung und die Student's t -Verteilung. Zusätzlich wird am Beispiel des Pearsonschen Verteilungssystem gezeigt, warum sich ein Großteil der Systemverteilungen, zu denen die Beta-Verteilung, die Gamma-Verteilung und die Exponentialverteilung gehören, nicht eignen.

Der effizienten Implementierung eines Zufallszahlengenerators, der Ξ -verteilte Zufallszahlen erzeugt, ist ein eigenes Kapitel gewidmet. In diesem Kapitel werden einführend wichtige Eigenschaften eines Zufallszahlengenerators erläutert. Darauf aufbauend werden unterschiedliche Methoden dargestellt, mit denen Zufallszahlen erzeugt werden können. Bedingt durch den Aufbau der Verteilungsfunktion der Ξ -Verteilungen bietet sich eine Kombination aus der Kompositionsmethode und der Transformation der Inversen an. Für den Einsatz dieser Methoden ist neben der Kenntnis der abschnittsweisen Definition der inversen Verteilungsfunktion auch die Approximation der inversen Fehlerfunktion notwendig. Eine effiziente und genaue Approximation wird in diesem Kapitel beschrieben. Als Abschluss des Implementierungskapitels wird der Algorithmus des Zufallszahlengenerators zur Erzeugung Ξ -verteilter Zufallszahlen angegeben sowie seine Eigenschaften in Bezug auf Genauigkeit, Robustheit und Effizienz bewertet.

Die bereits erwähnten Kapitel dienen der Realisierung einer asymmetrischen Evolutionsstrategie. Der essentielle Kern einer asymmetrischen Evolutionsstrategie wird durch die Gleichung (6.16) beschrieben:

$$x_{i,j}^{n(t)} = x_{i,j}^{(t)} + A_{\gamma}(c_{i,j}^{n(t)}, \sigma_{i,j}^{n(t)})$$

Diese Gleichung gibt an, wie die Objektkomponente eines Individuums in Abhängigkeit der Schrittweiten und der Schiefemaße verändert wird. In diese Gleichung geht eine $\Xi_{c, \sigma}$ -verteilte Zufallszahl in Form der Funktion $A_{\gamma}(c, \sigma)$ ein, die den in dieser Arbeit entwickelten asymmetrischen Zufallszahlengenerator verwendet.

Mit der asymmetrischen Evolutionsstrategie sollte ein Optimierverfahren geschaffen werden, das in der Lage ist sowohl Benchmarkprobleme als auch Probleme aus realen Anwendungen zu lösen. Der Schwerpunkt liegt dabei auf der Anwendbarkeit für reale Anwendungen. Die Lösung der Benchmarkprobleme soll in erster Linie einem Vergleich der asymmetrischen Evolutionsstrategie mit anderen Optimierverfahren dienen. Um diese Eigenschaften zu erfüllen, muss die asymmetrische Evolutionsstrategie über folgende Eigenschaften verfügen, die in der Einleitung ausführlich beschrieben wurden:

- geringe Anforderungen an mathematische Eigenschaften
- Robustheit
- Verzicht auf externe Steuerparameter
- lineare Steigerung des systembedingten Rechenaufwandes

Da die asymmetrische Evolutionsstrategie als großes Vorbild die Evolutionsstrategie mit Adaptation von n Schrittweiten hat, kann sie auch deren geringe Anforderungen an die mathematischen Eigenschaften übernehmen. Das dies auch für reale Anwendungen gilt, wird an den Beispielpunkten aus dem Bereich der unscharfen Logik deutlich. Sowohl bei der Optimierung unscharfer Mengen zur Farbbeschreibung als auch bei der Generierung unscharfer Inferenzsysteme sind keine Annahmen über Differenzierbarkeit, Stetigkeit oder andere Eigenschaften gemacht worden.

Die asymmetrische Evolutionsstrategie hat sich bei allen durchgeführten Experimenten als äußerst robust erwiesen. Mit Ausnahme des einfachen Kugelmodells und der einfachen Doppelsumme benötigt sie entweder die geringste Anzahl an Generationen oder unwesentlich mehr als die beste andere Variante. Für alle Fälle der Fehlinitialisierung findet sie das Optimum in der kürzesten Zeit. Zur statistischen Absicherung der experimentellen Resultate wurden für die Abschnitte 7.1 (Varianten des Kugelmodells, Seite 185) und 7.2 (Problemklassen nach Hansen und Ostermeier, Seite 197) pro Testfunktion und Variante 100 Experimente durchgeführt. In Summe bedeutet dies, dass für jede untersuchte Variante 2200 Experimente bewertet wurden. Die asymmetrische Evolutionsstrategie ist dabei die einzige Variante, die immer das Optimum finden konnte. Diese Robustheit zeigt die asymmetrische Evolutionsstrategie auch bei den realen Anwendungen.

Die asymmetrische Evolutionsstrategie benötigt keine externen Steuerparameter. Weder für die Lernraten der Schrittweiten noch für die der Schiefemaße oder den globalen Parameter γ müssen problemspezifische Belegungen gefunden werden. Alle Experimente wurden ohne Änderung der in den entsprechenden Abschnitten gegebenen Empfehlungen durchgeführt.

Der zusätzliche Aufwand, der durch die Einführung der Schiefemaße entsteht, ist gering. Die Anzahl der Strategieparameter steigt von n auf $2n$. Der Rechenaufwand zur Erzeugung asymmetrisch verteilter Zufallszahlen, in der in dieser Arbeit geforderten Form, ist gering. Die Zeitkomplexität steigt ebenfalls nur linear mit der Anzahl der Komponenten in einem Individuum.

In dieser Arbeit wurde die Forderung nach Verzicht auf externe Steuerparameter strikt eingehalten. Der Grund hierfür ist die Tatsache, dass ein Optimierverfahren, bei dem erst brauchbare Steuerparameter mittels Raten, Heuristiken oder gar Meta-Algorithmen aufwendig bestimmt werden müssen, weniger akzeptiert wird, als ein Verfahren, das ohne diese Parameter auskommt und trotzdem zuverlässig arbeitet. Wenn diese Forderung wegfällt, kann mittels optimaler Lernraten die Konvergenzgeschwindigkeit der asymmetrischen Evolutionsstrategie wahrscheinlich verbessert werden. Untersuchungen in diese Richtung, wie sie z. B. für symmetrische Evolutionsstrategien von Kursawe als Dissertation bearbeitet wurden, sind jedoch noch nicht durchgeführt worden.

Das generelle Konvergenzverhalten scheint sich nicht wesentlich von dem der Evolutionsstrategie mit Adaptation von n Schrittweiten zu unterscheiden. Diese Tatsache belegen die experimentellen Resultate, als auch die Sätze über das Grenzverhalten von Summen asymmetrisch verteilter Zufallszahlen. Bei Problemen mit ausgeprägter Richtungscharakteristik scheint die asymmetrische Evolutionsstrategie durch Selbstadaptation größere Schrittweiten in Richtung des Optimums zu adaptieren. Dies kann durch die Tatsache begründet werden, dass jedes von 0 verschiedene Schiefemaß eine Vergrößerung der Schrittweite bewirkt. Dieser Mechanismus ist wahrscheinlich auch für die schnelle Überwindung von Bereichen identischer Zielfunktionswerte verantwortlich. In der Nähe des Optimums ist die asymmetrische Evolutionsstrategie auf Grund ihrer hohen Reversibilität in der Lage, Fehladaptation schnell auszugleichen. Dieser Effekt wird durch den Einsatz einer intermediären Rekombination der Schiefemaße verstärkt. Theoretische Untersuchungen zu diesen Bereichen, die eventuell in einer allgemeinen Bestimmung der Fortschrittsrate resultieren, sind jedoch noch nicht durchgeführt worden. Ob eine solche Theorie der asymmetrischen Evolutionsstrategien in der näheren Zukunft aufgestellt werden kann ist jedoch zweifelhaft, da sogar für die wesentlich besser erforschten symmetrischen Evolutionsstrategien entsprechende Fortschrittsmaße nur für wenige Testfunktionen bestimmt werden konnten.

8.2 Ausblick

Ein Ausblick enthält in der Regel eine kurze Abhandlung über Themen und Untersuchungen, die sich an die vorliegende Arbeit anschließen können. Eine solche Übersicht ist jedoch bereits in der Zusammenfassung gegeben und bezieht sich auf eine, eventuell unmögliche Theorie der asymmetrischen Evolutionsstrategien. In dieser Arbeit ist der Ausblick statt dessen ein kleiner Blick über den Tellerrand des Informatikers.

1988 haben Cairns, Overbaugh und Miller, Forscher an der Harvard School of Public Health, über Experimente berichtet, aus denen man ableiten kann, dass bestimmte Zellen über Mechanismen zur Auswahl einzelner Mutationsvarianten verfügen. Speziell unter Stressbedingungen sollen *E. Coli* Bakterien die Häufigkeit bestimmter Mutationen über alle bekannten Maße steigern können, um so möglichst viele, vorteilbringende Mutationen zu erzeugen. Cairns nannte diese Art der Mutation gerichtete Mutation. Sie steht in krassem Widerspruch zu der bis dato gültigen Annahme, dass Mutationen in einer Häufigkeit auftreten, die in keinsten Weise durch ihre Konsequenzen beeinflusst werden.

Seit der Veröffentlichung der Idee der gerichteten Mutation herrscht ein ständiger Schlagabtausch unter den Molekularbiologen, ob es sich bei dem beobachteten (und reproduzierbarem) Verhalten der *E. Coli* Bakterien um eine Reaktion der Bakterien auf spezielle Umgebungssituationen oder um einen systematischen Fehler bei der Beobachtung erfolgreicher oder erfolgloser Mutationen handelt. Zusätzlich zu diesen Varianten, die die beiden extremen Meinungen darstellen, gibt es weitere Erklärungsversuche. Begriffe wie Hypermutation, d. h. der spontane Anstieg der Mutationsrate verbunden mit der Unmöglichkeit, fehlgeschlagene Mutationen (auf Grund ihrer äußerst kurzen Lebensdauer) zu untersuchen, versuchen zwischen diesen beiden Extremen zu vermitteln. Da die

genetischen Prozesse, die für eine endgültige Klärung dieser Frage für einen Nicht-Molekularbiologen kaum nachvollziehbar sind, soll der Blick über den Tellerrand mit zwei Zitaten und einem persönlichen Kommentar enden.

Mutations that occur more often when they're useful than when they're not: that I can document any day, every day, in the laboratory.¹

In evolutionary theory there has been an overemphasis on the power of selection as opposed to the generation of diversity.²

Aus Sicht eines Informatikers, der eine Variante der Evolutionsstrategien entwickelt hat, die die Bedeutung der Selektion mindert und die Bedeutung der Mutation erhöht, ist es ein beruhigendes Gefühl zu wissen, dass die biologische Evolution dies vielleicht auch längst tut.

-
1. Hall, B. G.; *Adaptive Evolution that Requires Multiple Spontaneous Mutations: Mutations Involving Base Substitutions*. Proc. Natl. Acad. Sci. USA, Vol. 88, 1991.
Prof. Dr. Hall ist Professor am Fachbereich Biologie der Rochester Universität
 2. Thaler, D. S. in Beardsley, T.; *Evolution Evolving*; Scientific American, No. 9, 1997.
Dr. Thaler ist Mikrobiologe an der Rockefeller Universität in New York

Kapitel 9 Literatur

- [Ada80] Adamo, J. M.; *Fuzzy Decision Trees*; Fuzzy Sets and Systems, Vol. 4, 1980. (207 - 219)
- [Alp91] Alpsector, J.; Gannet, J. W.; Haber, S.; Parker, M. B.; Chu, R.; *A VLSI-efficient Technique for Generating Multiple Uncorrelated Noise Sources and its Application to Stochastic Neural Networks*; IEEE Transactions on Circuits and Systems, Vol. 38, No. 1, 1991. (109 - 123)
- [Ang95] Anguita, D.; Rovetta, S.; Zunino, R.; *Compact Digital Pseudorandom Number Generator*; Electronic Letter, Vol. 31, No. 12, 1995. (956 - 958)
- [Ash70] Ash, R. B.; *Real Analysis and Probability*; Series on Probability and Mathematical Statistics, Vol. 11, Academic Press, New York, 1970.
- [Baa77] Baas, S. M.; Kwakernak, H.; *Rating and Ranking of Multiple Aspect Alternative Using Fuzzy Sets*; Automatica, Vol. 13, 1977. (47 - 58)
- [Bac98] Backhaus, W. G. K.; *Physiological and Psychophysical Simulations of Color Vision in Humans and Animals*; in: Backhaus, W. G. K.; Kliegl, R.; Werner, J. S. (Hrsg.); *Color Vision – Perspectives from Different Disciplines*; Walter de Gruyter, Berlin, 1998. (55 - 77)
- [Bac91] Bäck, Th.; Hoffmeister F.; *Extended Selection Mechanismus in Genetic Algorithms*; in: Belew, R. K.; Booker, L. B. (Hrsg.); *4th International Conference on Genetic Algorithms and their Applications*, Morgan Kaufmann, San Mateo, 1991. (92 - 99)
- [Bac92] Bäck, Th.; *The Interaction of Mutation Rate, Selection, and Self-Adaptation Within a Genetic Algorithm*; in: Männer, R.; Manderick, B. (Hrsg.); *Parallel Problem Solving from Nature II*, Elsevier, Amsterdam, 1992. (85 - 94).
- [Bac94] Bäck, Th.; *Evolutionary Algorithms in Theory and Practice*; Dissertation am Fachbereich Informatik, Universität Dortmund, Dortmund, 1994.
- [Bac95] Bäck, Th.; Schütz, M.; *Evolution Strategies for Mixed-Integer Optimization of Optical Multilayer Systems*; in: McDonnell, J. R.; Reynolds, R. G.; Fogel, D. B. (Hrsg.); *4th Annual Conference on Evolutionary Programming*, MIT Press, Cambridge, 1995. (33-51)

- [Bäc96] Bäck, Th.; *Evolutionary Algorithms in Theory and Practice*; Oxford University Press, New York, 1996.
- [Bae74] Baer, R. M.; Martinez, H. M.; *Automata and Biology*; Annual Reviews of Biophysics and Bioengineering, Vol. 3, 1974. (255 - 291)
- [Bal94] Baluja, S.; *Population-based Incremental Learning: A Method for Integrating Genetic Search Based Function Optimization and Competitive Learning*; Technical Report CMU-CS-94-163, Carnegie Mellon University, Pittsburg, 1994.
- [Bal97] Baluja, S.; Davies, S.; *Using Optimal Dependency-trees for Combinatorial Optimization: Learning the Structure of the Search Space*; 14th International Conference on Machine Learning, Morgan Kaufmann, San Mateo, 1997. (30 - 38)
- [Ban98] Banzhaf, W.; Francone, F. D.; Keller, R. E.; Nordin, P.; *Genetic Programming – An Introduction*; Morgan Kaufmann, San Francisco, 1998.
- [Bey88] Beyer, O.; et al; *Wahrscheinlichkeitsrechnung und mathematische Statistik*; 5. Auflage, BSB B. G. Teubner Verlagsgesellschaft, Leipzig, 1988.
- [Bey00] Beyer, H.-G.; Deb, K.; *On the Desired Behaviors of Self-Adaptive Evolutionary Algorithms*; in: Schoenauer, M.; Deb, K.; Rudolph, G.; Yao, X.; Lutton, E.; Merelo, J. J.; Schwefel, H.-P. (Hrsg.); *Parallel Problem Solving from Nature – PPSN VI*, Lecture Notes in Computer Science, Vol. 1917, Springer, Berlin, 2000. (59 - 68)
- [Bey01] Beyer, H.-G.; *The Theory of Evolution Strategies*; Natural Computing Series, Springer, Berlin, 2001.
- [Bie80] Biess, G.; Erfurth, H.; Zeidler G.; *Optimale Prozesse und Systeme*; 2. Auflage, BSB B. G. Teubner Verlagsgesellschaft, Leipzig, 1980.
- [Bie97] Biewer, B.; *Fuzzy-Methoden: Praxisrelevante Rechenmodelle und Fuzzy-Programmiersprachen*; Springer, Berlin, 1997.
- [Bla96] Bland, I. M.; Megson, G. M.; *Systolic Random Number Generation for Genetic Algorithms*; Electronic Letter, Vol. 32, No. 12, 1996. (1069 - 1070)
- [Böh93] Böhme G.; *Fuzzy-Logik - Einführung in die algebraischen und logischen Grundlagen*, Springer, Berlin, 1993.
- [Bor78] Born, J.; *Evolutionsstrategien zur numerischen Lösung von Adaptionsaufgaben*; Dissertation an der Humboldt-Universität, Berlin, 1978.
- [Bor85] Bortolan, G., Degani, R.; *A Review of some Methods for Ranking Fuzzy Subsets*; Fuzzy Sets and Systems, Vol. 15, 1985. (1 - 19).
- [Box58] Box, G. E. P.; Muller, M. E.; *A Note on the Generation of Random Normal Deviates*; Ann. Math. Statist., Vol. 29, 1958. (610 - 611)
- [Bra92] Brandt, S.; *Datenanalyse*; 3. Auflage, BI Wissenschaftsverlag, Mannheim, 1992.

-
- [Bre62] Bremermann, H. J.; *Optimization Through Evolution and Recombination*; in: Yovitz, M. C.; Jacobi, G. T.; Goldstein, G. D. (Hrsg.) ; Self-Organizing Systems 1962; Conference on Self-Organizing Systems, Spartan Books, Washington D. C., 1962. (93 - 106)
- [Bro89] Bronstein, I. N.; *Taschenbuch der Mathematik*; 24. Auflage, BSB B. G. Teubner Verlagsgesellschaft, Leipzig, 1989.
- [Bro90] Bronstein, I. N.; *Taschenbuch der Mathematik, ergänzende Kapitel*; 6. Auflage, BSB B. G. Teubner Verlagsgesellschaft, Leipzig, 1990.
- [Bru06] Bruns, H.; *Wahrscheinlichkeitsrechnung und Kollektivmasslehre*; Teubner, Leipzig, 1906.
- [Bur42] Burr, I. W.; *Cumulative Frequency Functions*; Annals of Mathematical Statistics, Vol. 13, 1942. (215 - 232)
- [Bur68] Burr, I. W.; *On General Systems of Distributions III. The Sample Range*; Journal of American Statistical Association, Vol. 63, 1968. (636 - 643)
- [Cox98] Cox, E.; *The Fuzzy Systems Handbook*; Academic Press, San Diego, 1998.
- [Cun69] Cunningham, S. W.; *From Normal Integral do Derivate*; Applied Statistics; Vol. 28, No. 2, 1969. (290 - 293)
- [Dav72] Davies, O. L.; Goldsmith, P. L.; *Statistical Methods in Research and Production*; 4. Auflage, Oliver and Boyd, Edinburgh, 1972.
- [Dav91] Davis, L.; *Handbook of Genetic Algorithms*; Van Nostrand Reinhold, New York, 1991.
- [Daw76] Dawkins, R.; *The Selfish Gene*; Oxford University Press, Oxford, 1976.
- [Daw86] Dawkins, R.; *The Blind Watchmaker*; Longman, Oxford, 1986.
- [Deb95a] Deb, K.; Agrawal, R. B.; *Simulated Binary Crossover for Continuous Search Space*; Complex Systems, Vol. 9, 1995. (115 - 148)
- [Deb95b] Deb, K.; Kumar, A.; *Real-Coded Genetic Algorithms with Simulated Binary crossover: Studies on Multi-Modal and Multi-Objective Problems*; Complex Systems, Vol. 9, No. 6, 1995. (431 - 454)
- [Deb99] Deb, K.; Beyer, H.-G.; *Self-Adaptive Genetic Algorithms with Simulated Binary Crossover*; Technischer Bericht Nr. CI-61/99, Universität Dortmund, Dortmund, 1999.
- [Deb01] Deb, K. Beyer, H.-G.; *Self-Adaptive Genetic Algorithms with Simulated Binary Crossover*; Evolutionary Computation Journal, Vol. 9, No. 2, 2001. (197 - 221)
- [DeJ75] De Jong, K. A.; *An Analysis of the Behaviour of a Class of Genetic Adaptive Systems*; PhD Thesis, University of Michigan, 1975.
- [Dev86] Devroye, L.; *Non-uniform Random Variate Generation*; Springer, New York, 1986.

- [Die94] Dierkes, S.; Reusch, B.; Temme, K.-H.; *Hybrid Reasoning with FUZZYNEX*; in: Kruse, R.; Gebhardt, J.; Palm, R. (Hrsg.); *Fuzzy Systems in Computer Science*; Vieweg, Braunschweig, 1994. (183 - 191)
- [Dre90] Dreißig, B.; *Ein Beitrag zur Auslegung von trockenlaufenden Schraubenmotoren*, Dissertation der Fakultät Maschinenbau, Universität Dortmund, Dortmund, 1990.
- [Dro98] Droste, S.; Jansen, Th.; Wegener, I.; *A Rigorous Complexity Analysis of the (1+1) Evolutionary Algorithm for Linear Functions With Boolean Inputs*; in: Fogel, D. B.; Schwefel, H.-P.; Bäck, Th.; Yao, X.; (Hrsg.); *IEEE International Conference on Evolutionary Computation – ICEC '98*, IEEE Press, Piscataway, 1998. (499 - 504)
- [Dub93] Dubois, D.; Prade, H.; Yager, R. R.; *Fuzzy Sets for Intelligent Systems*, Morgan Kaufmann, San Mateo, 1993.
- [Edg07] Edgeworth, F. Y.; *On the Representation of Statistical Frequency by a Series*; *Journal of the Royal Society, Series A*, Vol. 70, 1907. (102 - 106)
- [Edg16] Edgeworth, F. Y.; *On the Mathematical Representation of Statistical Data*; *Journal of the Royal Society, Series A*, Vol. 79, 1916. (455 - 500)
- [Els80] Elster, K.-H.; *Nichtlineare Optimierung*; 2. Auflage, BSB B. G. Teubner Verlagsgesellschaft, Leipzig, 1980.
- [Esh93] Eshelman, L. J.; Schaffer, J. D.; *Real-Coded Genetic Algorithms and Interval Schemata*; in: Whitley, D. (Hrsg.); *Foundations of Genetic Algorithms II*, 1993. (187-202)
- [Etx99] Etxeberria, R.; Larrañaga, P.; *Global Optimization with Bayesian Networks*; *II Symposium on Artificial Intelligence, CIFMA99*, 1999. (332 - 339)
- [Eva93] Evans, M.; Hastings, N.; Peacock, B.; *Statistical Distribution*; 2. Auflage, John Wiley & Sons, New York, 1993.
- [Fat94a] Fathi, M.; Hildebrand, L.; *Using Fuzzy Logic to Increase the Performance of FEM-Modules*; *International Journal on Cybernetics & Systems*, Taylor & Francis, Vol. 25, No. 2, 1994. (207 - 216)
- [Fat94b] Fathi, M.; Hildebrand, L.; *Evolutionary Strategies for Optimizing of Fuzzy Rules*; in: Bouchon-Meunier, B.; Yager, R. R. (Hrsg.); *Fifth International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems – IPMU, Paris, 4.-8. Juli, 1994*. (671 - 674)
- [Fat95] Fathi, M.; Hildebrand, L.; *The Application of Evolution Strategies to the Problem of Parameter Optimization in Fuzzy Rule Based Systems*; in: IEEE Computer Society (Hrsg.); *IEEE International Conference on Evolutionary Computing – ICEC '95*, IEEE Press, Piscataway, 1995. (825 - 830)
- [Fat96] Fathi, M.; Hildebrand, L.; *Function Approximation using Fuzzy Logic and Evolutionary Strategies*; in: Jamshidi, M.; Fathi, M.; Pierrot, F. (Hrsg.); *Soft Computing with Industrial Applications, World Automation Congress – WAC'96*, TSI Press, Albuquerque, 1996. (197 - 202)

-
- [Fat97] Fathi, M.; Hildebrand, L.; *Modell Free Optimization of Fuzzy Rule Based Systems Using Evolution Strategies*; IEEE International Journal on Transaction on Systems, Man and Cybernetics, 1997. (270 - 277)
- [Fis73] Fisz, M.; *Wahrscheinlichkeitsrechnung und mathematische Statistik*; 7. Auflage, VEB Deutscher Verlag der Wissenschaften, Berlin, 1973.
- [Fle87] Fletcher, R.; *Practical Methods of Optimization*; John Wiley & Sons, Chichester, 1987.
- [Fog95] Fogel, L. J.; Angeline, P. J.; Fogel, D. B.; *An Evolutionary Programming Approach to Selfadaptation on Finite State Machines*; in: McDonnell, J. R.; Reynolds, R. G.; Fogel, D. B. (Hrsg.); 4th Annual Conference on Evolutionary Programming, MIT Press, Cambridge, 1995. (355 - 365)
- [Fog99] Fogel, L. J.; *Intelligence Through Simulated Evolution – Forty Years of Evolutionary Programming*; Series on Intelligent Systems, John Wiley & Sons, New York, 1999.
- [Fre99] Freitag, K.; Hildebrand, L.; Moraga, C.; *Quaternary Coded Genetic Algorithms*; in: IEEE Computer Society (Hrsg.); 29th IEEE International Symposium on Multiple-valued Logic – ISMVL 99, IEEE Press, Piscataway, 1999. (194 - 199)
- [Gar99] Garrett, G.; et al.; *Parallel Real-Valued Genetic Algorithms for Bio-Remediation Optimization of TCE-Contaminated Groundwater*; in: IEEE Computer Society (Hrsg.); Congress on Evolutionary Computation – CEC99, IEEE Press, Piscataway, 1999. (2183 - 2189)
- [Gho96] Ghozeil, A.; Fogel, D. B.; *A Preliminary Investigation into Directed Mutation in Evolutionary Algorithms*; in: Voigt, H.; Ebeling, W.; Rechenberg, I. (Hrsg.); Parallel Problem Solving from Nature – PPSN IV, Lecture Notes in Computer Science, Springer, Berlin, 1996. (329 - 335)
- [Gne60] Gnedenko, B. W.; Kolmogorov, A. N.; *Grenzverteilungen von Summen unabhängiger Zufallsgrößen*; 2. Auflage, Akademie-Verlag, Berlin, 1960.
- [Gne68] Gnedenko, B. W.; Kolmogorov, A. N.; *Limit Distributions for Sums of Independent Random Variables*; überarbeitete Auflage, Addison-Wesley, Reading, 1968.
- [Gor98] Gorges-Schleuter, M.; *A Comparative Study of Global and Local Selection in Evolution Strategies*; in: Eiben, A. E.; Bäck, Th.; Schoenauer, M.; Schwefel, H.-P. (Hrsg.); Parallel Problem Solving from Nature – PPSN V, Lecture Notes in Computer Science, Vol. 1498, Springer, Berlin, 1998. (367 - 377)
- [Gul94] Gulbins, J.; Obermayr, K.; *Desktop Publishing mit FrameMaker*, 2. Auflage, Springer, Berlin, 1994.
- [Han95a] Hansen, N.; Ostermeier, A.; Gawelczyk, A.; *Über die Adaptation von allgemeinen, Koordinatensystem-unabhängigen, normalverteilten Mutationen in der Evolutionsstrategie: Die Erzeugendensystemadaptation*; Technischer Report TR-02-95, Institut für Bionik und Evolutionstechnik der Technischen Universität Berlin, 1995.
- [Han95b] Hansen, N.; Ostermeier, A.; Gawelczyk, A.; *On the Adaptation of Arbitrary Normal Mutation Distributions in Evolution Strategies: The Generating Set Adaptation*; in:

- Eshelman, L. (Hrsg.); Sixth International Conference on Genetic Algorithms – ICGA '95, Morgan Kaufmann, San Mateo, 1995. (57 - 64)
- [Han98] Hansen, N.; *Verallgemeinerte individuelle Schrittweitenregelung in der Evolutionsstrategie*; Dissertation an der Technischen Universität Berlin; Mensch und Buch Verlag, Berlin, 1998.
- [Har97] Harik, G. R.; Lobo, F. G.; Goldberg, D. E.; *The Compact Genetic Algorithm*; IlliGAL Report No. 97006, Illinois Genetic Algorithms Laboratory, University of Illinois at Urbana-Champaign, Illinois, 1997.
- [Har98] Hardin, C. L.; *Basic Color Terms and Basic Color Categories*; in: Backhaus, W. G. K.; Kliegl, R.; Werner, J. S. (Hrsg.); *Color Vision – Perspectives from Different Disciplines*; Walter de Gruyter, Berlin, 1998. (207 - 217)
- [Har99] Harik, G. R.; *Linkage Learning via Probabilistic Modelling in the ECGA*; IlliGAL Report No. 99010, Illinois Genetic Algorithms Laboratory, University of Illinois at Urbana-Champaign, Illinois, 1999.
- [Heb49] Hebb, D.; *Organization of Behaviour*; John Wiley & Sons, New York, 1949.
- [Hec88] Hecht-Nielsen, R.; *Neurocomputing*; Addison-Wesley, Reading, 1988.
- [Hel98] Hellekalek, P.; *Good Random Number Generators are (not so) Easy to Find*; Mathematics and Computers in Simulation, Vol. 46, Elsevier, 1998. (485 - 505)
- [Her98] Herrera, F.; Lozano M.; Moraga C.; *Hybrid Distributed Real-Coded Genetic Algorithms*; in: Eiben, A. E.; Bäck, Th.; Schoenauer, M.; Schwefel, H.-P. (Hrsg.); *Parallel Problem Solving from Nature – PPSN V, Lecture Notes in Computer Science*, Vol. 1498, Springer, Berlin, 1998. (603-612)
- [Heß82] Heß, D.; *Genetik: Grundlagen - Erkenntnisse*; Herder, Freiburg, 1982.
- [Hil73] Hill, G. W.; Davis, A. W.; *Algorithm 442, Normal Deviate*; Collected algorithms from CACM, 1973.
- [Hil96] Hildebrand, L.; *Schiefe - Ein Ansatz zur Erweiterung von Evolutionsstrategien um gerichtete Mutation*; in: Reusch, B. (Hrsg.); 16. Workshop Interdisziplinäre Methoden der Informatik, Forschungsberichte des Fachbereichs Informatik, Bd. 639, Universität Dortmund, Dortmund, 1996. (56 - 64)
- [Hil97a] Hildebrand, L.; Fathi, M.; *Knowledge Based Modelling for the Design of Screw Rotors*; in: Hui, D. (Hrsg.); *International Conference on Composite Engineering – ICCE/4*, 1997. (1035 - 1036)
- [Hil97b] Hildebrand, L.; *Evolutionsstrategien: Ein Vergleich von gerichteter und ungerichteter Mutation*; in: Reusch, B. (Hrsg.); 17. Workshop Interdisziplinäre Methoden der Informatik, Forschungsberichte des Fachbereichs Informatik, Bd. 682, Universität Dortmund, Dortmund, 1997. (3 - 18)

-
- [Hil98a] Hildebrand, L.; Fathi, M.; *Evolutionary Design of Screw Rotor Profiles*; in: Hui, D. (Hrsg.); International Conference on Composite Engineering – ICCE/5, 1998. (941 - 942)
- [Hil98b] Hildebrand, L.; *Unschärfe Farbverarbeitung*; in: Reusch, B. (Hrsg.); 18. Workshop Interdisziplinäre Methoden der Informatik, Forschungsberichte des Fachbereichs Informatik, Bd. 718, Universität Dortmund, Dortmund, 1998. (69 -83)
- [Hil99a] Hildebrand, L.; Reusch, B.; Fathi, M.; *Directed Mutation – A New Selfadaptation for Evolution Strategies*; in: IEEE Computer Society (Hrsg.); Congress on Evolutionary Computation – CEC99, IEEE Press, Piscataway, 1999. (1550 - 1557)
- [Hil99b] Hildebrand, L.; *Eigenschaften gerichteter Mutationsoperatoren am Beispiel des Kugelmodells*; in: Reusch, B. (Hrsg.); 19. Workshop Interdisziplinäre Methoden der Informatik, Forschungsberichte des Fachbereichs Informatik, Bd. 729, Universität Dortmund, Dortmund, 2000. (43 -48)
- [Hil00a] Hildebrand, L.; Reusch, B.; *Fuzzy Color Processing*; in: Kerre, E.; Nachtgael, M. (Hrsg.); Fuzzy Techniques in Image Processing, Studies in Fuziness and Soft Computing, Vol. 52, Physica-Verlag, Heidelberg, 2000. (267 - 286)
- [Hil00b] Hildebrand, L.; Fathi, M.; *Linguistic Color Processing for Human-Like Vision Systems*; in: SPIE – The International Society for Optical Engineering (Hrsg.); Electronic Imaging 2000, IS&T/SPIE 12th International Symposium, SPIE Press, Bellingham, 2000. (3959-32 S7)
- [Hil00c] Hildebrand, L.; *Detection of 3-Dimensional Image Features Using a Single Camera*; in: Jamshidi, M.; Fathi, M.; Furuhashi, T. (Hrsg.); International Forum on Multimedia and Image Processing – IFMIP, World Automation Congress – WAC2000, TSI Press, Albuquerque, 2000. (834 - 842)
- [Hof71] Hoffmann, U.; Hofmann, H.; *Einführung in die Optimierung*; Verlag Chemie GmbH, Weinheim, 1971.
- [Hof92] Hoffmeister, F.; Bäck, Th.; *Genetic Algorithms and Evolution Strategies: Similarities and Differences*; Technischer Bericht, Nr. SYS-1/92, Universität Dortmund, Fachbereich Informatik, LS11, Dortmund, 1992.
- [Hol75] Holland, J. H.; *Adaption in Natural and Artificial Systems*; University of Michigan Press, Ann Arbor, 1975.
- [IMS87] IMSL, Inc.; *User's Manual: Stat/Library*; Vol 3, IMSL, Houston, 1987. (911 - 912)
- [Joh49] Johnson, N. L.; *Systems of Frequency Curves Generated by Methods of Translations*; Biometrika, Vol. 36, 1949. (149 - 176)
- [Joh70] Johnson, N. L.; Kotz, S.; *Continuous Univariate Distributions*; Vol. 1, John Wiley & Sons, New York, 1970.
- [Jul99] Julstrom, B. A.; *It's All the Same to Me: Revisiting Rank-Based Probabilities and Tournaments*; in: IEEE Computer Society (Hrsg.); Congress on Evolutionary Computation – CEC99, IEEE Press, Piscataway, 1999. (1501 - 1505)

- [Kah95] Kahlert, J.; *Fuzzy Control für Ingenieure*; Vieweg, Braunschweig, 1995.
- [Kan92] Kandel, A.; *Fuzzy Expert Systems*; CRC Press, 1992.
- [Kap96] Kappler, C.; *Are Evolutionary Algorithms Improved by Large Mutations?*; in: Voigt, H.-M.; Ebeling, W.; Rechenberg, I.; Schwefel, H.-P. (Hrsg.); *Parallel Problem Solving from Nature – PPSN IV*, Lecture Notes in Computer Science, Vol. 1141, Springer, Berlin, 1996. (346 - 355)
- [Kau00] Kauder, K.; Reusch, B.; *Methode zur Auslegung trockenlaufender Rotationsverdrängermaschinen*; DFG-Projekt Ka 663/23-1; Fakultät Maschinenbau & Fachbereich Informatik, Universität Dortmund, Dortmund, 2000.
- [Kel99] Kelley, C. T.; *Iterative Methods for Optimization*; Frontiers in Applied Mathematics, Vol. 18, SIAM – Society for Industrial and Applied Mathematics, Philadelphia, 1999.
- [Ker82] Kerre, E. E.; *The Use of Fuzzy Logic in Electrocardiological Diagnostics*; in: Gupta, M. M.; Sanchez, E. (Hrsg.); *Approximate Reasoning in Decision Analysis*, North-Holland-Publishing Company, 1982. (277 - 282)
- [Ker83] Kernighan, B. W.; Ritchie, D. M.; *Programmieren in C*; Carl Hanser, München, 1983.
- [Kie98] Kiencke, U.; *Signale und Systeme*; Oldenbourg Verlag, München, 1998.
- [Kir83] Kirkpatrick S.; Gelatt D.; Vecchi, M. P.; *Optimization by Simmulated Annealing*; Science, Vol. 220, No. 4598, 1983. (671-680)
- [Kli88] Klir, G. J.; Folger, T. A.; *Fuzzy Sets, Uncertainty, and Information*; Prentice-Hall, Englewood Cliffs, 1988.
- [Kur99] Kursawe, F.; *Grundlegende empirische Untersuchungen der Parameter von Evolutionsstrategien – Metastrategien*; Dissertation am Fachbereich Informatik, Universität Dortmund, Dortmund, 1999.
- [Kva96] Kvasnicka, V.; Pelikan, M.; Pospichal, J.; *Hill Climbing with Learning (An Abstraction of Genetic Algorithms)*; Neural Network World, No. 6, 1996. (773 - 796)
- [Lar99] Larrañaga, P.; Inza, I.; Sierra, B.; Etxeberria, R.; Lozano, J. A.; Peña, J. M.; *Representing the Behaviour of Supervised Classification Learning Algorithms by Bayesian Networks*; Pattern Recognition Letters, Vol. 20, No. 11-13, 1999. (1201 - 1210)
- [Lar00] Larrañaga, P.; Etxeberria, R.; Lozano, J. A.; Peña, J. M.; *Combinatorial Optimization by Learning and Simulation of Bayesian Networks*; Conference on Uncertainty in Artificial Intelligence, UAI 2000, 2000. (343 - 352)
- [Law91] Law, A. M.; Kelton, W. D.; *Simulation Modelling & Analysis*; McGraw-Hill Series in Industrial Engineering and Management Science, McGraw-Hill, Singapur, 1991.
- [Leg99] Leguizamon, G.; Michalewicz, Z.; *A New Version of Ant System for Subset Problems*; in: IEEE Computer Society (Hrsg.); *Congress on Evolutionary Computation – CEC99*, IEEE Press, Piscataway, 1999. (1459 - 1464)

-
- [Lel88] Lelgemann, K. D.; *Einsatz von Schraubenkompressoren, insbesondere in der chemischen und petrochemischen Industrie*, in Vetter, G. (Hrsg.); Handbuch Verdichter, 1. Ausgabe, Vulkan Verlag, Essen, 1988.
- [Lüt97] Lütkepohl, H.; *Handbook of Matrices*; John Wiley & Sons, Chichester, 1997.
- [Man00] Mandischer, M.; *An Empirical Investigation of Neural Networks, Evolution Strategies and Evolutionary-Tained Neural Networks and their Application to Chemical Engineering*; Dissertation am Fachbereich Informatik, Universität Dortmund, 2000.
- [Mar83] Marse, K; Roberts, S. D.; *Implementing a Portable FORTRAN Uniform(0,1) Generator*; Simulation, Vol. 41, 1983. (135-139)
- [May93] Mayer a.; et al., *Fuzzy Logic - Einführung und Leitfaden zur praktischen Anwendung*, Addison-Wesley, Bonn, 1993.
- [Mic96] Michalewicz, Z.; *Genetic Algorithms + Data Structures = Evolution Programs*; Springer, Berlin, 1996.
- [Mic00] Michalewicz, Z.; Fogel, D. B.; *How to Solve it: Modern Heuristics*; Springer, Berlin, 2000.
- [Müh97] Mühlenbein, H.; *The Equation for Response to Selection and its Use for Prediction*; Evolutionary Computing, Vol. 5, No .3, 1997. (303 - 346)
- [Müh99a] Mühlenbein, H.; Rodriguez, A. O.; Mahnig, T.; *Schemata, Distributions, and Graphical Models in Evolutionary Optimization*; Journal of Heuristics, Vol. 5, 1999. (215 - 247)
- [Müh99b] Mühlenbein, H.; Mahnig, T.; *FDA – A Scalable Evolutionary Algorithm for the Optimization of Additively Decomposed Functions*; Evolutionary Computation, Vol. 7, 1999. (353 - 376)
- [Müh99c] Mühlenbein, H.; Mahnig, T.; *Convergence Theory and Application of Factorized Distribution Algorithms*; Journal of Computing and Information Technology, Vol. 7, 1999. (19 - 32)
- [Ngu89] Nguyen, H. T.; Rogers, G. S.; *Fundamentals of Mathematical Statistics, Volume 1: Probability and Statistics*; Springer, New York, 1989.
- [Ode74] Odeh, R. E.; Evans, J. O.; *The Percentage Points of the Normal Distribution*; Appl. Statist., Vol. 23, 1974. (96 - 97)
- [Ohk98] Ohkura, K.; Matsumura, Y.; Ueda, K.; *Robust Evolution Strategies*; in McKay, B.; Yao, X.; Newton, C. S.; Kim, J.-H.; Furuhashi, T. (Hrsg.); Second Asia-Pacific Conference on Simulated Evolution and Learning – SEAL'98, Lecture Notes in Computer Science, Vol. 1585, Springer, Berlin, 1998. (10 - 17)
- [Opp89] Oppenheim, A. V.; Willsky, A. S.; *Signale und Systeme*; VCH Verlagsgesellschaft, Weinheim, 1989.

- [Ost94] Ostermeier, A.; Gawelczyk, A.; Hansen, N.; *A Derandomized Approach to Self Adaptation of Evolution Strategies*; Evolutionary Computation, Vol. 2, No. 4, 1994. (369 - 380)
- [Ost97] Ostermeier, A.; *Schrittweisenadaptation in der Evolutionsstrategie mit einem entstochastisierten Ansatz*; Dissertation an der Technischen Universität Berlin, Berlin, 1997.
- [Pea95] Pearson, K.; *Contribution to the Mathematical Theory of Evolution. II. Skew Variations in Homogeneous Material*; Philosophical Transactions of the Royal Society of London, Series A, Vol. 186, 1895. (343 - 414)
- [Pel98] Pelikan, M.; Goldberg, D. E.; Cantú-Paz, E.; *Linkage Problem, Distribution Estimation, and Bayesian Estimation Networks*; IlliGAL Report No. 98013, Illinois Genetic Algorithms Laboratory, University of Illinois at Urbana-Champaign, Illinois, 1998.
- [Pel99] Pelikan, M.; Mühlenbein, H.; *The Bivariate Marginal Distribution Algorithm*; in: Roy, R.; Furuhashi, T.; Chawdhry, P. K. (Hrsg.), *Advances in Soft Computing - Engineering Design and Manufacturing*, Springer, London, 1999. (521 - 535)
- [Pow98] Powell, M. J. D.; *Direct Search Algorithms for Optimization Calculations*; Acta Numerica, Vol. 7, Cambridge University Press, Cambridge, 1998. (287 - 336)
- [Ras95] Rasch, D.; *Mathematische Statistik*; Johann Ambrosius Barth Verlag, Heidelberg, 1995.
- [Rec70] Rechenberg, I.; *Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*; Dissertation an der Technischen Universität Berlin, Berlin, 1970.
- [Rec73] Rechenberg, I.; *Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*; Frommann-Holzboog, Stuttgart, 1973.
- [Rec94] Rechenberg, I.; *Evolutionsstrategie '94*; Frommann-Holzboog, Stuttgart, 1994.
- [Rec99] Rechenberg, P.; Pomberger, G.; *Informatik-Handbuch*; 2. aktualisierte und erweiterte Auflage, Carl Hanser, München, 1999.
- [Reu94] Reusch, B.; *Fuzzy Logik - Theorie und Praxis*; Tagungsband der 4. Dortmunder Fuzzy-Tage, Springer, Berlin, 1994.
- [Reu98] Reusch, B.; Fathi, M.; Hildebrand, L.; *Fuzzy Color Processing for Quality Improvement*; in: Jamshidi, M.; Bien, Z.; Fathi, M. (Hrsg.); *International Forum on Multimedia and Image Processing – IFMIP, World Automation Congress – WAC'98*, TSI Press, Albuquerque, 1998. (841 - 848)
- [Reu99] Reusch, B.; Hildebrand, L.; *Intelligente Systeme in der Schweisstechnik*, Spektrum der Wissenschaft - Digest: Computersoftware, Heidelberg, 1999. (64 - 67)
- [Ros89] Ross, S. M.; *Introduction to Probability Models*; 4. Auflage, Academic Press, San Diego, 1989.
- [Ros94] Ross, P.; Corne, D.; Fang, H.-L.; *Improving Evolutionary Timetabling with Delta Evaluation and Directed Mutation*; DAI Research Paper, No. 707, Department of Artificial Intelligence, University of Edinburgh, 1994.

-
- [Rud94] Rudolph, G.; *An Evolutionary Algorithm for Integer Programming*; in: Davidor, Y.; Schwefel, H.-P.; Männer, R. (Hrsg.); *Parallel Problem Solving From Nature – PPSN III*, Lecture Notes in Computer Science, Vol. 866, Springer, Berlin, 1994. (139-148)
- [Rud96] Rudolph, G.; *Convergence Properties of Evolutionary Algorithms*; Dissertation am Fachbereich Informatik, Universität Dortmund, Dortmund, 1996.
- [Rud97] Rudolph, G.; *Local Convergence Rates of Simple Evolutionary Algorithms with Cauchy Mutations*; *IEEE Transactions on Evolutionary Computation* Vol. 1, No. 4, 1997. (249-258)
- [Rud98] Rudolph, G.; *Asymptotical Convergence Rates of Simple Evolutionary Algorithms under Factorizing Mutation Distributions*; in: Hao, J.-K.; Lutton, E.; Ronald, E.; Schoenauer, M.; Snyers, D. (Hrsg.); *Third European Conference on Artificial Evolution – AE '97*, Lecture Notes in Computer Science, Vol. 1363, Springer, Berlin, 1998. (223-233)
- [Rud99] Rudolph, G.; *Self-Adaptation and Global Convergence: A Counter-Example*; in: *IEEE Computer Society (Hrsg.); Congress on Evolutionary Computation – CEC99*, IEEE Press, Piscataway, 1999. (646 - 651)
- [Sal98a] Salomon, R.; *The Evolutionary-Gradient-Search Procedure*; in: Koza, J. R.; Banzhaf, W.; Chellapilla, K.; Deb, K.; Dorigo, M.; Fogel, D. B.; Garzon, M. H.; Goldberg, D. E.; Iba, H. (Hrsg.); *Third Annual Conference on Genetic Programming*, Morgan Kaufmann, San Mateo, 1998. (852 - 861)
- [Sal98b] Salomon, R.; *Accelerating the Evolutionary-Gradient-Search Procedure: Individual Step Sizes*; in: Eiben, A. E.; Bäck, Th.; Schoenauer, M.; Schwefel, H.-P. (Hrsg.); *Parallel Problem Solving from Nature – PPSN V*, Lecture Notes in Computer Science, Vol. 1498, Springer, Berlin, 1998. (408 - 417)
- [Sch75] Schwefel, H.-P.; *Evolutionsstrategie und numerische Optimierung*; Dissertation an der Technischen Universität Berlin, Berlin, 1975.
- [Sch94] Schwefel, H.-P.; *Evolution and Optimum Seeking*; John Wiley & Sons, New York, 1994.
- [Sch95] Schöneburg E.; Heinzmann F.; Feddersen S.; *Genetische Algorithmen und Evolutionsstrategien*; Addison-Wesley, Bonn, 1995.
- [Sha87a] Shapiro, S. C.; et al.; *Encyclopedia of Artificial Intelligence*; Ausgabe 1, 4. Auflage, John Wiley & Sons, New-York, 1987.
- [Sha87b] Shapiro, S. C.; et al.; *Encyclopedia of Artificial Intelligence*; Ausgabe 2, 4. Auflage, John Wiley & Sons, New-York, 1987.
- [Sim96] Simonnet, M.; *Measures and Probabilities*; Springer, New York, 1996.
- [Sto95] Storn, R.; Price, K.; *Differential Evolution – A Simple and Efficient Adaptive Scheme for Global Optimization Over Continuous Spaces*; Technischer Bericht TR-95-012, International Computer Science Institute (ICSI), Berkeley, 1995.

- [Sto96] Storn, R.; *System Design by Constraint Adaptation and Differential Evolution*; Technischer Bericht TR-96-039, International Computer Science Institute (ICSI), Berkeley, 1996.
- [Sto98] Stoerig, P.; *Wavelength Information Processing Versus Color Perception: Evidence from Blindsight and Color-Blind Sight*; in: Backhaus, W. G. K.; Kliegl, R.; Werner, J. S. (Hrsg.); *Color Vision – Perspectives from Different Disciplines*; Walter de Gruyter, Berlin, 1998. (131 - 147)
- [Sys93] Syswerda, G.; *Simulated Crossover in Genetic Algorithms*; Foundations of Genetic Algorithms, Morgan Kaufmann, San Mateo, 1993. (239 - 255)
- [Tor52] Toranzos, F. I.; *An Asymmetric Bell-shaped Frequency Curve*; Annals of Mathematical Statistics, Vol. 23, 1952. (467 - 469)
- [Wei99] Weicker, K.; Weicker, N.; *On Evolution Strategy Optimization in Dynamic Environments*; in: IEEE Computer Society (Hrsg.); Congress on Evolutionary Computation – CEC99, IEEE Press, Piscataway, 1999. (2039 - 2046)
- [Wei01] Weicker, K.; *Problem Difficulty in Real-valued Dynamic Problems*; in: Reusch, B. (Hrsg.), Computational Intelligence – Theory and Application, Lecture Notes in Computer Science, Vol. 2206, Springer, Berlin, 2001. (313 - 325)
- [Wri91] Wright, A.; *Genetic Algorithms for Real Parameter Optimization*; Rawlins, in: G. J. E. (Hrsg.); First Workshop on the Foundations of Genetic Algorithms; Morgan Kaufmann, San Mateo, 1991. (205 - 218)
- [Wol91] Wolfram, S.; *Mathematica*; 4. Auflage, Addison-Wesley, Reading, 1991.
- [Wol97] Wolpert, D.; Macready, W.; *No Free Lunch Theorems for Optimization*; IEEE Transactions on Evolutionary Computation, 1 (1), 1997. (67 - 82)
- [Yag78] Yager, R. R.; *Ranking Fuzzy Subsets Over the Unit Interval*; in: IEEE Computer Society (Hrsg.); IEEE International Conference on Cybernetics and Society – ICCS'78, IEEE Press, Piscataway, 1978. (1735 - 1737)
- [Yag81] Yager, R. R.; *A New Methodology for Ordinal Multiple Aspect Decisions Based on Fuzzy Sets*; Decision Sciences, Vol. 12, 1981. (589 - 600)
- [Yao97] Yao, X.; Liu, Y.; *Fast Evolution Strategies*; Control and Cybernetics, 26 (3), 1997. (467 - 496).
- [Yao99] Yao, X.; Liu, Y.; Lin, G.; *Evolutionary Programming Made Faster*; IEEE Transactions on Evolutionary Computation, Vol. 3, No. 2, 1999. (82 - 102).
- [Yov62] Yovitz, M. C.; Jacobi, G. T.; Goldstein, G. D.; *Self-Organizing Systems 1962*; Conference on Self-Organizing Systems, Spartan Books, Washington DC, 1962.
- [Zad92] Zadeh, L. A.; Kacprzyk, J. (Hrsg.); *Fuzzy Logic for the Management of Uncertainty*; John Wiley & Sons, New York, 1992.
- [Zub72] Zubrzycki, S.; *Lectures in Probability Theory and Mathematical Statistics*; Elsevier, Amsterdam, 1972.

Anhang A Verwendete Symbole

In diesem Abschnitt werden die verwendeten Symbole aufgeführt und kurz erläutert. Generell gelten in dieser Arbeit folgende Notationen:

- Vektoren werden durch fett gesetzte Buchstaben und Zahlen markiert, z. B.: $\mathbf{0}$, \mathbf{v} und \mathbf{Z}
- Generationen werden durch einen Exponenten in Klammern dargestellt, z. B.: $s^{(t)}$
- Individuen einer Generation werden durch einen Index unterschieden, z. B.: \mathbf{I}_i
- Komponenten eines Vektors werden ebenfalls durch einen Index markiert, z. B.: x_j . Sollte durch den Kontext nicht deutlich werden, ob ein spezielles Individuum oder eine Komponente eines Vektors beschrieben wird, wird die Bedeutung explizit im Text angegeben.
- Komponenten eines Individuums werden in der Reihenfolge Individuum, Komponente als Index dargestellt, z. B.: beschreibt $I_{j,i}$ die i -te Komponente des j -ten Individuums
- Bereichseinschränkungen für Funktionen werden als Index vor dem Funktionssymbol angegeben, z. B.: $_{(c < 0)}f(x)$

Zitierte Gleichungen aus der Literatur wurden, falls notwendig, an diese Notation angepasst. Die folgende Liste der verwendeten Symbole ist alphabetisch geordnet, griechische Symbole sind vor den lateinischen eingeordnet. Sollte ein Symbol nur in einem speziellen Kontext oder Themenabschnitt gültig sein oder sollte ein Symbol mehrere Bedeutungen haben, so wird dies in der entsprechenden Erläuterung angegeben. Zusätzlich ist zu jedem Symbol der Abschnitt angegeben, in dem es definiert bzw. erstmalig verwendet wurde.

Symbol	Erläuterung	Kapitel
$\mathbf{0}$	n -dimensionaler Nullvektor	2.4, Seite 27
$\mathbf{1}$	$n \times m$ -dimensionale Einheitsmatrix	2.4, Seite 27
$\mathbf{1}$	n -dimensionaler Einheitsvektor	2.4, Seite 27
$\langle \rangle_m$	Modulofunktion bzgl. m	3.4-2, Seite 45
$\ \mathbf{v}\ $	Länge des Vektors \mathbf{v}	2.2, Seite 21

Symbol	Erläuterung	Kapitel
\sim	verteilt nach Beispiel für eine Zufallsvariable X , die normalverteilt ist: $X \sim N(0, 1)$	4.7-1, Seite 129
*	Faltung	4.7-6, Seite 134
$(\mu/1, \lambda)$ -CMA-ES	Evolutionstrategie mit μ Eltern, λ Nachkommen, intermediärer Rekombination aus allen Individuen und Kovarianzmatrixadaptation	3.3, Seite 41
$(1 + 1)$ -ES	Evolutionstrategie mit 1 Elter und 1 Nachkommen	2.1-6, Seite 20
$(1 + \lambda)$ -ES	Evolutionstrategie mit 1 Elter und λ Nachkommen	2.1-6, Seite 20
$(1, \lambda)$ -ES	Evolutionstrategie mit 1 Elter und λ Nachkommen	2.1-6, Seite 20
(μ, λ) -ES	Evolutionstrategie mit μ Eltern und λ Nachkommen	2.1-6, Seite 20
$(\mu + \lambda)$ -ES	Evolutionstrategie mit μ Eltern und λ Nachkommen	2.1-6, Seite 20
$(\mu/\rho, \lambda)$ -ES	Evolutionstrategie mit μ Eltern und λ Nachkommen und lokaler Selektion	3.2, Seite 37
$(\mu/1, \lambda)$ -KSA-ES	Evolutionstrategie mit μ Eltern, λ Nachkommen, keiner Rekombination und kumulativer Schrittweitenadaptation	3.2, Seite 37
$(\mu/1, \lambda)$ -KSA-ES	Evolutionstrategie mit μ Eltern, λ Nachkommen, intermediärer Rekombination und kumulativer Schrittweitenadaptation	3.2, Seite 37
$A(\sigma, c)$	asymmetrisch verteilte Zufallszahl mit Verteilungsfunktion $\Xi_{c, \sigma}$	5, Seite 149
$A_\gamma(\sigma, c)$	asymmetrisch verteilte Zufallszahl mit Verteilungsfunktion $\Xi_{c, \sigma}^\gamma$	5, Seite 149
α	Lagewinkelvektor (auch Lagewinkelparameter) eines Individuums I	2.1-2, Seite 17

Symbol	Erläuterung	Kapitel
α	<i>nur Simulated Binary Crossover:</i> Steuerparameter	3.9, Seite 55
α_i	i -te Komponente des Lagewinkelvektors α	2.1-2, Seite 17
\mathbf{a}	Lageparametervektor der n -dimensionalen Normalverteilung	2.1-1, Seite 14
a	Lageparameter der Normalverteilung	2.1-1, Seite 14
a_3	Schiefemaß einer Zufallsvariablen	4.1-6, Seite 69
β	<i>nur Evolutionary Gradient Search:</i> Steuerparameter zur Schrittweitenadaptation	3.5-2, Seite 48
β_i	<i>nur Simulated Binary Crossover:</i> Ausbreitungsfaktor	3.9-1, Seite 57
β_{q_i}	<i>nur Simulated Binary Crossover:</i> Ordinatenwert	3.9-1, Seite 57
\mathbf{C}	Kovarianzmatrix	2.1-1, Seite 14
$\mathbf{C}^{(t)}$	Kovarianzmatrix eines beliebigen Individuums der Generation t	2.4, Seite 27
c	Strategieparameter zur Steuerung der Asymmetrie	4.3, Seite 72
c	<i>nur kumulative Schrittweitenadaptation:</i> Kumulationszeitraum	3.2, Seite 37
$c_{1-\alpha}$	<i>nur Kolmogorov-Smirnov-Test:</i> Testkriterium mit Signifikanz α	5.1-1, Seite 150
c_d	<i>nur klassische Evolutionsstrategien:</i> Schrittweitenkontrolle der (1 + 1)-ES	2.2, Seite 21
c_i	<i>nur klassische Evolutionsstrategien:</i> Schrittweitenkontrolle der (1 + 1)-ES	2.2, Seite 21
$c_{jk}(\mathbf{x})$	Elemente der Kovarianzmatrix	2.1-1, Seite 14
$c_{ii}(\mathbf{x})$	Varianzen der Kovarianzmatrix	2.1-1, Seite 14
c_r	<i>nur Differentielle Evolution:</i> Rekombinationsrate	3.4-2, Seite 45
c_u	<i>nur kumulative Schrittweitenadaptation:</i> Normierungsfaktor der Varianzen	3.2, Seite 37

Symbol	Erläuterung	Kapitel
co_i	nur Lokale Selektion: Operator zur Auswahl der Eltern	3.7-1, Seite 53
$cov(x_i, x_j)$	Kovarianz der Komponenten i und j des n -dimensionalen Zufallsvektors \mathbf{x}	2.1-1, Seite 14
$\delta^{(t)}$	nur kumulative Schrittweitenadaptation: Schrittweite der Generation t	3.2-3, Seite 40
$\delta_i^{(t)}$	nur kumulative Schrittweitenadaptation: Schrittweite des i -ten Individuums der Generation t	3.2-2, Seite 38
D	nur kumulative Schrittweitenadaptation: Dämpfungsparameter	3.2, Seite 37
D_n	charakteristischer Wert des Kolmogorov-Smirnov-Test	5.1-1, Seite 150
$det()$	Determinante einer Matrix	2.1-1, Seite 14
E_X	Erwartungswert der Zufallsvariablen X	2.1-1, Seite 14
$E_X _a^b$	partieller Erwartungswert	4.1-4, Seite 67
E_{X^2}	2. Moment einer Zufallsvariable (auch m_2)	4.1-5, Seite 68
E_X^2	Quadrat des 1. Momentes (Erwartungswert) einer Zufallsvariable (auch m_1^2)	4.1-5, Seite 68
E_X^3	3. Potenz des 1. Momentes (Erwartungswert) einer Zufallsvariable (auch m_1^3)	4.1-6, Seite 69
\mathbf{e}	Einheitsvektor	3.5-1, Seite 47
e^x	Exponentialfunktion	4.3, Seite 72
$erf(x)$	Fehlerfunktion	4.3-1, Seite 76
ε	nur Robust Evolution Strategies: untere Schranke für robuste Adaptation	3.6-2, Seite 50
$\zeta(x)$	Riemannsche Zeta-Funktion	4.6-1, Seite 114
ζ	nur Evolutionary Gradient Search: Steuerparameter	3.5-1, Seite 47

Symbol	Erläuterung	Kapitel
ζ	nur 1/5-Erfolgsregel: Verhältnis von erfolgreichen zu allen Mutationen	2.2-2, Seite 23
ζ_i	nur kumulative Schrittweitenadaptation: Index eines Individuums, das zur entstochastisierten Mutation gewählt wurde	3.2, Seite 37
F	nur Differentielle Evolution: Skalierungsfaktor	3.4-1, Seite 44
$(c < 0)f(x)$	Funktion $f(x)$ für $c < 0$	4.3, Seite 72
$(c \geq 0)f(x)$	Funktion $f(x)$ für $c \geq 0$	4.3, Seite 72
η_k	Standardisiertes k -tes Moment einer Verteilungsfunktion	4.1-3, Seite 66
η_3	Koeffizient der Schiefe (auch standardisiertes 3 -tes Moment einer Verteilungsfunktion)	4.1-3, Seite 66
η_4	Koeffizient der Kurtosis (auch standardisiertes 4 -tes Moment einer Verteilungsfunktion)	4.1-3, Seite 66
η_i	nur Evolutionary Gradient Search: Steuerparameter des Individuums i	3.5-2, Seite 48
η	nur Evolutionary Gradient Search: Schrittweitenvektor	3.5-2, Seite 48
η_i	nur Evolutionary Gradient Search: i -te Komponente des Schrittweitenvektor	3.5-2, Seite 48
γ	nur Rekombination: Kontrollparameter der intermediären Rekombination	2.1-5, Seite 18
γ	globaler Steuerparameter der asymmetrischen Verteilungsfunktion	4.5, Seite 102
γ_i	nur Simulated Binary Crossover: Steuerparameter	3.9, Seite 55
γ_{Euler}	Euler'sche Konstante	4.6-1, Seite 114
G_n	gleichverteilte, diskrete Zufallsvariable über $[1, n]$	2.1-5, Seite 18
$G(a, b)$	gleichverteilte, kontinuierliche Zufallsvariable über $[a, b]$	5.1-1, Seite 150

Symbol	Erläuterung	Kapitel
$\tilde{\mathbf{g}}^{(t)}$	<i>nur Evolutionary Gradient Search:</i> geschätzter Gradient	3.5-1, Seite 47
$g_j(\mathbf{x})$	allgemeine Nebenbedingung	2.2, Seite 21
\mathbf{I}	Individuum einer Evolutionsstrategie	2.1-2, Seite 17
$\tilde{\mathbf{I}}$	<i>nur Differentielle Evolution:</i> Tastindividuum	3.4-1, Seite 44
$\mathbf{I}^{(t)}$	beliebiges Individuum der Generation t	2.4, Seite 27
$I_{j,i}^{(t)}$	i -te Komponente des j -ten Individuums der Generation t	2.4, Seite 27
$\mathbf{I}_{sel}^{(t)}$	<i>nur kumulative Schrittweitenadaptation:</i> Menge der in Generation t selektierten μ Individuen	3.2-3, Seite 40
\inf_x	Infimum über alle x	5.2-1, Seite 154
$I_i(x)$	Indexfunktion der Kompositionsmethode	5.2-2, Seite 155
κ	Abbruchkriterium aller ES-Varianten	2.2, Seite 21
κ	<i>nur Lokale Selektion:</i> maximale Lebenserwartung eines Elters	3.7, Seite 52
k_{circle}	<i>nur Adaptation einer nicht-zentralen Normalverteilung:</i> Überlappungsindex	3.8, Seite 54
$k_{corridor}$	<i>nur Adaptation einer nicht-zentralen Normalverteilung:</i> Überlappungsindex	3.8, Seite 54
λ	Anzahl der Nachkommen einer ES	2.4, Seite 27
λ	<i>nur Differentielle Evolution:</i> Steuerparameter zur Fokussierung	3.4-2, Seite 45
L	<i>nur Differentielle Evolution:</i> Länge des Rekombinationsbereiches	3.4-2, Seite 45
$LN(a, \sigma)$	logarithmisch normalverteilte Zufallsvariable mit Parametern a und σ	2.4, Seite 27

Symbol	Erläuterung	Kapitel
$\lim_{x \rightarrow x_0 - 0} f(x)$	linksseitiger Grenzwert der Funktion $f(x)$ an der Stelle x_0	4.7-4, Seite 131
$\lim_{x \rightarrow x_0 + 0} f(x)$	rechtsseitiger Grenzwert der Funktion $f(x)$ an der Stelle x_0	4.7-4, Seite 131
μ	Anzahl der Eltern einer ES	2.4, Seite 27
μ_k	k -tes zentrales Moment einer Zufallsvariablen	4.1-3, Seite 66
μ_2	Varianz (2. zentrales Moment) einer Zufallsvariablen	4.1-5, Seite 68
mod	Modulofunktion	5.1-1, Seite 150
mut	Mutationsoperator	2.1-4, Seite 18
m_k	k -tes Moment einer Zufallsvariablen	4.1-3, Seite 66
m_1	1 -tes Moment (Erwartungswert) einer Zufallsvariablen	4.1-3, Seite 66
m_2	2 -tes Moment einer Zufallsvariablen	4.1-6, Seite 69
m_3	3 -tes Moment einer Zufallsvariablen	4.1-6, Seite 69
m_1^2	Quadrat des 1. Momentes (Erwartungswert) einer Zufallsvariable (auch E_X^2)	4.1-6, Seite 69
m_1^3	3. Potenz des 1. Momentes (Erwartungswert) einer Zufallsvariable (auch E_X^3)	4.1-6, Seite 69
v	nur Lokale Selektion: Größe der Nachbarschaft	3.7, Seite 52
v_i	nur Lokale Selektion: Nachbarschaft des Individuums i	3.7, Seite 52
$N(a, \sigma)$	Normalverteilung mit Erwartungswert a und Standardabweichung σ	2.1-1, Seite 14
$\xi_c(x)$	asymmetrische Dichtefunktion einer Zufallsvariablen mit Steuerparameter c	4.3, Seite 72
$(c < 0)\xi_c(x)$	abschnittsweise definierte, asymmetrische Dichtefunktion einer Zufallsvariablen mit Steuerparameter $c < 0$	4.3, Seite 72

Symbol	Erläuterung	Kapitel
${}_{(c \geq 0)}\xi_c(x)$	abschnittsweise definierte, asymmetrische Dichtefunktion einer Zufallsvariablen mit Steuerparameter $c \geq 0$	4.3, Seite 72
$\xi_{c, \sigma}(x)$	asymmetrische Dichtefunktion einer Zufallsvariablen mit den Steuerparametern c und σ (siehe auch ${}_{(c < 0)}\xi_c(x)$ und ${}_{(c \geq 0)}\xi_c(x)$)	4.4, Seite 90
$\xi_{c, \sigma}^\gamma(x)$	asymmetrische Dichtefunktion einer Zufallsvariablen mit den Steuerparametern c , σ und γ (siehe auch ${}_{(c < 0)}\xi_c(x)$ und ${}_{(c \geq 0)}\xi_c(x)$)	4.5, Seite 102
$\xi_1^-(t)$	abkürzende Schreibweise für das Faltungsintegral $\xi_{c_1, \sigma_1}^\gamma(t)$ für $t < 0$	4.7-6, Seite 134
$\xi_1^+(t)$	abkürzende Schreibweise für das Faltungsintegral $\xi_{c_1, \sigma_1}^\gamma(t)$ für $t \geq 0$	4.7-6, Seite 134
$\Xi_c(x)$	asymmetrische Verteilungsfunktion einer Zufallsvariablen mit den Steuerparametern c und σ (siehe auch ${}_{(c < 0)}\xi_c(x)$ und ${}_{(c \geq 0)}\xi_c(x)$)	4.3-1, Seite 76
$\Xi_{c, \sigma}(x)$	asymmetrische Verteilungsfunktion einer Zufallsvariablen mit den Steuerparametern c und σ (siehe auch ${}_{(c < 0)}\xi_c(x)$ und ${}_{(c \geq 0)}\xi_c(x)$)	4.4-1, Seite 92
$\Xi_{c, \sigma}^\gamma(x)$	asymmetrische Verteilungsfunktion einer Zufallsvariablen mit den Steuerparametern c , σ und γ (siehe auch ${}_{(c < 0)}\xi_c(x)$ und ${}_{(c \geq 0)}\xi_c(x)$)	4.5-1, Seite 103
$\Xi_c(y)$	inverse, asymmetrische Verteilungsfunktion einer Zufallsvariablen mit Steuerparameter c (siehe auch ${}_{(c < 0)}\xi_c(x)$ und ${}_{(c \geq 0)}\xi_c(x)$)	4.3-3, Seite 80
$\Xi_{c, \sigma}(y)$	inverse, asymmetrische Verteilungsfunktion einer Zufallsvariablen mit den Steuerparametern c und σ (siehe auch ${}_{(c < 0)}\xi_c(x)$ und ${}_{(c \geq 0)}\xi_c(x)$)	4.4-3, Seite 97
$\Xi_{c, \sigma}^\gamma(y)$	inverse, asymmetrische Verteilungsfunktion einer Zufallsvariablen mit den Steuerparametern c , σ und γ (siehe auch ${}_{(c < 0)}\xi_c(x)$ und ${}_{(c \geq 0)}\xi_c(x)$)	4.5-3, Seite 108

Symbol	Erläuterung	Kapitel
O_{dub}	<i>nur Robust Evolution Strategies:</i> Duplikationsoperator	3.6-2, Seite 50
O_{del}	<i>nur Robust Evolution Strategies:</i> Deletationsoperator	3.6-2, Seite 50
O_{inv}	<i>nur Robust Evolution Strategies:</i> Inversionsoperator	3.6-2, Seite 50
$\Phi_X(x)$	Verteilungsfunktion der Zufallsvariablen X	2.1-1, Seite 14
$\Phi_{\mathbf{X}}(\mathbf{x})$	Verteilungsfunktion des Zufallsvektors \mathbf{X}	2.1-1, Seite 14
$\Phi_t(x)$	<i>nur Robust Evolution Strategies:</i> allgemeine Cauchy-Verteilung mit Parameter t	3.6-1, Seite 49
$\Phi_1(x)$	<i>nur Robust Evolution Strategies:</i> Cauchy-Verteilung mit Parameter $t = 1$	3.6-1, Seite 49
$\Phi_{\text{Gumbel}}(x)$	Gumbel-Verteilung	4.6-1, Seite 114
$\Phi_{G(a,b)}(x)$	kontinuierliche Gleichverteilung über $[a, b]$	5.1-1, Seite 150
Φ	inverse Verteilungsfunktion	5.2-1, Seite 154
$\Phi_n(\bar{x})$	<i>nur Kolmogorov-Smirnov-Test:</i> empirische Verteilungsfunktion	5.1-1, Seite 150
$\varphi(\beta)$	<i>nur Simulated Binary Crossover:</i> Wahrscheinlichkeitsdichtefunktion	3.9-1, Seite 57
$\varphi_X(x)$	Dichtefunktion der Zufallsvariablen X	2.1-1, Seite 14
$\varphi_{\mathbf{X}}(\mathbf{x})$	Dichtefunktion des Zufallsvektors \mathbf{X}	2.1-1, Seite 14
$(c < 0) \varphi(x)$	Funktion $\varphi(x)$ für $c < 0$	4.3, Seite 72
$(c \geq 0) \varphi(x)$	Funktion $\varphi(x)$ für $c \geq 0$	4.3, Seite 72
$\varphi_t(x)$	<i>nur Robust Evolution Strategies:</i> Dichtefunktion der allgemeinen Cauchy-Verteilung mit Parameter t	3.6-1, Seite 49
$\varphi_1(x)$	<i>nur Robust Evolution Strategies:</i> Dichtefunktion der Cauchy-Verteilung mit Parameter $t = 1$	3.6-1, Seite 49

Symbol	Erläuterung	Kapitel
$\varphi_{\text{Gumbel}}(x)$	Dichtefunktion einer Gumbel-verteilten Zufallszahl	4.6-1, Seite 114
$\varphi_{\text{Cauchy}}(x)$	Dichtefunktion einer Cauchy-verteilten Zufallszahl	4.6-3, Seite 127
$\varphi_{\text{Fehler}}(x)$	Dichtefunktion einer Fehler-verteilten Zufallszahl	4.6-3, Seite 127
$\varphi_{\text{Logistisch}}(x)$	Dichtefunktion einer logistisch verteilten Zufallszahl	4.6-3, Seite 127
$\varphi_{\text{Student}}(x)$	Dichtefunktion einer Student's t-verteilten Zufallszahl	4.6-3, Seite 127
$\varphi_{G(a,b)}(x)$	Dichtefunktion einer über $[a, b]$ gleichverteilten Zufallsvariablen	5.1-1, Seite 150
φ	Erwartungswert der Fortschrittsgeschwindigkeit	2.2-2, Seite 23
$\varphi_{\text{Korridor}}$	Erwartungswert der Fortschrittsgeschwindigkeit für das Korridor-Modell	2.2-2, Seite 23
φ_{Kugel}	Erwartungswert der Fortschrittsgeschwindigkeit für das Kugel-Modell	2.2-2, Seite 23
$\Psi(\mathbf{I})$	Fitness des Individuums \mathbf{I}	2.1-3, Seite 18
$P(E)$	Wahrscheinlichkeit für Ereignis E	2.1-1, Seite 14
$P^{(t)}$	Population der Generation t	2.1, Seite 14
$P^{(0)}$	Initialpopulation	2.1, Seite 14
q	nur Kovarianzmatrixadaptation: Dämpfungsparameter	3.3, Seite 41
q_α	α -Quantil einer Zufallsvariable	4.1-2, Seite 65
ρ	nur Lokale Selektion: Anzahl der Eltern pro Nachkomme	3.7, Seite 52
ρ	Reversibilität	6.2-3, Seite 177
\mathbb{R}	Menge der reellen Zahlen	2.1, Seite 14
\mathbb{R}_+	Menge der positiven reellen Zahlen ohne 0	2.1, Seite 14
\mathbb{R}^n	n -dimensionaler Körper der reellen Zahlen	2.1, Seite 14

Symbol	Erläuterung	Kapitel
\mathbb{R}_+^n	n -dimensionaler Körper der positiven reellen Zahlen ohne 0	2.1, Seite 14
$r(x)$	Dichtefunktion der Akzeptanz/Ablehnungsmethode	5.2-4, Seite 156
r_1, r_2, r_3	nur Differentielle Evolution: Indizes der Tastindividuen	3.4-1, Seite 44
rek	Rekombinationsoperator	2.1-5, Seite 18
rek^v	nur Lokale Selektion: Rekombinationsoperator	3.7-1, Seite 53
re	nur Lokale Selektion: Operator zur Generierung von Nachkommen	3.7-1, Seite 53
σ_X	Standardabweichung der Zufallsvariablen X	2.1-1, Seite 14
σ	Schrittweitenvektor (auch Schrittweitenparameter) eines Individuums \mathbf{I}	2.1-2, Seite 17
σ_i	i -te Komponente des Schrittweitenvektors σ	2.1-2, Seite 17
$\sigma_{i,j}$	j -te Komponente des Schrittweitenvektors σ_i des i -ten Individuums	2.4-1, Seite 27
$\sigma_{i,j}^{(t)}$	j -te Komponente des Schrittweitenvektors σ_i des i -ten Individuums der Generation t	2.4-1, Seite 27
σ_{opt}	optimale Schrittweite	2.2-2, Seite 23
σ^*	nur Robust Evolution Strategies: aktive Strategiekomponente	3.6-2, Seite 50
σ_i^*	nur Robust Evolution Strategies: i -te Komponente der aktiven Strategiekomponente	3.6-2, Seite 50
σ_{max}	nur Robust Evolution Strategies: obere Schranke für robuste Adaptation	3.6-2, Seite 50
SD_X	Standardabweichung einer Zufallsvariablen	2.1-1, Seite 14
$\mathbf{s}^{(t)}$	nur kumulative Schrittweitenadaptation: Evolutionsspfad der Generation t	3.2-3, Seite 40

Symbol	Erläuterung	Kapitel
$\mathbf{s}_i^{(t)}$	nur kumulative Schrittweitenadaptation: Evolutionspfad des i -ten Individuums der Generation t	3.2-2, Seite 38
sel	Selektionsoperator	2.1-6, Seite 20
sel_i	nur Lokale Selektion: Selektionsoperator	3.7-2, Seite 53
\sup_x	Supremum über alle x	5.1-1, Seite 150
τ	Steuerparameter der ES	2.4, Seite 27
$t(x)$	nur Akzeptanz/Ablehnungs-Methode: majorisierende Funktion	5.2-4, Seite 156
\mathbf{u}	nur Differentielle Evolution: Rekombinationsvektor	3.4-1, Seite 44
u_i	nur Simulated Binary Crossover: $G(0, 1)$ -verteilte Zufallsvariable	3.9, Seite 55
V	nur Simulated Binary Crossover: Elternintervall	3.9, Seite 55
\mathbf{v}^T	Transponierte des Vektors \mathbf{v}	2.1, Seite 14
V_X	Varianz der Zufallsvariablen X	2.1-1, Seite 14
V_X^3	3. Potenz der Varianz der Zufallsvariablen X	4.3-6, Seite 88
w	Anzahl der Lagewinkel bei korrelierter Mutation	2.1-2, Seite 17
X	allgemeine Zufallsvariable	2.1-1, Seite 14
\bar{X}	nur Testen von Zufallszahlen: erzeugte Zufallszahl	5.1, Seite 149
\mathbf{X}	allgemeiner Zufallsvektor	2.1, Seite 14
χ_n	Erwartungswert der Länge eines $(\mathbf{0}, \mathbf{1})$ -normalverteilten Zufallsvektors	3.2, Seite 37
χ^2	Testwert für die Nullhypothese H_0 des χ^2 -Tests	5.1-1, Seite 150
$\chi_{k-1, 1-\alpha}^2$	Kriterium für die Nullhypothese H_0 des χ^2 -Tests bei $k-1$ Freiheitsgraden und Signifikanz α	5.1-1, Seite 150

Symbol	Erläuterung	Kapitel
$\chi^2(n)$	χ^2 -Verteilung mit n Freiheitsgraden	2.2-2, Seite 23
\mathbf{x}	Objektvektor (auch Objektparameter) eines Individuums \mathbf{I}	2.1-2, Seite 17
$\tilde{\mathbf{x}}$	nur Differentielle Evolution: Objektkomponente des Tastindividuum	3.4-1, Seite 44
$\mathbf{x}^{(t)}$	Objektvektor eines beliebigen Individuums der Generation t	2.4, Seite 27
$\langle \mathbf{x} \rangle_{\mu}^{(t)}$	nur kumulative Schrittweitenadaptation: Schwerpunkt der in Generation t selektierten μ Individuen	3.2-3, Seite 40
x_i	i -te Komponente der Objektvektors \mathbf{x}	2.1-2, Seite 17
$x_{j,i}^{(t)}$	i -te Komponente des Objektvektors des j -ten Individuums der Generation t	2.4, Seite 27
Y	allgemeine Zufallsvariable	2.1, Seite 14
Z_i	nur Zufallszahlengeneratoren: mittels eines PMMLC-Generators im i -ten Schritt erzeugte gleichverteilte, diskrete Zufallszahl	5.1-1, Seite 150
Z_i	nur Kovarianzmatrixadaptation: (0, 1)-normalverteilte Zufallsvariablen	3.3, Seite 41
\mathbf{z}	allgemeiner Zufallsvektor	2.1, Seite 14
z	allgemeiner Zufallsvariable	2.1, Seite 14
z	nur (μ, λ)- und $(\mu + \lambda)$-ES Zufallsvariable der globalen Lernrate für die Schrittweiten	2.4, Seite 27
z^*	nur (μ, λ)-aES Zufallsvariable der globalen Lernrate für die Schiefemaße	6.1-4, Seite 169
$\mathbf{z}_i^{(t)}$	nur kumulative Schrittweitenadaptation: normalverteilter Zufallsvektor mit Erwartungswert $\mathbf{0}$ und Standardabweichung $\mathbf{1}$, zur Mutation des i -ten Individuums der Generation t	3.2, Seite 37

Symbol	Erläuterung	Kapitel
$\mathbf{z}_{sel}^{(i)}$	nur Kovarianzmatrixadaptation: Vektor, der in der Generation i das Individuum erzeugt hat, das selektiert wurde	3.3, Seite 41
z_i	i -te Komponente des Zufallsvektors \mathbf{z}	2.1, Seite 14
z_i	nur (μ, λ)- und $(\mu + \lambda)$-ES Zufallsvariable der lokalen Lernrate für die Schrittweiten	2.4, Seite 27
z_i^*	nur (μ, λ)-aES Zufallsvariable der lokalen Lernrate für die Schiefeße	6.1-4, Seite 169

Anhang B Abkürzungsverzeichnis

In diesem Abschnitt werden die verwendeten Abkürzungen, die nicht zum allgemeinen Gebrauch der deutschen Sprache gehören, erläutert. In der Spalte Kapitel wird zudem die Definition bzw. die erste Verwendung der entsprechenden Abkürzung angegeben.

Abkürzung	Erläuterung	Kapitel
aES	asymmetrische Evolutionsstrategie	6.1, Seite 167
BBSC	Bit Based Simulated Crossover	3.10, Seite 59
BLX	Blend-Crossover	3.9, Seite 55
BMDA	Bivariate Marginal Distribution Algorithm	3.10, Seite 59
BOA	Bayesian Optimization Algorithm	3.10, Seite 59
cGA	Compact Genetic Algorithm	3.10, Seite 59
CMA	Kovarianzmatrixadaptation	3.3, Seite 41
(μ/λ) -CMA-ES	Evolutionsstrategie mit Kovarianzmatrixadaptation, μ Eltern, λ Nachkommen und globaler Rekombination	3.3-2, Seite 42
DE	Differentielle Evolution	3.4, Seite 44
ECGA	Extended Compact Genetic Algorithm	3.10, Seite 59
EDA	Estimation Distribution Algorithm	3.10, Seite 59
EGS	Evolutionary Gradient Search	3.5, Seite 46
ES	Evolutionsstrategie	2.1, Seite 14
FDA	Factorized Distribution Algorithm	3.10, Seite 59
FES	Fast Evolution Strategies	3.6, Seite 49
FEP	Fast Evolutionary Programming	3.6, Seite 49
GA	Genetische Algorithmen	3, Seite 35
IFEP	Improved Fast Evolutionary Programming	3.11, Seite 61

Abkürzung	Erläuterung	Kapitel
KSA	kumulative Schrittweitenregelung	3.2, Seite 37
KSA-ES	Evolutionstrategie mit kumulativer Schrittweitenregelung	3.2, Seite 37
$(\mu/1, \lambda)$ -KSA-ES	Evolutionstrategie mit kumulativer Schrittweitenregelung, μ Eltern, λ Nachkommen und keiner Rekombination	3.2-2, Seite 38
$(\mu/\mu, \lambda)$ -KSA-ES	Evolutionstrategie mit kumulativer Schrittweitenregelung, μ Eltern, λ Nachkommen und globaler Rekombination	3.2-3, Seite 40
LES	Lokale Selektion	3.7, Seite 52
MIMIC	Mutual Information Maximizing Input Clustering	3.10, Seite 59
NFL	No Free Lunch-Theorem	1, Seite 9
PBIL	population based incremental learning	3.10, Seite 59
PMMLC	prime modulus multiplicative linera congruential	5.1, Seite 149
RES	Robust Evolution Strategies	3.6, Seite 49
SBX	Simulated Binary Crossover	3.9, Seite 55
UMDA	Univariate Marginal Distribution Algorithm	3.10, Seite 59
VLSI	Very Large Scale Integration	5.1, Seite 149

Anhang C Abbildungsverzeichnis

Abb. 2.1:	Normalverteilung für unterschiedliche σ	16
Abb. 2.2:	Adaptation einer Schrittweite	30
Abb. 2.3:	Adaptation n Schrittweiten	32
Abb. 2.4:	Adaptation von Schrittweiten und Lagewinkeln	33
Abb. 3.1:	Verlauf der von Rudolph untersuchten Dichtefunktionen	36
Abb. 3.2:	Typische Evolutionspfade bei der kumulativen Schrittweitenadaptation	38
Abb. 3.3:	Konstruktion der Mutationsverteilung für Generation vier	42
Abb. 3.4:	Tastschritt der Differentiellen Evolution	45
Abb. 3.5:	Rekombination der Differentiellen Evolution	46
Abb. 3.6:	Schätzung des Gradienten in der Evolutionary-Gradient-Search	48
Abb. 3.7:	Verlauf von Normalverteilung (blau) und Cauchy-Verteilung (rot)	50
Abb. 3.8:	Arbeitsweise des SBX-Operators	58
Abb. 3.9:	Stochastisches Modell mit unabhängigen Variablen	60
Abb. 3.10:	Stochastische Modelle mit paarweise abhängigen Variablen	60
Abb. 3.11:	Stochastische Modelle mit mehrfach abhängigen Variablen	61
Abb. 4.1:	Verlauf der asymmetrischen Funktion $f(x)$	74
Abb. 4.2:	Verlauf der Wahrscheinlichkeitsdichtefunktion φ für unterschiedliche c	76
Abb. 4.3:	Dichtefunktion $\xi_c(x)$ für variierende c	76
Abb. 4.4:	Verlauf der Verteilungsfunktion $\Xi_c(x)$ für variierende c	78
Abb. 4.5:	Verlauf der inversen Verteilungsfunktion $\bar{\Xi}_c(y)$	82
Abb. 4.6:	Erwartungswert einer $\Xi_c(x)$ -verteilten Zufallsvariablen	83
Abb. 4.7:	Varianz einer $\Xi_c(x)$ -verteilten Zufallsvariablen	87
Abb. 4.8:	Standardabweichung einer $\Xi_c(x)$ -verteilten Zufallsvariablen	87
Abb. 4.9:	Verlauf des Schiefemaßes η_3 in Abhängigkeit von c	90
Abb. 4.10:	Verlauf des Schiefemaßes m_3	90

Abb. 4.11: Verlauf der Dichtefunktion $\xi_{c,\sigma}(x)$ für unterschiedliche σ	92
Abb. 4.12: Dichtefunktion $\xi_{c,\sigma}(x)$ für variierende σ	92
Abb. 4.13: Verlauf der Verteilungsfunktion $\Xi_{c,\sigma}(x)$ für unterschiedliche σ	94
Abb. 4.14: Verlauf der inversen Verteilungsfunktion $\bar{\Xi}_{c,\sigma}(y)$ für unterschiedliche σ	98
Abb. 4.15: Verlauf des Erwartungswertes einer $\Xi_{c,\sigma}(x)$ -verteilten Zufallsvariablen	99
Abb. 4.16: Varianz einer $\Xi_{c,\sigma}(x)$ -verteilten Zufallsvariablen	100
Abb. 4.17: Standardabweichung einer $\Xi_{c,\sigma}(x)$ -verteilten Zufallsvariablen	100
Abb. 4.18: Verlauf des Schiefemaßes η_3	101
Abb. 4.19: Verlauf des Schiefemaßes m_3	102
Abb. 4.20: Verlauf der Dichtefunktion $\xi_{c,\sigma}^\gamma(x)$ für unterschiedliche γ	103
Abb. 4.21: Verlauf der Verteilungsfunktion $\Xi_{c,\sigma}^\gamma(x)$	105
Abb. 4.22: Verlauf der inversen Verteilungsfunktion $\bar{\Xi}_{c,\sigma}^\gamma(y)$ für variierende γ	109
Abb. 4.23: Verlauf des Erwartungswertes einer $\Xi_{c,\sigma}^\gamma(x)$ -verteilten Zufallsvariablen	110
Abb. 4.24: Varianz einer $\Xi_{c,\sigma}^\gamma(x)$ -verteilten Zufallsvariablen	111
Abb. 4.25: Standardabweichung einer $\Xi_{c,\sigma}^\gamma(x)$ -verteilten Zufallsvariablen	112
Abb. 4.26: Verlauf des Schiefemaßes η_3 für unterschiedliche γ	113
Abb. 4.27: Verlauf des Schiefemaßes m_3	113
Abb. 4.28: Verlauf der Dichtefunktion einer Gumbel-verteilten Zufallsvariablen	115
Abb. 4.29: Dichtefunktionen der Beta-Verteilung mit $p, q > 1$ und $p=q$ (links) und $p, q < 1$ (rechts)	118
Abb. 4.30: Dichtefunktionen der Beta-Verteilung mit $p, q > 1$ (links) und $p > 1, q < 1$ (rechts)	118
Abb. 4.31: Dichtefunktionen der Typ II Verteilung mit $p, q > 1$ (links) und $p, q < 1$ (rechts)	119
Abb. 4.32: Dichtefunktionen der Gamma-Verteilungen mit variierendem Parameter λ	120
Abb. 4.33: Dichtefunktionen der Typ IV Verteilungen mit variierendem Parametern	121
Abb. 4.34: Dichtefunktionen der Typ V Verteilungen mit variierenden Parametern	122
Abb. 4.35: Dichtefunktionen der Typ VI Verteilungen mit variierenden Parametern	122
Abb. 4.36: Dichtefunktionen der Typ VI Verteilungen mit variierenden Parametern	123
Abb. 4.37: Dichtefunktionen der Student's t-Verteilung mit variierenden Parametern	124
Abb. 4.38: Dichtefunktionen der Typ VIII Verteilungen mit variierenden Parametern	124
Abb. 4.39: Dichtefunktionen der Typ IX Verteilungen mit variierenden Parametern	125
Abb. 4.40: Dichtefunktionen der Typ XI Verteilungen mit variierenden Parametern	125
Abb. 4.41: Dichtefunktionen der Typ XII Verteilungen mit variierenden Parametern	126
Abb. 4.46: Darstellung der zu faltenden Funktionen $\xi_{c1,\sigma1}^\gamma(t)$ und $\xi_{c2,\sigma2}^\gamma(t)$	136
Abb. 4.47: Dichtefunktion zweier symmetrischer Zufallsvariablen (links) sowie die Dichtefunktion der Summe (rechts)	141

Abb. 4.48: Dichtefunktion einer links- und einer rechtsschiefen Zufallsvariablen mit identischem Betrag des Schiefemaßes (links) sowie die Dichtefunktion der Summe (rechts)	142
Abb. 4.49: Dichtefunktion zweier rechtsschiefer Zufallsvariablen (links) sowie die Dichtefunktion der Summe (rechts)	143
Abb. 4.50: Dichtefunktion zweier linksschiefer Zufallsvariablen (links) sowie die Dichtefunktion der Summe (rechts)	143
Abb. 4.51: Dichtefunktion zweier rechtsschiefer Zufallsvariablen (links) sowie die Dichtefunktion der Summe (rechts)	143
Abb. 5.1: Verlauf der Funktionen $\Phi_0(x)$ (rot) und $\operatorname{erf}(x)$ (blau)	159
Abb. 5.2: Verlauf der Verteilungsfunktion $\Xi_{c,\sigma}^\gamma(x)$	162
Abb. 5.3: Verlauf der inversen Verteilungsfunktionen $\Phi_1(y)$ und $\Phi_2(y)$	163
Abb. 6.6: Definition der Variablen zur Steuerung der Schiefemaße	176
Abb. 6.7: Position von Elter und Nachkomme	177
Abb. 6.8: Reversibilität der nicht-zentralen Normalverteilung mit Lageparameter	178
Abb. 6.9: Reversibilität der -Verteilung mit in Abhängigkeit des Erwartungswertes	180
Abb. 6.10: Reversibilitätsfenster bei 15 Eltern und 100 Nachkommen	180
Abb. 6.11: Grenze des Reversibilitätsfenster für die Adaptation einer nicht-zentralen Normalverteilung	181
Abb. 6.12: Grenze des Reversibilitätsfenster für die -Verteilung	181
Abb. 6.13: Mutationsraum für unterschiedliche Belegungen des globalen Parameters γ	182
Abb. 7.1: Verlauf des Kugelmodells f_1 für $n=2$ Dimensionen	187
Abb. 7.2: Experimentelle Resultate zur Testfunktion f_1	188
Abb. 7.3: Experimentelle Resultate zur Testfunktion f_1 mit Fehlinitialisierung	189
Abb. 7.4: Verlauf des Kugelmodells f_2 mit projizierten Koordinaten	189
Abb. 7.5: Experimentelle Resultate zur Testfunktion f_2	190
Abb. 7.6: Experimentelle Resultate zur Testfunktion f_2 mit Fehlinitialisierung	190
Abb. 7.7: Verlauf des Kugelmodells f_3 mit projizierten Koordinaten	191
Abb. 7.8: Experimentelle Resultate zur Testfunktion f_3	191
Abb. 7.9: Experimentelle Resultate zur Testfunktion f_3 mit Fehlinitialisierung	192
Abb. 7.10: Verlauf des Kugelmodells f_4 mit projizierten Koordinaten	192
Abb. 7.11: Experimentelle Resultate zur Testfunktion f_4	193
Abb. 7.12: Experimentelle Resultate zur Testfunktion f_4 mit Fehlinitialisierung	193
Abb. 7.13: Verlauf des Kugelmodells f_5 mit projizierten Koordinaten	194
Abb. 7.14: Experimentelle Resultate zur Testfunktion f_5	194

Abb. 7.15: Experimentelle Resultate zur Testfunktion f_5 mit Fehlinitialisierung	195
Abb. 7.16: Verlauf des Kugelmodells f_6 mit projizierten Koordinaten	195
Abb. 7.17: Experimentelle Resultate zur Testfunktion f_6	196
Abb. 7.18: Experimentelle Resultate zur Testfunktion f_6 mit Fehlinitialisierung	196
Abb. 7.19: Experimentelle Resultate zur Testfunktion B_1	198
Abb. 7.20: Experimentelle Resultate zur Testfunktion B_2	199
Abb. 7.21: Experimentelle Resultate zur Testfunktion B_4	200
Abb. 7.22: Verlauf der Rosenbrock-Funktion B_5	201
Abb. 7.23: Experimentelle Resultate zur Testfunktion B_5	201
Abb. 7.24: Verlauf der Ellipse B_6	202
Abb. 7.25: Experimentelle Resultate zur Testfunktion B_6	202
Abb. 7.26: Verlauf der Zigarre B_7 mit Skalierung	203
Abb. 7.27: Verlauf der Zigarre B_7 ohne Skalierung	203
Abb. 7.28: Experimentelle Resultate zur Testfunktion B_7	204
Abb. 7.29: Verlauf der Tablette B_8 , Skalierung der y -Achse	205
Abb. 7.30: Verlauf der Tablette B_8	205
Abb. 7.31: Experimentelle Resultate zur Testfunktion B_8	205
Abb. 7.32: Verlauf der Summe verschiedener Potenzen B_9	206
Abb. 7.33: Experimentelle Resultate zur Testfunktion B_9	207
Abb. 7.34: Verlauf des Parabelgrates B_{10}	208
Abb. 7.35: Experimentelle Resultate zur Testfunktion B_{10}	208
Abb. 7.36: Verlauf des spitzen Grates B_{11}	209
Abb. 7.37: Experimentelle Resultate zur Testfunktion B_{11}	209
Abb. 7.38: Ebenenmodell für hohe Dimensionen	211
Abb. 7.39: Kugelmodell für hohe Dimensionen	212
Abb. 7.40: Doppelsumme für hohe Dimensionen	213
Abb. 7.41: Informationstragende Bereiche einer Punktschweißverbindung	214
Abb. 7.42: Schweißpunkte mit guter (links) und mangelhafter (rechts) Qualität	215
Abb. 7.43: Definition der unscharfen Menge blau	215
Abb. 7.44: Häufigkeitsverteilung der Farben eines informationstragenden Bereiches	218
Abb. 7.45: Experimentelle Resultate der unscharfen Farbverarbeitung	220
Abb. 7.46: Transformationssystem für unscharfe Inferenzsysteme	222
Abb. 7.47: Struktur eines unscharfen regelbasierten Systems	222
Abb. 7.48: Unscharfes Inferenzsystem mit einer Eingangs- und einer Ausgangsvariablen	224
Abb. 7.49: Kodierung einer unscharfen Menge	224

Abb. 7.50: Initialisierung der Objektkomponente	226
Abb. 7.51: Sollkennlinien Linear und Linear-invers	227
Abb. 7.52: Sollkennlinien Sprung und Sprung-invers	227
Abb. 7.53: Experimentelle Resultate der Kennlinien Linear und Linear-invers	228
Abb. 7.54: Experimentelle Resultate der Kennlinien Sprung und Sprung-invers	228
Abb. 7.55: Schraubenverdichter	230
Abb. 7.56: Verzahnung von Haupt- und Nebenläufer	230
Abb. 7.57: Schraubenrotorpaar	231
Abb. E.1: Anzahl der Funktionsauswertung 10-dim. Ebenenmodell	277
Abb. E.2: Anzahl der Funktionsauswertung 20-dim. Ebenenmodell	278
Abb. E.3: Anzahl der Funktionsauswertung 30-dim. Ebenenmodell	278
Abb. E.4: Anzahl der Funktionsauswertung 40-dim. Ebenenmodell	278
Abb. E.5: Anzahl der Funktionsauswertung 50-dim. Ebenenmodell	279
Abb. E.6: Anzahl der Funktionsauswertung 60-dim. Ebenenmodell	279
Abb. E.7: Anzahl der Funktionsauswertung 70-dim. Ebenenmodell	279
Abb. E.8: Anzahl der Funktionsauswertung 80-dim. Ebenenmodell	280
Abb. E.9: Anzahl der Funktionsauswertung 90-dim. Ebenenmodell	280
Abb. E.10: Anzahl der Funktionsauswertung 100-dim. Ebenenmodell	280
Abb. E.11: Anzahl der Funktionsauswertung 150-dim. Ebenenmodell	281
Abb. E.12: Anzahl der Funktionsauswertung 10-dim. Kugelmodell	281
Abb. E.13: Anzahl der Funktionsauswertung 20-dim. Kugelmodell	282
Abb. E.14: Anzahl der Funktionsauswertung 30-dim. Kugelmodell	282
Abb. E.15: Anzahl der Funktionsauswertung 40-dim. Kugelmodell	282
Abb. E.16: Anzahl der Funktionsauswertung 50-dim. Kugelmodell	283
Abb. E.17: Anzahl der Funktionsauswertung 60-dim. Kugelmodell	283
Abb. E.18: Anzahl der Funktionsauswertung 70-dim. Kugelmodell	283
Abb. E.19: Anzahl der Funktionsauswertung 80-dim. Kugelmodell	284
Abb. E.20: Anzahl der Funktionsauswertung 90-dim. Kugelmodell	284
Abb. E.21: Anzahl der Funktionsauswertung 100-dim. Kugelmodell	284
Abb. E.22: Anzahl der Funktionsauswertung 10-dim. Doppelsumme	285
Abb. E.23: Anzahl der Funktionsauswertung 20-dim. Doppelsumme	285
Abb. E.24: Anzahl der Funktionsauswertung 30-dim. Doppelsumme	286
Abb. E.25: Anzahl der Funktionsauswertung 40-dim. Doppelsumme	286
Abb. E.26: Anzahl der Funktionsauswertung 50-dim. Doppelsumme	286
Abb. E.27: Anzahl der Funktionsauswertung 60-dim. Doppelsumme	287

Abb. E.28: Anzahl der Funktionsauswertung 70-dim. Doppelsumme	287
Abb. E.29: Anzahl der Funktionsauswertung 80-dim. Doppelsumme	287
Abb. E.30: Anzahl der Funktionsauswertung 90-dim. Doppelsumme	288
Abb. E.31: Anzahl der Funktionsauswertung 100-dim. Doppelsumme	288

Anhang D Algorithmenverzeichnis

Alg. 2-1:	Ablauf einer (1+1)-Evolutionstrategie	22
Alg. 2-2:	Ablauf einer ($\mu+1$)-Evolutionstrategie	26
Alg. 2-3:	Ablauf einer ($\mu+\lambda$)-Evolutionstrategie	28
Alg. 2-4:	Drehung der Objektkomponente	34
Alg. 3-1:	Ablauf einer ($\mu/1, \lambda$)-KSA-ES	38
Alg. 5-1:	Transformation der Inversen	154
Alg. 5-2:	Kompositionsmethode	155
Alg. 5-3:	Konvolutionsmethode	156
Alg. 5-4:	Akzeptanz/Ablehnungs-Methode	157
Alg. 5-5:	Approximation der inversen, zentralen und normierten Normalverteilung	160
Alg. 5-6:	Hilfsfunktion $freq(x)$	160
Alg. 5-7:	Erzeugung einer $\Xi_{c, \sigma}^{\gamma}(\mathbf{x})$ -verteilten Zufallszahl	164
Alg. 6-1:	Ablauf einer (μ, λ)-asymmetrischen Evolutionstrategie	168

Anhang E Vollständige Übersicht über die Experimente

In dem Kapitel 7 (Experimentelle Resultate), Seite 185 sind die Resultate zu einigen Experimenten in stark verdichteter Fassung angegeben. Damit auch für diese Experimente eine vollständige Übersicht über die bestimmten Eigenschaften gegeben werden kann, finden sich in diesem Anhang alle zusätzlichen Resultate.

E.1 B1 - Ebenenmodell in hohen Dimensionen

Die Testfunktion B_1 ist ausführlich im Abschnitt 7.2-1 (B1 - Ebenenmodell), Seite 197 erläutert. B_1 ist n -dimensional. Es wird jedoch nur eine Koordinate zur Berechnung des Funktionswertes verwendet. Es gilt:

$$B_1(\mathbf{x}) = -x_1, \text{ mit } \mathbf{x} \in \mathbb{R}^n \quad (\text{E.1})$$

Das Optimum von B_1 befindet sich an den Stellen $\mathbf{x}^* = (\infty, x_2, \dots, x_n)^T$, mit $B_1(\mathbf{x}^*) = -\infty$, die Werte für x_2 bis x_n sind beliebig. Zur Bestimmung der benötigten Zielfunktionsauswertungen ist in einer ersten Untersuchung die optimale Anzahl an Eltern μ und Nachkommen λ in Abhängigkeit der Dimensionalität n der Testfunktion B_1 bestimmt worden. Die Resultate dieser Untersuchung sind in diesem Abschnitt dargestellt.

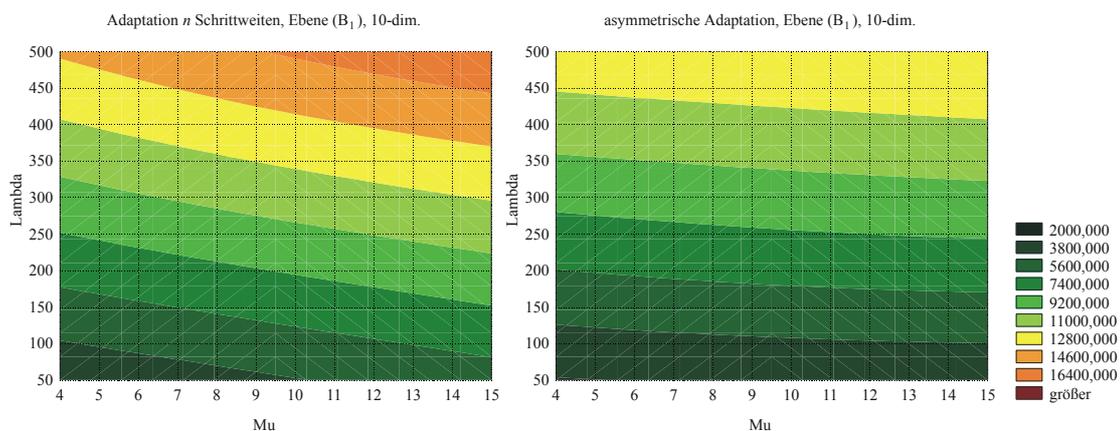


Abb. E.1: Anzahl der Funktionsauswertung 10-dim. Ebenenmodell

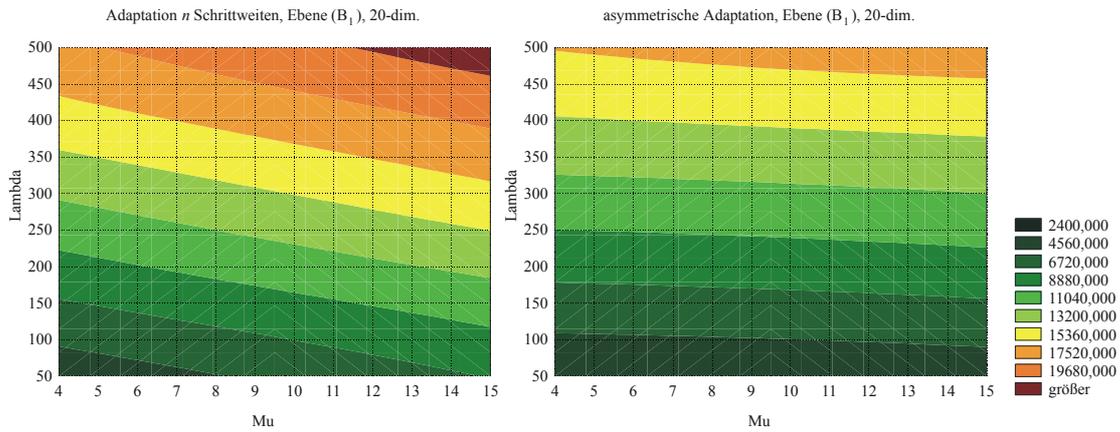


Abb. E.2: Anzahl der Funktionsauswertung 20-dim. Ebenenmodell

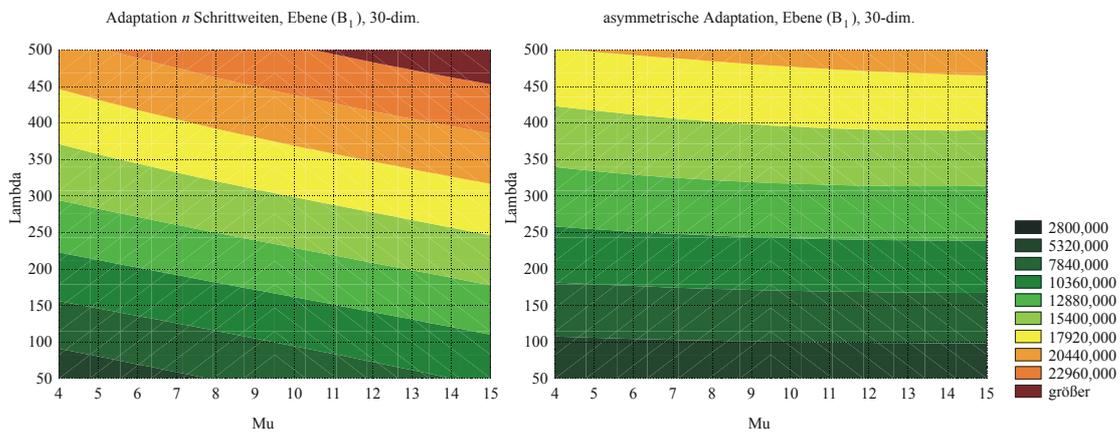


Abb. E.3: Anzahl der Funktionsauswertung 30-dim. Ebenenmodell

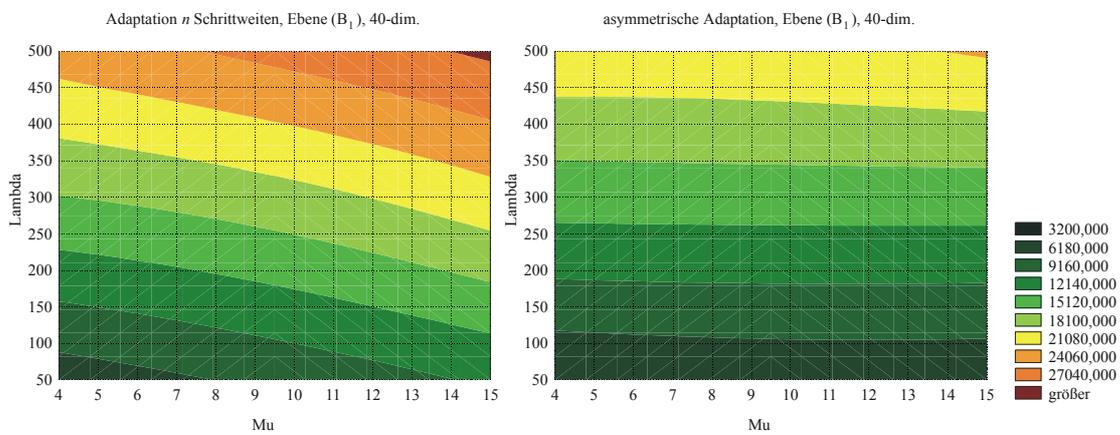


Abb. E.4: Anzahl der Funktionsauswertung 40-dim. Ebenenmodell

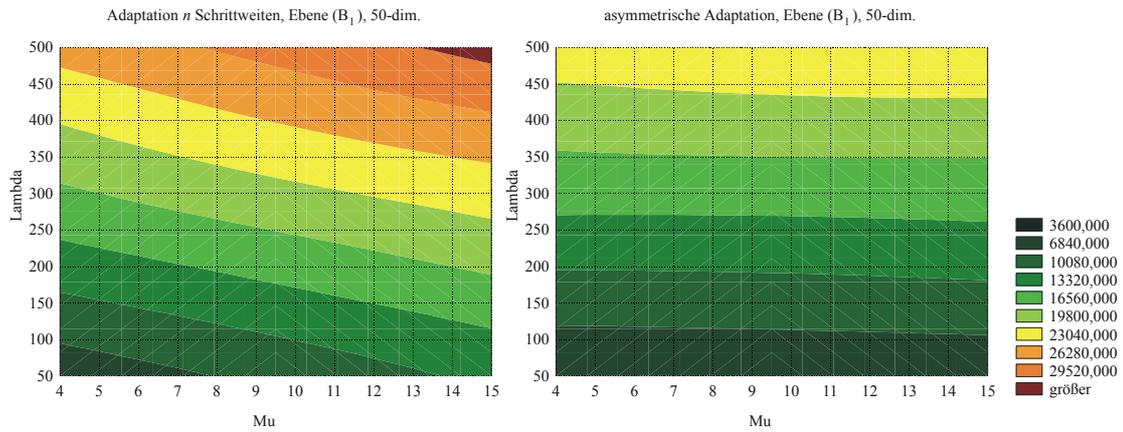


Abb. E.5: Anzahl der Funktionsauswertung 50-dim. Ebenenmodell

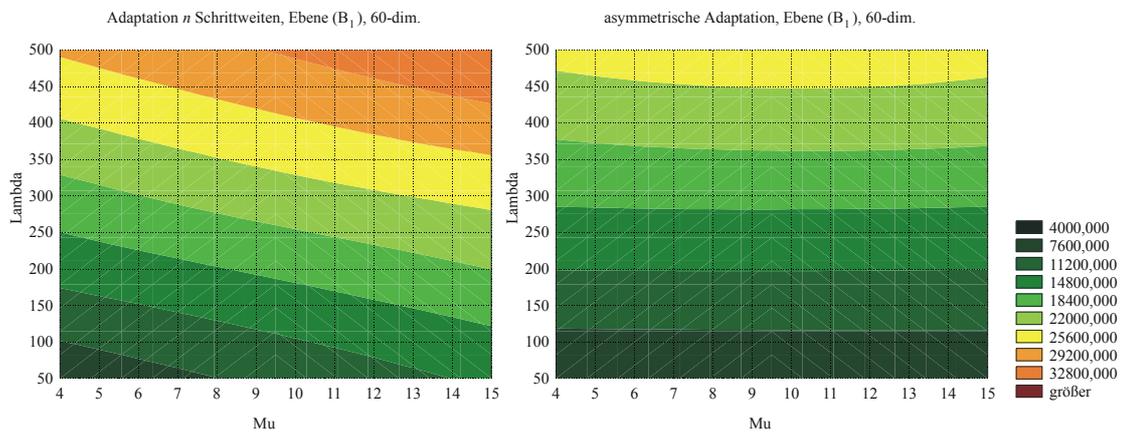


Abb. E.6: Anzahl der Funktionsauswertung 60-dim. Ebenenmodell

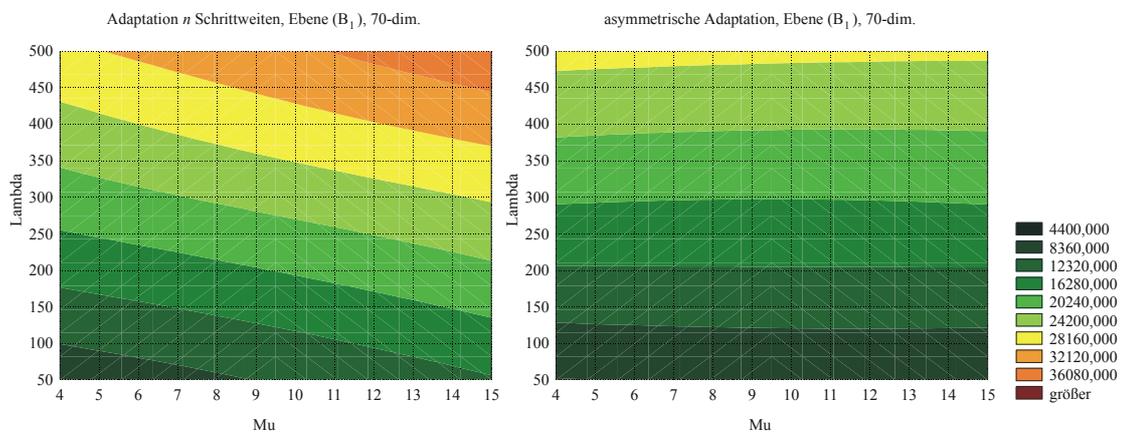


Abb. E.7: Anzahl der Funktionsauswertung 70-dim. Ebenenmodell

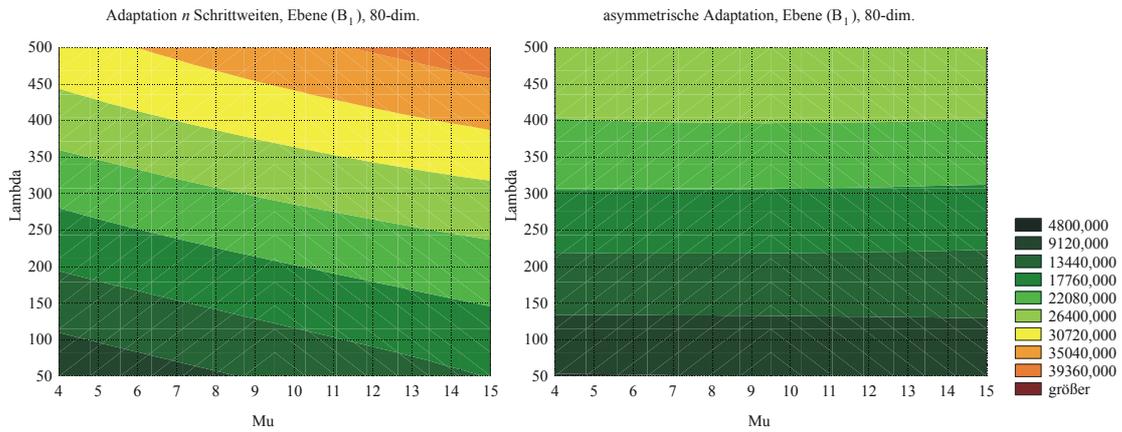


Abb. E.8: Anzahl der Funktionsauswertung 80-dim. Ebenenmodell

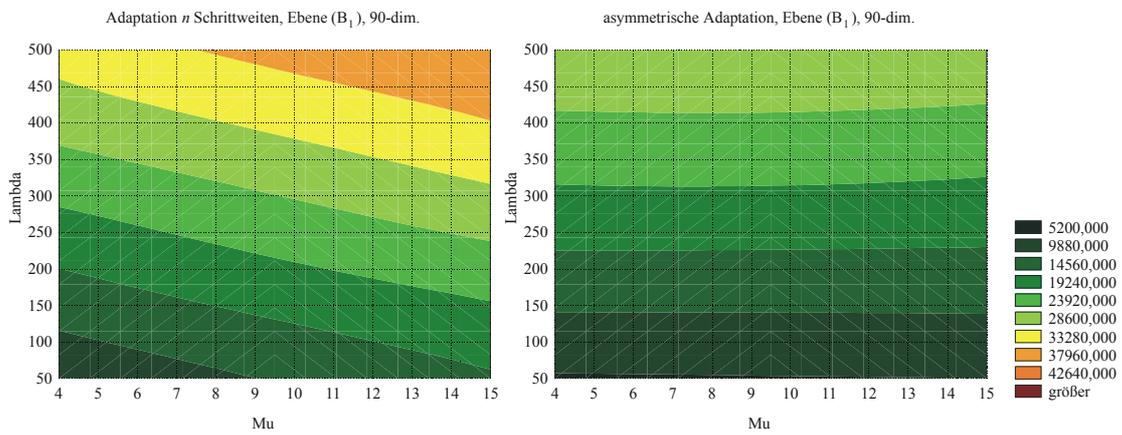


Abb. E.9: Anzahl der Funktionsauswertung 90-dim. Ebenenmodell

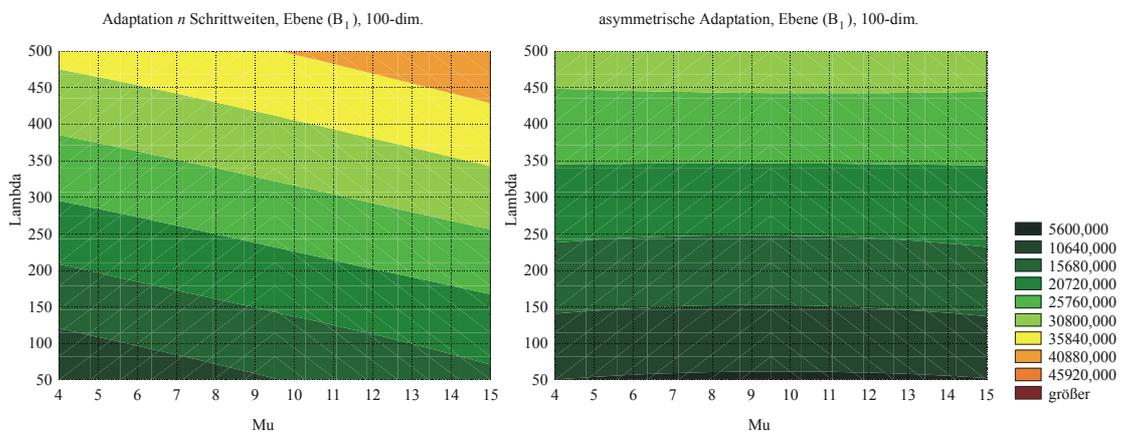


Abb. E.10: Anzahl der Funktionsauswertung 100-dim. Ebenenmodell

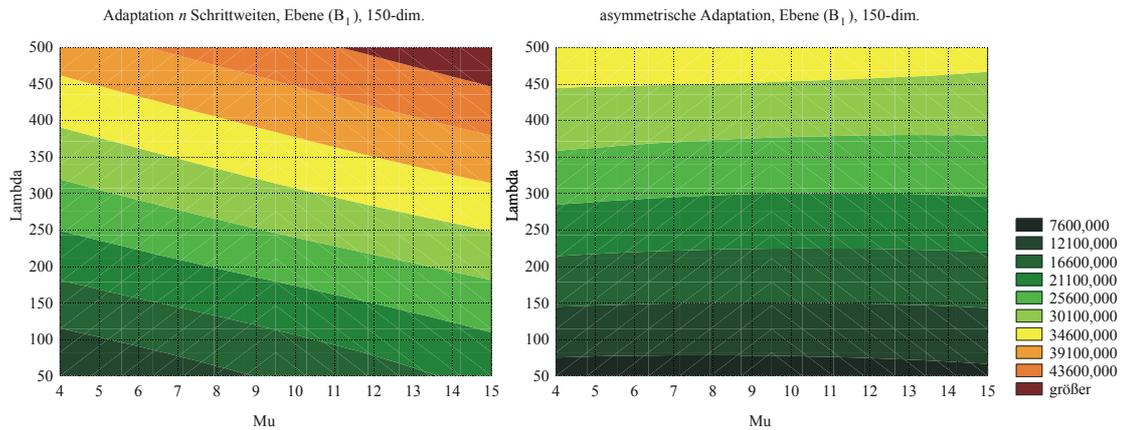


Abb. E.11: Anzahl der Funktionsauswertung 150-dim. Ebenenmodell

E.2 f_1 - Kugelmodell in hohen Dimensionen

Die Testfunktion f_1 ist ausführlich im Abschnitt 7.1-1 (f_1 - Kugelmodell), Seite 187 erläutert. Das Kugelmodell ist eine n -dimensionale quadratische Funktion, die kontinuierlich, streng konvex und unimodal ist. Sie ist definiert durch:

$$f_1(\mathbf{x}) = \sum_{i=1}^n x_i^2, \text{ mit } \mathbf{x} \in \mathbb{R}^n \quad (\text{E.2})$$

Das Optimum von f_1 befindet sich an der Stelle $\mathbf{x}^* = (0, \dots, 0)^T$, mit $f(\mathbf{x}^*) = 0$.

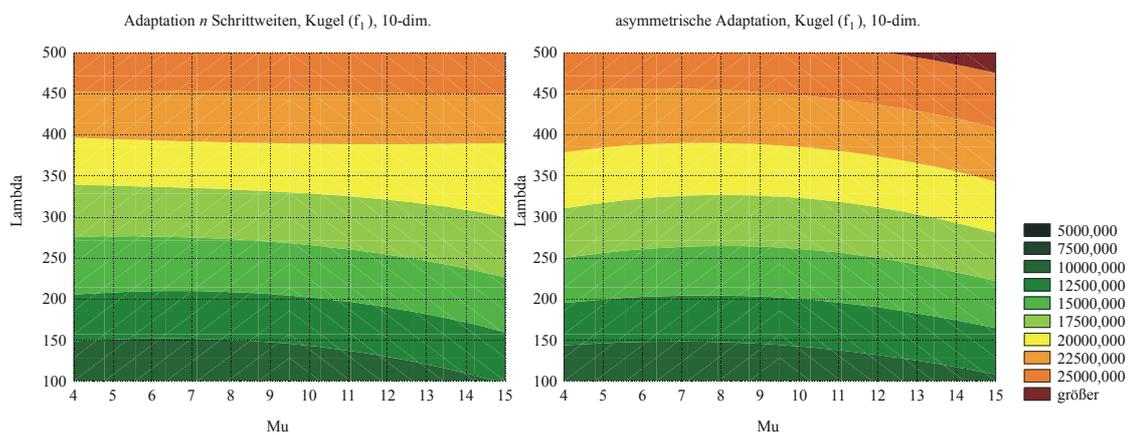


Abb. E.12: Anzahl der Funktionsauswertung 10-dim. Kugelmodell

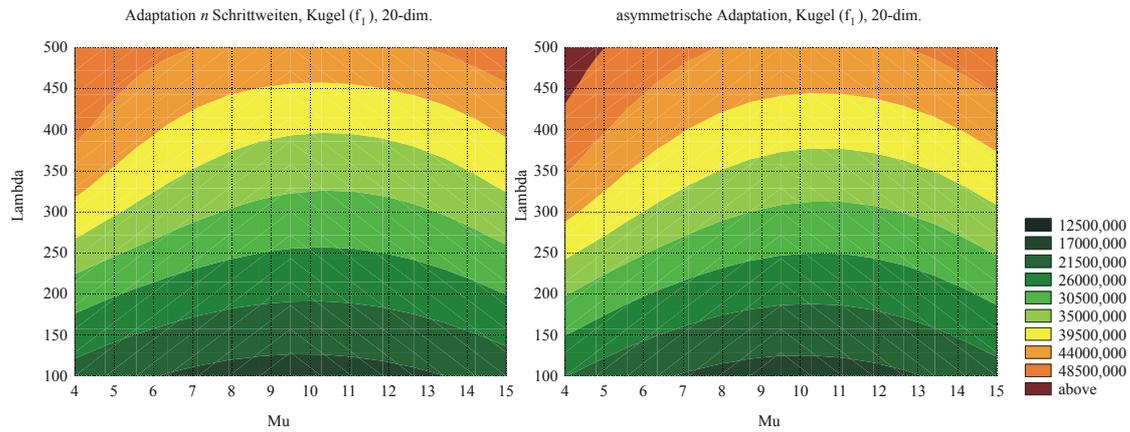


Abb. E.13: Anzahl der Funktionsauswertung 20-dim. Kugelmodell

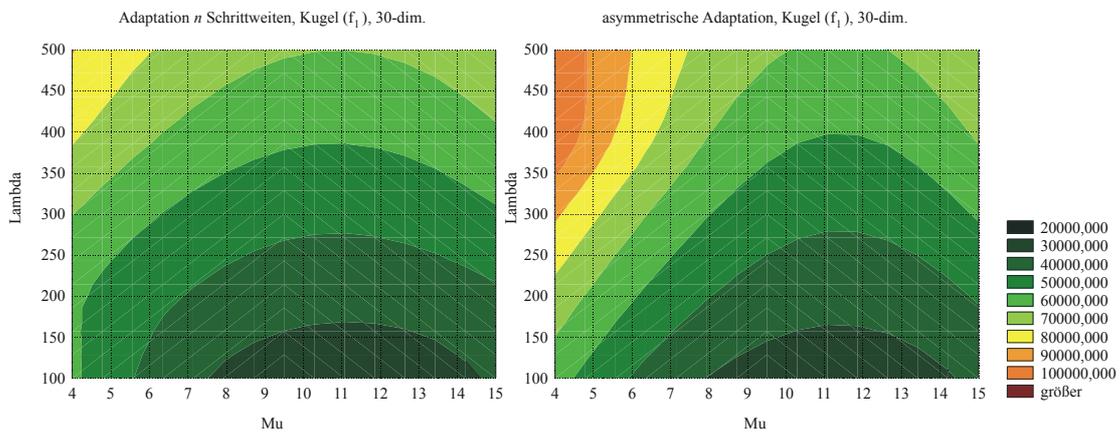


Abb. E.14: Anzahl der Funktionsauswertung 30-dim. Kugelmodell

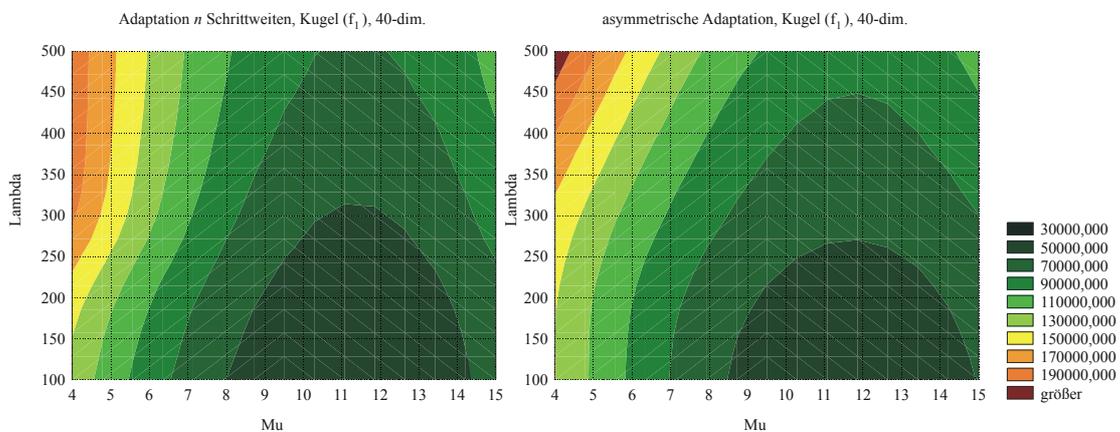


Abb. E.15: Anzahl der Funktionsauswertung 40-dim. Kugelmodell

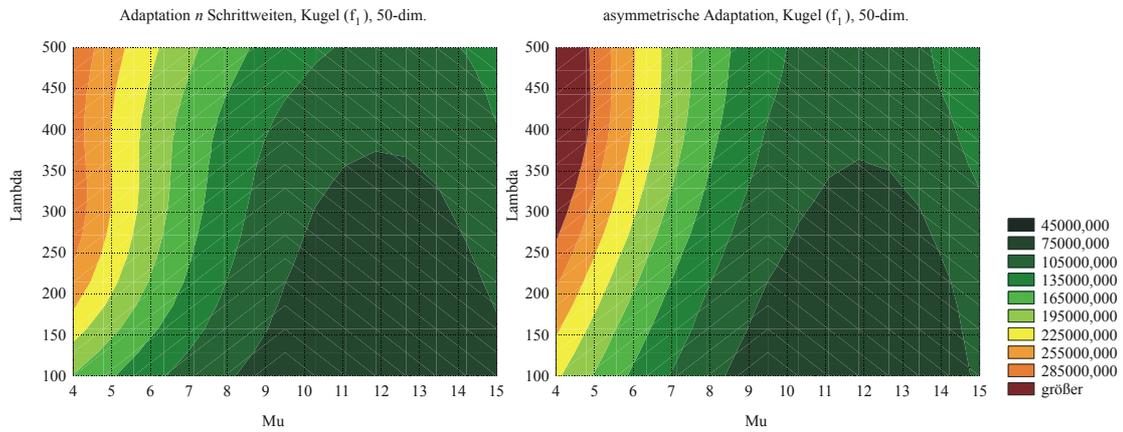


Abb. E.16: Anzahl der Funktionsauswertung 50-dim. Kugelmodell

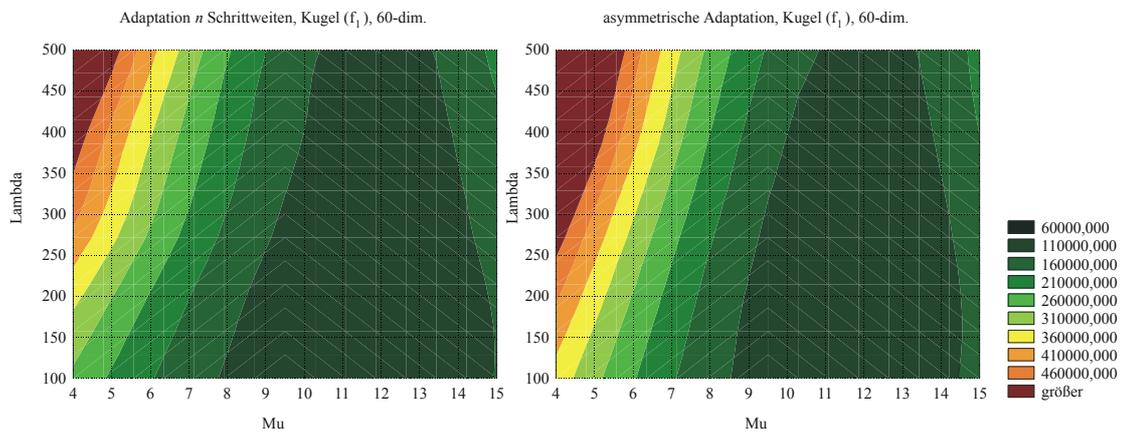


Abb. E.17: Anzahl der Funktionsauswertung 60-dim. Kugelmodell

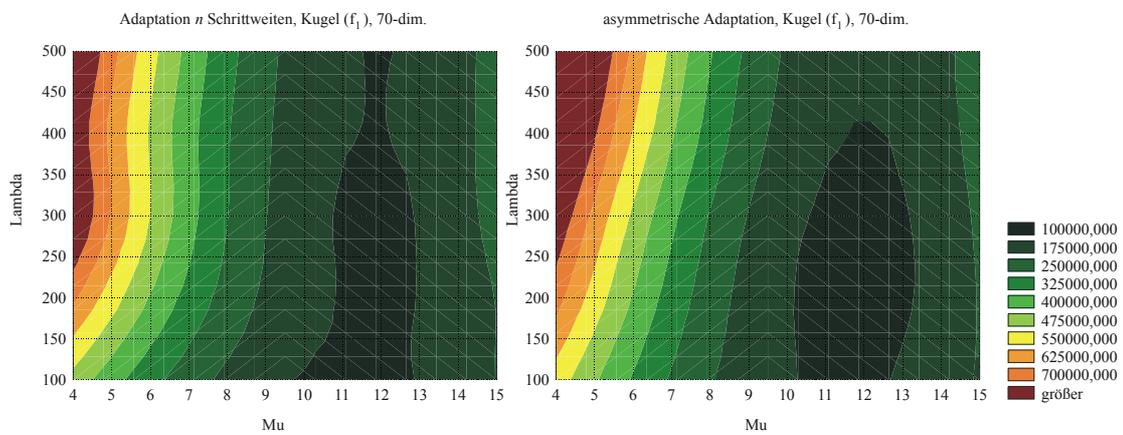


Abb. E.18: Anzahl der Funktionsauswertung 70-dim. Kugelmodell

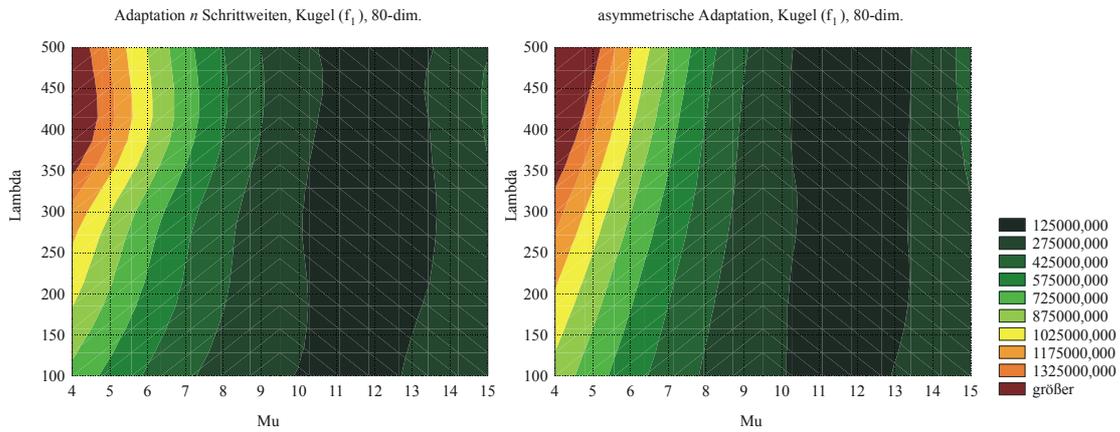


Abb. E.19: Anzahl der Funktionsauswertung 80-dim. Kugelmodell

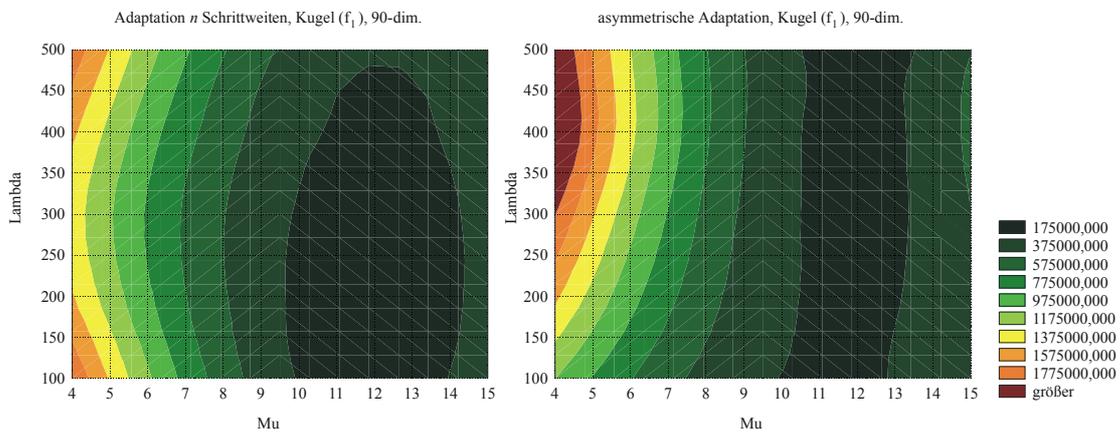


Abb. E.20: Anzahl der Funktionsauswertung 90-dim. Kugelmodell

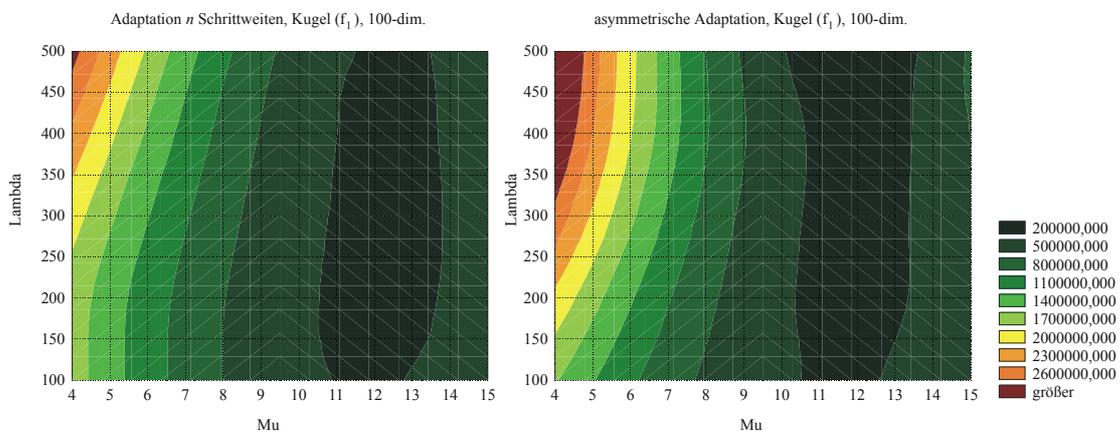


Abb. E.21: Anzahl der Funktionsauswertung 100-dim. Kugelmodell

E.3 f_4 - Doppelsumme in hohen Dimensionen

Die Testfunktion f_1 ist ausführlich im Abschnitt 7.1-4 (f_4 - Doppelsumme), Seite 192 erläutert. Die Grundfunktion f_4 lautet:

$$f_4(\mathbf{x}) = \sum_{i=1}^n \left(\sum_{j=1}^i x_j \right)^2, \text{ mit } \mathbf{x} \in \mathbb{R}^n. \quad (\text{E.3})$$

Das Optimum von f_4 befindet sich an der Stelle $\mathbf{x}^* = (0, \dots, 0)^T$, mit $f_4(\mathbf{x}^*) = 0$.

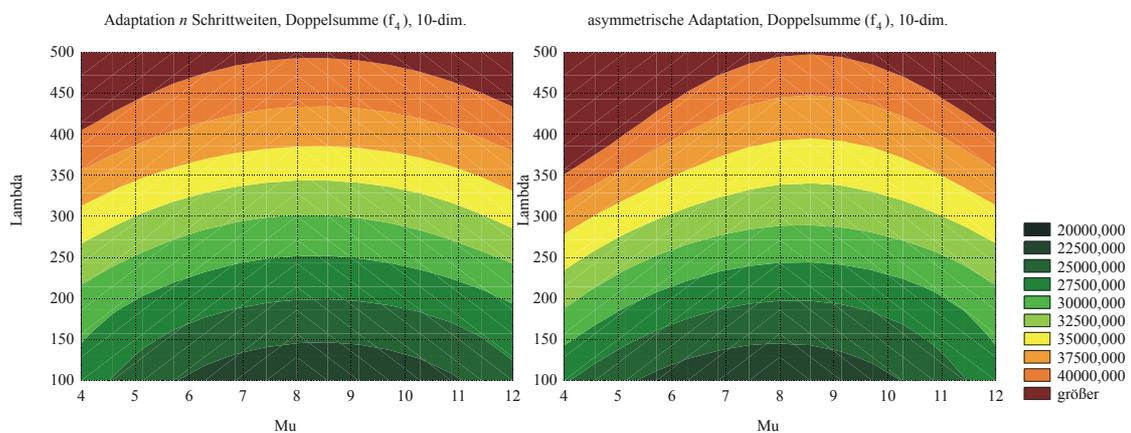


Abb. E.22: Anzahl der Funktionsauswertung 10-dim. Doppelsumme

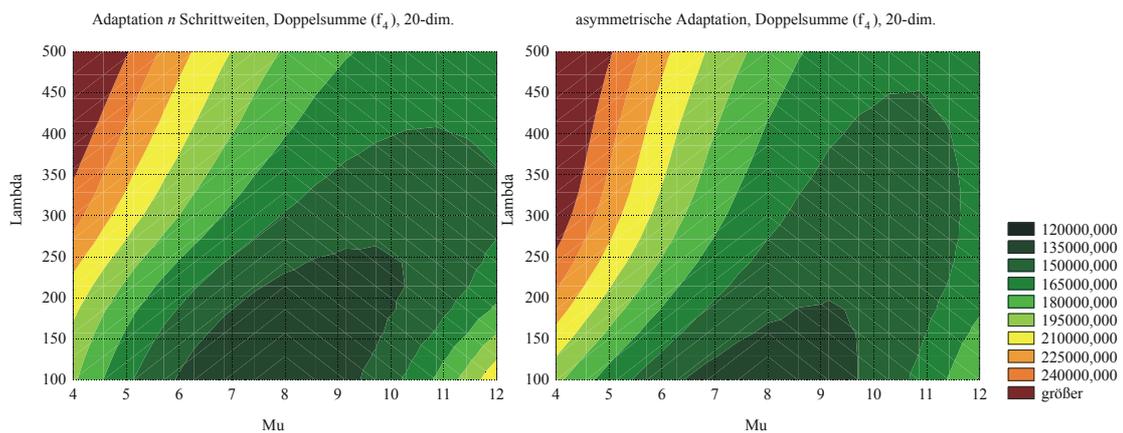


Abb. E.23: Anzahl der Funktionsauswertung 20-dim. Doppelsumme

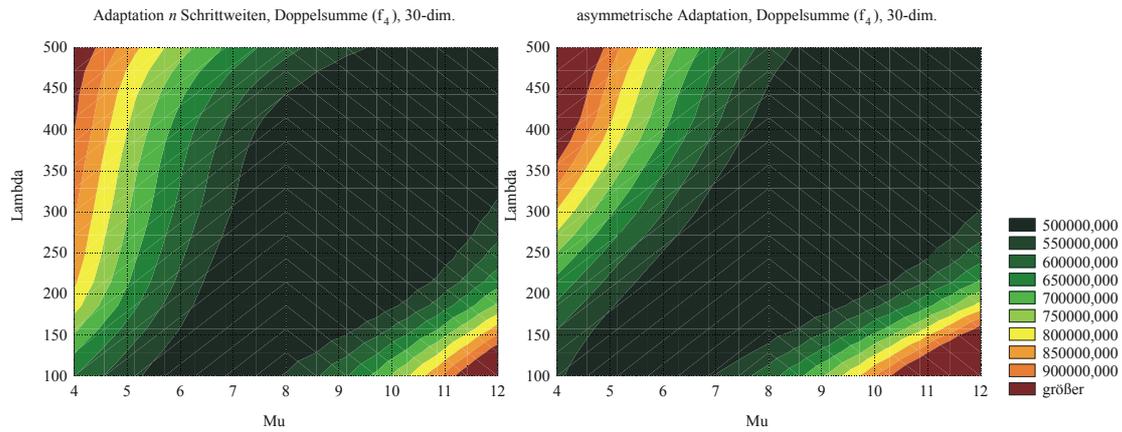


Abb. E.24: Anzahl der Funktionsauswertung 30-dim. Doppelsumme

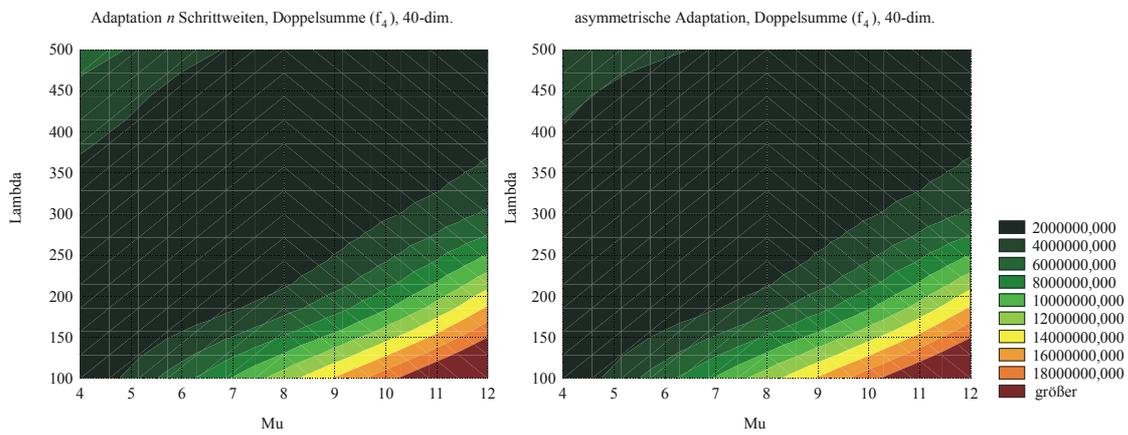


Abb. E.25: Anzahl der Funktionsauswertung 40-dim. Doppelsumme

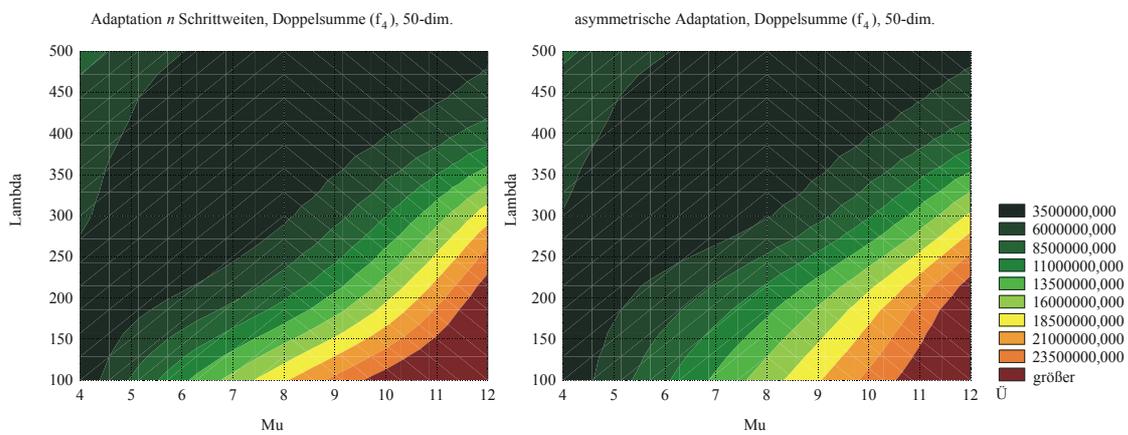


Abb. E.26: Anzahl der Funktionsauswertung 50-dim. Doppelsumme

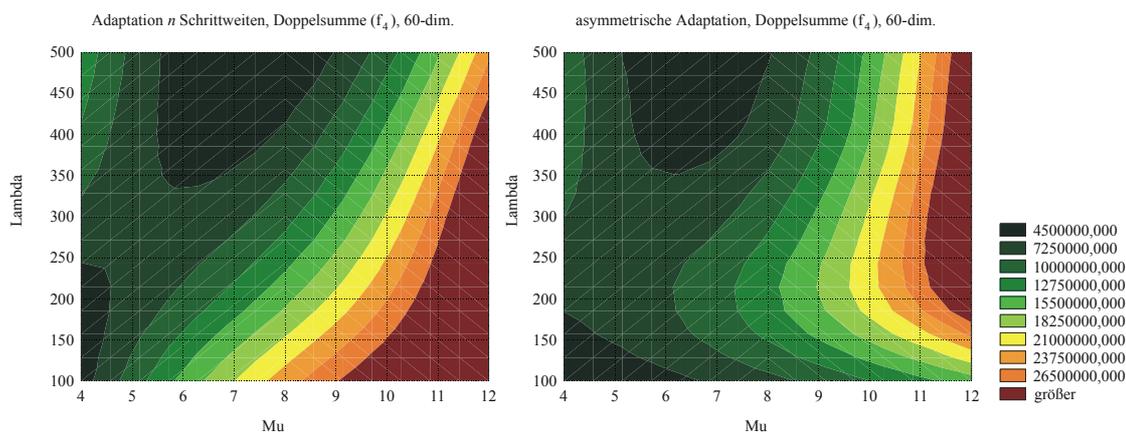


Abb. E.27: Anzahl der Funktionsauswertung 60-dim. Doppelsumme

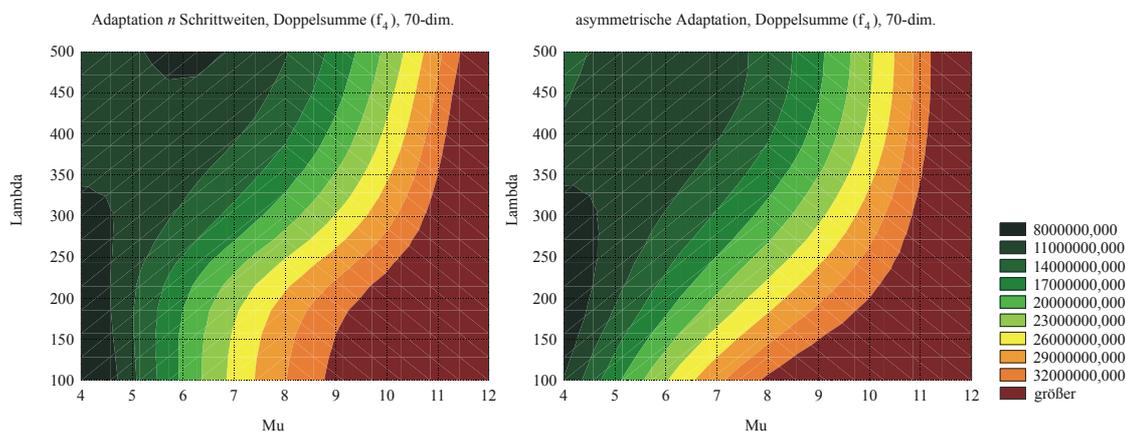


Abb. E.28: Anzahl der Funktionsauswertung 70-dim. Doppelsumme

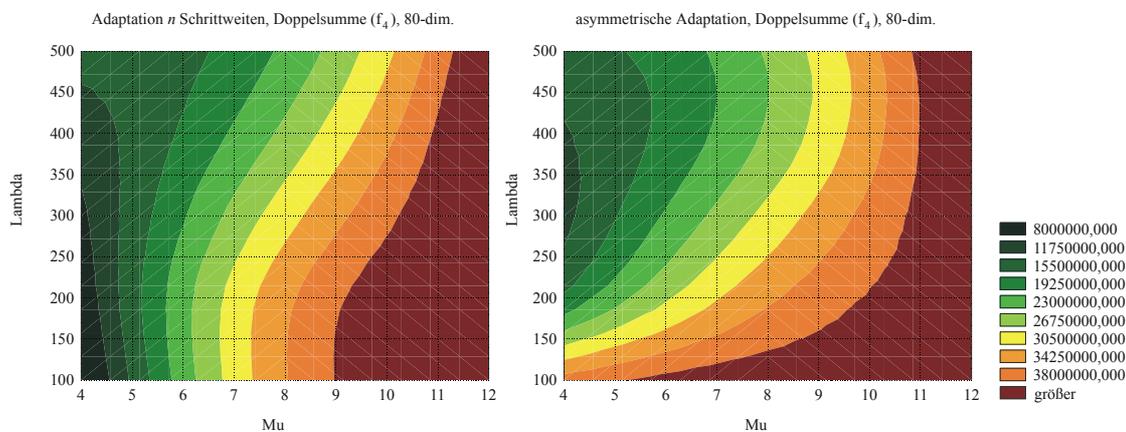


Abb. E.29: Anzahl der Funktionsauswertung 80-dim. Doppelsumme

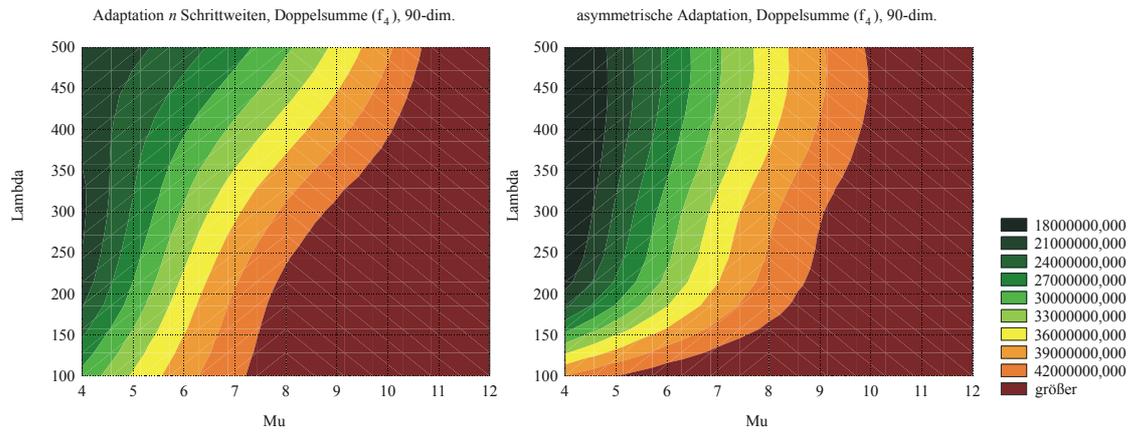


Abb. E.30: Anzahl der Funktionsauswertung 90-dim. Doppelsumme

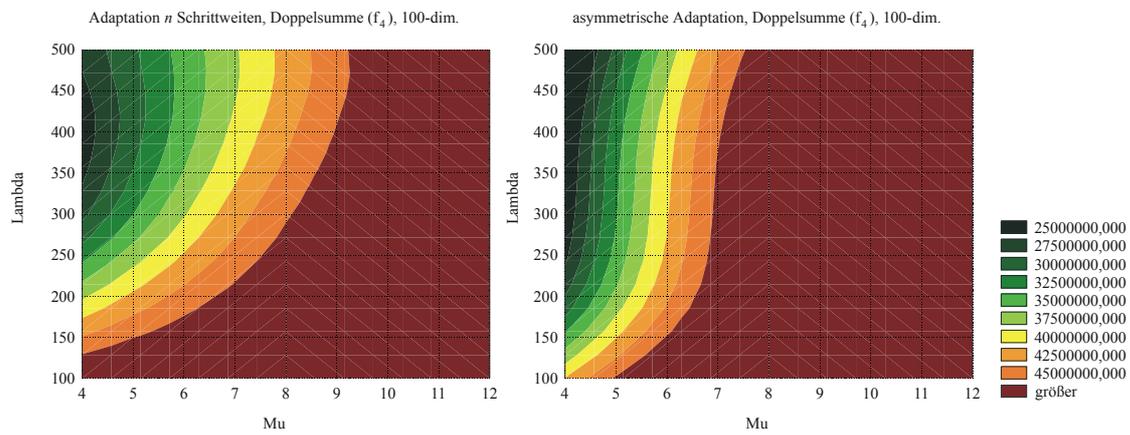


Abb. E.31: Anzahl der Funktionsauswertung 100-dim. Doppelsumme

Anhang F

Stichwortverzeichnis

(1+1)-Evolutionsstrategie	17, 21, 36	(μ, λ)-Evolutionsstrategie	39, 167
Algorithmus	22	\Rightarrow ($\mu+, \lambda$)-Evolutionsstrategie	27
formale Notation	21	Selbstadaptation	167
Fortschritts­geschwindigkeit	23	($\mu/1, \lambda$)-KSA-ES	39
globale Konvergenz	22	($\mu/1, \lambda$)-KSA-Evolutionsstrategie	
Mutation	22	Algorithmus	38
Objekt­komponente	22	Dämpfung	39
optimale Schrittweite	23	Kumulationszeitraum	39
Selektion	22	Mutation	39
Strategie­komponente	22	Objekt­komponente	39
(1, λ)-Evolutionsstrategie	36, 37	Rekombination	38
($\mu+, \lambda$)-Evolutionsstrategie	27	Repräsentation	39
Algorithmus	28	Schrittweite	39
formale Notation	27	Selektion	39
Lagewinkel	28	Strategie­komponente	39
Mutation	29, 30, 31, 33	($\mu/\mu, \lambda$)-CMA-Evolutionsstrategie	42, 197
Objekt­komponente	28	Algorithmus	42
Rekombination	29	Dämpfung	44
Repräsentation	28	Kumulationszeitraum	43
Schrittweiten	28	Mutation	42
Selbstadaptation	29	Objekt­komponente	42
Selektion	28	Summationsvektor	43
Strategie­komponente	28	($\mu/\mu, \lambda$)-XMA-Επολυτιονστρατηγες	
($\mu+1$)-Evolutionsstrategie	25	Selektion	42
Algorithmus	25	Strategie­komponente	43
formale Notation	25	($\mu/\mu, \lambda$)-KSA-ES	41
Mutation	26	($\mu/\mu, \lambda$)-KSA-Evolutionsstrategie	40, 197
Objekt­komponente	26	Dämpfung	41
Rekombination	25	Dämpfungsparameter	41
Schrittweiten	26	Eigenschaften	41
Selektion	26	Kumulationszeitraum	40
Strategie­komponente	26	Mutation	40
($\mu+\lambda$)-Evolutionsstrategie		Objekt­komponente	40
\Rightarrow ($\mu+, \lambda$)-Evolutionsstrategie	27	Repräsentation	40

Schrittweite	40	Eigenschaften	173
Selektion	40	formale Notation	167
0-9		globale Skalierung	171
1/5-Erfolgsregel	18, 22, 24, 26, 54	Mutation	169, 170
A		Objektkomponente	168
Abbruchkriterium	21, 25, 27, 168, 185, 197	Rekombination	167, 168
Abhängigkeiten		Repräsentation	219
zwischen Variablen	59	Reversibilität	177
Adamo, J. M.	225	Schrittweite	167
$A_\gamma(c, \sigma)$ -Generator		Selbstadaptation	64, 167, 170, 177
Algorithmus	163	Strategiekomponenten	168
Einsatz	170	Strategieparameter	167
inverse Verteilungsfunktion	163	Asymmetrische Verteilungsfunktion	70, 76, 167
Komplexität	164	als Mutationsfunktion	78, 94
Kompositionsmethode	161	Alternativen	114, 127
Realisierung	161	Äquivalenz zur Normalverteilung	130, 131
Reproduzierbarkeit	165	Asymmetriemaß	63
Robustheit	165	Beispiele	76, 78, 92, 103
Speicherverbrauch	164	Definition	72, 76, 90, 102
Transformation der Inversen	161	Dichtefunktion	72, 75, 91, 103
$A_\gamma(c, \sigma)$ -Generator	161, 170	Differenzierbarkeit	80, 96, 107, 133
Aktive Strategiekomponente	51	Eigenschaften	72
Akzeptanz/Ablehnungsmethode	154, 156, 161	Entwicklung	63
Alpsector, J.	153	Erwartungswert	83, 98, 109
Ameisensysteme	13	Faltung	134
Angeline, P.	55	globale Skalierung	63, 102
Anguita, D.	153	Grenzverteilung	134, 145
Anwendungsbeispiel		Grundfunktion	72, 73, 114
Funktionsapproximatoren	221	inverse Verteilungsfunktion	80, 97, 108, 130
Schraubenrotorprofile	229	inverse Verteilungsfunktionen	163
unscharfe Farbverarbeitung	213	Modus	80, 96, 107
Approximationen		Notation	64
Gaußsche Fehlerfunktion	149	richtungsunabhängige Skalierung	63, 90
inverse Fehlerfunktion	149	Schiefe	88, 100, 112
Asymmetrie	69	Skalierbarkeit	80, 96, 107
Asymmetriemaß	63, 167	Stetigkeit	79, 95, 106, 131
Asymmetrische Dichtefunktion		strenge Monotonie	79, 95, 106
\Leftrightarrow Asymmetrische Verteilungsfunktion		Strukturerhaltung	142
Asymmetrische Evolutionsstrategie	64, 167, 213, 219	Summe	134
		Symmetrie	129
		Symmetrieerhaltung	141
		Symmetrieerzeugung	142
		Varianz	84, 99, 111, 130
		Verteilungsfunktion	76, 92, 103, 130
		Asymmetrische Verteilungsfunktion	
		globale Skalierung	168

inverse Verteilungsfunktion	158	Chintschin (Satz von)	144
Auswahlfunktion	19	Chu, R.	153
B		Compact Genetic Algorithm	59
Baas, S. M.	225	Corne, D.	61
Bäck, Th.	13, 21, 32, 55, 169, 188	Cox, E.	215
Backhaus, W. G. K.	214	Crossover	56, 59
Baer, R. M.	14	Cunningham, S. W.	159
Baluja, S.	35, 59, 60	D	
Banzhaf, W.	13	Dämpfung	
Bayesian Optimization Algorithm	61	($\mu/\mu, \lambda$)-CMA-Evolutionsstrategier	44
Bayessche Netzwerke	61	Dämpfungswert	39, 41, 42
Beta-Verteilung	114, 117, 124, 126	Davies, S.	35, 60
Beispiele	118	Davis, A. W.	159
Bewertung		Davis, L.	21
direkte	37	Dawkins, R.	14
indirekte	18, 28, 37	De Jong, K.	13
Beyer, H.-G.	35, 55, 70	Deb, K.	35, 55
Beyer, O.	14, 159	Defuzzifizierung	222
Bit Based Simulated Crossover	59	Degani, R.	225
Bivariate Marginal Distribution Algorithm	60	Denormalisierung	222
Bland, I. M.	153	Devroye, L.	154, 156
Blend-Crossover	56, 58	Dezil	66
Born, J.	22	Diagonalkomponenten	
Bortolan, G.	225	\Leftrightarrow Kovarianzmatrix	
Box, G. E. P.	164	Dichtefunktion	15
Bronstein, I. N.	14, 132	Definition	64
Bruns, H.	116	Eigenschaften	64
building blocks	59	Zufallsvariable	15
Burr System	116	Differentielle Evolution	44
Burr, I. W.	116	Fokussierung	46
C		lokale	46
χ^2 -Test	151	Objektkomponente	44
χ^2 -Verteilung	23, 114	Rekombination	45
Cantú-Paz, E.	35, 61	Rekombinationsparameter	44
Cauchy-Verteilung	37, 49, 127	Rekombinationsrate	46
nicht-sphärische	37	Repräsentation	44
sphärische	37	Selektion	45
centric parent selection	53	Skalierungsfaktor	46
		Skalierungsparameter	44
		Steuerparameter	46
		Diffusionsmodelle	52

Direkte		Grundlagen	14
Bewertung	37	Mechanismen	13
Direkte Bewertung	170	Evolutionäre Algorithmen	13
Diskrete Rekombination		Repräsentation	13
⇒ Rekombination		Evolutionäre Operationen	
Doppelsumme		⇒ Mutation	
Bewertung der Resultate	196	⇒ Rekombination	
Definition	192, 199	⇒ Selektion	
experimentelle Resultate	192, 193, 195, 199, 212	Evolutionäres Programmieren	13
Fehlinitialisierung	193, 195, 196	Evolutionary-Gradient-Search	46
hoch-dimensional	212	Adaptation	48
skalierte Varianten	193, 195	Gradientenschätzung	47
Dreißig, B.	230	Iterationsschritt	47
		Mutation	47
		Schrittweitensteuerung	48
		Steuerparameter	49
		Tastschritte	47
E		Evolutionspfad	37, 40, 43
Ebenenmodell		Evolutionstrategien	9, 13
Definition	197	(1+1)-ES	17, 21, 36
experimentelle Resultate	197, 210	(1,λ)-ES	36, 37
hoch-dimensional	210	(μ+,λ)-ES	27
Edgeworth, F. Y.	116	(μ+1)-ES	25
Ellipsenmodell		(μ,λ)-ES	39, 167
Definition	201	(μ/μ,λ)-CMA-ES	42, 197
experimentelle Resultate	201	(μ/μ,λ)-KSA-ES	40, 197
Entstochastisierung	37	Adaptation n Schrittweiten	173
Erreichbarkeit	71	Adaptation Schrittweiten und Lagewinkeln	173
Erwartungswert	16	asymmetrische	
Definition	67	⇒ Asymmetrische	
partieller	68, 83, 84, 98, 109	Evolutionstrategie	
Rechenregeln	67, 69	Erweiterungen	35
Eshelman, L. J.	56	experimentelle Resultate	12
Estimation Distribution Algorithms		Fitness	18
Crossover	59	Generation	14
graphische Notation	60	Grundlagen	14
Modellierung von Abhängigkeiten	60	Individuum	14
Mutation	59	klassische	
Estimation of Distribution Algorithms	35, 59	⇒ Evolutionstrategien	
Etzberger, R.	61	symmetrische	
Eulersche Beta-Funktion	118	Konvergenz	23
Eulersche Konstante	115	Mindestschrittweite	49
Evans, J. O.	159	Mutation	18
Evans, M.	23, 114	Objektkomponente	17, 22, 26, 28
Evolution	13	Population	14
		Rekombination	18

- ⇒ Rekombination
- Global intermediäre Rekombination
 - ⇒ Rekombination
- Gnedenko, B. W.** 134, 144
- Goldberg, D. E.** 35, 59, 60, 61
- Gorges-Schleuter, M.** 35, 52, 53
- Granularität 225
- Grenzverteilung 134, 144, 145
- Grenzwertsatz von Bernoulli 144
- Grenzwertsatz von de Moivre 144
- Grenzwertsatz von Tschebyschew 144
- Gumbel-Verteilung
 - Definition 114
 - Dichtefunktion 114
 - Lageparameter 114
 - Schiefemaß 115
 - Skalierungsparameter 114
 - Verlauf 115
- Γ -Verteilung 114

- H**
- Haber, S.** 153
- Hansen, N.** 35, 41, 188, 197, 198
- Hardin, C. L.** 214
- Harik, G. R.** 35, 59, 60
- Hastings, N.** 23, 114
- Häufigkeitsverteilung 213, 218
- Hebb, D.** 59
- Hebbsche Lernregel 59
- Hecht-Nielsen, R.** 13
- Hellekalek, P.** 150
- Herrera, F.** 13
- Hildebrand, L.** 72, 90, 102, 161, 163, 213, 215, 221
- Hill, G. W.** 159
- Hoffmann, U.** 34
- Hoffmeister, F.** 21, 169, 188
- Hofmann, H.** 34
- Holland, J. H.** 13
- HSI-Farbmodell 216
- Hybride Expertensysteme 223

- I**
- Inaktive Strategiekomponente 51
- Improved Fast Evolutionary Programming 61
- Indirekte Bewertung 37, 170
- indirekte Bewertung 18, 28
- Individuum
 - ⇒ Evolutionsstrategien
 - ⇒ Genetische Algorithmen
 - ⇒ kumulierte Schrittweitenadaptation
 - Kovarianzmatrixadaptation 12
 - Objektkomponente 12
- Inferenz 222
- Intermediäre Rekombination
 - ⇒ Rekombination
- Inverse Fehlerfunktion 82, 149
- Inza, I.** 61

- J**
- Johnson, N. L.** 116
- Julstrom, B. A.** 21

- K**
- Kacprzyk, J.** 215
- Kappler, C.** 36
- Kartesischer Pyramidenstumpf 216
- Kauder, K.** 231
- Keller, R. E.** 13
- Kelley, C. T.** 34
- Kelton, W. D.** 156
- Kelton, W. D.** 150, 151, 157
- Kerre, E. E.** 225
- Kiencke, U.** 134, 135
- Kirkpatrick, S.** 13
- Kolmogorov, A. N.** 134, 144
- Kolmogorov-Smirnov-Test 151
- Kompositionsmethode 154, 155, 156, 161
- Konvolutionsmethode 154, 156
- Koordinatenachse 63
- Koordinatenhalbachse 63

Koordinatenrichtung	63	L	
Korridormodell		Lageparameter	114
Definition	23	Lagevektor	54
Konvergenz	23	Lagewinkel	17, 29, 33, 185, 197
Kotz, S.	116	Lagewinkelmatrix	32
Kovarianz	17	Laplace-Verteilung	36
Kovarianzmatrix	17, 32, 43	Larrañaga, P.	61
Kovarianzmatrixadaptation	43	Law, A. M.	151
Kovarianzmatrixadaptation	37, 41	Law, A. M.	150, 151, 156, 157
Algorithmus	42	Leguizamon, G.	13
Dämpfungsparameter	42	Lin, G.	49, 61
Evolutionspfad	43	Lindeberg-Lévy (Satz von)	144
Mutation	42	Linguistische Variable	224
nicht-zentrale Normalverteilung	54	Linguistischer Term	224
Objektkomponente	42	linkage problem	59
Kugelmodell	169, 185	Liu, Y.	49, 52, 61
alternative Initialisierung	188	Ljapunow (Satz von)	145
Bewertung der Resultate	196	Lobo, F. G.	35, 59, 60
Definition	24, 187, 198	local parent selection	53
experimentelle Resultate	187, 189, 190, 198, 211	Logistische Verteilung	36, 128
Fehlinitialisierung	190, 192	Lokale Differentielle Evolution	46
hoch-dimensional	211	Steuerparameter	46
Konvergenz	24	Lokale Rekombination	53
skalierte Varianten	189, 190	Lokale Selektion	52
Varianten	185	Mutation	52
Kumulationszeitraum	39	Nachbarschaft	52
Kumulative Schrittweitenregelung	37	Notation	52
Algorithmus	38	Rekombination	52, 53
Dämpfungsparameter	39, 41	Selektion	52, 53
Eigenschaften	41	Steuerparameter	53
Evolutionspfad	37, 39	Lozano, J. A.	61
$(\mu/\mu, \lambda)$ -KSA-Evolutionsstrategie	40	Lozano, M.	13
intermediären Rekombination	40	M	
Kumulationszeitraum	39, 40	Majorisierende Funktion	157, 161
Mutation	40	Mandischer, M.	29
Objektkomponente	39, 40	Marse, K.	151
Repräsentation	39, 40	Martinez, H. M.	14
Schrittweite	39, 40	Matsumura, Y.	35, 49
Selbstadaptation	37	Maximum-Entropie-Prinzip	71
Selektion	39		
Künstliche neuronale Netze	13, 59		
Kursawe, F.	14		
Kvasnicka, V.	35, 59		
Kwakernak, H.	225		

Median	66	Mutationsoperator	21, 25, 27, 168
Megson, G. M.	153		
Michalewicz, Z.	13, 21	N	
Modus		Nachbarschaft	52
Definition	65	Natur-analoge Algorithmen	13
Momente		Nebenbedingungen	21, 25, 27, 168
\Rightarrow Erwartungswert		No Free Lunch-Theorem	11
\Rightarrow Schiefemaße		Nordin, P.	13
\Rightarrow Varianz		Normalisierung	222
absolute	67	Normalverteilung	28, 36, 158
Definition	66	Definition	15
k -ter Ordnung	66	korreliert	28, 39
standardisierte	67	nicht-zentrale	54
zentrale	66, 70	normierte	15
Moraga, C.	13	zentrale	15, 158
moving circle	55		
moving corridor	55	O	
Mühlenbein, H.	35, 59, 60	Objektkomponente	17, 22, 26, 28, 219, 226
Muller, M. E.	164	$(\mu/\lambda, \lambda)$ -CMA-Evolutionsstrategie	42
Mutation	14, 18, 54, 59	Differentielle Evolution	44
$(\mu/\lambda, \lambda)$ -CMA-Evolutionsstrategie	42	Kovarianzmatrixadaptation	42
$(\mu/\lambda, \lambda)$ -KSA-Evolutionsstrategie	40	Kumulative Schrittweitenregelung	39, 40
\Rightarrow asymmetrische Mutation		Robust Evolution Strategies	50
Cauchy verteilt	37	Odeh, R. E.	159
Evolutionary-Gradient-Search	47	Ohkura, K.	35, 49
Kovarianzmatrixadaptation	42	Oppenheim, A. V.	135
Kumulative Schrittweitenregelung	39, 40	Orthonormalbasis	42
Laplace-verteilt	36	Ostermeier, A.	41
logistisch verteilte	36		
Lokale Selektion (Diffusionsmodelle)	52	P	
normalverteilt	36	Parabelgrat	
Strategiematrix	30, 31, 33	Definition	207
symmetrische	18	experimentelle Resultate	207
Mutationsfunktion	35, 71	Parker, M. B.	153
Cauchy-Verteilung	37, 49	Partieller Erwartungswert	68, 83
faktorisierend	35	Erweiterung	68
im engeren Sinne	71	Peacock, B.	23, 114
im weiteren Sinne	71, 78, 94, 105	Pearson System	116
Laplace-Verteilung	36	allg. Wahrscheinlichkeitsdichtefunktion	116
logistische Verteilung	36	116	
Normalverteilung	36	Pearson Typ I Verteilungen	117
Skalierbarkeit	71		
Student's t-Verteilung	36		
Symmetrie	71		
systematische Tendenzen	71		
Varianten	35		

Pearson Typ II Verteilungen	118	Lokale Selektion (Diffusionsmodelle)	53
Pearson Typ III Verteilungen	119	Objektkomponente	19
Pearson Typ IV Verteilungen	120	Strategiekomponente	19
Pearson Typ IX Verteilungen	124	symmetrische	18
Pearson Typ V Verteilungen	121, 122	Rekombinationsoperator	25, 26, 27, 168
Pearson Typ VI Verteilungen	122	rellwertige	56
Pearson Typ VII Verteilungen	123	Reparaturalgorithmen	55
Pearson Typ VIII Verteilungen	124	Repräsentation	17, 224
Pearson Typ X Verteilungen	125	($\mu/\mu, \lambda$)-KSA-Evolutionsstrategie	40
Pearson Typ XI Verteilungen	125	Differentielle Evolution	44
Pearson Typ XII Verteilungen	125	Evolutionäre Algorithmen	13
Pearson, K.	116	Evolutionsstrategien	13
Pelikan, M.	35, 60, 61	Genetische Algorithmen	13
Peña, J. M.	61	Kovarianzmatrixadaptation	42
PMMLC-Generator	153	Kumulative Schrittweitenregelung	39, 40
Pomberger, G.	59	Robust Evolution Strategies	50
Population	21	Reusch, B.	213, 215, 231
⇒ Evolutionsstrategien		Reversibilität	72, 177
⇒ Genetische Algorithmen		Reversibilitätsfenster	180
Population Based Incremental Learning	59	RGB-Farbmodell	216
Powell, M. J. D.	34	Richtungsadaptation	54, 177, 179
Punktschweißverbindung	214	Riemannsche Zeta-Funktion	115
		Roberts, S. D.	151
Q		Robust Evolution Strategies	49
Quantil	65	aktive Strategiekomponente	51
Quartil	66	inaktive Strategiekomponente	51
		Mutation	51
R		Mutationsoperator	49
Rasch, D.	116	Objektkomponente	50
Rayleigh-Verteilung	114	Repräsentation	50
Rechenberg, I.	13, 19	Robuste Adaptation	50
Rechenberg, P.	59	Steuerparameter	51
Regelbasierte Systeme	222	Strategiekomponente	50
Rekombination	14, 19, 25, 29, 38, 168	Strategiemutationen	51
Differentielle Evolution	44, 45	Robuste Adaptation	50
diskrete	19, 26	Rosenbrock-Funktion	
global diskrete	20	Definition	200
global intermediäre	20	experimentelle Resultate	200
intermediäre	20	Ross, P.	61
Kumulative Schrittweitenregelung	39, 40	Ross, S. M.	156
lokale	53	Rovetta, S.	153
Lokale Selektion	52	Rudolph, G.	13, 35, 70

S

Salomon, R.	35, 46	Selbstadaptation	55
Schaffer, J. D.	56	Simulated Binary Crossovers	59
Schiefe	185, 197	Skalierbarkeit	71
Definition	69	Skalierungsparameter	114
Koeffizient der	69	spitzer Grat	
Schiefemaß	167	Definition	208
Schrittweite	17, 22, 29, 33, 54, 185, 197	experimentelle Resultate	208
Schrittweitenkontrolle	21, 25	Standardabweichung	16, 87
Schwefel, H.-P.	13, 21, 23, 27, 32, 34	Definition	69
Sebstadaptation	37	Startpopulation	21, 25, 27, 167
Selbstadaptation	10, 18, 29, 30, 31, 33, 36, 197	Stoerig, P.	215
Adaptation einer Schrittweite	29, 173	Storn, R.	35, 44, 46
Adaptation n Schrittweiten	31	Strategiekomponente	17, 22, 26, 28
Adaptation Schrittweiten und Lagewinkel	32	Asymmetriemaß	63
Adaptation von n Schrittweiten	219	globaler Skalierungsfaktor	63
Asymmetrische Evolutionsstrategie	64	Kumulative Schrittweitenregelung	39, 40
Genetische Algorithmen	56	Lagewinkel	17
genetische Algorithmen	55	richtungsunabhängiger Skalierungsfaktor	63
Kovarianzmatrixadaptation	54	Robust Evolution Strategies	50
Lagewinkel	29	Schrittweite	17
Mutation	30, 31, 33	Strategiematrix	29
nicht-zentrale Normalverteilung	54	Strategiematrix	
Schrittweite	29	Mutation	30, 31, 33
Strategiematrix	30, 31, 33	Strategieparameter	167
Selektion	14, 20, 22, 28	Streuungsmaße	66
(μ/λ)-KSA-Evolutionsstrategie	40	Student's t-Verteilung	36, 128, 234
begrenzte Lebensspanne	21	Summe verschiedener Potenzen	
deterministische	21	Definition	206
Evolutionary-Gradient-Search	47	experimentelle Resultate	206
Komma	21, 27, 52, 53	Symmetrie	71, 158
Kumulative Schrittweitenregelung	39	Systematische Tendenzen	71
lokale	21, 52, 53	Syswerda, G.	35, 59
Lokale Selektion (Diffusionsmodelle)	53		
Notation	21	T	
Plus	21, 22, 26, 53	Tablettenmodell	
probabilistische	21	Definition	204
Selektionsoperator	21, 25, 27, 168	experimentelle Resultate	204
Sequenzlänge	150	Tastindividuum	44
Sierra, B.	61	Tastschritte	47
Simulated Annealing	13	Testfunktionen	
Simulated Binary Crossover	35, 55	Bewertung	196, 210
Definition	57	Doppelsumme	59, 192, 193, 195, 199, 212
Eigenschaften	58		

F -Verteilung	114	Definition	203
Laplace-Verteilung	36	experimentelle Resultate	203
Logistische Verteilung	128	Zubrzycki, S.	144, 145
logistische Verteilung	36	Zufallsvariable	14
Rayleigh-Verteilung	114	Dezil	66
Student's t-Verteilung	36, 128, 234	diskrete	15
Weibull-Verteilung	114	Erwartungswert	16
Verteilungsfunktionen		Kovarianz	17
linksschiefe	63	Kovarianzmatrix	17
rechtsschiefe	63	Median	66
Verteilungssysteme	115	Modus	65
Burr System	116	normalverteilt	15
Expansionssystem	116	Quantil	65
Pearson System	116	Quartil	66
Toranzos System	116	Standardabweichung	16
transformierte Verteilungen	116	stetige	15
		Streuungsmaße	66
W		Varianz	16
Weibull-Verteilung	114	Verteilungsfunktion	14
Weicker, K.	35, 54, 55	Zufallsvektoren	17
Weicker, N.	54	Zufallsvektor	
Widerstandspunktschweißen	213	Kovarianz	17
Willsky, A. S.	135	Kovarianzmatrix	17
Wissensakquisition	214, 222	Zufallsvektoren	17
Wissensbasierte Systeme	222	Zufallszahlengeneratoren	
Wissensbasierte Verarbeitung	214	Akzeptanz/Ablehnungsmethode	154, 156,
Wissenstransformation	222	161	
Wright, A.	56	χ^2 -Test	151
		Effizienz	150, 152
Y		Eigenschaften	149
Yager, R. R.	225	Genauigkeit	150
Yao, X.	49, 52, 61	Kolmogorov-Smirnov-Test	151
		Kompositionsmethode	154, 155, 156, 161
Z		Konvolutionsmethode	154, 156
Zadeh, L. A.	215	PMMLC-Generator	153
Zentraler Grenzwertsatz	36, 71, 144	Reproduzierbarkeit	153
Zerstörende Prüfmethode	214	Robustheit	150, 153
Zerstörungsfreie Prüfmethode	214	Sequenzlänge	150
Zeta-Funktion	115	Transformation der Inversen	154, 156, 161,
Zielfunktion	21, 25, 27, 168	163	
Zigarrenmodell		Zunino, R.	153
		Zylinderkoordinatensystem	216