

**Untersuchungen
zum adaptiven Verhalten evolutionärer Algorithmen
in dynamischen Umgebungen**

Dissertation

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

von

Lutz Schönemann

Dortmund
2005

Tag der mündlichen Prüfung: 8.12.2005

Dekan: Prof. Dr. Bernhard Steffen

Gutachter: Prof. Dr. Günter Rudolph und Prof. Dr. Ingo Wegener

Inhaltsverzeichnis

Einleitung	1
I Grundlagen	3
1 Globale Optimierung	5
1.1 Statische Optimierung	5
1.2 Dynamische Optimierung	6
1.3 Optimierungskriterien	10
2 Evolutionäre Algorithmen	13
2.1 Repräsentation und Fitnessberechnung	13
2.2 Initialisierung der Startpopulation	14
2.3 Reproduktion und Variation	15
2.4 Adaptation	17
2.5 Selektion	20
2.6 Abbruchkriterium	20
3 Evolutionäre Algorithmen für dynamische Umgebungen	23
3.1 Vergleich zwischen Dynamik und Rauschen	23
3.2 Diversitätserhaltung	26
3.3 Adaptive und selbstadaptive Mechanismen	28
3.4 Speicherbasierte Verfahren	28
3.5 Bekannte theoretische Ergebnisse	29
3.6 Fazit	32
4 Möglichkeiten des Vergleichs der Ergebnisse verschiedener Strategien	33
4.1 Darstellung der Einzelergebnisse	34
4.2 Vergleich mehrerer Verfahren	39
4.3 Gütemaße	44
4.4 Fazit	51

II Experimentelle Untersuchungen	53
5 Aufbau der Experimente	55
5.1 Testfunktionen	55
5.2 Vorgehensweise und EA-Parametrisierung	61
6 Geeignete Wahl der Anzahl der Schrittweiten	63
6.1 Einführung	63
6.2 Experimentelle Untersuchungen	63
6.3 Fazit	70
7 Ergebnisse auf dem Kugelmodell	71
7.1 Statische Umgebung	71
7.2 Dynamische Umgebungen	83
7.3 Fazit	96
8 Ergebnisse auf multimodalen Funktionen	99
8.1 Ackley-Funktion	99
8.2 Rastrigin-Funktion	107
8.3 Fazit	116
9 Betrachtungen zur Diversität	121
9.1 Einführung	121
9.2 Kugelmodell	122
9.3 Ackley- und Rastrigin-Funktion	123
9.4 Fazit	128
10 PEACE – Ein neuartiger paralleler evolutionärer Algorithmus für dynamische Umgebungen	129
10.1 Parallele evolutionäre Algorithmen	129
10.2 PEACE im Detail	130
10.3 Experimentelle Resultate	131
10.4 Fazit	140
Zusammenfassung und Ausblick	143
Literaturverzeichnis	147

Einleitung

Kaum einer unserer Lebensbereiche ist von dauerhafter statischer Natur. Dies führt dazu, dass wir nicht nur statischen sondern auch sehr vielen dynamischen Optimierungsproblemen gegenüber stehen. Ebenso groß wie die Zahl der Anwendungsgebiete ist dabei die Anzahl der unterschiedlichen Dynamikszenerarien.

Evolutionäre Algorithmen (EA) bilden eine Klasse probabilistischer Optimierungsverfahren, die auf dem Darwinschen Evolutionsparadigma basieren. Durch die wesentlichen Operationen der Rekombination, Mutation und Selektion verbessern sie schrittweise eine Menge von Lösungen. Weil sie dabei nur wenige Voraussetzungen an die Optimierungsaufgabe stellen, erzielen sie häufig auch in solchen Situationen gute Erfolge, in denen andere — insbesondere klassische — Verfahren versagen.

Die von evolutionären Algorithmen zur Optimierung von Funktionen mit reellwertigem Suchraum verwendeten Operatoren besitzen zumeist verschiedene Freiheitsgrade. Beispiele für geeignet zu wählende Einstellungen betreffen

- den Rekombinationstyp (diskret, intermediär),
- die Mutationsverteilung (Normal-, Cauchyverteilung),
- das Selektionsverfahren (deterministisch, stochastisch),
- die Strategieparameter (Lernraten, Schrittweiten) etc.

Einzelne Varianten unterscheiden sich dabei oftmals nur durch wenige Details. Wenn genaue Kenntnisse der Zusammenhänge zwischen unterschiedlichen Einstellungen und den erzielten Ergebnissen fehlen, wird vielfach eine Standardparametrisierung gewählt. Allerdings können bereits kleine Unterschiede zu wesentlich differierenden Ergebnissen führen. Womöglich entscheiden diese Unterschiede sogar darüber, ob ein Verfahren gegen das Optimum konvergiert oder divergiert.

Während sich evolutionäre Algorithmen bereits in vielen Anwendungen als äußerst vorteilhaft erwiesen haben, liegen dennoch nur wenige systematische Untersuchungen zum adaptiven Verhalten evolutionärer Algorithmen in den in dieser Arbeit im Vordergrund stehenden dynamischen Umgebungen vor. Betrachten wir die Auswirkungen einzelner Parametrisierungen so ist noch ein erheblicher Untersuchungsaufwand notwendig. Insbesondere existieren zu den Auswirkungen unterschiedlicher Populationsgrößen kaum angemessene Untersuchungen. Etwas besser sieht die Situation in statischen Umgebungen aus, bei denen für einzelne Strategien bereits erste Resultate erzielt werden konnten. Auf unseren Untersuchungsgegenstand bezogen seien an dieser Stelle die Untersuchungen von Affenzeller und Wagner (2004) genannt, die eine einfache Populationsgrößensteuerung in genetischen Algorithmen (GA) zur Lösung von Traveling Salesperson Problemstellungen einsetzen.

In der vorliegenden Arbeit sollen wesentlich umfangreichere Untersuchungen zum adaptiven Verhalten einer Variante evolutionärer Algorithmen, den Evolutionsstrategien (ES), durchgeführt werden. Dabei gehen wir vom Szenario einer Blackbox-Optimierung aus. Dies bedeutet, dass wir

neben den an ausgewählten Punkten erhaltenen Funktionswerten keine Informationen über die zugrunde liegende Funktion besitzen. Allerdings setzen wir voraus, dass es sich um eine Funktion mit einer gewissen Kausalität der Funktionswerte zwischen benachbarten Punkten handelt, da andernfalls eine sinnvolle Optimierung unmöglich erscheint. Aus dem gleichen Grund soll auch die Bewegungsstärke in den dynamischen Fällen relativ moderat sein.

Zur Optimierung wird eine übliche ES eingesetzt, weil durch den Einsatz eines auf dynamischen Funktionen spezialisierten Verfahrens Nachteile bei der Optimierung statischer Funktionen zu erwarten sind, die jedoch vermieden werden sollen. Besonderes Augenmerk widmen wir dabei den Populationsgrößenparametern μ und λ respektive dem Selektionsdruck λ/μ . Bei den Betrachtungen stehen die Auswirkungen der Wahl dieser Parameter auf das Konvergenzverhalten der Evolutionsstrategien im Vordergrund. Das Konvergenzverhalten wird dabei durch zwei Komponenten bestimmt, die sich zum Teil widersprechen. Zum einen handelt es sich um die Konvergenzsicherheit, zum anderen um die Konvergenzgeschwindigkeit.

Eine in den ES benutzte Einstellung mit gleichen Populationsgrößen für unterschiedliche Problemstellungen und -dimensionen lässt beispielsweise nachfolgende Punkte unberücksichtigt. Mit zunehmender Problemdimension n wird

- eine erhöhte Zeit zur Einstellung der so genannten Mutationsschrittweiten σ_i benötigt,
- eine geringere Konvergenzgeschwindigkeit zu verzeichnen sein,
- eine erhöhte Anzahl an Funktionsauswertungen benötigt, um das Ergebnis mit einer gleichen absoluten Genauigkeit zu bestimmen,
- die Wahrscheinlichkeit einer Verbesserung bei gleicher Schrittweite immer kleiner werden (Rechenberg 1994, S. 161).

Darüber hinaus hat das Ausmaß einer eventuell vorhandenen Multimodalität der Zielfunktion ebenfalls größeren Einfluss auf die Schwierigkeit der Optimierung dieser Funktion.

Somit steht die Untersuchung einer guten Wahl der Parameter μ und λ insbesondere bei der oftmals verwendeten paarweisen Rekombination noch aus. Da theoretische Ergebnisse zum Verhalten EA selbst in statischen Umgebungen sehr schwer zu ermitteln sind, wird hier auf theoretische Untersuchungen gänzlich verzichtet. Stattdessen sollen die durch empirische Untersuchungen erhaltenen Ergebnisse möglichst umfassend interpretiert und wenn sinnvoll, durch Testverfahren statistisch abgesichert werden. Die durchgeführten Experimente sollen einerseits zu grundsätzlichen Erkenntnissen führen. Andererseits münden sie in der Entwicklung eines neuen parallelen EA zur Optimierung speziell in dynamischen Umgebungen.

Die Arbeit ist wie folgt gegliedert: Nach einer kurzen Vorstellung der Problemstellung der globalen Optimierung im ersten Kapitel wird das zweite Kapitel die Grundlagen evolutionärer Algorithmen darstellen. Gängige Verfahren für dynamische Umgebungen zeigt das dritte Kapitel. Betrachtungen zur Durchführung empirischer Untersuchungen mit EA schließen sich im vierten Kapitel an. Das fünfte Kapitel stellt die hier benutzten Testfunktionen und Parametrisierungen der eingesetzten EA vor. Während im sechsten Kapitel Experimente zur Wahl der Anzahl der Schrittweiten durchgeführt werden, präsentieren die Kapitel sieben und acht die Ergebnisse der Untersuchungen zur Einstellung der Populationsgrößen auf den Testfunktionen in statischen und dynamischen Umgebungen. Das neunte Kapitel geht auf die Problematik der Diversität in Populationen ein. Im zehnten Kapitel wird der neue parallele EA im Detail vorgestellt. Abschließend werden noch einmal die erhaltenen Ergebnisse zusammengefasst und ein Ausblick auf noch offene Punkte gegeben.

Teil I

Grundlagen

Kapitel 1

Globale Optimierung

Optimierungsaufgaben begegnen uns auf zahlreichen Gebieten. Die Struktur der Aufgaben ist dabei ebenso vielfältig wie von unterschiedlichem Schwierigkeitsgrad. Von prinzipiell unterschiedlichem Charakter erweisen sich statische und dynamische Optimierungsprobleme. Aus diesem Grund bietet es sich an, eine Unterteilung dieser beiden Optimierungsprobleme vorzunehmen, weshalb im Abschnitt 1.1 zunächst das statische Optimierungsproblem vorgestellt und im Abschnitt 1.2 das dynamische Optimierungsproblem betrachtet wird.

1.1 Statische Optimierung

Die grundsätzliche Struktur der überwiegenden Anzahl der statischen Optimierungsprobleme gestaltet sich ähnlich. Für eine Menge beeinflussbarer Parameter muss eine Einstellung gefunden werden, die unter Einhaltung bestimmter Nebenbedingungen eine Zielfunktion minimiert oder maximiert. Wie in der gängigen Literatur üblich, wird im Folgenden ein Optimierungsproblem mit der Minimierung einer Zielfunktion gleichgesetzt, sodass die Begriffe Optimierung und Minimierung in der vorliegenden Arbeit äquivalent benutzt werden. Für eine mögliche Maximierung sind die Begriffe entsprechend anzupassen, denn die ausschließliche Berücksichtigung der Minimierung bedeutet keine Beschränkung der Allgemeinheit, da eine Maximierung mühelos auf eine Minimierung zurückgeführt werden kann mittels

$$\max_x f(x) = - \min_x -f(x) .$$

Insgesamt kommen wir zu folgender allgemein gebräuchlichen Definition eines globalen Minimums:

Definition 1.1 (Globales Minimum) *Gegeben sei eine statische Zielfunktion $f_s : \mathbb{R}^n \rightarrow \mathbb{R}$. An der Stelle $x^* \in \mathbb{R}^n$ (**Optimalstelle**) existiert ein globales Minimum mit Wert $f_s(x^*) > -\infty$ genau dann, wenn gilt:*

$$\forall x \in \mathbb{R}^n : f_s(x^*) \leq f_s(x) .$$

Definition 1.2 (Statisches globales Optimierungsproblem) *Gegeben seien die Voraussetzungen aus Definition 1.1. Das statische globale Optimierungsproblem besteht in der Aufgabe, die Stelle (eine Stelle) des globalen Minimums (der globalen Minima) der statischen Funktion f_s zu bestimmen.*

Im Folgenden wird davon ausgegangen, dass das globale Optimum eindeutig bestimmt ist. D.h. die Funktion besitzt nur eine Optimalstelle.

Zur Definition eines lokalen Minimums benötigen wir den Begriff der ϵ -Umgebung um x' . Sie ist definiert als

$$U_\epsilon(x') = \{x \in \mathbb{R}^n \mid \|x - x'\| < \epsilon\} \text{ für } \epsilon > 0 .$$

Dabei bezeichnet $\|\cdot\|$ die euklidische Norm im \mathbb{R}^n .

Definition 1.3 (Lokales Minimum) Gegeben sei eine statische Zielfunktion $f_s : \mathbb{R}^n \rightarrow \mathbb{R}$. An der Stelle $x' \in \mathbb{R}^n$ existiert ein lokales Minimum genau dann, wenn gilt:

$$\exists \epsilon \in \mathbb{R}, \epsilon > 0 \forall x \in U_\epsilon(x') : f_s(x') \leq f_s(x) .$$

Aus den Definitionen 1.1 bis 1.3 geht hervor, dass ein globales Minimum immer gleichzeitig auch ein lokales darstellt. In der vorliegenden Arbeit soll jedoch ein lokales Optimum eine Lösung bezeichnen, die ein lokales Minimum darstellt, das nicht das globale Optimum bildet. Die Stelle des globalen Minimums selbst wird nachfolgend auch optimale Lösung genannt.

Der zulässige Parameterbereich kann durch Angabe von Restriktionen

$$g_i(x) \geq 0 \wedge h_j(x) = 0, \text{ für } i \in \{1, \dots, k\}, j \in \{1, \dots, l\}$$

beschränkt sein. Da solche Restriktionen die Optimierung wesentlich erschweren können, werden oftmals spezialisierte Verfahren angewendet. In dynamischen Umgebungen ergäbe sich durch die Hinzunahme von Restriktionen die Schwierigkeit, dass bereits erhaltene zulässige Lösungen ungültig werden. Da in der vorliegenden Arbeit aber nur grundsätzliche Untersuchungen zum adaptiven Verhalten EA durchgeführt werden sollen, wird im Folgenden von einem unrestringierten Parameterraum \mathbb{R}^n ausgegangen.

Zur Erklärung der Begriffe Konvergenz und Divergenz schließen wir uns der Beschreibung von (Beyer 2001, s. 14) an, der die Konvergenz als das Erreichen des Optimums bezeichnet. Meinen wir eine Konvergenz gegen ein lokales Optimum wird dies explizit benannt oder als Stagnation bezeichnet. Von einer Divergenz sprechen wir, wenn ein Verfahren sich immer mehr vom Optimum entfernt, ohne dass für die Entfernung eine obere Schranke existiert.

1.2 Dynamische Optimierung

1.2.1 Formen der Dynamik

Dynamische Optimierungsprobleme umfassen eine Vielzahl denkbarer Formen der Dynamik. Verschiedene Autoren versuchen diese zu kategorisieren und erstellen einen kleinen Katalog unterschiedlicher Typen. So trifft beispielsweise (Abk. bspw.) De Jong (1999) eine Unterteilung dynamischer Problemstellungen in die für ihn wichtigsten vier Kategorien:

- Funktionen der ersten Kategorie ändern sich nur sehr langsam und schwach.
- In der zweiten Kategorie befinden sich Funktionen, die sich schneller und im Laufe der Zeit auch wesentlich verändern. Es entstehen neue Minima an bisher „uninteressanten“ Positionen und Stellen mit bisherigen Minima erhalten schlechtere Werte.
- Der dritte Typ dynamischer Funktionen besteht aus oszillierenden Funktionen, bei denen das Minimum zwischen verschiedenen (aber endlich vielen) Stellen wechselt.
- Funktionen der vierten Kategorie unterliegen dagegen abrupten und extremen Änderungen.

Im Gegensatz zu den meisten anderen Autoren unterscheiden Trojanowski und Michalewicz (1999a) zwei zusätzliche Spezialfälle, die sich durch eine zeitliche Dynamik auszeichnen. Zum einen erwähnen sie eine Veränderung der Problemdimension n . Dabei kann verschiedentlich eine Änderung der Umgebung auch eine Änderung der Kodierung der Problemstellung erfordern. Um einen solchen Fall handelt es sich u.a. bei der Bearbeitung von Aufträgen, bei denen die Ankunft eines neuen Auftrags auch eine Vergrößerung der zu verarbeitenden Aufträge erzwingt (Branke 1999). Zum anderen können sich auch Restriktionen ändern. Dies beinhaltet neben dem Intervallbereich der Variablen auch die Anzahl der Restriktionen. Im Fall der Änderungen von Restriktionen können von einem Schritt zum nächsten theoretisch sämtliche Lösungen ihren Status von zulässig nach unzulässig und umgekehrt ändern.

Durch Kombination mit verschiedenen Formen der Änderung der Zielfunktion gelangen Trojanowski und Michalewicz (1999a) zu insgesamt vier verschiedenen Klassen dynamischer Problemstellungen. Als wichtigste Unterscheidungskriterien der Dynamik der Zielfunktionen sehen sie zufällige Änderungen gegenüber nicht zufälligen, aber dennoch nicht voraussagbaren Änderungen. Daneben unterscheiden sie kontinuierliche und diskrete Änderungen.

Eine Möglichkeit zur Klassifizierung dynamischer Funktionen könnte sich aus der von Branke und Schmeck (2003) vorgestellten Messung der Stärke der Änderungen ergeben. Findet nicht nur eine Verschiebung des Optimums statt, sondern ändert sich die Funktion in einem größeren Rahmen, kann die Korrelation zwischen den Fitnesswerten aller Punkte im Suchraum vor der Änderung und nach der Änderung Anhaltspunkte für die Stärke der erfolgten Änderungen geben. Eine hohe Korrelation deutet auf eine kleine Änderungsstärke hin, eine niedrige Korrelation wird bei einer großen Änderungsstärke vermutet.

Allen Klassifizierungen gemeinsam ist jedoch die Unvollständigkeit der Unterteilungen. Auch De Jong (1999) erwähnt, dass es noch weitere Typen dynamischer Funktionen gibt. Allerdings misst er diesen keine hohe Relevanz bei. Der Vollständigkeit halber soll an dieser Stelle jedoch die Menge der häufigsten Formen dynamischer Funktionen etwas differenzierter unterteilt werden, wobei allerdings lediglich eine kurze Aufzählung präsentiert wird. Die unterschiedlichen Formen der Dynamik betreffen folgende Parameter und Bereiche:

- Häufigkeit der Änderungen:
 - Einmalige Änderung.
 - Mehrmalige (endlich oft) Änderung.
 - Unendlich viele Änderungen, unterbrochen durch konstante Phasen.
 - Änderungen erfolgen kontinuierlich.
- Zeitpunkt der Änderungen:
 - Kontinuierliche Dynamik oder längere konstante Phasen.
 - Feste oder variable Abstände zwischen den Änderungszeitpunkten.
 - Vorhersagbarer oder nicht vorhersagbarer Zeitpunkt der Änderung.
 - Interaktion zwischen der Umwelt und dem Algorithmus oder keine Interaktion.
- Arten der Änderungen:
 - Verschiebung der Zielfunktion im Variablenraum (Transformationen der Art: $f_{t+1}(x) = f_t(x + c)$).

- Über die Verschiebung hinausgehende, durch elementare Transformation beschreibbare Änderung der Zielfunktion ($f_{t+1}(x) = c_1 \cdot f_t(c_2 \cdot x)$).
 - Änderung des Charakters der Zielfunktionstopologie, dabei können zusätzliche lokale oder globale Optima entstehen oder bestehende verschwinden.
 - Änderungen betreffen nur einen Teilbereich der Zielfunktion bzw. allgemeiner mehrere, aber wenige Teilbereiche oder sie betreffen den gesamten oder überwiegenden Bereich der Zielfunktion.
 - Änderungen des Bereichs zulässiger Lösungen (Restriktionen).
 - Gleichmäßige oder variable Geschwindigkeit der Bewegung.
 - Oszillierende Änderungen, d.h. permanenter Wechsel zwischen mehreren Zuständen.
- Umfang der Änderungen:
 - Konstante oder variable Größe der Änderungen.
 - Moderate oder abrupte Änderungen.

Bei der Betrachtung dynamischer Optimierungsprobleme sind zahlreiche Kombinationen obiger Formen möglich. Neben diesen Formen haben aber Weicker und Weicker (1999, 2000) eine weitere Form der Dynamik präsentiert. Der zusätzlich eingeführte Aspekt besteht darin, dass zu jedem Zeitpunkt nur ein kleiner, jeweils anderer Ausschnitt der Gesamtfunktion sichtbar ist. Der „unsichtbare“ Bereich erhält jeweils einen konstanten Wert zugewiesen. Allerdings bleibt das Optimum immer konstant und an einem festen Ort, was nicht den allgemeinen Fall dynamischer Problemstellungen darstellt. Der Aspekt der wechselnden eingeschränkten Sichtbarkeit ist interessant, findet in der vorliegenden Arbeit jedoch keine Berücksichtigung.

Branke (1998) zählt auch Funktionen des Typs $f(x + \delta)$ zu den dynamischen Funktionen, wobei δ eine für jeden Funktionsaufruf unterschiedliche normalverteilte Zufallszahl bezeichnet. Dieser Funktionstyp des Parameterrauschens ist jedoch als Spezialfall des Rauschens anzusehen, denn es findet weder eine Änderung der Funktionstopologie statt, noch bewegt sich die Optimalstelle. Dieses Rauschen erinnert mehr an den Fall des klassischen weißen Rauschens $f_{Rauschen} = f(x) + \delta$, da lediglich jede Funktionsberechnung mit einem Störterm der Objektvariablen behaftet ist. Gehorchen die Störterme δ ebenso wie die Funktion selbst gewissen Bedingungen, kann deshalb für eine Stelle auch der echte Funktionswert durch eine mehrfache Funktionsberechnung geschätzt werden. Zur Ähnlichkeit der Optimierung gestörter und echt dynamischer Funktionen siehe Abschnitt 3.1.

1.2.2 Erkennen von Veränderungen

Bei vielen Optimierungsverfahren wird mit trickreichen Methoden versucht, die Dynamik zu beherrschen. Vereinzelt Methoden können von der Kenntnis einer eingetretenen Änderung der Umgebung profitieren. Ein dabei von vielen Autoren vernachlässigter Punkt stellt jedoch das Erkennen von Änderungen der Umgebung dar. Manche Autoren machen es sich leicht, wenn sie voraussetzen, dass der Algorithmus bzw. das System alle notwendigen Informationen über eine Änderung der Umwelt erhält. Von dieser Möglichkeit kann jedoch allgemein nicht ausgegangen werden. Anregungen zu dieser Thematik geben Mori, Kita und Nishikawa (1998) sowie Mori, Kita und Nishikawa (1996).

Eine Möglichkeit besteht sicherlich darin, bereits berechnete Lösungen erneut auszuwerten. Eine Umgebungsänderung hat dann stattgefunden, wenn sich einer dieser Werte geändert hat. Zwei Schwierigkeiten treten bei diesem Vorgehen auf. Zum einen kann sich die Zielfunktion in bestimmten

Bereichen geändert haben, in denen sich keine aktuelle oder früher gefundene Lösung befindet. In diesem Fall würde das Verfahren also davon ausgehen, dass sich die Zielfunktion nicht geändert hätte. Aus Sicht der für diese Untersuchung angenommenen Blackbox-Optimierung bereitet diese Möglichkeit jedoch keine Probleme. Zum anderen können die Funktionsauswertungen gestört sein. Dies führt dazu, dass auch bei identischer Zielfunktion eine wiederholte Funktionsauswertung einer Lösung unterschiedliche Funktionswerte liefert, die abhängig von der Stärke der Störung sehr weit differieren können, weshalb irrtümlich von einer Änderung ausgegangen wird.

Zur Unterscheidung dieser beiden Fälle ist ein alternativer Gedanke, den Fitnessverlauf als stochastischen Prozess zu interpretieren. Wird an ausgewählten gleichen Stellen ausgewertet und ändert sich die Zielfunktion um einen Mittelwert, es handelt sich also um eine stationäre Zeitreihe. Bei Änderung der Zielfunktion sollte ein Test auf Stationarität die Änderung feststellen, da hier mit einer steten Verschlechterung der Fitness zu rechnen ist. Für einen ausführlicheren Vergleich der Situationen im gestörten und im dynamischen Fall siehe Abschnitt 3.1.

In jedem Fall stellt sich unmittelbar die Frage, wieviele bereits berechnete Lösungen erneut ausgewertet werden sollen. Treten Änderungen nur sporadisch und an wenigen Punkten auf, sind zur Feststellung von Änderungen sicherlich mehr Lösungen erneut auszuwerten als bei häufigen und globalen Änderungen. Insgesamt wird durch eine Neuauswertung früherer Lösungen die Gesamtzahl der Funktionsauswertungen pro Zeitintervall erhöht, die meist beschränkt bleiben soll.

Bei populationsbasierten Verfahren lautet eine Empfehlung, eine Änderung anhand einer zurückgehenden Durchschnittsfitness der Population festzumachen. Allerdings kann sich in verschiedenen Varianten eines EA die durchschnittliche Fitness der Population auch verringern, ohne dass sich die Zielfunktion geändert hat. Eine Unterscheidung beider Situationen durch den Algorithmus ist jedoch schwierig.

Ein anderes Vorgehen zum Erkennen von Umweltänderungen benutzt Karr (1999). Zur Erstellung eines Fuzzy Logic Controllers zur Prozesssteuerung vergleicht er das Verhalten des vorhandenen echten Systems mit dem eines internen Computermodells. Unterscheidet sich über eine gewisse Zeit die Verhaltensweise des echten Systems von dem des Computermodells um mehr als einen bestimmten Schwellwert, wird auf eine Änderung der Umgebung geschlossen. Eine Schwierigkeit für ein solches Vorgehen besteht dabei, ein gutes internes Modell zu besitzen.

1.2.3 Definition des dynamischen Optimierungsproblems

Die Betrachtung verschiedener Beiträge zur dynamischen Optimierung ergibt, dass momentan noch keine allgemein anerkannte Definition für die dynamische Optimierung existiert. Manche Autoren umgehen diese Problematik, indem sie nicht einmal den Versuch dazu unternehmen, sondern stattdessen nur eine etwas ungenaue Umschreibung angeben. In anderen Fällen unterscheiden sich die Definitionen stark voneinander. In keinem Fall wird eine allgemein gültige Definition gegeben, die alle denkbaren oder auch nur gängigen Situationen erfasst. Stattdessen wird jeweils nur ein kleiner Teilbereich der dynamischen Zielfunktionen abgedeckt.

Für Trojanowski und Michalewicz (1999a) besteht ein Modell eines (Optimierungs-)Problems aus einem Dreitupel $M(P) = (D, F, C)$. Dabei bezeichnet D den Wertebereich der Variablen, der sich über der Zeit ändern kann, und C eine Menge von Restriktionen. Den n Objektvariablen fügen die Autoren als weitere Dimension die Zeitvariable hinzu. Als Zielfunktion F verwenden sie eine zeitabhängige problemspezifische Funktion $f(x, t) : D \rightarrow \mathbb{R}$. Ein hier wie in vielen anderen Publikationen übliches Einbeziehen der Dynamik durch Hinzufügen einer die Zeit modellierenden Variable zur Menge der Parameter erscheint wenig sinnvoll. Andernfalls könnte durch die Darstellung von

$\min f(x, t)$ der Eindruck erweckt werden, dass sich das Problem lediglich in der Dimension geändert hätte (vom n -dimensionalen in den $n+1$ -dimensionalen Suchraum). In einem solchen Fall wäre die dynamische Optimierung auf den Fall der statischen Optimierung zurückgeführt und es würden sich keine wesentlichen Änderungen ergeben.

Stattdessen geben wir für das dynamische Optimierungsproblem eine Definition an, die durch Einbeziehung des Optimums $x^*(t)$ andeuten soll, dass der entsprechende Zeitparameter effektiv nicht zu beeinflussen ist. Eine Beeinflussung dieses Parameters ist nur möglich, indem vor der nächsten Funktionsauswertung absichtlich eine bestimmte Zeitspanne gewartet wird. Dieses Vorgehen könnte dann Sinn machen, wenn aufgrund einer Analyse bekannt ist, dass sich die Funktion zu bestimmten — vorher bekannten — festen Zeitpunkten in vorteilhafter Weise ändert. In den meisten anderen Situationen wird dieses Vorgehen aber eher kontraproduktiv sein.

In der vorliegenden Arbeit wird für das dynamische globale Optimierungsproblem folgende Definition verwendet:

Definition 1.4 (Dynamisches globales Optimierungsproblem) *Gegeben sei eine statische Zielfunktion f_s wie in Definition 1.1 bezeichnet. O.B.d.A. sei angenommen, dass die statische Zielfunktion das Optimum im Ursprung annimmt. Für das dynamische globale Optimierungsproblem gehen wir nun von Zielfunktionen des Typs*

$$f_d : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$$

mit $f_d(x; x^*(t)) = f_s(x - x^*(t))$ aus. Dabei bezeichnet $x^*(t)$ die Stelle des globalen Minimums zum Zeitpunkt t .

Das dynamische globale Optimierungsproblem besteht nun in der Aufgabe, zu jedem Zeitpunkt t die Stelle des globalen Minimums $x^*(t)$ der Funktion f_d zu finden und somit dem Optimum über dem Zeitverlauf zu folgen. Darf ab einem Zeitpunkt t_0 eine vorgegebene Schranke η nicht überschritten werden ist zu jedem Zeitpunkt t eine Lösung $x^\circledast(t)$ zu finden, sodass folgende Ungleichung erfüllt ist:

$$\forall t > t_0, t, t_0 \in \mathbb{N} : |f_d(x^\circledast(t); x^*(t)) - f_d(x^*(t); x^*(t))| < \eta .$$

Die obige Definition geht also davon aus, dass sich die Lage der Funktion im Raum ändert, aber nicht das Aussehen der Funktion. Der Parameter t enthält eine Angabe der Zeit und ist für jeden Funktionsaufruf unterschiedlich. Gilt $t_1 \neq t_2 \Rightarrow x^*(t_1) \neq x^*(t_2)$, so unterscheiden sich für jeden Zeitpunkt t die Ergebnisse der Funktionsaufrufe. Abhängig von der Reihenfolge der Funktionsaufrufe innerhalb einer Generation könnten also unterschiedliche Gesamtergebnisse erzielt werden. Insbesondere wären am Ende einer Generation die bereits zu Beginn erhaltenen Funktionswerte veraltet. Auch wenn dieser Fall von Interesse sein kann, wird in dieser Arbeit ein einfacherer Fall vorausgesetzt. Es wird angenommen, dass sich die Werte für $x^*(t)$ innerhalb einer Generation nicht unterscheiden. Der statische Fall wird o.B.d.A. mittels $x^*(t) := 0 \forall t$ ebenfalls durch die obige Definition abgedeckt.

1.3 Optimierungskriterien

Die bei der Lösung unterschiedlicher Optimierungsaufgaben im Vordergrund stehenden verschiedensten Kriterien können sich teilweise widersprechen und oftmals nicht gleichzeitig erfüllt werden. Welches dieser Kriterien besondere Gewichtung benötigt, hängt vom jeweiligen Einzelfall ab.

Beim statischen globalen Optimierungsproblem wird die Stelle des globalen Optimums gesucht. Bei zufallsbasierten Lösungsstrategien wie den EA ist es aber sehr wahrscheinlich, dass die Optimalstelle nicht genau erreicht wird. Es könnte also passieren, dass die Optimalstelle mit einer ϵ -Umgebung approximiert wird, die Qualität des Funktionswertes jedoch zu wünschen übrig lässt. Deshalb liegt häufig das Hauptaugenmerk auf einer Reduzierung der Qualitätseinbußen, wodurch eine möglichst große Fitness erreicht werden soll. Liegt die erzielte Fitness in dem von uns vorgegebenen Rahmen, dann ist es unerheblich, ob sich der zugehörige Punkt nahe an der Stelle des globalen Optimums befindet oder ob es sich um eine Stelle bei einem lokalen Optimum handelt. Darüber hinaus ist es noch denkbar, dass beide vorherigen Varianten kombiniert werden, sodass wir eine sehr gute Lösung in unmittelbarer Umgebung des Optimums benötigen.

Zur Bestimmung der Güte oder Eignung eines Optimierungsverfahrens gibt es mehrere Einzelkriterien, die sich meist konträr zueinander verhalten (siehe auch Ott (1993, S. 13)):

- **Konvergenzsicherheit:** Die Konvergenzsicherheit dient als Maß, mit welcher Wahrscheinlichkeit das Verfahren das globale Optimum mit einer gewünschten Genauigkeit annähern kann.
- **Konvergenzgeschwindigkeit:** Zusätzlich zur Konvergenzsicherheit ermöglicht die Konvergenzgeschwindigkeit Aussagen der zum Erreichen einer geforderten Genauigkeit potenziell benötigten Anzahl Funktionsauswertungen.
- **Robustheit:** Die Robustheit dient zur Abschätzung, welchen Einfluss bereits kleine Änderungen der Parameter auf die Qualität des Funktionswertes haben.
- **Ressourcenverbrauch:** Verschiedene klassische Optimierungsverfahren greifen auf speicherplatz- und rechenintensive Operationen wie der Matrizeninvertierung zurück. Aus Effizienzgründen sollte bei einer Optimierung auf solche Operationen verzichtet werden.
- **Informationsbedarf/Anwendungsbreite:** Die zur Anwendung eines Verfahrens notwendigen Voraussetzungen sollten so gering wie möglich ausfallen. Beispielsweise verhindert die Erfordernis der Kenntnis der Ableitungen erster und zweiter Ordnung eine allgemeine Nutzung eines Verfahrens in den Situationen, wo diese Informationen nicht vorliegen.

Das Kriterium der Robustheit einer Lösung gewinnt im Fall der dynamischen Optimierung eine besondere Bedeutung. Robustheit soll in dieser Situation heißen, dass bei Änderungen der Fitnesslandschaft eine gefundene Lösung nicht unbrauchbar werden sollte, da üblicherweise die neue Suche bei der bisher besten gefundenen Lösung fortgesetzt wird. In dynamischen Umgebungen sollte besonders auf rechenintensive Operationen verzichtet werden, weil ansonsten nicht nur die Optimierung verzögert wird, sondern oftmals aufgrund zwischenzeitlicher Änderungen auch das Optimierungsergebnis verschlechtert wird.

Von jedem Einzelfall abhängig sind weitere Einzelkriterien denkbar. Während die obigen Kriterien auch für statische Optimierungsprobleme gelten, kommt bei der Optimierung dynamischer Problemstellungen noch mindestens ein weiteres Kriterium hinzu:

- **Schwerpunkt der Optimierung:** Je nachdem, welcher der beiden Phasen (Einschwingphase oder Verfolgungsphase) eine höhere Priorität zugemessen wird, sind an das Optimierungsverfahren unterschiedliche Kriterien zu stellen. Im ersten Fall ist die Anzahl der Funktionsauswertungen bis zum Erreichen einer Mindestgüte zu minimieren, im zweiten Fall ist nach Erreichen einer Mindestgüte das Optimum in einem möglichst geringen Abstand zu verfolgen.

Kapitel 2

Evolutionäre Algorithmen

Evolutionäre Algorithmen haben sich bereits bei einer Vielzahl von Anwendungsproblemen etabliert. Neben den durch ihre historische Entwicklung bedingten einzelnen Varianten evolutionärer Algorithmen gibt es für unterschiedliche Problemstellungen differierende EA. Diese basieren jedoch alle auf einem gemeinsamen Basisalgorithmus.

Aufgrund des Vorbilds der natürlichen Evolution haben sich zur Beschreibung der Mechanismen evolutionärer Algorithmen die spezifischen Termini der Evolutionsbiologie etabliert. So wird jede zulässige Lösung als Individuum bezeichnet und die Menge der μ aktuellen Lösungen als Population. Eine von λ neu erzeugten und aus anderen Lösungen hervorgegangene Lösung heißt Nachkomme der entsprechenden Eltern. Die Güte einer Lösung wird mit deren Fitness assoziiert. Dem natürlichen Vorbild folgend wird das Aussehen der Funktion deshalb auch als Fitnesslandschaft bezeichnet. Die Fitness selbst beeinflusst die Reproduktionswahrscheinlichkeit des Individuums. Mit den aus der natürlichen Evolution bekannten Operatoren Rekombination, Mutation und Selektion wird die Menge der Lösungen so entwickelt, dass eine möglichst gute bzw. optimale Lösung approximiert wird.

Die Abbildung 2.1 bietet eine Darstellung des Grundalgorithmus evolutionärer Algorithmen (nach Schwefel und Kursawe 1998). Eine genauere Darstellung der einzelnen Schritte wird in den nächsten Abschnitten gegeben. Sie beschränkt sich aber im Wesentlichen auf die in dieser Arbeit verwendeten Evolutionsstrategien. Für weitergehende Einzelheiten sei der Leser auf (Beyer und Schwefel 2002) verwiesen.

2.1 Repräsentation und Fitnessberechnung

Die von Rechenberg und Schwefel initiierten Evolutionsstrategien wurden zunächst für experimentelle (diskrete) Problemstellungen eingesetzt (Rechenberg 1965; Schwefel 1965). Später standen numerische (reellwertige) Optimierungsaufgaben im Mittelpunkt der ES. Dafür wird zur Kodierung der Objekt- oder Entscheidungsvariablen ein Vektor mit n reellen Zahlen x_i verwendet. Dieser Vektor wurde sehr bald durch weitere n_σ Parameter σ_i ergänzt, die Standardabweichungen für die Mutation darstellen und der Selbstadaptation der ES dienen. Ihre genaue Verwendung wird im Abschnitt 2.3 näher beschrieben.

In manchen Implementierungen existiert darüber hinaus noch ein weiterer Satz von bis zu $n_\alpha = n(n-1)/2$ Rotationswinkeln. Wie Rudolph (1992) gezeigt hat, steigt bei Verwendung des kompletten Parametersatzes die zum Erreichen einer bestimmten Genauigkeit benötigte Rechenzeit des EA stark an. So werden Rotationswinkel seltener eingesetzt, obwohl sie sich in verschiedenen Si-

tuationen bewährt haben. Mutationsschrittweiten und Rotationswinkel zusammen werden auch als Strategieparameter des EA bezeichnet. Da die Varianten ohne Rotationswinkel die größte Verbreitung gewonnen haben und sich bewährt haben, beschränken wir uns in der weiteren Beschreibung auf die Verwendung von Mutationsschrittweiten als einzige Strategieparameter.

Zur Berechnung der Fitness einer Lösung wird ausschließlich der Satz der Entscheidungsvariablen verwendet, die Strategieparameter beeinflussen hingegen lediglich die Erzeugung der Nachkommen. Aufgrund des Minimierungsproblems bedeutet ein kleiner Funktionswert eine große Fitness.

2.2 Initialisierung der Startpopulation

Die Initialisierung der Startpopulation kann auf mindestens zweierlei Weise erfolgen. Liegen keine näheren Informationen bzgl. der Lage des globalen Optimums vor, so wird die Startpopulation oftmals gleichverteilt über den gesamten oder einen eingeschränkten Teil des Wertebereich/s der Objektvariablen initialisiert. Ist dagegen aufgrund früherer Untersuchungen bereits eine gute Lösung oder ein schwieriger Startpunkt bekannt, so kann diese/r als Startpunkt der Optimierung dienen.

Neben den Objektvariablen sind zu Beginn auch die Strategieparameter durch Werte zu belegen. Die Initialisierung der Strategieparameter besitzt neben der Beeinflussung der Konvergenzgeschwindigkeit auch Auswirkungen auf die Konvergenzsicherheit. Gute Einstellungen sind dabei i.Allg. sehr schwierig, weshalb die Initialisierung meist aufgrund heuristischer Überlegungen geschieht. Dies führt manchmal zu einer vom Suchbereich unabhängigen und willkürlich aussehenden Festlegung vor dem Start. Zur Lösung einer Problemstellung aus dem Bereich der Traveling Salesperson Probleme (TSP) mithilfe des Simulated Annealing empfehlen Michalewicz und Fogel (2004, S. 124) dagegen z.B. die Standardabweichungen auf ein Sechstel des Definitionsbereichs zu initialisieren. Ungeklärt ist dann aber immer noch die Initialisierung bei einem unbeschränkten Suchraum wie dem \mathbb{R}^n .

Die nachfolgenden Schritte der EA-Schleife werden solange wiederholt bis ein Abbruchkriterium erfüllt ist. Ein vollständiger Zyklus der Schleife wird als Generation bezeichnet.

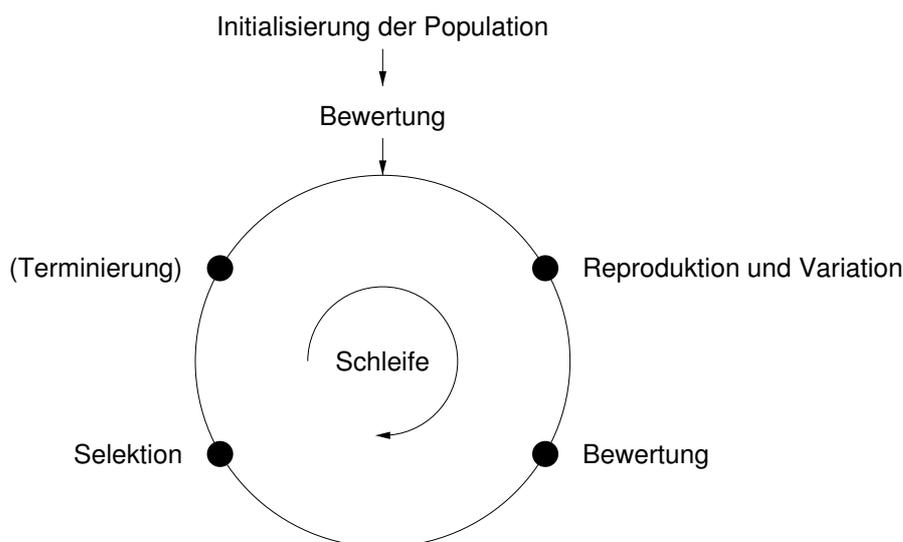


Abbildung 2.1: Grundalgorithmus der EA (leicht verändert aus Schwefel und Kursawe 1998)

2.3 Reproduktion und Variation

Zur Erzeugung eines Nachkommen werden aus der aktuellen Population zuerst Elemente ausgewählt, rekombiniert und anschließend so verändert, dass ein guter Kompromiss zwischen einer volumen- und einer pfadorientierten Suche gewährleistet ist.

Theoretische Analysen und praktische Studien haben gezeigt, dass für eine gute Adaptation EA ein gewisser Nachkommenüberschuss vonnöten ist (Schwefel 1995, S. 133,141,148). Für diese in der Kommastrategie (siehe Abschnitt 2.5) auch als Selektionsdruck¹ bezeichnete Größe hat sich ein Wert von $\lambda/\mu \approx 7$ als eine oftmals benutzte Variante eingebürgert, sodass in jeder Generation insgesamt $\lambda \approx 7\mu$ Nachkommen generiert werden. Die erhöhte Anzahl der Nachkommen wird anschließend in der Selektion wieder auf die Größe μ der Elternpopulation reduziert.

Zur Unterstützung der Rekombination gibt es zwei grundsätzlich verschiedene Sichtweisen (Beyer und Schwefel 2002). Einerseits wird die Rekombination als vorteilhaft angesehen, um Nachkommen zu erzeugen, welche gute Variablen-Cluster der jeweiligen Eltern in einem Individuum kombinieren. Diese als Building Block Hypothese (Goldberg 1989a) bezeichnete Sichtweise wird kontrovers diskutiert (Beyer 1997). Andererseits wird die Rekombination auch als Möglichkeit der genetischen Reparatur gesehen, mit der nachteilige Einstellungen ausgemittelt werden (Beyer 1995).

Für die Rekombination werden in den ES üblicherweise $\rho = 2$ oder $\rho = \mu$ Eltern ausgewählt. Die Arbeit von Beyer (1996) lässt die nähere Untersuchung eines optimalen Wertes für ρ wünschenswert erscheinen. Die Auswahl der zu rekombinierenden Eltern selbst wird als Partnerselektion bezeichnet, spielt in den ES aber nur eine fitnessunabhängige Rolle, bei der die Elemente gleichverteilt aus der Überlebendenpopulation ausgewählt werden.

Bei zwei oder mehr Eltern werden im Fall der so genannten intermediären Rekombination die Merkmalsausprägungen eines Nachkommen mithilfe eines gewichteten Durchschnitts aus den Merkmalsausprägungen der ausgewählten Eltern zusammengesetzt. Im Fall zweier Eltern a und b ergibt sich die Komponente² $i \in \{1, \dots, 2n\}$ des Nachkommen s zu:

$$s_i = z_i * a_i + (1 - z_i) * b_i \text{ ,}$$

wobei z_i eine für jede Komponente einzeln (evtl. zufällig) gewählte Zahl aus dem Intervall $[0, 1]$ bezeichnet. Folgende daraus abgeleitete Rekombinationsmöglichkeiten sind in Gebrauch (siehe Bäck 1996, S. 73):

Zuweisung	Abk.	Bemerkung
$s_i = a_i$	r_-	keine Rekombination
$s_i = a_i \vee s_i = b_i$	r_d	paarweise diskrete Rekombination
$s_i = a_i \vee s_i = b_i^{(i)}$	r_D	globale diskrete Rekombination
$s_i = (a_i + b_i)/2$	r_i	paarweise intermediäre Rekombination
$s_i = (a_i + b_i^{(i)})/2$	r_I	globale intermediäre Rekombination

Dabei bezeichnet $b^{(i)}$ ein für jede Komponente anderes Elternindividuum. Für $\rho > 2$ spricht man von einer Multirekombination, wobei im intermediären Fall die Ausprägungen von ρ Eltern kombiniert werden. Für die Untersuchungen dieser Arbeit wird $\rho = 2$ gewählt und es erfolgt eine diskrete

¹Der Quotient λ/μ wird üblicherweise nur in der Kommastrategie als Selektionsdruck bezeichnet. Wir erweitern hier die Verwendung dieses Begriffes auch auf die Plusstrategie, bei der der Selektionsdruck den Quotienten $(\lambda + \mu)/\mu$ bezeichnet.

²In den ES unterliegen die Strategieparameter ebenso den Variationsoperatoren wie die Objektvariablen. Unter der Annahme der Verwendung von n Schrittweiten besitzt ein Individuum somit insgesamt $2n$ Komponenten.

Rekombination der Objektvariablen und eine intermediäre Rekombination der Strategieparameter. Diese Rekombinationsart empfiehlt Schwefel (1995, S. 146–148) aufgrund praktischer Untersuchungen und findet regelmäßig Verwendung. Wie Kursawe (1999, S. 53) zeigt, kann die Wahl des Rekombinationsoperators jedoch einen größeren Einfluss auf die zu erzielenden Ergebnisse haben.

Die Mutation bildet den zweiten Variationsoperator der EA und wird in den ES immer nach der Anwendung der Rekombination ausgeführt. Der Mutationsoperator sorgt für eine größere Variation der Merkmalsausprägungen nach der Rekombination. Ein Maß für die durchschnittliche Änderung der Objektvariablen während der Mutation bilden die Strategieparameter σ_i . Sie unterliegen in adaptiven ES ihrerseits einer Anpassung während des Ablaufs des Algorithmus. Wichtig ist bei der Durchführung der Mutation, dass zunächst die Schrittweiten und erst danach die Objektvariablen verändert werden. Dies hat den Vorteil, dass die zur Fitnessberechnung herangezogenen Objektvariablen bereits mit den neuen Strategieparametern verändert wurden.

Sei also $a = (x, \sigma)$ das zu mutierende Element. Dann ergibt sich das neue Element $a^{(neu)} = (x^{(neu)}, \sigma^{(neu)})$ zu:

$$\begin{aligned}\sigma_i^{(neu)} &= \sigma_i \cdot \exp(\tau' \cdot N(0, 1) + \tau \cdot N_i(0, 1)) \quad \forall i \in \{1, \dots, n\} \\ x_i^{(neu)} &= x_i + N(0, \sigma_i^{(neu)}) \quad \forall i \in \{1, \dots, n\} ,\end{aligned}$$

wobei $N(\cdot)$ eine für alle i gleiche, $N_i(\cdot)$ eine für jedes i einzeln gewählte normalverteilte Zufallszahl bezeichnet. $\tau, \tau' > 0$ stellen feste Strategieparameter dar, die auch als Lernraten bezeichnet werden und für deren Wahl nach Beyer und Schwefel (2002) folgende Regeln zu beachten sind:

$$\tau = \frac{c}{\sqrt{2\sqrt{n}}} \quad \text{und} \quad \tau' = \frac{c}{\sqrt{2n}} .$$

Das optimale c hängt dabei sowohl von den Populationsgrößen als auch der Zielfunktion ab, ist aber nur für das Kugelmodell bekannt. Bäck (1996, S. 72) bemerkt, dass die Proportionalitätsfaktoren von τ und τ' häufig auf den Wert Eins gesetzt werden.

Die Änderungen der Strategieparameter unterliegen somit einer auf den mehrdimensionalen Raum erweiterten logarithmischen Normalverteilung (Beyer und Schwefel 2002), die dafür sorgt, dass nur positive Werte angenommen werden können. In praktischen Anwendungen muss zusätzlich darauf geachtet werden, dass die Mutationsschrittweiten nicht unter ein bestimmtes Minimum fallen, weil sonst wegen der beschränkten Darstellungsgenauigkeit keine Änderung möglich wäre. Schwefel (1975, S. 132) gibt hierfür die zwei Restriktionen

$$\sigma_i \geq \epsilon_a > 0 \quad \text{und} \quad \sigma_i \geq \epsilon_b |x_i| \text{ mit } 1 + \epsilon_b > 1$$

an, wobei ϵ_a, ϵ_b in Abhängigkeit der jeweils benutzten Gleitkomma-Architektur gewählt werden müssen.

Für den Fall, dass für alle Parameter nur eine Standardabweichung benutzt wird, reduziert sich die Berechnung der neuen Standardabweichung auf

$$\sigma^{(neu)} = \sigma \cdot \exp(\tau_0 \cdot N(0, 1))$$

wobei nach Beyer und Schwefel (2002) für die Wahl von τ_0 folgende Regel beachtet werden sollte:

$$\tau_0 \propto \frac{1}{\sqrt{n}} .$$

Für die Änderung der Objektvariablen wird statt der Normalverteilung bisweilen alternativ die Verwendung einer anderen Verteilung wie der Cauchyverteilung untersucht (Rudolph 1997, 1998). Die Ränder der Cauchyverteilung sind wesentlich dichter besetzt als die der Normalverteilung, weshalb die Cauchyverteilung keinen Erwartungswert besitzt. Aufgrund der gegenüber einer Normalverteilung höheren Wahrscheinlichkeit für große (absolute) Ereignisse erhofft man dadurch, leichter ein lokales Optimum verlassen zu können. Allgemein hat sich die Benutzung alternativer Verteilungen jedoch nicht durchsetzen können.

2.4 Adaptation

Das Verhalten EA wird durch eine Vielzahl von Parametereinstellungen beeinflusst. Viele dieser Parameter werden vor dem Start des Algorithmus aufgrund bestimmter Überlegungen heraus festgelegt und ändern sich üblicherweise während der Suche nicht. Demgegenüber ist es bereits in allen modernen EA-Varianten üblich, andere Strategieparameter während des Laufs zu verändern. Dies resultiert aus der Erkenntnis, dass für eine gegebene Problemstellung evtl. zwar eine gute durchschnittliche Parametereinstellung gefunden werden kann, die sich aber schon bei einer geringen Veränderung der Problemstellung als ungünstig erweisen kann. Darüber hinaus wird es in den meisten Problemstellungen keine Parametereinstellung geben, die für die gesamte Laufzeit des EA optimal ist.

Dieser Umstand impliziert, dass sich die Parameter im Verlauf des Algorithmus anpassen sollten. Um eine bessere allgemeine Anwendbarkeit zu besitzen, sollte der EA dabei in der Lage sein, die besten Parametereinstellungen selbstständig zu finden. EA mit dieser Eigenschaft heißen selbstadaptiv und entsprechen mittlerweile dem Stand der Technik.

In den folgenden Abschnitten soll ein Überblick über die verschiedenen Möglichkeiten der Adaptation sowie der am meisten lohnenswert erscheinenden adaptierbaren Strategieparameter gegeben werden.

2.4.1 Adaptierbare Parameter

Grundsätzlich eignen sich alle zu Beginn eines Algorithmus festzulegenden Parameter für eine Adaptation. Abhängig vom Verlauf des EA können Veränderungen von der Anzahl der Eltern über die Anpassung der Selektion bis hin zur Repräsentation der Individuen wünschenswert, sinnvoll oder gar notwendig sein.

Eine Konzentration auf den Mutationsoperator ist deshalb gerechtfertigt, weil dieser Operator in der Literatur am häufigsten untersucht wurde und somit zahlreiche Varianten verfügbar sind. Der Adaptation des Rekombinationsoperators wurde dagegen bislang weniger Aufmerksamkeit geschenkt. Dies erklärt Bäck (1998) damit, dass der Zusammenhang zwischen der Wahl unterschiedlicher Strategieparameter des Rekombinationsoperators und der Fitness der Nachkommen nicht so offensichtlich ist wie bei den Strategieparametern der Mutation. Dennoch gibt es Untersuchungen, die die Adaptation des Rekombinationsoperators beinhalten. Bäck (1998) stellt die gängigsten in der Literatur genannten Adaptationsmöglichkeiten für den Rekombinationsoperator vor. Der Autor berichtet von den frühen Untersuchungen von Schaffer und Morishima (1987), die in einem GA die Stellen des Crossovers durch einen zusätzlichen Binärvektor variieren lassen. Dieser unterliegt seinerseits einer Rekombination und anschließender Mutation. Die Experimente mit einer zu geringen Problemdimension erlauben jedoch keine Schlüsse über den wirklichen Nutzen dieses Vorgehens.

Eine andere interessante Form der Adaptation haben Ostermeier, Gawelczyk und Hansen (1994) eingeführt. Bei dieser derandomisierten Mutationsschrittweitenanpassung erfolgt die Anpassung der Mutationsschrittweiten aufgrund des bisherigen Verlaufs des EA. Dabei wird ein gewichtetes Mittel aus dem bisherigen langfristigen Wert und dem Wert des aktuell besten Individuums gebildet. Dieses Vorgehen wird u.a. damit begründet, dass es bei einer für jede Komponente unterschiedlichen Mutationsschrittweite nicht immer einen direkten Zusammenhang zwischen der Mutationsschrittweite und der Länge des ausgeführten Schrittes gibt. So kann aufgrund stochastischer Schwankungen ein großer Schritt trotz einer kleinen Schrittweite ausgeführt werden. Bei der derandomisierten Schrittweitenanpassung wird die neue Schrittweite aber abhängig von dem zurückgelegten Schritt des besten Individuums angepasst. Ebenso wie die u.a. in (Ostermeier und Hansen 1999) vorgestellten Erweiterungen hat sich diese Technik in der EA-Gemeinde bislang aber noch nicht als Standard durchsetzen können.

Wie Bäck (1999) erwähnt, wird von dem nach dem Prinzip der selbstadaptiven Mutationsschrittweite arbeitenden Mutationsoperator in den EA ein ausgezeichneter Strategieparameter (τ in den ES) verwendet, der wiederum einen Einfluss auf die Funktionsweise der Mutation besitzt. Dass auch diese und die weiteren üblicherweise fest gewählten Parameter wie die Anzahl der Eltern und die Anzahl der Nachkommen variiert werden können, zeigen u.a. zahlreiche Untersuchungen zu Meta-Strategien. Gemeinsam ist diesen, dass verschiedene Strategien miteinander konkurrieren und ein übergeordneter EA dazu benutzt wird, die besten Strategien zu selektieren und deren feste Parameter weiter zu evolvieren. Neben dem grundlegenden Konzept und einer formalen Definition stellt Freisleben (1997) auch einige konkrete Varianten vor. Zu den Meta-Strategien führen unter anderem auch Kursawe (1999) und Schlierkamp-Voosen (1999) umfangreiche Untersuchungen durch.

2.4.2 Typen der Adaptation

Die folgende Einteilung adaptiver Verfahren basiert im Wesentlichen auf den Ausführungen von Hinterding, Michalewicz und Eiben (1997), die sich auf den Artikel von Angeline (1995) stützen und die dort vorgenommene Einteilung erweitern. Weitere Übersichten und Anregungen enthalten auch (Bäck 1992; Bäck und Schütz 1996; Bäck 1997; Bäck 1998).

Hinterding et al. (1997) treffen zunächst eine grundsätzliche Unterscheidung der Adaptation in eine statische und eine dynamische Adaptation. Bei der statischen Adaptation erfolgt vor dem Start des EA eine Wahl der Parameter. Diese behalten im gesamten anschließenden Lauf des EA ihre Werte. Im Rahmen der vorliegenden Arbeit soll dieser Typ nicht zur Adaptation gezählt werden, da während des EA-Laufs keine Veränderung der Parametereinstellungen stattfindet.

Zu den von Hinterding et al. (1997) gezählten dynamischen Adaptationsmechanismen gehören deterministische, adaptive und selbstadaptive Methoden. Deterministische Adaptationsmechanismen verändern die Parameterwerte unabhängig vom EA-Verlauf nach festen Regeln. Die Autoren erwähnen als Beispiel hierfür eine von der Generation abhängige Reduzierung der Mutationswahrscheinlichkeit, wie sie in vielen GA benutzt wird. Ein deterministisches Scheduling findet sich bspw. auch beim Simulated Annealing oder Threshold Accepting. Ohne auf die Details dieser Verfahren einzugehen sei hier angemerkt, dass beim Simulated Annealing die bei der Selektion berücksichtigte Temperatur T mit zunehmender Generationszahl reduziert wird (Metropolis et al. 1953; Kirkpatrick et al. 1983; Dueck und Scheuer 1990).

Werden Zwischenergebnisse des EA benutzt, um die Parametereinstellungen zu variieren, gelangen wir zu adaptiven Schedules. Hierbei erfolgt eine Überwachung des Evolutionsprozesses und

abhängig vom bisher erzielten Fortschritt werden die Strategieparameter geeignet angepasst. Ein typisches Beispiel bildet die 1/5-Erfolgsregel von Rechenberg, bei der die Mutationsschrittweite abhängig von der relativen Häufigkeit erfolgreicher Mutationen zur Gesamtzahl durchgeführter Mutationen um einen konstanten Faktor erhöht oder verringert wird. Verschiedene Autoren klassifizieren solche Schedules häufig als selbstadaptiv. Dieser Einteilung schließen wir uns in der vorliegenden Arbeit jedoch nicht an.

Im Fall der Selbstadaptation werden die Parameter mit im Individuum kodiert und unterliegen wie die Variablen der Optimierungsaufgabe ebenfalls den Variationsoperatoren. Obwohl die Parameterwerte die Fitness nicht direkt beeinflussen, sollten Individuen mit *guten* Parametereinstellungen im Verhältnis zu Individuen mit *schlechten* Parametereinstellungen zu Nachkommen mit einer größeren Überlebenswahrscheinlichkeit führen, wodurch sich auch die guten Parameterwerte in der Population ausbreiten sollten. Eine große Bedeutung hat dabei der Selektionsoperator, der entscheidet, welches Individuum in die nächste Generation übernommen wird. Da der Selektionsoperator die Strategieparameter beim Selektionsprozess unberücksichtigt lässt, ist ein indirekter Einfluss der Strategieparameter auf die Fitness notwendig. Dies erfolgt meist dadurch, dass die Objektvariablen eines Individuums mit den bereits veränderten Strategieparametern variiert werden. Wie in Abschnitt 2.3 gezeigt, hat Schwefel diesen Mechanismus bereits frühzeitig in die ES integriert, wo neben den Objektvariablen u.a. auch die Werte zur Mutationsschrittweite mit im Individuum kodiert werden.

Verschiedene Anpassungsmethoden können nicht eindeutig einer der genannten Varianten zugeordnet werden. So bezeichnet Bäck (1998) bspw. die derandomisierte Mutationsschrittweitenanpassung als zwischen einer adaptiven und einer selbstadaptiven Anpassung angesiedelt. Bereits weiter oben wurde angedeutet, dass darüber hinaus manche Strategien von einzelnen Autoren dem einen Typ, von anderen Autoren aber einem anderen Typ zugeordnet werden. Diese Problematik soll an dieser Stelle jedoch nicht weiter vertieft werden.

2.4.3 Stufen der Adaptation

Neben den verschiedenen Typen der Adaptation unterscheiden Hinterding et al. (1997) vier Stufen, auf denen die Adaptation wirkt.

Zunächst kann die Adaptation auf der Ebene der Umwelt stattfinden. Denkbar ist hierbei etwa, dass sich der in den meisten EA bei Restriktionsverletzungen zu addierende Strafterm mit der Generationenzahl erhöht, bis zuletzt jede unzulässige Lösung eine schlechtere Fitness besitzt als jede zulässige Lösung.

Die Adaptation kann aber auch Parameter betreffen, die Auswirkungen auf die gesamte Population besitzen. Als Beispiel einer solchen adaptiv genannten Adaptationsregel ist auch hier die 1/5-Erfolgsregel von Rechenberg zu nennen. Die Autoren bemerken, dass auf dieser Ebene keine Selbstadaptation bekannt ist.

Erfolgt die Adaptation der Mutationsschrittweiten mit nur einer Schrittweite für alle Objektvariablen eines Individuums, so geschieht die Adaptation auf der Ebene der Individuen. Im Unterschied zur individuenbasierten Adaptation erfolgt bei einer komponentenweisen Anpassung die Adaptation für jede Komponente eines Individuums getrennt.

2.5 Selektion

Die zuweilen als Umweltselektion bezeichnete Selektion zur Auswahl der neuen Elternpopulation wird benötigt, um den in den ES vorhandenen Nachkommenüberschuss auf die Anzahl der Eltern zu reduzieren. In dieser Phase werden die besten (d.h. mit der höchsten Fitness ausgestatteten) Individuen zur Bildung der nachfolgenden Elternpopulation ausgewählt. Hierbei unterschied man früher zwischen den Varianten der Plus-Selektion ($\mu + \lambda$) und der Komma-Selektion (μ, λ). Im ersten Fall besteht die nächste Elternpopulation aus den besten Individuen der bisherigen Eltern und der Nachkommen. Dies hat den Vorteil, dass im Laufe der Evolution kein Rückschritt möglich ist. Andererseits birgt es die Gefahr, dass ein mögliches „Superindividuum“ (bezogen auf den Fitnesswert, das aber zugleich ungeeignete Strategieparameter besitzt) nachfolgende Generationen dominiert und damit die Selbstadaptation behindert. Diese Möglichkeit wird durch die Komma-Selektion vermieden. In diesem Fall wird die nächste Elterngeneration aus den besten Individuen der Nachkommen gebildet. Allerdings kann es hierbei zu einem Rückschritt kommen, und zwar dann, wenn die Wahrscheinlichkeit zu gering ist, um in jeder Generation mindestens einen gleich guten Nachkommen wie den besten Elter zu generieren. Schwefel (1987) empfiehlt die Komma-Selektion für unbeschränkte reelle Suchräume wie sie in der vorliegenden Untersuchung zugrunde liegen. Beide Selektionsformen werden auch unter $(\mu \ddagger \lambda)$ -ES zusammengefasst. Unter Einbeziehung der Multirekombination aller Eltern subsummiert Beyer (2001) beide Varianten zur $(\mu/\mu \ddagger \lambda)$ -ES.

Die Plus- und Komma-Selektion sind Spezialfälle der von Schwefel und Rudolph (1995) eingeführten $(\mu, \kappa, \lambda, \rho)$ -ES. Während der Parameter ρ wiederum die Anzahl der für jeden Nachkommen zu rekombinierenden Eltern bezeichnet, gibt der Parameter κ die maximale Lebensdauer eines einzelnen Individuums an. Dabei ergibt sich mit $\kappa = 1$ die Komma-Selektion und mit $\kappa = \infty$ die Plus-Selektion. Es ist möglich, dass dazwischen liegende Werte für κ in verschiedenen Situationen Vorteile gegenüber den beiden Extrema besitzen. Ansatzweise wurde dies bereits von Wörmann (2000) betrachtet. Für dynamische Umgebungen könnte sich eine κ -Selektion als vorteilhaft erweisen, wenn sich das Optimum mit einer Bewegungsfrequenz Δg ändert, die mit dem Wert κ übereinstimmt. Erste eigene diesbezügliche Untersuchungen waren jedoch wenig viel versprechend. Auch in umfangreicheren Studien im statischen Fall konnten keine positiven Auswirkungen festgestellt werden (Schönemann 2004b).

2.6 Abbruchkriterium

Zur Aufstellung geeigneter Abbruchkriterien sind zunächst zwei Situationen zu unterscheiden. In praktischen Anwendungen kann es zum einen erforderlich sein, eine bestimmte Güte zu erreichen. Zum anderen wird die Ausführungsdauer des Algorithmus durch einen maximalen Zeitfaktor begrenzt sein. Für Untersuchungen sind diese ad hoc-Kriterien jedoch ungeeignet. Hier muss das Abbruchkriterium individuell auf die zu untersuchende Problematik abgestimmt und unabhängig von der gewählten Implementierung sein.

Folgende Abbruchkriterien sind üblich:

- Beenden des Algorithmus nach Ablauf einer zu Beginn festgelegten eventuell unnötig großen maximalen Anzahl von Reproduktionszyklen oder Funktionsauswertungen. Wenn die Funktionsauswertungen eine unterschiedlich lange Ausführungszeit benötigen, ist alternativ der Abbruch nach Ablauf einer festgelegten Zeitspanne (CPU-Zeit) denkbar. Beide Kriterien sind eher ungünstig, weil sie unabhängig vom erzielten Fortschritt sind.

- Den beiden vorangehenden Kriterien vorzuziehen sind daher immer problemspezifische Abbruchbedingungen. Hierzu zählt insbesondere das Beenden des Algorithmus beim Erreichen eines bestimmten Fitnesswertes.
- In bestimmten EA-Varianten ist der Abbruch beim Erreichen einer Homogenität der Population angeraten, da in diesem Fall ein Fortschritt nur noch geringe Wahrscheinlichkeit besitzt.
- Für die ES bietet sich dagegen an, den Algorithmus abzubrechen, wenn in den letzten Generationen kein Fortschritt mehr erzielt wurde. Dabei sollte die Anzahl der relevanten Generationen von der Dimension des Variablenraums abhängen.

Kapitel 3

Evolutionäre Algorithmen für dynamische Umgebungen

3.1 Vergleich zwischen Dynamik und Rauschen

Das Einfachste ist es, die von der statischen Optimierung her bekannten Verfahren auch zur Optimierung dynamischer Problemstellungen einzusetzen. Hierbei gilt es jedoch einige Punkte zu berücksichtigen.

Da bei der statischen Optimierung die zu optimierende Funktion stets fest ist, wird jede zusätzliche Funktionsauswertung dazu beitragen, die Kenntnisse über diese Funktion zu erweitern. Dadurch kann ein internes Modell der Funktion erstellt werden. Ein solches Vorgehen wird für statische Optimierungsprobleme von verschiedenen Autoren vorgeschlagen, um die Zahl zeitaufwändiger Funktionsauswertungen zu reduzieren (Ratle 1998; Beltagy et al. 1999). Demgegenüber tritt bei der dynamischen Optimierung der Fall auf, dass sich bereits berechnete Funktionswerte ändern, weshalb sich die Erstellung eines internen Modells der Funktion noch schwieriger gestaltet. In manchen Situationen ist es sogar so, dass ein bereits berechneter Funktionswert unmittelbar nach Berechnung seine Gültigkeit verliert, d.h. schon bei der Betrachtung des Wertes nicht mehr korrekt ist.

Obwohl es im gestörten Fall im Gegensatz zum dynamischen Fall meistens möglich ist, durch mehrfache Auswertung einer Stelle den entsprechenden Funktionswert mit sehr hoher Genauigkeit zu bestimmen, gibt es zwischen der Optimierung gestörter und dynamischer Funktionen auffällige Gemeinsamkeiten. Dies wird auch von anderen Autoren erkannt. Allerdings gibt es dabei zahlreiche Unterschiede. So erwähnt Branke (1998) zwar auch die Ähnlichkeit mit der Optimierung verrauschter Funktionen, bezeichnet aber folgende Unterschiede:

$$f_{\text{Rauschen}} = f(x) + \delta \text{ gegenüber } f_{\text{Dynamik}} = f(x + \delta) .$$

Hierbei wird aber keine Änderung der Funktionstopologie berücksichtigt, es ändert sich nicht einmal die Optimalstelle, es wird lediglich jede Funktionsberechnung mit einem Störterm der Objektvariablen behaftet. Insgesamt handelt es sich nur um eine spezielle Form der Störung, nämlich die der Parameterstörung.

Eine kurze Überlegung soll nun eine Anregung für den prinzipiellen Unterschied liefern. Im Fall der Parameterstörung hängt die Streuung des Funktionswertes stark von der Zielfunktion ab. Zum Vergleich beider Formen legen wir das n -dimensionale Kugelmodell ($f(x) = \sum_{i=1}^n x_i^2$, siehe

Kapitel 5.1) zugrunde. Eine Parameterstörung führt dann zu folgender Gleichung:

$$f(x + \delta) = \sum_{i=1}^n (x_i + \delta_i)^2 .$$

Für eine Störung in einer (hier der n -ten) Dimension folgt

$$\begin{aligned} f(x + \delta) &= \sum_{i=1}^{n-1} x_i^2 + (x_n + \delta_n)^2 \\ &= f(x) + (2x_n\delta_n + \delta_n^2) . \end{aligned}$$

Findet eine Störung in allen Koordinaten mit unterschiedlichen Stärken statt, so gestaltet sich obige Formel aufwändiger.

Nun kann eine Bewegung des Optimums zwar am ehesten mit einer Parameterstörung verglichen werden, dennoch existieren zwischen beiden doch Unterschiede. Während im Fall der Störung δ als normalverteilt angenommen wird, ist im dynamischen Fall die Bewegungsstärke oftmals konstant. Die stochastischen Terme δ_i werden dann durch die vom Mutationsoperator abhängenden Streuungen der Nachkommen um das neue Optimum erzeugt. In diesem Fall ist also die Mutation das entscheidende Element. Eine approximative Bestimmung, wie die Terme δ zu wählen sind, damit in den Fällen des Rauschens und der Parameterstörung der Funktionswert um den gleichen Mittelwert schwankt, sollte auch für eine (μ, λ) -ES mit $\mu > 2$ möglich sein. Der Versuch, das Verhalten eines realen EA im Fall des Rauschens und der Dynamik analytisch zu vergleichen, gestaltet sich jedoch schwieriger und steht nicht im Mittelpunkt der vorliegenden Untersuchungen. Dieses Ziel wird deshalb hier nicht weiter verfolgt. Stattdessen beschränken wir uns auf einen exemplarischen experimentellen Vergleich beider Formen, jedoch ohne den umfangreichen Untersuchungen des Teils II (ab Seite 55) vorzugreifen.

Beim Kugelmodell mit nur geringem Rauschen bzw. nur geringer Dynamik weisen die beiden Darstellungen der Fitness eines Einzellaufs ein ähnliches Aussehen auf (Abb. 3.1 oben). Die in der Abbildung betrachtete dynamische Funktion lässt sich jedoch bei der hier gewählten Größe der Änderungsterme etwas besser optimieren als die verrauschte Funktion. Dabei fällt auf, dass die verrauschte Funktion stärkere Schwankungen von einer Generation zur nächsten aufweist. Allerdings ist die Gesamtstreuung bei der dynamischen Funktion etwas größer. Werden die Änderungsterme anders gewählt, sind die Funktionswerte teilweise sehr unterschiedlich. Wie der unteren Abbildung 3.1 zu entnehmen ist, lässt sich die dynamische Funktion mit der Bewegungsstärke 1.0 schwieriger optimieren als die verrauschte Funktion mit einer Varianz des Störterms in gleicher Größe.

Abhängig von der Größe des Änderungsterms sind die Auswirkungen im Fitnessraum ähnlich, während die Betrachtung des Objektraums stärkere Unterschiede zeigt. Die Abbildung 3.2 verdeutlicht dies anschaulich. Im verrauschten Fall sind die Punkte relativ gleichmäßig um das im Ursprung liegende Optimum verteilt. Im dynamischen Fall geben die Werte die jeweilige Abweichung vom Optimum an, sodass sie vergleichbar mit dem verrauschten Fall sind. Deutlich ist für $\delta \geq 0.01$ die Lage der meisten Individuen im III. Quadranten zu erkennen. Dies zeigt, dass das beste Individuum in den meisten Fällen etwas hinter dem zu verfolgenden Optimum hinterherläuft. Extrem ist die Situation für $\delta = 100.0$, für das die Menge der besten Punkte im verrauschten Fall zu einem Punkt zu degenerieren scheint, während im dynamischen Fall eine große Punktwolke links unten erkennbar ist. Das Fazit für diese Untersuchungen lautet, dass für kleines δ die dynamische Funktion etwas leichter zu optimieren ist als die verrauschte Funktion. Mit wachsendem δ kehrt sich dieses Verhalten aber um.

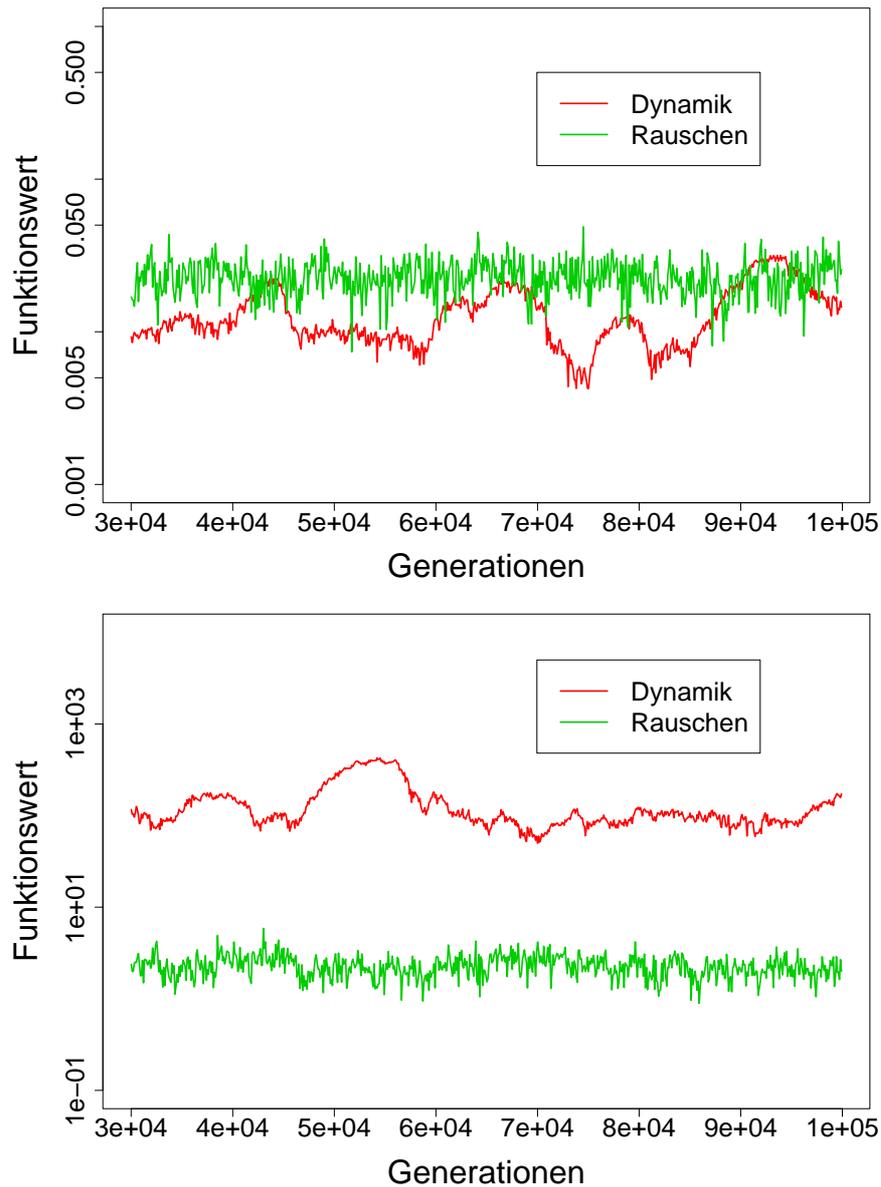


Abbildung 3.1: Vergleich der Fitness des jeweils besten Individuums einer (15, 100)-ES beim dynamischen und gestörten 30-dimensionalen Kugelmodell. Bei der Störung handelt es sich um eine additive $N(0, 0.01)$ (oben) bzw. $N(0, 1.0)$ (unten) verteilte Störung. Im dynamischen Fall erfolgt in jeder Generation eine Verschiebung in allen Dimensionen mit einer konstanten Gesamtstärke von 0.01 (oben) bzw. 1.0 (unten).

In jedem Fall wird im Gegensatz zur statischen Optimierung gestörter Funktionen bei der dynamischen Optimierung der durch zusätzliche Funktionsauswertungen gewonnene Informationsgewinn oftmals beschränkt sein, da sich das Optimum in der Zwischenzeit bereits bewegt haben könnte. Daraus ergeben sich zahlreiche Konsequenzen, zu denen zählt, dass der Wert einer Funktionsauswertung eine nur noch zeitlich begrenzte Aussagekraft besitzt. Deshalb ist auch der Vergleich zweier Funktionswerte nicht mehr uneingeschränkt möglich, dies umso mehr, je weiter die beiden Funktionsauswertungen zeitlich auseinander liegen.

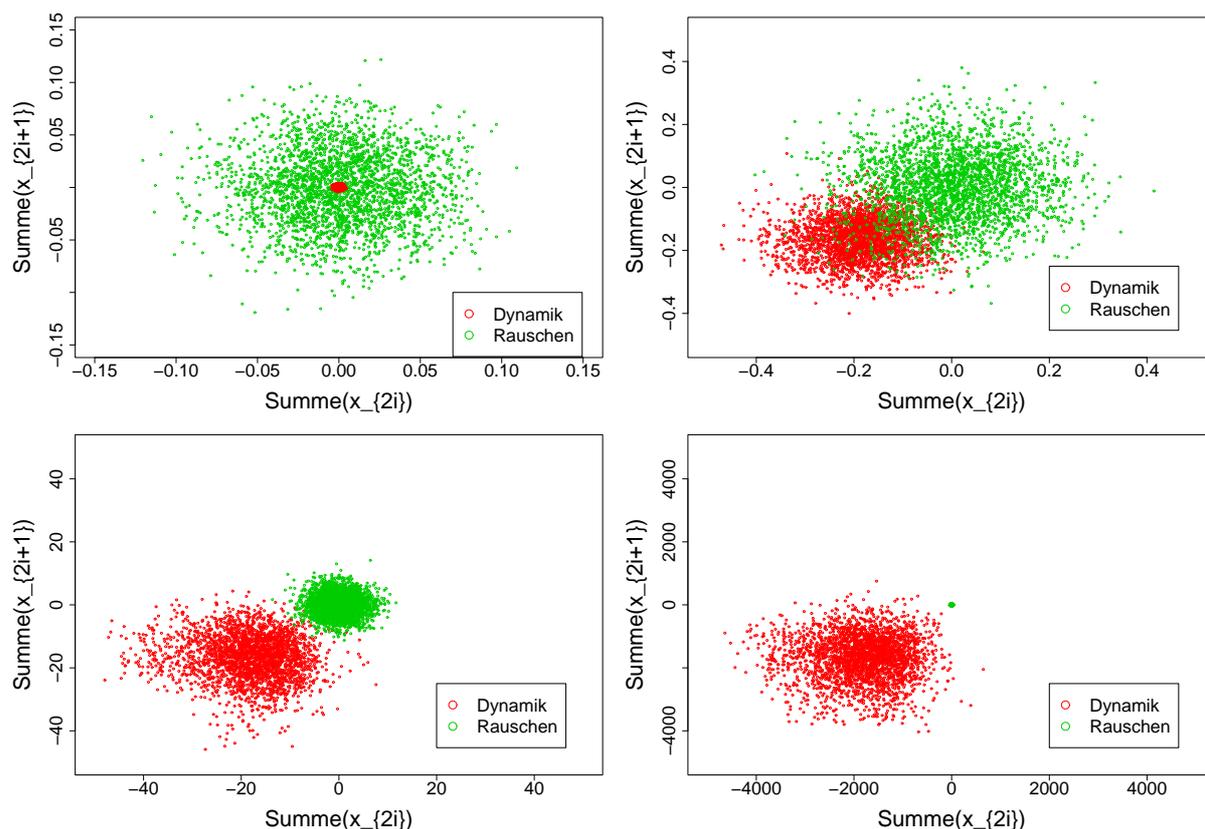


Abbildung 3.2: Vergleich des jeweils besten Individuums jeder Generation einer (15, 100)-ES beim dynamischen und gestörten 30-dimensionalen Kugelmodell. Aufgetragen ist jeweils die Summe der Komponenten mit geradem Index (x -Achse) bzw. mit ungeradem Index (y -Achse). Bei der Störung handelt es sich um eine additive $N(0, \delta)$ -verteilte Störung. Im dynamischen Fall erfolgt in jeder Generation eine Änderung in allen Dimensionen mit einer konstanten Gesamtstärke von δ . Der Änderungsterm lautet $\delta = 0.001$ (oben links), $\delta = 0.01$ (oben rechts), $\delta = 1.0$ (unten links), $\delta = 100.0$ (unten rechts).

Der Vergleich mit gestörten Zielfunktionswerten soll an dieser Stelle lediglich dazu dienen, Überlegungen zur Verwendung speziell für diese Funktionstypen entwickelter Operatoren auch im Bereich der dynamischen Optimierung anzuregen. Ein denkbare Verfahren ist bspw. die von Markon et al. (2001) betrachtete Threshold-Selektion.

Neben dem durchgehenden Einsatz eines Standard-EA kann bei einer Änderung der Umgebung alternativ ein Neustart des Standard-EA erfolgen oder ein Übergang zu einer speziellen Variante, von denen einige der gebräuchlichsten in den folgenden Abschnitten 3.2 bis 3.4 kurz vorgestellt werden sollen.

3.2 Diversitätserhaltung

Eine immer wieder geforderte Eigenschaft stellt eine hohe Diversität innerhalb der Population dar. Auf statische Problemstellungen angewandte Standard-EA haben die Eigenschaft, mit zunehmender Generationenzahl gegen eine einzelne Lösung zu konvergieren. Diese Konvergenz führt dazu, dass die Population aus nur noch gering voneinander abweichenden Individuen besteht. Handelt es sich bei dieser Lösung um eine Lösung nahe dem globalen Optimum, so ist eine Konvergenz gegen diese

Lösung erwünscht. Problematisch ist aber der häufig auftretende Fall, dass eine frühzeitige Stagnation in einem lokalen Optimum erfolgt. Aufgrund der in den traditionellen GA unterrepräsentierten Mutation führt eine homogene Population zu einem Stillstand in der weiteren Entwicklung, weshalb gerade für diese EA-Variante Mechanismen zur Erhöhung der Diversität entwickelt wurden.

In dynamischen Umgebungen kann eine Diversitätserhöhung aber auch für andere EA-Varianten vorteilhaft sein. In diesem Fall ist eine schnellere Reaktion auf eine Änderung der Umwelt möglich, wenn der Suchbereich der neuen Optimalstelle bereits mit Individuen belegt ist oder von anderen Orten schneller zum neuen Ziel gelangt werden kann. Für den Bereich der Diversitätserhaltung unterscheiden Trojanowski und Michalewicz (1999a) mehrere Varianten:

Sharing und Crowding: Zur Vermeidung zu großer Cluster in der Population wird beim Sharing die Fitness eines Individuums verringert, wenn sich in unmittelbarer Umgebung weitere Individuen befinden. Hierdurch soll vermieden werden, dass sich die Lösungen in einem lokalen Optimum zu schnell vermehren. Beim Crowding ersetzt ein mit den üblichen Verfahren generiertes neues Individuum dasjenige Element aus einer zufällig gewählten Untermenge der Eltern, das die größte Ähnlichkeit mit dem neuen Element aufweist.

Nicheing: Beim Nicheing werden bewusst Cluster gesucht, um eine intensivere Suche in mehreren Bereichen voranzutreiben. Eine Rekombination kann dann nur zwischen den Elementen eines Clusters erfolgen. Dadurch bilden sich mit der Zeit an verschiedenen Stellen Subpopulationen aus.

Multipopulationen: Der Übergang zu Multipopulationsstrukturen (Sprave 1999) bringt zunächst den Vorteil einer möglichen Parallelisierung mit sich. Darüber hinaus kann es auch noch zu einer größeren Diversität führen. Solche Verfahren sind recht beliebt (Branke 2002, S. 58).

Maximierung der Entropie: Eng mit dem ersten Punkt verbunden ist die Technik der Maximierung der Entropie. Hier wird mit einer einzigen Population gearbeitet, wobei aus der Menge aller potenziellen Eltern diejenigen Elemente präferiert werden, für die sich eine Maximierung der Entropie ergibt. Während die Fitness zur Diversität im Phänotypraum benutzt wird, dient die Berücksichtigung der Entropie zur Diversitätserhaltung im Genotypraum, da hier auch die Seltenheit bestimmter Gene in einzelnen Komponenten in den Selektionsprozess einfließen (Mori, Kita und Nishikawa 1998). Ein Mindestmaß an Diversität wird bereits dadurch erreicht, dass das mehrfache Auftreten eines Individuums in der Population verhindert wird (Jansen und Wegener 2002).

Berücksichtigung des Alters der Individuen: Im Abschnitt 2.4.1 wurde erwähnt, dass neben den in den ES üblichen beiden Selektionsformen Komma und Plus eine Einbeziehung des Alters κ vorgeschlagen wurde. Dieser Aspekt gewinnt insbesondere in der dynamischen Optimierung an Bedeutung. Je nachdem, wie stark die Dynamik ist, werden bereits ausgewertete Individuen einerseits einen veränderten Fitnesswert besitzen, andererseits werden sie mehr oder weniger weit vom neuen Optimum entfernt liegen und somit lediglich einen die Diversität betreffenden positiven Effekt für den EA haben. In dieser Situationen ist es sinnvoll, eine bereits gefundene Lösung nach Ablauf einer gewissen Zeit aus der Population zu entfernen.

Benutzung zufälliger Immigranten: Zum permanenten Aufrechterhalten der Diversität können in die Population ständig neue, zufällig gewählte Individuen immigrieren. Abhängig von der Stärke der Dynamik kann durch Variieren des Prozentsatzes neuer Individuen an der Gesamtzahl aller Individuen auch die Größe der Diversität gesteuert werden (Grefenstette 1992).

Erhöhung der Mutationsrate (Hypermuation): Ebenso kann bei einer festgestellten größeren Änderung der Umgebung bewusst eine höhere Mutationsrate eingestellt werden. Nach einer Änderung kann eine Diversitätserzeugung auch durch eine kurzfristige Erhöhung der Mutationsrate erfolgen (Grefenstette 1999). Runarsson (2000, S. 82) führt dagegen eine besonders starke Mutation für bestimmte Komponenten einzelner Individuen durch.

Populationsgrößen: Durch die Wahl großer Populationsparameter (μ und λ) erhofft man sich bessere Ergebnisse aufgrund einer stärkeren Diversität über den gesamten Lauf (Weicker und Weicker 1999; Weicker 2000). Die Überprüfung dieses Arguments soll auch Gegenstand dieser Arbeit sein.

3.3 Adaptive und selbstadaptive Mechanismen

Für die Adaptation von EA in dynamischen Umgebungen bemerkt Angeline (1997), dass nach einer Änderung der Umwelt eine gefundene Einstellung eines Strategieparameters veraltet oder sogar nachteilig sein kann. Der Autor erwähnt, dass in Extremfällen der EA durch eine falsche Einstellung der Strategieparameter auch in eine völlig falsche Richtung gelenkt werden kann.

Theoretische Aussagen zur Adaptationsdauer der Mutationsschrittweiten leitet Beyer (1995, 1996) her. Er erkennt die Zahl der für die Anpassung benötigten Generationen umgekehrt proportional zu τ_0^2 (somit proportional zur Problemdimension n , was schon von Schwefel (1975) gezeigt wurde). Einen Vergleich zwischen einer multiplikativen und der von Fogel (1991) vorgeschlagenen additiven Schrittweitanpassung in einer einfachen dynamischen Umgebung führen Bäck und Schwefel (1996) durch. Beyer (1996) zeigt analytisch, dass beide Mechanismen für kleine Lernraten zu identischen Ergebnissen führen sollten, was auch die Resultate der von Bäck und Schwefel (1996) durchgeführten Experimente erklärt.

3.4 Speicherbasierte Verfahren

Die Hinzunahme speicherbasierter Verfahren ist insbesondere dann vorteilhaft, wenn das Optimum zwischen mehreren Zuständen wechselt. In diesem Fall kann eine bereits berechnete gute Lösung später bei Bedarf abgerufen werden.

Trojanowski und Michalewicz (1999a) erkennen mehrere Typen speicherbasierter Verfahren:

Numerischer Speicher: Bei diesem Typ werden zusätzliche numerische Parameter benutzt, die bspw. die Funktion der Variationsoperatoren beeinflussen. Dabei besteht eine enge Verwandtschaft zu adaptiven Strategien, denn auch dort beeinflussen Strategieparameter die Funktionsweise der Operatoren.

Symbolischer Speicher: Bei dieser Technik konstruiert der Algorithmus Regeln, um damit anschließend die Funktionsweise der Operatoren zu beeinflussen.

Exakter Speicher: Hier werden frühere gute Lösungen in einer geeigneten Form gespeichert. Neben der Möglichkeit, frühere Lösungen explizit in einer Datenbank zu speichern, gehört zu diesem Typ auch die Diploidie oder Polyploidie. Mattfeld und Bierwirth (1999) erwähnen aber die Schwierigkeit, einen wie von Goldberg und Smith (1987) vorgestellten diploiden EA in Umgebungen zu benutzen, in denen das Optimum zwischen mehr als zwei Zuständen oszilliert.

Für die Lösung dynamischer Probleme wurden in der Literatur bereits diverse Verfahren vorgestellt. Diese sind häufig auf das Auftreten einer bestimmten Form der Dynamik zugeschnitten und deshalb oftmals nur für eine kleine Auswahl elementar unterschiedlicher Problemstellungen einsetzbar. Alternativ könnten deshalb auch einige der obigen Ansätze kombiniert werden. So ist es möglich, nach einem Wechsel der Umgebung mit einer Kombination aus gespeicherten Individuen und Zufallsimmigranten zu arbeiten. Dabei klingt die Idee viel versprechend, nach unterschiedlichen Strategien arbeitende Populationen parallel zu entwickeln.

3.5 Bekannte theoretische Ergebnisse

Mittlerweile existiert bereits eine Vielzahl theoretischer Ergebnisse zu EA. Allerdings beschränken sich diese zumeist auf einfache Zielfunktionen. Zudem entzieht sich die Komplexität heutiger EA in dynamischen Umgebungen bisher einer detaillierten Untersuchung. Speziellere Analysen, wann welche Adaptationsformen zu bevorzugen sind oder wann auf ein adaptives Verhalten verzichtet werden soll, sind bisher eher selten. Verschiedene Analysen in dieser Richtung stammen von Rudolph. In (Rudolph 1999; Rudolph 2001) zeigt er, dass es unmöglich ist, mit einem die Mutationsschrittweite nach der 1/5-Erfolgsregel von Rechenberg adaptierenden EA eine Konvergenz gegen das globale Optimum mit Wahrscheinlichkeit Eins zu garantieren. Ausgehend von einem $(1 + 1)$ -EA überträgt er die Aussagen zunächst auf eine $(1 + \lambda)$ - und danach auf eine Multistartstrategie. Dass ein selbstadaptiver EA eine Konvergenz gegen das globale Optimum nicht sicher garantieren kann, resultiert nach seinen Analysen aus der Eigenschaft, dass im weiteren Verlauf des EA die Mutationsschrittweiten zu schnell klein werden, um ein lokales Optimum sicher verlassen zu können. Diese Aussagen haben selbst dann noch Bestand, wenn der EA unendlich lange läuft. Durch die in praktischen Anwendungen beschränkte Zeit wird diese Eigenschaft somit noch mehr zur Geltung kommen. Auch ein Übergang zu einer cauchyverteilten Änderung der Objektvariablen führt zu keiner Änderung der Aussage. Im Gegensatz dazu hat Rudolph (1997, S. 204) für bestimmte Zielfunktionen gezeigt, dass ein elitärer EA mit Wahrscheinlichkeit Eins gegen das globale Optimum konvergiert, wenn sichergestellt ist, dass die Mutationsschrittweite nicht unter ein bestimmtes Minimum fällt.

Hervorzuheben sind auch die von Rudolph (1999) in seinem Fazit gemachten Bemerkungen. Zum einen zeigen theoretische Analysen für nicht-selbstadaptive EA zwar, dass diese EA mit Wahrscheinlichkeit 1 gegen das globale Optimum konvergieren. Dabei benötigt ein solcher EA allerdings eine für praktische Anwendungen zu lange Zeit. Zum anderen ist die Konvergenzgeschwindigkeit in selbstadaptiven EA wesentlich größer. Zusätzlich wächst in Multistartstrategien die Wahrscheinlichkeit, das globale Optimum zu finden, exponentiell mit der Anzahl der Läufe.

Darüber hinaus beweist Rudolph (2001), dass es zur Erzielung einer hohen Konvergenzgeschwindigkeit unter gleichzeitiger Bewahrung der Möglichkeit, ein lokales Optimum zu verlassen, ausreichend ist, eine übliche $(1 + \lambda)$ -ES in Kombination mit einem zusätzlichen Nachkommen unter Verwendung einer festen Cauchyverteilung zu benutzen. Ohne einen vollständigen Beweis anzugeben glaubt er, seine Ergebnisse auf einen echt selbstadaptiven EA übertragen zu können.

Theoretische Abhängigkeiten zwischen der Wahl des Lernparameters τ und der Anzahl der Generationen bis zur Einstellung einer optimalen Mutationsschrittweite hat Beyer (1996) für das Kugelmodell erarbeitet. In seinen Untersuchungen ergibt sich, dass die Anzahl der bis zum Erreichen eines optimalen Strategieparameters σ benötigten Generationen umgekehrt proportional zur Wahl von τ sind. Dies heißt umgekehrt, dass große Werte für τ vermutlich zu einer zu schnellen Abnahme von σ im Falle multimodaler Funktionen führen werden. Für praktische Probleme großer Dimension

($n \geq 1000$) empfiehlt er τ zuerst groß und später optimal zu wählen.

Bekanntlich gibt es bereits fundierte theoretische Aussagen zur optimalen Einstellung bestimmter Strategieparameter insbesondere beim Kugelmodell. Wo theoretische Analysen nicht möglich sind, werden experimentelle Untersuchungen zur Erarbeitung allgemeiner Aussagen durchgeführt. Wie bspw. Kursawe (1999, S. 59-66) zeigt, besitzt die Einstellung der so genannten Lernraten τ_0, τ_i auch auf anderen Funktionen erhebliche Auswirkungen auf das Konvergenzverhalten. Aus experimentellen Untersuchungen am skalierten Kugelmodell resultiert die Empfehlung von Kursawe (1999), die beiden Parameter größer als üblich zu wählen. Da ein Anwender i.d.R. aber keine genauen Kenntnisse über die zu optimierende Funktion besitzt, wird er auf eine allgemeine Einstellung der Strategieparameter zurückgreifen.

Die Einbeziehung der Rotationswinkel in die Repräsentation der Individuen führt dazu, dass eine Adaptation des EA an die Problemstellung erheblich länger dauert. Zudem ist eine wesentlich größere Anzahl für μ und λ notwendig als bei einem Verzicht auf die Rotationswinkel. Besitzt die Zielfunktion in einer Umgebung eines Punktes x_0 eine positive Hessematrix, so hat Rudolph (1992) gezeigt, dass zur Berechnung eines Kleinste-Quadrate-Schätzers der Hessematrix eine Anzahl von $\frac{1}{2}(n^2 + 3n + 4)$ Funktionsaufrufen nicht unterschritten werden darf. Soll die Hessematrix in jeder Generation berechnet werden, muss die Summe aus Eltern und Nachkommen mit dieser Anzahl übereinstimmen. Bei einer identischen Gesamtzahl von Funktionsauswertungen führt dies bei einem 30-dimensionalen Problem gegenüber der üblichen Wahl von $\mu = 15, \lambda = 100$ zu einer Reduzierung der Generationen auf ungefähr ein Viertel. In jedem Einzelfall ist dann abzuwägen, ob es durch die Hinzunahme der Rotationswinkel möglich ist, mit einer solch reduzierten Generationenzahl ähnlich gute oder sogar bessere Ergebnisse gegenüber der Standardvariante zu erzielen. Die von Kursawe (1999, S. 43) angegebene Begründung zur Vernachlässigung korrelierter Mutationen im Fall $n_\sigma = n$ aufgrund der Erhöhung notwendiger Generationen und Funktionsauswertungen ist deshalb schlüssig und führt in der vorliegenden Arbeit ebenfalls dazu, diesen ES-Typ auszuschließen.

Daneben ist auch der Einfluss der Rekombination noch nicht hinreichend untersucht. Theoretische Analysen wie die von Beyer (1995) beziehen sich zumeist auf die Multirekombination, bei der der Rekombinationsoperator den Durchschnitt aller Eltern berechnet. Dieses Vorgehen vereinfacht die Analysen enorm, stellt jedoch nur eine der möglichen Varianten dar. Zusätzlich sollen die experimentellen Erkenntnisse von Rechenberg (1994, S. 146) nicht unerwähnt bleiben, dass die Rekombination aller Eltern erst für $n = 20.000$ beim Kugelmodell die theoretisch berechneten Erfolge zeigte.

Die auch in dieser Arbeit betrachtete Fluktuation der Strategieparameter ist bekannt und wurde bereits von Beyer und Schwefel (2002) beschrieben. Zur Vermeidung dieses Verhaltens empfehlen sie einerseits die intermediäre (μ/μ)-Rekombination der Strategieparameter und andererseits die Methode der „rescaled mutations“ (siehe auch (Rechenberg 1994) und (Beyer 2000)).

Betrachten wir die einschlägige Literatur so können wir feststellen, dass die Konzentration der meisten Autoren im Bereich der ES auf der Untersuchung neuer Operatoren liegt. Dabei betrachten viele Untersuchungen zwar eine Vielzahl an Testfunktionen mit teils abweichender Problemdimension, allerdings benutzen sie immer die gleichen Standardeinstellungen. Insbesondere werden für μ und λ häufig die Standardwerte $\mu = 15$ und $\lambda = 100$ verwendet. Diese Festlegung erfolgt meist ohne Begründung.

Die Überlegungen zur Wahl von μ und λ führen automatisch zu der Frage, in welchen Situationen eine echte Population ($\mu > 1$) überhaupt notwendig ist. Wird zum Vergleich zweier Verfahren die erreichte Güte nach einer bestimmten Anzahl Funktionsauswertungen betrachtet, so ist es durchaus

möglich, dass ein $(1+1)$ -EA auf bestimmten Zielfunktionen bessere Ergebnisse als eine $(15, 100)$ -ES liefert. Als wesentlicher Vorteil populationsbasierter Verfahren wird immer wieder auf den positiven Effekt der Rekombination verwiesen. Die Theorie zu dieser Problematik ist aber ebenfalls noch nicht ausgeprägt. So leitet Beyer (2001, S. 224) bei der Begründung des Effekts der genetischen Reparatur Formeln zur Berechnung der optimalen Anzahl Eltern bei Benutzung der Multirekombination mit $\rho = \mu$ her. Allerdings hängt μ sowohl von der zuvor festgelegten Anzahl Nachkommen als auch einer optimal eingestellten Schrittweite ab, die in praktischen Anwendungen jedoch in den seltensten Fällen bekannt sein dürfte. Weitere Ansätze stammen u.a. von Jansen und Wegener (2001), der eine Funktion vorstellt, für die allein schon der Einsatz einer echten Population (d.h. ohne Verwendung der Rekombination) vorteilhaft ist.

Bisherige Untersuchungen zur optimalen Wahl der Populationsgrößenparameter beziehen sich zumeist auf spezielle Problemstellungen der klassischen GA wie das Traveling Salesperson Problem (Julstrom 1996) oder die Auftragsbearbeitung (Nakano, Davidor und Yamada 1994). Von einer speziellen Problemstellung unabhängige Untersuchungen sind dagegen auch für GA selten, auch wenn einige Autoren solche Untersuchungen durchgeführt haben (De Jong 1975; Goldberg 1985; Goldberg 1989b; Reeves 1993). Obwohl Smith (1997) behauptet, die für die klassischen GA gewonnenen Erkenntnisse auch auf andere Varianten der EA anwenden zu können, sind diese häufig auf dem Schematheorem¹ basierenden Analysen nur schwer auf die reellwertige Optimierung mit ES übertragbar. Weitere Untersuchungen die Wahl der Populationsgrößen betreffend finden sich bspw. bei Jansen und Wegener (2001) sowie Witt (2004). Gemein ist diesen Untersuchungen aber, dass meist nur diskrete oder kombinatorische Problemstellungen betrachtet werden. Dagegen leiten Jägersküpper und Witt (2005) Laufzeitberechnungen für eine nach der 1/5-Erfolgsregel von Rechenberg arbeitende $(\mu + 1)$ -ES auf dem Kugelmodell her.

Eine eindrucksvolle Unterstützung für die Erfordernis einer etwas sensibleren Einstellung für μ und λ liefert Kursawe (1999). Er stellt für das Kugelmodell fest, dass die Standardparametrisierung mit $\mu = 15$ und $\lambda = 100$ im 100-dimensionalen Kugelmodell divergiert. Die von ihm speziell betrachteten Meta-ES (Kursawe 1999, S.69-86) zeigen dagegen, dass mit anderen μ - und λ -Einstellungen dort sehr wohl eine Konvergenz gegen das Optimum zu verzeichnen ist. Obwohl in diesen Meta-ES auch Anpassungen anderer Parameter vorgenommen werden, deuten die Ergebnisse doch die Abhängigkeit zwischen der Problemdimension n und einer sinnvollen Wahl von μ und λ an.

Somit steht die Untersuchung der optimalen Wahl der Parameter μ und λ insbesondere bei Verwendung der üblicherweise als vorteilhaft angesehenen Rekombination noch aus. Deshalb soll in dieser Arbeit eine genauere Untersuchung des Zusammenhangs zwischen den Parametern μ und λ und dem Konvergenzverhalten von ES bei verschiedenen Problemdimensionen im Vordergrund stehen. Die durchgeführten Experimente sollen einerseits zu grundsätzlichen Erkenntnissen führen, andererseits Anhaltspunkte zur Wahl dieser Parameter bei verschiedenen Problemstellungen liefern.

Wie bereits in vorhergehenden Abschnitten beschrieben, sind theoretische Analysen in dynamischen Umgebungen bisher erst sehr selten. Eine aktuelle Veröffentlichung stammt von Jansen und Schellbach (2005). Im Blickpunkt ihres Interesses steht ein $(1 + \lambda)$ -EA, der auf eine diskrete Problemstellung angewendet wird. Ohnehin beziehen sich existierende Arbeiten häufig auf den diskreten Fall. Mit zunehmender Verbesserung der Möglichkeiten zur Durchführung theoretischer Analysen wird aber auch der Bereich der kontinuierlichen Dynamik stärker in den Mittelpunkt des Interesses gelangen.

¹Beyer (2001, S. 4) erwähnt, dass in den letzten Jahren die Anhänger der GA zum Schematheorem alternative Ansätze verfolgen.

3.6 Fazit

In diesem Kapitel wurden die gängigsten spezialisierten Verfahren zur Optimierung dynamischer Funktionen vorgestellt. Die Diversitätserhaltung ist in vielen dieser Algorithmen stark ausgeprägt, weshalb wir diesen Aspekt im Kapitel 9 noch einmal aufgreifen werden. Spezialisierte Verfahren besitzen jedoch oftmals weitere Freiheitsgrade zur Auswahl zusätzlicher Strategieparameter. Da deren Einfluss nicht immer genau bekannt ist, wird oftmals auf Standard-Strategien zurückgegriffen. Diese besitzen zudem den Vorteil, dass sie in statischen Umgebungen ihre Berechtigung bewiesen haben.

Die Übertragung zur Optimierung von gestörten Zielfunktionen entwickelter Verfahren ist überlegenswert. Allerdings muss dabei berücksichtigt werden, dass trotz Ähnlichkeiten im Fitnessraum die Unterschiede im Suchraum erheblich sind. In den entsprechenden Experimenten wurde deutlich gezeigt, dass die Populationen der Evolutionsstrategien einem sich bewegenden Optimum meistens folgen, d.h. mit einem gewissen Abstand hinterherlaufen.

Kapitel 4

Möglichkeiten des Vergleichs der Ergebnisse verschiedener Strategien

Solange ausreichende fundierte theoretische Erkenntnisse zum Verhalten EA ausstehen, bleiben empirische Untersuchungen notwendig. Trotz dieser Tatsache fand bis vor kurzem die Theorie empirischer Untersuchungen evolutionärer Algorithmen wenig Beachtung. Mittlerweile haben sich verschiedene Autoren dieser Thematik angenommen. Obwohl die Veröffentlichung des Artikels von Hooker (1994) schon einige Jahre zurück liegt, ist der Inhalt noch immer aktuell. Hooker begründet die Notwendigkeit empirischer Untersuchungen besonders mit der beschränkten Aussagekraft bisheriger theoretischer Analysen. Zwar gibt es theoretische Analysen, oftmals besitzen solche Resultate jedoch lediglich einen asymptotischen oder worst-case Charakter. Zudem beruhen sie häufig auf unrealistischen Annahmen sowie auf durch die Praxis nicht gestützten Wahrscheinlichkeitsverteilungen der Problemstellungen. Darüber hinaus werden meist nur für die einfachsten Algorithmen theoretische Ergebnisse erzielt.

Neben den ohnehin schon komplexen Standard-Algorithmen sind die für praktische Problemstellungen spezialisierten Verfahren einer theoretischen Analyse oftmals nicht zugänglich. Allgemein stellen theoretische Analysen zwar einen wichtigen Schritt zu einem besseren Verständnis der Funktionsweise von EA dar, allerdings bleibt das Verhalten eines realen, allgemein anwendbaren EA noch mit zu vielen Unbekannten versehen. Eine auf theoretische Untersuchungen gestützte akzeptable Wahl geeigneter Parameterwerte ist häufig nicht möglich, weshalb experimentelle Untersuchungen weiterhin unerlässlich sind.

In einem späteren, nicht speziell auf evolutionäre Algorithmen bezogenen Artikel hat Hooker (1995) seine Ausführungen durch weitere Hinweise ergänzt. Zunächst bemängelt er die gängige Praxis vieler Wissenschaftler, lediglich verfeinerte und schnellere Algorithmen zu entwickeln, ohne nach den Gründen für das bessere Verhalten zu forschen. Er schlägt vor, alle Einzeleffekte durch entsprechende Experimente zu isolieren und das Testen EA so zu einer Wissenschaft auszubauen.

Nicht verschwiegen werden sollen aber auch Nachteile empirischer Untersuchungen. Dazu zählt Hooker insbesondere die Abhängigkeit der gewonnenen Resultate von der gewählten Rechnerarchitektur und der Implementierung des Algorithmus. Daneben bringt er noch einen weiteren interessanten Aspekt ein, indem er bemerkt, dass negative Resultate in der Informatik allgemein nicht anerkannt würden. Dies ist seiner Ansicht nach sehr zu bedauern, können sie doch Anhaltspunkte geben, wann ein Algorithmus nicht funktioniert, und damit andere Anwender vor ähnlichen Fehlern bewahren. Dies schließt auch die Angabe von Problemklassen ein, bei denen ein sonst gut funktionierender Algorithmus weniger erfolgreich ist.

Die von ihm und anderen (Barr et al. 1995) erwähnten unterschiedlichen Ausführungszeiten durch besondere Programmiertechniken sind in EA ausgeschlossen, da Algorithmenvergleiche üblicherweise aufgrund der gleichen Anzahl Funktionsauswertungen erfolgen. Bzgl. der von Gent et al. (1997) angesprochenen Beobachtung zu den Zufallszahlengeneratoren ist zu sagen, dass Gronek (1997) empirisch ermittelt hat, dass die verwendeten Pseudo-Zufallszahlengeneratoren in den EA einen zu vernachlässigenden Effekt haben, sofern sie gewissen, leicht zu berücksichtigenden Anforderungen genügen. Die von Hooker (1994) gemachten Vorbehalte gegen Benchmarksuiten sind dagegen weit verbreitet und nicht vollkommen zu widerlegen.

Um empirische Ergebnisse für ein breites Publikum nutzbar zu machen, ist ein strukturiertes Vorgehen notwendig. Dazu muss zunächst überlegt werden, welche Aussagen über den EA getroffen werden sollen. Anschließend ist die genaue Vorgehensweise zur Durchführung der Experimente festzulegen. Dies beginnt zunächst mit der angemessenen Wahl der Testfunktionen. Auf die Darstellung dieses komplexen Teilgebiets soll an dieser Stelle verzichtet werden. Stattdessen verweisen wir auf den Abschnitt 5.1, der sich diesem Themenkomplex widmet, sowie die dort verzeichneten Literaturstellen.

Nach Auswahl der Testfunktionen ist zu entscheiden, wieviele Testläufe mit jedem Einzelalgorithmus durchgeführt werden müssen, um aussagekräftige Ergebnisse zu erhalten. Hierzu zählt die Notwendigkeit, zur Beurteilung der Signifikanz unterschiedlicher Ergebnisse statistische Tests durchzuführen.

Problematisch sind bei allen experimentellen Untersuchungen die Einstellungen der verwendeten Strategieparameter und Variationsoperatoren der Verfahren. So wird des Öfteren die Art der Rekombination nicht genauer spezifiziert. Ist bspw. bei der Rekombination das Cloning erlaubt, bei dem zwei identische Eltern rekombiniert werden? Findet bei der Verwendung von n Strategieparametern für die Mutationsschrittweiten eine gekoppelte Rekombination statt (also identische Eltern für die Rekombination der Objektvariablen und der Strategieparameter)? Abschließend geklärt werden muss noch, ob für Strategieparameter übliche Werte benutzt wurden. Darüber hinaus bleibt es manchmal unklar, mit welcher Problemdimension n die Experimente durchgeführt wurden.

Insgesamt ist diese Thematik in den vergangenen Jahren zu Recht stärker in den Blickpunkt des allgemeinen Interesses geraten. Zur Motivierung der Darstellungsformen im Teil II beschränken wir uns in den nächsten Abschnitten auf eine kurze Zusammenfassung allgemein gebräuchlicher Methoden wie sie von der Statistik her bekannt sind.

4.1 Darstellung der Einzelergebnisse

Zum Vergleich zweier unterschiedlicher Verfahren müssen prinzipiell die jeweils üblichen oder aus heuristischen Überlegungen heraus besten Einstellungen der verwendeten Verfahren benutzt werden. Eine absichtliche Wahl einer guten Parametereinstellung bei einem der Verfahren und einer identischen (aber erfahrungsgemäß schlechten) Einstellung bei dem zweiten Verfahren ist unfair.

Bei der Darstellung der Ergebnisse unterschiedlicher Verfahren ist eine Vergleichbarkeit zu gewährleisten. Für den Einsatz EA ist darauf zu achten, dass für jede Generation eine identische Anzahl Funktionsauswertungen benutzt wird. Ist die Anzahl der Funktionsauswertungen pro Generation bei den Verfahren unterschiedlich, ist hierauf besonders hinzuweisen. In diesen Fällen bietet sich eine Darstellung der Fitness in Abhängigkeit der durchgeführten Funktionsauswertungen an. Dies steht in engem Zusammenhang mit der Möglichkeit, bei zwei Verfahren jeweils nur den besten von m Durchläufen zu vergleichen. Sollen die Ergebnisse einer Einzelstrategie mit denen

einer Multistartstrategie verglichen werden, bietet sich ein Vergleich der Ergebnisse eines einzelnen Durchlaufs einer (15, 100)-ES nach 1000 Generationen mit dem besten Ergebnis aus 10 Durchläufen eines (3, 20)-EA nach 500 Generationen an.

Jedes Vorgehen hat seine Vor- und Nachteile. In jedem Einzelfall ist das gewählte Vorgehen zu begründen. Hinweise auf mögliche abweichende Ergebnisse bei einem anderen Vorgehen sind ratsam.

Im weiteren Verlauf dieses Abschnitts stellen wir verschiedene Verfahren zur Beurteilung einzelner Algorithmen vor. Der Abschnitt 4.2 zeigt dagegen, wie die Ergebnisse mehrerer Verfahren miteinander verglichen werden können.

4.1.1 Numerische Darstellungen

Lage- und Streuungsmaße

Für kompakte Darstellungen der erreichten Fitnesswerte bieten sich Lage- und Streuungsmaße an. Hierbei handelt es sich meist um das arithmetische Mittel in Verbindung mit der Standardabweichung. Diese sind theoretisch untermauert und lassen sich äußerst leicht und effizient berechnen.

Ein für alle EA benutzbares Maß ist die bei Eiben und Jelasity (2002) erwähnte mittlere beste Fitness (**MBF**). Für jeden Durchlauf wird nach Beendigung der beste Fitnesswert gespeichert. Das arithmetische Mittel dieser Werte von allen Durchläufen bildet dann die mittlere beste Fitness.

Ein üblicher Fehler liegt jedoch in der „unreflektierten“ Verwendung des Mittelwertes. Dieser führt bei stark unterschiedlichen Stichprobenwerten zu einem sehr labilen Maß. D.h. insbesondere, dass einzelne starke Ausreißer den so gemessenen Wert dominieren können. Die zusätzlich angegebene Standardabweichung deutet in solchen Fällen vielleicht auf Unregelmäßigkeiten oder Besonderheiten hin, reicht aber nicht immer aus. Dann werden zusätzliche Angaben wie höhere Momente und Histogramme benötigt. Dies trifft umso mehr auf transformierte Daten zu, bei denen die Mittelwertbildung der transformierten Daten häufig eine starke Verzerrung der zugrunde liegenden Daten ergeben kann. In diesem Fall ist eine Rücktransformation der Daten notwendig, für die sich die so genannten Box-Cox-Transformationen anbieten (Schlittgen 2001, S. 27).

Eine mehrfache Ausführung eines EA auf einer bestimmten Funktion führt meist zu einer gewissen Streuung der geordneten Fitnesswerte. Der Mittelwert und die Standardabweichung betragen für mehrere Durchläufe eines fiktiven (15, 100)-EA auf dem Kugelmodell $\bar{x} = 1.20 \cdot 10^{-18}$ und $\sigma_x = 4.94 \cdot 10^{-18}$. Mittelt man über den Logarithmus (zur Basis 10) der Daten, so ergibt sich der Mittelwert $\overline{x_{\log}} = -18.88$. Die Rücktransformation durch Berechnung der 10er-Potenz führt nun zum Mittelwert $10^{\overline{x_{\log}}} = 1.33 \cdot 10^{-19}$.

Der Mittelwert ist anfällig gegenüber einzelnen extremen Ausreißern. Ein erster Ausweg zur Vermeidung dadurch bedingter zu starker Verzerrungen ist die Streichung von Daten. Oftmals werden vor der Berechnung des Mittelwertes die jeweils 5% besten und schlechtesten Werte aus einer Stichprobe eliminiert. Dieses pauschale Vorgehen findet jedoch nicht überall Akzeptanz, weshalb häufig der Übergang zu einem anderen Lagemaß erfolgt.

Der Median \tilde{x} erweist sich als robuster gegenüber Ausreißern. Er ist definiert als ein Spezialfall des α -Quantils ($0 \leq \alpha \leq 1$)

$$q_{x,\alpha} = (1 - f) \cdot x_{(i)} + f \cdot x_{(i+1)} \quad ,$$

$$\text{wobei } i = \lfloor r \rfloor, \quad f = r - i, \quad r = 1 + (n - 1) \cdot \alpha \quad \text{und} \quad x_{(n+1)} := x_{(n)} \quad ,$$

mit $\alpha = 0.5$. Das α -Quantil $q_{x,\alpha}$ interpoliert den Wert der geordneten Stichprobe für den $100\alpha\%$ der Daten kleiner oder gleich und $100(1 - \alpha)\%$ der Daten größer oder gleich diesem Wert sind. Für

das obige Beispiel ergibt sich für den Median der Wert $\tilde{x} = 1.18 \cdot 10^{-19}$. Da der Logarithmus eine monotone Funktion ist, führt die Bildung des Logarithmus mit anschließender Rücktransformation zum gleichen Resultat. Der Median stimmt wesentlich besser mit dem über die Transformation berechneten Mittelwert als mit dem Mittelwert der Originaldaten überein.

Mithilfe der α -Quantile kann darüber hinaus ein Streuungsmaß angegeben werden. Der Quartilsabstand ist als Spezialfall des α -Quantilsabstands

$$d_{x,\alpha} = q_{x,1-\alpha} - q_{x,\alpha}, \quad \alpha \in]0, 0.5[$$

mit $\alpha = 0.25$ definiert:

$$d_{x,0.25} = q_{x,0.75} - q_{x,0.25} \quad .$$

Der Quartilsabstand zeigt demnach die Spannbreite der Werte der in der Mitte zentrierten Hälfte der geordneten Stichprobe an und ist wie der Median unempfindlich gegenüber einzelnen Ausreißern. Für das Beispiel ergibt sich der Quartilsabstand zu $d_{x,0.25} = 5.99 \cdot 10^{-19} - 3.47 \cdot 10^{-20} = 5.64 \cdot 10^{-19}$. Zu beachten ist allerdings, dass der Quartilsabstand im vorliegenden Fall nur wenig Aussagekraft besitzt. Sinnvoller ist hier der Abstand $d_{x_{\log},0.25} = 1.2$ der logarithmierten Daten. Die Daten liegen also ca. 1,2 Zehnerpotenzen auseinander.

Konfidenzintervalle

Bei genügend großer Stichprobe lassen sich mithilfe der Quantile auch weitere, über den mittleren zu erwartenden Fitnesswert hinaus gehende Angaben ableiten. Hierzu gehören (approximative) Konfidenzintervalle für den erreichbaren Fitnesswert. Ein Konfidenzintervall bezeichnet zwei Grenzen, innerhalb derer ein Wert mit einer gewissen vorgegebenen Wahrscheinlichkeit liegt.

Ein Konfidenzintervall (KI) kann dabei einerseits für ein einzelnes bestimmtes Lagemaß angegeben werden. Bspw. soll ein Konfidenzintervall bestimmt werden, das den Mittelwert oder Median der Grundgesamtheit mit einer gewissen Wahrscheinlichkeit enthält. Je größer die Stichprobe (d.h. die Anzahl der durchgeführten unabhängigen Läufe), desto kleiner (bzw. schmaler) wird das zugehörige Konfidenzintervall, wodurch sich der jeweilige Wert genauer bestimmen lässt.

Andererseits interessiert man sich für ein Intervall, in dem der Fitnesswert eines einzelnen Durchlaufs mit einer gewissen Wahrscheinlichkeit liegt. Hierbei führt die Vergrößerung des Stichprobenumfangs im Gegensatz zum ersten Fall i.d.R. nicht zu einer Verkleinerung des entsprechenden Intervalls. Stattdessen werden die Grenzen des echten Intervalls mit zunehmendem Stichprobenumfang immer genauer getroffen, wobei sich das Intervall durch Hinzunahme eines weiteren Wertes jedoch auch vergrößern kann.

Als ein Beispiel eines solchen Intervalls haben wir im vorigen Abschnitt bereits den Quartilsabstand kennen gelernt. Analog berechnet man das 0.9-Quantil für den ungefähren Fitnesswert, der mit einer 90%igen Wahrscheinlichkeit mindestens erreicht wird. Möchte man dagegen wissen, innerhalb welcher Werte der erreichte Fitnesswert in 90% aller Fälle liegt, so berechnet man die 0.05- und 0.95-Quantile als Grenzen des Intervalls. Für die Daten des obigen Beispiels ergeben sich die Grenzen eines solchen Konfidenzintervalls zu $q_{x,0.05} = 4.04 \cdot 10^{-21}$ und $q_{x,0.95} = 3.55 \cdot 10^{-18}$.

Neben einem Konfidenzintervall für den Fitnesswert nach Ablauf des Algorithmus ist es auch denkbar, ein solches KI für den Fitnesswert jeder Generation anzugeben. Dies führt dann zu dem Begriff des Konfidenzstreifens. Der Konfidenzstreifen wird dann für jede Generation aus den Fitnesswerten dieser Generation gebildet. Ein zentrierter 90%-Konfidenzstreifen enthält dann die Fitnesswerte jeder Generation, die zwischen dem 0.05- und dem 0.95-Quantil liegen.

4.1.2 Grafische Darstellungen

Einen anschaulicheren Eindruck der Verhältnisse gestatten oft grafische Darstellungsarten, welche die Daten der gesamten Stichprobe widerspiegeln. Hierbei handelt es sich üblicherweise um Darstellungen, die auf der empirischen Verteilung der ermittelten Fitnesswerte verschiedener Läufe basieren.

Fitnessverlauf

Natürlich kann zunächst der Fitnessverlauf eines einzelnen EA-Durchlaufs auf dem betrachteten Problem dargestellt werden. Für das Kugelmodell mit einer (15, 100)-ES ergibt sich ein ähnlicher Verlauf wie er in der Abbildung 4.1 dargestellt ist. Da sich beim Kugelmodell die erreichten

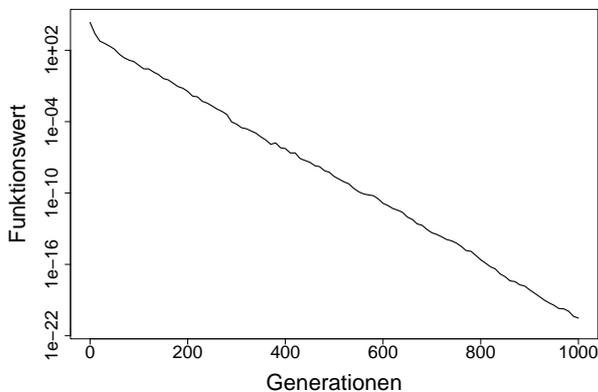


Abbildung 4.1: Einzelllauf.

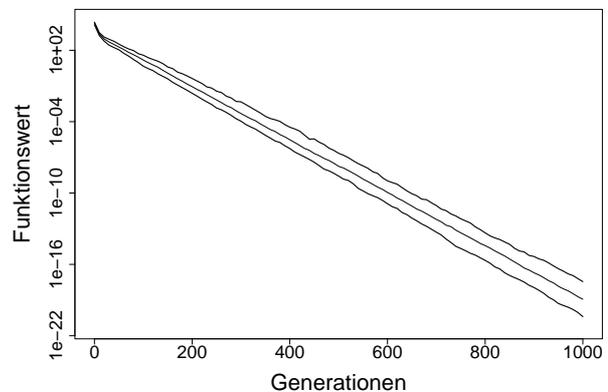


Abbildung 4.2: Median und Konf.-streifen.

Abbildungen 4.1 und 4.2: Einzelllauf sowie zentrierter 90%-Konfidenzstreifen und Median von 101 Durchläufen einer (15, 100)-ES auf dem 30-dimensionalen Kugelmodell.

Funktionswerte immer mehr der Null nähern, ist zur Beobachtung der Konvergenzordnung eine Darstellung der Funktionswerte mit einer logarithmischen Skala angebracht.

Entfernt man sich vom Einzelllauf, stehen verschiedene Möglichkeiten offen. Bei der einen werden in der Grafik beispielhafte Fitnessverläufe dargestellt. Abbildung 4.2 zeigt ein Beispiel für den mittleren (Median) Verlauf einer (15, 100)-ES auf dem Kugelmodell zusammen mit dem zentrierten 90%-Konfidenzstreifen der Fitnesswerte. Zentriert heißt hier, dass der Konfidenzstreifen die Fitnesswerte jeder Generation umschließt, die zwischen dem 0.05- und 0.95-Quantil liegen. Somit erhält man einen schnellen Eindruck von repräsentativen Einzelläufen. Dabei wird aber außer Acht gelassen, dass die Einzelläufe im Normalfall größere Schwankungen aufweisen und sich zeitweise immer wieder außerhalb beider 90%-Grenzen bewegen können. Möglich ist auch, dass sich ein kompletter Einzellauf oberhalb oder unterhalb dieser Grenzen bewegt.

Histogramme

Insbesondere im statischen Fall interessiert i.d.R. nicht der genaue Fitnessverlauf eines einzelnen Durchlaufs. Wichtiger ist stattdessen der nach einer vorgegebenen Anzahl von Funktionsauswertungen erreichte Fitnesswert. Dieser Wert fungiert dann auch als das entscheidende Vergleichsmaß

zweier Algorithmen. Auch in diesem Fall wird man nicht vom Ergebnis nur eines einzelnen Durchlaufs auf die Qualität des Verfahrens schließen.

Eine Möglichkeit der grafischen Darstellung der erreichten Fitnesswerte bilden Histogramme. Dafür wird das Intervall der erreichbaren Fitnesswerte in einzelne Klassen unterteilt und beobachtet, wieviele Fitnesswerte innerhalb jeder Klasse liegen. Wird der Fitnessverlauf mit einer logarithmischen Skala gezeichnet, ist auch bei der Darstellung mittels Histogramm eine Verwendung der logarithmierten Fitnesswerte empfehlenswert (siehe Abb. 4.3 und 4.4).

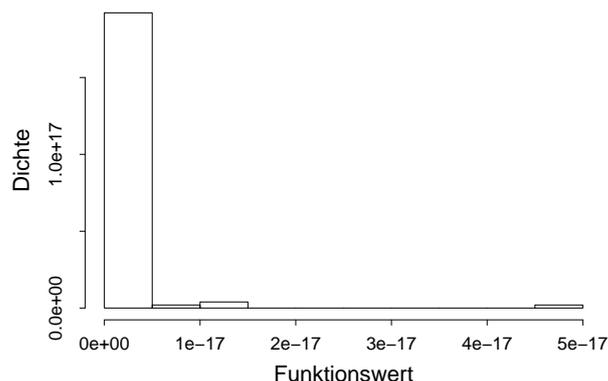


Abbildung 4.3: Originaldaten.

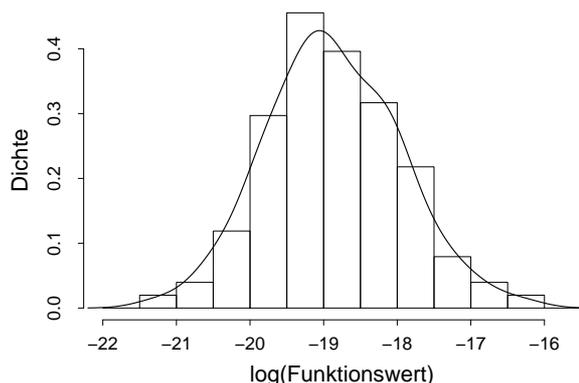


Abbildung 4.4: Logarithmierte Daten.

Abbildungen 4.3 und 4.4: Histogramm der erreichten Fitnesswerte auf dem 30-dimensionalen Kugelmodell. Abb. 4.4 enthält zusätzlich eine mit einem Gauß-Kernschätzer berechnete approximative Dichte.

Bei einer nicht zu kleinen Stichprobe ist auch der Übergang zur Dichtefunktion möglich. Dabei wird das Histogramm mit einer Kerndichteschätzung (Bünig und Trenkler 1994, S. 160) überlagert und dadurch ein stetiger Verlauf der Verteilung erzielt. Abhängig von dem verwendeten Kernschätzer ergibt sich für das obige Beispiel eine wie in der Abbildung 4.4 dargestellte ähnliche Dichte. In dieser Abbildung ist nun deutlicher die einer Normalverteilung ähnliche Verteilung der logarithmierten Fitnesswerte zu erkennen.

Boxplot

Eine sehr kompakte Darstellung der wesentlichen deskriptiven Daten bietet ein Boxplot (siehe Abb. 4.5 und 4.6). Er stellt in prägnanter Form den Median (horizontale Linie innerhalb des Rechtecks), das erste und dritte Quartil (als oberer und unterer Rand des Rechtecks), ein Streuungsmaß (äußere Querstriche) sowie potenzielle Ausreißer (einzelne Kreise) dar. Die Grenzen des Streuungsmaßes ergeben sich hier aus den Datenpunkten, die nicht mehr als den 1,5-fachen Quartilsabstand von der Box entfernt liegen. Alle außerhalb befindlichen Datenpunkte werden als Ausreißer betrachtet. Die Wahl des Faktors zur Berechnung des Streuungsmaßes ist relativ willkürlich und dient lediglich als Anhaltspunkt zur Erkennung von Ausreißern.

Zur Abschätzung der Unsicherheit bzgl. der Lage des Medians zeigt der hier dargestellte Boxplot mittels Einkerbungen zusätzlich ein 95%iges Konfidenzintervall für den Median an. Die in der Abbildung 4.5 zu sehende Box besitzt eine besonders auffällige Gestalt. Das untere Quartil liegt oberhalb der für das 95%ige Konfidenzintervall berechneten unteren Grenze. Ein solches Verhal-

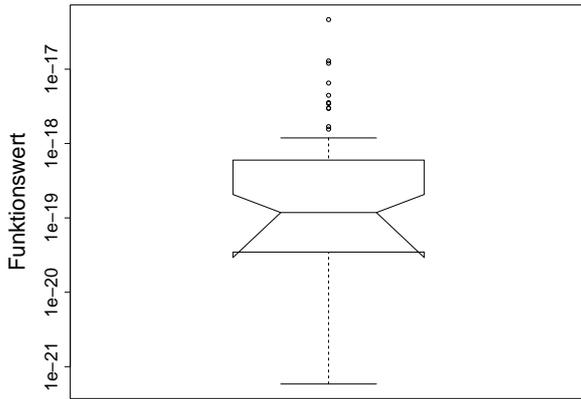


Abbildung 4.5: Originaldaten.

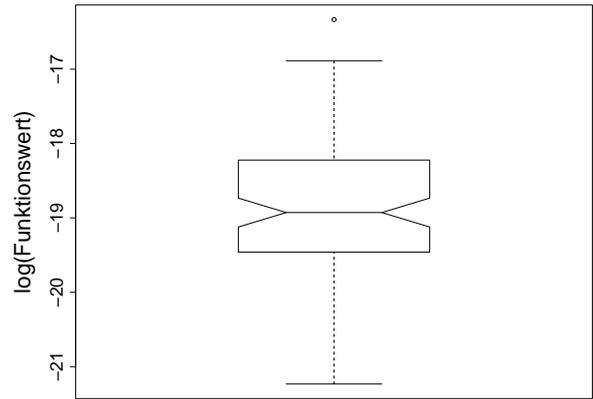


Abbildung 4.6: Logarithmierte Daten.

Abbildungen 4.5 und 4.6: Boxplot der Beispieldaten.

ten tritt üblicherweise nicht auf und verschwindet bei der Verwendung der logarithmierten Daten (Abb. 4.6).

4.2 Vergleich mehrerer Verfahren

4.2.1 Grafische Darstellung

Beim Entwurf neuer Algorithmen ist man daran interessiert festzustellen, wie sie sich im Vergleich zu bereits bestehenden Verfahren verhalten. Erste Anhaltspunkte lassen sich bereits durch den Vergleich der obigen Einzelergebnisse gewinnen. Deutlicher werden die Aussagen jedoch, wenn die Ergebnisse zweier Verfahren in einer gemeinsamen Grafik dargestellt werden. Überschneiden sich in einem Boxplot die Einkerbungen der Mediane nicht, so unterscheiden sich zumindest die Mediane der beiden Verfahren signifikant voneinander.

Für das folgende Beispiel seien einige wichtige statistische Werte der Ergebnisse eines fiktiven (14, 95)-EA wie folgt:

$$\bar{y} = 3.14 \cdot 10^{-19}, \sigma = 8.26 \cdot 10^{-19} ,$$

$$\tilde{y} = 7.49 \cdot 10^{-20}, d_{y,0.25} = 2.50 \cdot 10^{-19} .$$

Für die logarithmierten Daten gilt:

$$\bar{y}_{\log} = -19.22 \text{ und } 10^{\bar{y}_{\log}} = 6.08 \cdot 10^{-20} .$$

Auch hier beträgt der Quartilsabstand ungefähr $d_{y_{\log},0.25} = 1.2$.

Die gemeinsame Darstellung der beiden Boxplots ist in den Abbildungen 4.7 und 4.8 dargestellt. Während für die Ausgangsdaten kein signifikanter Unterschied des Medians festgestellt werden kann, ist der Unterschied der logarithmierten Daten signifikant zum 5%-Niveau. Dies ergibt sich aus der Abb. 4.8, da sich dort die Einkerbungen zwar überschneiden, sich allerdings keiner der beiden Mediane im eingekerbten Bereich des jeweils anderen Boxplots befindet.

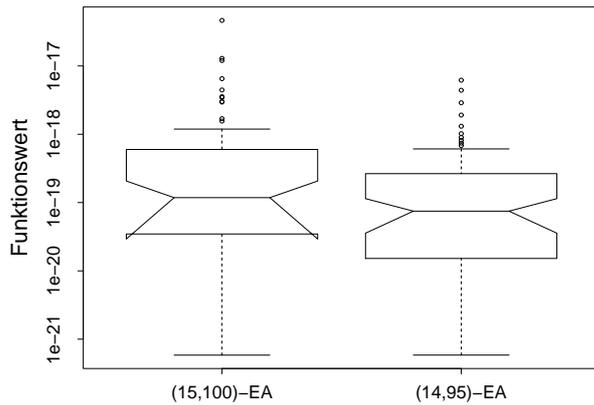


Abbildung 4.7: Originaldaten.

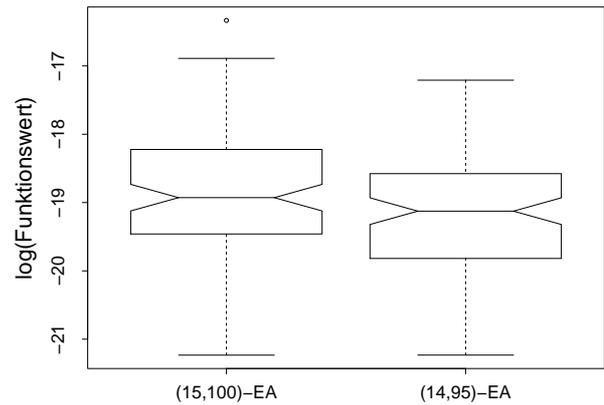


Abbildung 4.8: Logarithmierte Daten.

Abbildungen 4.7 und 4.8: Gemeinsame Darstellung der Boxplots der beiden Beispieldaten.

4.2.2 Statistische Tests

Ein statistisch signifikanter Unterschied der Lageparameter lässt sich am besten mithilfe statistischer Tests zeigen. Diese sind in der Lage, für eine aufgestellte Hypothese und ein vorgegebenes Signifikanzniveau α zu entscheiden, ob die Hypothese durch die Daten unterstützt wird oder nicht.

Zu unterscheiden sind zunächst der Einstichproben- und der Zweistichprobenfall. Im ersten Fall setzen wir bereits einen Standardalgorithmus voraus, dessen durchschnittlicher Fitnesswert für eine gegebene Problemklasse bekannt ist. Nun haben wir einen zweiten Algorithmus entworfen, dessen Verhalten wir mit dem des ersten vergleichen möchten. Wir führen also verschiedene unabhängige Durchläufe durch und überprüfen, ob die neuen durchschnittlichen Ergebnisse besser als die bisherigen sind.

Im Zweistichprobenfall haben wir zwei Algorithmen, deren durchschnittliche Fitnesswerte wir bisher noch nicht kennen. Ein Zweistichprobentest wird uns Anhaltspunkte darüber liefern, welcher der beiden Algorithmen vorzuziehen ist. Hierzu benötigen wir jeweils eine Stichprobe von beiden Algorithmen.

Um die Tests anwenden zu können, müssen verschiedene Voraussetzungen erfüllt sein. Eine wesentliche und für alle Tests gleichermaßen notwendige Voraussetzung ist dabei die Unabhängigkeit der einzelnen Stichprobenelemente. Auch wenn die Resultate nur durch Computersimulationen erzielt werden, sollte dies in allen Situationen der Fall sein, in denen mehrere Durchläufe eines Algorithmus mit einem neuen zufälligen Startwert des verwendeten Zufallszahlengenerators gestartet werden.

Darüber hinaus müssen oftmals weitere Voraussetzungen erfüllt sein. Hierbei kann es sich zum einen um eine stetige Verteilung oder sogar eine normalverteilte Grundgesamtheit handeln. Zum anderen muss zur Durchführung statistischer Tests ein bestimmter Stichprobenumfang vorhanden sein. Damit bspw. der Mittelwert normalverteilt ist, wird häufig ein Stichprobenumfang von mindestens 30 vorausgesetzt (Hartung et al. 1999, S. 505–508).

Aufgrund erster Experimente hat man eine gewisse Annahme über das Verhalten des Algorithmus bzw. der Algorithmen. Statistische Tests haben nun die Eigenschaft, eine Annahme nicht direkt bestätigen zu können. Stattdessen deuten sie lediglich auf statistisch signifikante Abwei-

chungen hin. Im Folgenden verwenden wir die aus der statistischen Testtheorie bekannten Begriffe der Nullhypothese und der Alternativhypothese. Die Nullhypothese ist oftmals eine Negierung der Vermutung und soll durch den Test abgelehnt werden, um die als Alternativhypothese formulierte eigentliche Vermutung signifikant zu stützen.

Bedingt durch statistische Schwankungen kann es passieren, dass die Verläufe für einen bestimmten EA sehr ungünstig verlaufen. Auch wenn dies mit zunehmender Zahl unabhängiger Durchläufe reduziert werden sollte, könnte somit ein Test trotz Gleichheit einen signifikanten Unterschied feststellen. In diesem Fall würde die Nullhypothese verworfen und wir begingen einen so genannten Fehler 1. Art. Die Wahrscheinlichkeit hierfür ist beherrschbar und wird durch das vorgegebene Signifikanzniveau α des Tests bestimmt. Bei einem üblichen $\alpha = 0.05$ beträgt die Wahrscheinlichkeit, die Nullhypothese trotz Zutreffens abzulehnen, also 5%. Umgekehrt ist es auch möglich, dass die Nullhypothese zwar falsch ist, aber unser Test nicht in der Lage ist, dies zu identifizieren. Wir werden dann die Nullhypothese beibehalten und einen so genannten Fehler 2. Art begehen. Die Wahrscheinlichkeit diesen Fehler zu begehen beträgt β und ist im Gegensatz zum Fehler 1. Art wesentlich schwieriger zu bestimmen und zu beherrschen, weshalb ihm meistens weniger Beachtung geschenkt wird. Kann man diese Fehlerwahrscheinlichkeit bestimmen, so wird der Term $1 - \beta$ als Teststärke oder Power des Tests bezeichnet, die angibt, mit welcher Wahrscheinlichkeit eine nicht zutreffende Nullhypothese abgelehnt wird.

Kolmogorow-Smirnow- und χ^2 -Test

Verschiedene Tests stellen bestimmte Anforderungen an die Stichprobenwerte. Hierzu gehört bspw., dass die Stichprobe aus einer Normalverteilung stammt. Davon kann i.d.R. aber nicht ausgegangen werden. Um die Vermutung einer Normalverteilung zu überprüfen, können der Kolmogorow-Smirnow-Test (KS-Test) oder der χ^2 -Test eingesetzt werden. Beide Tests gehören zu den nicht-parametrischen Verfahren und sind wie die anderen ebenfalls in diesem Abschnitt dargestellten Verfahren ausführlich bei Büning und Trenkler (1994) erläutert.

Der KS-Test vergleicht im Einstichprobenfall die Verteilung der Stichprobe mit der angenommenen Verteilung. Dabei bildet er an allen Stellen des Wertebereichs der Verteilungen die Differenz zwischen beiden Verteilungen. Der maximale absolute Abstand wird dann als Kriterium benutzt, ob die beiden Verteilungen identisch sein können.

Da der KS-Test voraussetzt, dass die zugrunde liegende Verteilung stetig ist, sollten in der Stichprobe keine Bindungen (gleiche Werte für verschiedene Stichprobenelemente) auftreten. Ist dies dennoch der Fall, kann der Test nicht die vorgegebenen Niveaus einhalten.

Ist die Zahl der Bindungen nicht zu groß, kann der Test mit gewisser Vorsicht dennoch ausgeführt werden. Für den Fall nicht zu vieler Bindungen ist eine Alternative denkbar, die auch in anderen Zusammenhängen benutzt werden kann. Dabei addiert man per Zufallsprozess auf alle Werte zufällige Störterme, deren Größen von der minimalen absoluten Differenz der Daten abhängen und dafür sorgen, dass alle Bindungen aufgehoben werden, aber nicht die Reihenfolge bereits unterschiedlicher Werte verändern. Ist die Anzahl der Bindungen jedoch zu groß, sollte auf dieses Verfahren und die Anwendung des KS-Tests verzichtet werden, da dann die zugrunde liegende Verteilung kaum stetig und eine Normalverteilung eher ausgeschlossen ist. Hiervon ist insbesondere bei der Optimierung einer multimodalen Funktion auszugehen, da die Optimierungsalgorithmen bei solchen Funktionen häufig in den gleichen lokalen Optima stecken bleiben. Aufgrund der addierten zufälligen Terme können sich die Resultate für verschiedene Tests mit den gleichen Ausgangsdaten natürlich leicht unterscheiden.

Darüber hinaus haben solche Tests einen weiteren Nachteil. Beim KS-Test muss die Vergleichsverteilung genau spezifiziert werden. Es reicht also nicht aus anzugeben, dass auf das Vorliegen einer Normalverteilung geprüft werden soll. Stattdessen müssen der Mittelwert und die Standardabweichung der Normalverteilung angegeben werden. Aufgrund fehlender Kenntnisse bedient man sich hierzu der aus der Stichprobe geschätzten Werte. Dies führt allerdings dazu, dass der Test konservativ wird, d.h. der Fehler 1. Art verringert sich, wodurch die Nullhypothese weniger häufig abgelehnt wird. Diese Problematik lässt sich nur schwer handhaben und kann hier nicht detailliert erörtert werden. Wird ein anderer Test dem eigentlichen Test vorgeschaltet, ist jedoch grundsätzlich äußerste Vorsicht geboten. Denn wie Sachs (1997, S. 200) erläutert, „ist die Irrtumswahrscheinlichkeit des eigentlichen Tests nicht mehr korrekt, da sie durch die Vortests in undurchsichtiger Weise modifiziert worden ist.“

Während zur Anwendung des KS-Tests die zugrunde liegende Verteilung stetig sein muss, ist dies für den χ^2 -Test nicht erforderlich. Bei diesem Test wird der Wertebereich möglichst so in ca. \sqrt{n} Klassen aufgeteilt, dass bei Vorliegen der angenommenen Verteilung in jeder Klasse mindestens fünf Elemente der Stichprobe liegen. Wegen der Klassenbildung sind auch Bindungen in der Stichprobe ohne Belang. Anschließend wird eine Teststatistik berechnet, die auf den quadrierten Differenzen zwischen angenommenen und realisierten Klassenstärken basiert. Ist der Wert der Teststatistik zu groß, muss davon ausgegangen werden, dass die Stichprobe nicht der angenommenen Verteilung entstammt. Auch beim χ^2 -Test werden oftmals fehlende Verteilungsparameter aus der Stichprobe geschätzt. Um das Testniveau einzuhalten, wird in diesem Fall die Anzahl der Freiheitsgrade um die Zahl der geschätzten Parameter reduziert.

Gauß- und t-Test

Die Standardtests zur Ermittlung statistisch signifikanter Unterschiede der Mittelwerte zweier Stichproben sind der Gauß- und der t-Test. Beide sind theoretisch gut fundiert und in jedem Standardwerk beschrieben (Hartung et al. 1999). Der t-Test unterscheidet sich vom Gaußtest dadurch, dass zu seiner Durchführung die Standardabweichung σ der zugrunde liegenden Normalverteilung aus der Stichprobe durch s geschätzt wird, während beim Gaußtest die Standardabweichung bekannt sein muss. Als Teststatistik wird in den beiden Einstichprobenfällen

$$\frac{\bar{x} - \mu}{\sigma} \cdot \sqrt{n} \sim N(0, 1) \quad (\text{bekannte Varianz}) \text{ bzw.}$$

$$\frac{\bar{x} - \mu}{s} \cdot \sqrt{n} \sim t_{n-1} \quad (\text{unbekannte Varianz})$$

benutzt. Dabei bezeichnet μ den Mittelwert der angenommenen Verteilung und t_{n-1} die T-Verteilung mit $n - 1$ Freiheitsgraden. Eine bekannte Standardabweichung kann jedoch bei experimentellen Untersuchungen von EA als nahezu ausgeschlossen gelten.

Eine wesentliche Voraussetzung zur Anwendung der Tests ist eine normalverteilte Grundgesamtheit. Wird nur der Mittelwert geprüft, sind die Tests bereits anwendbar, wenn die Verteilung des Mittelwertes einer Normalverteilung genügt. Wie bereits erwähnt wird dies ab 30 Stichprobenelementen angenommen.

Der Zweistichproben-t-Test dient dazu, Unterschiede des Mittelwertes zweier Stichproben aufzuzeigen. Sind die Daten normalverteilt, handelt es sich um den aus statistischer Sicht besten Test.

Wilcoxon-Rangsummentest

Sind die Voraussetzungen des Gaußtests nicht erfüllt oder ist die Unsicherheit über die Vermutung zu groß, andere Verfahren zu benutzen. Allgemein anwendbare Verfahren sind die bereits angesprochenen nichtparametrischen Verfahren (Büning und Trenkler 1994), die meist nur geringe Anforderungen an die zugrunde liegenden Daten stellen. Bei diesen Voraussetzungen handelt es sich bspw. um die Stetigkeit der Verteilung oder das Vorliegen eines mindestens ordinalen Skalenniveaus der Daten. Dieses liegt dann vor, wenn auf den Daten eine Ordnung definiert ist.

Um einen solchen verteilungsunabhängigen Test handelt es sich bei dem Wilcoxon-Rangsummentest, der äquivalent zum Mann-Whitney-U-Test ist. Mit dem Wilcoxon-Rangsummentest wird getestet, ob die Mediane zweier Verteilungen mit mindestens ordinalskaliertem Messniveau gleich sind. Den Wilcoxon-Test gibt es für den Einstichproben- und Zweistichprobenfall. In beiden Fällen wird als Voraussetzung gefordert, dass die Verteilungen stetig und symmetrisch um den Median sind. Für Stichprobenumfänge $n \geq 10$ sind wenige Bindungen nicht sehr bedeutsam.

Der Wilcoxon-Vorzeichentest oder -Rangsummentest ist auch bei Vorliegen einer Normalverteilung dem t-Test kaum unterlegen. Schon bei kleineren Abweichungen von der Normalverteilung ist er dem t-Test überlegen und sollte diesem vorgezogen werden.

Van der Waerden-Test

Um einen weiteren Zweistichproben-Test handelt es sich bei dem van der Waerden- oder auch X_N genannten Test. Dieser verbindet parametrische wie auch nichtparametrische Elemente miteinander. Der X_N -Test prüft wiederum auf Gleichheit der beiden zugrunde liegenden Verteilungen. Als parametrischen Anteil benutzt er zur Gewichtung der Ränge die Quantile der Standardnormalverteilung. Bei Vorliegen von Bindungen können Durchschnittsränge vergeben werden. Der Test von van der Waerden ist bei Abweichung von der Normalverteilung ebenfalls dem t-Test vorzuziehen, findet aber seltener Anwendung als der Wilcoxon-Test, weil er rechenaufwändiger ist.

Simultane Vergleiche

Prinzipiell können die bisher angesprochenen Techniken auch dann verwendet werden, wenn mehr als zwei Strategien miteinander verglichen werden sollen. Allerdings ergibt sich hierbei eine Schwierigkeit, die besonders beachtet werden muss. Bei jedem Test ist im Vorhinein ein Signifikanzniveau α anzugeben. Dieses ist im Fall von zwei zu vergleichenden Datensätzen präzise. Im Fall von mehr als zwei Datensätzen dürfte eigentlich kein paarweiser Vergleich der Strategien mehr stattfinden. Dies liegt an der Tatsache, dass für jeden Einzeltest eine Irrtumswahrscheinlichkeit von $1 - \alpha$ vorliegt. Durch mehrfache Tests wird diese Irrtumswahrscheinlichkeit u.U. wesentlich erhöht. Darüber hinaus ist zu beachten, dass die paarweisen Tests nicht unabhängig voneinander sind. Die Testtheorie gibt verschiedene Hinweise zur Behandlung dieses unter dem Titel simultane oder multiple Vergleiche bezeichneten Problems, bekannte Verfahren stammen von Tukey, Scheffé und Bonferroni. Für eine ausführlichere Darstellung dieser hier aus Platzgründen nicht näher behandelten Thematik siehe bspw. Sachs (1997, S. 649ff) und Kockelkorn (2000, S. 543ff). Weitere Hinweise zu den in diesem Kapitel vorgestellten sowie zusätzlichen Tests enthalten viele Lehrbücher der Statistik. Dazu zählen u.a. die Werke von Hartung et al. (1999), Kreienbrock (1993) sowie Kockelkorn (2000).

4.3 Gütemaße

Zum Vergleich zweier oder mehrerer Algorithmen ist ein Gütemaß¹ erforderlich. Ein bestimmtes Maß sollte dabei die besonderen Anforderungen der Untersuchung berücksichtigen. So kann es einerseits wichtig sein, über den gesamten Zeitraum einen möglichst guten durchschnittlichen Fitnesswert zu erzielen. Andererseits könnte gefordert werden, dass der Algorithmus zu jedem Zeitpunkt einen bestimmten Funktionswert nicht überschreitet. Jede Anforderung bedingt ein unterschiedliches Vorgehen beim Vergleich und kann ein anderes Gütemaß erfordern.

Insbesondere bei reellwertigen Problemstellungen wird ein Algorithmus nur selten das genaue Optimum erreichen. Statt das Optimum genau zu treffen, kann es dann auch akzeptabel sein, das Optimum mit einer geringen Abweichung zu bestimmen. Wird als Maß die prozentuale Erfolgsquote verwendet genügt es dann, den akzeptablen Bereich um das Optimum zu treffen.

Prinzipiell kann auch die Anzahl der Funktionsauswertungen maßgeblich sein, die zum Erreichen einer bestimmten Güte benötigt wird. Eiben und Jelasity (2002) definieren dazu die durchschnittliche Anzahl Funktionsauswertungen (**A**verage number of **E**valuations to a **S**olution) bis zum Erreichen einer bestimmten Güte. Allerdings muss dieses Maß immer in Kombination mit der Erfolgsrate angegeben werden, da sich die AES nur auf erfolgreiche Läufe bezieht. Vor allem in dynamischen Umgebungen ist ein solches Maß aber von untergeordneter Bedeutung. Deshalb wird es dort für Vergleichszwecke wesentlich seltener benutzt.

Aufgrund der gegebenen Anforderungen kommen unterschiedliche Maße in Betracht, die in den folgenden Abschnitten erläutert werden. Zweierlei Berechnungen fallen dabei an. Einerseits müssen die Daten eines Laufes zu einem Maß komprimiert werden. Andererseits sind die Daten mehrerer unabhängiger Läufe geeignet zusammenzufassen. Dabei könnte es sinnvoll sein, von den Einzelläufen zu abstrahieren und bspw. als Maß das arithmetische Mittel aller Funktionswerte des jeweils besten Individuums aller Generationen und aller Durchläufe zu bilden. Alternativ ist es denkbar, jeden Durchlauf als Einheit zu betrachten und einen durchschnittlichen, aber echten Durchlauf als Repräsentanten auszuwählen und aus diesem ein geeignetes Maß zu berechnen.

Eiben und Jelasity (2002) unterscheiden die so genannte Spitzen- und die Durchschnittsperformance. Während letztere selbsterklärend ist, wird bei der ersten von mehreren Läufen nur der beste Lauf betrachtet und zur Berechnung des Maßes herangezogen. Dies ist dann aber stark von der Anzahl der durchgeführten Läufe abhängig, denn mit zunehmender Anzahl wird dieser Wert immer besser. Ein Durchschnittswert wird unter der Voraussetzung der Erwartungstreue dagegen immer den gleichen Wert liefern. Hier wirkt sich die Anzahl der Läufe lediglich auf die Streuung um den echten Wert aus. Unter diesem Aspekt ist die implizite Empfehlung von Eiben und Jelasity (2002) diskussionswürdig, als Maß eher den Spitzenwert zu nehmen.

Die Skalierbarkeit eines Algorithmus wird üblicherweise durch die benötigte Rechenzeit in Abhängigkeit von der Problemdimension gemessen. Dies ist sicherlich ein wichtiges Maß, Barr et al. (1995) schlagen aber vor, zusätzlich die Qualität der besten Lösung in Abhängigkeit der Problemdimension anzugeben. Damit zusammenhängend ist es nützlich, die erreichte Güte in Abhängigkeit von der benötigten Rechenzeit aufzutragen. Weitere hilfreiche Angaben umfassen die benötigte Rechenzeit in Abhängigkeit von der Problemdimension und die Robustheit des Algorithmus in Abhängigkeit von der erreichbaren Güte.

Beyer (2001, S. 14-18) unterscheidet die lokale und die globale Konvergenz. Unter der globalen Konvergenz versteht er das Verhalten des EA, gegen das (globale) Optimum zu konvergieren. Zur

¹In der gängigen Literatur wird auch der Begriff des Performance-Maßes zur Beurteilung der Leistung eines Algorithmus verwendet. In dieser Arbeit wird jedoch ausschließlich der neutralere Begriff des Gütemaßes benutzt.

Approximation des Optimums mit einer gewissen Abweichung gibt er die wünschenswerte Eigenschaft so an:

$$\lim_{g \rightarrow \infty} \mathbb{P} (|f(x_g) - f(x^*)| < \epsilon) = 1 ,$$

für ein passendes $\epsilon > 0$.

Den Nutzen solcher Analysen für praktische Anwendungen schränkt Beyer (2001, S. 16) selbst ein, weil die Durchführung über $\lim g \rightarrow \infty$ Generationen nicht möglich ist. Somit ist eine Betrachtung über einem endlichen Zeitraum notwendig. EA sind dafür bekannt, in kurzer Zeit akzeptable Lösungen zu liefern. Diese Eigenschaft steht in Opposition zur Konvergenzsicherheit. Um das Optimum sicher zu finden, muss der gesamte Suchraum überprüft werden.

Neben der globalen Konvergenz geht Beyer (2001, S. 17) auch auf die lokalen Aspekte der Konvergenz ein. Durch die Variationsoperatoren evolutionärer Algorithmen wird eine lokale Nachbarschaft definiert. Diese muss eng begrenzt sein, weil die Wahrscheinlichkeit einer Verbesserung mit zunehmender Nachbarschaftsgröße i.d.R. immer kleiner wird. Als Maß für den lokalen Fortschritt unterscheidet er die Fortschrittsrate φ im Parameterraum und den Qualitätsgewinn \bar{Q} im Fitnessraum.

Die Konzentration auf den Fitnesswert wird auch durch verschiedene Maße der Konvergenzgeschwindigkeit begründet. So stützt sich Bäck (1996, S. 151) auf eine Definition von Schwefel (1988), die dieser bereits vorher (Schwefel 1987) eingeführt hat. Während Bäck als Maß

$$\ln \sqrt{\frac{f_{\min}^{(0)}}{f_{\min}^{(g)}}}$$

verwendet, ist im Original von Schwefel der Logarithmus zur Basis 10 benutzt worden. Die Verwendung des Logarithmus zur Basis 10 hat den Vorteil, dass das Fortschrittsmaß die Verbesserung in 10er-Potenzen angibt. Dieses Vorgehen wird auch von Kursawe (1999, S. 50) gestützt. Er verwendet neben dem von Schwefel (1988) definierten Konvergenzmaß auch

$$\frac{n}{\text{Generationen}} \cdot \log \sqrt{\frac{f_{\min}^{(0)}}{f_{\min}^{(g)}}} ,$$

das berücksichtigt, dass eine Konvergenz bei höherdimensionalen Problemstellungen i.Allg. schwieriger ist.

Dagegen berechnet Mück (1989) den logarithmierten Quotienten

$$\log \sqrt{\frac{f(x^{(0)})}{\frac{1}{\mu} \sum_{i=1}^{\mu} f(x_i^{(g)})}}$$

aus der Wurzel des Funktionswertes des Startvektors und dem arithmetischen Mittel der Funktionswerte der Population nach g Generationen. Man erhält somit ein Maß für die auf den Abstand zum Ziel bezogene Verbesserung vom Start aus in Zehnerpotenzen.

Auch in dynamischen Umgebungen besteht eine der einfachsten Möglichkeiten des Vergleichs zweier oder mehrerer Verfahren darin, jedem Verfahren einen einfachen Zahlenwert seiner Leistungsfähigkeit zuzuordnen. Hierbei könnte es sich bspw. um eine Abschätzung der Genauigkeit, sich einem Optimum anzupassen oder seiner Konvergenzgeschwindigkeit handeln, jeweils gemittelt über ein mehr oder weniger großes Spektrum von Zielfunktionen. Die Eigenschaft in der letzten Situation ist besonders dann von Interesse, wenn sich die Umgebung selten ändert, dafür aber

größere Anpassungen notwendig sind. In diesem Fall könnte der Schwerpunkt darauf liegen, dass sich der EA schnell an die neue Umgebung anpasst.

Einen ersten Ansatzpunkt für auf dynamische Umgebungen angepasste Maße liefern die von Trojanowski und Michalewicz (1999a) eingeführten und auf De Jong (1975) zurückgehenden Maße ACC und ADA. De Jong (1975, S. 11-12) unterscheidet zwei lokale Gütemaße:

- Die on-line Performance einer Strategie ist durch den Fall motiviert, dass ein System in Echtzeit optimiert werden muss. Hierbei ist entscheidend, dass sich ein gutes durchschnittliches Verhalten ergibt.
- Die off-line Performance einer Strategie ist durch den Fall motiviert, dass ein System zunächst analysiert und anschließend in Echtzeit mit der besten Lösung betrieben werden kann.

Diese Maße wurden von Trojanowski und Michalewicz (1999a) aufgegriffen und auf den Fall übertragen, dass in dynamischen Umgebungen die Änderungen in gleichen Generationsabständen l erfolgen:

- Genauigkeit (accuracy):

$$ACC = \frac{1}{K} \sum_{i=1}^K (err_{i,l-1}) .$$

Hierbei handelt es sich um die durchschnittliche Abweichung des besten Individuums einer jeden Generation vom Optimum zum Zeitpunkt jeweils vor einer Änderung der Umwelt. Erfolgt ein EA-Lauf bspw. über 100 Generationen und ändert sich die Fitness alle 20 Generationen, so wird der Durchschnitt von insgesamt 5 Werten benutzt, nämlich den Abweichungen zu den Zeitpunkten $t = 19, 39, 59, 79, 99$.

- Anpassungsfähigkeit (adaptability):

$$ADA = \frac{1}{K} \sum_{i=1}^K \frac{1}{l} \sum_{j=0}^{l-1} (err_{i,j}) .$$

Die Anpassungsfähigkeit misst die durchschnittliche Abweichung des jeweils besten Individuums einer Generation vom Optimum über den gesamten EA-Lauf.

In beiden Fällen bezeichnet $err_{i,j}$ die Differenz der Fitness zwischen der besten Lösung der j -ten Generation nach der letzten (der i -ten) Änderung und dem optimalen Wert nach der i -ten Änderung ($i \in \{0, \dots, K-1\}$, $j \in \{0, \dots, l-1\}$). Dabei bezeichnet K die Anzahl der Änderungen in einem Lauf und l gibt die Anzahl der Generationen zwischen zwei Änderungen an. Für beide Maße (ACC und ADA) sind diejenigen Werte besser, die näher an Null liegen.

Einen allgemeineren Überblick über diese Thematik bietet der ähnliche Artikel (Trojanowski und Michalewicz 1999b) der gleichen Autoren. Während das Maß der Anpassungsfähigkeit auch bei kontinuierlichen Änderungen anwendbar ist, versagt das Genauigkeitsmaß in diesen Situationen. Die von De Jong (1975) eingeführten Maße werden u.a. auch von Branke (2002) sowie Branke und Schmeck (2003) aufgegriffen und direkt auf den dynamischen Fall übertragen.

Weicker (2002) definiert zunächst drei Ziele der dynamischen Optimierung. Diese bestehen zum einen darin, die Optimierungsgenauigkeit zu einem Zeitpunkt t zu maximieren. In diesen Wert fließen sowohl der minimale als auch maximale Funktionswert im Suchraum zum Zeitpunkt t ein. Dies erscheint wenig sinnvoll, da eine bloße Vergrößerung des Suchraums eine Änderung dieses

Wertes zur Folge hat. Zum anderen sollte die Stabilität des Verfahrens verbessert werden. Diese definiert er als den Wert der Verschlechterung der Optimierungsgenauigkeit von einer Generation zur nächsten. Kann sich der EA sogar verbessern, so setzt er den Wert der Stabilität auf Null, was gleichzeitig den besten Wert darstellt. Bzgl. der Stabilität erwähnt er selbst, dass diese die absoluten Funktionswerte unberücksichtigt lässt. Von besonderer Bedeutung in Situationen, in denen sich längere statische Phasen mit kurzen Phasen und großen Änderungen abwechseln, ist zum dritten die ϵ -Reaktionsfähigkeit zum Zeitpunkt t . Sie gibt an, wie schnell ein Algorithmus nach einer Umgebungsänderung bis auf einen Faktor ϵ wieder an die letzte gemessene Genauigkeit herankommen kann.

Alle drei Kriterien basieren auf der Kenntnis des besten Funktionswertes. Dieser muss jedoch nicht zwangsläufig bekannt sein. Des Weiteren ist es möglich, dass sich der beste Funktionswert nach einer Umgebungsänderung ebenfalls geändert hat, sodass die Werte der Maße nicht mehr vergleichbar sind.

Einen guten Überblick über die in der gängigen Literatur verwendeten Gütemaße bietet auch Morrison (2003). Generell ist zu bemerken, dass die wenigsten der Veröffentlichungen zwischen den beiden Phasen bei der dynamischen Optimierung unterscheiden, die eigentlich unterschiedliche Gütemaße erfordern. Verschiedene Algorithmen werden eine unterschiedlich lange Suchphase bis zum Erreichen der Verfolgungsphase benötigen. Aus nahe liegenden Gründen kann auch im dynamischen Fall ein Algorithmus nur über eine begrenzte Anzahl Generationen ausgeführt werden. Ein Algorithmus mit einer langen Suchphase, aber einer sehr nahen Verfolgung des Optimums könnte bei Mittelung über beide Phasen einem anderen Algorithmus benachteiligt sein, der sich dem Optimum zwar sehr schnell nähert, diesem aber in einem größeren Abstand folgt.

Aufgrund der für jedes bestehende Gütemaß existierenden Nachteile entwickeln wir hier ein neues Gütemaß, den so genannten **Average Best Function Value**. Im Gegensatz zu anderen Maßen berücksichtigt er speziell die Resultate der Verfolgungsphase. Dabei wird davon ausgegangen, dass für einen Anwender die erreichten Funktionswerte das entscheidende Kriterium zur Beurteilung der Güte eines Verfahrens darstellen. Während in multimodalen Landschaften der Abstand zum Optimum nicht unbedingt aussagekräftig ist, sollten auch hier die Funktionswerte das entscheidende Kriterium für den Anwender sein und Aussagen zur Robustheit des Verfahrens ermöglichen. Gleichzeitig sollten jedoch einzelne extreme Ausreißer das Maß nicht beeinflussen und so Aussagen möglicherweise verfälschen. Der ABFV berücksichtigt diese Anforderungen, indem er sich alleine auf die während der Optimierung erreichten Funktionswerte stützt und andere Kriterien wie den Abstand zum Optimum unberücksichtigt lässt. Daneben greift er zur Bildung des Durchschnitts auch auf den Median zurück, der unempfindlicher gegenüber extremen Ausreißern ist als der Mittelwert. Im Folgenden wird dieses Maß näher erläutert.

In der Abbildung 4.9 sind die Ausschnitte einiger typischer Verläufe der Suchphase in dynamischen Umgebungen dargestellt. Ein Einzellauf ist dabei gekennzeichnet durch seine Schwankungen um einen durchschnittlichen Funktionswert. Mehrere Einzelläufe werden sich dann von Zeit zu Zeit überschneiden, sollten aber ein prinzipiell gleiches Verhalten aufweisen.

Zur Definition des ABFV abstrahieren wir nun von einzelnen realen Durchläufen einer ES. Die aus m Durchläufen für jede Generation erhaltenen m Funktionswerte werden zunächst sortiert. Anschließend werden die Funktionswerte jeder Generation verbunden, die dem gleichen Sortierniveau entsprechen. Das bedeutet, dass alle besten Werte miteinander verbunden werden, alle zweitbesten, alle drittbesten usw. bis zum Verbinden aller schlechtesten Funktionswerte (Abb. 4.10). Auf diese Weise erhalten wir m sortierte virtuelle Läufe. Nach der Berechnung der Mittelwerte dieser m Läufe

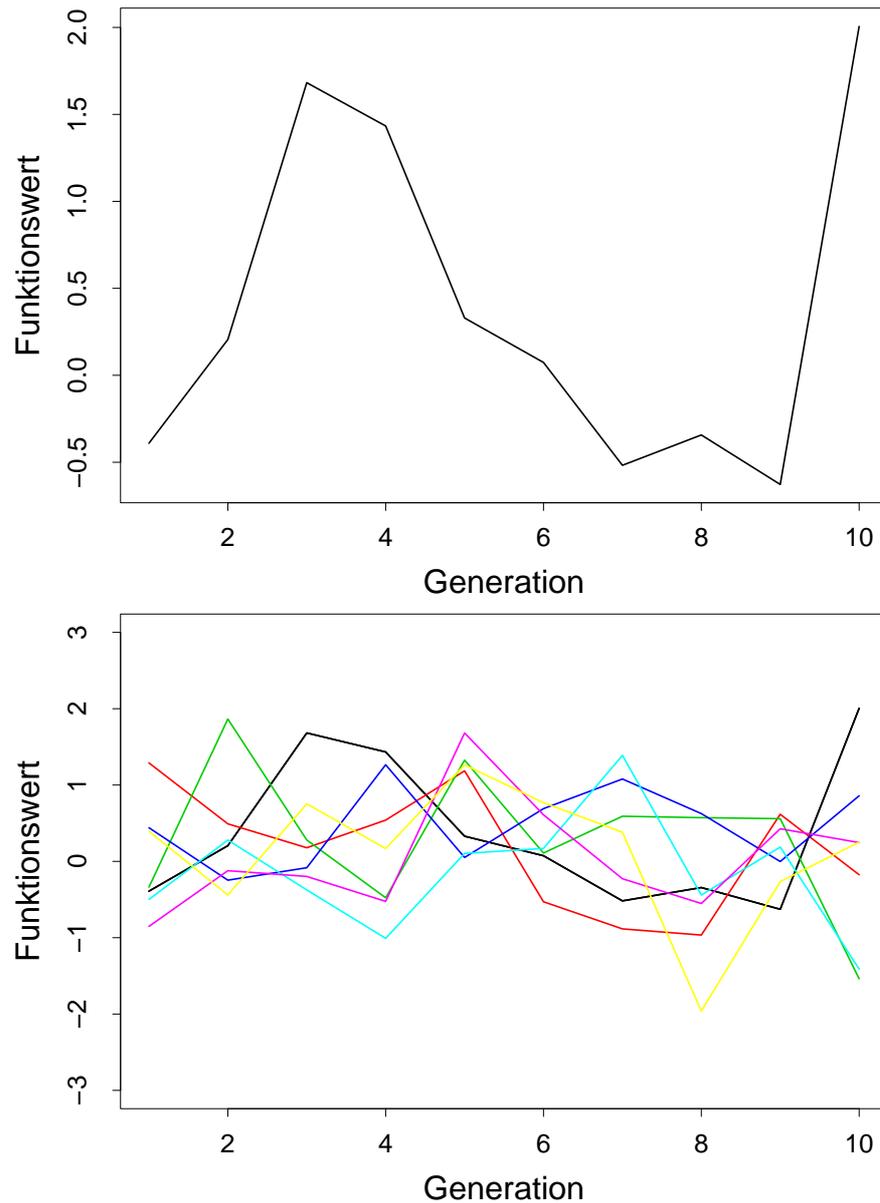


Abbildung 4.9: Typische Verläufe eines fiktiven EA während der Verfolgungsphase in einer dynamischen Umgebung: Einzellauf (oben) und Mehrfachläufe (unten).

lassen wir die Läufe selbst unberücksichtigt und betrachten einzig noch die berechneten Mittelwerte. Der Median der so berechneten m Mittelwerte dient uns nun als Gütemaß ABFV. Dieses Maß ist aufgrund der Berechnungen sehr robust gegenüber auch extremen einzelnen Ausreißern in jedem Einzellauf.

Eine formale Definition des ABFV wird im Folgenden geschildert. Sei $f_i^{(\gamma)}$ der Funktionswert des besten Individuums der Generation γ des i -ten Laufs. Dann können die Funktionswerte als Matrix F geschrieben werden:

$$F := \begin{pmatrix} f_1^{(1)} & f_1^{(2)} & \dots & f_1^{(g)} \\ f_2^{(1)} & f_2^{(2)} & \dots & f_2^{(g)} \\ \dots & \dots & \dots & \dots \\ f_m^{(1)} & f_m^{(2)} & \dots & f_m^{(g)} \end{pmatrix},$$

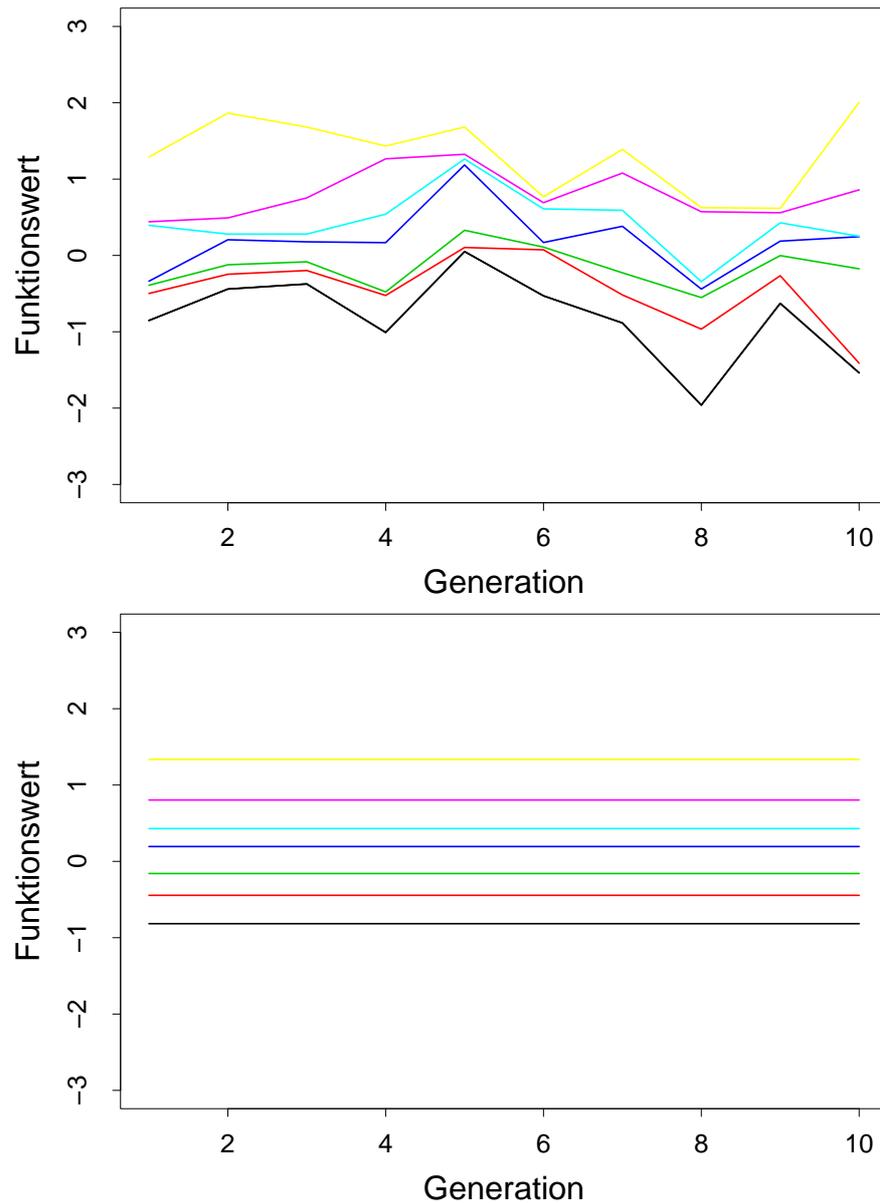


Abbildung 4.10: Sortierte virtuelle Läufe (oben) und deren Mittelwerte (unten).

wobei g die Anzahl der ausgeführten Generationen pro Lauf und m die Anzahl der Durchläufe bezeichnet. Zur Berechnung des ABFV müssen diese Werte für jede Generation geordnet werden:

$$F' := \begin{pmatrix} f_{(1)}^{(1)} & f_{(1)}^{(2)} & \cdots & f_{(1)}^{(g)} \\ f_{(2)}^{(1)} & f_{(2)}^{(2)} & \cdots & f_{(2)}^{(g)} \\ \vdots & \vdots & \ddots & \vdots \\ f_{(m)}^{(1)} & f_{(m)}^{(2)} & \cdots & f_{(m)}^{(g)} \end{pmatrix},$$

Der Wert $f_{(i)}^{(\gamma)}$ bezeichnet also den i . besten aller Werte $f_i^{(\gamma)}$ der Generation γ . Die Werte $f_{(i)}^{(\cdot)}$ stellen dann den i . sortierten virtuellen Lauf dar. Mit $i := \lfloor (m+1)/2 \rfloor$ erhalten wir den mittleren sortierten virtuellen Lauf. Der Durchschnitt

$$\bar{f}_{\lfloor (m+1)/2 \rfloor} := \frac{1}{g} \sum_{j=1}^g f_{\lfloor (m+1)/2 \rfloor}^{(j)}$$

ist dann der gesuchte ABFV. Für gerades m ist der Median nicht eindeutig definiert. Hier wird der Einfachheit halber der bessere $\bar{f}_{(\lfloor(m+1)/2\rfloor)}$ der beiden mittleren Werte verwendet, wobei sich bei genügend großer Anzahl Durchläufe jedoch keine spürbaren Änderungen durch eine Verwendung des anderen mittleren Wertes $\bar{f}_{(\lceil(m+1)/2\rceil)}$ ergeben. In unseren Experimenten werden wir jedoch immer eine ungerade Anzahl Durchläufe ausführen, sodass der Median eindeutig definiert ist.

An dieser Stelle soll ein kurzer Vergleich der Eigenschaften des ABFV mit den in den vorigen Absätzen angesprochenen anderen Gütemaßen erfolgen. Die Definition des ABFV lehnt sich dabei an die aus der Zeitreihenanalyse bekannte Mittelwertfunktion an, wo zu jedem Zeitpunkt t das arithmetische Mittel aller Elemente benutzt wird (Schlittgen und Streitberg 1989, S. 74). Aufgrund der noch zu beobachtenden großen Schwankungen innerhalb eines EA-Laufs wird beim ABFV jedoch der Median benutzt. Dieser ist insbesondere dann näher am echten mittleren Wert, wenn das Optimum den Funktionswert Null besitzt. Beim arithmetischen Mittel würde schon ein einziger großer Ausreißer nach oben genügen, um einen wesentlich verfälschten Eindruck des mittleren Werts zu liefern. Ein weiterer Vorteil dieses Maßes besteht darin, dass im Gegensatz zu bspw. dem ADA der Wert des globalen Optimums nicht bekannt sein muss. Es ist somit auch für bisher wenig untersuchte Problemstellungen geeignet, bei denen das genaue Optimum unbekannt ist.

Zwei Aspekte bilden ein Alleinstellungsmerkmal des ABFV gegenüber den anderen Gütemaßen. Zum einen werden von vornherein m Durchläufe als Einheit betrachtet. So benutzt z.B. der ADA wie viele andere Maße auch das arithmetische Mittel eines einzelnen Durchlaufs. Werden mehrere Durchläufe ausgeführt, wird über diese ebenfalls gemittelt. Der Einfluss einzelner Ausreißer bleibt jedoch auch bei der Mittelung über alle Läufe bestehen, sodass gerade bei großen Schwankungen die Durchschnittsfitness eines Laufs nicht unbedingt eine hinreichende Aussagekraft besitzt. Im Gegensatz dazu betrachtet man beim ABFV sämtliche Läufe zusammen und bildet das arithmetische Mittel eines mittleren Laufs.

Der Vergleich der ABFV für unterschiedliche Strategien dient nun als ein Aspekt zur Einteilung der Verfahren. Um weitere Aussagen zur Güte eines Verfahrens treffen zu können ist es darüber hinaus aber auch möglich, alle m Mittelwerte zu verwenden. Diese spiegeln dann jeweils unterschiedliche Aspekte der Verfahren wider. Beispielsweise zeigt der Mittelwert der schlechtesten Werte an, wie stark sich die Extremwerte bemerkbar machen. Interessant ist dieser Wert bspw. dann, wenn ein Verfahren nur möglichst geringe Abweichungen vom Durchschnittswert haben soll. Die Verwendung aller m Mittelwerte gestattet auch die Durchführung weitergehender statistischer Tests wie sie in den vorangehenden Abschnitten dieses Kapitels dargestellt wurden. In den grafischen Darstellungen der nachfolgenden Kapitel werden häufig neben dem ABFV auch die 0.05%- und 0.95%-Quantile des ABFV aufgetragen. Dadurch ergibt sich ein approximatives 90%-Konfidenzintervall des ABFV. Auch hierbei handelt es sich um ein wesentliches Element, das von anderen Autoren unbetrachtet bleibt. Diese definieren Gütemaße, die nur einen einzelnen Wert liefern und keine weiteren die statistischen Schwankungen berücksichtigenden Vergleiche ermöglichen.

Ein Aspekt dieses Maßes muss an dieser Stelle aber noch genannt werden. Es handelt sich um die Mittelwerte virtueller Läufe, d.h. die realen Läufe können sich durchaus stark von den sortierten Läufen unterscheiden, wodurch natürlich auch das Gütemaß unter diesem Aspekt benutzt werden muss. Trotzdem bietet sich die Vorgehensweise an, um ein möglichst robustes, aber dennoch einfaches Maß zu verwenden.

Das Maß ABFV ist hier zwar nur für dynamische Umgebungen eingeführt worden, kann aber leicht auf den statischen Fall erweitert werden. Dazu muss nur entschieden werden, wie der ABFV in den denkbaren Situationen berechnet werden soll. Zu den möglichen Situationen gehören:

- Stagnation: Stagniert das Verfahren in einem lokalen Optimum, lässt sich der ABFV in analoger Weise betrachten, als ob es sich um die Verfolgungsphase handelt.
- Konvergenz: Schreitet das Verfahren über den gesamten Zeitraum der Optimierung voran, ist zu entscheiden, ob das Verfahren so lange fortgeführt wird bis eine Stagnation einsetzt. Ist dies nicht möglich, bieten sich zwei Vorgehensweisen an. Zum einen kann auch hier ein gewisser Bereich der letzten Funktionswerte als Grundlage des Maßes genommen werden. Zum anderen ist es denkbar, nur den letzten Funktionswert zu nehmen, wenn sich ansonsten eine zu große Diskrepanz ergäbe.
- Divergenz: Hier gilt das Entsprechende wie für die Konvergenz.

In dieser Arbeit wird in statischen Umgebungen der Einfachheit halber und weil die Ergebnisse im statischen Fall nicht im Vordergrund stehen, nur der letzte Wert jedes Durchlaufs zur Berechnung des ABFV verwendet. Bei einer genügenden Anzahl Durchläufe sollte in statischen Umgebungen auch der letzte Wert eine hohe statistische Aussagekraft besitzen, da in diesen Situationen die Schwankungen nicht so groß sind. Die durch statistische Erkenntnisse begründete Zahl von jeweils mindestens 31 durchgeführten Läufen sollte dies gewährleisten.

4.4 Fazit

In diesem Kapitel wurden die Möglichkeiten zur Beurteilung und zum Vergleich verschiedener Strategien aufgrund empirischer Daten vorgestellt. Ein wesentlicher Bestandteil bildete die Darstellung der zur Einschätzung eines Verfahrens notwendigen Gütemaße. Hierzu wurden die gängigsten existierenden Gütemaße für statische und dynamische Umgebungen vorgestellt. Dies beinhaltet eine ausführliche Diskussion der Vor- und Nachteile der einzelnen Maße. Eines der Mankos dieser Maße besteht in der Anfälligkeit gegenüber einzelnen starken Ausreißern, die gerade in dynamischen Umgebungen auftreten. Als Abhilfe wurde deshalb ein neues Maß entwickelt, das unempfindlich auch gegenüber mehreren extremen Ausreißern ist. Darüber hinaus gestattet das neue Maß weitergehendere statistische Tests zur Beurteilung der Güte der eingesetzten Algorithmen.

Teil II

Experimentelle Untersuchungen

Kapitel 5

Aufbau der Experimente

5.1 Testfunktionen

Der Problematik der Auswahl einer geeigneten Suite von Testfunktionen ist in der Literatur eine Reihe von Artikeln gewidmet. Beispielhaft seien hier die frühen Beiträge von Schwefel (1975, S. 319ff), De Jong (1975, S. 196ff) sowie Törn und Žilinskas (1989, S. 183ff) genannt. Dazu gehören aber auch Beiträge zur Erzeugung konkreter Probleminstanzen wie sie beispielsweise von Michalewicz, Deb, Schmidt und Stidsen (2000) vorgestellt werden. Darüber hinaus findet sich im Internet eine Vielzahl weiterer Problemgeneratoren mit zahlreichen zusätzlichen Hinweisen und Lösungsverfahren (z.B. <http://www.cs.uwo.edu/~wspears/multi.kennedy.html> sowie <http://www.mat.univie.ac.at/~neum/glopt/test.html>). Einen umfangreichen neueren Katalog von Testfunktionen hat Suganthan (2005) mit Unterstützung zahlreicher anderer Autoren erstellt.

Jede einzelne dieser Testsuiten ist aus Sicht der jeweiligen Autoren gut begründet. Dennoch unterscheiden sich die gewählten Zusammenstellungen teilweise sehr. Die Berücksichtigung aller Richtlinien fällt einem Experimentator deshalb schwer. Auch wenn die jeweiligen Testfunktionen im eigenen Fall aus dem einen oder anderen Grund nicht in Betracht kommen, sollte man dennoch versuchen, die Argumentationskette nachzuvollziehen und daraus eigene Schlüsse zu ziehen. Exemplarisch soll an dieser Stelle nur einer der zahlreichen Artikel herausgegriffen werden.

Einen wesentlichen Teil ihres Artikels widmen Eiben und Jelasity (2002) den Testfunktionen. Sie bemängeln, dass zur Einordnung eines Verfahrens oftmals ein Vergleich auf einer größeren Anzahl Testfunktionen durchgeführt wird, die ohne Begründung aus der Literatur übernommen wurden. Stattdessen sollte das Testen eines Algorithmus auf genau definierten Problemklassen erfolgen. Dabei gelte es aber, solche nutzlosen allgemeinen Problemklassen wie die der häufig benutzten *NP-harten* zu vermeiden. Ein erforderliches Merkmal der Problemklassen muss die Ähnlichkeit des Verhaltens eines Algorithmus auf den in einer Problemklasse enthaltenen Problemen sein. So erwähnen die Autoren eine eigene Veröffentlichung, in der ein bestimmter EA innerhalb einer Problemklasse ein signifikant unterschiedliches Verhalten zeigte. Ohne das unterschiedliche Verhalten genauer zu spezifizieren, sei dies ein deutliches Indiz dafür, die Festlegung der Problemklassen zu überdenken.

Die Autoren erwähnen, dass den Problemgruppen Klassen von Fitnesslandschaften zugeordnet werden können. Problematisch sei dieses Vorgehen jedoch dann, wenn vom eigentlichen Problem abstrahiert wird und nur noch die Fitnesslandschaft betrachtet wird. Die Fitnesslandschaft wird dann durch die benutzten Repräsentationen und Operatoren der EA quasi transformiert. Darüber hinaus sind die Fitnesslandschaften von solchen Problemen nur schwer zu klassifizieren, deren Funk-

tionswerte wir nur durch umfangreiche Berechnungen an endlich vielen Stellen erhalten können.

Zur Beurteilung der Algorithmen empfehlen Eiben und Jelasity (2002) ein Vorgehen aus dem maschinellen Lernen, bei dem die Algorithmen für gewisse Probleme der Problemklasse (Trainingsdaten) optimiert werden und anschließend anhand anderer Probleme derselben Problemklasse (Testdaten) miteinander verglichen werden.

Oftmals werden Zielfunktionen verwendet, die durch Überlagerung mehrerer Peaks entstehen, die sich frei im Raum bewegen können. Solche Testfunktionen benutzen bspw. Morrison (2004, S. 72) und Branke (2002, S. 81). Insgesamt bieten solche Funktionstypen eine Vielzahl interessanter Möglichkeiten zu denen die Variation der Höhe der Peaks und die Änderung der Anzahl der lokalen Optima durch Überlagerung einzelner Peaks gehören.

Weitere Hinweise zur Verwendung geeigneter Testfunktionen finden sich bei Schönemann (1999) und Kleefeld (2004). Erstere Arbeit stellt neben einzelnen existierenden Klassifikationsversuchen auch eine Übersicht der Anforderungen an Testfunktionen sowie ausgewählter Testsuiten vor. Die Arbeit von Kleefeld bezieht sich insbesondere auch auf den dynamischen Fall und gibt detaillierte Hinweise auf unterschiedliche Bewegungsarten.

Die für die vorliegende Untersuchung durchgeführte Auswahl geeigneter Testfunktionen resultiert im Wesentlichen aus den folgenden Überlegungen. Einerseits soll die Vergleichbarkeit der eigenen Ergebnisse mit denen anderer Autoren gewährleistet sein. Deshalb wird die Verwendung gängiger Testprobleme propagiert. Andererseits soll ein möglichst breites Spektrum häufig auftretender Problemklassen berücksichtigt werden. Dadurch soll das Verhalten der Algorithmen bei Auftreten unterschiedlicher Eigenschaften der Probleme ermittelt werden.

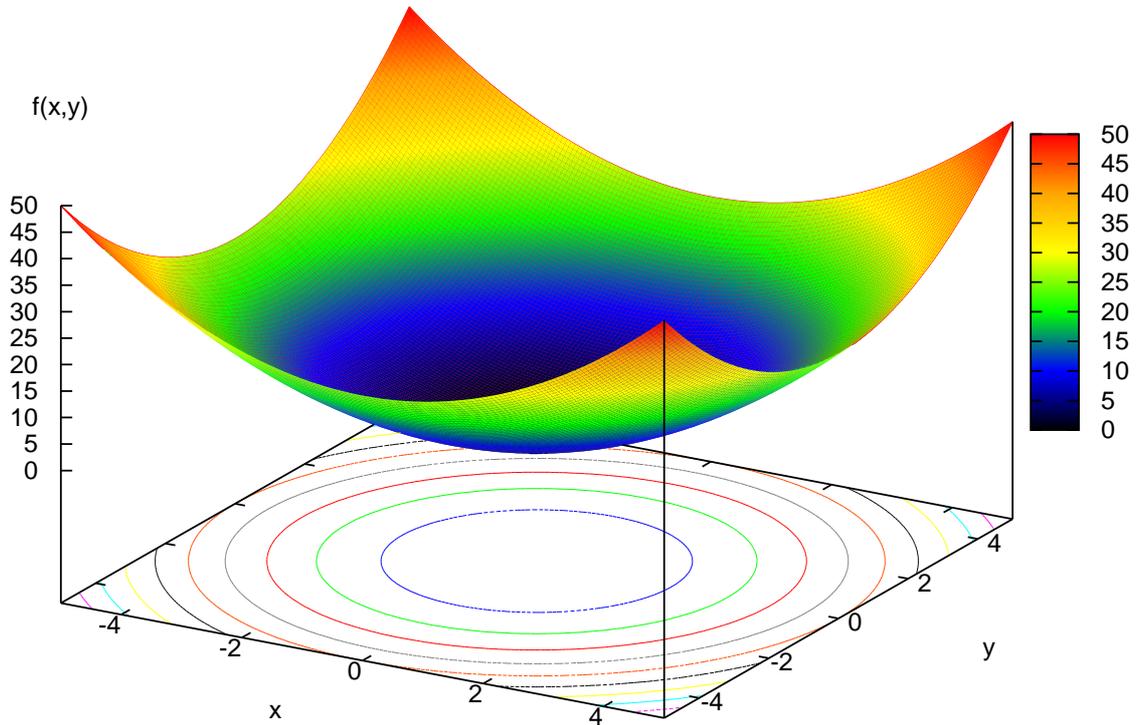
Die in dieser Arbeit erfolgte Beschränkung auf eine relativ kleine Anzahl von Testproblemen resultiert aus der Berücksichtigung eines praktischen Gesichtspunktes. In statischen Umgebungen zeigt ein Algorithmus oftmals ein relativ gleichmäßiges Verhalten während des gesamten Laufs. Nach einer Phase mit stetigen Verbesserungen gelangt er in ein lokales Optimum. Die Adaptation der Strategieparameter sorgt dann dafür, dass er dieses nur selten verlassen kann und stattdessen stagniert. Demgegenüber werden bei der Optimierung in dynamischen Umgebungen während eines Laufs häufig unterschiedliche Phasen beobachtet, weshalb eine Optimierung über eine längere Zeit durchgeführt werden muss. Aus demselben Grund müssen zur möglichst vollständigen Beurteilung eines Verfahrens für jede Generation sämtliche relevanten Zwischenergebnisse protokolliert werden. Des Weiteren können sich einzelne Läufe etwas stärker voneinander unterscheiden als im statischen Fall, weshalb für jeden einzelnen Lauf eine größere Zahl Generationen zu wählen ist. Beide Eigenschaften führen zu einem erhöhten Ressourcenverbrauch (Speicherbedarf und Rechenzeit).

5.1.1 Kugelmodell

Als Standard für nahezu alle Testsuiten im Bereich evolutionärer Algorithmen hat sich das Kugelmodell

$$f_{Kugel}(x) = \sum_{i=1}^n x_i^2$$

etabliert. Es hat für die Dimension $n = 2$ das in der Abbildung 5.1 dargestellte Aussehen und besitzt wie alle im Folgenden vorgestellten Funktionen das Minimum $f(x^*) = 0$, wobei $x^* = (0, \dots, 0)$. Das Kugelmodell spiegelt die Eigenschaften der meisten Optimierungsprobleme in unmittelbarer Umgebung des globalen Optimums wider. Im statischen Fall dient es zur Messung der Konvergenzgeschwindigkeit eines Verfahrens. Um vergleichbare Startbedingungen zu gewährleisten, aber dennoch eine Zufallskomponente zu integrieren, werden die Objektvariablen wie auch bei den anderen

Abbildung 5.1: Kugelmodell für die Dimension $n = 2$.

hier benutzten Funktionen zufällig innerhalb eines Intervalls initialisiert. Während des Laufs sind die Variablen dagegen unbeschränkt, was insbesondere zur Feststellung einer Divergenz notwendig ist. Für die Startpopulation im Kugelmodell wird in Anlehnung an Bäck (1996) eine gleichverteilte zufällige Initialisierung aller Objektvariablen aus dem Intervall $[-40, 60]^n$ gewählt.

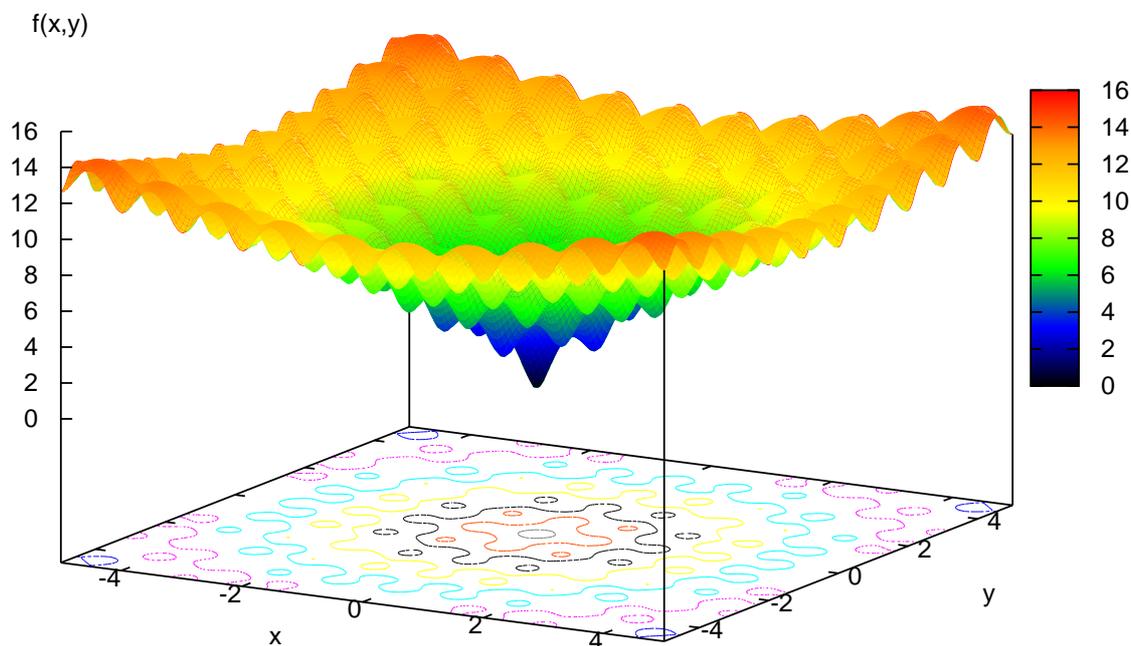
5.1.2 Ackley-Funktion

Die von Bäck (1996, S. 143) benutzte generalisierte Ackley-Funktion

$$f_{\text{Ackley}}(x) = -a \cdot \exp \left(-b \cdot \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2} \right) - \exp \left(\frac{1}{n} \sum_{i=1}^n \cos(c \cdot x_i) \right) + a + e$$

$$\text{mit } a = 20, b = 0.2, c = 2\pi,$$

besitzt eine moderate Multimodalität und sollte den meisten Algorithmen keine besonderen Schwierigkeiten bei der Optimierung bereiten. Die Abbildung 5.2 zeigt das Aussehen der Funktion für die Problemdimension $n = 2$. Die Initialisierung der Startpopulation erfolgt im Intervall $[-20, 30]$. Die Ackley-Funktion wird in diesen Testkatalog integriert, um das Verhalten EA auf leicht multimodalen Funktionen zu beobachten. Aufgrund der Problemstruktur sollten die Abweichungen zum Verhalten auf dem Kugelmodell relativ gering ausfallen.

Abbildung 5.2: Generalisierte Funktion von Ackley für $n = 2$.

5.1.3 Rastrigin-Funktion

Bäck und Hammel (1994) verwenden in ihren Experimenten folgende verallgemeinerte Funktion von Rastrigin

$$f_{\text{Rastrigin}}(x) = n \cdot A + \sum_{i=1}^n (x_i^2 - A \cdot \cos(\omega \cdot x_i))$$

$$\text{mit } A = 3, \omega = 2\pi,$$

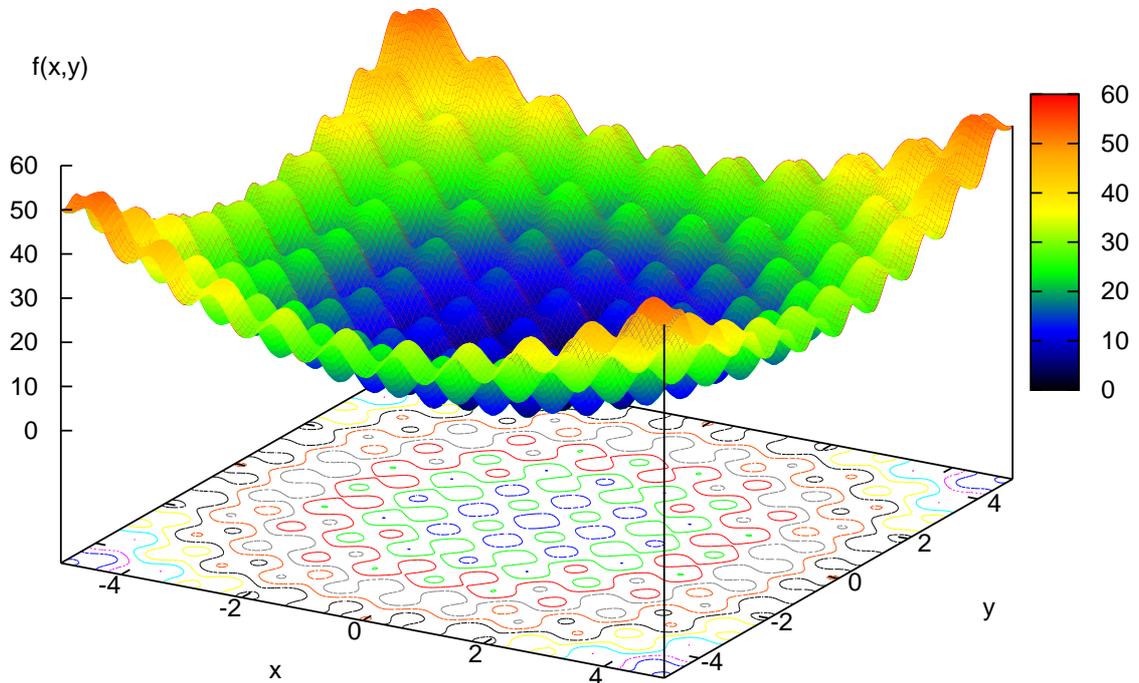
die für den zweidimensionalen Fall in der Abbildung 5.3 dargestellt ist. Wie bei den anderen beiden Funktionen erfolgt die Beschränkung der Objektvariablen nur für die Initialisierung. Hierfür gilt $-30 \leq x_i \leq 30$. Im Gegensatz zur Ackley-Funktion ist die Rastrigin-Funktion jedoch durch eine wesentlich größere Multimodalität geprägt und nach Bäck und Hammel dementsprechend schwieriger zu optimieren.

Zu beachten ist bei der Rastrigin-Funktion, dass sie ab einem bestimmten Schwellwert der Objektvariablen keine lokalen Optima mehr besitzt. Dies gilt, wenn alle Einzelterme

$$f_{x_i} = x_i^2 - A \cdot \cos(\omega \cdot x_i)$$

monoton wachsen oder fallen. Dies ist dann erfüllt, wenn die jeweiligen Ableitungen

$$f'_{x_i} = 2x_i + A\omega \cdot \sin(\omega \cdot x_i)$$

Abbildung 5.3: Generalisierte Funktion von Rastrigin für $n = 2$.

alle entweder immer größer als Null oder immer kleiner als Null sind. Da $\sin(\cdot) \in [-1, 1]$ ist das der Fall, wenn die Objektvariablen den gesamten Term dominieren, also wenn für alle $i \in \{1, \dots, n\}$ gilt (mit $A > 0, \omega > 0$):

$$2x_i > A\omega \Rightarrow x_i > \frac{A\omega}{2} \text{ oder}$$

$$2x_i < -A\omega \Rightarrow x_i < -\frac{A\omega}{2} .$$

Ab $|x_i| > A\omega/2 \forall i$ kann die Rastrigin-Funktion durch das Kugelmodell angenähert werden. Dennoch sollte die größere Multimodalität der Rastrigin-Funktion dazu führen, dass die Ergebnisse stärker von den auf den beiden anderen Testfunktionen erreichten Resultaten abweichen. Was eine günstige Wahl der Populationsgrößen anbelangt, wird hier interessant sein, ob es größere Abweichungen zwischen dem statischen und dem dynamischen Fall gibt.

Für zukünftige Untersuchungen sollten auch Funktionen integriert werden, die im statischen Fall sehr schwierig zu optimieren sind. Hierdurch könnte überprüft werden, ob es benennbare Problemeigenschaften gibt, die die Optimierung im dynamischen Fall erleichtern. Eine Eigenschaft solcher Problemstellungen könnte bspw. eine extrem multimodale Fitnesslandschaft sein, da aufgrund der angenommenen permanenten Dynamik eine Stagnation in einem lokalen Optimum verhindert werden sollte. Dass dieses Phänomen überhaupt auftritt, wurde bereits von Kleefeld (2004, S. 91) beschrieben.

5.1.4 Verwendete Bewegungsarten

Nach der Auswahl der zu benutzenden Testfunktionen müssen noch die zu verwendenden Bewegungsarten festgelegt werden. Wie von Kleefeld (2004) dargestellt, gibt es eine Vielzahl unterschiedlichster Bewegungsformen. Sieht man von solch komplexen Änderungen wie der Transformation von einer Funktion in eine andere ab, sind hier u.a. die Verschiebung oder Drehung in der Ebene denkbar.

Für eine erste Untersuchung beschränken wir uns in der vorliegenden Arbeit auf eine einfache Verschiebung der benutzten Funktionen. Trotz dieser Beschränkung besitzen wir im Folgenden weitere zahlreiche Freiheitsgrade. Neben der obligatorischen Bewegungsstärke¹ s ist es bspw. möglich, die Bewegung linear, zyklisch oder zufällig zu gestalten. Bereits im linearen Fall sind weitere Unterscheidungen nach der Bewegungsrichtung möglich.

Wir konzentrieren uns auf Bewegungsmuster, bei denen sich das Optimum in eine gleich bleibende Richtung bewegt. Genauer gesagt wird die gesamte Funktion verschoben, der Einfachheit halber sprechen wir jedoch nur davon, dass sich das Optimum bewegt. Folgende Bewegungstypen werden verwendet:

Typ I: Das Optimum bewegt sich in genau einer Dimension mit der Stärke s .

Typ II: Das Optimum bewegt sich in allen Dimensionen gleich mit der Gesamtstärke s .

Typ III: Das Optimum bewegt sich in allen Dimensionen mit der Gesamtstärke s , wobei die Bewegung in den Dimensionen unterschiedlich ist.

Die Bewegung erfolgt bei allen Varianten in festen Abständen Δg . Durch eine Wahl $\Delta g > 1$ können auch längere statische Phasen realisiert werden. Die Bewegungsstärke s bezeichnet die zurückgelegte Entfernung des Optimums bei jeder Bewegung. Die während der Optimierung über g Generationen insgesamt zurückgelegte Entfernung beträgt dann $s \cdot g / \Delta g$. Ungleichmäßige Bewegungen lassen sich dadurch realisieren, dass die Bewegungsstärke nicht mehr konstant gehalten wird, sondern nunmehr normalverteilt mit einer Standardabweichung d um s schwankt. In diesem Fall sorgt eine große Zahl von Generationen aber ebenfalls für eine ungefähre Gesamtentfernung von $s \cdot g / \Delta g$, sodass eine merkbare Beeinflussung der Ergebnisse, die aus unterschiedlichen zurückgelegten Gesamtentfernungen resultiert, ausgeschlossen werden kann.

Für alle drei Varianten soll das jeweilige Bewegungsmuster anhand des zweidimensionalen Falles grafisch verdeutlicht werden (Abbildungen 5.4 bis 5.6). Die Bewegung des Typs I erfolgt in nur einer (i.Allg. der ersten) Dimension, alle anderen Dimensionen bleiben konstant. Dagegen bewegt sich das Optimum bei den Typen II und III in allen Dimensionen. Während beim Typ II die Bewegung in allen Dimensionen gleich geschieht, differieren beim Bewegungstyp III die Entfernungen in den Dimensionen. Die Bewegung ist dabei so gestaltet, dass sich in jeder Koordinate die Entfernung erhöht. Konkret bewegt sich das Optimum in der ersten Koordinate um einen Betrag a , in der zweiten um einen Betrag $\sqrt{2} \cdot a$, in der dritten um einen Betrag $\sqrt{3} \cdot a$, ..., und in der n -ten Koordinate um einen Betrag $\sqrt{n} \cdot a$. Der Term a hängt dabei sowohl von der Bewegungsstärke s als auch der Problemdimension n ab und berechnet sich wie folgt:

$$\sqrt{\sum_{i=1}^n (\sqrt{i} \cdot a)^2} = \sqrt{a^2 \cdot \sum_{i=1}^n i} = a \cdot \sqrt{\frac{n^2 + n}{2}} \stackrel{!}{=} s$$

¹Die Bewegungsstärke ist hier als die pro Generation zurückgelegte Entfernung im \mathbb{R}^n definiert.

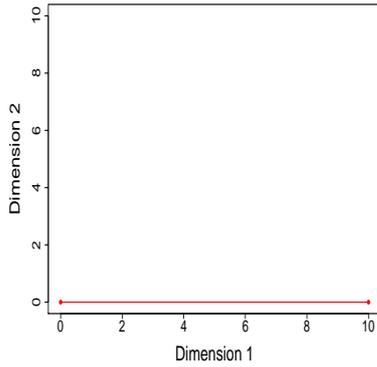


Abbildung 5.4: Typ I

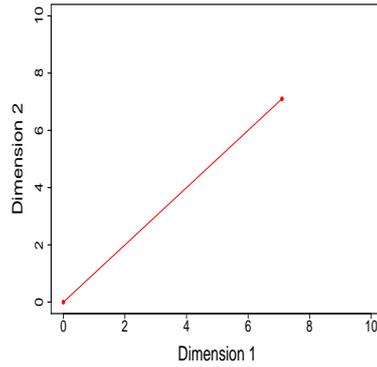


Abbildung 5.5: Typ II

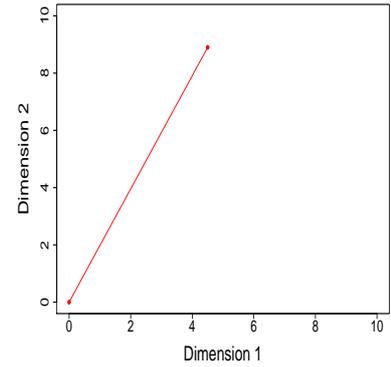


Abbildung 5.6: Typ III

Abbildungen 5.4 bis 5.6: Grafische Verdeutlichung der unterschiedlichen Bewegungsarten. In allen Fällen ist eine Bewegungsstärke von $s = 10$ angenommen.

$$\implies a = s \cdot \sqrt{\frac{2}{n^2 + n}} .$$

Zur Verdeutlichung sei eine 30-dimensionale Funktion mit einer Bewegungsstärke $s = 1$ angenommen. Daraus folgt $a = 1/\sqrt{465}$. Das Optimum bewegt sich in der ersten Koordinate um den Wert $\sqrt{1/465}$, in der zweiten um den Wert $\sqrt{2/465}$ usw. und in der letzten Koordinate um den Wert $\sqrt{30/465}$. Die zurückgelegte Entfernung pro Einzelschritt ergibt sich dadurch wie gewünscht zu

$$\sqrt{\sum_{i=1}^{30} \left(\sqrt{i/465}\right)^2} = \sqrt{\frac{1}{465} \sum_{i=1}^{30} i} = \sqrt{\frac{1}{465} \cdot 465} = 1 .$$

5.2 Vorgehensweise und EA-Parametrisierung

In den nachfolgenden Untersuchungen wird für jedes Optimierungsproblem zunächst eine hier als Standard-ES bezeichnete Evolutionsstrategie eingesetzt, deren Parametrisierung in Tabelle 5.1 angegeben ist. Anschließend werden einzelne Experimente gestartet, die in ausgesuchten Parametern

Tabelle 5.1: Standardparametrisierung einer (15, 100)-ES.

Parameter	Einstellung
μ	15
λ	100
n_σ	n
τ_i	$\frac{1}{\sqrt{2 \cdot \sqrt{n}}}$
τ_0	$\frac{1}{\sqrt{2 \cdot n}}$
$\sigma_i^{(0)}$	3
Rekombination	Objektvariable: paarweise diskret Strategieparameter: global-intermediär
Selektion	Komma

von dieser Standardeinstellung abweichen. Dies dient dazu, den möglichen Einfluss einzelner Aspekte zu isolieren und im Idealfall quantifizieren zu können. Die von der Standardparametrisierung ab-

weichenden Einstellungen eines Experiments sind jeweils deutlich bezeichnet, sodass die Resultate reproduzierbar sind.

Bei den vorliegenden Untersuchungen wird die Problemdimension immer über den Bereich $n = \{5, 10, 20, 30, 40, \dots, 200\}$ variiert. Zusätzliche Problemdimensionen werden berücksichtigt, wenn dies notwendig oder sinnvoll erscheint. Aus Gründen der Übersichtlichkeit finden sich in vielen Abbildungen nur die Darstellungen für wenige ausgewählte Problemdimensionen. In den meisten Fällen liegen die Kurven für die anderen n zwischen den Darstellungen der benachbarten Problemdimensionen. Bei erwähnenswerten Abweichungen wird dieser Umstand aufgezeigt.

Für die in der vorliegenden Untersuchung einen großen Rahmen einnehmende Betrachtung der durch unterschiedliche Populationsgrößen erreichbaren Ergebnisse werden in den meisten Experimenten die Populationsgrößen auf die Werte

$$\mu = \{2, 3, 5, 8, 10, 15, 20, 30, 40\} \text{ und}$$

$$\lambda = \{10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 100, 120, 140, 160, 180, 200\}$$

gesetzt. In besonders interessant erscheinenden Situationen werden für μ und λ weitere Werte benutzt. Bekannt ist, dass für das Funktionieren der (Selbst-)Adaptation der Selektionsdruck einen bestimmten Wert besitzen muss. Dem Autor ist keine Untersuchung bekannt, in der sich ein Selektionsdruck kleiner drei als optimal erwies. Deshalb werden hier nur Kombinationen der obigen Werte mit $\lambda \geq 3\mu$ verwendet.

Zur Erzielung einer guten statistischen Mittelung werden für alle Strategien mindestens 31 unabhängige Durchläufe verwendet. In den Experimentreihen mit unterschiedlichen Problemdimensionen n wird die Anzahl der Funktionsauswertungen in Abhängigkeit von n gewählt. Dieses Vorgehen soll dem höheren Optimierungsbedarf aufgrund der wachsenden Problemdimension Rechnung tragen. In den statischen Umgebungen beträgt die Anzahl der Funktionsauswertungen $2000n$. Soweit nichts anderes gesagt, wird in dynamischen Umgebungen die Anzahl auszuführender Funktionsauswertungen auf $10000n$ festgelegt. Da nur vollständige Generationen (bspw. $\lfloor 2000n/\lambda \rfloor$) ausgeführt werden, kann sich die echte Anzahl von Funktionsauswertungen abhängig von der gewählten Anzahl Nachkommen geringfügig unterscheiden. Dies ist insbesondere beim Vergleich zweier EA-Varianten bei gleicher Problemdimension zu berücksichtigen, hat jedoch, was hier vorweggenommen werden soll, keinen spürbaren Einfluss auf die Ergebnisse.

In dynamischen Umgebungen wird in den Experimentreihen mit einer unterschiedlichen Anzahl Nachkommen die zurückgelegte Gesamtentfernung identisch gewählt. Dies bedeutet z.B. dass bei einer Halbierung der Anzahl der Nachkommen die pro Generation zurückgelegte Entfernung ebenfalls halbiert wird. In den Experimenten mit verschiedenen Bewegungsfrequenzen Δg kann die pro Generation zurückgelegte Entfernung dagegen identisch gewählt werden, wenn sich bei einer Halbierung der Nachkommen die Bewegungsfrequenz verdoppelt.

Darüber hinaus kann es bei der Verwendung einer variablen Bewegungsstärke wegen der stochastischen Schwankungen ebenfalls zu abweichenden zurückgelegten Gesamtentfernungen kommen. In den Experimenten ist die Anzahl der Generationen jedoch so groß gewählt, dass die Unterschiede kleiner als 2% sind und sich nicht wesentlich auf den erreichten ABFV auswirken.

Die Initialisierung der Startpopulation erfolgt im statischen Fall zufällig gleichverteilt über dem gesamten Wertebereich. Im dynamischen Fall wird die Startpopulation immer im oder nahe dem Optimum initialisiert, wobei die Strategieparameter ebenfalls nahezu optimal gewählt werden. Dennoch wird dem Algorithmus ein gewisser Zeitrahmen eingeräumt, in dem eine mögliche Adaptation der Strategieparameter erfolgen kann.

Kapitel 6

Geeignete Wahl der Anzahl der Schrittweiten

6.1 Einführung

In statischen Umgebungen hängt die optimale Anzahl n_σ der zu verwendenden Schrittweiten stark von der zu optimierenden Funktion und dem gewählten Startpunkt ab. So erweist es sich als vorteilhaft, bei symmetrischen Funktionen und einer in allen Achsen gleichen Entfernung zum Optimum mit nur einer Schrittweite zu arbeiten. Dagegen führt eine solche Strategie bei stark unsymmetrischen Funktionen zu schlechteren Ergebnissen.

In diesem Kapitel werden nun einige Untersuchungen durchgeführt, die zeigen sollen, wie viel versprechend eine unterschiedliche Wahl der Anzahl der Schrittweiten im dynamischen Fall ist. Um die Anzahl der durchzuführenden Experimente zu beschränken, werden zunächst nur die beiden Extremfälle $n_\sigma = 1$ und $n_\sigma = n$ betrachtet. Die Hoffnung dabei ist, dass aufgrund der erzielten Ergebnisse auch qualitative Aussagen für die dazwischen liegenden Varianten $1 < n_\sigma < n$ gemacht werden können. Zur Verifizierung der gemachten Vermutungen erfolgen ausgewählte Experimente mit einer alternativen Anzahl Schrittweiten.

6.2 Experimentelle Untersuchungen

Als Zielfunktionen kommen die im Kapitel 5 vorgestellten drei Testfunktionen für mehrere Problemdimensionen zum Einsatz. Zur Gestaltung der Dynamik werden die im gleichen Kapitel definierten drei Bewegungstypen mit verschiedenen Stärken s eingesetzt.

Zunächst verwenden wir das Kugelmodell, bei dem sich das Optimum in nur einer Richtung bewegt. Die Verwendung einer ES mit nur einer Schrittweite ist hier unangebracht, da in der Koordinate der Bewegung eine Schrittweite $\sigma = s$ optimal ist, in den anderen Dimensionen aber ein Wert nahe Null. Zur Verifizierung unserer Vermutung und Abschätzung der quantitativen Unterschiede müssen wir die Ergebnisse beider Strategien vergleichen. Die Abbildung 6.1 zeigt für $n = 10$ die Werte des ABFV in Abhängigkeit von der Bewegungsstärke s sowohl für eine ES mit einer Schrittweite als auch mit n Schrittweiten. Während eine lineare Skalierung der y -Achse keine deutlichen Hinweise auf eine Beziehung zulässt, zeigt die hier dargestellte Grafik einen eindeutigen Zusammenhang. Der erreichte ABFV hängt in dem hier gewählten Bereich quadratisch von der Bewegungsstärke s ab. Die auf dem reellwertigen Suchraum arbeitende ES mit einer Schrittweite ist also auch bei einer größeren zurückgelegten Entfernung in der Lage, dem Optimum zu folgen – auch

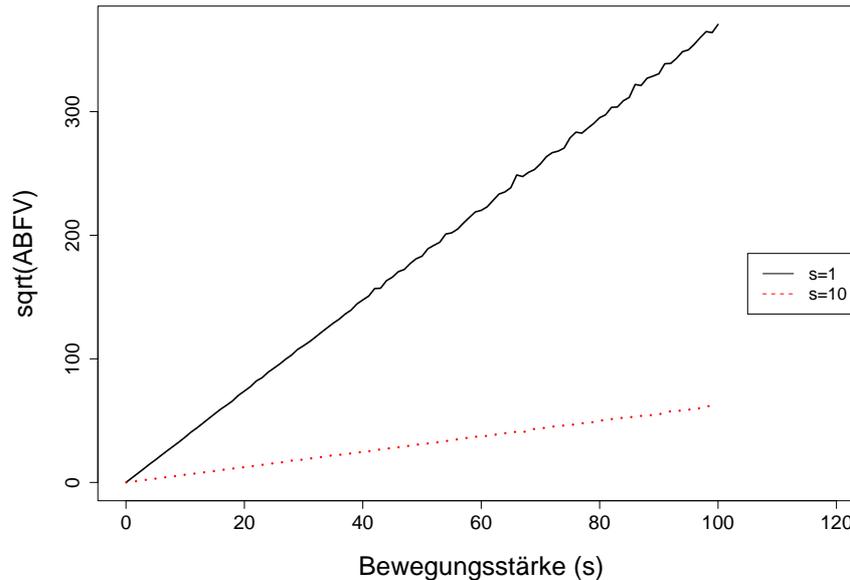


Abbildung 6.1: Abhängigkeit des erreichten ABFV von der Bewegungsstärke für eine (15, 100)-ES mit einer Schrittweite und n Schrittweiten auf dem 10-dimensionalen dynamischen Kugelmodell bei einer Bewegung in einer Dimension (Typ I).

wenn nur mit einer in der Bewegungsstärke quadratisch zunehmenden Distanz. Diese Eigenschaft steht im Gegensatz zu Beobachtungen mit einem auf einer OneMax-Problemstellung eingesetzten binären EA, wo eine Bewegungsstärke zu verzeichnen war, ab der der EA dem Optimum nicht mehr folgen konnte (Droste 2003). Wie die Werte einzuschätzen sind, zeigt der Vergleich mit einer ES mit mehr als einer Schrittweite. Zu erwarten ist, dass eine solche ES besser in der Lage ist, sich an den Bewegungstyp anzupassen, da die unterschiedlichen Schrittweiten für die einzelnen Dimensionen ein dichteres Verfolgen des Optimums gestatten sollten. Diese Vermutung wird auch durch die entsprechenden Ergebnisse bestätigt. Unter den gewählten Testbedingungen macht sich also die Verwendung einer ES mit n Schrittweiten bezahlt.

Aus den in der obigen Abbildung zu erkennenden Schwankungen ist darüber hinaus ersichtlich, dass für eine genauere Bestimmung des ABFV in diesen Fällen weitere Läufe notwendig wären. Daneben ist eine Extrapolation der Resultate auf eine Bewegung mit extrem großen Sprüngen pro Generation an dieser Stelle nicht zulässig. Dies ist aber auch eher von untergeordnetem Interesse, da in solchen Situationen eine angemessene Optimierung kaum möglich ist.

Im nächsten Experiment wollen wir untersuchen, ob die ES mit n Schrittweiten gegenüber der ES mit nur einer Schrittweite benachteiligt ist, wenn sich das Optimum in allen Dimensionen bewegt, die zurückgelegte Entfernung aber in allen Dimensionen identisch ist (Bewegungstyp II). Eine erste Vermutung könnte dies bejahen, weil die Adaptation nur einer Schrittweite leichter gelingen sollte als die gleichzeitige Adaptation von n Schrittweiten. Die Abbildung 6.2 zeigt jedoch, dass zwischen den Ergebnissen beider Varianten kaum Unterschiede festzustellen sind.

Dieses zunächst vielleicht etwas überraschende Ergebnis besitzt eine einfache Erklärung (Schönmann 2003). Wie im Abschnitt 7.2 durch weitere Experimente unterlegt, kann dieses Verhalten damit begründet werden, dass aufgrund der verwendeten Schrittweitensteuerung (siehe Abschnitt 2.3 ab Seite 15) größere prozentuale Schwankungen in den einzelnen Schrittweiten auftreten. Diese sind absolut sehr klein, wenn die Schrittweiten einen Wert nahe Null besitzen. Für größere Werte fallen die absoluten Schwankungen in der Regel jedoch stärker aus. Somit kann es selbst bei den hier

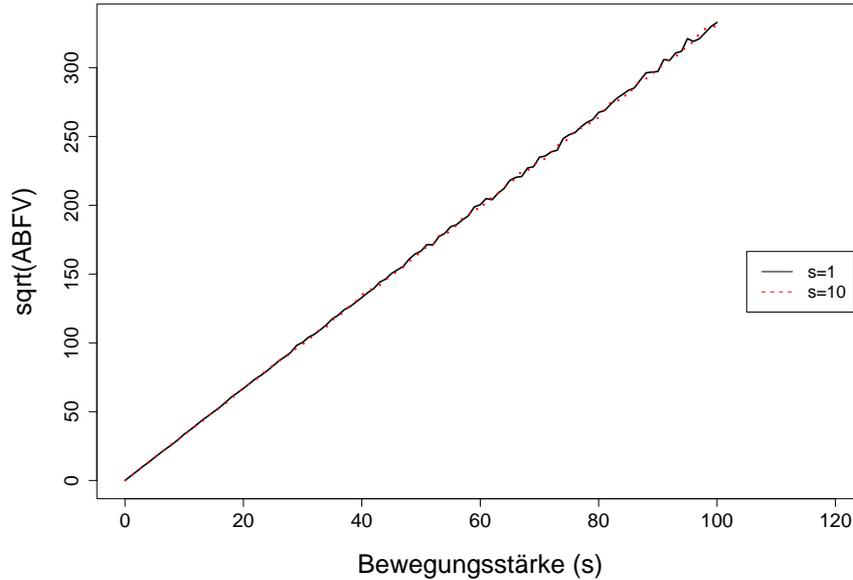


Abbildung 6.2: Abhängigkeit des erreichten ABFV von der Bewegungsstärke für eine (15, 100)-ES mit einer Schrittweite und mit n Schrittweiten auf dem 10-dimensionalen dynamischen Kugelmodell bei einer gleichmäßigen Bewegung in allen Dimensionen (Typ II).

verwendeten 30 Dimensionen dann immer einmal vorkommen, dass die zur Bildung der einen oder anderen Schrittweite gezogene Zufallszahl stark von Null abweicht, sodass die durch Differenzen in den n Schrittweiten hervorgerufenen Schwankungen überlagert werden. Falls diese Vermutung allgemein gültig ist, sollte für größeres n eine höhere Wahrscheinlichkeit existieren, dass sich unter den neuen Schrittweiten eines Individuums zumindest ein Ausreißer befindet, der den Funktionswert stark verschlechtert. Bevor wir auf diesen Punkt bei der ausführlichen Erörterung der Ergebnisse im dynamischen Kugelmodell zurückkommen (Abschnitt 7.2) soll eine theoretische Überlegung diesen Sachverhalt verdeutlichen.

Sind die Schrittweiten erst einmal richtig eingestellt, so ist eine Veränderung der Schrittweiten um einen zu großen Faktor in den wenigsten Fällen sinnvoll. Dennoch wird z.B. eine Verdoppelung häufig auftreten. Denn:

$$\begin{aligned}\sigma_{\text{neu}} &= \sigma_{\text{alt}} \cdot \exp(\tau_g \cdot N_0(0, 1) + \tau_l \cdot N_i(0, 1)) \stackrel{!}{=} 2 \cdot \sigma_{\text{alt}} \\ &\Rightarrow \exp(\tau_g \cdot N_0(0, 1) + \tau_l \cdot N_i(0, 1)) = 2 \\ &\Rightarrow \tau_g \cdot N_0(0, 1) + \tau_l \cdot N_i(0, 1) = 0.69 \ .\end{aligned}$$

Unter der Annahme, dass sich beide Terme um den gleichen Faktor ($N(0, 1) := N_0(0, 1) = N_i(0, 1)$) erhöhen, folgt

$$N(0, 1) = 0.69 / (\tau_g + \tau_l) \ .$$

Tabelle 6.1 zeigt für ausgewählte Problemdimensionen die notwendigen Werte der gezogenen Zufallszahlen, um eine Verdoppelung zu erhalten: Allgemein beträgt die Wahrscheinlichkeit einer Verdoppelung also P_N^2 , weil sich beide Zufallszahlen ändern müssen. Da bei n Schrittweiten jedes Individuum n Komponenten besitzt, beträgt für nicht zu kleines n (≥ 10) unter diesen Annahmen die Wahrscheinlichkeit grob abgeschätzt $n \cdot P_N^2$, dass sich eine solche Änderung in mindestens einer Komponente ergibt. Ein solcher Fall tritt also für $n = 10$ ungefähr bei jedem 6. Nachkommen auf.

Tabelle 6.1: Die für eine Verdoppelung der Schrittweite notwendigen Zufallszahlen und die Wahrscheinlichkeit der Realisierung eines Wertes größer oder gleich dieser Zufallszahl.

n	τ_g	τ_l	$N(0, 1)$	P_N
10	0.22	0.40	1.11	$1.3 \cdot 10^{-1}$
30	0.13	0.30	1.60	$5.5 \cdot 10^{-2}$
50	0.10	0.27	1.89	$3.0 \cdot 10^{-2}$
100	0.07	0.22	2.34	$9.5 \cdot 10^{-3}$

Für $n = 100$ ergibt sich ein solches Ereignis ungefähr für alle ≈ 111 Nachkommen. Mit zunehmender Problemdimension verringert sich zwar die Wahrscheinlichkeit einer solchen Zunahme. Trotzdem tritt selbst für $n = 100$ ein solcher Fall fast einmal pro Generation auf (bei $\lambda = 100$ Nachkommen).

Der Vollständigkeit halber sollen die obigen Experimente auch für den Bewegungstyp III durchgeführt werden. In der Abbildung 6.3 sind die entsprechenden Ergebnisse dargestellt. Trotz der

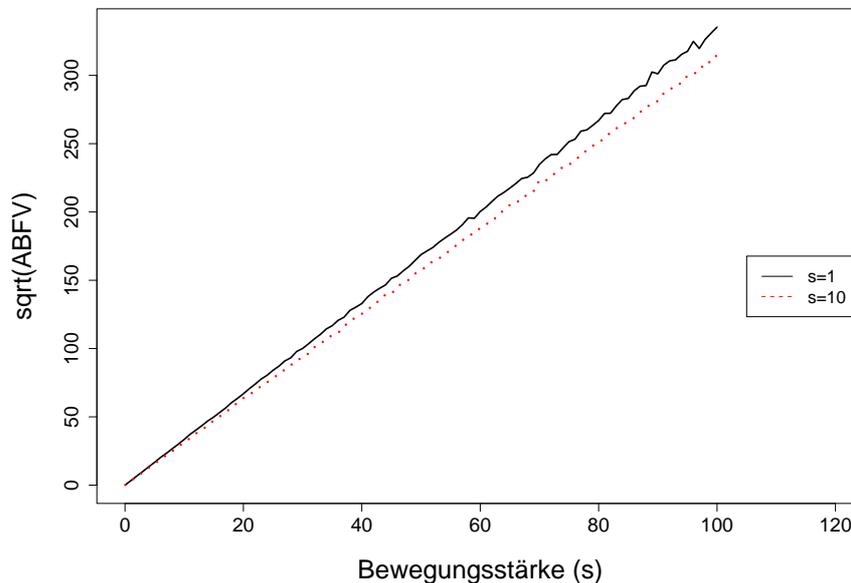


Abbildung 6.3: Abhängigkeit des erreichten ABFV von der Bewegungsstärke für eine (15, 100)-ES mit einer Schrittweite und mit n Schrittweiten auf dem 10-dimensionalen dynamischen Kugelmodell bei einer Bewegung in einer schrägen Diagonalen über allen Dimensionen (Typ III).

unterschiedlichen Bewegungsgeschwindigkeiten in den einzelnen Komponenten besitzt die ES mit n Schrittweiten keine wesentlichen Vorteile gegenüber der Variante mit nur einer Schrittweite.

Die hier auf dem Kugelmodell mit $n = 10$ gezeigten Ergebnisse werden auch für $n = 20$ und $n = 30$ in ähnlicher Form bestätigt. Allerdings zeigt die Abbildung 6.4, dass die zunehmende Anzahl einzustellender Parameter die Optimierung schwieriger gestaltet, denn die Ergebnisse der ES mit n Schrittweiten fallen nun etwas schlechter aus als bei einer Schrittweite.

Für das Kugelmodell halten wir fest, dass eine ES mit n Schrittweiten abhängig vom durchgeführten Versuch entweder wesentlich besser oder aber nur unwesentlich schlechter als eine ES mit nur einer Schrittweite ist. Darüber hinaus führen weitere Experimente auf der Ackley- sowie der Rastrigin-Funktion zu ähnlichen Ergebnissen. Exemplarisch werden deshalb in den Abbildungen 6.5 und 6.6 nur die Ergebnisse für die Rastrigin-Funktion mit $n = 30$ und den Bewegungstypen II

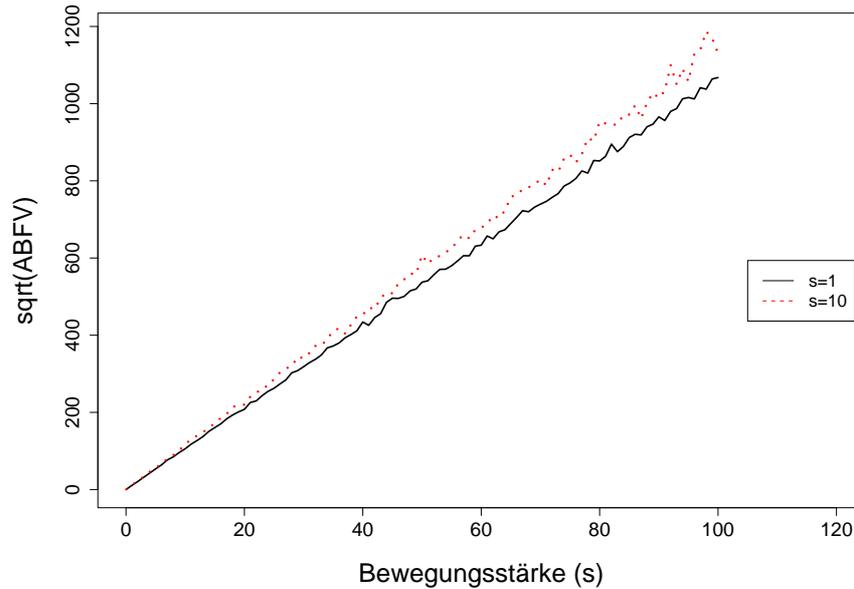


Abbildung 6.4: Abhängigkeit des erreichten ABFV von der Bewegungsstärke für eine (15, 100)-ES mit einer Schrittweite und mit n Schrittweiten auf dem 30-dimensionalen dynamischen Kugelmodell bei einer Bewegung in einer schrägen Diagonalen über allen Dimensionen (Typ III).

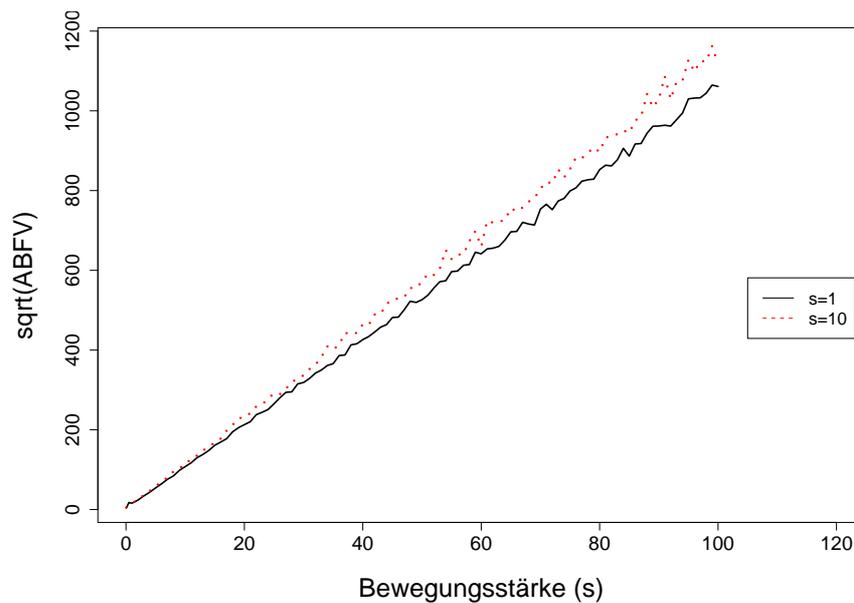


Abbildung 6.5: Abhängigkeit des erreichten ABFV von der Bewegungsstärke für eine (15, 100)-ES mit einer Schrittweite und mit n Schrittweiten auf der 30-dimensionalen dynamischen Rastrigin-Funktion bei einer gleichmäßigen Bewegung in allen Dimensionen (Typ II).

und III dargestellt. Für den Bewegungstyp II sind die Ergebnisse der ES mit n Schrittweiten etwas schlechter als die einer ES mit nur einer Schrittweite. Da die Rastrigin-Funktion stark multimodal ist, gestaltet sich die Anpassung vieler Schrittweiten schwieriger als bei der Kugelfunktion. Mithin führt die Verwendung einer zunehmenden Anzahl einzustellender Parameter dazu, dass die Ergebnisse schlechter ausfallen. Interessant ist darüber hinaus auch die Tatsache, dass die absoluten Werte des ABFV im Gegensatz zum statischen Fall nun nahezu identisch zum 30-dimensionalen Kugelmo-

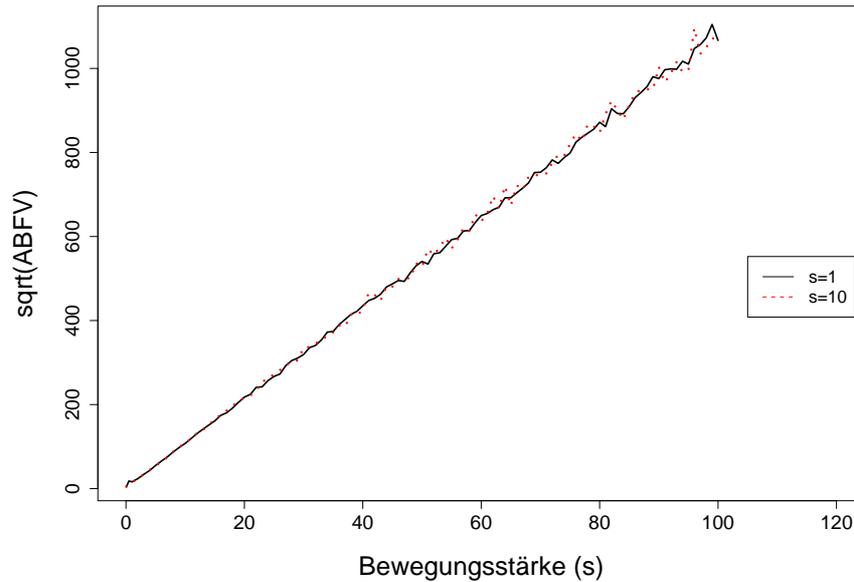


Abbildung 6.6: Abhängigkeit des erreichten ABFV von der Bewegungsstärke für eine (15, 100)-ES mit einer Schrittweite und mit n Schrittweiten auf der 30-dimensionalen dynamischen Rastrigin-Funktion bei einer Bewegung in einer schrägen Diagonalen über allen Dimensionen (Typ III).

dell sind. Während für $|x_i| > A\omega/2$ bereits oben gezeigt wurde, dass die Rastrigin-Funktion stark dem Kugelmodell ähnelt, wird in dem anderen Definitionsbereich die durch die Multimodalität hervorgerufene erhöhte Problemschwierigkeit in dieser Situation durch die Dynamik dominiert. Dieses Verhalten ist auch nachvollziehbar, da die ES bei ausreichender Dynamik nicht so leicht in einem lokalen Optimum stecken bleiben kann.

Bei einer Bewegung des Typs III stimmen die Ergebnisse beider ES-Varianten ziemlich genau überein. Das heißt wiederum, dass die ES auch bei der Rastrigin-Funktion mit $n = 30$ den Vorteil individueller Schrittweiten gegenüber einer ES mit einer Schrittweite nicht in bessere Ergebnisse ummünzen kann. Allerdings sind hier die Ergebnisse beider Varianten nahezu gleich, währenddessen im Kugelmodell die ES mit einer Schrittweite leichte Vorteile besaß.

Obwohl es auf der Ackley-Funktion gewisse Unterschiede zu den Ergebnissen auf den anderen beiden hier verwendeten Testfunktionen gibt, wird auf eine ausführliche Darstellung der dort erzielten Ergebnisse verzichtet. Es sei lediglich angemerkt, dass sich bei der Ackley-Funktion ein vergleichbares Bild ergibt. Auch dort ist die Verwendung einer ES mit n Schrittweiten gegenüber einer ES mit einer Schrittweite unter den gegebenen Versuchsbedingungen in einigen Fällen bevorzugt und in den anderen nicht oder nur wenig benachteiligt. Allerdings hat die Ackley-Funktion den Nachteil, dass sich die Funktionswerte ab einer gewissen Distanz zum Optimum im Rahmen der Rechengenauigkeit nicht mehr unterscheiden und somit für die ES keine Ansatzmöglichkeit für Verbesserungen bieten.

Eine differenzierte Festlegung der Anzahl der Schrittweiten könnte in einzelnen Situationen zu besseren Ergebnissen führen. Hierzu die folgenden Überlegungen. Wird eine ES mit $1 < n_\sigma < n$ Schrittweiten verwendet, muss eine Zuordnung der Schrittweiten zu den Koordinaten erfolgen. Wird bei zwei Schrittweiten eine explizit für eine Koordinate verwendet und die zweite für die restlichen Koordinaten gibt es n Möglichkeiten, die einzelne Schrittweite einer der n Koordinaten zuzuordnen. Bei einer entsprechenden Verwendung von drei Schrittweiten gibt es dagegen $\binom{n}{2} = n \cdot (n - 1)/2$ alternative Belegungen. Für den allgemeinen Fall müsste die ES also sowohl die Anzahl der Schritt-

weiten als auch die Zuordnung zu den Koordinaten selbstständig lernen. Weiß man um die Schwierigkeit, optimale Schrittweiten einzustellen, so gestaltet sich dieser Lernprozess umso aufwändiger. Die nächsten Experimente sollen einen kleinen Eindruck der Auswirkungen verschiedener Zuordnungen der Schrittweiten zu den Koordinaten liefern.

Im folgenden Experiment bewegt sich das Optimum immer in der ersten Koordinate (Bewegungstyp I). Die Zuordnung der Schrittweiten zu den einzelnen Koordinaten sieht so aus, dass die ersten $n_\sigma - 1$ Koordinaten eine eigene Schrittweite erhalten und sich die restlichen $n - n_\sigma + 1$ Koordinaten eine Schrittweite teilen. Die Abbildung 6.7 zeigt die entsprechenden Ergebnisse. Wie

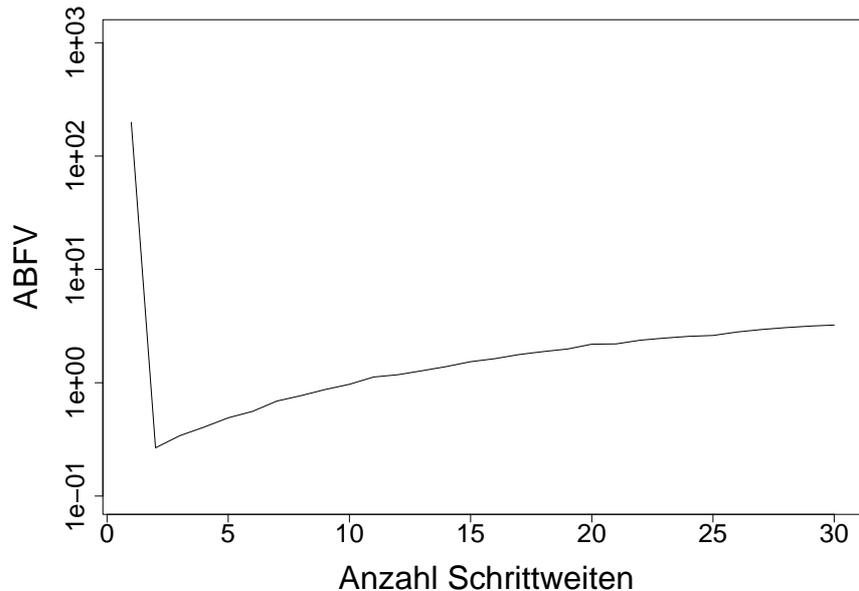


Abbildung 6.7: Abhängigkeit des erreichten ABFV einer (15, 100)-ES von der Anzahl der Schrittweiten auf dem 30-dimensionalen dynamischen Kugelmodell bei einer Bewegung in einer Dimension (Typ I). (Zur näheren Erklärung siehe Text.)

erwartet sind die Ergebnisse mit zwei Schrittweiten die besten. Mit zunehmender Anzahl Schrittweiten verschlechtern sich die Ergebnisse zwar, allerdings sind sie selbst für n Schrittweiten noch relativ gemäßigt. Die schlechtesten Ergebnisse erreicht eine ES mit nur einer Schrittweite.

In der nächsten Testreihe bewegt sich das Optimum weiterhin in der ersten Koordinate. Allerdings werden die Schrittweiten zu den Koordinaten anders zugeordnet als bisher. Die Zuordnung beginnt nun bei der letzten Koordinate. Bei n_σ Schrittweiten besitzen die letzten $n_\sigma - 1$ Koordinaten jeweils eine eigene Schrittweite, die ersten $n - n_\sigma + 1$ Koordinaten teilen sich dagegen eine Schrittweite. Eine ES mit nur zwei Schrittweiten ist mit der hier getroffenen Zuordnung also wenig optimal. Die Abbildung 6.8 bestätigt diese Annahme. Mit zunehmender Anzahl Schrittweiten verbessert sich das Ergebnis zusehends. Erst bei einer maximalen Anzahl Schrittweiten, wenn also die sich verändernde Koordinate eine eigene Schrittweite erhält, ist der Funktionswert am besten. Auffallend ist aber ein weiteres Resultat. Das Maximum der Kurve wird nicht für $n_\sigma = 1$ angenommen, sondern für $n_\sigma = 8$. Hier gibt es offensichtlich einen Trade-off. Zwar wird die für die sich bewegende Koordinate benutzte Schrittweite nur noch für $n - n_\sigma + 1 = 23$ Koordinaten verwendet, aber der erhöhte Aufwand für acht weitere gegenüber weniger Schrittweiten dominiert den erreichbaren Funktionswert. Mit einer weiter zunehmenden Anzahl Schrittweiten macht sich jedoch der Vorteil bemerkbar, dass die für die sich verändernde Koordinate verwendete Schrittweite nur für

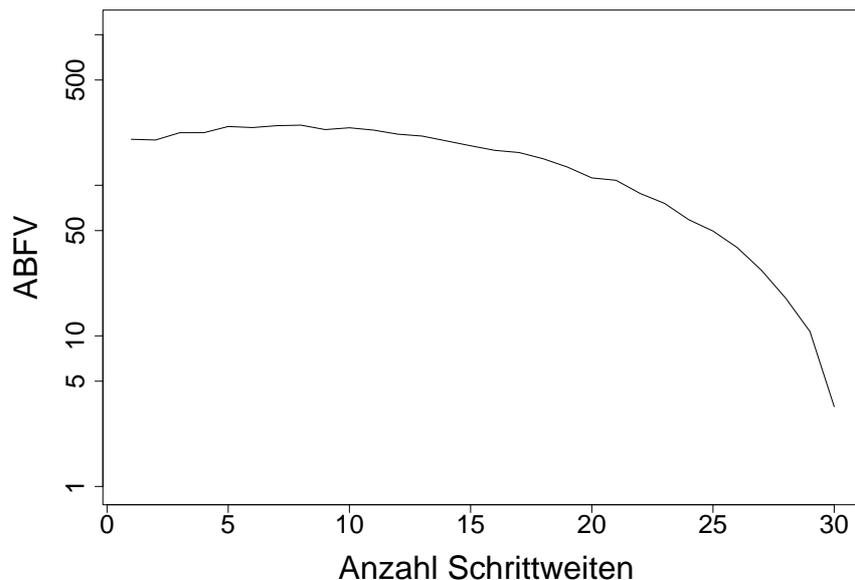


Abbildung 6.8: Abhängigkeit des erreichten ABFV einer (15, 100)-ES von der Anzahl der Schrittweiten auf dem 30-dimensionalen dynamischen Kugelmodell bei einer Bewegung in einer Dimension (Typ I). (Zur näheren Erklärung siehe Text.)

wenige weitere Koordinaten benutzt wird.

6.3 Fazit

Die grundsätzliche Empfehlung aus den obigen Tests lautet nun, jeder sich verändernden Koordinate eine eigene Schrittweite zuzuordnen. Da die vorliegende Bewegungsform inklusive der Anzahl der sich bewegenden Koordinaten sowie deren Bewegungsstärke in der Regel zu Beginn nicht bekannt sein dürfte, sollte eine ES mit maximaler Anzahl Schrittweiten gewählt werden, da die hiermit erreichbaren Ergebnisse im Durchschnitt am besten sind. Eine Anpassung der Anzahl der Schrittweiten während eines Optimierungslaufs erscheint in dynamischen Umgebungen hingegen wenig viel versprechend. Die Lerndauer bis zur Einstellung der richtigen Anzahl Schrittweiten und der richtigen Zuordnung zu den Koordinaten dürfte zumindest sehr lang sein, wenn nicht gar unmöglich. Als vorteilhaft ist ein solcher Mechanismus dagegen vielleicht in statischen Umgebungen oder in Situationen, in denen sich das Optimum selten, aber dafür in größeren Schritten bewegt. Soll also eine ES mit mutativer Schrittweitenanpassung verwendet werden, die für das in dieser Arbeit im Mittelpunkt stehende Untersuchungsgebiet ein möglichst breites Einsatzspektrum besitzt und dabei im Allgemeinen „gute“ Resultate liefert, ist aufgrund der hier durchgeführten Versuche die Verwendung einer ES mit n Schrittweiten empfehlenswert. Aufgrund dieser Überlegungen wird in den Experimenten der nächsten Kapitel deshalb eine ES mit n Schrittweiten verwendet. Während für $|x_i| > A\omega/2$ bereits im Abschnitt 5.1.3 gezeigt wurde, dass die Rastrigin-Funktion stark dem Kugelmodell ähnelt, haben wir in diesem Abschnitt gesehen, dass in dem anderen Definitionsbereich die durch die Multimodalität bedingte erhöhte Problemschwierigkeit durch die Dynamik dominiert wird.

Kapitel 7

Ergebnisse auf dem Kugelmodell

Zunächst wird anhand des statischen Kugelmodells der Einfluss unterschiedlicher Parametrisierungen beobachtet. Anschließend werden diese Einstellungen auch im dynamischen Kugelmodell verwendet, um den direkten Vergleich zu erhalten. Daran schließen sich Untersuchungen der Auswirkungen spezieller Einstellungen für μ und λ auf das Konvergenzverhalten der ES bei unterschiedlichen Problemdimensionen n an. Aufgrund des teils exemplarischen Charakters der Ergebnisse des Kugelmodells werden diese Ergebnisse etwas detaillierter dargestellt. Da sich die Resultate auf den anderen Funktionen zum Teil sehr ähneln, kann dort auf eine umfangreiche Darstellung verzichtet werden.

Zur Darstellung der Ergebnisse werden wir insbesondere auf die im Kapitel 4 vorgestellten grafischen Verfahren zurückgreifen. Das Auftragen der gleichen Ergebnisse mit unterschiedlichen Achsenparametern wird uns dabei einen Einblick aus unterschiedlichen Blickwinkeln gestatten.

7.1 Statische Umgebung

In einer Experimentreihe untersucht Kursawe (1999) u.a. die Auswirkungen unterschiedlicher Rekombinationstypen auf das Konvergenzverhalten evolutionärer Algorithmen. Bei der Untersuchung des Kugelmodells stieß er auf das Phänomen, dass der übliche Rekombinationstyp (diskrete Rekombination der Objektvariablen und intermediäre Rekombination der Strategieparameter aller Eltern) bei einer Problemdimension von $n = 30$ zwar die bekannt ausreichenden Resultate liefert. Bei einer Problemdimension von $n = 100$ stellte er aber mit diesem Rekombinationstyp eine Divergenz fest, währenddessen ein anderer Rekombinationstyp bei sonst gleichen Bedingungen zu einer Konvergenz führte (Kursawe 1999, S. 53). Neben dem Rekombinationstyp und vielen anderen Interpretationsmöglichkeiten drängt sich aber auch die Frage auf, ob bei einem Übergang von $n = 30$ nach $n = 100$ eine ES mit identischen Populationsgrößen und daraus resultierendem gleichen Selektionsdruck noch angemessen ist.

Ein einfaches Experiment soll erste Ansätze zur Beantwortung dieser Frage liefern. Abbildung 7.1 zeigt das Konvergenzverhalten einer (15, 100)-Standard-ES auf dem statischen Kugelmodell in Abhängigkeit von der Problemdimension n . Deutlich zu erkennen ist der von Kursawe (1999, S. 50) genannte Effekt der Divergenz für große Problemdimensionen n . Während für kleine n eine Verbesserung gegenüber dem Startwert zu verzeichnen ist, erfolgt in dem Bereich $100 \leq n \leq 120$ der Übergang zur Divergenz. Aufgrund der Versuchsanordnung ist es nicht möglich, den genauen Wert für n anzugeben, ab dem die Divergenz beginnt. Dies liegt u.a. darin begründet, dass dazu einerseits mehr Einzelläufe pro Parametereinstellung durchgeführt werden müssten, andererseits

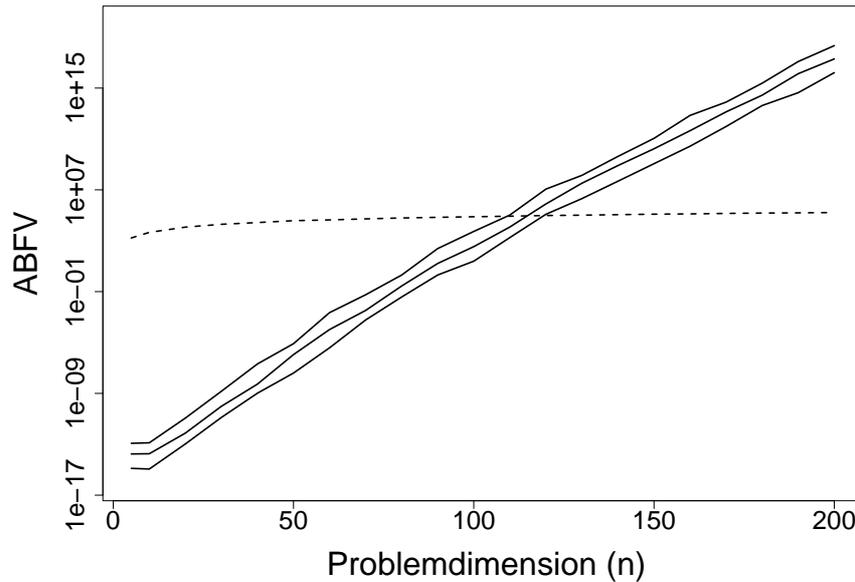


Abbildung 7.1: Konvergenzverhalten einer $(15,100)$ -ES auf dem statischen Kugelmodell in Abhängigkeit von der Problemdimension n . Neben dem ABFV sind noch die 0.05- und 0.95-Quantile des ABFV dargestellt. Die gestrichelte Linie zeigt darüber hinaus den durchschnittlichen Startwert der ES. Die Initialisierung der Startpopulation erfolgte gemäß Abschnitt 5.1.1.

die Einzelläufe über mehr Funktionsauswertungen ausgeführt werden müssten, was jedoch in dieser Studie nicht möglich ist.

Zur besseren Einschätzung des beobachteten Verhaltens betrachten wir die einzelnen Durchläufe für ausgewählte Problemdimensionen etwas genauer. Die Abbildung 7.2 zeigt den Verlauf der 31 unabhängigen Durchläufe der $(15,100)$ -ES für drei Problemdimensionen. Während sich die Verläufe bei geringer und hoher Dimension meist ähneln, sind die Verläufe für $n = 110$ sehr unterschiedlich, wobei keine einheitlichen Aussagen über eine Konvergenz oder Divergenz gemacht werden können. Eine genauere Analyse zeigt, dass sich nicht nur die verschiedenen Läufe stark voneinander unterscheiden, sondern dass auch einzelne Läufe zunächst gegen das Optimum konvergieren, anschließend aber divergieren. Auch der umgekehrte Fall ist zu beobachten. Die Betrachtung der Schrittweiten zeigt, dass diese in diesen Situationen sehr labil sind und ständig wechseln. Eine Interpretation ist, dass zur Erzielung bzw. Aufrechterhaltung günstiger Schrittweiten eine gewisse Diversität in der Population vorhanden sein muss. Andernfalls können aufgrund auftretender stochastischer Schwankungen schlecht eingestellte Schrittweiten die Gesamtpopulation dominieren. Stochastische Schwankungen werden dabei mit wachsender Problemdimension zunehmen, weshalb in diesen Fällen eine größere Population notwendig sein wird. Darüber hinaus muss auch die Anzahl der Nachkommen steigen, um genügend gute Individuen für die nächste Elternpopulation zu erzeugen. Denn auch bei einer großen Elternpopulation werden vereinzelte ungünstige Strategieparameterinstellungen für Nachkommen mit schlechter Fitness sorgen.

Interessant ist, dass sich trotz der für $n = 110$ in der Abb. 7.2 zu beobachtenden Unregelmäßigkeiten die in der Abb. 7.1 dargestellten Quantile auch für einen Bereich um $n \approx 110$ in einem gleichbleibend engen Band um den ABFV legen. Unter diesem Gesichtspunkt weisen die einzelnen Läufe also ein ähnliches Verhalten auf.

Um einen von einer ganz speziellen Parametrisierung abhängenden Effekt auszuschließen, folgen zunächst Experimente, die bis auf einzelne Änderungen identische Einstellungen besitzen. Als

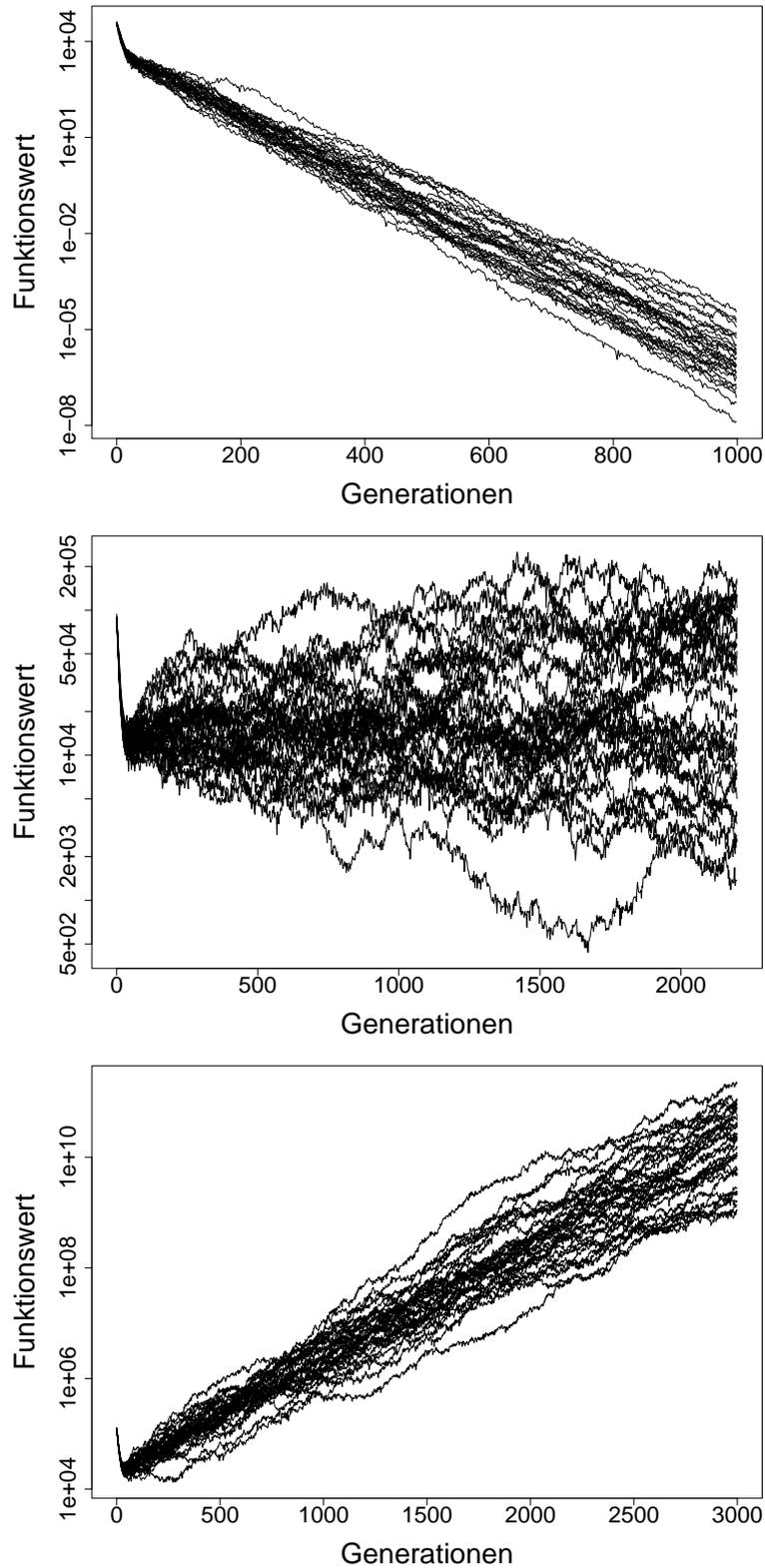


Abbildung 7.2: Alle 31 unabhängigen Durchläufe einer (15, 100)-ES auf dem statischen Kugelmodell für $n = 50$ (oben), $n = 110$ (Mitte) und $n = 150$ (unten).

erstes beobachten wir den Verlauf der (15, 100)-ES auf dem n -dimensionalen Kugelmodell bei einer Initialisierung der Startpopulation in einem Punkt weit entfernt vom Optimum. Dabei sollte

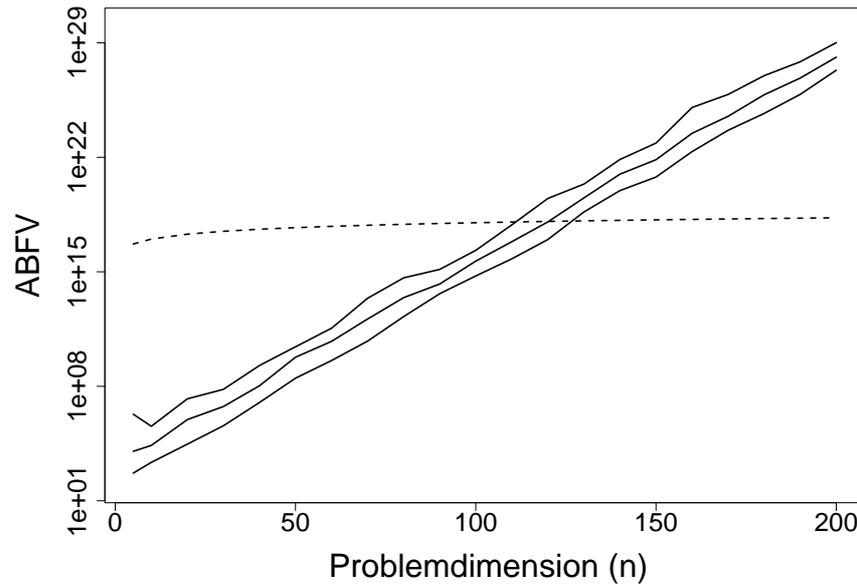


Abbildung 7.3: Konvergenzverhalten einer (15, 100)-ES bei Initialisierung der Startpopulation mit $x_i = 10^8$ und $\sigma_i = 10^4$ auf dem statischen Kugelmodell in Abhängigkeit von der Problemdimension n . Neben dem ABFV sind noch die 0.05- und 0.95-Quantile des ABFV dargestellt. Die gestrichelte Linie zeigt darüber hinaus den durchschnittlichen Startwert der ES.

sich ein vergleichbarer Konvergenzverlauf ergeben. Die Abbildung 7.3 bestätigt diese Vermutung. Auch die jeweiligen Einzelläufe entsprechen den in der Abbildung 7.2 dargestellten Verläufen mit Ausnahme der erreichten Funktionswerte, die wegen des weiter vom Optimum entfernt liegenden Startpunkts natürlich schlechter sind. In Verbindung mit weiteren hier nicht dargestellten Experimenten kann ein lediglich von der Initialisierung der Objektvariablen x_i und der Strategieparameter σ_i abhängender Effekt ausgeschlossen werden.

Bereits im Abschnitt 5.2 wurde angedeutet, dass Rechenberg (1994, S.161) erwähnt, dass wegen der geringen Trefferwahrscheinlichkeit im hochdimensionalen Kugelmodell eine Reduzierung der Mutationsschrittweiten helfen sollte. Ob die selbstadaptive (15, 100)-ES auch für große Problemdimensionen eine passende Einstellung der Mutationsschrittweiten bewahren kann, soll mit dem folgenden Experiment beantwortet werden. Dazu kehren wir wieder zur ursprünglichen Initialisierung der Objektvariablen zurück und setzen die Anfangsschrittweiten auf den Wert 0.001. Die Abbildung 7.4 zeigt das entsprechende Konvergenzverhalten in Abhängigkeit von der Problemdimension n . In der Abbildung 7.5 sind für die beiden alternativen Initialisierungen die jeweiligen Läufe für $n = 110$ dargestellt. Bei der Initialisierung mit einer kleinen Mutationsschrittweite ist zu erkennen, dass aufgrund der kleineren Anfangseinstellungen der Mutationsschrittweiten für längere Zeit ein Fortschritt erzielt werden kann (im Vergleich zur Abbildung 7.2). Jedoch setzt auch hier nach einiger Zeit die Phase der großen Schwankungen ein.

Detailliertere Beobachtungen der Mutationsschrittweiten bestätigen die Vermutung, dass bei großen Problemdimensionen die verwendete ES nicht in der Lage ist, die für eine Konvergenz geeigneten Einstellungen der Mutationsschrittweiten dauerhaft zu bewahren. Zunächst beginnen einzelne Mutationsschrittweiten sich stark zu erhöhen, sodass bald kein Fortschritt mehr erzielt wird. Nachfolgend erhöhen sich auch die Mutationsschrittweiten der anderen Komponenten, sodass sich die ES im Weiteren immer mehr vom Optimum entfernt. Dieses Ergebnis deutet auf gewisse Schwierigkeiten der in dieser Art verwendeten selbstadaptiven Mutationsschrittweitenanpassung

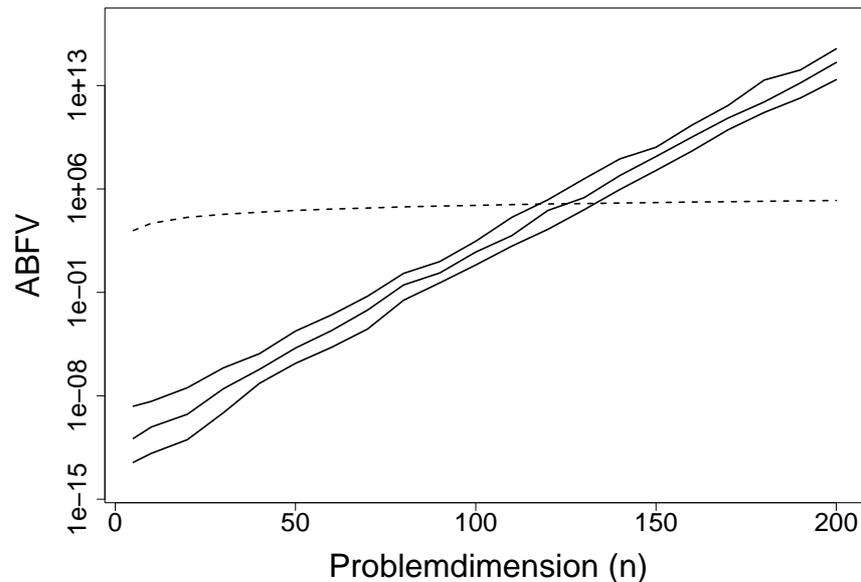


Abbildung 7.4: Konvergenzverhalten einer (15,100)-ES bei Initialisierung der Schrittweiten mit $\sigma_i = 0.001$ auf dem statischen Kugelmodell in Abhängigkeit von der Problemdimension n . Neben dem ABFV sind noch die 0.05- und 0.95-Quantile des ABFV dargestellt. Die gestrichelte Linie zeigt darüber hinaus den durchschnittlichen Startwert der ES.

hin. Ob die Probleme auch mit einer anderen Mutationsschrittweitensteuerung behoben werden können, bedarf weiterer eingehender Untersuchungen, ist aber nicht Gegenstand der vorliegenden Arbeit. Erste Experimente mit einer geometrischen Mittelung der Schrittweiten (siehe Bäck, Fogel und Michalewicz, S. C3.3:13) verhindern zwar die Divergenz für größere Problemdimensionen, zeigen aber den Nachteil, dass sie in den anderen Situationen keine mit der üblichen Schrittweitanpassung vergleichbaren guten Ergebnisse erreichen.

Experimentelle Untersuchungen von Kursawe (1999) zeigen, dass eine differenzierte Einstellung der Lernraten τ und τ' vorteilhaft ist. Darüber hinaus haben theoretische Analysen von Beyer (2001, S. 303) gezeigt, dass die optimalen Lernraten auch von den Populationsgrößen abhängen. Können die optimalen Lernraten für einzelne Funktionen noch theoretisch hergeleitet werden, sind wir auf eine heuristische Wahl angewiesen, wenn die zu optimierende Funktion unbekannt ist und aus einer größeren Funktionsklasse stammt. Weitere hier vorgenommene Experimente betreffen deshalb eine alternative Einstellung der Lernraten τ_0, τ_i . In verschiedenen Experimenten zeigte sich dabei, dass eine alternative Einstellung zu prinzipiell ähnlichen Resultaten führt, wobei die genaue Einstellung aber den Bereich beeinflusst, ab dem eine Divergenz zu verzeichnen ist. Exemplarisch seien hier die Ergebnisse mit den um den Faktor 3 erhöhten τ -Werten

$$\tau_0 = \frac{3}{\sqrt{2 \cdot n}}, \quad \tau_i = \frac{3}{\sqrt{2 \cdot \sqrt{n}}}$$

dargestellt. Für alle σ_i wurde wieder zum Startwert 3.0 zurückgekehrt und die einzelnen Komponenten der Objektvariablen wurden wiederum gleichverteilt aus dem Intervall $-40 \leq x_i \leq 60$ gezogen.

Aus der Abbildung 7.6 ist erkennbar, dass auch in diesem Fall eine Verbesserung gegenüber dem Startwert mit zunehmender Problemdimension immer schwieriger wird. Zunächst ähnelt der Verlauf der Kurve dem aus der Abbildung 7.1 bekannten Verlauf. Aufgrund der höheren Lernraten

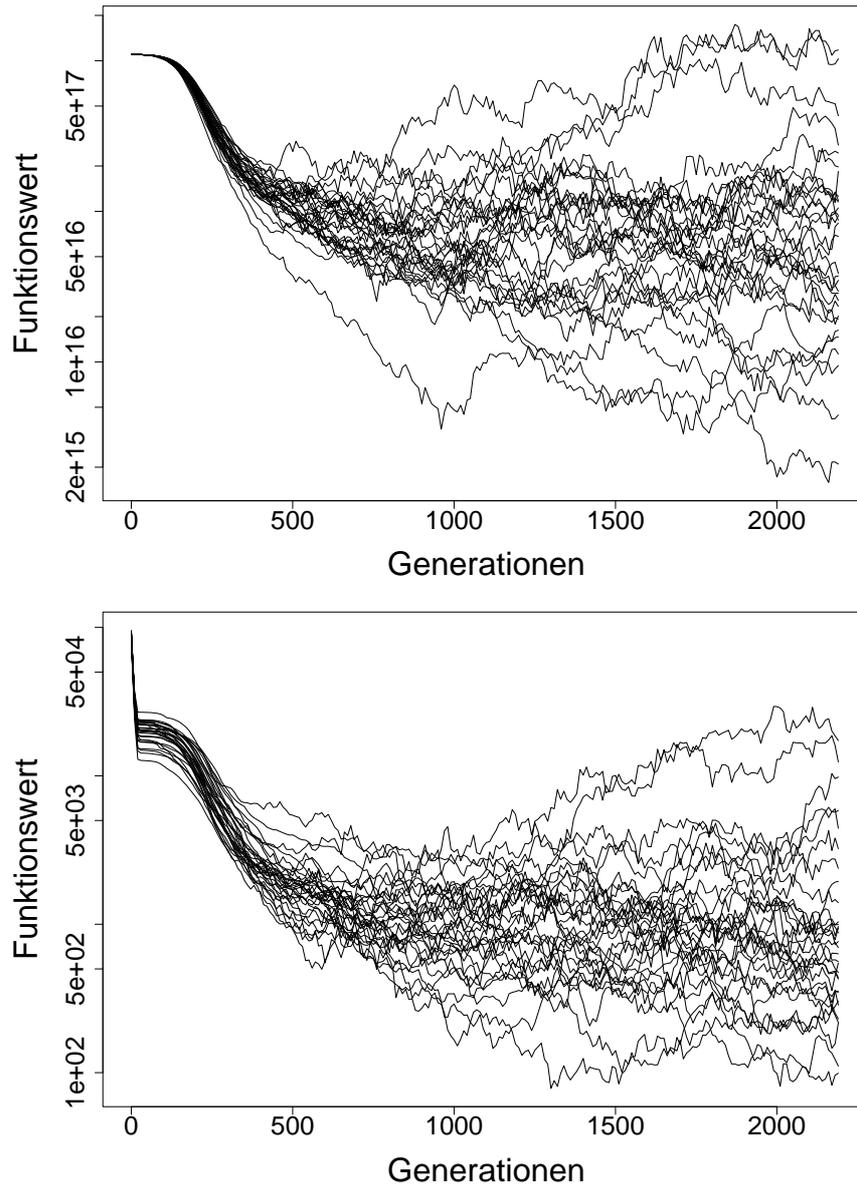


Abbildung 7.5: Darstellung der 31 unabhängigen Durchläufe einer (15,100)-ES auf dem 110-dimensionalen Kugelmodell bei Initialisierung mit $x_i = 10^8$, $\sigma_i = 10^4$ (oben) und $x_i \in [-40, 60]$, $\sigma_i = 0.001$ (unten).

verändern sich die Schrittweiten jedoch stärker. Dies führt zum einen dazu, dass sich die ES für kleine n wesentlich schneller dem Optimum nähert. Zum anderen verschiebt sich der Übergang zur Divergenz jetzt in den Bereich ab der Problemdimension $n > 70$. Die ES wird also bereits für kleinere n anfälliger gegenüber Schwankungen in den für die Schrittweitenberechnungen gezogenen Zufallszahlen. Darüber hinaus sind die erreichten Funktionswerte für große n viel schlechter, da sich die ES deutlich schneller vom Optimum entfernt. Ohne hier auf die genauen Resultate einzugehen sei betont, dass eine Reduzierung der Lernraten dazu führt, den Bereich der beginnenden Divergenz nach hinten zu verschieben. Allerdings mit dem negativen Nebeneffekt, dass die für kleine Problemdimensionen erreichten Funktionswerte schlechter sind. Daneben erscheint eine zu kleine Wahl der Schrittweiten insbesondere in den in dieser Arbeit im Mittelpunkt stehenden dynamischen

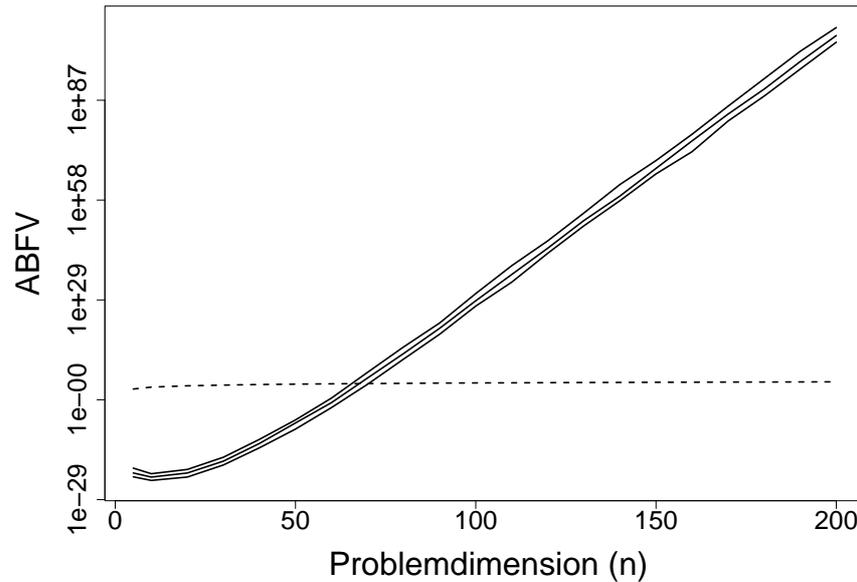


Abbildung 7.6: Konvergenzverhalten einer (15,100)-ES auf dem statischen Kugelmodell in Abhängigkeit von der Problemdimension n . Die Lernraten wurden gegenüber den anderen Experimentreihen mit einem Proportionalitätsfaktor von 3.0 versehen ($\tau_0 = \frac{3}{\sqrt{2 \cdot n}}$ und $\tau_i = \frac{3}{\sqrt{2 \cdot \sqrt{n}}}$). Neben dem ABFV sind noch die 0.05- und 0.95-Quantile des ABFV dargestellt. Die gestrichelte Linie zeigt darüber hinaus den durchschnittlichen Startwert der ES.

Umgebungen wenig sinnvoll, ist dort doch eine schnelle Anpassung an eine sich verändernde Umgebung notwendig. Insgesamt zeigen die Untersuchungen mit verschiedenen Lernraten, dass einfache Regeln für eine optimale Wahl der Lernraten für die meisten Testfunktionen noch ausstehen. Sehr vereinfacht formuliert ist als Ergebnis aus den obigen Untersuchungen zu folgern, dass die Lernraten für kleine Problemdimensionen eher größer und für große Problemdimensionen eher kleiner als üblich zu wählen sind. Darüber hinaus ist für zukünftige Untersuchungen zu überdenken, ob zu Vergleichszwecken nicht auf andere Schrittweitensteuerungen zurückgegriffen werden kann.

Wir wenden uns nun einer ersten Abschätzung des Einflusses der Populationsgrößen auf das Konvergenzverhalten der ES zu. Aufgrund der vorherigen Experimente dürfen wir dabei hoffen, auf eine Standardeinstellung der sonstigen Parameter zurückgreifen zu können, ohne die allgemeine Gültigkeit der Aussagen zu stark zu reduzieren. Den Einfluss unterschiedlicher Populationsgrößen hat bereits Kursawe (1999, S. 53) beobachtet. Für eine bestimmte Problemstellung stellt er fest, dass auf dem 100-dimensionalen Kugelmodell eine (10,100)-ES im Gegensatz zu einer (15,100)-ES die Divergenz verhindern kann. Wir möchten diese Aussagen nun etwas verallgemeinern und wiederholen deshalb das obige Grundexperiment mit einer (10,100)-ES.

Die Abbildung 7.7 zeigt das Konvergenzverhalten einer (10,100)-ES im statischen Kugelmodell in Abhängigkeit von der Problemdimension n . Die sonstigen Einstellungen sind identisch zu den Versuchen aus Abbildung 7.1. Die mit dieser Strategie erhaltenen Werte sind besser als die mit der (15,100)-ES erzielten Werte. Daneben hat sich die zuvor für $n > 100$ beobachtete Divergenz nun in den Bereich um $n \approx 160$ verschoben. Die sehr unterschiedlichen ABFV-Werte für die beiden Problemdimensionen zeigt exemplarisch die Abbildung 7.8, in der die zu beiden Strategien für $n = 50$ erhaltenen Werte in Boxplots aufgetragen sind. Die beiden Boxplots liegen weit auseinander. Sogar einschließlich der extremsten Ausreißer überschneiden sich die beiden Graphen nicht. Insgesamt

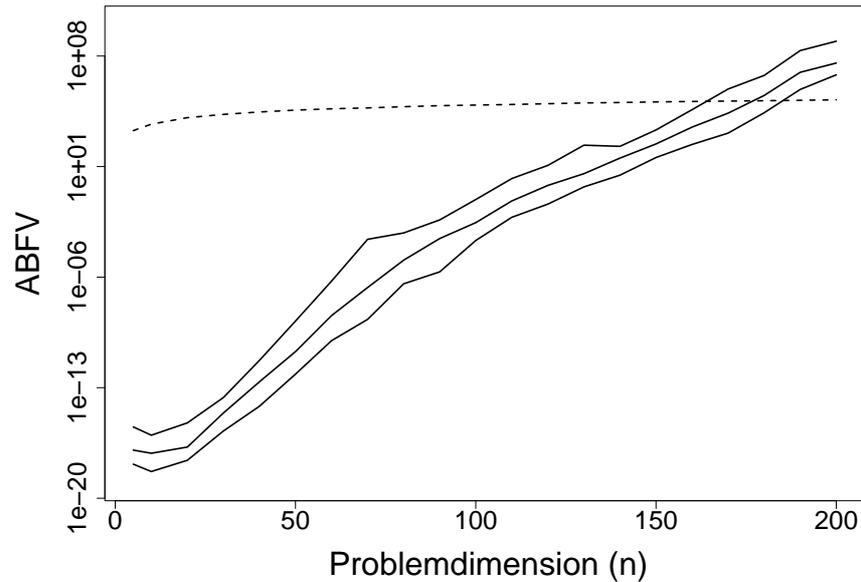


Abbildung 7.7: Konvergenzverhalten einer (10,100)-ES auf dem statischen Kugelmodell in Abhängigkeit von der Problemdimension n . Neben dem ABFV sind noch die 0.05- und 0.95-Quantile des ABFV dargestellt. Die gestrichelte Linie zeigt darüber hinaus den durchschnittlichen Startwert der ES.

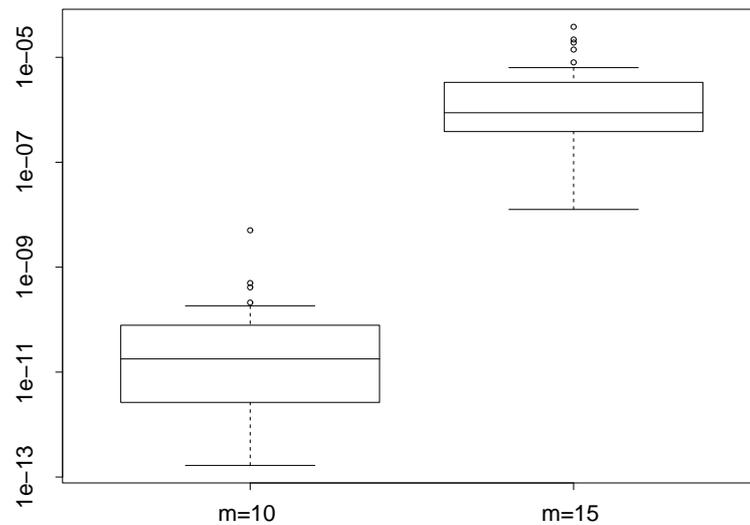


Abbildung 7.8: Boxplots der mit einer (10,100)- und (15,100)-ES auf dem 50-dimensionalen statischen Kugelmodell erreichten Funktionswerte in 31 unabhängigen Durchläufen.

sind die Ergebnisse für $\mu = 10$ über den gesamten hier getesteten Bereich n auf dem 5%-Niveau signifikant besser.

Alles in allem zeigen die bisherigen Experimente, dass das grundsätzliche Verhalten unabhängig von der gewählten Initialisierung der Startpopulation der ES ist. Zwar hat die Initialisierung einen nicht unerheblichen Einfluss auf die in einer vorgegebenen Zeit erreichbaren Ergebnisse. Allerdings kann dadurch nicht beeinflusst werden, ob eine Konvergenz oder Divergenz stattfindet. Demgegenüber führt eine unterschiedliche Einstellung anderer Strategieparameter zu einem wesentlich differenzierteren Verhalten der ES. Dies gilt zum einen für die Lernraten. Dabei sei aber nochmals

betont, dass eine geeignete Wahl von τ und τ' zwar zu besseren Ergebnissen führt, eine solche Wahl in praktischen Anwendungen aber unbekannt sein dürfte. Hier nicht weiter untersucht, aber zu vermuten ist, dass die Beobachtungen von Kursawe (1999, S. 53) bestätigt werden könnten, dass die Wahl des Rekombinationstyps einen häufig nicht zu unterschätzenden Einfluss auf das Verhalten des EA besitzt. Zum anderen hat aber auch die Wahl der in dieser Arbeit im Vordergrund stehenden Populationsgrößen einen stärkeren Einfluss auf das Konvergenzverhalten der ES. Die weiteren Ausführungen sollen zu qualitativen Aussagen bzgl. der Auswirkungen der unterschiedlichen Einstellungen der Populationsgrößen führen.

Zur Abschätzung, wie sich für eine feste Wahl μ ein veränderter Selektionsdruck λ/μ auswirkt, wird zunächst für verschiedene Problemdimensionen n und unterschiedliche Parameter λ der ABFV beobachtet. Bis auf die unterschiedliche Anzahl λ der Nachkommen handelt es sich um eine Standard-ES mit $\mu = 15$ Eltern. Nochmals sei betont, da nur vollständige Generationen ausgeführt werden, dass die Anzahl der Generationen $g = \lfloor \frac{2000n}{\lambda} \rfloor$ beträgt. Die Abbildung 7.9 zeigt für die verwendeten Parameterkombinationen die Abhängigkeit des ABFV von der Problemdimension n und der Anzahl der Nachkommen.

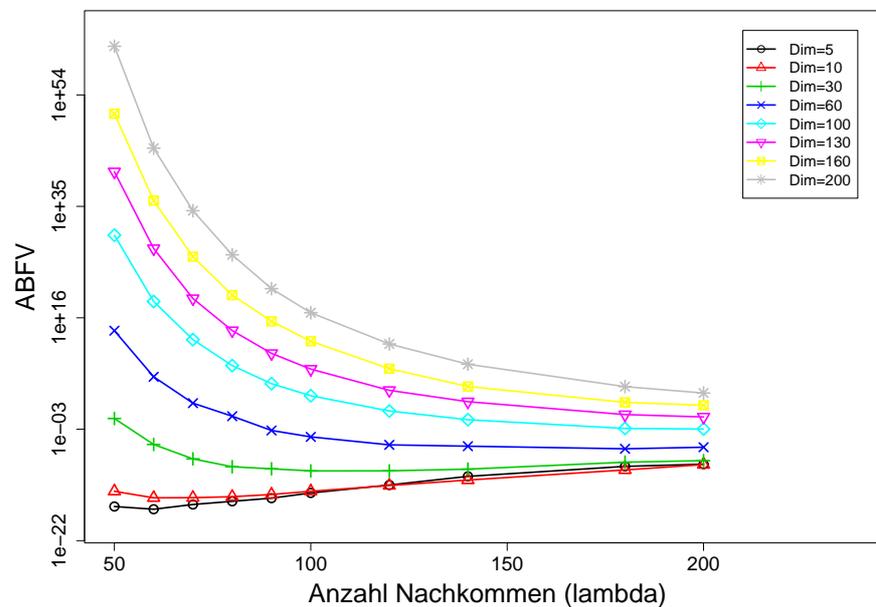


Abbildung 7.9: ABFV einer $(15, \lambda)$ -ES im statischen Kugelmodell in Abhängigkeit von der Problemdimension n und der Anzahl Nachkommen.

Deutlich zu erkennen ist, dass es für alle Problemdimensionen Bereiche für die Anzahl der Nachkommen gibt, in denen der ABFV gute Werte annimmt. Diese Bereiche sind für kleine Problemdimensionen schmal, was darauf hindeutet, dass es dort unter den Versuchsbedingungen eindeutige optimale Einstellungen für λ gibt. Für höhere Problemdimensionen wird dagegen der Bereich ähnlich guter Einstellungen für λ immer breiter. Hier unterscheiden sich die erzielten ABFV nicht so stark voneinander. Dennoch ist anzunehmen, dass auch hier mit einer bestimmten Anzahl an Nachkommen der ABFV minimiert werden kann.

Aus der Abbildung kann auch abgelesen werden, dass die optimale Anzahl der Nachkommen mit zunehmender Problemdimension immer größer wird. Während für $n = 5$ noch mit ca. 50 Nachkommen die besten Ergebnisse erzielt werden, steigt diese Anzahl für $n = 10$ bereits auf ca. 70 und für $n = 30$ auf ca. 120. Eine weitere Erhöhung von n lässt diesen Wert immer weiter an-

steigen. Eine genaue Spezifizierung der jeweiligen optimalen Anzahl Nachkommen erfordert jedoch weitaus mehr Läufe über eine längere Zeit. Vor allem aber müssten die Experimente für jedes $\lambda \in \{\lambda_{min}, \dots, \lambda_{max}\}$ durchgeführt werden.

Weitere Experimente mit anderen festen Einstellungen für die Anzahl der Eltern und variabler Anzahl der Nachkommen führen zu vergleichbaren Resultaten. D.h. für eine feste Problemdimension und eine feste Anzahl Eltern zeigt die Kurve bei der gewählten Skalierung eine einer Parabel ähnliche Abhängigkeit des ABFV von der Anzahl der Nachkommen. Insbesondere gibt es eine Einstellung der Nachkommen, die den ABFV minimiert.

Eine ähnliche Beobachtung lässt sich feststellen, wenn wir die Anzahl der Nachkommen fest wählen und stattdessen die Anzahl der Eltern variieren. Für $\lambda = 100$ wird für verschiedene Problemdimensionen n und unterschiedliche Parameter μ der ABFV beobachtet. Die Abbildung 7.10 zeigt für die verwendeten Parameterkombinationen die Abhängigkeit des ABFV von der Problem-

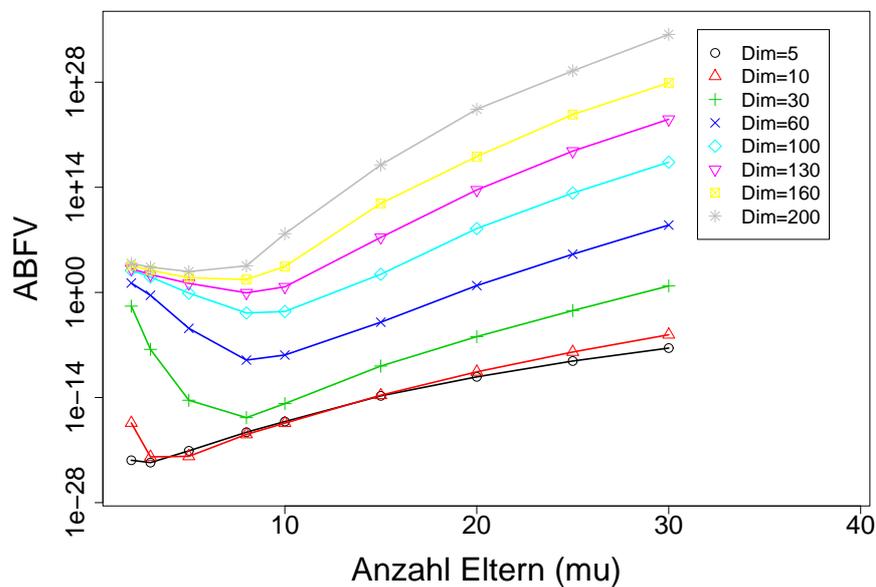


Abbildung 7.10: ABFV einer $(\mu, 100)$ -ES im statischen Kugelmodell in Abhängigkeit von der Problemdimension n und der Anzahl Eltern.

dimension n und der Anzahl der Eltern. Auch in diesem Fall lassen sich deutliche Strukturen erkennen. Während für kleine Problemdimensionen der ABFV bereits für $\mu = 3$ minimiert wird, sind für größere n höhere Werte für die Anzahl der Eltern erforderlich. Für $n = 10$ liegt der Wert bei $\mu = 5$ und steigt für die weiteren Problemdimensionen auf $\mu = 8$ an. Auch diese Experimentreihe lässt sich mit anderen festen Einstellungen für die Anzahl der Nachkommen wiederholen und liefert dabei ebenfalls vergleichbare Resultate.

Aus den obigen Experimenten wird eines deutlich: Wird neben der Problemdimension einer der Parameter μ oder λ fest gewählt, so gibt es einen Wert für den nun freien Populationsparameter, der den ABFV minimiert. Eine durch die Untersuchungen von Cvetković und Mühlenbein (1994) motivierte Vermutung lautet, dass es einen einfachen funktionalen Zusammenhang zwischen diesen Größen geben könnte. Cvetković und Mühlenbein haben für die Optimierung der OneMax-Funktion eine komplexe Formel zur Berechnung der optimalen Populationsgröße vorgestellt. Für die vorliegende Problemstellung wurden bereits für $\mu = 15$ erste diesbezügliche Untersuchungen gestartet. Dabei hat sich allerdings keine einfach zu erkennende Abhängigkeit der Anzahl Nachkommen von

der Problemdimension ergeben. Zudem können aufgrund des bereits oben Gesagten keine eindeutigen Aussagen getroffen werden. Zum einen besitzen für vereinzelte Strategien die Werte des ABFV aufgrund zu weniger Einzelläufe noch keine ausreichende Genauigkeit. Zum anderen wurden die Experimente nur in teils größeren Abstufungen für λ vorgenommen. Das führt zu nicht abschätzbaren Ungenauigkeiten bei einer Regression. Dass die gefundenen Zusammenhänge zudem nicht auf andere Werte für μ übertragbar sind, zeigt bereits eine entsprechende Untersuchung für $\mu = 10$. Hieraus deutete sich bereits ein anderer Zusammenhang an, was zur Erkenntnis führt, dass ein Zusammenhang aus einer gemeinsamen Regression für μ und λ resultieren muss. Solcherlei Ergebnisse können allenfalls nur aus wesentlich umfangreicheren Experimenten gefolgert werden, die jedoch aufgrund der beschriebenen Schwierigkeiten mit einem enormen Ressourcenverbrauch (Speicherplatzbedarf und Rechenzeit) verbunden wären.

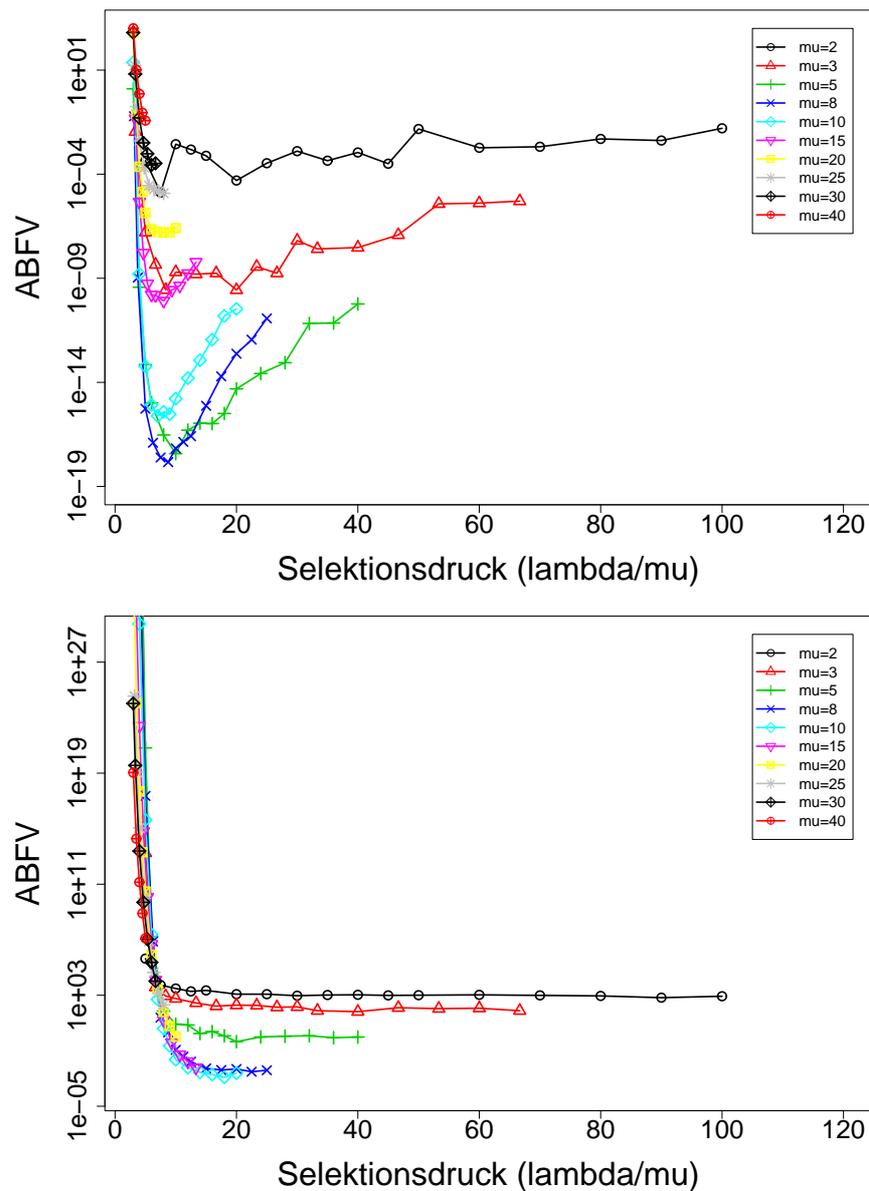


Abbildung 7.11: ABFV einer (μ, λ) -ES im 30-dimensionalen (oben) und 110-dimensionalen (unten) statischen Kugelmodell.

Im Folgenden soll deshalb ein anderer Vergleich der verwendeten Strategien gegeben werden. Dazu wird von der gesonderten Betrachtung von μ und λ abgewichen und stattdessen der Selektionsdruck λ/μ berücksichtigt. Für einzelne Problemdimensionen soll jetzt untersucht werden, ob für die verwendeten Werte für μ ein ähnlicher Selektionsdruck herrschen muss, um den ABFV zu minimieren. Dabei wird sich auch zeigen, welche der getesteten Kombinationen von μ und λ zu den besten Ergebnissen führen.

Die Abbildung 7.11 zeigt diesen Sachverhalt für verschiedene Problemdimensionen. Zunächst gilt es hier zu beachten, dass in den Abbildungen der Bereich des Selektionsdrucks für die einzelnen Kurven sehr unterschiedlich ist. Dies liegt an der Versuchsanordnung, da die Anzahl der Nachkommen auf 200 begrenzt wurde und bspw. für $\mu = 40$ deshalb nur ein maximaler Selektionsdruck von 5 erreichbar ist. Meistens existieren für jede Problemdimension Kombinationen aus μ und λ mit

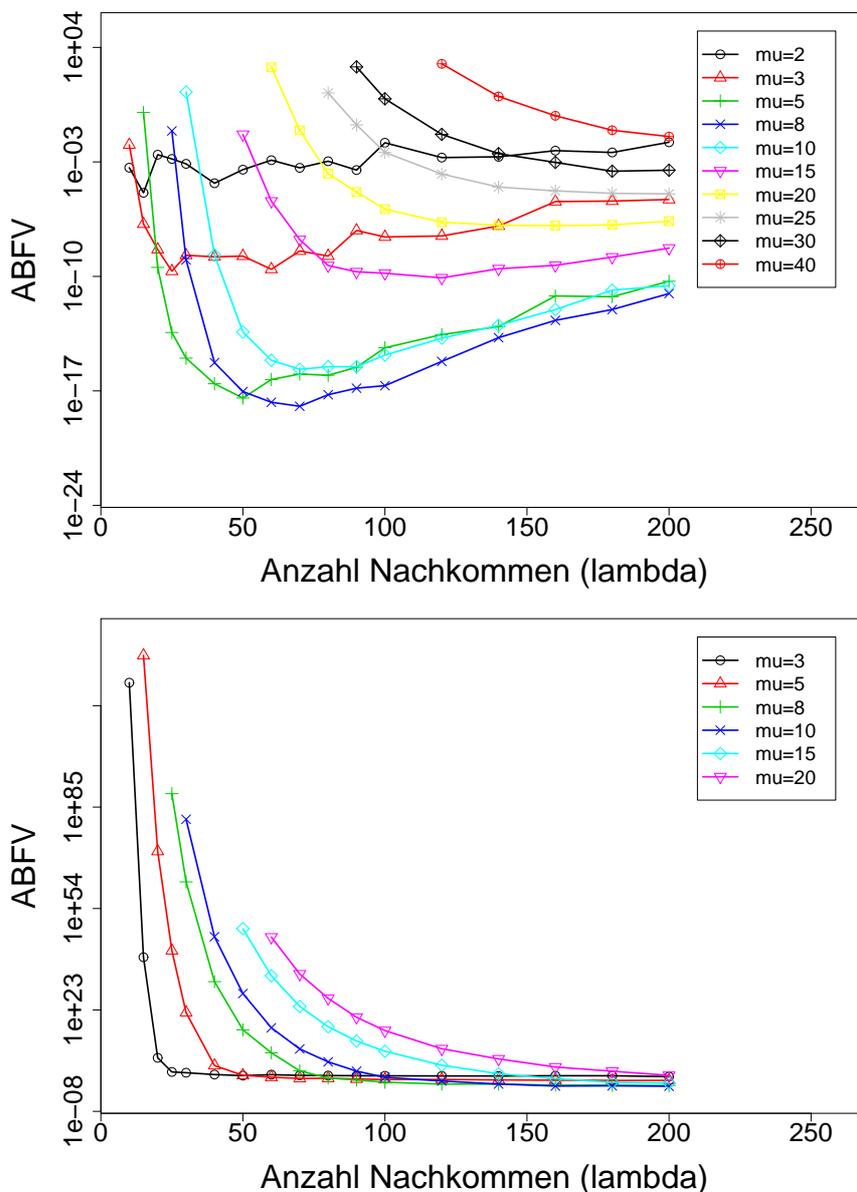


Abbildung 7.12: ABFV einer (μ, λ) -ES im 30-dimensionalen (oben) und 150-dimensionalen (unten) statischen Kugelmodell.

größeren Vorteilen in Bezug auf den ABFV gegenüber den anderen Kombinationen. Eindeutig zu erkennen ist, dass es offensichtlich nicht ausreicht, für jede Wahl von μ oder λ einfach den Selektionsdruck anzupassen. Es ist zu vermuten, dass es für jede Problemdimension ein Paar μ, λ gibt, das den ABFV minimiert. Dabei müssen mit wachsendem n sowohl die Anzahl der Eltern als auch der verwendete Selektionsdruck erhöht werden. Die Abbildung 7.12 zeigt für zwei Problemdimensionen deshalb den ABFV für einzelne Kombinationen für μ und λ .

Abschließend soll noch nach möglichen Gründen für das beobachtete Verhalten gesucht werden. Auffallend ist das schlechte Abschneiden für $\mu = 2$ und $\mu = 3$. Es scheint so zu sein, dass die für die sinnvolle Funktionsweise der Rekombination notwendige, hier aber fehlende Diversität in der Population bessere Ergebnisse verhindert. Für $\mu > 3$ kann sich wohl der bereits von Beyer (1995) angesprochene Effekt der „genetischen Reparatur“ auch im Fall der hier benutzten Rekombination von jeweils zwei Eltern bemerkbar machen. Die Resultate geben Hinweise darauf, dass mit steigender Problemdimension die Anzahl der Eltern ebenfalls wachsen muss. Zum Vergleich folgt deshalb noch die Abbildung 7.13, die die Ergebnisse auf dem 10-dimensionalen Kugelmodell enthält. Es

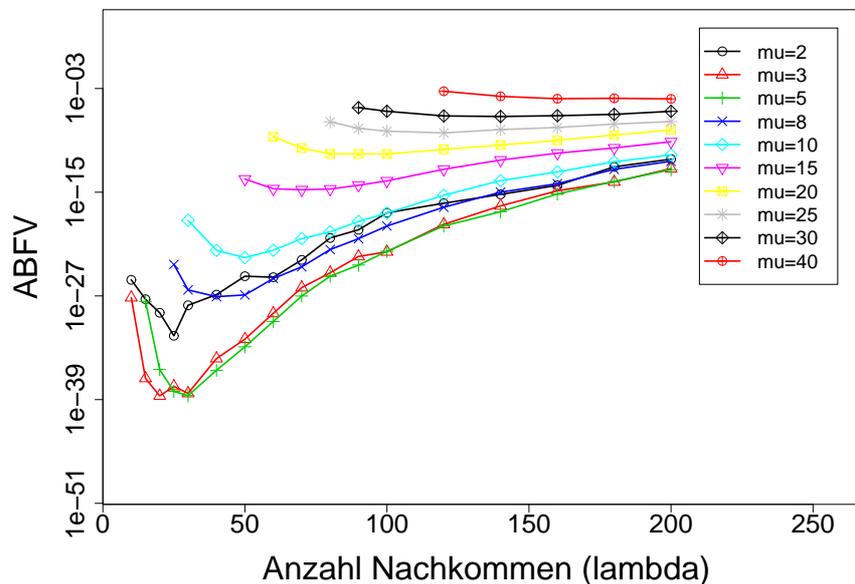


Abbildung 7.13: ABFV einer (μ, λ) -ES im 10-dimensionalen statischen Kugelmodell.

zeigt sich sehr deutlich, dass in diesem Fall bereits drei Eltern ausreichen, um den besten ABFV zu erreichen. Auch für $\mu = 2$ sind die Ergebnisse wesentlich besser, allerdings fallen sie gegenüber den für $\mu \in \{3, 5\}$ deutlich zurück.

7.2 Dynamische Umgebungen

Gemäß unseres Untersuchungsgegenstands führen wir nun erste Experimente auf dem dynamischen Kugelmodell durch. Wir beobachten zunächst wieder das Verhalten einer $(15, 100)$ -ES. Das Optimum bewegt sich dabei in jeder Generation in einer Dimension um die konstante Entfernung $s = 1.0$. Die Ergebnisse in Abhängigkeit von der Problemdimension n sind in der Abbildung 7.14 verzeichnet. Um die Einschwingphase zu minimieren, wurde die ES nahe dem Optimum gestartet und mit nahezu optimalen Einstellungen der Schrittweiten initialisiert. Wenn der ABFV der Strategie besser oder gleich dem Startwert der ES ist, so konnte die ES dem Optimum sehr gut folgen.

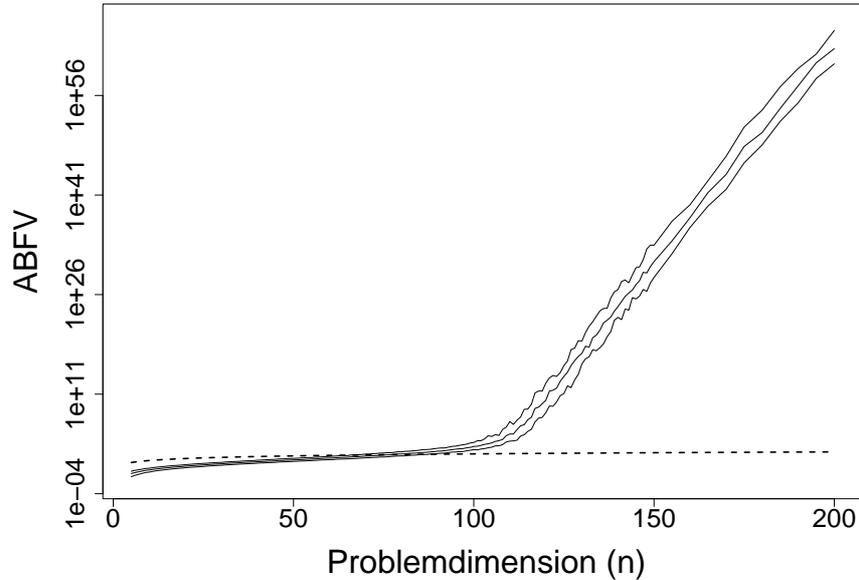


Abbildung 7.14: Konvergenzverhalten einer (15, 100)-ES auf dem dynamischen Kugelmodell in Abhängigkeit von der Problemdimension n . Neben dem ABFV sind noch die 0.05- und 0.95-Quantile des ABFV dargestellt. Die gestrichelte Linie zeigt darüber hinaus den durchschnittlichen Startwert der ES. Das Optimum bewegt sich jede Generation in einer Dimension um die Entfernung $s = 1.0$ (Bewegungstyp I).

Die Abbildung zeigt, dass dies für $n < 70$ der Fall ist. Auch wenn die ES im Bereich $70 < n < 100$ die geringe anfängliche Entfernung zum Optimum nicht halten kann, so ist sie dennoch in der Lage, dem Optimum in akzeptabler Entfernung zu folgen. Wie im statischen Fall liegt der Bereich der beginnenden Divergenz bei $n \approx 100$. Für $n = 110$ ist das Verhalten der Einzelläufe vergleichbar mit dem im statischen Kugelmodell beobachteten Verlauf. D.h. in einigen Läufen konnte die ES dem Optimum folgen, während in anderen eine Divergenz zu verzeichnen ist. Für größere n (≥ 120) kann die ES die Dynamik dagegen nicht beherrschen und divergiert für alle Einzelläufe vom Optimum. Bzgl. der Divergenz können wir festhalten, dass sich bei der verwendeten Dynamik kaum ein Unterschied zum statischen Fall ergibt. Mithin ist die hier aufgetretene Divergenz nicht durch die Dynamik begründet.

Das prinzipielle Verhalten ist unabhängig vom Typ der Dynamik und der Stärke der Bewegung. Dies wird durch die Abbildung 7.15 dokumentiert. Im Gegensatz zu den vorherigen Experimenten haben wir hier einerseits die Bewegungsstärke reduziert. Andererseits erfolgte ein Wechsel zum Bewegungstyp III, bei dem sich das Optimum in einer schrägen Diagonalen in allen Dimensionen bewegt. Neben den abweichenden Funktionswerten ist auf den ersten Blick kein gravierender Unterschied festzustellen. Die Abbildung 7.16 illustriert für $n = 110$ die Differenzen eines Einzellaufs für $s = 0.1$ gegenüber $s = 1.0$. Im Fall der geringeren Bewegungsstärke schwanken die meisten Läufe um den Startwert. Für die größere Bewegungsstärke benötigt die ES eine größere Einschwingphase bis der schwankende Verlauf erreicht ist. Dass der Startwert beim Bewegungstyp III bereits für geringere Problemdimensionen überschritten wird, liegt natürlich daran, dass sich die Optimierung einer Bewegung in n Dimensionen schwieriger gestaltet.

In der folgenden Testserie führen wir einige Evolutionsstrategien mit variierenden Populationsgrößen auf unterschiedlichen Problemdimensionen durch ($n \in \{10, 30, 50\}$). Wie wir bereits festgestellt haben kann eine (15, 100)-ES für diese n dem Optimum in einem konstanten Abstand

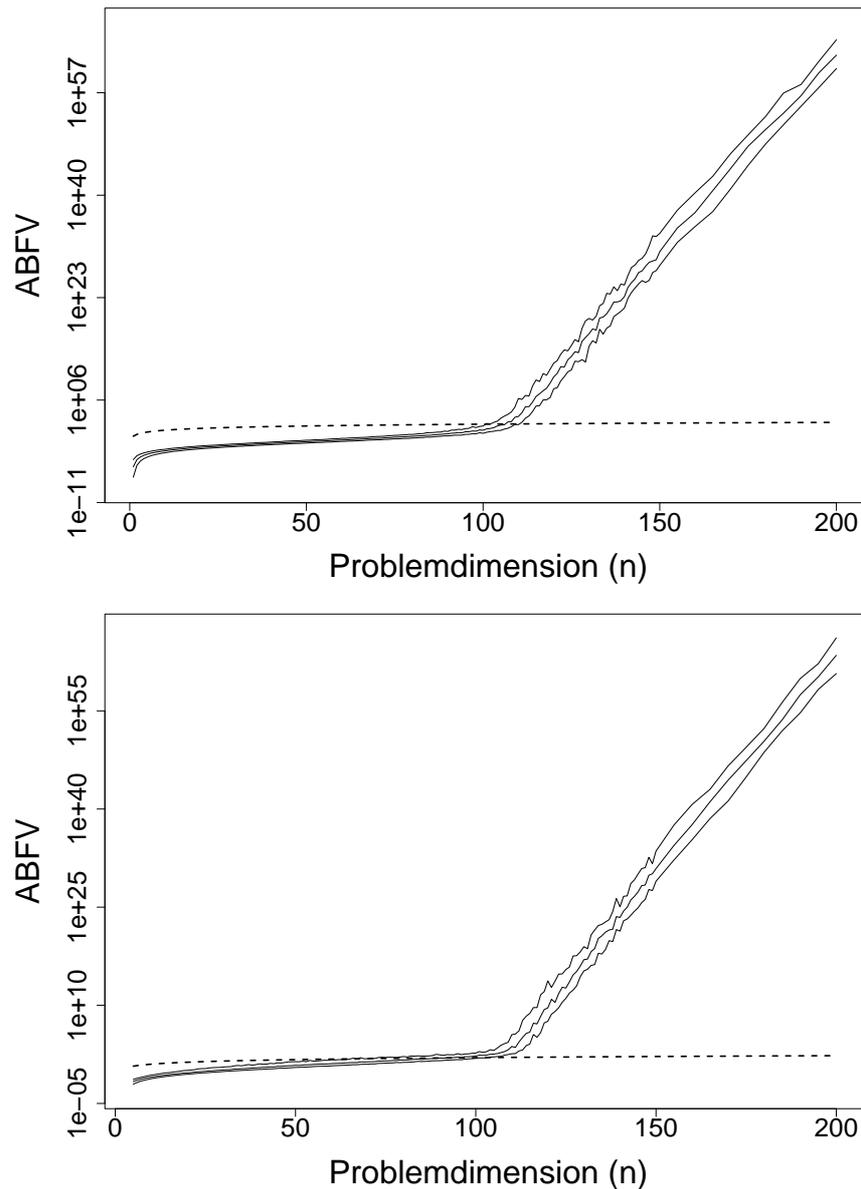


Abbildung 7.15: Konvergenzverhalten einer (15,100)-ES auf dem dynamischen Kugelmodell in Abhängigkeit von der Problemdimension n . Neben dem ABFV sind noch die 0.05- und 0.95-Quantile der 31 unabhängigen Läufe dargestellt. Die gestrichelte Linie zeigt darüber hinaus den durchschnittlichen Startwert der ES. Das Optimum bewegt sich jede Generation in einer Dimension (Bewegungstyp I, oben) und allen Dimensionen (Typ III, unten) mit einer Stärke $s = 0.1$.

folgen. Die obere Grafik der Abbildung 7.17 zeigt die Ergebnisse von verschiedenen (μ, λ) -ES auf dem 30-dimensionalen dynamischen Kugelmodell, wenn sich das Optimum in einer Dimension bewegt. Für diese Untersuchungen wird ein konstanter Selektionsdruck von $\lambda/\mu = 10$ benutzt. Jede Strategie wird über 300.000 Funktionsauswertungen ausgeführt, wobei die Bewegungsfrequenz Δg zwischen $\{1, 2, 3, 5, 10\}$ wechselt. Die Bewegungsstärke s pro Generation wird so gesetzt, dass die zurückgelegte Gesamtentfernung konstant bei $s_{\text{gesamt}} = s \cdot g / \Delta g = 300$ liegt. Dies führt für ein Verfahren zu einer Bewegungsstärke von $s = 1/1000 \cdot \lambda \cdot \Delta g$ pro Generation. Wie die Abbildung zeigt werden die besten Ergebnisse erreicht, wenn sich das Optimum jede Generation bewegt ($\Delta g = 1$).

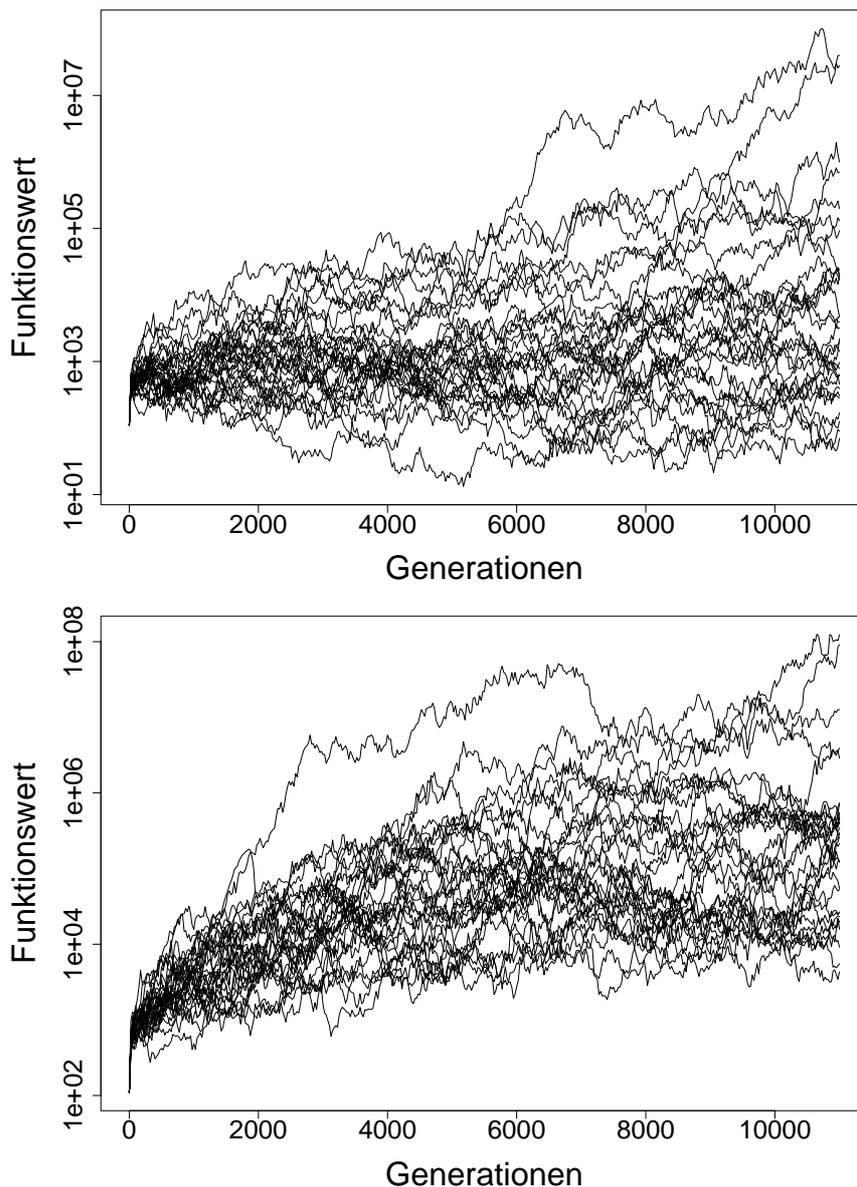


Abbildung 7.16: Alle 31 unabhängigen Durchläufe einer $(15, 100)$ -ES auf dem dynamischen Kugelmodell mit $n = 110$. Das Optimum bewegt sich jede Generation in einer Dimension (Typ I) mit der Bewegungsstärke $s = 0.1$ (oben) bzw. $s = 1.0$ (unten).

Mit zunehmender Bewegungsfrequenz erhöht sich auch der erreichte ABFV. Die besten Funktionswerte werden jeweils mit einer $(2, 20)$ -ES erzielt. Offenbar ist es in dieser Situation ausreichend, eine kleine Diversität für die Rekombination zu gewährleisten. Hier nicht dargestellte Experimente mit einer ES mit $\mu = 1$ führten dagegen zu wesentlich schlechteren Resultaten.

Wenn sich das Optimum in allen Dimensionen bewegt, ergibt sich ein wenig anderes Bild. Die besten Ergebnisse werden hier mit einer $(10, 100)$ - und $(20, 200)$ -ES erzielt (untere Grafik der Abbildung 7.17), also mit einer größeren Anzahl Eltern. Die mit unterschiedlichen Bewegungsfrequenzen Δg erhaltenen Ergebnisse unterscheiden sich dagegen weniger. Für den Fall der Bewegung in einer Dimension unterschieden sich die durch verschiedene Bewegungsfrequenzen resultierenden Funktionswerte um ca. 0,3. Wird diese absolute Differenz auch für die Bewegung in allen Dimensionen

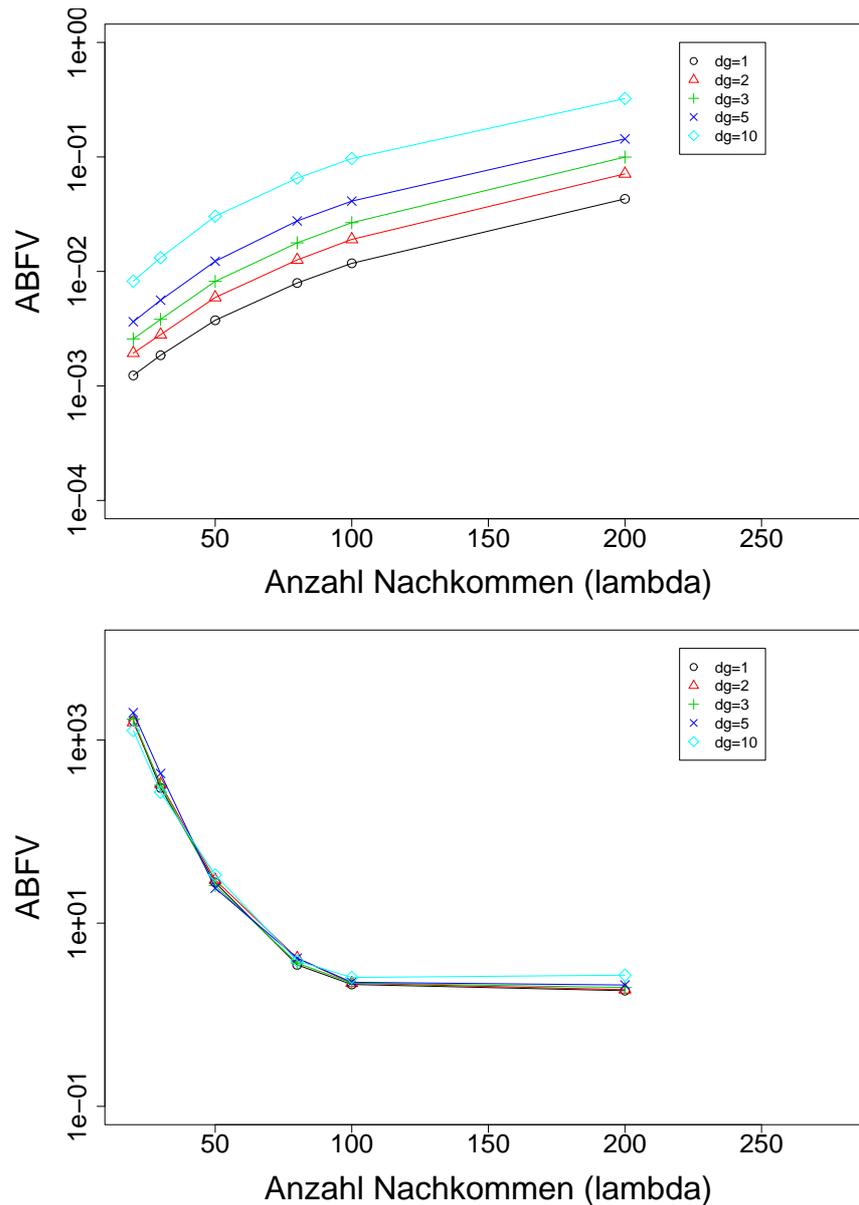


Abbildung 7.17: ABFV verschiedener (μ, λ) -ES mit gleichem Selektionsdruck $\lambda/\mu = 10$ auf dem 30-dimensionalen dynamischen Kugelmodell. Das Optimum bewegt sich in einer Dimension (Typ I, oben) oder in allen Dimensionen (Typ III, unten) und einer Bewegungsfrequenz $\Delta g \in \{1, 2, 3, 5, 10\}$. Die Anzahl der Funktionsauswertungen beträgt 300.000 und die zurückgelegte Gesamtdistanz 300.

angenommen, so fällt der Einfluss der Bewegungsfrequenz aufgrund der höheren Funktionswerte kaum ins Gewicht. Der erreichbare Funktionswert wird hier offensichtlich durch den komplexeren Bewegungstyp dominiert.

Die Vermutung, dass der Bewegungstyp den erreichbaren Funktionswert sehr stark bestimmt, wird durch folgende Überlegung zum Mutationsoperator unterstützt. Üblicherweise werden die Schrittweiten und Objektvariablen wie folgt mutiert:

$$\sigma'_i = \sigma_i \cdot \exp(\tau_0 \cdot N(0,1) + \tau \cdot N_i(0,1)), \quad x'_i = x_i + \sigma'_i \cdot N_i(0,1) .$$

Bei einer Bewegung in einer Dimension heißt das für nahezu optimal eingestellte Schrittweiten:

$\sigma_1 = 1, \sigma_2 = \dots \sigma_n = \epsilon$ mit ϵ klein. Im Fall einer gleichmäßigen Bewegung in allen Dimensionen haben alle den ungefähren Wert $\sigma_1 = \dots = \sigma_n = \sqrt{1/n} + \epsilon$. Abhängig von den Einstellungen der Lernraten τ_0 und τ hat eine mögliche Änderung von $\sigma_i \gg \epsilon$ um einen Faktor c einen größeren negativen Einfluss auf den Funktionswert als die gleiche prozentuale Änderung eines $\sigma_i \approx \epsilon$. In diesem Sinn ist im ersten Fall nur eine Schrittweite anfällig, im zweiten Fall sind dagegen alle Schrittweiten anfällig. Aus diesem Grund macht sich bei einer Bewegung in allen Dimensionen (Bewegungstyp II und III) eine größere Diversität in der Population positiv bemerkbar.

Wir kommen noch einmal auf die Experimente aus dem Kapitel 6 zurück. Wie wir dort gesehen haben ist eine ES mit n Schrittweiten einer ES mit nur einer Schrittweite selbst dann nicht unterlegen, wenn sich das Optimum gleichmäßig in allen Richtungen bewegt. Dieses Verhalten kann nun damit begründet werden, dass aufgrund der größeren Schrittweiten die Schwankungen der Objektvariablen in den einzelnen Komponenten sehr groß sind. Selbst bei den hier verwendeten 30 Dimensionen kommt es dann immer einmal vor, dass die zur Bildung der einen oder anderen Objektvariable gezogene Zufallszahl stark von Null abweicht, sodass die durch Differenzen in den n Schrittweiten hervorgerufenen Schwankungen überlagert werden. Falls diese Vermutung allgemein gültig ist, sollte für größeres n eine höhere Wahrscheinlichkeit existieren, dass sich unter den neuen Schrittweiten eines Individuums zumindest ein Ausreißer befindet, der den Funktionswert stark verschlechtert. Obwohl theoretische Untersuchungen diese Vermutung aufgreifen sollten, werden wir zu diesem Punkt bei der Erörterung der Ergebnisse für die Problemdimensionen $n = 10, 50$ zurückkommen.

Um signifikante Differenzen zwischen den Ergebnissen der einzelnen Strategien mit unterschiedlichen Populationsgrößen zu erkennen, sind in der Abbildung 7.18 exemplarisch die Ergebnisse einer

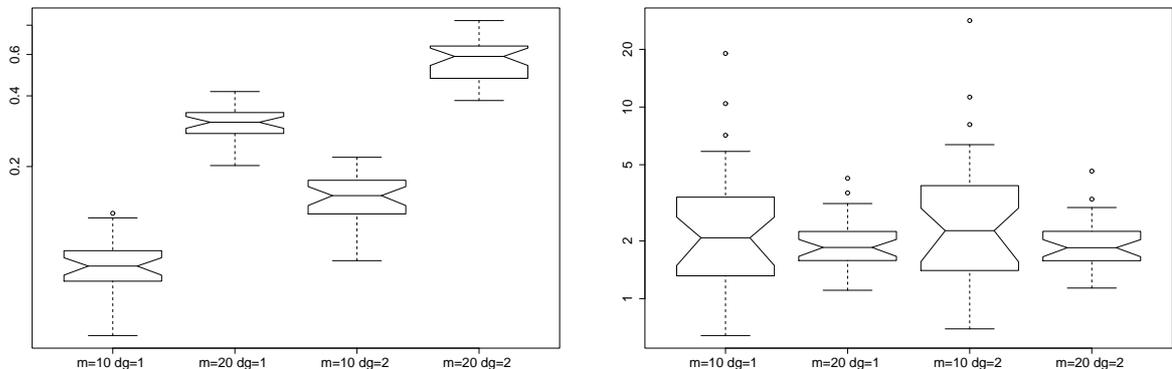


Abbildung 7.18: Boxplot der in jeweils 31 unabhängigen Durchläufen erreichten mittleren Funktionswerte einer (10, 100)- und (20, 200)-ES auf dem dynamischen Kugelmodell mit $n = 30$. Das Optimum bewegt sich jede ($\Delta g = 1$) und jede zweite ($\Delta g = 2$) Generation in einer Dimension (Typ I, links) und in allen Dimensionen (Typ III, rechts). Die zurückgelegte Gesamtdistanz beträgt für alle Strategien 300.

(10, 100)- und (20, 200)-ES in Boxplots dargestellt. Wie der Abbildung zu entnehmen ist, sind für die Bewegung in einer Dimension sämtliche Unterschiede auf dem 5%-Niveau signifikant. Es existieren also signifikante Unterschiede sowohl zwischen den beiden ES-Varianten als auch zwischen den Ergebnissen mit den verschiedenen Bewegungsfrequenzen $\Delta g = 1$ und $\Delta g = 2$. Im Gegensatz dazu sind die für eine Bewegung in allen Dimensionen ersichtlichen Mittelwerte nicht signifikant unterschiedlich. Hier macht sich die größere Schwierigkeit der Optimierung in n Dimensionen be-

merkbar. Auffallend ist dabei einerseits, dass der Boxplot für $\mu = 10$ jeweils wesentlich größer ist als für $\mu = 20$. Dies interpretieren wir als einen deutlichen Hinweis darauf, dass die Anzahl der Eltern zu gering ist, um einheitliche Ergebnisse zu erzielen. Aufgrund der Streuung werden hiermit andererseits aber auch trotz der schlechteren Mittelwerte in einzelnen Durchläufen bessere Ergebnisse erreicht als mit 20 Eltern.

Mit einem extremen Beispiel sollen die bei den beiden Bewegungstypen stark abweichenden erreichbaren Ergebnisse verdeutlicht werden. Die Abbildung 7.19 zeigt die auf dem dynamischen Kugelmodell mit einer (2,20)-ES erreichten Ergebnisse. Im Fall der Bewegung in einer Dimension folgt die ES dem Optimum in einem geringen Abstand. Wenn sich das Optimum in einer schrägen Diagonalen bewegt kann die ES jedoch nur noch in einem größeren Abstand folgen. Bemerkenswert ist, dass die Zeit bis zur erfolgten „Anpassung“ an die Problemstellung in beiden Fällen nahezu gleich ist ($\approx 15,000$ Generationen). Wie den Einzelläufen bei der Bewegung in allen Dimensionen zu entnehmen ist, verläuft der Anstieg bis zur Anpassung geradlinig, um anschließend sehr stark zu schwanken. Weiterhin ist bereits in dem in der Abbildung 7.20 dargestellten Ausschnitt erkennbar, dass es gewisse zyklische Schwankungen gibt, in denen der Funktionswert besser angenähert werden kann, nach einer gewissen Zeit tritt aber wiederum eine stetige Verschlechterung ein. Das gleiche Verhalten zeigt sich auch in den anderen Einzelläufen, wobei sich lediglich die Zyklenlängen und Schwankungsamplituden voneinander unterscheiden. Diese Beobachtung lässt sich ebenfalls durch die im vorigen Abschnitt erwähnte Schrittweitensteuerung erklären. Mit größerer Regelmäßigkeit tritt immer wieder ein ungünstiges Ereignis ein, das dafür sorgt, dass in der nachfolgenden Generation viele Nachkommen schlecht adaptiert sind und sich dieser Effekt über mehrere Generationen negativ bemerkbar macht.

Für $n = 30$ betrachten wir nun den Fortschritt verschiedener ES in Abhängigkeit vom Selektionsdruck λ/μ . Die Abbildung 7.21 zeigt einerseits den ABFV im Fall einer Bewegung in einer Dimension. Den besten Wert erhalten wir hier mit einer (2,20)-ES. Für jedes μ verbessert sich der

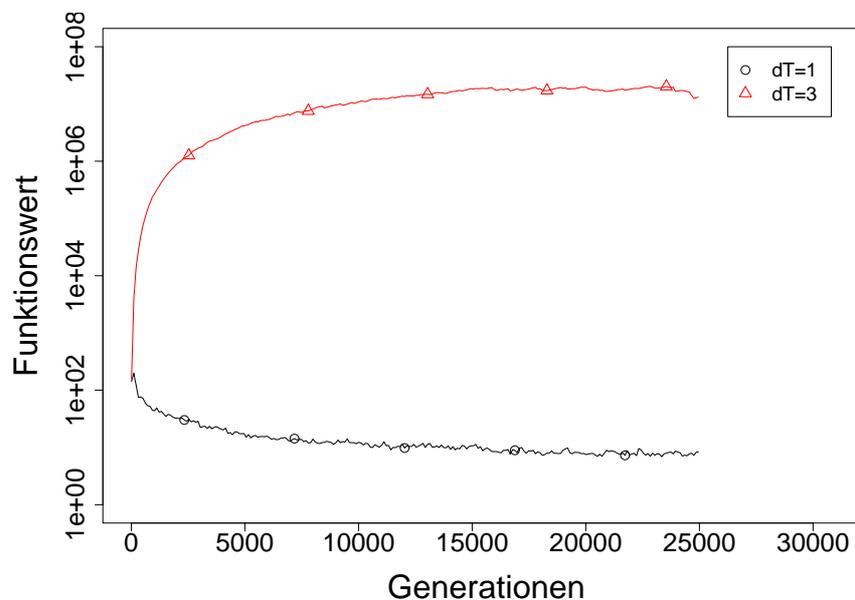


Abbildung 7.19: Mittlerer Lauf einer (2,20)-ES auf dem 50-dimensionalen dynamischen Kugelmodell. Das Optimum bewegt sich jede Generation in einer (Typ I: $dT = 1$) oder allen (Typ III: $dT = 3$) Dimension/en mit einer Bewegungsstärke $s = 1$.

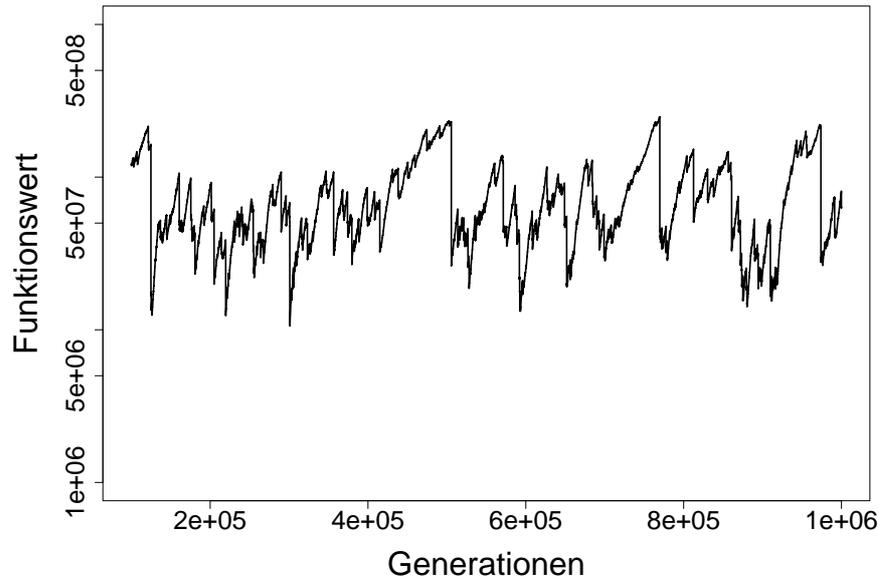


Abbildung 7.20: Einzelllauf einer $(2, 20)$ -ES auf dem 50-dimensionalen dynamischen Kugelmodell. Das Optimum bewegt sich jede Generation in allen Dimensionen (Typ III) mit einer Bewegungsstärke $s = 1$.

ABFV mit wachsendem Selektionsdruck. Um dieses Ergebnis auch für größere Selektionsdrücke zu überprüfen, wären jedoch Experimente mit einer wesentlich erhöhten Anzahl Funktionsauswertungen notwendig. Andererseits werden beim Bewegungstyp III die besten Ergebnisse mit einer Anzahl Eltern von $\mu \approx 20$ erreicht. Für den Selektionsdruck scheint hier ein deutliches Optimum bei einer Kombination von $4 \leq \lambda/\mu \leq 5$ zu liegen. Interessant ist zudem der direkte Vergleich einzelner Kurven. So fällt es auf, dass mit wachsendem μ die Kurven zunächst steil abfallen, in einem kleinen Bereich ihr Minimum annehmen, und dann wieder stärker steigen. Dies führt bspw. dazu, dass sich die Kurven für $\mu = 8$ und $\mu = 40$ zweimal schneiden. Dabei muss allerdings beachtet werden, dass bei einer Erhöhung des Selektionsdruckes um Eins die Anzahl der Nachkommen pro Generation einmal um 8 und einmal um 40 zunimmt.

In den nächsten Abschnitten präsentieren wir weitere Ergebnisse auf dem dynamischen Kugelmodell für $n = 10$ und $n = 50$. Zunächst sind in der Abbildung 7.22 für das 10-dimensionale dynamische Kugelmodell die Ergebnisse verschiedener (μ, λ) -ES mit $\lambda/\mu = 10$ dargestellt, wenn sich das Optimum in einer Dimension bewegt und die Bewegungsfrequenz Δg zwischen $\{1, 2, 3, 5, 10\}$ variiert. Wie schon für $n = 30$ beobachtet, wächst für jedes μ der ABFV mit steigendem Δg . Bewegt sich das Optimum dagegen in einer schrägen Diagonalen weisen die Kurven ein parabelähnliches Aussehen auf. Der ABFV wird dann für $\mu = 8$ minimiert.

Für das 50-dimensionale Kugelmodell erhalten wir ähnliche Ergebnisse wie für $n = 30$. Auch hier werden in den beiden Fällen die besten Ergebnisse für $\mu = 2$ bzw. $\mu \approx 20$ erzielt (Abbildung 7.23). Abweichend von den beiden zuvor betrachteten Problemdimensionen sind einerseits auf der x -Achse die Anzahl Eltern aufgetragen, welche jedoch aufgrund des einheitlichen Selektionsdrucks mit der Anzahl der Nachkommen korrespondieren. Andererseits wurden die Bewegungsfrequenzen zwischen $\Delta g \in \{1, 2, 5, 10, 20\}$ variiert. Wegen des flachen Verlaufs einzelner Kurven für $n = 30$ wurden die Experimente hier mit bis zu 500 Nachkommen durchgeführt. Wie zu sehen, zeichnet sich ein Minimum jeweils für $10 < \mu < 30$ ab. Für die Bewegung in allen Dimensionen sind zwischen den Ergebnissen mit differierender Bewegungsfrequenz, aber gleicher Anzahl Eltern, mit bloßem

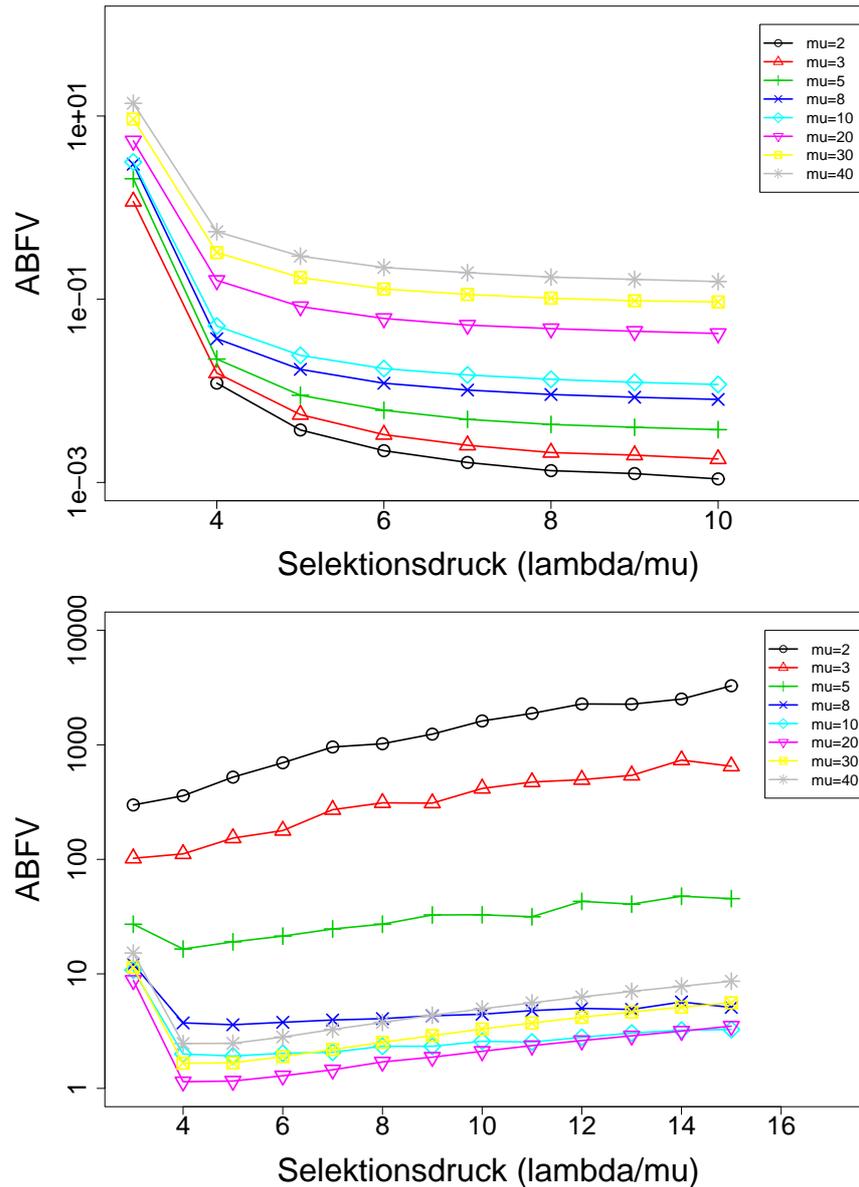


Abbildung 7.21: ABFV verschiedener (μ, λ) -ES in Abhängigkeit vom Selektionsdruck λ/μ auf dem 30-dimensionalen dynamischen Kugelmodell bei einer Bewegung in einer Dimension (Typ I, oben) und in allen Dimensionen (Typ III, unten). Die zurückgelegte Gesamtentfernung beträgt 600 bei 600.000 Funktionsauswertungen.

Auge kaum Unterschiede feststellbar. Lediglich für $\Delta g = 20$ zeichnet sich eine deutliche Trennung ab.

Für unterschiedliche Selektionsdrücke enthält die Abbildung 7.24 den ABFV der (μ, λ) -ES auf dem dynamischen Kugelmodell mit $n = 10$ und einer Bewegung in einer Dimension. Die Kurven zeigen keine Besonderheiten und korrespondieren zu denen in Abbildung 7.21. Für eine schräge diagonale Bewegung in allen Dimensionen erhalten wir etwas überraschende Ergebnisse. Einige Strategien ($\mu \in \{3, 5\}$) zeigen ein wesentlich anderes Verhalten als der Rest. Mit steigendem Selektionsdruck fallen diese Kurven über einen größeren Bereich. Dagegen besitzen die Kurven für größeres μ ihr Minimum bei kleinem Selektionsdruck. Dieses Verhalten kann wie folgt begründet

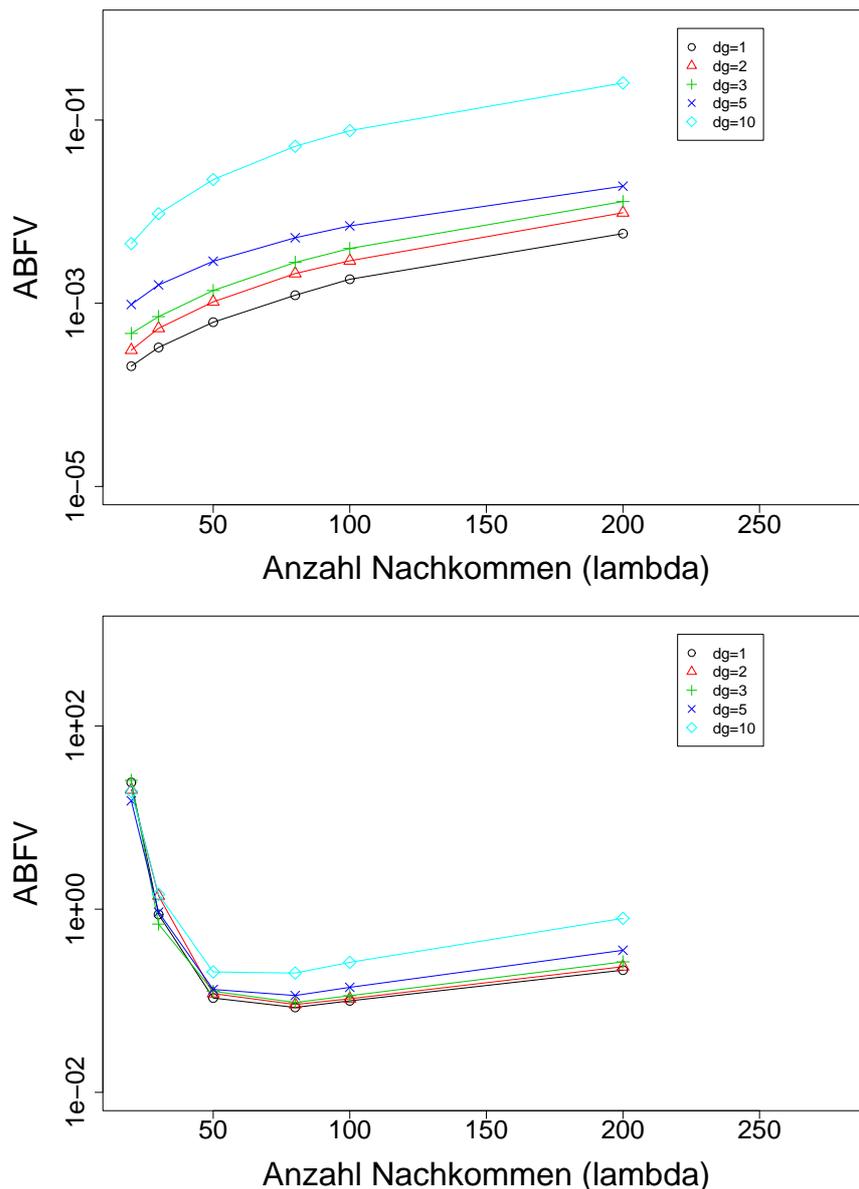


Abbildung 7.22: ABFV verschiedener (μ, λ) -ES mit gleichem Selektionsdruck $\lambda/\mu = 10$ auf dem 10-dimensionalen dynamischen Kugelmodell bei einer Bewegung in einer Dimension (Typ I, oben) und in einer schrägen Diagonalen (Typ III, unten). Die Bewegungsfrequenz variiert zwischen $\Delta g \in \{1, 2, 3, 5, 10\}$, wobei die zurückgelegte Entfernung 100 bei 100.000 Funktionsauswertungen beträgt.

werden. Eine aufgrund der kleinen Anzahl Eltern existierende zu geringe Diversität in der Population kann für $n = 10$ noch dadurch ausgeglichen werden, dass eine erhöhte Anzahl Nachkommen generiert wird. Dies ist im Fall $n = 30$ jedoch nicht mehr möglich, da dort die Wahrscheinlichkeit für einen Erfolg zu gering ist.

An dieser Stelle möchten wir ein etwas anderes Experiment durchführen. Es dient zur Beurteilung, ob der von Beyer (1995) als „genetische Reparatur“ bezeichnete Effekt hier positiv genutzt werden kann. Dazu werden die Schrittweiten eines Nachkommen durch global-intermediäre Rekombination der Schrittweiten aller Eltern erzeugt, während für die Objektvariablen die diskrete

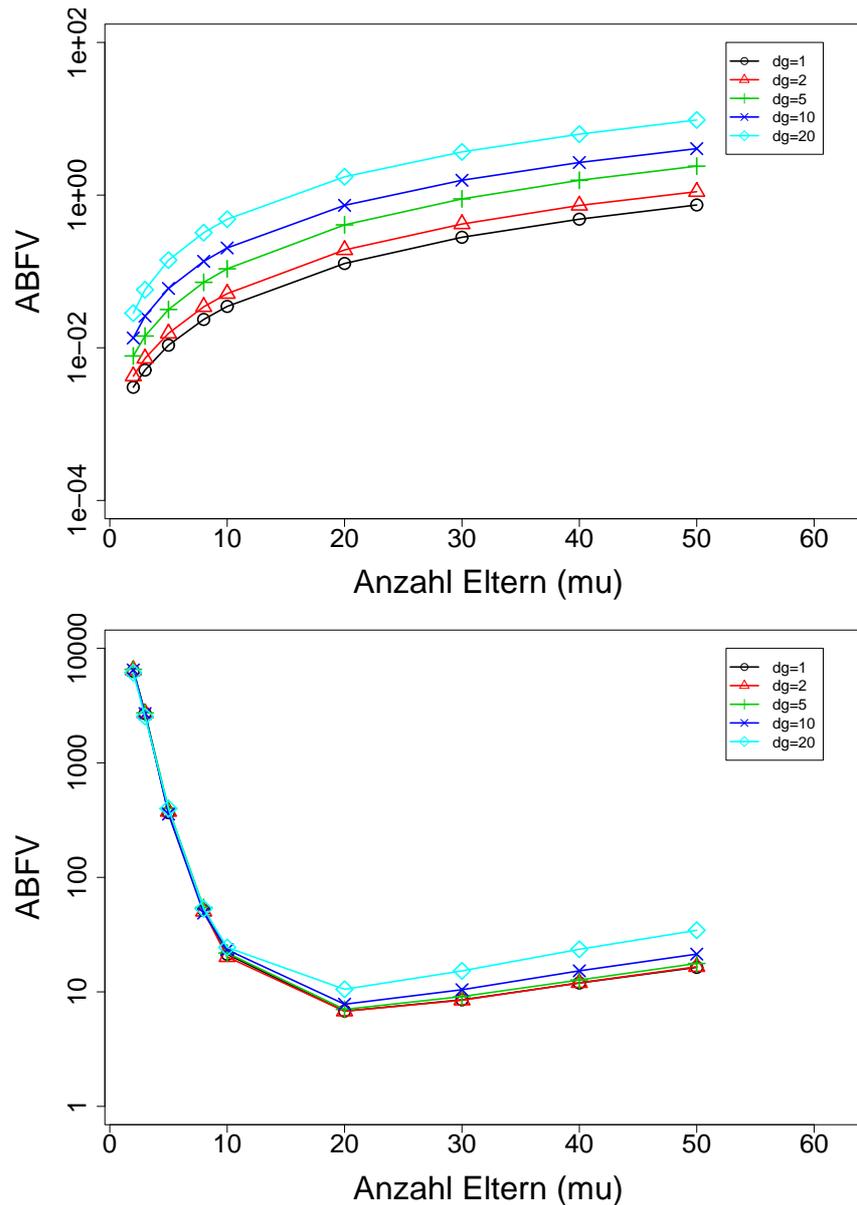


Abbildung 7.23: ABFV verschiedener (μ, λ) -ES mit gleichem Selektionsdruck $\lambda/\mu = 10$ auf dem 50-dimensionalen dynamischen Kugelmodell bei einer Bewegung in einer Dimension (Typ I, oben) und in allen Dimensionen (Typ III, unten). Die zurückgelegte Gesamtentfernung beträgt 1.000 bei 1.000.000 Funktionsauswertungen. Die Bewegungsfrequenz variiert zwischen $\Delta g \in \{1, 2, 5, 10, 20\}$.

Rekombination beibehalten wird. Die Abbildung 7.25 zeigt, dass einige Strategien bessere Ergebnisse erzielen. Verbesserungen werden insbesondere bei einer kleinen Anzahl Eltern erreicht. Für $\mu \in \{2, 3\}$ sind die Ergebnisse merklich besser als bei Verwendung der üblichen Rekombination der Schrittweiten. Allerdings weichen die Ergebnisse bei den anderen Strategien nur unwesentlich von denen des anderen Rekombinationstyps ab. Eine Begründung für dieses Verhalten ist nicht evident. Bei vielen Eltern könnte man vermuten, dass unter den vielen Eltern immer wieder Ausreißer auftreten, die bei der Rekombination aller Eltern zwangsläufig eingebunden werden. Demgegenüber sollten sich aber bei wenigen Eltern bereits einzelne Ausreißer stark bemerkbar machen, da sie nicht ausgemittelt werden können. Da die Auswirkungen dieses Rekombinationsoperators zu kei-

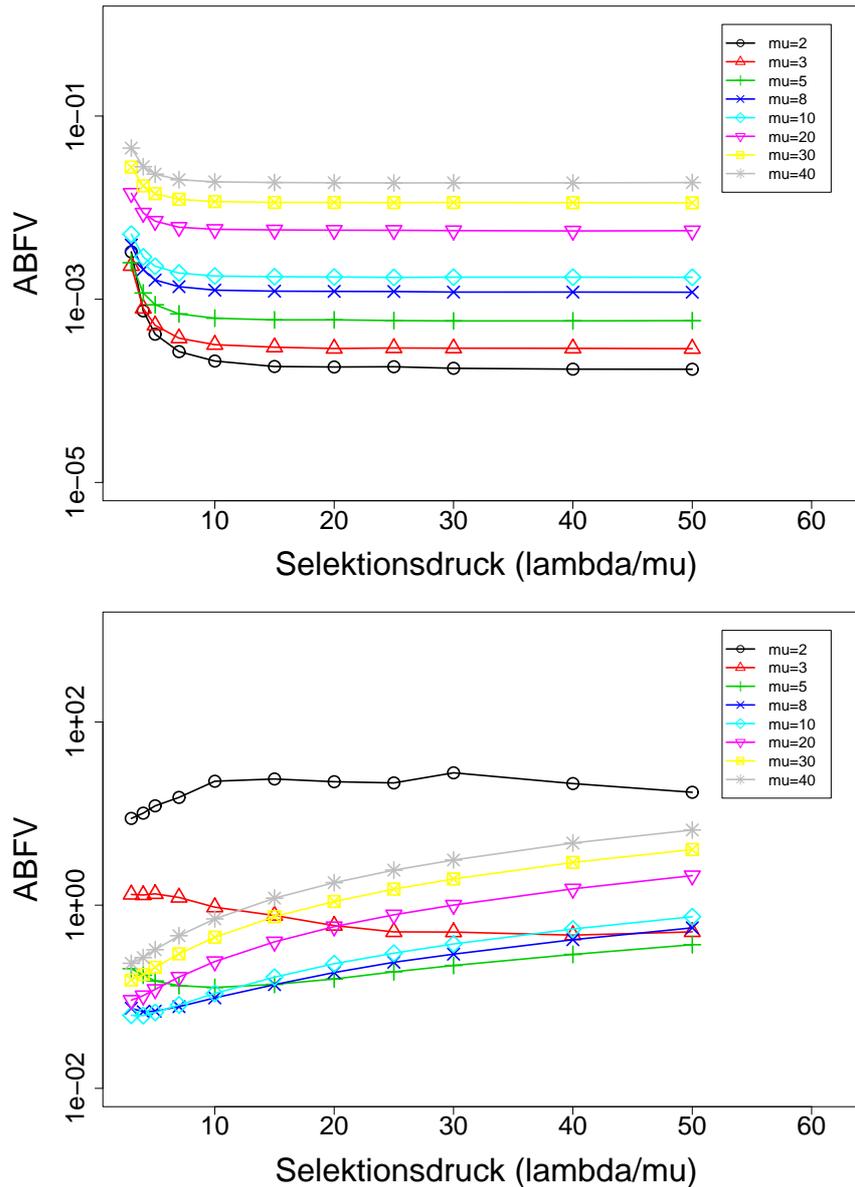


Abbildung 7.24: ABFV verschiedener (μ, λ) -ES in Abhängigkeit vom Selektionsdruck λ/μ auf dem 10-dimensionalen dynamischen Kugelmodell bei einer Bewegung in einer Dimension (Typ I, oben) und in allen Dimensionen (Typ III, unten). Die zurückgelegte Gesamtentfernung beträgt 2.400 bei 2.400.000 Funktionsauswertungen.

nen wesentlich anderen Ergebnissen führt und dementsprechend kein eindeutiger Zusammenhang erkennbar ist, wird im Folgenden von der weiteren Verwendung dieses Rekombinationstyps abgesehen.

Die für $n = 50$ mit unterschiedlichen Populationsgrößen erhaltenen Kurven sind in der Abbildung 7.26 dargestellt. Das Aussehen der Kurven entspricht im Wesentlichen dem Aussehen der für $n = 30$ erhaltenen Kurven und wird deshalb hier nicht so intensiv diskutiert. Während man für die Bewegung in einer Dimension mit $\mu = 2$ die besten Resultate erhält, scheint dagegen für den Bewegungstyp III wiederum ein Wert von $\mu \approx 20$ Eltern eine sehr gute Wahl zu sein. Hier kann die ES meistens erst wieder für $\lambda/\mu \geq 4$ dem Optimum in einem tolerierbaren Abstand folgen.

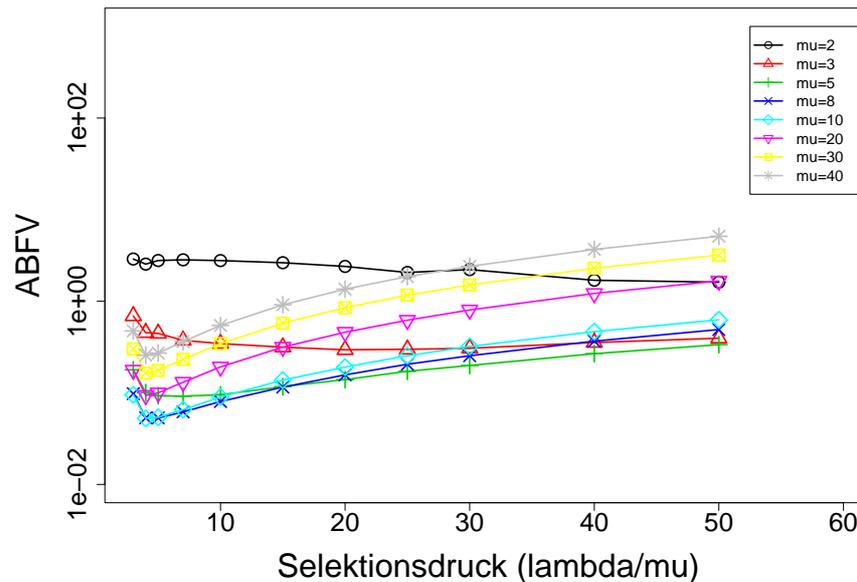


Abbildung 7.25: ABFV verschiedener (μ, λ) -ES in Abhängigkeit vom Selektionsdruck auf dem 10-dimensionalen dynamischen Kugelmodell mit einer Bewegung in einer schrägen Diagonalen (Typ III). Die zurückgelegte Gesamtentfernung beträgt 2.400 bei 2.400.000 Funktionsauswertungen. Für die Schrittweiten wurde eine global-intermediäre Rekombination aller Eltern verwendet.

Zuletzt soll an dieser Stelle noch kurz auf das Verhalten bei nicht konstanten Bewegungsstärken eingegangen werden. Für die hierzu durchgeführten Experimentreihen wurden verschiedene Strategien auf dem dynamischen Kugelmodell ausgeführt. Zur Simulation einer variablen Bewegungsstärke wird in jeder Generation zu dem fest gewählten Wert s eine normalverteilte Zufallsvariable mit Standardabweichung d unterschiedlicher Größe hinzuaddiert. Um die eigentliche Bewegung nicht zu dominieren, wurde der Term im Bereich $d \in [0, s]$ variiert. Durch die stochastische Bewegungsstärke kann die zurückgelegte Gesamtentfernung von einer Strategie zur anderen etwas abweichen. Aufgrund der hohen Zahl Generationen werden diese Abweichungen jedoch auf ein Minimum reduziert. Konkret beträgt die maximale Abweichung der zurückgelegten Gesamtentfernung weniger als 2%. Den ABFV für verschiedene Werte der Bewegungsstreuung zeigt die Abbildung 7.27. Bei beiden Bewegungstypen sind die Abweichungen relativ klein. Während für die Bewegung in einer Dimension eine (exponentiell) zunehmende Verschlechterung des ABFV zu verzeichnen ist, werden die Schwankungen überlagert, wenn sich das Optimum in allen Dimensionen bewegt. Hier dominieren die durch den Bewegungstyp verursachten schlechten Funktionswerte die Abweichungen aufgrund der variierenden Entfernungen. D.h. in dem vorliegenden Versuchsaufbau hat die variierende Bewegungsstärke lediglich bei einer Bewegung des Typs I einen deutlich erkennbaren Einfluss. Weitere Experimente zeigen, dass bei Verwendung einer normalverteilten Bewegungsstärke der Bereich der Problemdimension n , ab dem die Divergenz einsetzt, identisch ist zu dem Fall der konstanten Bewegungsstärke.

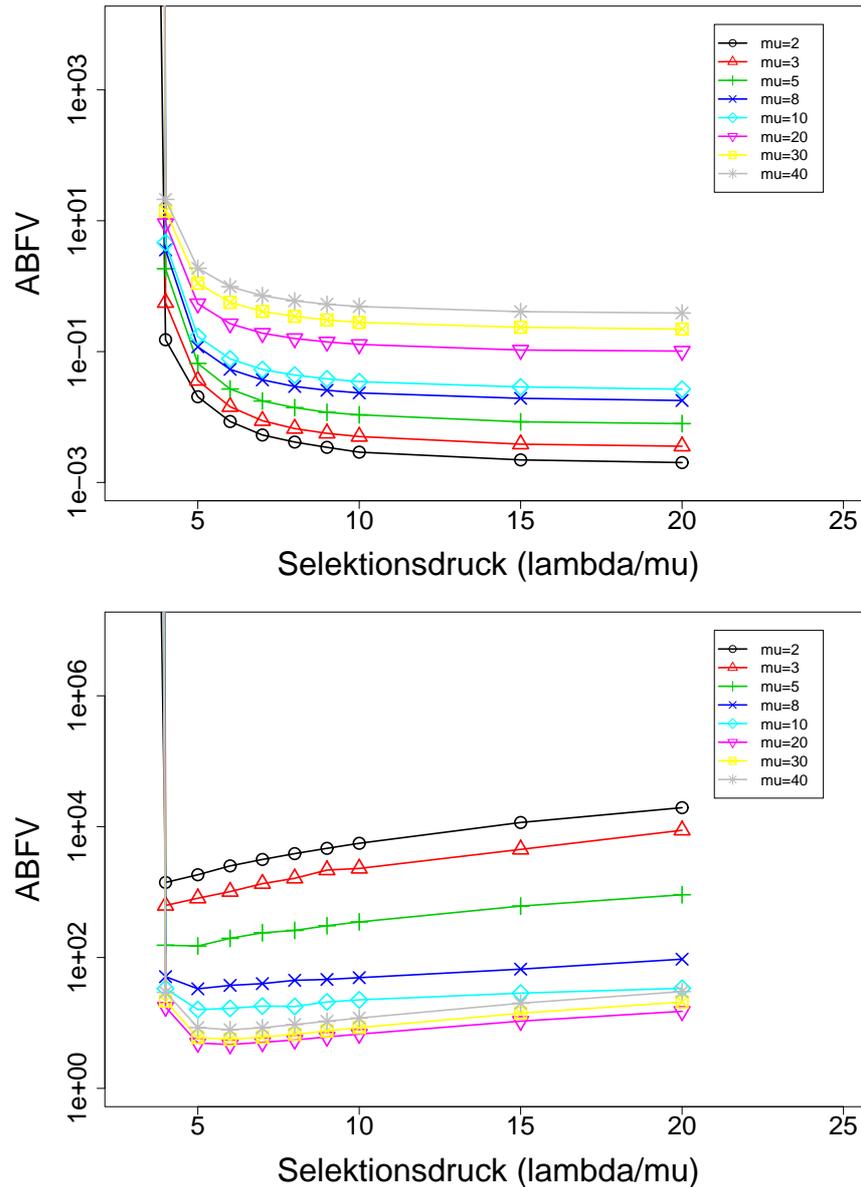


Abbildung 7.26: ABFV verschiedener (μ, λ) -ES in Abhängigkeit vom Selektionsdruck auf dem 50-dimensionalen dynamischen Kugelmodell bei einer Bewegung in einer Dimension (Typ I, oben) und in allen Dimensionen (Typ III, unten). Die zurückgelegte Gesamtentfernung beträgt 2.000 bei 2.000.000 Funktionsauswertungen. Für die Schrittweiten wurde eine global-intermediäre Rekombination aller Eltern verwendet.

7.3 Fazit

Nach diesen sehr umfangreichen Untersuchungen auf dem Kugelmodell fassen wir zunächst noch einmal einige wichtige gewonnene Erkenntnisse für den statischen Fall stichpunktartig zusammen:

- Für kleine Problemdimensionen ($n \approx 10$) werden die besten Ergebnisse mit einer kleinen Anzahl Eltern und Nachkommen erzielt. Hier reichen schon 3–5 Eltern und 30 Nachkommen aus.
- Mit leicht steigender Problemdimension muss zunächst bei gleicher Anzahl Nachkommen die

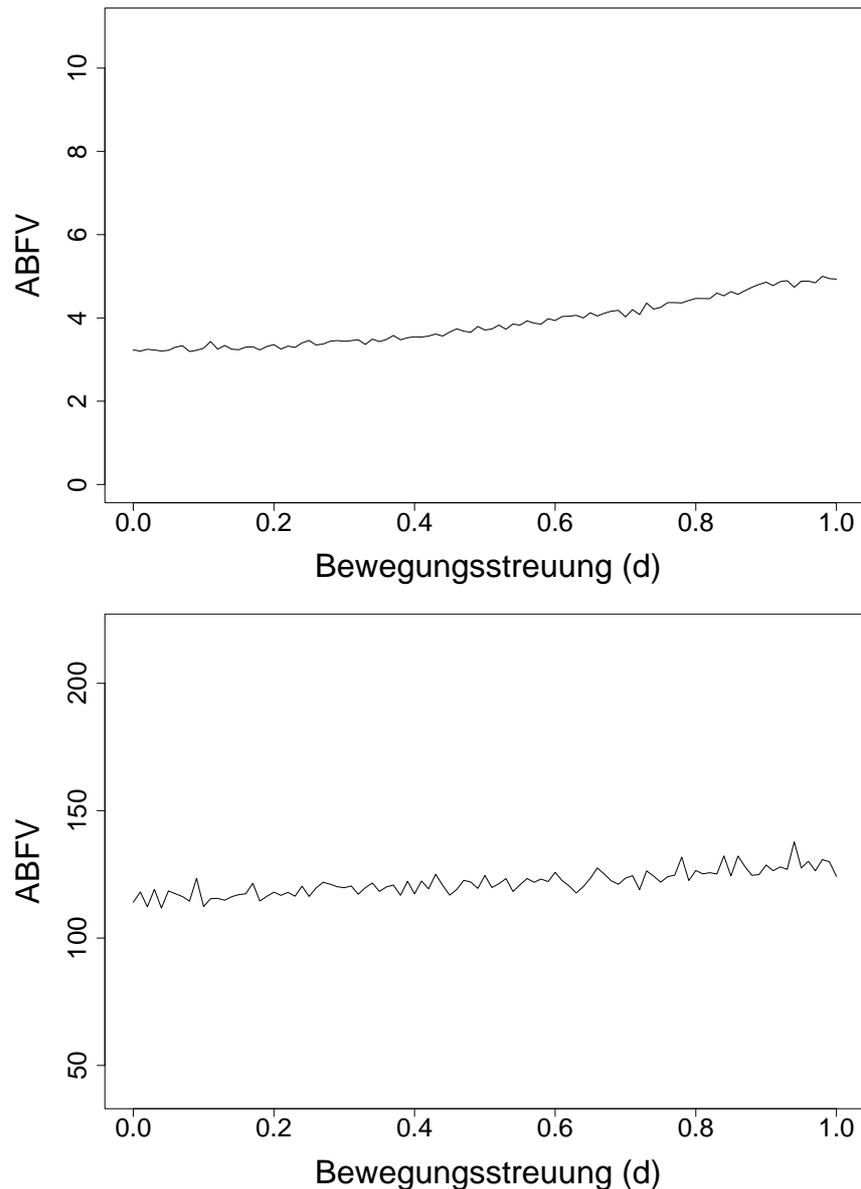


Abbildung 7.27: ABFV einer (15, 100)-ES in Abhängigkeit von der Stärke der Bewegungsstreuung auf dem 30-dimensionalen dynamischen Kugelmodell. Das Optimum bewegt sich in einer Dimension (Typ I, oben) bzw. in allen Dimensionen (Typ III, unten), wobei die zurückgelegte Gesamtentfernung ca. 3.000 beträgt bei 300.000 Funktionsauswertungen (entspricht $s = 1$).

Anzahl der Eltern etwas erhöht werden. Bei weiter steigender Problemdimension ist auch die Zahl der Nachkommen zu erhöhen, wobei ebenfalls der Selektionsdruck steigen muss. Für mittlere Problemdimensionen ($n \approx 30$) erweist sich dann eine (10, 100)-ES als angemessen.

- Für hohe Problemdimensionen ($n > 100$) ist es das wesentliche Element, mit großen Populationsgrößen (insbesondere λ) die Divergenz zu verhindern. Die erreichbaren Funktionswerte unterscheiden sich dann aber nicht mehr sehr stark voneinander.
- Ein direkter Zusammenhang zwischen der Problemdimension und den zu wählenden Populationsgrößen ist nicht erkennbar.

Für die betrachteten Situationen im dynamischen Fall sind unter anderem folgende wesentliche Ergebnisse erwähnenswert:

- Der Bereich der Problemdimension, ab dem eine Divergenz zu verzeichnen ist, unterscheidet sich im dynamischen Fall nicht von dem des statischen Falls.
- Eine große Bewegungsstärke führt zu einer schwerer zu optimierenden Problemstellung als eine kleine Bewegungsstärke. Zwar sind die erreichbaren Funktionswerte schlechter als bei einer kleinen Bewegungsstärke. Demgegenüber gibt es aber kaum Unterschiede, was den Bereich der Problemdimension anbelangt, ab dem eine Divergenz zu verzeichnen ist.
- Mit den hier verwendeten Strategien ist es leichter eine Funktion zu optimieren, deren Optimum sich bei gleicher Bewegungsstärke nur in einer Dimension bewegt als eine, die sich in allen Dimensionen bewegt. Dabei spielt es kaum eine Rolle, ob die Bewegung gleichmäßig in allen Dimensionen erfolgt oder in einer schrägen Diagonalen.
- Die Unterschiede zwischen der Optimierung einer Funktion mit konstanter Bewegungsstärke und einer Funktion mit variierender Bewegungsstärke sind relativ gering (solange die Variationen im Verhältnis zur Bewegungsstärke klein sind). Die Unterschiede verschwinden nahezu, wenn sich das Optimum in allen Dimensionen bewegt.
- Was die Populationsgrößen betrifft, treffen die schon im statischen Fall aufgestellten Regeln zu. Allerdings sind auf dem dynamischen Kugelmodell gute Einstellungen immer etwas größer zu wählen als im statischen Fall. Dies gilt insbesondere für kleine Problemdimensionen, wo im dynamischen Fall statt der 3 – 5 eher 8 – 10 Eltern ratsam sind. Dann ist aber auch die Anzahl der Nachkommen zu erhöhen. Für große Problemdimensionen bleibt auch hier das Hauptziel, die beginnende Divergenz zu verhindern.

Kapitel 8

Ergebnisse auf multimodalen Funktionen

Aufgrund der auf dem Kugelmodell erhaltenen Ergebnisse können wir uns bei der Betrachtung multimodaler Funktionen auf das Verhalten einiger weniger Strategien konzentrieren. Sollten die Resultate mit dem prinzipiellen Verhalten auf dem Kugelmodell übereinstimmen, dürfen wir nämlich mit einiger Gewissheit annehmen, dass andere Initialisierungen und Parametereinstellungen ähnliche Auswirkungen haben wie die auf dem Kugelmodell beobachteten.

8.1 Ackley-Funktion

Wie schon beim Kugelmodell schauen wir uns zunächst das Konvergenzverhalten einer (15, 100)-ES für unterschiedliche Problemdimensionen auf der statischen Ackley-Funktion an. Die Abbildung 8.1 zeigt, dass auch für die Ackley-Funktion ein ähnlicher Verlauf zu erkennen ist. Für Problemdimensionen

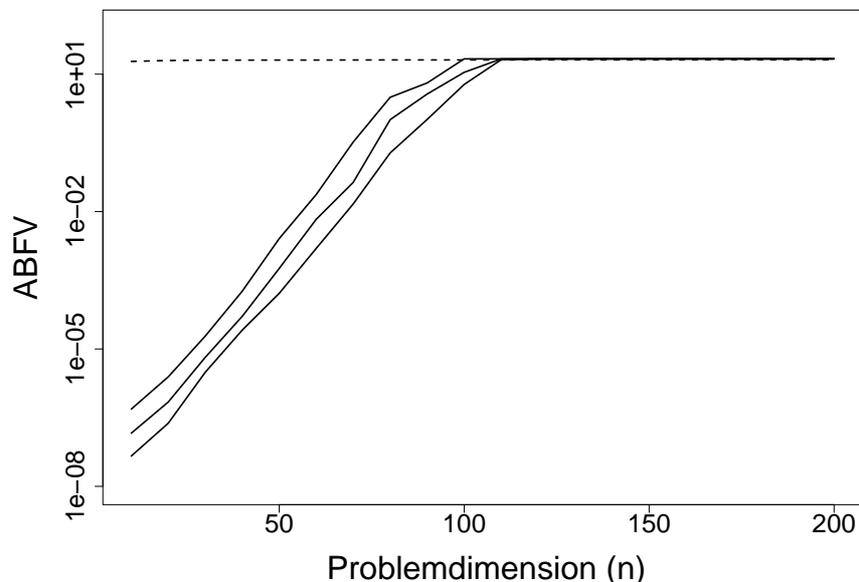


Abbildung 8.1: Konvergenzverhalten einer (15, 100)-ES auf der statischen Ackley-Funktion in Abhängigkeit von der Problemdimension. Neben dem ABFV sind noch die 0.05- und 0.95-Quantile dargestellt. Die gestrichelte Linie zeigt darüber hinaus den durchschnittlichen Startwert der ES.

sionen $n < 100$ kann die ES Fortschritte erzielen, wobei diese mit zunehmender Problemdimension immer kleiner ausfallen. Ab dem Bereich $n \approx 100$ fängt die ES jedoch auch auf der Ackley-Funktion an zu divergieren. Nicht überraschen darf uns der Verlauf der Kurve für $n > 110$. Hier macht sich der nach oben durch $a + e$ beschränkte Wertebereich der Ackley-Funktion bemerkbar. Ein näherer Blick auf die Objektvariablen zeigt uns, dass sich auch hier die ES immer weiter vom Optimum entfernt.

Für andere Initialisierungen der Objektvariablen und Schrittweiten sowie der Parametrisierungen einzelner Strategieparameter ist kein vom prinzipiellen Verlauf auf dem Kugelmodell abweichendes Verhalten erkennbar. Deshalb kommen wir an dieser Stelle direkt zu der Darstellung der Ergebnisse mit unterschiedlichen Populationsgrößen und den sonstigen Einstellungen der Standard-ES. Zunächst beantwortet die Abbildung 8.2 die Frage, ob auch hier eine $(10, 100)$ -ES bessere Er-

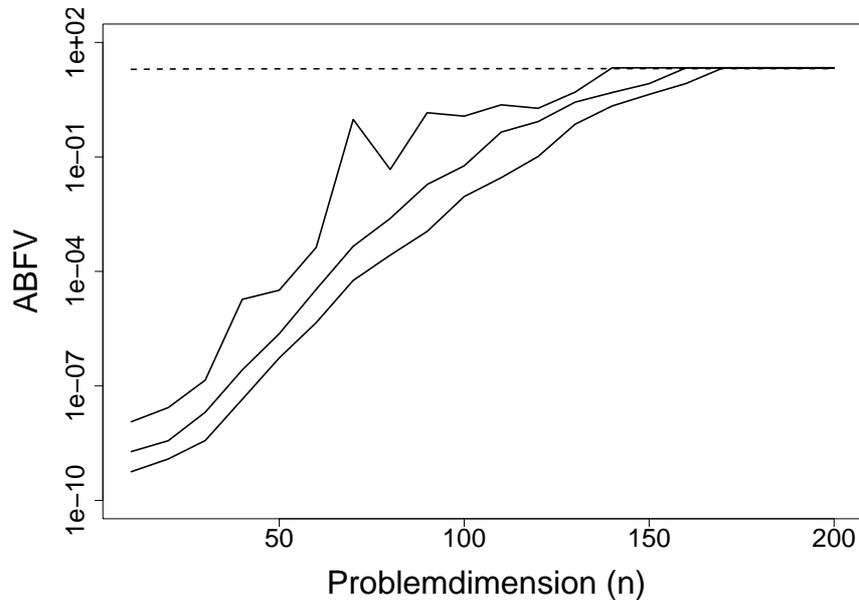


Abbildung 8.2: Konvergenzverhalten einer $(10, 100)$ -ES auf der statischen Ackley-Funktion in Abhängigkeit von der Problemdimension. Neben dem ABFV sind noch die 0.05- und 0.95-Quantile dargestellt. Die gestrichelte Linie zeigt darüber hinaus den durchschnittlichen Startwert der ES.

gebnisse liefert. Dies ist eindeutig zu bejahen. Der Bereich der beginnenden Divergenz der ES hat sich nach $n > 130$ verschoben. Ein statistischer Test zeigt, dass die Unterschiede zwischen beiden Strategien für alle verwendeten n auf dem 5%-Niveau signifikant sind.

Die nächsten beiden Abbildungen zeigen für einen festen Populationsparameter die Abhängigkeit des ABFV von der Problemdimension n und dem jeweils anderen Populationsparameter. Die Abbildung 8.3 zeigt dies für $\mu = 15$ Eltern und die Abbildung 8.4 für $\lambda = 100$ Nachkommen. In beiden Fällen zeichnen sich deutliche Strukturen ab. Bei fester Wahl eines der beiden Populationsparameter gibt es für jede Problemdimension einen Wert für den jeweils anderen Populationsparameter, der den ABFV minimiert. Bei der $(15, \lambda)$ -ES steigt dieser Wert mit zunehmender Problemdimension immer weiter an. Dies entspricht unseren Erwartungen, ist doch die Optimierung einer Funktion mit größerer Problemdimension in der Regel immer schwieriger. Demgegenüber ergibt sich bei fester Wahl von $\lambda = 100$ ein etwas anderes Bild. Für eine Problemdimension $n \leq 100$ steigt der Wert für μ , der den ABFV minimiert, zuerst immer weiter an. Danach findet jedoch eine Umkehrung statt, d.h. für $n > 100$ wird der Wert für μ immer kleiner. Dieser zunächst überraschende Effekt

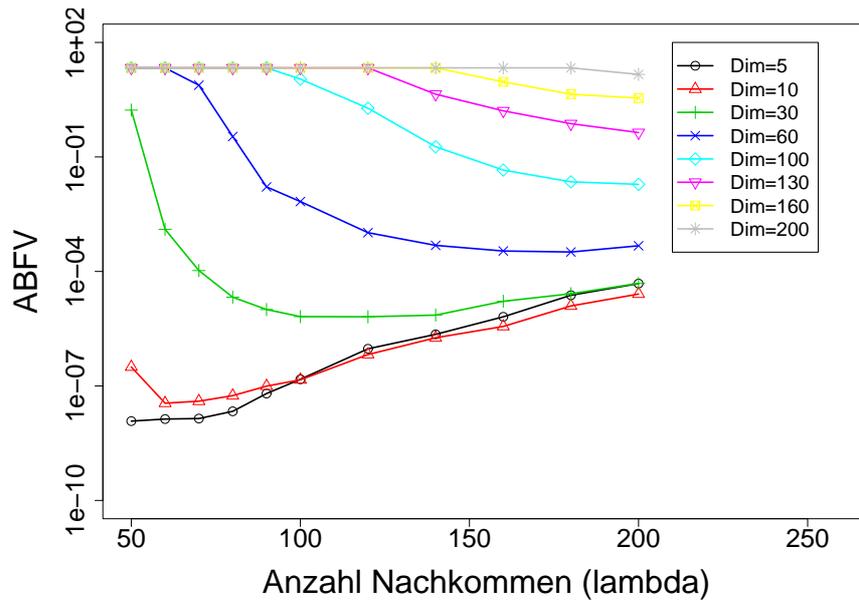


Abbildung 8.3: ABFV einer $(15, \lambda)$ -ES auf der statischen Ackley-Funktion in Abhängigkeit von der Problemdimension und der Anzahl Nachkommen.

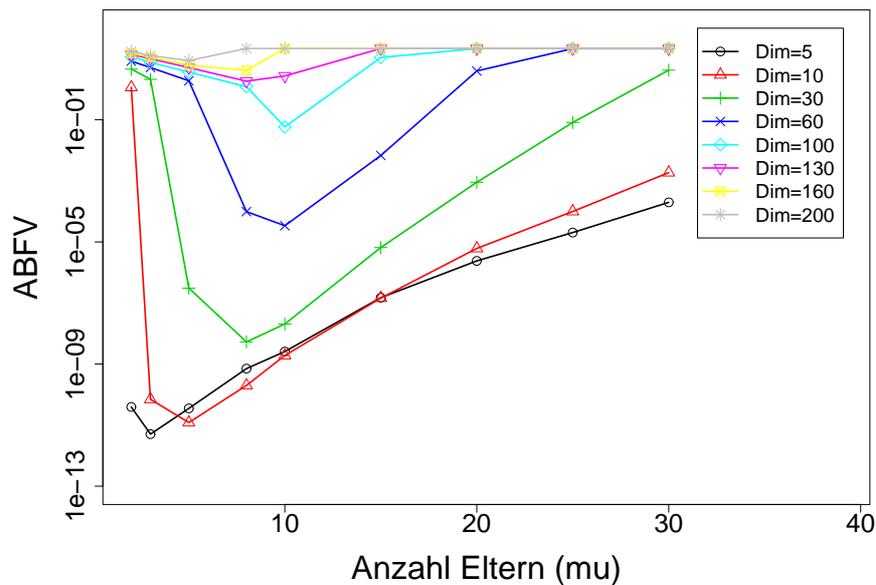


Abbildung 8.4: ABFV einer $(\mu, 100)$ -ES auf der statischen Ackley-Funktion in Abhängigkeit von der Problemdimension und der Anzahl Eltern.

hat jedoch einen einfachen Grund. Für die größeren Problemdimensionen ($n > 100$) ist eine Wahl von $\lambda = 100$ nicht mehr ausreichend, um einen adäquaten Fortschritt zu erzielen. Bei weiterer Erhöhung von μ wäre nun wegen des kleineren Selektionsdrucks ein Fortschritt gar nicht mehr möglich. Stattdessen stellt nun die Wahl eines kleinen μ aufgrund des größeren Selektionsdrucks das kleinere Übel dar und gewährleistet, dass überhaupt noch Fortschritte erzielt werden.

In den Abbildungen 8.5 bis 8.7 ist der ABFV für einzelne Kombinationen von μ und λ für drei Problemdimensionen dargestellt. Für $n = 30$ fällt auf, dass die Strategien mit $\mu \in \{2, 3\}$ zu relativ schlechten Ergebnissen führen. Eine Variation der Anzahl der Nachkommen hat hier kaum

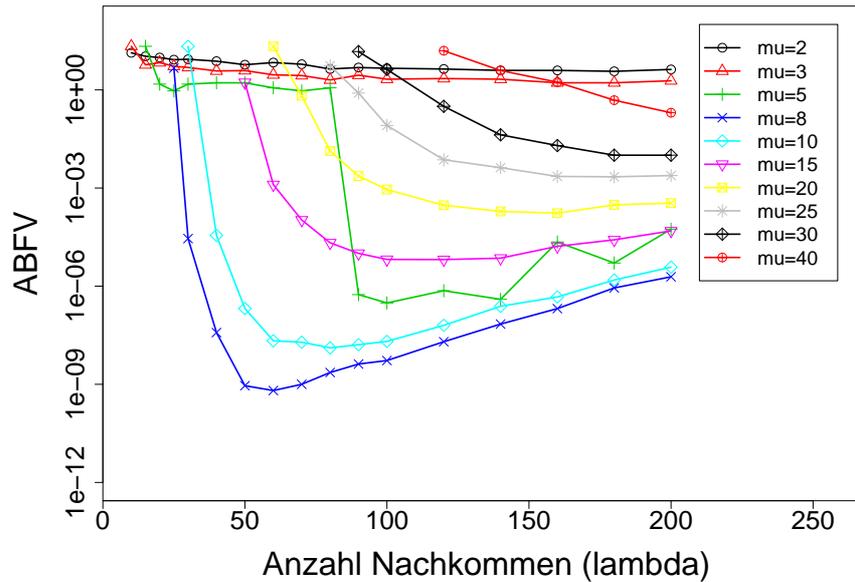


Abbildung 8.5: ABFV einer (μ, λ) -ES auf der 30-dimensionalen statischen Ackley-Funktion.

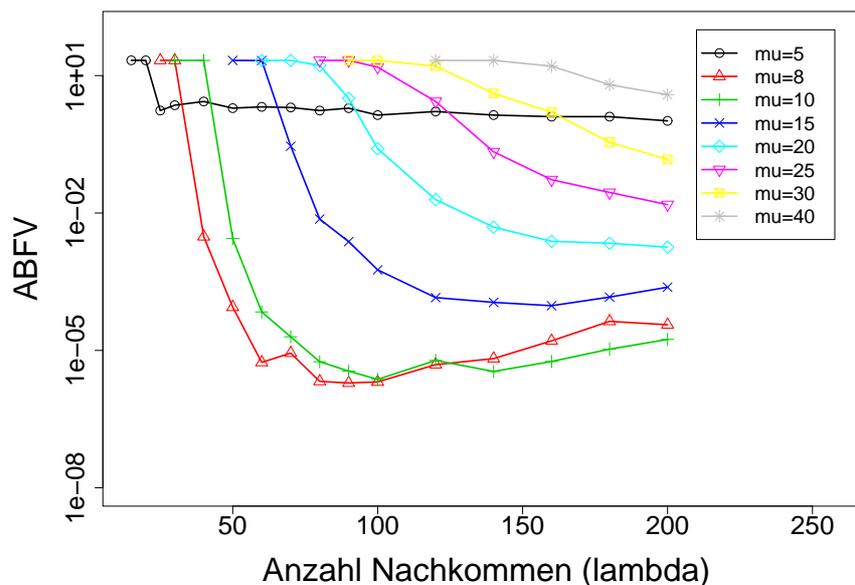


Abbildung 8.6: ABFV einer (μ, λ) -ES auf der 50-dimensionalen statischen Ackley-Funktion.

Auswirkungen. Es ist zu vermuten, dass die geringe Zahl der Eltern für eine zu kleine Diversität sorgt, sodass lokale Optima nur schwer verlassen werden können. Die Kurve für $\mu = 5$ zeigt einen besonderen Verlauf. Bei bis zu 80 Nachkommen sind die Fortschritte relativ moderat. Erst mit mehr Nachkommen können bessere Ergebnisse erzielt werden, die jedoch nicht so gut wie die mit $\mu = 8$ oder $\mu = 10$ erreichten sind. Hier scheint die Schwelle für μ zu liegen, ab der auf der statischen Ackley-Funktion ein passabler Fortschritt erreicht werden kann. Für die anderen Einstellungen für μ sind die erzielten Ergebnisse meistens wesentlich besser. Für alle μ zeichnen sich dabei auch auf der Ackley-Funktion deutliche Werte für λ ab, mit denen der ABFV minimiert wird. Die insgesamt besten Ergebnisse werden mit $\mu = 8$ erzielt. Für $\mu > 8$ verlaufen die Kurven jeweils oberhalb der jeweils verwendeten niedrigeren Einstellung für μ .

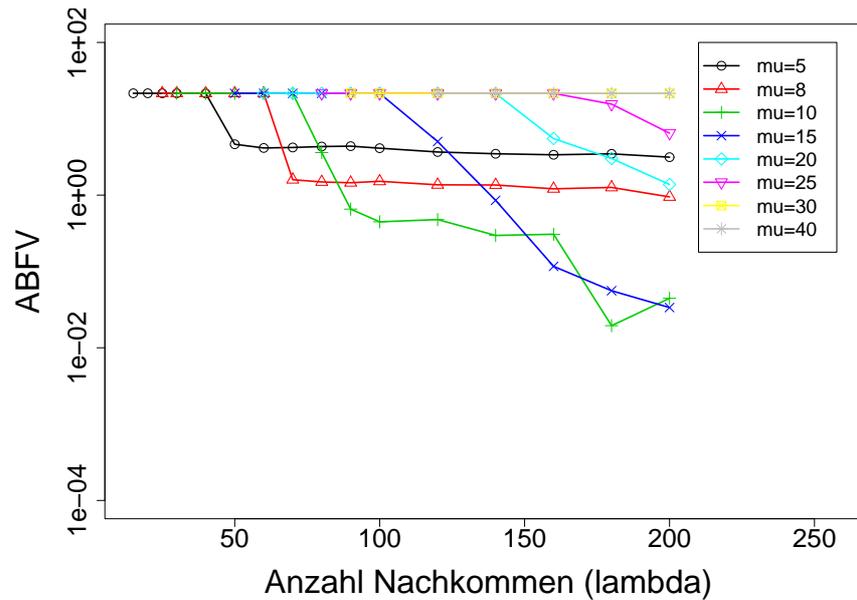


Abbildung 8.7: ABFV einer (μ, λ) -ES auf der 110-dimensionalen statischen Ackley-Funktion.

Da bereits für $n = 30$ die Ergebnisse mit $\mu \in \{2, 3\}$ sehr schlecht ausfielen und sich für $n \geq 50$ weiter verschlechterten, werden die entsprechenden Kurven in den folgenden Grafiken nicht dargestellt. Für $\mu = 5$ sind die Ergebnisse bei kleinem Selektionsdruck zwar besser als die Ergebnisse mit mehr Eltern, allerdings werden damit auch bei einer Erhöhung des Selektionsdrucks keine besseren Ergebnisse erzielt. Für die anderen μ gilt nun das bereits für $n = 30$ Gesagte. Es zeichnen sich wiederum deutliche Strukturen ab, wobei für jedes μ ein zum Erreichen des besten ABFV optimales λ zu existieren scheint, das mit zunehmendem μ ebenfalls ansteigt. Für $n = 110$ zeigt sich, dass die Ackley-Funktion mit zunehmender Problemdimension immer schwieriger zu optimieren wird. Eine $(8, \lambda)$ -ES ist nun nicht mehr in der Lage für eine ausreichende Qualität des ABFV zu sorgen. Stattdessen gleicht der Verlauf der Kurve dem der für $\mu = 5$ erhaltenen Kurve, wobei die mit $\mu = 8$ erhaltenen Ergebnisse aber besser sind.

Bei den multimodalen Funktionen wird immer wieder auf die Art der Initialisierung hingewiesen. Wird die Startpopulation gleichverteilt über dem gesamten Wertebereich initialisiert, so scheinen diejenigen Strategien bevorteilt, die mit einer größeren Anzahl Eltern arbeiten, da dann die Wahrscheinlichkeit größer ist, einen Elter in der Nähe des Optimums zu platzieren. Zudem wird vereinzelt behauptet, dass sich dann die Rekombination besonders positiv bemerkbar macht. Um diesen Effekt abschätzen zu können, wiederholen wir hier exemplarisch die Experimente auf der 30-dimensionalen Ackley-Funktion mit einer Initialisierung aller Individuen im Punkt $x_i = 20$ und $\sigma_i = 1$. Die Abbildung 8.8 zeigt die entsprechenden Ergebnisse, wobei im Vergleich zur Abbildung 8.5 kaum Unterschiede vorhanden sind. Besonders auffallend ist lediglich die Kurve für $\mu = 5$. Diese hat nun den zuvor schon für größeres n gesehenen flachen Verlauf. Selbstverständlich werden die Ergebnisse insgesamt anders ausfallen, wenn bei der Initialisierung eine sehr kleine Schrittweite verwendet wird. Dann wird die ES jedoch nicht mehr aus einem lokalen Optimum herausfinden. Dieser Effekt soll hier aber nicht näher untersucht werden, da bei der später zu untersuchenden dynamischen Optimierung eine Stagnation (mit immer kleiner werdender Populationsdiversität) in einem lokalen Optimum kaum auftreten kann, sodass sich dort eine Initialisierung in einem Startpunkt ohnehin nur unwesentlich bemerkbar machen sollte. Wir halten also auch für eine Ini-

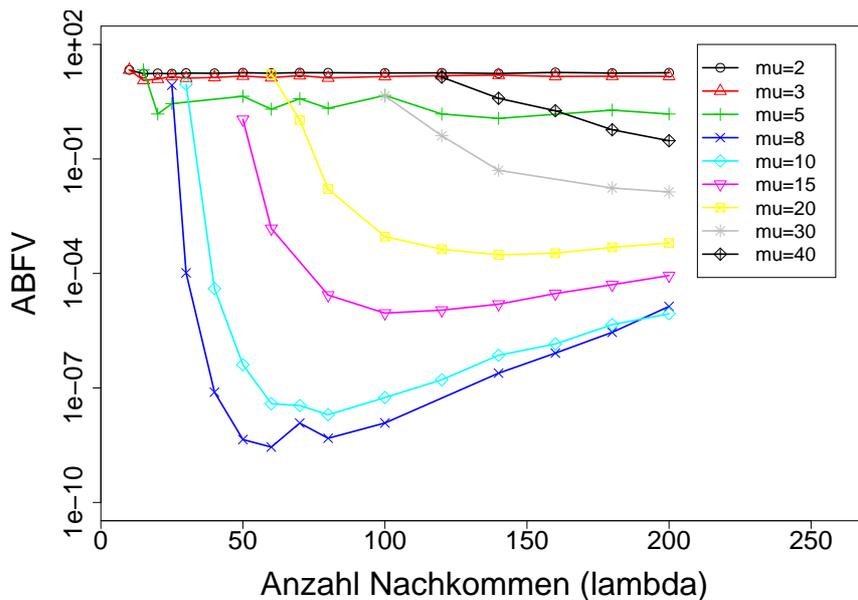


Abbildung 8.8: ABFV einer (μ, λ) -ES bei Initialisierung der Startpopulation mit $x_i = 20$ und $\sigma_i = 1$ auf der 30-dimensionalen statischen Ackley-Funktion.

tialisierung der Startpopulation in genau einem Punkt fest, dass die Auswirkungen im statischen Fall nur marginal sind.

Eine etwas andere Darstellung der mit mehreren Kombinationen der Populationsgrößen erreichbaren Ergebnisse auf der statischen Ackley-Funktion zeigen die Abbildungen 8.9 bis 8.11. Dort ist für verschiedene Problemdimensionen jeweils der ABFV in Abhängigkeit vom Selektionsdruck aufgetragen. Für jedes λ ist auch hier ein Wert für μ erkennbar, der den ABFV minimiert.

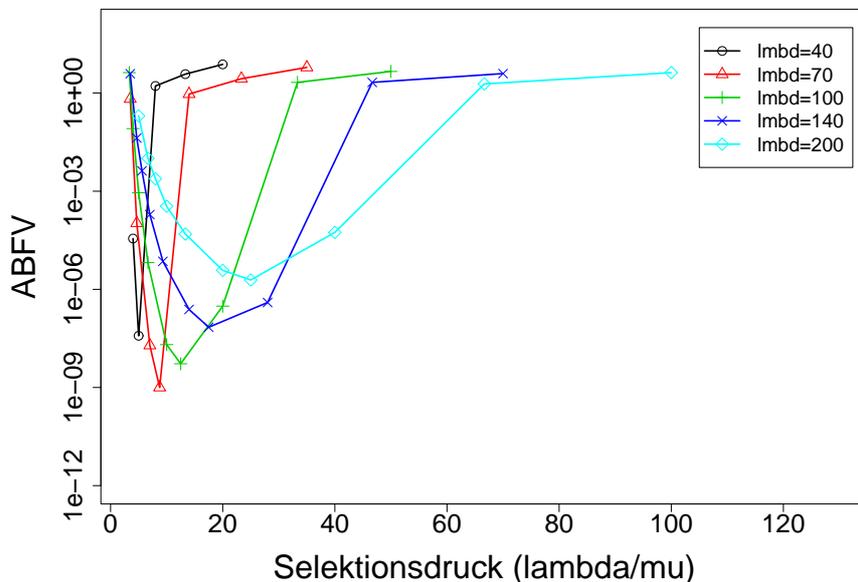
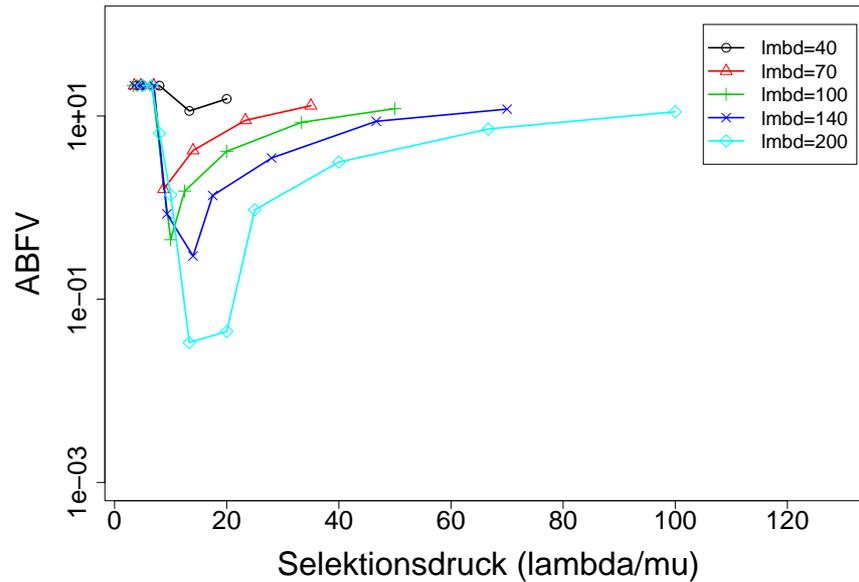
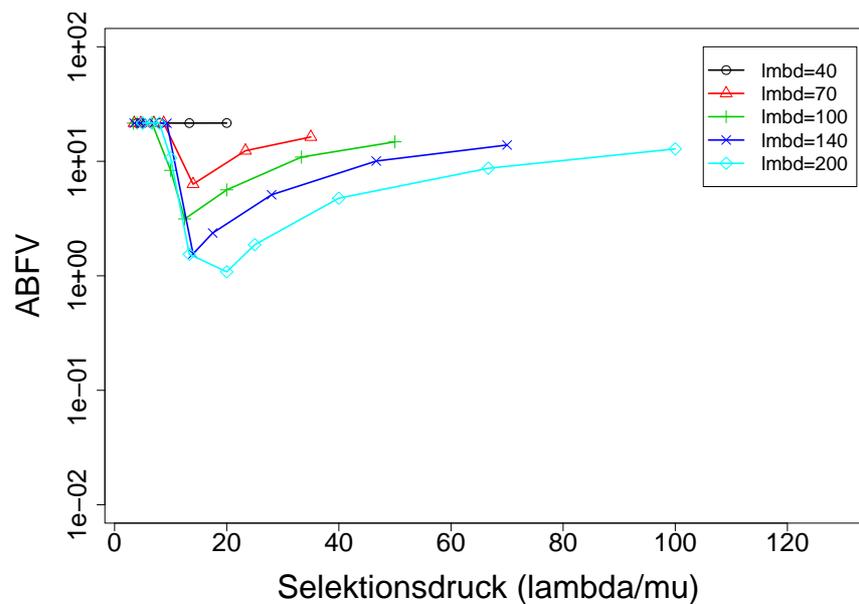


Abbildung 8.9: ABFV einer (μ, λ) -ES auf der 30-dimensionalen statischen Ackley-Funktion.

Was das Konvergenzverhalten der ES auf der statischen Ackley-Funktion anbelangt, weisen die Ergebnisse zu den auf dem Kugelmodell erhaltenen insgesamt eine große Ähnlichkeit auf. Für beide

Abbildung 8.10: ABFV einer (μ, λ) -ES auf der 110-dimensionalen statischen Ackley-Funktion.Abbildung 8.11: ABFV einer (μ, λ) -ES auf der 150-dimensionalen statischen Ackley-Funktion.

Funktionen liegt der Bereich der beginnenden Divergenz für eine $(15, 100)$ -ES bei $n \approx 100$. Für beide Funktionen konnte eine $(10, 100)$ -ES signifikant bessere Ergebnisse erzielen. Des Weiteren sind die Ergebnisse verschiedener ES mit unterschiedlichen Populationsgrößen vom prinzipiellen Aussehen sehr ähnlich. Größere Unterschiede sind für kleine μ erkennbar. Hier macht sich insbesondere für steigende Problemdimensionen die erhöhte Schwierigkeit aufgrund der Multimodalität bemerkbar. Ansonsten sind für jede Problemdimension bei fester Wahl eines Populationsparameters wiederum *optimale* Werte für den anderen Populationsparameter feststellbar, mit denen der ABFV minimiert wird.

Gemäß Definition 1.4 ist die dynamische Variante der generalisierten Ackley-Funktion wie folgt definiert:

$$f_{Ackley}(x, x^*(t)) = -a \cdot \exp\left(-b \cdot \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - x_i^*(t))^2}\right) - \exp\left(\frac{1}{n} \sum_{i=1}^n \cos(c(x_i - x_i^*(t)))\right) + a + e,$$

wobei $a = 20, b = 0.2, c = 2\pi, -20 \leq x_i \leq 30$.

Die Abbildung 8.12 zeigt, dass wir bei der Optimierung der dynamischen Ackley-Funktion mit einer Bewegung in einer Dimension eine ähnliche Situation bzgl. des Bereichs der beginnenden Divergenz erhalten wie im statischen Fall. Der teils sehr eng am Optimum gefasste Startpunkt der ES kann bereits ab einer Problemdimension $n \approx 50$ nur noch schwer gehalten werden. Dennoch erfolgt eine Divergenz auch hier erst ab $n \approx 100$.

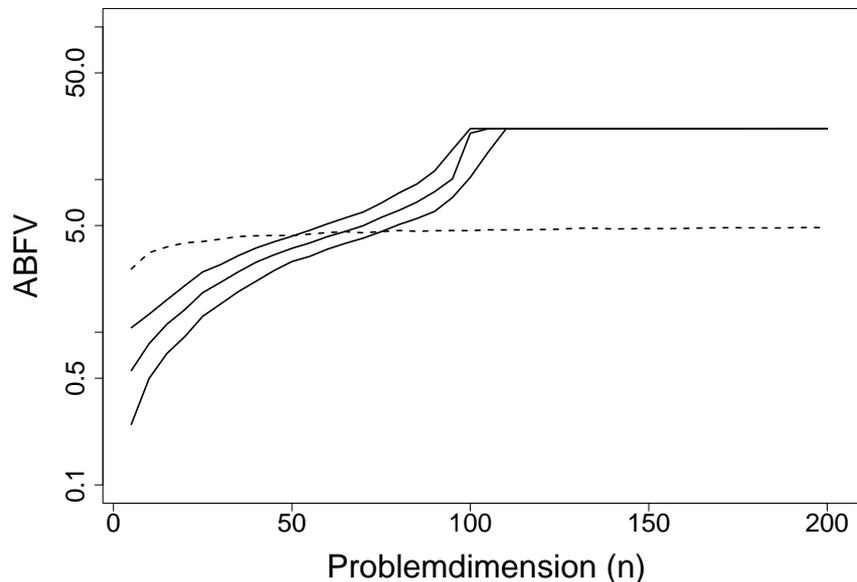


Abbildung 8.12: Konvergenzverhalten einer (15, 100)-ES auf der dynamischen Ackley-Funktion in Abhängigkeit von der Problemdimension. Neben dem ABFV sind noch die 0.05- und 0.95-Quantile dargestellt. Die gestrichelte Linie zeigt darüber hinaus den durchschnittlichen Startwert der ES. Das Optimum bewegte sich jede Generation in einer Dimension um die Entfernung $s = 1.0$.

Wird der Bewegungstyp III verwendet, ändert sich das Bild enorm. In der Abbildung 8.13 ist scheinbar eine ab $n \approx 60$ beginnende Divergenz zu erkennen. Allerdings macht sich hier der stark beschränkte Wertebereich der Ackley-Funktion bemerkbar. Die Form der Bewegung ist nun so schwierig, dass sich die ES bereits frühzeitig dem maximalen Funktionswert $a + e$ annähert. Es ist davon auszugehen, dass für einen Anwender diese erzielten Ergebnisse unbefriedigend wären. Für die hier durchgeführte Untersuchung blicken wir jedoch zusätzlich noch auf die Entfernung zum Optimum. Dabei zeigt sich, dass diese auch für ein festes n mit $60 < n < 100$ relativ konstant bleibt, d.h. die ES divergiert wiederum erst ab $n \approx 100$. Von diesem Standpunkt aus betrachtet ergeben sich für die Ackley-Funktion nur wenig Unterschiede zum Verhalten auf dem Kugelmodell.

Wir sehen deshalb davon ab, weitere Experimente auf der Ackley-Funktion durchzuführen. Stattdessen konzentrieren wir uns im Weiteren auf die Rastrigin-Funktion.

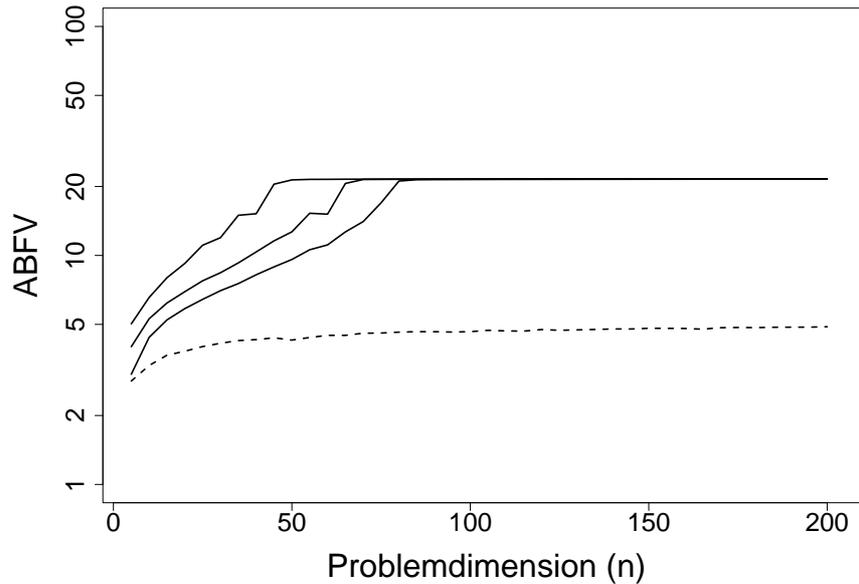


Abbildung 8.13: Konvergenzverhalten einer $(15, 100)$ -ES auf der dynamischen Ackley-Funktion in Abhängigkeit von der Problemdimension. Neben dem ABFV sind noch die 0.05- und 0.95-Quantile dargestellt. Die gestrichelte Linie zeigt darüber hinaus den durchschnittlichen Startwert der ES. Das Optimum bewegte sich jede Generation in einer schrägen Diagonalen um die Entfernung $s = 1.0$.

8.2 Rastrigin-Funktion

Wir betrachten zunächst wieder das Konvergenzverhalten einer $(15, 100)$ -ES auf der statischen Funktion. Die Abbildung 8.14 zeigt, dass auch für die Rastrigin-Funktion ein ähnliches Verhalten zu erkennen ist. Für kleine Problemdimensionen $n < 100$ kann die ES Fortschritte erzielen, wobei

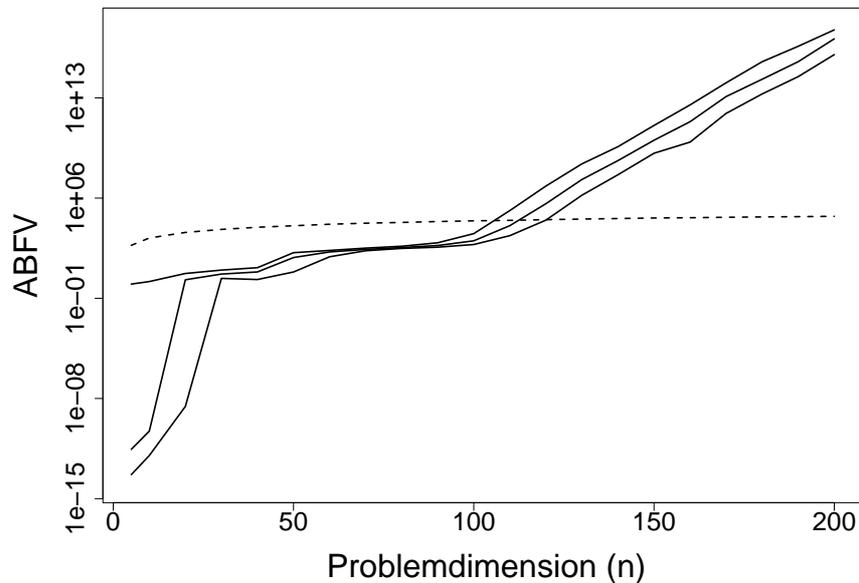


Abbildung 8.14: Konvergenzverhalten einer $(15, 100)$ -ES auf der statischen Rastrigin-Funktion in Abhängigkeit von der Problemdimension. Neben dem ABFV sind noch die 0.05- und 0.95-Quantile dargestellt. Die gestrichelte Linie zeigt darüber hinaus den durchschnittlichen Startwert der ES.

diese wegen der starken Multimodalität der Funktion wesentlich kleiner sind. Ab dem Bereich $n \approx 100$ fängt die ES jedoch auch auf der Rastrigin-Funktion an zu divergieren.

Wie schon bei den beiden anderen Funktionen sind auch hier für andere Initialisierungen der Objektvariablen und Schrittweiten sowie der Parametrisierungen der Strategieparameter keine besonderen Abweichungen vom sonstigen Verhalten erkennbar. Deshalb kommen wir an dieser Stelle direkt zu der Darstellung der Ergebnisse mit unterschiedlichen Populationsgrößen, aber identischen sonstigen Einstellungen der Standard-ES. Zunächst beantwortet die Abbildung 8.15 die Frage, ob

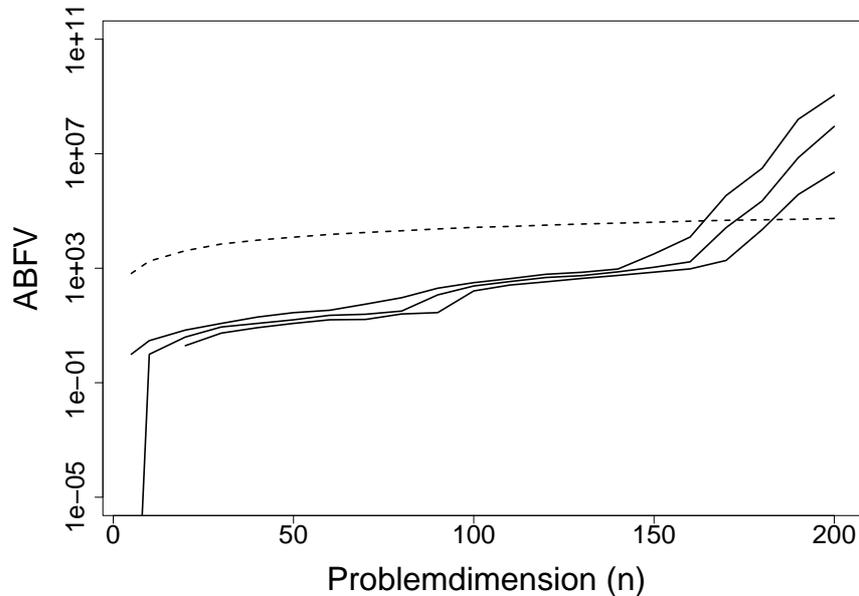


Abbildung 8.15: Konvergenzverhalten einer (10, 100)-ES auf der statischen Rastrigin-Funktion in Abhängigkeit von der Problemdimension. Neben dem ABFV sind noch die 0.05- und 0.95-Quantile dargestellt. Die gestrichelte Linie zeigt darüber hinaus den durchschnittlichen Startwert der ES.

auch hier eine (10, 100)-ES bessere Ergebnisse liefert. Dies ist eindeutig zu bejahen. Der Bereich der beginnenden Divergenz der ES hat sich nun nach $n > 150$ verschoben. Für $n < 100$ sind die Unterschiede in den Funktionswerten aber nicht sehr groß. Auch die (10, 100)-ES stagniert in der Regel frühzeitig in einem lokalen Optimum gleicher Größenordnung. Die veränderte Anzahl der Eltern wirkt sich also hierfür nicht positiv aus.

Die Abhängigkeit des ABFV von der Anzahl der Nachkommen sowie der Problemdimension ist in der Abbildung 8.16 dargestellt. Für die meisten Problemdimensionen fällt der ABFV mit zunehmender Anzahl Nachkommen immer besser aus, wobei die ES dann alle in einem lokalen Optimum gleicher Qualität stagnieren. Hier tritt also zum ersten Mal eine größere Abweichung zu den beiden anderen Funktionen auf. Die Abweichungen werden noch deutlicher, wenn wir uns die Abbildung 8.17 betrachten, in der der ABFV in Abhängigkeit von der Anzahl der Eltern und der Problemdimension aufgetragen ist. Während es bei den beiden anderen Funktionen einen eindeutigen Wert für μ gibt, der den ABFV minimiert, gilt dies hier nur für $n \in \{5, 10\}$. Aufgrund der Stagnation in einem lokalen Optimum gleicher Qualität lässt für $n \geq 30$ zwischen den mit den einzelnen Strategien erreichten ABFV nur geringe Unterschiede auftreten. In vielen Fällen divergiert die $(\mu, 100)$ -ES sogar. In der Abbildung 8.16 ist noch der für $n = 10$ bei 80 bzw. 90 Nachkommen erkennbare Ausreißer erwähnenswert. Hier ist zu berücksichtigen, dass ab diesem Bereich die ES immer besser in der Lage ist, das lokale Optimum zu verlassen, in dem die ES mit kleinerer Anzahl

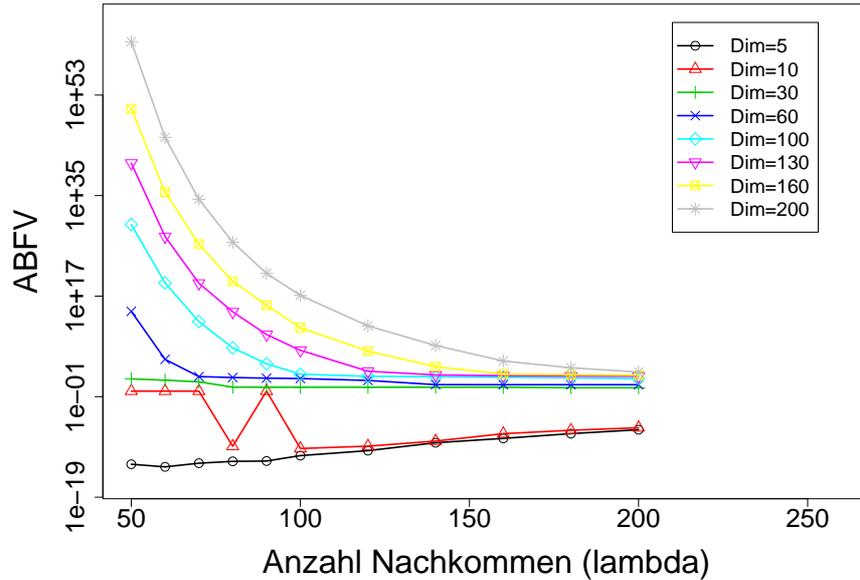


Abbildung 8.16: ABFV einer $(15, \lambda)$ -ES auf der statischen Rastrigin-Funktion in Abhängigkeit von der Problemdimension und der Anzahl Nachkommen.

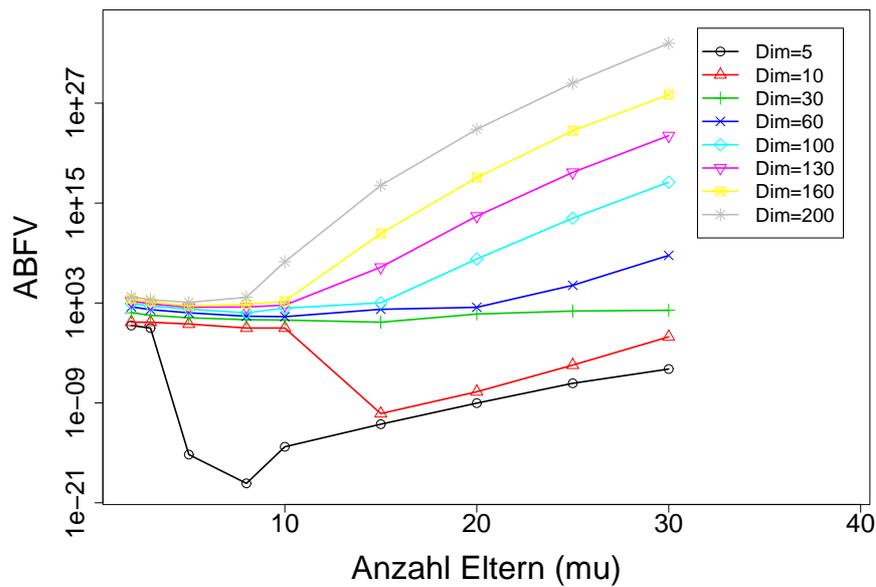


Abbildung 8.17: ABFV einer $(\mu, 100)$ -ES auf der statischen Rastrigin-Funktion in Abhängigkeit von der Problemdimension und der Anzahl Eltern.

Nachkommen frühzeitig stecken bleibt. Eine ES mit 15 Eltern benötigt einen gewissen Mindestselektionsdruck, um ausreichende Ergebnisse zu erzielen. Der Mindestselektionsdruck liegt hier bei $\lambda/\mu \approx 90/15 = 6$, wobei die einzelnen Durchläufe für die Strategien in diesem Bereich stärkere Schwankungen aufweisen. Wir können festhalten, dass es unter den gegebenen Versuchsbedingungen vorteilhafter ist, die ES mit mehr Nachkommen auszuführen, um den Mindestselektionsdruck auf keinen Fall zu unterschreiten.

Die Abbildungen 8.18 bis 8.22 zeigen für fünf ausgewählte Problemdimensionen den ABFV der einzelnen Kombinationen von μ und λ . Die hier gewählte Darstellungsart mit der auf der x -Achse

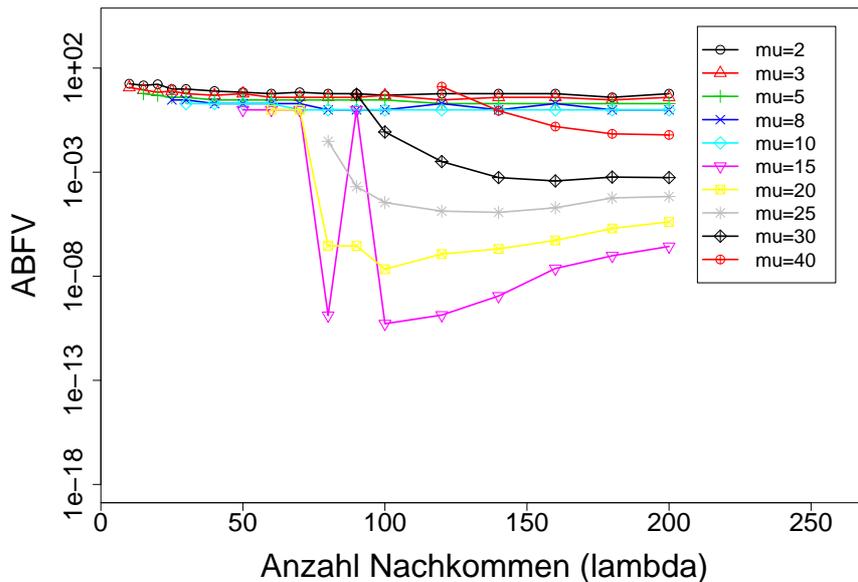


Abbildung 8.18: ABFV einer (μ, λ) -ES auf der 10-dimensionalen statischen Rastrigin-Funktion.

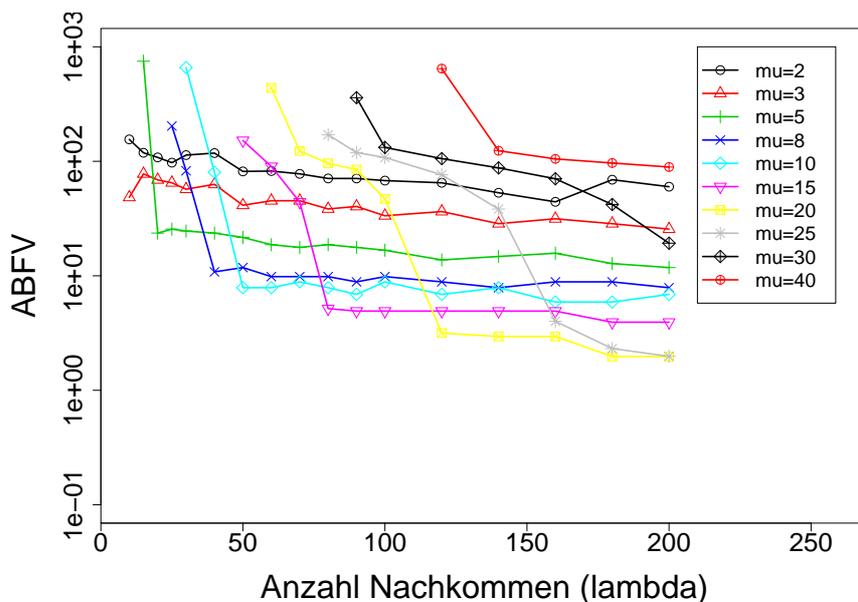


Abbildung 8.19: ABFV einer (μ, λ) -ES auf der 30-dimensionalen statischen Rastrigin-Funktion.

aufgetragenen Anzahl der Nachkommen lässt den eindeutigen Schluss zu, dass für den erreichbaren ABFV im Wesentlichen die geeignete Anzahl Eltern wichtig ist. Daneben muss lediglich gewährleistet sein, dass der Selektionsdruck einen Mindestwert von ca. 6 nicht unterschreitet. Diese Folgerung resultiert aus dem für jedes μ ab einem solchen Selektionsdruck horizontalen Verlauf der Kurven. Zu bemerken bleibt, dass sich für die einzelnen Problemdimensionen die Werte für μ unterscheiden, für die der beste ABFV erzielt wird. Entgegen einer ersten Vermutung steigen diese Werte nicht mit wachsender Problemdimension an. Wie die Tabelle 8.1 zeigt, fällt dieser Wert bis auf $n = 10$ mit zunehmender Problemdimension sogar. Eine Erklärung für dieses Verhalten steht noch aus.

Was das Konvergenzverhalten der ES auf der statischen Rastrigin-Funktion anbelangt, sind die Ergebnisse insgesamt mit den auf den beiden anderen Funktionen erhaltenen vergleichbar. Auch

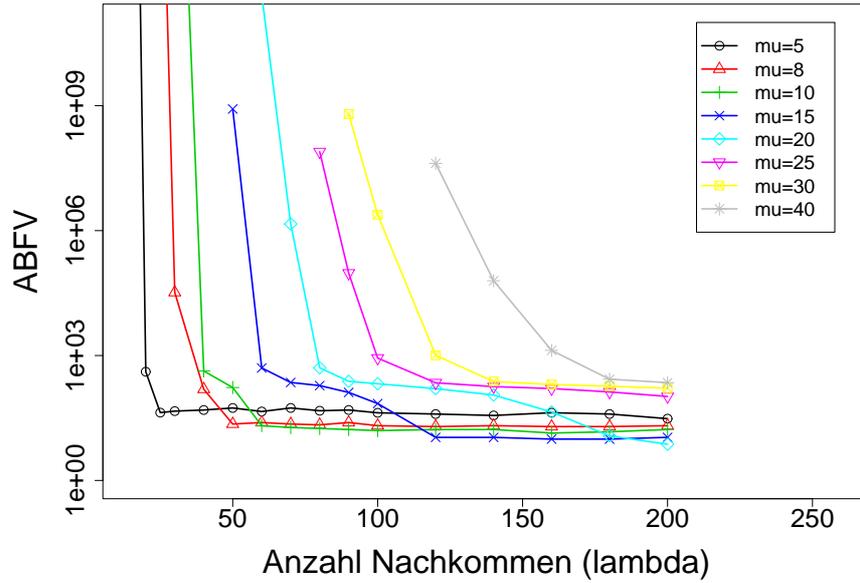


Abbildung 8.20: ABFV einer (μ, λ) -ES auf der 50-dimensionalen statischen Rastrigin-Funktion.

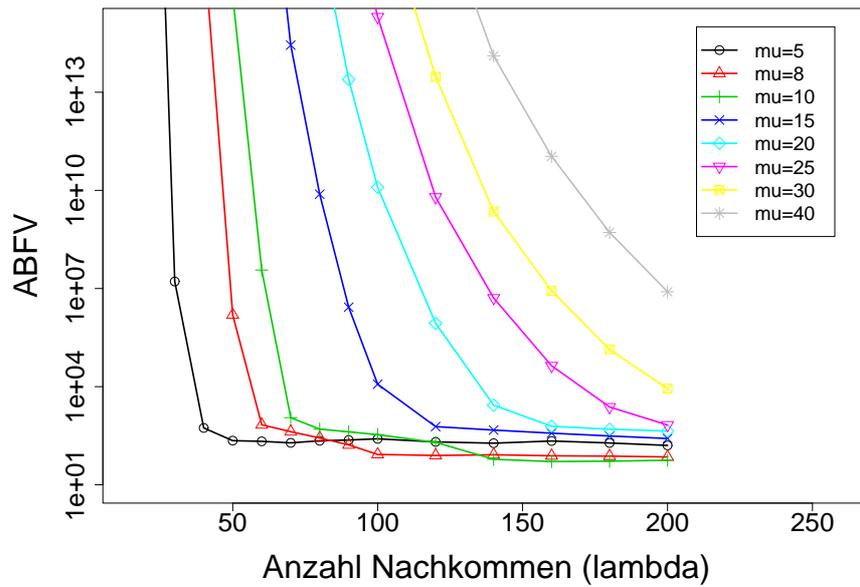


Abbildung 8.21: ABFV einer (μ, λ) -ES auf der 110-dimensionalen statischen Rastrigin-Funktion.

Tabelle 8.1: Die auf der statischen Rastrigin-Funktion zur Erzielung des besten ABFV benötigte Anzahl μ Eltern in Abhängigkeit von der Problemdimension n .

n	10	30	50	110	150
μ	15	20	20	10	8

für die Rastrigin-Funktion liegt der Bereich der beginnenden Divergenz für eine $(15, 100)$ -ES bei $n \approx 100$. Ebenso konnte eine $(10, 100)$ -ES diesen Bereich wesentlich nach hinten verschieben. Größere Unterschiede treten jedoch bei den Ergebnissen verschiedener ES mit unterschiedlichen Populationsgrößen auf. Wie zuvor beschrieben, hängen hier die erreichbaren Ergebnisse im Wesentlichen

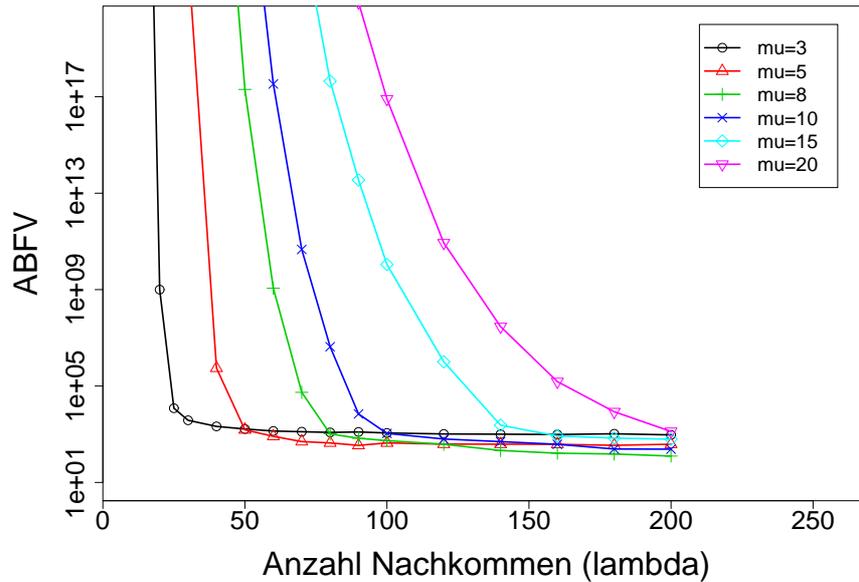


Abbildung 8.22: ABFV einer (μ, λ) -ES auf der 150-dimensionalen statischen Rastrigin-Funktion.

von der Anzahl der Eltern ab.

Auch für die Rastrigin-Funktion wollen wir noch einmal ein Experiment mit einer einheitlichen Startpopulation durchführen. Wiederum wählen wir als Initialisierung $x_i = 20$ und $\sigma_i = 1$. Vergleicht man die Ergebnisse in der Abbildung 8.23 mit denen aus der Abbildung 8.19 so ist

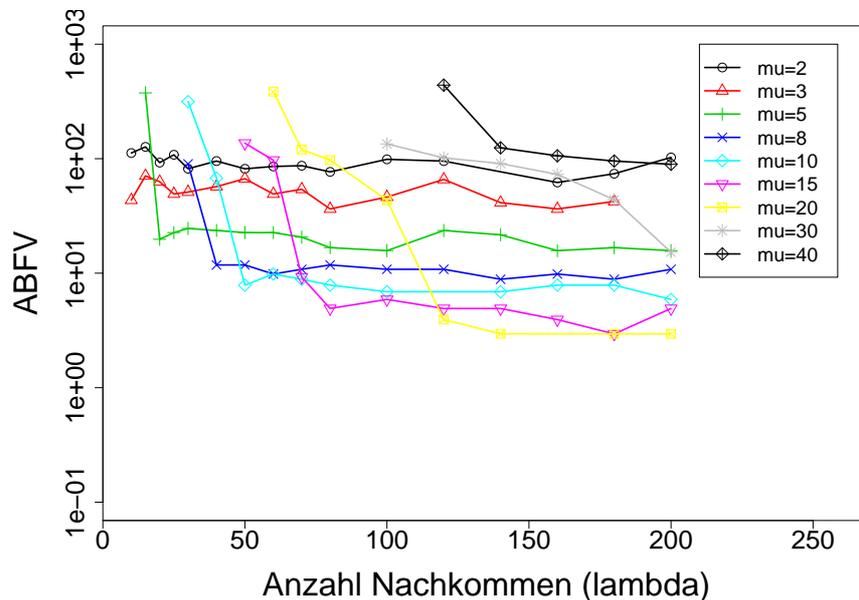


Abbildung 8.23: ABFV einer (μ, λ) -ES bei Initialisierung der Startpopulation mit $x_i = 20$ und $\sigma_i = 1$ auf der 30-dimensionalen statischen Rastrigin-Funktion.

festzustellen, dass die Differenzen nur gering sind. Während die Kurven für $\mu \in \{2, 3, 5\}$ wieder frühzeitig relativ flach verlaufen, flachen die Kurven für die meisten anderen μ erst recht spät ab. Insbesondere wird der beste ABFV auch hier mit 20 Eltern erzielt. Für die einzelnen Strategien ist dabei auch der jeweils erzielte absolute Wert des ABFV in beiden Situationen ziemlich gleich.

Rein spekulativ ist es, ob für $\mu = 30$ mit zunehmender Anzahl Nachkommen vielleicht noch bessere Ergebnisse erreicht werden können.

Im dynamischen Fall beginnen wir wieder mit der Darstellung des Konvergenzverhaltens einer (15, 100)-ES in Abhängigkeit von der Problemdimension n . Die Abbildung 8.24 zeigt den ABFV

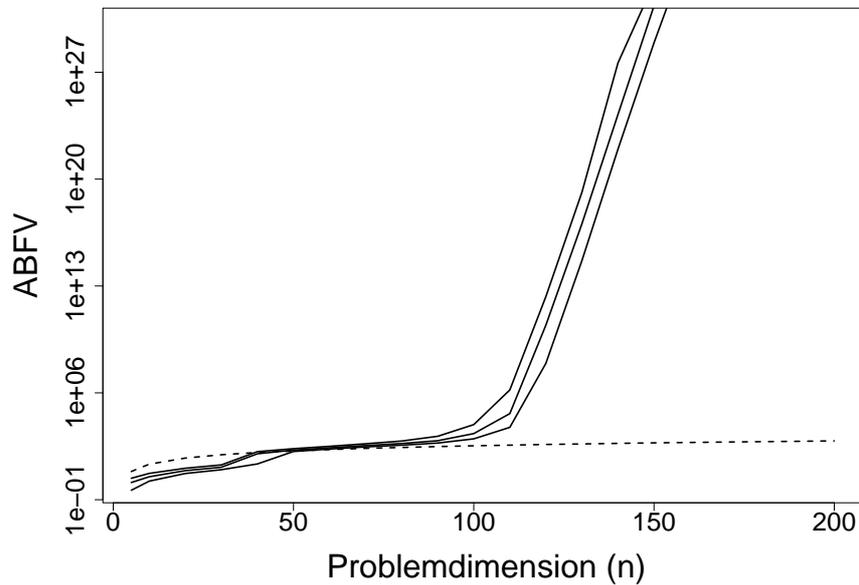


Abbildung 8.24: Konvergenzverhalten einer (15, 100)-Standard-ES auf der dynamischen Rastrigin-Funktion in Abhängigkeit von der Problemdimension n . Neben dem ABFV sind noch die 0.05- und 0.95-Quantile dargestellt. Die gestrichelte Linie zeigt darüber hinaus den durchschnittlichen Startwert der ES. Das Optimum bewegte sich jede Generation in einer Dimension um die konstante Entfernung $s = 1.0$.

für den Fall, dass sich das Optimum in jeder Generation in einer Dimension mit der Stärke $s = 1.0$ bewegt. Der Bereich der beginnenden Divergenz liegt wieder bei $n \approx 100$. Das prinzipielle Aussehen für kleinere Bewegungsstärken ($s = 0.1$) ist ähnlich, wobei dann natürlich die erreichten Funktionswerte etwas besser sind. Demgegenüber sind die von der ES beim Bewegungstyp III erreichten Funktionswerte schlechter. Allerdings ist der Bereich des Übergangs zur Divergenz identisch wie beim Bewegungstyp I. Was das Konvergenzverhalten anbelangt sind die Ergebnisse auf der dynamischen Rastrigin-Funktion ähnlich denen auf den beiden anderen Funktionen.

Auch für die Rastrigin-Funktion wurden Experimente bzgl. der Auswirkungen unterschiedlicher Dynamikfrequenzen durchgeführt. Dabei wurden hier wie auch bei der Ackley-Funktion die auf dem Kugelmodell erzielten Ergebnisse bestätigt. Mit zunehmender Dynamikfrequenz — bei gleicher zurückgelegter Gesamtentfernung — fallen die erreichten Funktionswerte immer schlechter aus. Wir können deshalb an dieser Stelle auf eine detaillierte Darstellung dieser Ergebnisse verzichten.

Die Abbildungen 8.25 bis 8.29 zeigen die Ergebnisse verschiedener (μ, λ) -ES für ausgewählte Problemdimensionen. Für $n = 5$ ergibt sich eine klare Einordnung der einzelnen Strategien. Wiederum ist die Anzahl der Nachkommen nicht entscheidend, soweit sie eine gewisse Mindestanzahl nicht unterschreiten. Die einzelnen Strategien können dem Optimum in Abhängigkeit von der Anzahl der Eltern in einem gewissen Abstand folgen. Mit zunehmendem μ steigt auch der jeweilige ABFV an. Diese einfachen Strukturen verschwinden mit wachsender Problemdimension. Für $n = 30$ existiert für jedes μ eine bestimmte Anzahl Nachkommen, mit der der ABFV minimiert wird. Demgegenüber

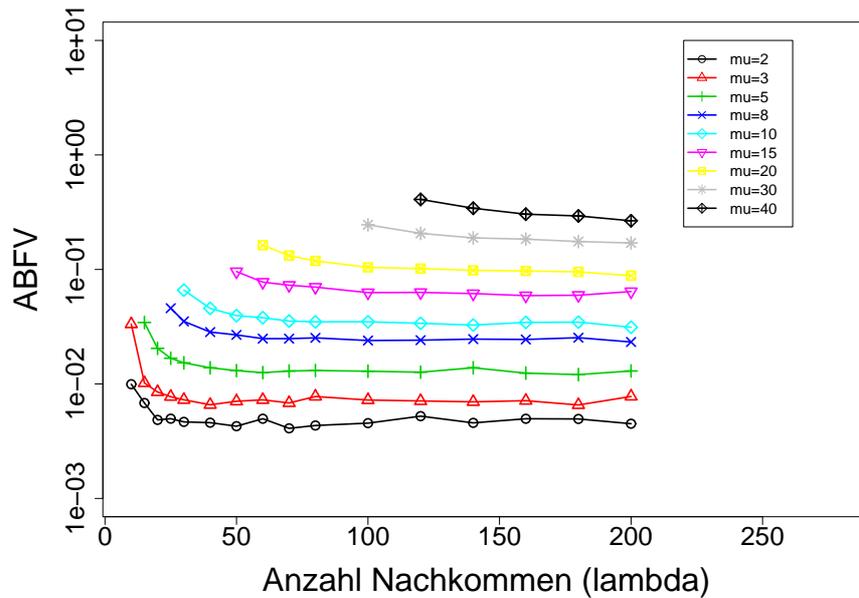


Abbildung 8.25: ABFV einer (μ, λ) -ES auf der 5-dimensionalen dynamischen Rastrigin-Funktion. Das Optimum bewegte sich in einer Dimension mit einer Gesamtentfernung von 50. Die Anzahl der Funktionsauswertungen beträgt 50.000.

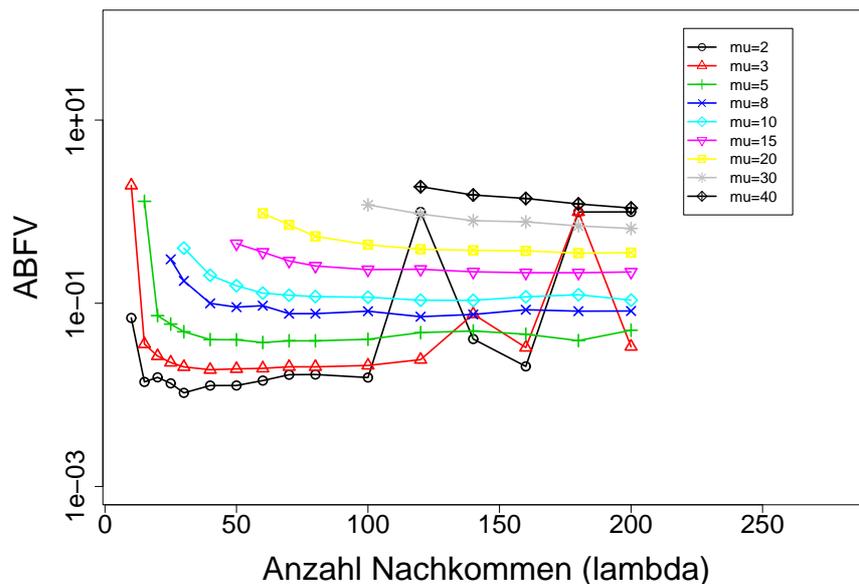


Abbildung 8.26: ABFV einer (μ, λ) -ES auf der 10-dimensionalen dynamischen Rastrigin-Funktion. Das Optimum bewegte sich in einer Dimension mit einer Gesamtentfernung von 100. Die Anzahl der Funktionsauswertungen beträgt 100.000.

gleichensich die Kurven für $n \in \{50, 90\}$ immer mehr an. Dies gipfelt für $n = 90$ darin, dass bis auf $\mu \geq 20$ für alle μ ein Wert für λ gefunden werden kann, der zu einem ABFV führt, der dicht am besten hier gemessenen Wert liegt. Wie die Abbildung andeutet, werden sich mit wachsendem λ auch die ABFV für $\mu \geq 20$ diesem Wert nähern. D.h. zwischen den einzelnen Strategien existieren nicht mehr so große Unterschiede.

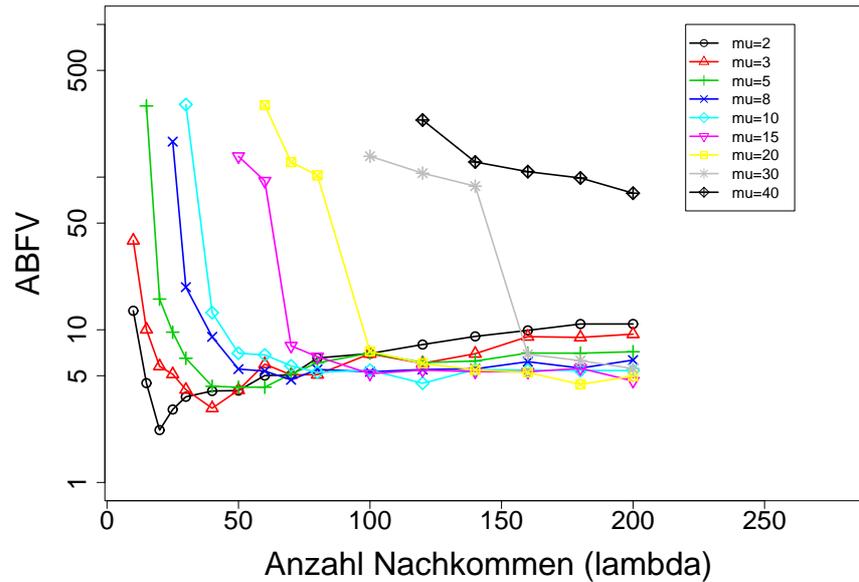


Abbildung 8.27: ABFV einer (μ, λ) -ES auf der 30-dimensionalen dynamischen Rastrigin-Funktion. Das Optimum bewegte sich in einer Dimension mit einer Gesamtentfernung von 300. Die Anzahl der Funktionsauswertungen beträgt 300.000.

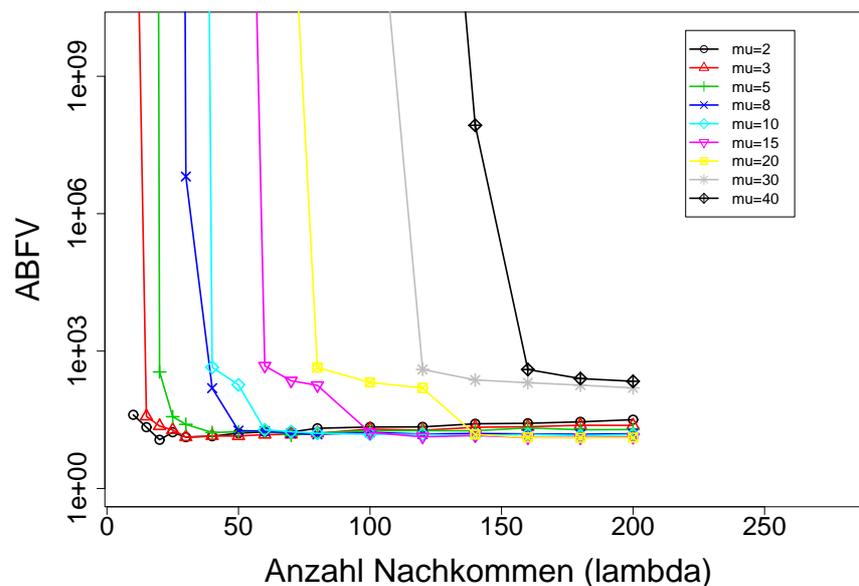


Abbildung 8.28: ABFV einer (μ, λ) -ES auf der 50-dimensionalen dynamischen Rastrigin-Funktion. Das Optimum bewegte sich in einer Dimension mit einer Gesamtentfernung von 500. Die Anzahl der Funktionsauswertungen beträgt 500.000.

Die bisherigen Experimente auf der dynamischen Rastrigin-Funktion bezogen sich alle auf den Bewegungstyp I. Bereits bei dieser Bewegung in nur einer Dimension zeigten sich wesentliche Unterschiede zu der Optimierung der beiden anderen Funktionen bei gleichem Bewegungstyp. Bei einer Bewegung des Typs III sind dementsprechend ebenfalls wesentliche Unterschiede zu erwarten. Größere Abweichungen erwarten wir wegen des schwierigeren Bewegungstyps auch im Verhältnis zu den Ergebnissen des Bewegungstyps I. Beim Bewegungstyp I konnten schon für kleine n Auffällig-

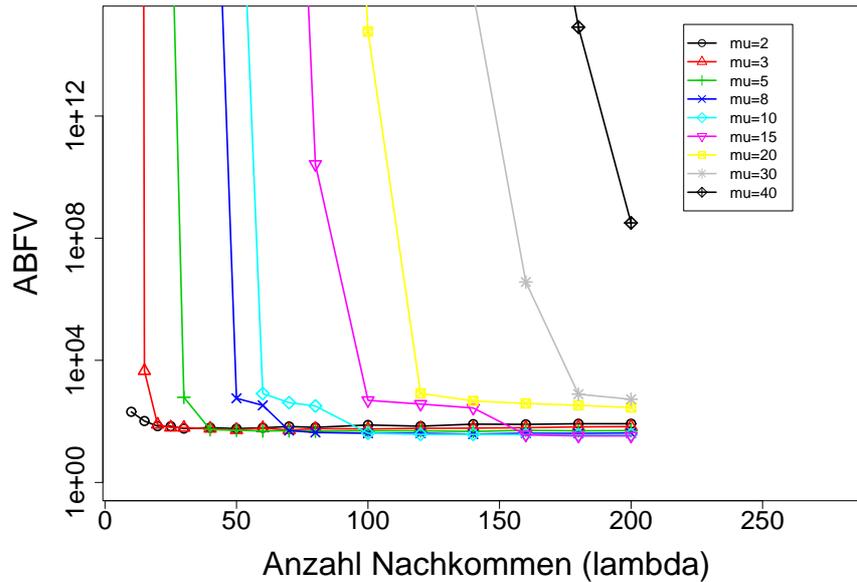


Abbildung 8.29: ABFV einer (μ, λ) -ES auf der 90-dimensionalen dynamischen Rastrigin-Funktion. Das Optimum bewegte sich in einer Dimension mit einer Gesamtentfernung von 900. Die Anzahl der Funktionsauswertungen beträgt 900.000.

keiten erkannt werden. Um einen direkten Vergleich beider Ergebnisse zu ermöglichen, werden die nachfolgenden Experimente mit den gleichen Problemdimensionen $n \in \{5, 10, 30, 50, 90\}$ durchgeführt. Die zugehörigen Ergebnisse sind in den Abbildungen 8.30 bis 8.34 dargestellt. Es fällt auf, dass die besten Ergebnisse ausschließlich mit größeren μ -Werten erreicht werden. Dies gilt bereits ab $n = 10$. Alle Strategien haben hier sowohl mit der starken Multimodalität zu kämpfen als auch mit der jetzt verschärften Dynamik. Hierfür scheint sich eine größere Diversität in der Population positiv bemerkbar zu machen.

8.3 Fazit

Die wichtigsten aus den Experimenten auf den getesteten multimodalen Funktionen gewonnenen Erkenntnisse seien hier aufgezählt:

- Die Optimierung auf der Ackley-Funktion führt zu ähnlichen Ergebnissen wie die Optimierung beim Kugelmodell, sodass die dort gemachten Aussagen auch hier zutreffen.
- Die Optimierung der Rastrigin-Funktion gestaltet sich aufgrund der stärkeren Multimodalität wesentlich schwieriger. Sehr viele Strategien bleiben im statischen Fall frühzeitig in einem lokalen Optimum stecken. Dies spiegelt sich auch in den Ergebnissen für die Optimierung im dynamischen Fall wider. Insbesondere wenn der Wert der Bewegungsfrequenz Δg hoch ist, macht sich die multimodale Struktur bemerkbar, sodass eine ES häufig in einem lokalen Optimum stecken bleibt. Allerdings ist es der ES bei entsprechender Bewegungsstärke nach einer Änderung auch hier möglich, dem Optimum zu folgen.
- Bei einer nahezu kontinuierlichen Bewegung (Δg klein) ähneln die erreichten Ergebnisse auf der Rastrigin-Funktion denen des Kugelmodells. Die Komplexität der Optimierung der multimodalen Funktion wird dann durch die Dynamik überlagert und die Eigenschaften des

Kugelmodells kommen stärker zum Tragen.

- Die beginnende Divergenz liegt für die Rastrigin-Funktion in einem ähnlichen Bereich der Problemdimension wie für die beiden anderen Funktionen. Dennoch tritt dieser Aspekt bei der Rastrigin-Funktion etwas früher auf.
- Insgesamt erweisen sich zur Optimierung der Rastrigin-Funktion im Vergleich zu den anderen beiden Testfunktionen größere Populationsgrößen als vorteilhaft.

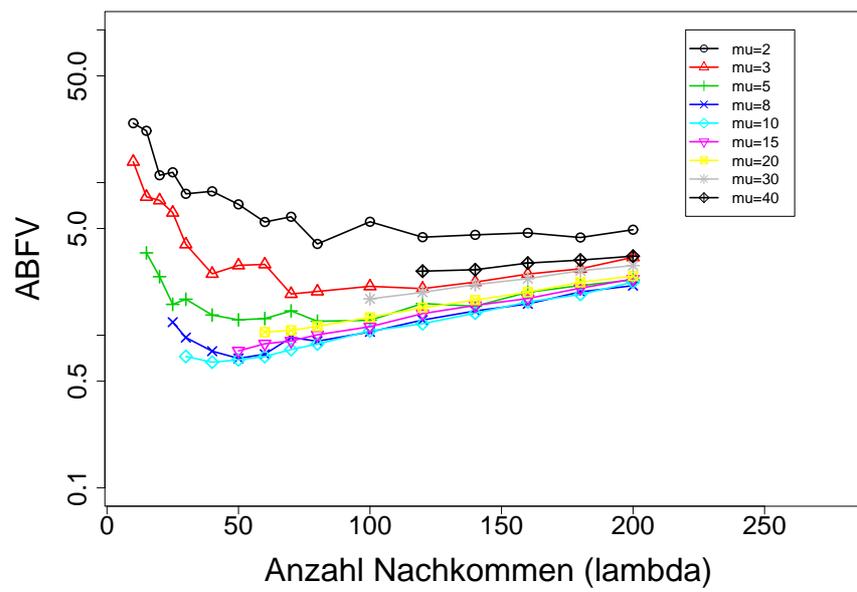


Abbildung 8.30: ABFV einer (μ, λ) -ES auf der 5-dimensionalen dynamischen Rastrigin-Funktion. Das Optimum bewegte sich in einer schrägen Diagonalen mit einer Gesamtentfernung von 50. Die Anzahl der Funktionsauswertungen beträgt 50.000.

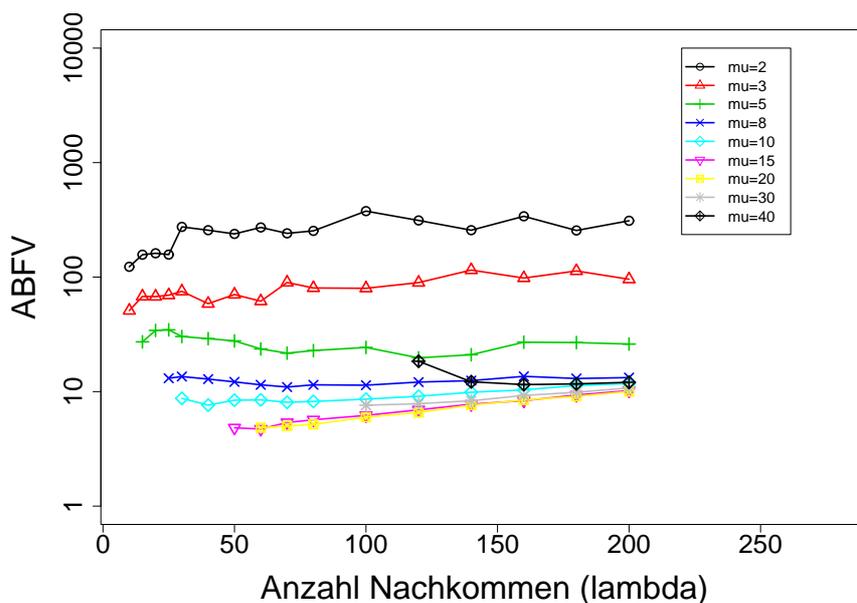


Abbildung 8.31: ABFV einer (μ, λ) -ES auf der 10-dimensionalen dynamischen Rastrigin-Funktion. Das Optimum bewegte sich in einer schrägen Diagonalen mit einer Gesamtdistanz von 100. Die Anzahl der Funktionsauswertungen beträgt 100.000.

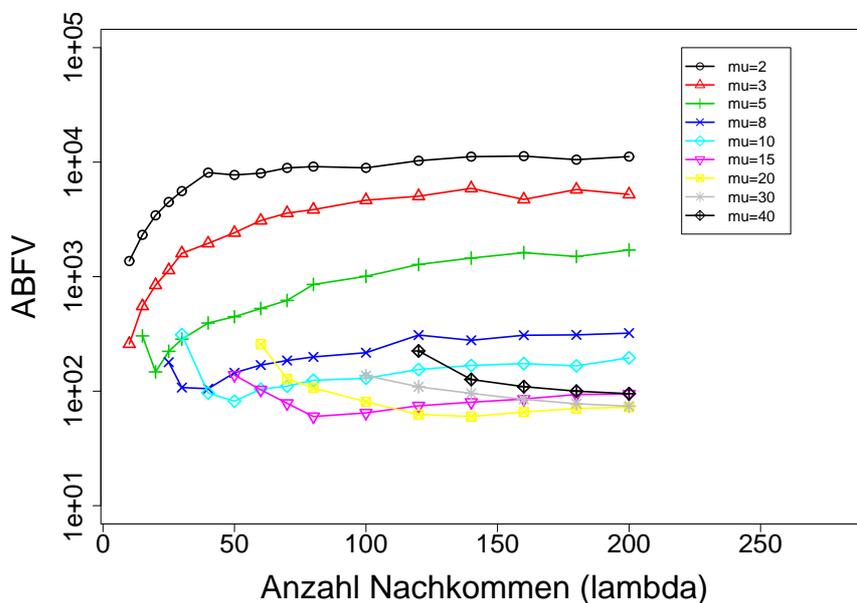


Abbildung 8.32: ABFV einer (μ, λ) -ES auf der 30-dimensionalen dynamischen Rastrigin-Funktion. Das Optimum bewegte sich in einer schrägen Diagonalen mit einer Gesamtdistanz von 300. Die Anzahl der Funktionsauswertungen beträgt 300.000.

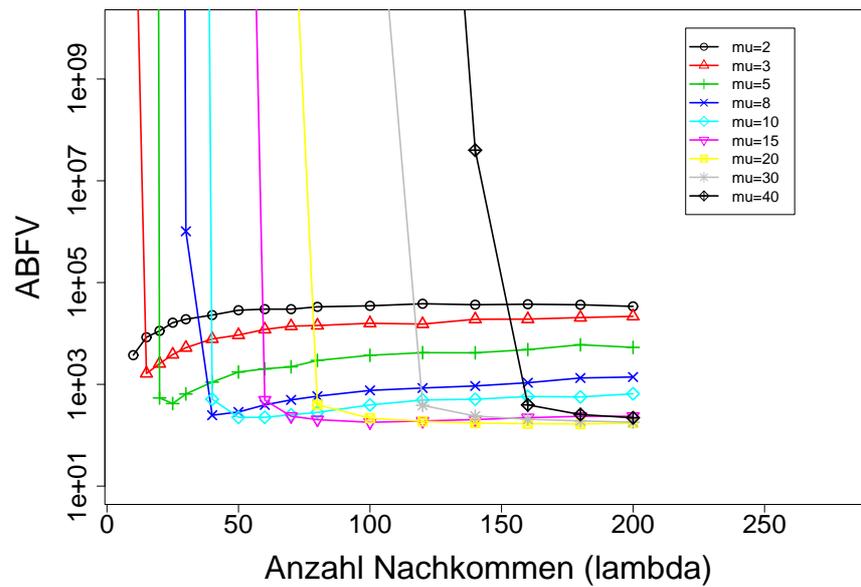


Abbildung 8.33: ABFV einer (μ, λ) -ES auf der 50-dimensionalen dynamischen Rastrigin-Funktion. Das Optimum bewegte sich in einer schrägen Diagonalen mit einer Gesamtdistanz von 500. Die Anzahl der Funktionsauswertungen beträgt 500.000.

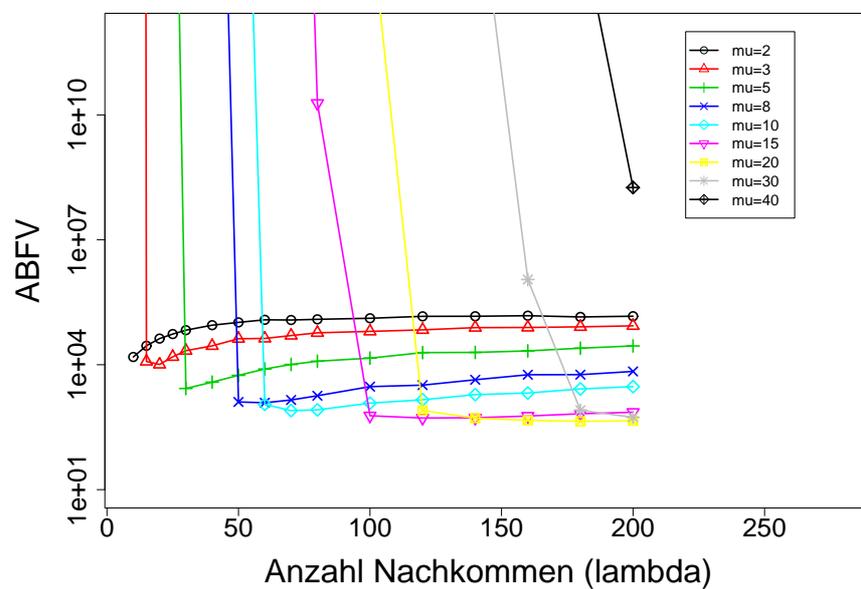


Abbildung 8.34: ABFV einer (μ, λ) -ES auf der 90-dimensionalen dynamischen Rastrigin-Funktion. Das Optimum bewegte sich in einer schrägen Diagonalen mit einer Gesamtdistanz von 900. Die Anzahl der Funktionsauswertungen beträgt 900.000.

Kapitel 9

Betrachtungen zur Diversität

9.1 Einführung

In diesem Kapitel betrachten wir das Verhalten EA in Bezug auf die Diversität innerhalb der Population. Diesem in früheren Jahren wenig Aufmerksamkeit geschenkten Aspekt evolutionärer Algorithmen widmen immer mehr Autoren eigene Beiträge. Beispielhaft seien hier Morrison (2004) und Wineberg und Oppacher (2003) genannt. Während ersterer das Hauptaugenmerk auf binäre Kodierungen legt, betrachten letztere auch reellwertige Codierungen. Als Maß der Populationsdiversität existieren demzufolge verschiedene Möglichkeiten. Beispielhaft seien hier die Folgenden aufgezählt:

Volumen: Das Volumen der Population, definiert als Produkt der Spannweiten in den Koordinaten, ist ungeeignet, da es den Wert 0 besitzt, wenn alle Individuen in einer Koordinate identisch sind.

Max. Abstand: Der maximale Abstand zwischen den einzelnen Individuen einer Population ist sehr empfindlich gegenüber einzelnen Ausreißern.

Summe der Spannweiten: Bei diesem Maß wird für jede Koordinate die Differenz des größten und kleinsten Wertes gebildet. Die Summe dieser Differenzen dient als grobes Maß zur Beurteilung der Diversität.

Mittlerer Abstand: Als mittlerer Abstand sind zwei Möglichkeiten gebräuchlich. Einerseits kann der mittlere Abstand aller Individuen zum Zentrum der Population berechnet werden. Andererseits verwendet man häufig den mittleren Abstand der Individuen untereinander.

Wie im Abschnitt 3.2 erwähnt, wird zur Beurteilung der Diversität oftmals auch die Entropie bestimmt. Dafür hat sich jedoch noch kein einheitliches Maß etabliert. Ein Grund hierfür liegt darin, dass die Entropie nur innerhalb eines definierten und beschränkten Raumes bestimmt werden kann. Wird dieser zu klein gewählt können einzelne Individuen außerhalb liegen. Wird er dagegen zu groß gewählt sind die Unterschiede der Entropie nicht aussagekräftig genug. Maße, die nur einen maximalen Abstand berücksichtigen, sind sehr anfällig gegenüber extremen Ausreißern. Liegt ein Element weit entfernt von den anderen, so wird z.B. das Volumen sehr groß, obwohl der größte Teil der Elemente einen kompakten Klumpen bilden kann.

Dass das genaue Maß keine so große Rolle spielt, zeigen Wineberg und Oppacher (2003). Insbesondere kommen sie zur Erkenntnis, dass die beiden Maße des mittleren Abstands äquivalent sind.

Aufgrund dieser Erkenntnis und weil es sich bei dem mittleren Abstand um ein häufig gebräuchtes Maß handelt, verwenden wir dieses hier wie auch schon in (Schönemann 2004a). Genauer gesagt betrachten wir den quadrierten mittleren euklidischen Abstand D^2 aller Individuen zum Populationszentrum. Dieser ist wie folgt definiert.

Definition 9.1 (Populationsdiversität D^2) Gegeben sei eine Menge $I^{(1)}, \dots, I^{(m)}$ von m Lösungen eines statischen oder dynamischen Optimierungsproblems aus den Definitionen 1.2 und 1.4.

Die j -te Komponente des Zentrums Z dieser auch Population genannten Menge ist $Z_j := \frac{1}{m} \sum_{i=1}^m I_j^{(i)}$,

d.h. der Mittelwert der j -ten Komponenten der Lösungen.

Die zugehörige Diversität der Population ist dann definiert als

$$D^2 := \left(\frac{1}{m} \sum_{i=1}^m d(I^{(i)}, Z) \right)^2,$$

wobei d der bekannte euklidische Abstand ist.

Ein ähnliches Konstrukt ist auch aus anderen Wissenschaften bekannt. So wird in der Statistik die Varianz als mittlerer Abstand der quadrierten Entfernungen zum Mittelwert berechnet. Darüber hinaus wird bei der Cluster-Einteilung darauf geachtet, dass die Elemente innerhalb eines Clusters homogen sind. Zur Bestimmung dieser Homogenität wird dort ein ähnliches Maß verwendet.

Die Verwendung des Maßes D^2 hat auch noch einen weiteren Vorteil. Die hier ermittelten empirischen Ergebnisse können als Ausgangsbasis für weitere theoretische Untersuchungen verwendet werden wie sie beispielsweise von Beyer (2001) sowie Beyer und Schwefel (2002) durchgeführt werden. Bei dieser für viele Analysen benutzten Multirekombination werden zur Erzeugung eines Nachkommen immer alle μ Eltern rekombiniert. Die Objektvariablen dieses Rekombinanten stimmen mit dem von uns benutzten Populationsmittelpunkt überein.

9.2 Kugelmodell

Im (statischen und dynamischen) Kugelmodell haben wir gesehen, dass für $n = 50$ alle Einzelläufe einer (15, 100)-ES sicher gegen das globale Optimum konvergierten. Für diese Problemdimension zeigt die Abbildung 9.1 den Verlauf des Funktionswertes und der zugehörigen Diversität für einen repräsentativen Einzellauf im dynamischen Fall. Beide Werte verlaufen während des gesamten Laufs im Gleichschritt. Eine Berechnung des zugehörigen Korrelationskoeffizienten bestätigt diesen Eindruck durch einen sehr hohen Wert ($r > 0.99$). Das gleiche Verhalten zeigt sich auch im statischen Fall, wenn die ES sicher gegen das Optimum konvergiert.

Für $n = 150$ divergierten alle Einzelläufe der (15, 100)-ES, wobei aber auch in dieser Situation beide Werte im Gleichschritt verlaufen. Demgegenüber war die (15, 100)-ES für $n = 110$ nicht in der Lage, in allen Einzelläufen zu konvergieren bzw. im dynamischen Fall dem Optimum zu folgen. Eine Idee geht dahin, in dieser Situation eine größere Abweichung zwischen dem Funktionswert und der Diversität zu vermuten. Wie die Abbildung 9.2 zeigt, stimmen der Verlauf des Funktionswertes und der Diversität aber auch in diesem Fall überein. Dies gilt sowohl für einen repräsentativen Einzellauf als auch für den Durchschnittslauf.

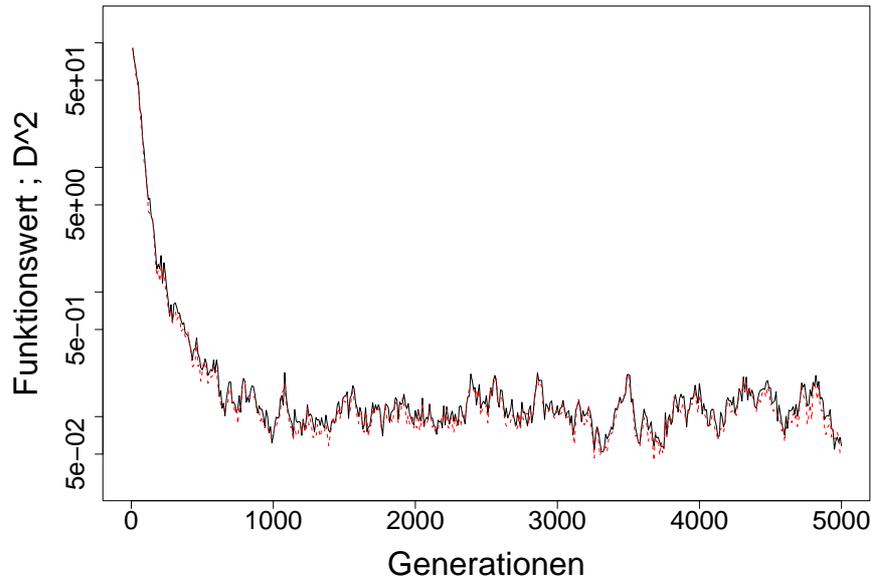


Abbildung 9.1: Funktionswert (schwarz) und Diversität (rot) eines repräsentativen Einzellaufs einer (15, 100)-ES auf dem dynamischen Kugelmodell mit $n = 50$. Das Optimum bewegte sich in einer Dimension mit einer Stärke von $s = 0.1$.

9.3 Ackley- und Rastrigin-Funktion

Die Optimierung der Ackley-Funktion ist ein wenig schwieriger als die des Kugelmodells. Dennoch zeigte sich in den vorhergehenden Kapiteln, dass die Optimierung auf beiden Funktionen ähnliche Eigenschaften besitzt. Wir möchten nun überprüfen, ob dies auch auf die Diversität zutrifft. Dies muss jedoch zunächst verneint werden. Das auf dem Kugelmodell verwendete Diversitätsmaß D^2 führt bei der Ackley-Funktion nur zu einer geringen Korrelation mit dem Funktionswert. Eine genauere Betrachtung führt jedoch dazu, statt D^2 als Maß D zu benutzen. Die sinnvolle Wahl dieses Maßes bei der Ackley-Funktion bestätigt bereits der in der Abbildung 9.3 dargestellte Verlauf der beiden Werte. Nun ergibt auch hier eine Korrelationsanalyse einen Wert $r > 0.99$. Allerdings ist hierbei zu beachten, dass die hohe Korrelation für einen Einzellauf nicht gilt, wenn die ES in einem lokalen Optimum stecken bleibt. Das Gleiche gilt auch für den Fall, dass die ES divergiert. Hier steigt das Populationsvolumen stark an, der Funktionswert ist aber nach oben beschränkt, sodass die Korrelation nur gering ist. Bei der im nächsten Abschnitt betrachteten Rastrigin-Funktion werden wir sehen, wie sich beide Werte verhalten, wenn die ES gegen ein lokales Optimum konvergiert.

Zur Berechnung der Populationsdiversität bei der Rastrigin-Funktion kehren wir zurück zum Maß D^2 . Für $n = 50$ zeigt die Abbildung 9.4 die Entwicklung des Funktionswertes und der Diversität für einen einzelnen Lauf. Während beide Werte zu Beginn im Gleichschritt verlaufen, beginnt die Diversität nach kurzer Zeit stärker zu schwanken und anschließend stark zu fallen. Während der Funktionswert noch für kurze Zeit fällt, stagniert er wenige Generationen später in einem lokalen Optimum. Die gesamte Population konzentriert sich somit innerhalb weniger Generationen vollständig in dem lokalen Optimum und ist nicht mehr in der Lage, dieses zu verlassen. Dabei spielen selbstverständlich zwei Gründe eine Rolle. Zum einen führt die Rekombination zu Nachkommen, die innerhalb der Populationshülle liegen. Zum anderen haben sich auch die Schrittweiten bereits soweit angepasst, dass ein Verlassen des lokalen Optimums nahezu ausgeschlossen ist. Die Korrelation ist in dieser Situation natürlich aufgrund des unterschiedlichen Verlaufs nur sehr gering.

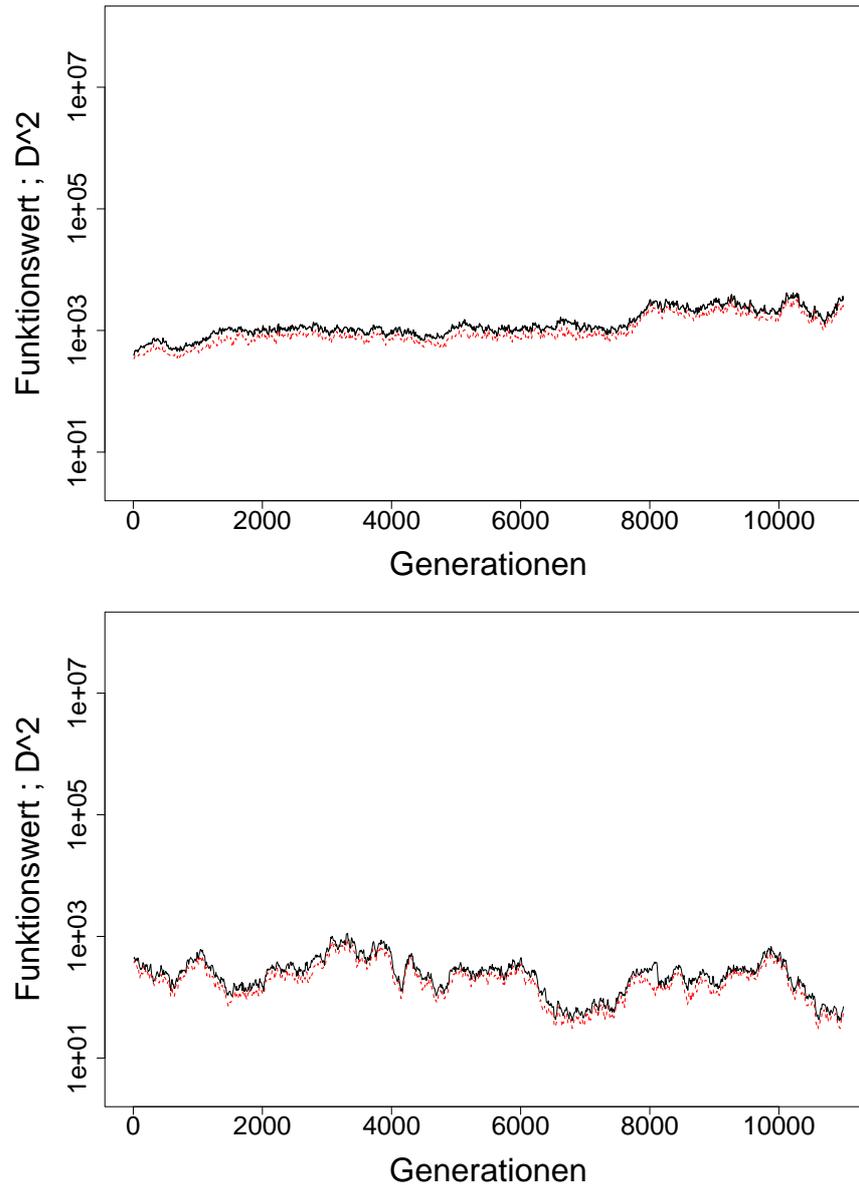


Abbildung 9.2: Funktionswert (schwarz) und Diversität (rot) des mittleren Laufs (oben) und eines repräsentativen Einzellaufs (unten) einer (15, 100)-ES auf dem dynamischen Kugelmodell mit $n=110$. Das Optimum bewegte sich in einer Dimension mit einer Stärke von $s = 0.1$.

Für $n = 100$ verdeutlicht die gleiche Abbildung die große Korrelation, wenn die ES nicht in einem lokalen Optimum stecken bleibt. Die Problemdimension $n = 100$ liegt bei der Rastrigin-Funktion genau in dem Bereich, ab dem die ES zu divergieren beginnt. Die verwendete ES ist hier offensichtlich nicht im Stande, gegen das Optimum zu konvergieren. Andererseits stagniert sie auch nicht in einem lokalen Optimum. Daraus ergeben sich Fluktuationen, die mit den Schwankungen übereinstimmen, die schon bei den anderen Funktionen in dynamischen Umgebungen zu beobachten waren. Auch für $n = 150$ ist die Korrelation wiederum sehr hoch, da hier in keinem Einzellauf eine Stagnation stattfindet, sondern die ES divergiert.

Wie dargelegt sind wir bei der Rastrigin-Funktion zur Analyse der Diversität wieder zum Maß D^2 zurückgekehrt. Es schließt sich natürlich sofort die Frage an, warum dies auch auf dem Kugel-

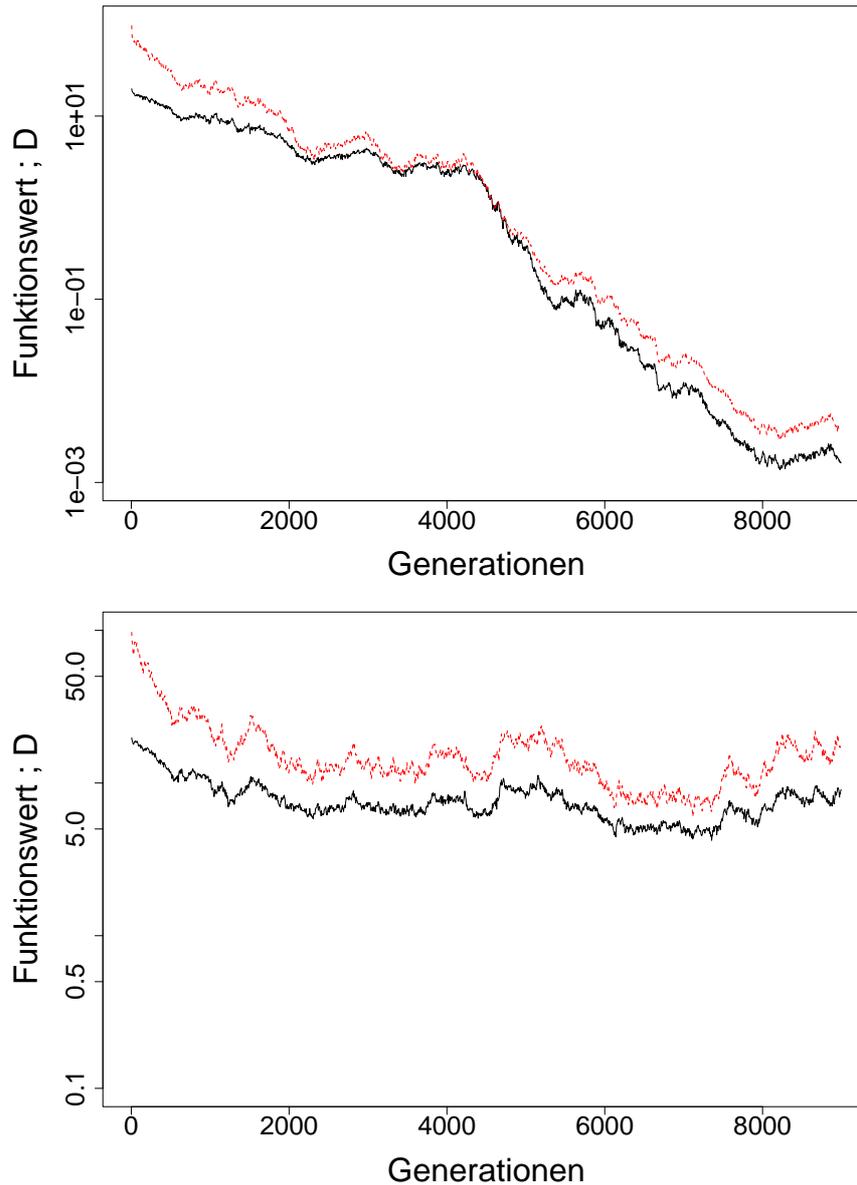


Abbildung 9.3: Funktionswert (schwarz) und Diversität (rot) eines repräsentativen Einzellaufs einer (15, 100)-ES auf der statischen (oben) und dynamischen (unten) Ackley-Funktion mit $n = 90$. Im dynamischen Fall bewegte sich das Optimum in einer Dimension mit einer Stärke von $s = 1.0$.

modell sinnvoll war, währenddessen bei der Ackley-Funktion D zu sinnvollen Ergebnissen führte. Ein genauerer Blick auf die Definitionen der drei Funktionen gibt uns einen ersten Anhaltspunkt für dieses Verhalten.

Sowohl beim Kugelmodell als auch bei der Rastrigin-Funktion fließen die Variablen in quadratischer Form in die Berechnung des Funktionswertes ein. Demgegenüber werden die Variablen bei der Ackley-Funktion zunächst nur in einem linearen Term eingebunden (genauer: $\sqrt{\sum_{i=1}^n x_i^2}$). Allerdings ist hierbei noch zu berücksichtigen, dass die Variablen anschließend noch weiteren Transformationen unterworfen werden. Dennoch könnte diese Beobachtung ein erster Ansatzpunkt für weitere Untersuchungen sein. Auch wenn sich diese Arbeit im Speziellen auf dynamische Umgebungen bezieht, so verspricht die Berücksichtigung der Populationsdiversität gerade in den Situationen

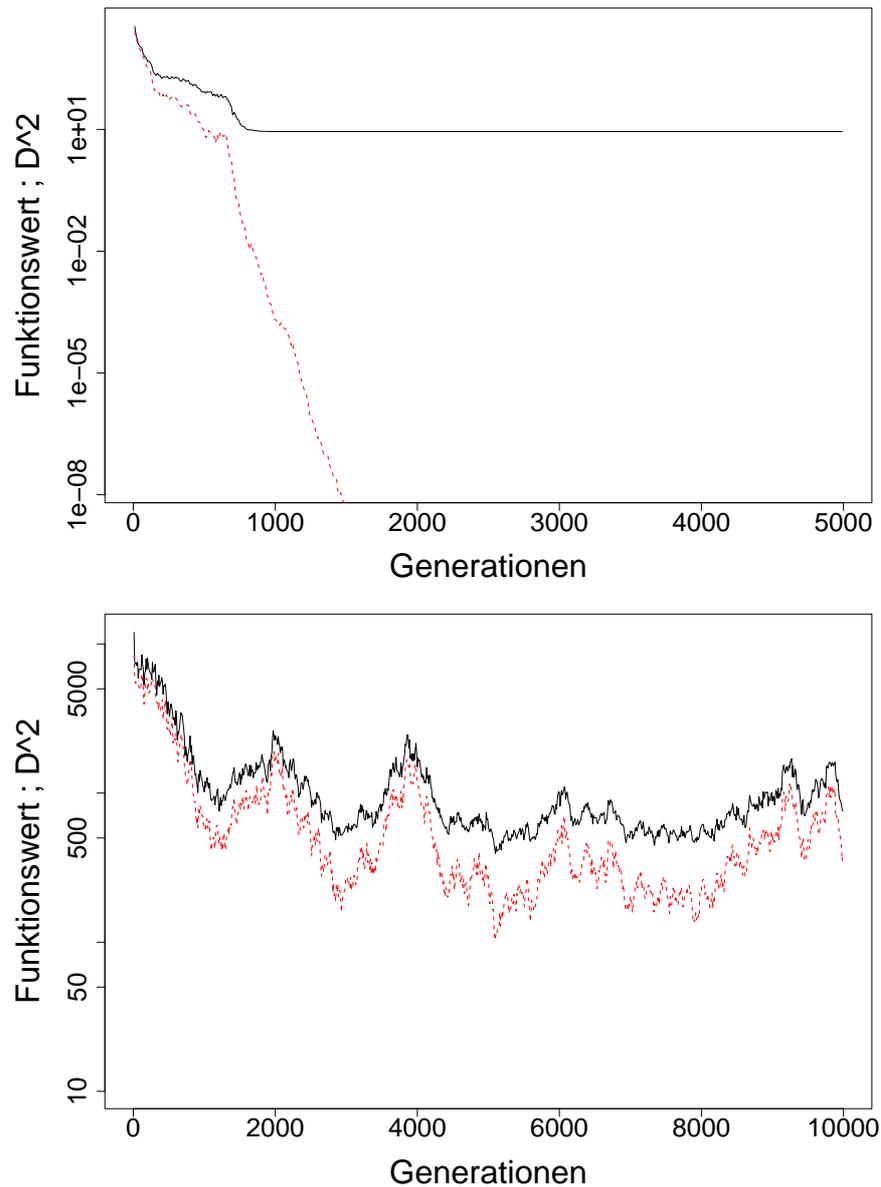


Abbildung 9.4: Funktionswert (schwarz) und Diversität (rot) eines Einzellaufs einer (15, 100)-ES auf der statischen Rastrigin-Funktion für $n = 50$ (oben) und $n = 100$ (unten).

Vorteile, in denen die ES gegen ein lokales Optimum konvergiert. Hier könnte ein starker Fall der beobachteten Diversität bei gleichzeitiger Konstanz des Funktionswertes auf eine frühzeitige Stagnation der ES hindeuten.

Auf der dynamischen Rastrigin-Funktion wurden zur Ermittlung eines Zusammenhangs Experimente mit einer (15, 100)-ES und einer Bewegungsstärke von $s = 0.1$ durchgeführt. Die für $n = 50$ und $n = 100$ erzielten Ergebnisse sind in der Abbildung 9.5 aufgeführt. Wie zuvor ist die Korrelation sehr hoch ($r \approx 0.95$). Das Aussehen der beiden Kurven für $n = 100$ ist ähnlich dem für die gleiche Problemdimension im statischen Fall beobachteten Verlauf. Durch die Dynamik wird in diesem Fall also keine zusätzliche wesentliche Schwierigkeit erzeugt, da die jeweiligen Funktionswerte auch um den gleichen Wert schwanken.

Die Abbildung 9.6 soll den Gleichlauf zwischen der Diversität und dem Funktionswert für den Fall aufzeigen, wenn die ES divergiert. Dazu wird die Rastrigin-Funktion mit $n = 150$ verwendet. Sowohl im hier nicht dargestellten statischen als auch im dynamischen Fall ist die hohe Korrelation bereits per in Augenscheinnahme erkennbar. Der Gleichlauf ist auch für andere Problemdimensionen vorhanden, vorausgesetzt die ES stagniert im statischen Fall nicht in einem lokalen Optimum. Tritt dies dennoch ein, ist wie bereits weiter oben beobachtet jeweils eine Abnahme der Diversität erkennbar. Zu unterscheiden ist eine Stagnation aber von dem gewollten Auffinden und Beibehalten des globalen Optimums mit $f(x^*) > 0$. In dieser Situation kommt es natürlich ebenfalls zum Auseinanderdriften des Funktionswertes und des Diversitätsmaßes.

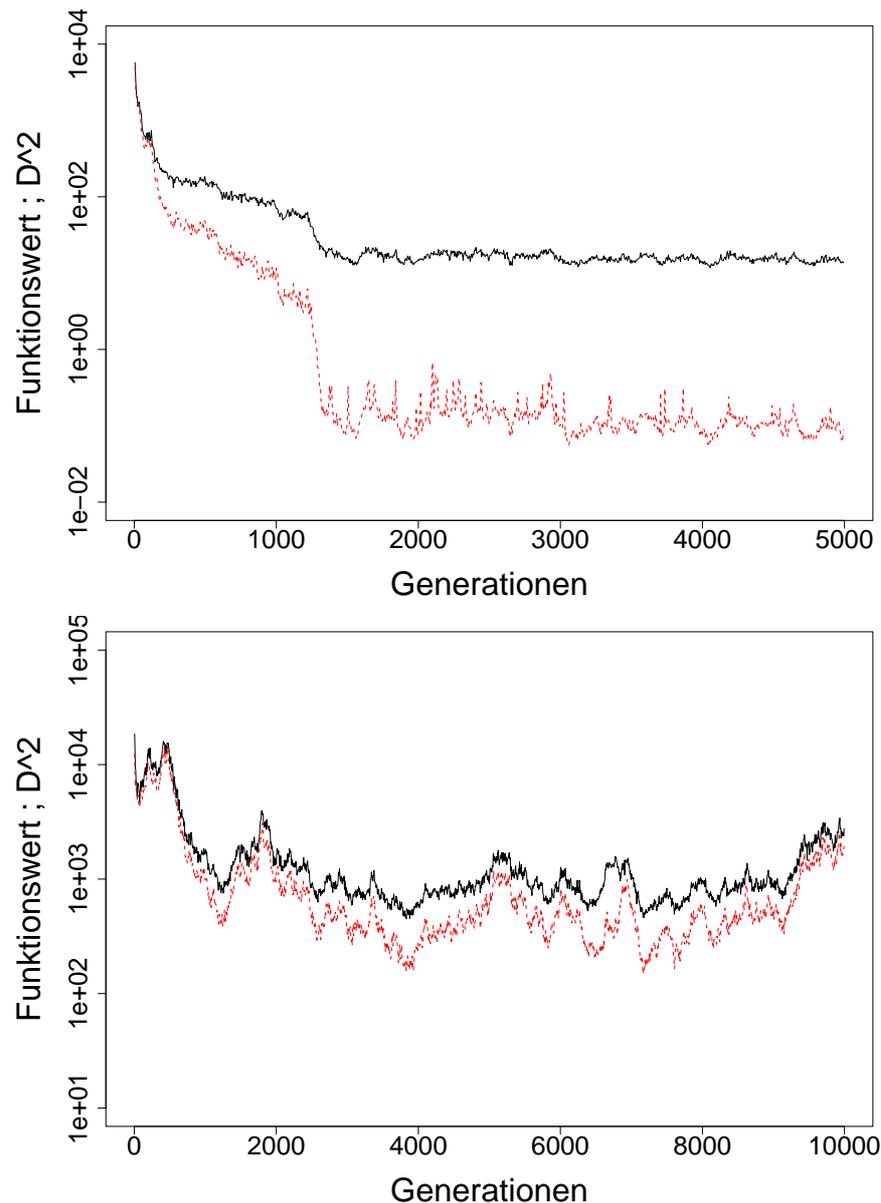


Abbildung 9.5: Funktionswert (schwarz) und Diversität (rot) eines Einzellaufs einer (15, 100)-ES auf der dynamischen Rastrigin-Funktion mit $n = 50$ (oben) und $n = 100$ (unten). Das Optimum bewegte sich jede Generation in einer Dimension mit einer Stärke $s = 0.1$.

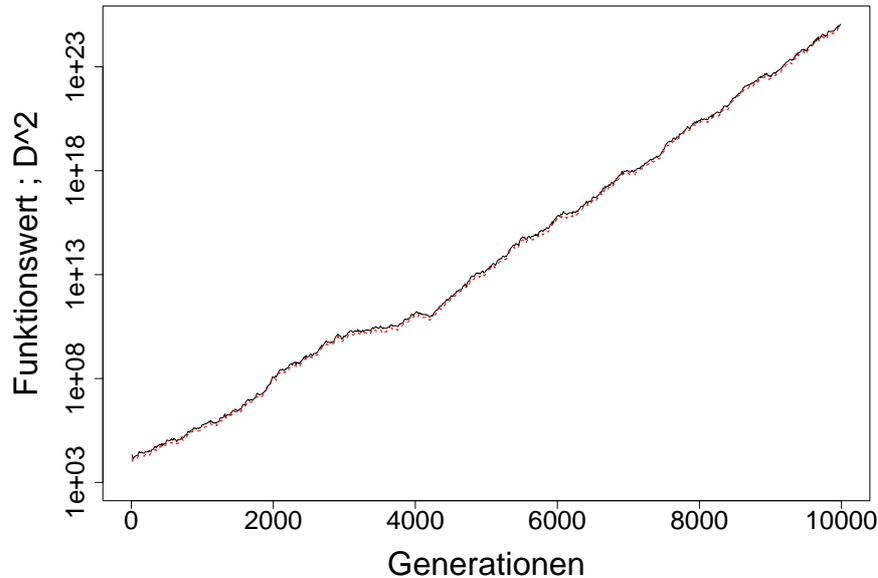


Abbildung 9.6: Funktionswert (schwarz) und Diversität (rot) eines Einzellaufs einer (15, 100)-ES auf der dynamischen Rastrigin-Funktion für $n = 150$. Das Optimum bewegte sich jede Generation in einer Dimension mit einer Stärke $s = 0.1$.

Eine besondere Erwähnung verdient aber noch die folgende Tatsache. Der Gleichlauf wird mit zwei unterschiedlichen Maßen erzielt. Einerseits handelt es sich um den Funktionswert des jeweils besten Individuums einer Generation, also von einem Individuum. Andererseits wird die Diversität innerhalb einer Population gemessen, also von insgesamt μ Individuen. Dass eine Korrelation zwischen beiden Werten existiert, ist leicht einsehbar. Dass sie aber so groß ist, ist dennoch überraschend und zeigt, dass auf dieser Erkenntnis zukünftige Algorithmen aufbauen sollten.

9.4 Fazit

Die Betrachtung der Diversität innerhalb der Population wird insbesondere in dynamischen Umgebungen seit längerem berücksichtigt. In der Vergangenheit geschah dies jedoch immer isoliert, d.h. insbesondere losgelöst von den Funktionswerten der Population. In diesem Kapitel wurde gezeigt, dass die Korrelation zwischen dem Funktionswert des besten Individuums der Population und der Diversität sehr groß ist. Das zu verwendende Diversitätsmaß hängt dabei von den Eigenschaften der zugrunde liegenden Funktion ab, d.h. von der Art, wie die Objektvariablen in den Funktionswert einfließen. Zum Erreichen des globalen Optimums erscheint es hilfreich, in zukünftigen Algorithmen die Korrelation zu nutzen, um eine frühzeitige Stagnation in einem lokalen Optimum zu erkennen und zu vermeiden. Durch eine detaillierte — auch theoretische — Untersuchung der genauen Abläufe könnte ein bevorstehender Übergang zur Stagnation in einem lokalen Optimum frühzeitig erkannt und durch geeignete Gegenmaßnahmen verhindert werden. Dazu gehört bspw. die künstliche Erhöhung der Schrittweiten.

Kapitel 10

PEACE – Ein neuartiger paralleler evolutionärer Algorithmus für dynamische Umgebungen

10.1 Parallele evolutionäre Algorithmen

In den Untersuchungen der vorangegangenen Kapitel haben wir verschiedene Evolutionsstrategien auf dynamischen Problemstellungen verglichen, bei denen sich die Bewegungsstärke nach der Anzahl der generierten Nachkommen pro Generation richtete. Dadurch war gewährleistet, dass die insgesamt zurückgelegte Entfernung sich nur nach der Anzahl der Funktionsauswertungen richtete und dementsprechend für alle Strategien identisch war, unabhängig davon, wieviele Nachkommen pro Generation erzeugt wurden.

In diesem Kapitel soll eine andere Variante, der so genannte **Parallel Evolutionary Algorithm for Changing Environments**, näher betrachtet werden. Hier wird die Bewegungsstärke pro Generation unabhängig von der Anzahl der erzeugten Nachkommen festgesetzt. Nutzt eine Strategie nicht die volle Anzahl maximaler Nachkommen aus, so bewegt sich das Optimum dennoch mit konstanter Stärke pro Generation. Die Grundannahme für dieses Vorgehen besteht darin, dass eine parallele Hardware-Architektur zur Verfügung steht, die es gestattet, bis zu λ_{gesamt} Funktionsauswertungen gleichzeitig durchzuführen. Untersucht wird nun, ob eine parallele Variante zweier Einzelstrategien mit zusammen λ_{gesamt} Nachkommen pro Generation bessere Ergebnisse erzielen kann als eine einzige Einzelstrategie mit der gleichen Anzahl Nachkommen. Zur Motivation dieser Fragestellung beginnen wir deshalb mit einer kurzen Übersicht bestehender paralleler evolutionärer Algorithmen.

In statischen Umgebungen besitzen parallele EA bereits seit längerer Zeit eine große Verbreitung. Die zahlreichen möglichen Varianten werden dabei unter wenigen Hauptformen zusammengefasst. Sprave (1999, S. 29) untergliedert die räumlichen Strukturen paralleler EA wie folgt:

- Im panmiktischen Modell existiert nur eine Population, sodass alle Individuen in direkter Konkurrenz zueinander stehen.
- Bei den Multipopulationsmodellen setzt sich die Population aus einzelnen, voneinander getrennten Teilpopulationen zusammen. Benachbarte Teilpopulationen tauschen von Zeit zu Zeit Individuen aus. Im Migrationsmodell werden die Individuen dabei verschoben, während sie im Pollinationsmodell kopiert werden.

- In den Nachbarschaftsmodellen wird eine so genannte Nachbarschaftsrelation definiert. Diese gibt an, zwischen welchen Individuen eine Interaktion stattfinden kann.

Trotz der in statischen Umgebungen erfolgreichen Anwendung ist dieses Konzept auf dynamische Umgebungen bisher erst selten übertragen worden. Sarma und De Jong (1999) vergleichen einen EA, der eine auf einem Gitter verteilte Population besitzt, mit einem herkömmlichen EA mit panmiktischer Population. Der Ansatz, verschiedene Teilpopulationen in unterschiedlichen Regionen des Suchraums zu platzieren, wird von mehreren Autoren benutzt (Branke 2002; Oppacher und Wineberg 1999; Ursem 2000). Ziel ist dabei eine Diversitätserhöhung, sodass jede Teilpopulation ein lokales Optimum verfolgt. Wird bei einer Änderung der Umgebung ein bisheriges lokales Optimum zu einem globalen, so ist dieses im Idealfall bereits mit einer Teilpopulation besetzt. Dabei existiert in der Regel ein übergeordneter EA, der die Einzelpopulationen erzeugt und beendet.

Für den in dieser Arbeit zugrunde gelegten Fall der Funktionsverschiebung genügt eine einfachere Variante eines parallelen evolutionären Algorithmus, die im Folgenden vorgestellt wird.

10.2 PEACE im Detail

10.2.1 Grundvariante von PEACE

Ausgangspunkt für den hier vorgestellten Algorithmus PEACE ist die Annahme, dass dem Anwender eine Hardware zur Verfügung steht, die die gleichzeitige Auswertung von λ_{gesamt} Individuen gestattet. Durch die Auswertung einer Lösung entstehen keine weiteren Kosten, sodass zur Informationsmaximierung in jedem Schritt auch λ_{gesamt} Individuen erzeugt werden sollten.

Die experimentellen Untersuchungen der Kapitel 7 und 8 haben gezeigt, dass es keinen EA mit „idealen“ Populationsgrößen gibt. Unter unterschiedlichen Bedingungen hat einmal die eine und einmal die andere Strategie Vorteile gegenüber den weiteren. Sinnvoll erscheint deshalb das Vorgehen, mehrere separate EA mit unterschiedlichen Strategien parallel laufen zu lassen und das jeweils beste Individuum aus allen Populationen zur aktuell verwendeten Lösung zu machen. Durch dieses Vorgehen hofft man nun die Vorteile der sich unabhängig voneinander entwickelnden EA kombinieren zu können. Durch das Zulassen des gegenseitigen Austauschs von Informationen über das beste Individuum könnte sogar das Ergebnis der jeweils besseren Einzelstrategie überboten werden. Zum Vergleich mit dem parallelen EA mit verschiedenen Teilpopulationen dient eine Einzelstrategie mit maximaler Anzahl Nachkommen.

Um die Anzahl der durchzuführenden Experimente in einem überschaubaren Rahmen zu halten beschränken wir uns in der folgenden Untersuchung auf die Betrachtung der aufgrund ihrer starken Multimodalität besonders interessanten Rastrigin-Funktion. Die Zahl der pro Generation insgesamt zur Verfügung stehenden Auswertungen wird auf $\lambda_{\text{gesamt}} = 200$ gesetzt. Diese Zahl wird gleichmäßig auf die beiden Algorithmen verteilt. Für die erste Strategie verwenden wir unsere bekannte (15, 100) Standard-ES mit n Schrittweiten (Strategie ES_1 genannt), die sich in vielen Fällen bewährt hat und auch im statischen Fall gute Resultate liefert.

Für die zweite Strategie (ES_2 genannt) kommt eine (1, 100)-ES mit einer Schrittweite zum Einsatz. Sie gestattet in einfachen Fällen eine schnelle und gute Annäherung an das Optimum. Dies gilt einerseits für statische unimodale Funktionen, aber insbesondere auch für dynamische Funktionen, die durch längere Phasen ohne Änderungen (großes Δg) gekennzeichnet sind. Für diese Kombination (also (15, 100)-ES mit (1, 100)-ES) werden wir zum einen die Resultate betrachten, die sich ergeben, wenn die parallel betriebenen EA separat evolvieren, also kein Austausch möglich ist.

Zum anderen gestatten wir in bestimmten Abständen einen Austausch zwischen beiden. Verglichen wird dabei zum einen der mittlere Lauf und zum anderen der ABFV der gemeinsamen Strategie mit den entsprechenden Daten der Einzelvarianten. Für die gemeinsame Strategie wird dann in jeder Generation der beste der Funktionswerte genommen, die in den beiden Einzelpopulationen vorhanden sind.

10.2.2 Austausch von Informationen zwischen den Einzelstrategien

Durch den teilweise zugelassenen Austausch von Informationen zwischen den beiden Einzelstrategien erhofft man sich einen zusätzlichen Performance-Gewinn. Das soll einer Strategie ermöglichen, näher an das Optimum zu gelangen und dann im Weiteren dem Optimum besser zu folgen. Neben dem Zeitintervall Δc , nach dessen Ablauf ein Austausch erfolgt, erfordert der Austausch von Individuen selbst eine besondere Beachtung. In dem hier beschriebenen neuen Algorithmus wird dabei wie folgt verfahren.

Sind die beiden Algorithmen ES_1 und ES_2 für Δc Generationen getrennt evolviert, wird für jeden der beiden EA das beste Individuum der aktuellen Population notiert (I_{ES_1} und I_{ES_2}). Sind die Funktionswerte beider Individuen identisch erfolgt kein Austausch und die beiden Algorithmen evolvieren wieder getrennt. Ist dagegen eines der beiden Individuen besser (o.B.d.A. I_{ES_1}) wird die gesamte andere Population mit μ_2 Kopien dieses Individuums re-initialisiert. Denkbar wäre natürlich auch, die zweite Population mit den μ_2 besten Individuen zu initialisieren (falls $\mu_1 \geq \mu_2$), um eine höhere Diversität zu gewährleisten. Dies dürfte im dynamischen Fall mit kleinem Δc jedoch relativ unproblematisch sein, da sich aufgrund der Dynamik nach kürzerer Zeit wieder eine gewisse Diversität eingestellt haben sollte. Bei großem Δc ist hingegen mit einer schnellen Konvergenz gegen ein lokales Optimum zu rechnen, was jedoch in Kauf genommen wird. Vom Kopieren nur eines (des besten) Individuums in die andere Population (mit $\mu_2 > 1$) wurde nach ersten Versuchen ebenso Abstand genommen. Dies deshalb, weil aufgrund der Rekombination das u.U. sehr viel bessere neue Individuum keinen positiven Einfluss auf die Gesamtpopulation nehmen konnte.

Bei Strategien mit einer unterschiedlichen Anzahl Schrittweiten pro Individuum muss bei einem Transfer jedoch auf die richtige Anzahl der Schrittweiten geachtet werden. In den hier benutzten Algorithmen wird/werden entweder eine oder n Schrittweite/n verwendet. Stimmt die Anzahl der in den Individuen implementierten Schrittweiten in beiden Strategien überein, ist ein direktes Kopieren möglich. Ist dies nicht der Fall, ergeben sich zwei zu unterscheidende Szenarien. Für den Fall, dass ein Individuum mit nur einer Schrittweite kopiert wird, werden alle Schrittweiten mit dem Wert der einen Schrittweite belegt. Im Fall, dass ein Individuum mit n Schrittweiten kopiert wird, erhält das neue Individuum als Schrittweite den Durchschnitt aus allen n Schrittweiten des kopierten Individuums. Die Alternative, genau eine Schrittweite zu kopieren, erwies sich dagegen in ersten Versuchen als wenig erfolgreich, sodass diese Möglichkeit nicht weiter verfolgt wird.

10.3 Experimentelle Resultate

10.3.1 PEACE ohne Austausch

Von den zahlreichen Experimenten sollen an dieser Stelle nur einige wenige exemplarische Resultate aufgeführt werden. Da sowohl auf dem Kugelmodell als auch auf der Ackley-Funktion bereits mit einfachen Strategien hinreichend gute Ergebnisse erzielt wurden, beschränken wir uns darüber hinaus auf die wesentlich interessantere, weil stärker multimodale, Rastrigin-Funktion. Neben der

im Abschnitt 5.1.3 dargestellten Rastrigin-Funktion mit der Amplitude $A = 3$ wird in diesem Kapitel eine zweite Form der Rastrigin-Funktion mit einer erhöhten Amplitude $A = 10$ verwendet, die zwischen dem Wert 3 und dem von Rudolph (1990, S. 72) benutzten Wert 50 liegt:

$$f_{\text{Rastrigin2}}(x) = n \cdot A + \sum_{i=1}^n (x_i^2 - A \cdot \cos(\omega \cdot x_i))$$

mit $A = 10$, $\omega = 2\pi$.

Die Frequenz ω wird auch bei dieser alternativen Definition identisch gewählt. Wie im Abschnitt 5.1.3 gezeigt wurde, ist für eine größere Amplitude A der Definitionsbereich größer, für den lokale Optima existieren. Konkret befinden sich für $A = 3$ die lokalen Optima im Bereich $|x_i| \lesssim 9.5$ und für $A = 10$ im Bereich $|x_i| \lesssim 31.5$. Die Abbildung 10.1 veranschaulicht das Aussehen für den eindimensionalen Fall. Dieser Vergleich zeigt deutlich, dass bereits in größerer Entfernung vom Optimum lokale Opti-

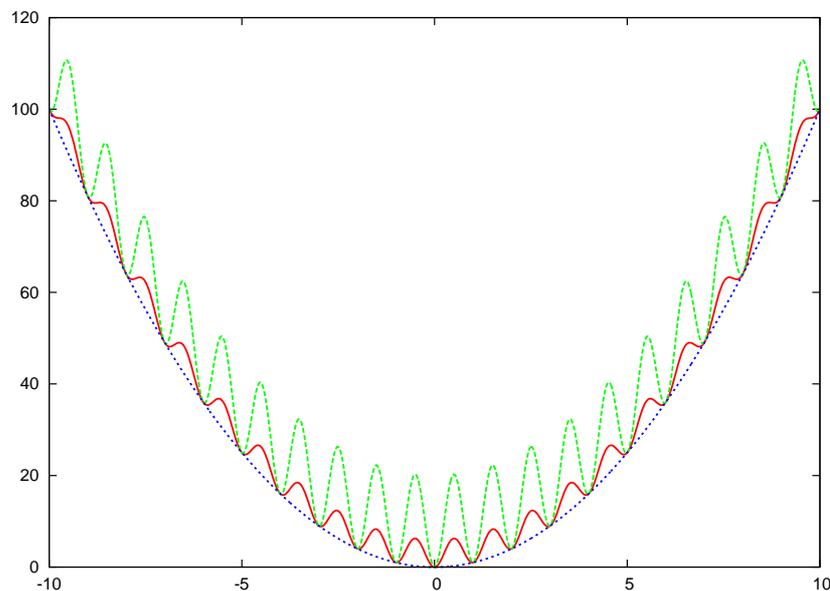


Abbildung 10.1: Aussehen der beiden Rastrigin-Funktionen $f_{\text{Rastrigin}}$ ($A = 3$, rot) und $f_{\text{Rastrigin2}}$ ($A = 10$, grün) sowie des Kugelmodells (blau) für $n = 1$.

ma kleinere Funktionswerte besitzen als der obere Bereich der Attraktoren besserer lokaler Optima. Dieser Umstand sollte dazu führen, dass die Funktion $f_{\text{Rastrigin2}}$ schwieriger zu optimieren ist.

Allgemein sind die Ergebnisse der kombinierten Strategie PEACE ohne Austausch immer mindestens so gut wie die der besseren Einzelstrategien. Allerdings ist es auch häufig so, dass die parallele Strategie bessere Ergebnisse erzielt als die bessere der Einzelstrategien. Dies liegt daran, dass eine Strategie A, die im Durchschnitt schlechter ist als Strategie B, in einzelnen Läufen oder in einzelnen Generationen durchaus bessere Resultate erzielen kann als Strategie B. Der Durchschnitt der parallelen Strategie kann somit auch besser sein als der Durchschnitt der besseren Einzelstrategie. Dies zeigt Abbildung 10.2.

Ähnliche Eigenschaften verdeutlicht auch der Verlauf der Kurven der Strategien in Abbildung 10.3. Während diesmal für $\mu = 15$ große Schwankungen zu erkennen sind, liegt für $\mu = 1$ die Kurve bis zur 3000. Generation weit unterhalb des danach erreichten Funktionswertes. Bei dieser Funktion benötigt die ES mit n Schrittweiten nach einem Wechsel zu viel Zeit, um sich der veränderten Situation anzupassen. Das führt dazu, dass die Funktionswerte häufig sehr stark schwanken.

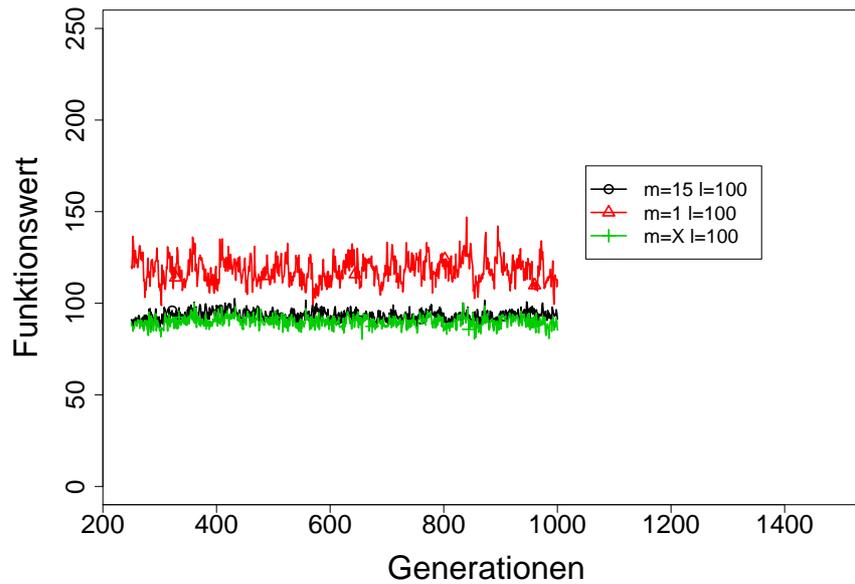


Abbildung 10.2: Mittlerer Lauf einer (1,100)-, (15,100)- und der parallelen ES auf der 10-dimensionalen Rastrigin-Funktion $f_{\text{Rastrigin2}}$ für den Bewegungstyp III bei einer Bewegungsfrequenz $\Delta g = 1$ und einer Bewegungsstärke $s = 1$.

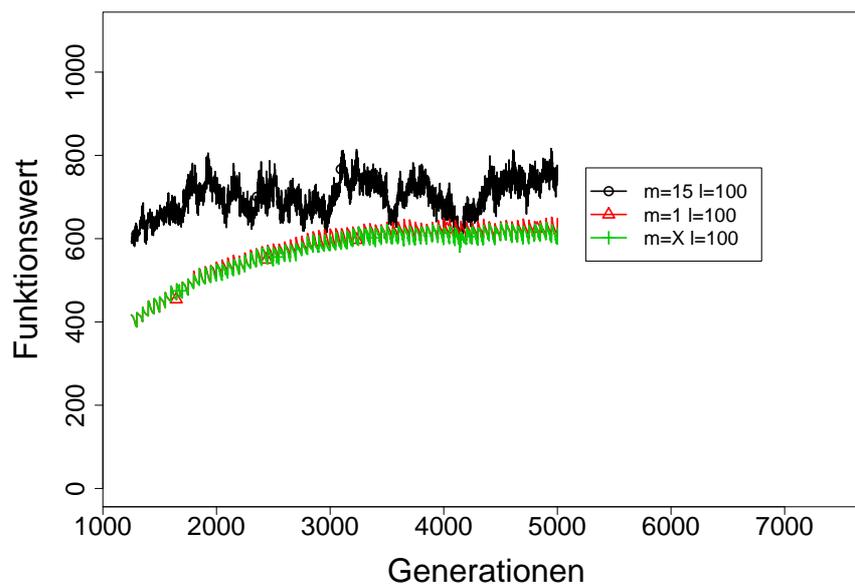


Abbildung 10.3: Mittlerer Lauf einer (1,100)-, (15,100)- und der parallelen ES auf der 100-dimensionalen Rastrigin-Funktion $f_{\text{Rastrigin}}$ für den Bewegungstyp I bei einer Bewegungsfrequenz $\Delta g = 50$ und einer Bewegungsstärke $s = 1$.

Dagegen kann die ES mit nur einer Schrittweite zwar unmöglich sehr gute Funktionswerte erreichen (siehe Erklärungen im Kapitel 6), sie sind aber durchschnittlich gut genug, um die andere Strategie auf den zweiten Platz zu verweisen. Der vergrößerte Ausschnitt in Abbildung 10.4 verdeutlicht die Situation. Für die (1,100)-ES ist deutlich die sägezahnartige Kurve zu erkennen, die sich aufgrund der sich alle 50 Generationen ändernden Umgebung ergibt. Die (15,100)-ES versucht nach einem Wechsel zunächst alle Strategieparameter zu adaptieren. Ist dies einigermaßen gelungen, kann mit

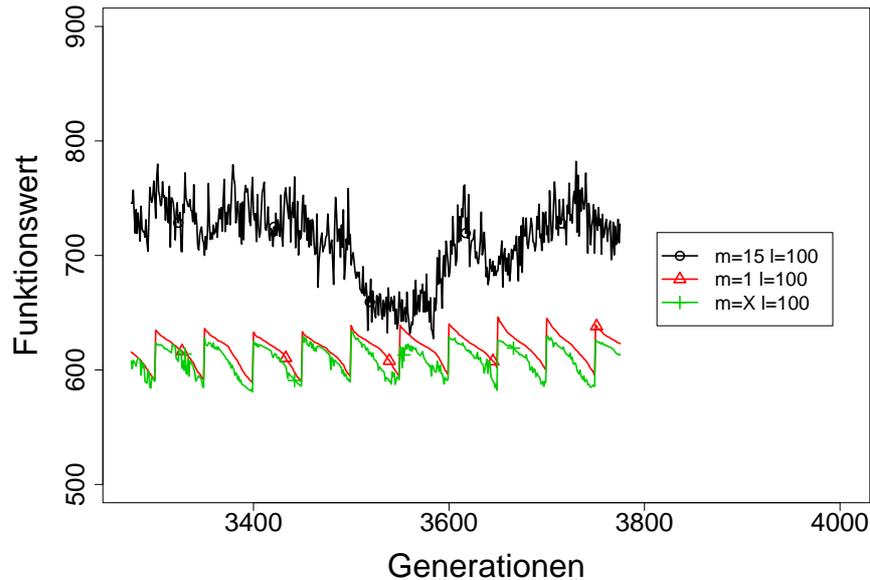


Abbildung 10.4: Ausschnitt eines mittleren Laufs einer $(1, 100)$ -, $(15, 100)$ - und der parallelen ES auf der 100-dimensionalen Rastrigin-Funktion $f_{\text{Rastrigin}}$ für den Bewegungstyp I bei einer Bewegungsfrequenz $\Delta g = 50$ und einer Bewegungsstärke $s = 1$.

den neuen Einstellungen jedoch kein wesentlicher Fortschritt erzielt werden, weil sich die Funktion erneut ändert. Darüber hinaus treten in der Abbildung auch noch einmal deutlich die (kleinen) Verbesserungen der kombinierten Strategie sogar gegenüber der $(1, 100)$ -ES zu Tage. Der kombinierte EA erreicht auf dieser Funktion zunächst die guten Resultate der $(1, 100)$ -ES. Im späteren Verlauf kann er sich sogar etwas von diesen Werten absetzen. Insgesamt zeigen die beiden untersuchten Fälle, dass der kombinierte EA die Stärken beider Einzelstrategien vereint. Man könnte sogar gewissermaßen sagen, die kombinierte Strategie ist „mehr als die Summe der Einzelstrategien“. Dies ist für den Bereich der parallelen und dynamischen Funktionen ein positives und neues Ergebnis.

Ein Vergleich auf Basis der Diversität ist nicht unbedingt sinnvoll. Es handelt sich hier um zwei getrennte Strategien, von denen lediglich das beste Individuum als Lösung der kombinierten Strategie genommen wird. Betrachten wir die beiden Populationen zusammen, so hat der kombinierte EA entweder eine ungefähr gleiche Diversität wie die Einzelpopulationen, nämlich dann, wenn sich beide Populationen in der Nähe des gleichen (lokalen) Optimums befinden. Oder es ist angemessener, die Diversität der beiden Populationen zu addieren, wenn sich die Populationen in den Attraktionsgebieten verschiedener (lokaler) Optima bewegen.

Der parallele EA wurde auch unter dem Aspekt entwickelt, besser als eine einfache Einzelstrategie zu sein. In hier nicht weiter dargestellten Voruntersuchungen hat sich gezeigt, dass eine Einzelstrategie mit weniger als der zur Verfügung stehenden Anzahl Nachkommen in der Regel schlechtere Ergebnisse erzielt als wenn die volle Anzahl benutzt wird. Deshalb verwenden wir zum Vergleich eine $(\mu, 200)$ -ES. Zur Festlegung der Anzahl der Eltern stützen wir uns für die hier benutzten beiden Varianten der Rastrigin-Funktion auf die Resultate des Kapitels 8, wo für multimodale Funktionen eine größere Anzahl Eltern empfohlen wurde. Den Selektionsdruck erhöhen wir dabei ebenfalls, so dass wir zu einer $(20, 200)$ -ES gelangen.

In den Abbildungen 10.5 und 10.6 sind für die beiden vorangegangenen Problemstellungen die Ergebnisse der parallelen ES und der $(20, 200)$ -ES im direkten Vergleich aufgetragen. In beiden Fällen schneidet die $(20, 200)$ -ES besser ab. Für die Funktion mit einer Bewegungsfrequenz

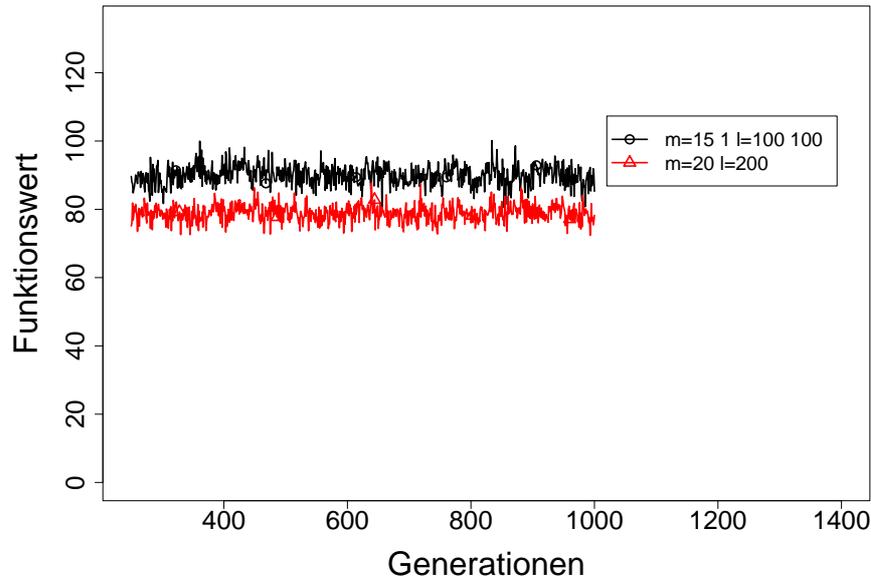


Abbildung 10.5: Mittlerer Lauf der parallelen $(1, 100)/(15, 100)$ -ES und einer $(20, 200)$ -ES auf der 10-dimensionalen Rastrigin-Funktion $f_{\text{Rastrigin2}}$ für den Bewegungstyp III bei einer Bewegungsfrequenz $\Delta g = 1$ und einer Bewegungsstärke $s = 1$.

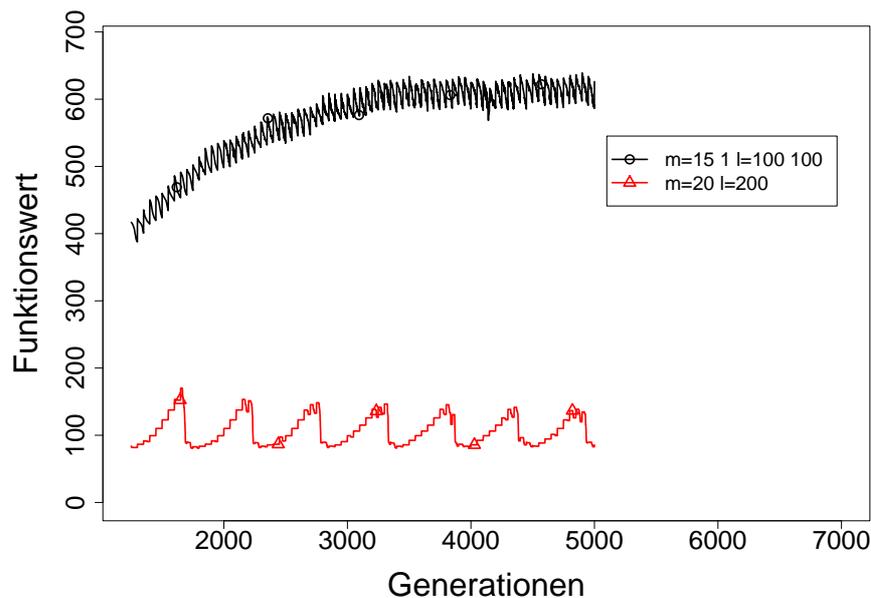


Abbildung 10.6: Mittlerer Lauf der parallelen $(1, 100)/(15, 100)$ -ES und einer $(20, 200)$ -ES auf der 100-dimensionalen Rastrigin-Funktion $f_{\text{Rastrigin}}$ für den Bewegungstyp I bei einer Bewegungsfrequenz $\Delta g = 50$ und einer Bewegungsstärke $s = 1$.

$\Delta g = 50$ ist dieser Unterschied sogar enorm. Offensichtlich kann die parallele Strategie nur unwesentlich über die Ergebnisse der beiden Einzelstrategien hinausgehen und fällt im Verhältnis zur Einzelstrategie mit größeren Populationen stark zurück. Ohne diese These durch explizite Ergebnisse zu untermauern kann gesagt werden, dass dieses Resultat auch für andere Problemstellungen zutrifft. Günstiger als zwei Einzelstrategien zu starten und das bessere der beiden Ergebnisse zu nehmen ist es meistens, direkt eine einzige Strategie mit voller Nachkommenschaft zu benutzen. Der

Vorteil eines parallelen Verfahrens macht sich hier also für eine panmiktische Population besonders bemerkbar.

Hervorzuheben ist auch der interessante Verlauf der Kurve der (20, 200)-ES in Abbildung 10.6. Aus diesem Verlauf könnte auf den ersten Blick geschlossen werden, dass die Strategie von den Änderungen profitiert, indem zu den Änderungszeitpunkten das Optimum näher an die Population kommt. Dies ist jedoch nicht der Fall wie die Abbildung 10.7 verdeutlicht. Zum einen erfolgt

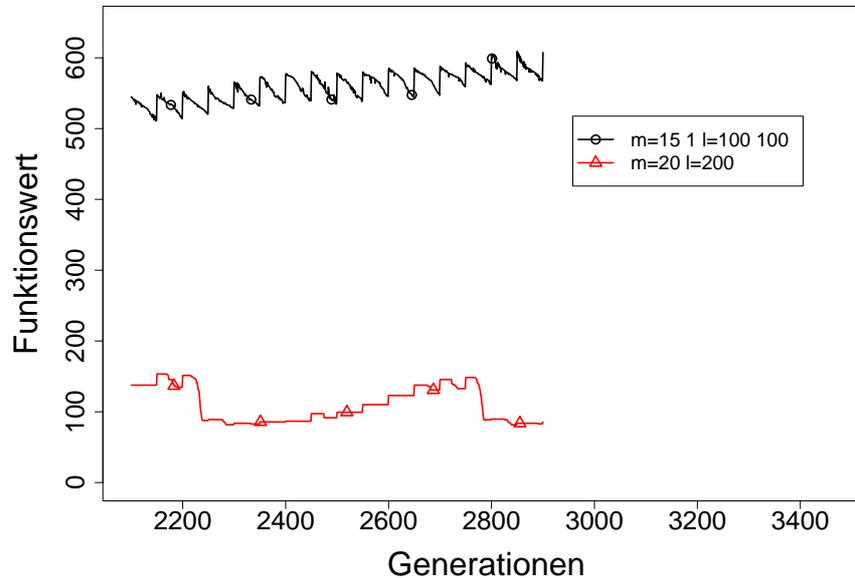


Abbildung 10.7: Mittlerer Lauf der parallelen (1, 100)/(15, 100)-ES und einer (20, 200)-ES auf der 100-dimensionalen Rastrigin-Funktion $f_{\text{Rastrigin}}$ für den Bewegungstyp I bei einer Bewegungsfrequenz $\Delta g = 50$ und einer Bewegungsstärke $s = 1$.

kein abrupter Fortschritt. Stattdessen verbessert sich der Funktionswert aufgrund sich ändernder Schrittweiten über mehrere Generationen. Zum anderen ist der Zeitpunkt der beginnenden Verbesserung unabhängig vom Änderungszeitpunkt. Dies wird auch dadurch belegt, dass die Abstände dieser „Zyklen“ nicht konstant sind und sich über einen größeren Zeitraum als $\Delta g = 50$ erstrecken. Zum dritten schwankt der Funktionswert ohnehin recht stark. Für eine Erklärung dieses Phänomens sei daran erinnert, dass die Rastrigin-Funktion stark multimodal ist, sodass die ES während der statischen Phasen leicht in einem lokalen Optimum stecken bleiben kann. Nach einer Bewegung des Optimums findet eine erneute Adaptation der ES statt. Hiervon sind alle Schrittweiten betroffen. Der Prozess, bis die ES erneut in einem lokalen Optimum stecken bleibt, kann länger dauern als der Abstand bis zur nächsten Funktionsänderung. In unregelmäßigen Abständen sind die Schrittweiten wieder besser angepasst, sodass es der ES gelingt, das Attraktionsgebiet eines besseren lokalen Optimums zu erreichen. Wir werden auf diesen Aspekt gleich noch einmal zurückkommen.

Im Folgenden soll genauer überprüft werden, für welches λ eine $(15, \lambda)$ -ES in den betrachteten Fällen mindestens genauso gute Ergebnisse liefert wie die kombinierte Variante. Zum übersichtlicheren Vergleich kehren wir deshalb zur Darstellung des gewohnten ABFV zurück. Die Tabelle 10.1 enthält die Daten ausgewählter getesteter Varianten. Die Rastrigin-Funktion $f_{\text{Rastrigin}}$ mit Bewegungsfrequenz $\Delta g = 50$ besitzt ausgesprochen lange konstante Phasen, die die Funktion schon fast zu einer statischen Problemstellung werden lassen. Eine $(15, 130)$ -ES kann nun die besten Ergebnisse erzielen. Mit zunehmender Anzahl Nachkommen werden die Ergebnisse wieder etwas schlechter,

Tabelle 10.1: ABFV verschiedener ES auf der Funktion $f_{\text{Rastrigin2}}$ mit $n = 10, \Delta g = 1, s = 1$, Bewegungstyp III (links) bzw. $f_{\text{Rastrigin}}$ mit $n = 100, \Delta g = 50, s = 1$, Bewegungstyp I (rechts).

Strategie	ABFV	Strategie	ABFV
(1, 100)/(15, 100)	89.8	(1, 100)/(15, 100)	602.4
(1, 100)	118.3	(1, 100)	611.8
(15, 100)	93.1	(15, 100)	708.6
(15, 110)	91.0	(15, 110)	523.6
(15, 120)	88.5	(15, 120)	376.4
(15, 130)	87.6	(15, 130)	89.4
(15, 140)	85.7	(15, 140)	94.7
(15, 150)	84.7	(15, 150)	104.9
(15, 160)	83.0	(15, 160)	111.3
(15, 170)	82.3	(15, 170)	114.4
(15, 180)	81.2	(15, 180)	113.3
(15, 190)	80.1	(15, 190)	118.0
(15, 200)	79.2	(15, 200)	118.6
(20, 200)	76.8	(20, 200)	107.7

da die ES dann aufgrund des höheren Selektionsdrucks schneller in einem lokalen Optimum stecken bleibt. Auch mit der (20, 200)-ES werden schlechtere Ergebnisse verzeichnet.

Aus den obigen Daten kann nicht geschlossen werden, dass die besten Ergebnisse mit ca. 130 Nachkommen erhalten werden. Dies ist lediglich für die hier benutzten $\mu = 15$ Eltern der Fall. Eine erhöhte Anzahl Nachkommen sollte mit einer anderen Wahl für μ durchaus bessere Ergebnisse liefern. Für die maximale Anzahl Nachkommen von 200 wurden deshalb mehrere Varianten getestet. Dabei zeigte sich das beste Ergebnis des ABFV von 87.5 mit $\mu = 24$ Eltern. Diese hat einen etwas kleineren Selektionsdruck als die andere ES ($200/24 \approx 8.3 < 130/15 \approx 8.7$). Für $\mu = 23$ Eltern (Selektionsdruck $200/23 \approx 8.7$) beträgt der ABFV = 103.0, ist also trotz der nur um Eins reduzierten Anzahl Eltern erheblich schlechter. Der Sprung der Ergebnisse von der Strategie (15, 120) zu (15, 130) ist so beeindruckend, dass eine kurze Erklärung angebracht ist. Wir erinnern uns, dass auf der Rastrigin-Funktion der Bereich der beginnenden Divergenz einer (15, 100)-ES bei $n \approx 100$ einsetzte. In diesem Bereich sind erhebliche Schwankungen sowohl zwischen als auch innerhalb der Läufe festzustellen. Wird die Anzahl der Nachkommen erhöht, reduzieren sich diese Schwankungen. Zur genaueren Eingrenzung einer guten Wahl der Anzahl der Nachkommen werden deshalb weitere Experimente mit dazwischen liegenden Werten der Nachkommen durchgeführt. Abbildung 10.8 gibt die Ergebnisse in einer grafischen Übersicht wieder. Ab einer Anzahl der Nachkommen von $\lambda > 120$ tritt eine deutliche Verbesserung ein. Während der ABFV bis zu $\lambda \approx 140$ noch immer erkennbar schwankt, ergibt sich für noch größere Werte von λ ein relativ gleich bleibender Wert. Diese Resultate korrespondieren somit mit den im Kapitel 8 erhaltenen Resultaten, wo sich der erhaltene ABFV bis zu einer bestimmten Anzahl Nachkommen reduzierte und sich mit weiter steigendem λ kaum noch veränderte. Die Betrachtung der Objektvariablen bestätigt die Interpretation. Es zeigt sich, dass die einzelnen Läufe einer (15, 100)-ES stark schwanken. Mit größer werdendem λ schwächt sich dieses Verhalten ab und die ES ist mehr und mehr in der Lage, sich dem Optimum dauerhaft zu nähern und besser zu folgen.

Die strukturellen Resultate sind für alle Dimensionen ähnlich. Unterschiedlich sind dagegen die Funktionswerte und genauen Verläufe. Das strukturelle Verhalten verdeutlicht die Abbildung 10.9

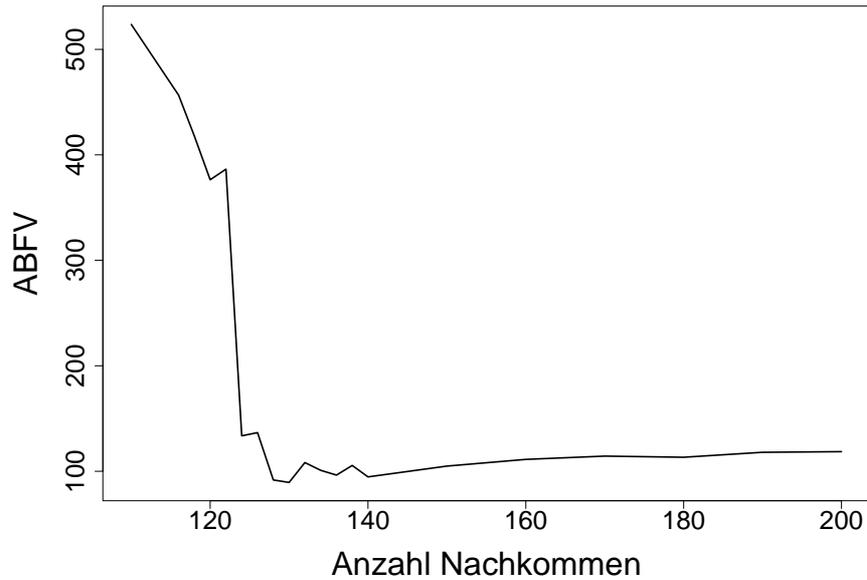


Abbildung 10.8: Darstellung des ABFV verschiedener $(15, \lambda)$ -ES auf der 100-dimensionalen Rastrigin-Funktion $f_{\text{Rastrigin}}$ für den Bewegungstyp I bei einer Bewegungsfrequenz $\Delta g = 50$ und einer Bewegungsstärke $s = 1$.

für den 10-dimensionalen Fall. Nach der ersten Adaptation an die Problemstellung gerät die ES wie im statischen Fall in ein lokales Optimum. Trotz der danach einsetzenden Dynamik ist die Phase der Konstanz zu groß ($\Delta g = 50$) und die Änderung zu klein ($s = 1$), sodass keine weitere Adaptation mehr stattfindet. Erst wenn eine Objektvariable aus dem Bereich der Multimodalität herausläuft ($x_i > |A\omega/2|$, siehe Abschnitt 5.1.3) kann die ES eine erneute Adaptation durchführen.

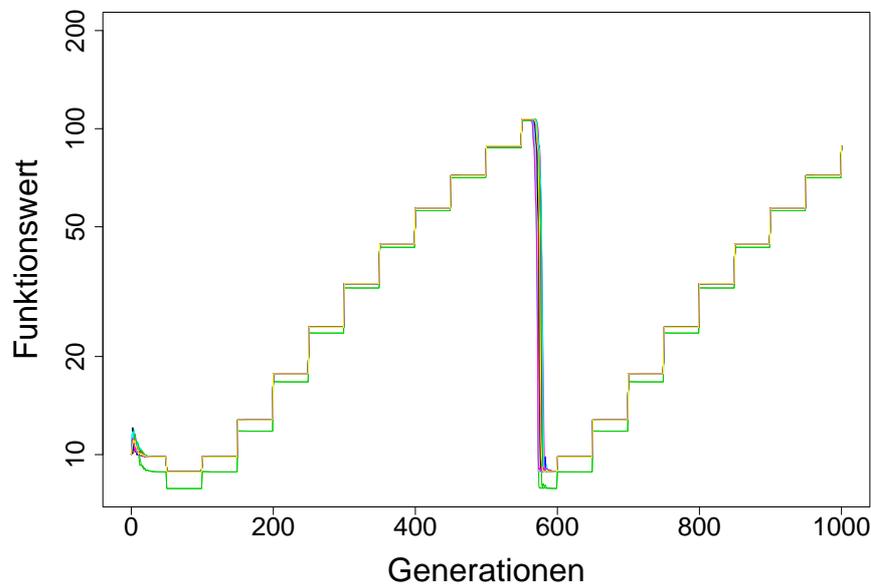


Abbildung 10.9: Darstellung aller 31 Läufe einer $(15, 100)$ -ES auf der 10-dimensionalen Rastrigin-Funktion $f_{\text{Rastrigin}}$ für den Bewegungstyp I bei einer Bewegungsfrequenz $\Delta g = 50$ und einer Bewegungsstärke $s = 1$.

Hier reicht es dann schon aus, dass auch nur die Objektvariable¹ der Bewegungskordinate (Bewegungstyp I) aus dem Bereich herausläuft, sodass die anschließende Adaptation alle anderen Objektvariablen ebenfalls erfasst. Mit der gleichen Begründung erklären wir jetzt auch das bereits in den Abbildungen 10.6 und 10.7 beobachtete Verhalten. Auch in den dortigen Experimenten war die Bewegungsfrequenz so hoch, dass die ES in ein lokales Optimum lief, aus dem es erst wieder herauskam als die der Bewegungskordinate entsprechende Objektvariable eine bestimmte Differenz zur Optimumkoordinate besaß. Danach konvergierte die ES erneut gegen ein lokales Optimum von etwa der gleichen Qualität.

Insgesamt handelt es sich hier um bemerkenswerte neue Resultate, die sich sowohl von den Ergebnissen mit kleinem Δg als auch denen für den statischen Fall unterscheiden.

10.3.2 PEACE mit Austausch

In der letzten Experimentreihe soll untersucht werden, welchen Einfluss der Austausch von Informationseinheiten in unserem parallelen EA besitzt. Zur Überprüfung des Einflusses werden die beiden Algorithmen wie bisher zunächst isoliert evolviert. Von Zeit zu Zeit (Δc Generationen) wird das beste Individuum beider Algorithmen zu dem anderen Algorithmus übertragen und besetzt dort die gesamte Population. In den dynamischen Umgebungen mit $\Delta g > 1$ wird dabei darauf geachtet, dass die Zeitspanne Δc nicht identisch mit Δg ist.

Für einen direkten Vergleich der Ergebnisse werden die Versuche des letzten Abschnitts mit der alternativen Strategie noch einmal wiederholt. In den Abbildungen 10.10 und 10.11 sind für die obigen Problemstellungen die mittleren Läufe der beiden parallelen Varianten mit und ohne Austausch zusammen aufgetragen. Wie der Tabelle 10.2 entnommen werden kann, erweist sich der hier benutzte Wert $\Delta c = 13$ Generationen als günstiger gegenüber $\Delta c = 23$. Ein wesentlich kleinerer

Strategie	Δc	ABFV	Strategie	Δc	ABFV
(1, 100)/(15, 100)	—	89.8	(1, 100)/(15, 100)	—	602.4
(1, 100)/(15, 100)	13	87.9	(1, 100)/(15, 100)	13	406.6
(1, 100)/(15, 100)	23	88.6	(1, 100)/(15, 100)	23	426.6
(1, 100)	—	118.3	(1, 100)	—	611.8
(15, 100)	—	93.1	(15, 100)	—	708.6
(15, 200)	—	79.2	(15, 200)	—	118.6
(20, 200)	—	76.8	(20, 200)	—	107.7

Tabelle 10.2: ABFV verschiedener ES auf der Funktion $f_{\text{Rastrigin2}}$ mit $n = 10, \Delta g = 1, s = 1$, Bewegungstyp III (links) bzw. $f_{\text{Rastrigin}}$ mit $n = 100, \Delta g = 50, s = 1$, Bewegungstyp I (rechts). Der Austausch erfolgte alle Δc Generationen.

Wert als $\Delta c = 13$ erscheint nicht sinnvoll, da dann keiner der beiden Algorithmen Gelegenheit besitzt, sich adäquat an die Problemstellung anzupassen. Bei der ersten Funktion $f_{\text{Rastrigin2}}$ kann die kombinierte Strategie mit Austausch keine besondere Leistungssteigerung erreichen. Dies ist zwar bei der zweiten benutzten Funktion $f_{\text{Rastrigin}}$ so, trotz der positiven Entwicklung ist es der parallelen Variante mit Austausch aber auch auf dieser Funktion nicht gelungen, besser als die (20, 200)-ES abzuschneiden.

¹Genauer formuliert bleibt die Objektvariable aufgrund der adaptierten kleinen Schrittweiten relativ konstant. Es verändert sich die Koordinate des sich bewegenden Optimums, sodass sich die Differenz zwischen beiden aus dem Bereich der Multimodalität bewegt, also $x_1 - x_1^* > |A\omega/2|$.

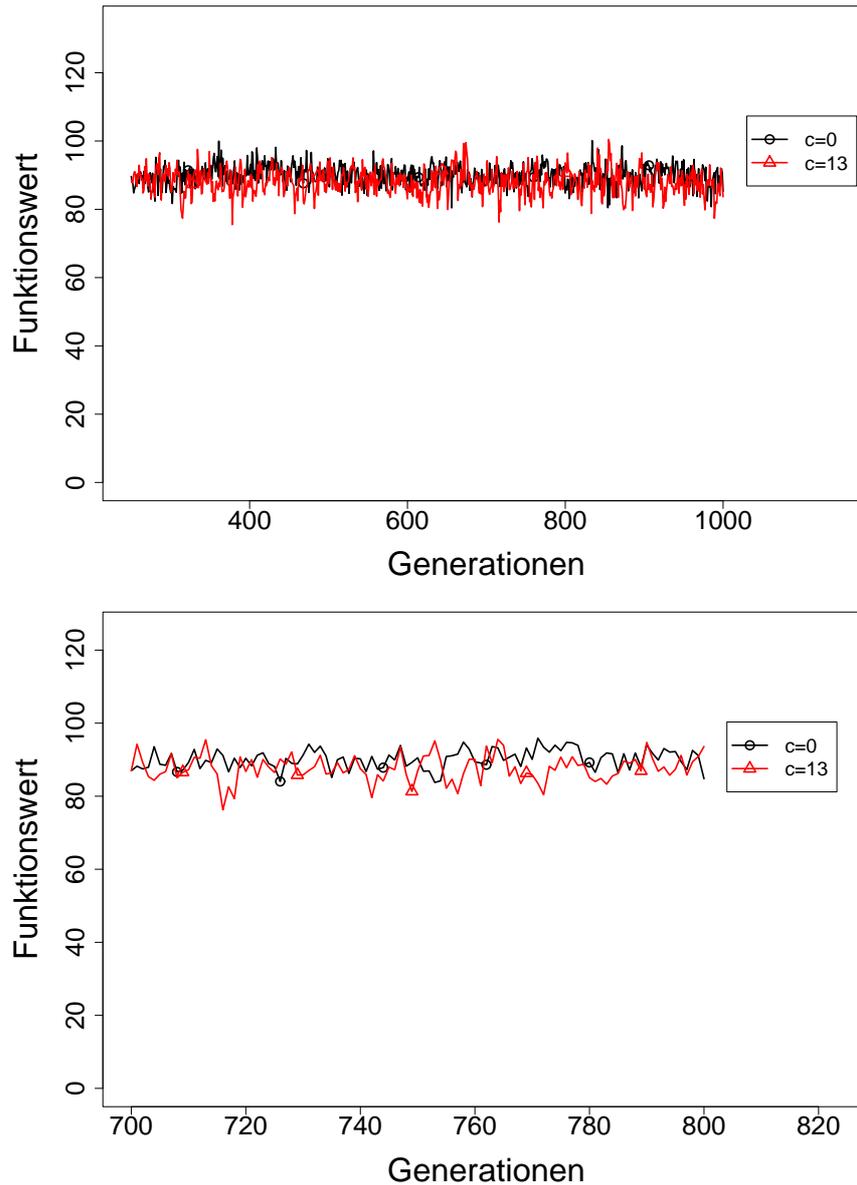


Abbildung 10.10: Mittlerer Lauf einer parallelen $(1, 100)/(15, 100)$ -ES mit und ohne Austausch auf der 10-dimensionalen Rastrigin-Funktion $f_{\text{Rastrigin2}}$ für den Bewegungstyp III bei einer Bewegungsfrequenz $\Delta g = 1$ und einer Bewegungsstärke $s = 1$. In der parallelen Variante mit Austausch erfolgte dieser alle $\Delta c = 13$ Generationen.

10.4 Fazit

Die kombinierte Strategie PEACE erreicht definitionsgemäß mindestens die gleichen Ergebnisse wie die beiden zugrunde liegenden Einzelstrategien. Oftmals kann sie sogar bessere Ergebnisse erzielen als die bessere der beiden Strategien. Was den Austausch von Informationseinheiten anbelangt, erzielt die kombinierte Strategie mit Austausch in der Regel bessere Ergebnisse als die ohne Austausch. Wird die Austauschfrequenz Δc so gewählt, dass den Einzelstrategien genügend Zeit zur Adaptation gelassen wird, so ist die kombinierte Strategie mit Austausch der ohne zu bevorzugen. Allerdings bleiben auch mit dieser Variante die strukturellen Resultate gleich. D.h. die Ergebnisse werden negativ durch die steigende Problemdimension und Bewegungsstärke beeinflusst. Eben-

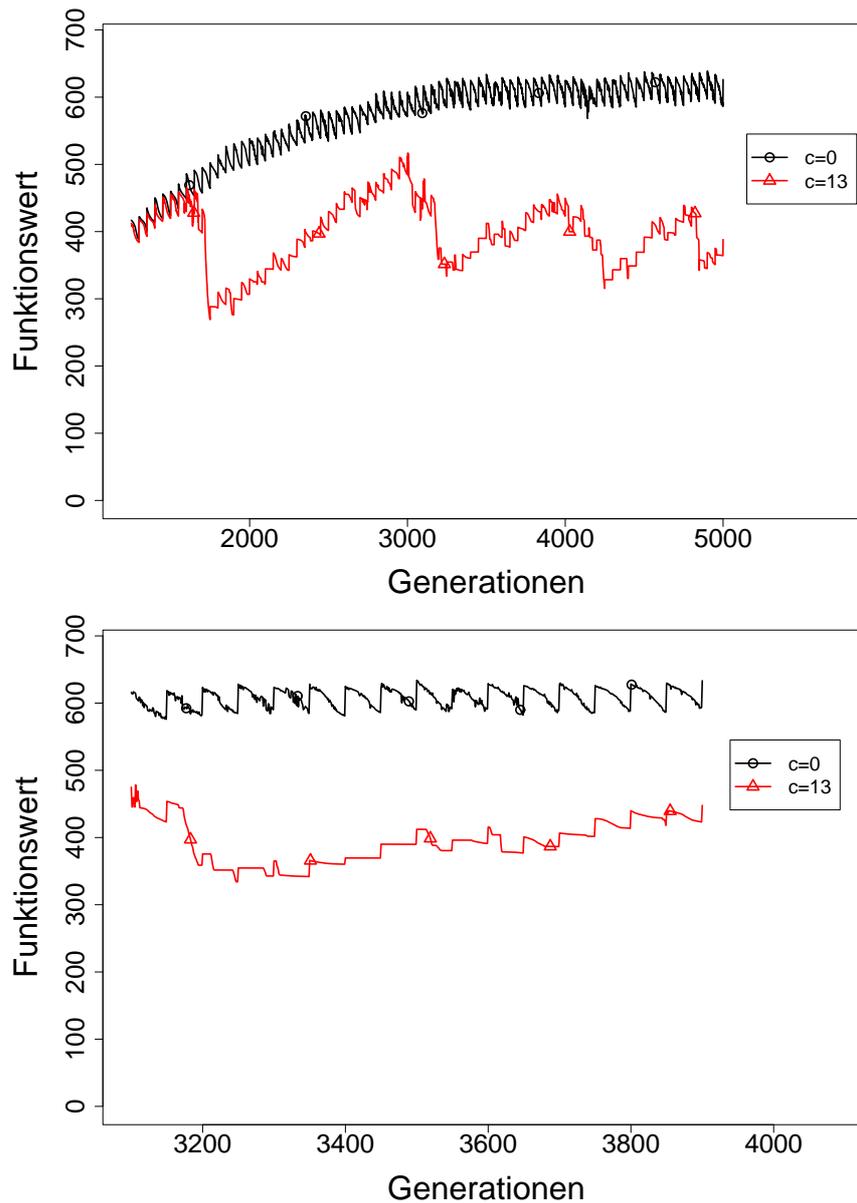


Abbildung 10.11: Mittlerer Lauf der parallelen $(1, 100)/(15, 100)$ -ES mit und ohne Austausch auf der 100-dimensionalen Rastrigin-Funktion $f_{\text{Rastrigin}}$ für den Bewegungstyp I bei einer Bewegungsfrequenz $\Delta g = 50$ und einer Bewegungsstärke $s = 1$. In der parallelen Variante mit Austausch erfolgte dieser alle $\Delta c = 13$ Generationen.

so gestaltet sich die Optimierung aufgrund einer erhöhten Multimodalität schwieriger. Daneben sind die Ergebnisse der parallelen Strategie im Allgemeinen schlechter als die einer panmiktischen Einzelstrategie mit gleicher Anzahl Nachkommen pro Generation. Für dynamische Umgebungen lohnt sich also der Einsatz einer kombinierten Strategie nicht. Stattdessen sollte eine parallele Architektur voll ausgenutzt werden, indem eine panmiktische Einzelstrategie mit maximaler Anzahl Nachkommen verwendet wird.

Zusammenfassung und Ausblick

Im Folgenden sollen die in dieser Arbeit herausgefundenen Resultate nochmals explizit aufgeführt werden. Bisherige vereinzelte Vermutungen wurden hier zum ersten Mal empirisch untermauert. Im Einzelnen lauten die Resultate wie folgt.

Die Verwendung der in dieser Arbeit in den meisten Experimenten eingesetzten ES mit n Schrittweiten empfiehlt sich aufgrund der grundsätzlichen Untersuchungen im Kapitel 6. Es zeigt sich, dass in dynamischen Umgebungen die Verwendung von n Schrittweiten in den meisten Fällen bessere Ergebnisse liefert als eine ES mit nur einer Schrittweite. In den anderen Situationen ist die Variante mit mehr Schrittweiten der anderen entweder gleichwertig oder nur wenig unterlegen. Viele der nachfolgenden Aussagen gelten aber auch für eine Variante mit nur einer Schrittweite, wobei aber die erreichbaren Ergebnisse dann meistens schlechter sind als die mit n Schrittweiten erzielbaren. Ein Lernen einer optimalen Anzahl von Schrittweiten und die dazugehörige Zuordnung der Schrittweiten zu den einzelnen Koordinaten kann für dynamische Umgebungen aufgrund der Untersuchungen nicht empfohlen werden.

In motivierenden Voruntersuchungen für statische Umgebungen wurde festgestellt, dass sich die Optimierung mit zunehmender Problemdimension immer schwieriger gestaltet. Dabei bestätigte sich, dass eine Evolutionsstrategie mit einer oftmals als Standard verwendeten Parametrisierung in hohen Problemdimensionen nicht günstig ist. Mit zunehmender Problemdimension werden die mit diesen Einstellungen erreichbaren Ergebnisse immer schlechter. Ab einer gewissen Problemdimension tritt dann sogar eine Divergenz auf. Zur Verhinderung der Divergenz ist es bei zunehmender Problemdimension notwendig, die Populationsgrößen immer größer zu wählen. Dies gilt noch moderat für die Anzahl der Eltern, trifft aber wesentlich stärker für die Anzahl der Nachkommen zu. Im Übergangsbereich zeichnen sich die einzelnen Läufe durch starke, teils extreme Schwankungen aus. Im Gegensatz dazu bietet die Darstellung der durchschnittlich erreichten Funktionswerte einen sehr gleichmäßigen Verlauf. Das hier entwickelte neue Gütemaß des Average Best Function Value (ABFV) wird von einem homogenen Band der Quantile umschlossen. Der ABFV wurde dabei in dem einführenden ersten Teil dieser Arbeit unter dem Aspekt entwickelt, dass er wesentlich robuster gegenüber großen Ausreißern ist, wie sie häufiger in dynamischen Umgebungen auftreten. In diesen Abschnitten wurden darüber hinaus Gemeinsamkeiten und Unterschiede zur Optimierung bei gestörten Zielfunktionen aufgezeigt.

In den folgenden Experimenten wurde zunächst intensiv untersucht, welche Einflussfaktoren die Optimierung in dynamischen Umgebungen schwierig machen. Hierzu zählt zuvorderst eine große Bewegungsstärke bei sonst gleichen Bedingungen. Dabei sind einerseits die erreichbaren Funktionswerte schlechter als bei einer kleineren Bewegungsstärke. Andererseits setzt für große Bewegungsstärken der Bereich der Problemdimension etwas früher ein, ab dem eine Divergenz zu verzeichnen ist. Dagegen sind die Unterschiede zwischen der Optimierung einer Funktion mit konstanter Bewegungsstärke und einer Funktion mit einer zu der Bewegungsstärke in einem angemessenen

Verhältnis variierenden Bewegungsstärke relativ klein. Die Unterschiede verschwinden sogar, wenn sich das Optimum in allen Dimensionen bewegt. Erkennbare Unterschiede im Schwierigkeitsgrad der Optimierung ergeben sich aber aufgrund der Bewegungsart. Hier ist eine Funktion leichter zu optimieren, wenn sie sich in nur einer Dimension bewegt anstatt in allen. Wird der Wert der Bewegungsfrequenz (Δg) bei gleicher zurückgelegter Gesamtdistanz erhöht, verschlechtern sich die erreichbaren Ergebnisse ebenfalls.

Unterschiedliche Initialisierungen der Objektvariablen oder Schrittweiten können das globale Verhalten der ES nicht wesentlich beeinflussen. Zwar hat die Initialisierung einen nicht unerheblichen Einfluss auf die in einer vorgegebenen Zeit erreichbaren Ergebnisse. Allerdings kann dadurch nicht beeinflusst werden, ob eine Konvergenz oder Divergenz stattfindet. In dynamischen Umgebungen ist auch keine frühzeitige Stagnation in einem weit vom globalen Optimum entfernt liegenden lokalen Optimum feststellbar.

Demgegenüber führt eine unterschiedliche Einstellung bestimmter Strategieparameter zu einem wesentlich differenzierteren Verhalten der ES. Dies gilt zum einen für die Lernraten als auch für die in dieser Arbeit näher betrachteten Populationsgrößen. Der Einfluss unterschiedlicher Lernraten auf die Optimierungsergebnisse in statischen Umgebungen wurde von Kursawe (1999) gezeigt. In der vorliegenden Arbeit wurde darüber hinaus untersucht, wie mit einer differenzierten Wahl der Lernraten auch der Bereich der Problemdimension der beginnenden Divergenz verschoben werden kann. Dies gilt gleichermaßen für statische als auch dynamische Umgebungen. Hierbei haben sich besonders kleine Lernraten als vorteilhaft erwiesen. Allerdings führen diese aber auch zu niedrigeren Konvergenzgeschwindigkeiten für kleine Problemdimensionen, in denen auch für große Lernraten keine Divergenz beobachtet wurde.

Für die Wahl der Populationsgrößen auf dem dynamischen Kugelmodell können folgende Aussagen getroffen werden. Mit steigender Problemdimension erweisen sich größere Populationsparameter als notwendig. Dabei bleibt der Wertebereich guter Einstellungen für die Anzahl der Eltern μ aber relativ moderat. Im Gegensatz dazu steigt der Bereich guter Einstellungen für die Anzahl der Nachkommen überproportional, das heißt es erhöht sich auch der notwendige Selektionsdruck. Wie schon im statischen Kugelmodell kann mit tendenziell größeren Populationsgrößen insbesondere auch der Bereich der beginnenden Divergenz nach hinten verschoben werden. Eine funktionale Abhängigkeit zwischen den Problemparametern (wie z.B. der Problemdimension) und den „optimalen“ Populationsgrößen ließ sich aber auch für eine einzelne Testfunktion nicht erkennen.

Bei den Experimenten auf den hier getesteten multimodalen Funktionen haben wir festgestellt, dass sich die Aussagen zur Optimierung des Kugelmodells auch auf die Ackley-Funktion übertragen lassen. Die Optimierung der Rastrigin-Funktion gestaltet sich dagegen aufgrund der stärkeren Multimodalität wesentlich schwieriger. Sehr viele Strategien bleiben im statischen Fall frühzeitig in einem lokalen Optimum stecken. Dies gilt auch für den dynamischen Fall, wenn die Bewegungsfrequenz hoch ist. Bei einer nahezu kontinuierlichen Bewegung (Δg klein) wird die Komplexität der Optimierung der multimodalen Funktion durch die Dynamik überlagert und die Eigenschaften der Optimierung des Kugelmodells kommen stärker zum Tragen. Die beginnende Divergenz liegt für die Rastrigin-Funktion in einem ähnlichen Bereich der Problemdimension wie für die beiden anderen Funktionen. Dennoch tritt dieser Aspekt bei der Rastrigin-Funktion erkennbar früher auf. Insgesamt erweisen sich auf der Rastrigin-Funktion aber wesentlich größere Populationsgrößen als vorteilhaft.

Für alle drei benutzten Testfunktionen ergibt sich eine hohe Korrelation zwischen dem Funktionswert des besten Individuums einer Generation und der Diversität der Population. Die geeignete

Wahl des Diversitätsmaßes spiegelte dabei direkt die Eigenschaften der jeweiligen Funktion wider. Zukünftig wird es interessant sein, diese Korrelation zum Erkennen einer frühzeitigen Stagnation in einem lokalen Optimum zu nutzen und somit das lokale Optimum zu verlassen.

Aus den Ergebnissen der grundlegenden Experimente auf den drei Testfunktionen im statischen und dynamischen Fall wurde anschließend die allgemein verwendbare Strategie PEACE für eine parallele Rechnerarchitektur entwickelt. Die kombinierte Strategie mit Austausch erreicht meistens bessere Ergebnisse als die bessere der beiden Einzelstrategien. Eine kombinierte Strategie mit Austausch von Informationseinheiten erzielt in der Regel bessere Ergebnisse als die ohne Austausch, wenn die Austauschfrequenz Δc so gewählt wird, dass den Einzelstrategien genügend Zeit zur Adaptation gelassen wird. Die strukturellen Resultate bleiben jedoch gleich. Darüber hinaus ist ein wesentlicher Aspekt, dass die Ergebnisse der parallelen Strategie im Allgemeinen schlechter sind als die einer panmiktischen Einzelstrategie mit gleicher Anzahl Nachkommen pro Generation. Auch wenn für den Einsatz einer kombinierten Strategie viele über die getesteten Varianten hinausgehende Möglichkeiten existieren, scheint sich für dynamische Umgebungen also der Einsatz einer kombinierten Strategie nicht zu lohnen. Stattdessen sollte eine parallele Architektur voll ausgenutzt werden, indem eine panmiktische Einzelstrategie mit maximaler Anzahl Nachkommen verwendet wird.

Sowohl für die Wahl guter Einstellungen der Populationsgrößen als auch der Beobachtung der sonstigen Zusammenhänge war das qualitative Verhalten auf den in dieser Arbeit gewählten Testfunktionen nahezu identisch. Anders wird es sicherlich aussehen, wenn in der Zukunft weitere Testfunktionen verwendet werden, die anderen Problemklassen zuzuordnen sind. Eine solche Testfunktion sollte insbesondere durch eine Asymmetrie gekennzeichnet sein. Interessante Einblicke können auch durch Untersuchungen gewonnen werden, bei denen sich das Optimum nach jeder Funktionsauswertung verschiebt. Statistische Schwankungen könnten sich hier verstärkt bemerkbar machen, wenn im Funktionsdiagramm der Funktionswert nach jeder Auswertung geschrieben wird, die neue Elternpopulation aber generationsweise selektiert wird. Wird das Optimum nach jeder Funktionsauswertung verschoben, ändert sich auch die Fitness des bisher besten Individuums und die Abweichung vergrößert sich mit jeder weiteren Funktionsauswertung vor der nächsten Selektion. Findet die Selektion erst nach einer sehr großen Anzahl von Funktionsauswertungen statt (entspricht in der Regel einer großen Anzahl Nachkommen), ist eine größere durchschnittliche Abweichung zu erwarten.

Nicht zuletzt ist die Untersuchung anderer als die hier getesteten Bewegungsformen interessant. Zu diesen Formen gehören neben zyklischen oder kreisförmigen Bewegungen auch Änderungen des Funktionstyps wie die von Kleefeld (2004) angesprochenen Funktionstransformationen. Die Betrachtung von Funktionen, die durch extreme Änderungen gekennzeichnet sind, verspricht weitere spannende Erkenntnisse. Hier wäre genauer zu betrachten, wann es sich nicht mehr lohnt, eine ES weiter fortzusetzen, sondern eine ES mit neuer Initialisierung zu starten.

Ein wesentlicher Aspekt bei jeglicher Optimierung muss natürlich die Verbesserung bestehender und die Entwicklung neuer Algorithmen sein. Dazu sollten genauere Einblicke in die Verfahrensweise derzeitiger Verfahren auf den eben geschilderten Problemstellungen hilfreich sein. Ein viel versprechender Ansatz ist der im Kapitel 9 angedeutete Zusammenhang zwischen der Diversität und dem Funktionswert. Die Nutzung dieser Informationen wird insbesondere in statischen Umgebungen eine beginnende frühzeitige Konvergenz gegen ein lokales Optimum zu erkennen und die Stagnation zu vermeiden helfen.

Momentan findet die mehrkriterielle Optimierung immer größere Aufmerksamkeit. Informativ wären Untersuchungen auch für diese Problemstellungen. Dabei ist jedoch zu beachten, dass in aktuellen mehrkriteriellen EA ohnehin bestimmte, eine gewisse Diversität sichernde Populationsgrößen nicht unterschritten werden sollten. In der statischen mehrkriteriellen Optimierung gilt es auch, die Pareto-Front möglichst in ihrer Gesamtbreite zu approximieren. Die hierdurch automatisch vorhandene Diversität sollte auch in dynamischen Umgebungen nutzbar sein, damit sich ständig zumindest einige Individuen in der Nähe der neuen Pareto-Front befinden.

Literatur

- M. Affenzeller und S. Wagner (2004). SASEGASA: A new generic parallel evolutionary algorithm for achieving highest quality results. *Journal of Heuristics - Special Issue on New Advances on Parallel Meta-Heuristics for Complex Problems 10*, 239–263.
- P. J. Angeline (1995). Adaptive and self-adaptive evolutionary computation. In M. Palaniswami, Y. Attikiouzel, R. Marks, D. Fogel und T. Fukuda (Hrsg.), *Computational Intelligence: A Dynamic System Perspective*, S. 152–163. IEEE Press, Piscataway, NJ.
- P. J. Angeline (1997). Tracking extrema in dynamic environments. In P. J. Angeline, R. G. Reynolds, J. R. McDonnell und R. Eberhart (Hrsg.), *Proc. of the Sixth Int'l Conf. on Evolutionary Programming*, S. 335–345. Springer, Berlin.
- P. J. Angeline (Hrsg.) (1999). *Int'l Congress on Evolutionary Computation 1999 (CEC'99)*. IEEE Press, Piscataway, NJ.
- T. Bäck (1992). Self-adaptation in genetic algorithms. In F. J. Varela und P. Bourguine (Hrsg.), *Proc. of the 1st European Conf. on Artificial Life*, S. 263–271. MIT Press, Cambridge, MA.
- T. Bäck (1996). *Evolutionary Algorithms in Theory and Practice*. Oxford University Press, New York.
- T. Bäck (1997). Self-adaptation. In Bäck, Fogel und Michalewicz (1997), S. C7.1:1–15.
- T. Bäck (1998). An overview of parameter control methods by self-adaptation in evolutionary algorithms. *Fundamenta Informaticae* (35), 51–66.
- T. Bäck (1999). Self-adaptive genetic algorithms for dynamic environments with slow dynamics. In Banzhaf, Daida, Eiben, Garzon, Honavar, Jakiela, und Smith (1999), S. 142–145.
- T. Bäck, D. B. Fogel und Z. Michalewicz (Hrsg.) (1997). *Handbook of Evolutionary Computation*. Oxford University Press, New York, and Institute of Physics Publishing, Bristol.
- T. Bäck und U. Hammel (1994). Evolution strategies applied to perturbed objective functions. In Z. Michalewicz, J. D. Schaffer, H.-P. Schwefel, D. B. Fogel und H. Kitano (Hrsg.), *Proc. of the First IEEE Conf. on Evolutionary Computation*, S. 40–45. IEEE Press, Piscataway, NJ.
- T. Bäck und H.-P. Schwefel (1996). Evolutionary computation: An overview. In *Proc. of the 1996 IEEE Int'l Conf. on Evolutionary Computation*, S. 20–29. IEEE Press, Piscataway, NJ.
- T. Bäck und M. Schütz (1996). Intelligent mutation rate control in canonical genetic algorithms. In Z. Ras und M. Michalewicz (Hrsg.), *Foundations of Intelligent Systems. 9th Int'l Symposium (ISMIS'96)*, S. 158–167. Springer, Berlin.
- W. Banzhaf, J. Daida, A. E. Eiben, M. H. Garzon, V. Honavar, M. Jakiela und R. E. Smith (Hrsg.) (1999). *Proc. of the Genetic and Evolutionary Computation Conf. GECCO'99*. Morgan Kaufmann, San Francisco, CA.

- R. S. Barr, B. L. Golden, J. Kelly, W. R. Stewart und M. G. C. Resende (1995). Designing and Reporting on Computational Experiments with Heuristic Methods. Technical report, Southern Methodist University, Department of Computer Science and Engineering, Dallas, TX. (Extended version of *Journal of Heuristics* 1(1)1995, S. 9–32 [cited: April 29, 2003], <http://faculty.smu.edu/barr/pubs/johReport.pdf>).
- A. Barry (Hrsg.) (2003). *2003 Genetic and Evolutionary Computation Conf. Workshop Program*.
- M. A. E. Beltagy, P. B. Nair und A. J. Keane (1999). Metamodeling techniques for evolutionary optimization of computationally expensive problems: Promises and limitations. In Banzhaf, Daida, Eiben, Garzon, Honavar, Jakiela, und Smith (1999), S. 196–203.
- H.-G. Beyer (1995). Toward a theory of evolution strategies: On the benefits of sex – the $(\mu/\mu, \lambda)$ -theory. *Evolutionary Computation* 3(1), 81–111.
- H.-G. Beyer (1996). Toward a theory of evolution strategies: Self-adaptation. *Evolutionary Computation* 3(3), 311–347.
- H.-G. Beyer (1997). An alternative explanation for the manner in which genetic algorithms operate. *BioSystems* 41, 1–15.
- H.-G. Beyer (2000). Evolutionary algorithms in noisy environments: Theoretical issues and guidelines for practice. *CMAME (Computer Methods in Applied Mechanics and Engineering)* 186, 239–267.
- H.-G. Beyer (2001). *The Theory of Evolution Strategies*. Springer, Berlin.
- H.-G. Beyer, U.-M. O’Reilly, D. V. Arnold, W. Banzhaf, C. Blum, E. W. Bonabeau, E. Cantú-Paz, D. Dasgupta, K. Deb, J. A. Foster, E. D. de Jong, H. Lipson, X. Llorca, S. Mancoridis, M. Pelikan, G. R. Raidl, T. Soule, A. Tyrrell, J.-P. Watson und E. Zitzler (Hrsg.) (2005). *Proc. of the Genetic and Evolutionary Computation Conf. (GECCO 2005)*. ACM Press, New York.
- H.-G. Beyer und H.-P. Schwefel (2002). Evolution strategies: A comprehensive introduction. *Natural Computing* 1(1), 3–52.
- J. Branke (1998). Creating robust solutions by means of evolutionary algorithms. In Schoenauer, Schwefel, Eiben und Bäck (1998), S. 119–128.
- J. Branke (1999). Evolutionary approaches to dynamic optimization problems – a survey. In Banzhaf, Daida, Eiben, Garzon, Honavar, Jakiela, und Smith (1999), S. 134–137.
- J. Branke (2002). *Evolutionary optimization in dynamic environments*. Kluwer, Dordrecht, NL.
- J. Branke und H. Schmeck (2003). Designing evolutionary algorithms for dynamic optimization problems. In S. Tsutsui und A. Ghosh (Hrsg.), *Theory and Application of Evolutionary Computation: Recent Trends*, S. 239–262. Springer, Berlin.
- H. Büning und G. Trenkler (1994). *Nichtparametrische statistische Methoden* (2. Auflage). de Gruyter, Berlin.
- E. Cantú-Paz (Hrsg.) (2003). *Proc. of the Genetic and Evolutionary Computation Conf. GECCO 2003*. Springer, Berlin.
- D. Cvetković und H. Mühlenbein (1994). The optimal population size for uniform crossover and truncation selection. Technischer Bericht, GMD-AS-TR-94-11, GMD, St. Augustin.
- K. A. De Jong (1975). *An analysis of the behavior of a class of genetic adaptive systems*. Dissertation, University of Michigan.

- K. A. De Jong (1999). Evolving in a changing world. In Ras und Skowron (1999), S. 513–519.
- S. Droste (2003). Analysis of the (1 + 1) EA for a dynamically bitwise changing OneMax. In Cantú-Paz (2003), S. 909–921.
- G. Dueck und T. Scheuer (1990). Threshold accepting: A general purpose optimization algorithm appearing superior to simulated annealing. *Journal of Computational Physics* 90(1), 161–175.
- Á. E. Eiben und M. Jelasity (2002). A critical note on experimental research methodology in EC. In D. B. Fogel, M. A. El-Sharkawi, X. Yao, G. Greenwood, H. Iba, P. Marrow und M. Shackleton (Hrsg.), *Proc. of the 2002 Congress on Evolutionary Computation (CEC 2002)*, S. 582–587. IEEE Press, Piscataway, NJ.
- D. B. Fogel (1991). Meta-evolutionary programming. In R. R. Chen (Hrsg.), *Proc. of the 25th Asilomar Conf. on Signals, Systems and Computers*, S. 540–545. Maple press, San Jose, CA.
- B. Freisleben (1997). Metaevolutionary approaches. In Bäck, Fogel und Michalewicz (1997), S. C7.2:1–8.
- I. A. Gent, S. A. Grant, E. MacIntyre, P. Prosser, P. Shaw, B. M. Smith und T. Walsh (1997). How Not To Do It. Research Report Series, Report 97.27, School of Computer Studies, University of Leeds.
- D. E. Goldberg (1985). Optimal initial population size for binary-coded genetic algorithms. Technical report, TCGA Report No. 85001, University of Alabama.
- D. E. Goldberg (1989a). *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley, Reading, MA.
- D. E. Goldberg (1989b). Sizing populations for serial and parallel genetic algorithms. In J. D. Schaffer (Hrsg.), *Proc. of the 3rd Int'l Conf. on Genetic Algorithms*, Los Altos, CA. Morgan Kaufmann.
- D. E. Goldberg und R. E. Smith (1987). Nonstationary function optimization using genetic algorithms with dominance and diploidy. In Grefenstette (1987), S. 59–68.
- J. J. Grefenstette (Hrsg.) (1987). *Genetic Algorithms and their Applications: Proc. of the Second Int'l Conf. on Genetic Algorithms*, Hillsdale, New Jersey. Lawrence Erlbaum Associates.
- J. J. Grefenstette (1992). Genetic algorithms for changing environments. In Männer und Manderick (1992), S. 137–144.
- J. J. Grefenstette (1999). Evolvability in dynamic fitness landscapes: A genetic algorithm approach. In Angeline (1999), S. 2031–2038.
- M. Gronek (1997). Der Einfluss von Pseudo-Zufallszahlengeneratoren auf das Verhalten Evolutionäre Algorithmen. Diplomarbeit, Universität Dortmund.
- J. Hartung, B. Elpelt und K.-H. Klösener (1999). *Statistik: Lehr- und Handbuch der angewandten Statistik* (12. Auflage). Oldenbourg, München.
- R. Hinterding, Z. Michalewicz und A. E. Eiben (1997). Adaptation in evolutionary computation: A survey. In R. Eberhart, P. Angeline und T. Bäck (Hrsg.), *Fourth IEEE Conf. on Evolutionary Computation*, S. 65–69. IEEE Press, Piscataway, NJ.
- J. N. Hooker (1994). Needed: An empirical science of algorithms. *Operations Research* 2(42), 201–212.

- J. N. Hooker (1995). Testing Heuristics: We Have It All Wrong. *Journal of Heuristics* 1(1), 33–42. Erweiterte Version unter <http://www.gsia.cmu.edu/afs/andrew/gsia/jh38/report96.html>.
- T. Jansen und U. Schellbach (2005). Theoretical analysis of a mutation-based evolutionary algorithm for a tracking problem in the lattice. In Beyer, O'Reilly, Arnold, Banzhaf, Blum, Bonabeau, Cantú-Paz, Dasgupta, Deb, Foster, de Jong, Lipson, Llorca, Mancoridis, Pelikan, Raidl, Soule, Tyrrell, Watson, und Zitzler (2005), S. 841–848.
- T. Jansen und I. Wegener (2001). On the utility of populations. In L. Spector, E. D. Goodman, A. Wu, W. B. Langdon, H.-M. Voigt, M. Gen, S. Sen, M. Dorigo, S. Pezeshk, M. H. Garzon und E. Burke (Hrsg.), *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2001)*, S. 375–382. Morgan Kaufmann, San Francisco, CA.
- T. Jansen und I. Wegener (2002). On the analysis of evolutionary algorithms — a proof that crossover really can help. *Algorithmica* 34(1), 47–66.
- B. Julstrom (1996). A simple estimate of population size in genetic algorithms for the traveling salesman problem. In J. T. Alander (Hrsg.), *Proc. of the Second Nordic Workshop on Genetic Algorithms and their Applications (2NWGA)*, Proc. of the University of Vaasa, Nro. 13, Vaasa, Finland, S. 3–14.
- J. Jägersküpper und C. Witt (2005). Rigorous runtime analysis of a $(\mu + 1)$ ES for the sphere function. In Beyer, O'Reilly, Arnold, Banzhaf, Blum, Bonabeau, Cantú-Paz, Dasgupta, Deb, Foster, de Jong, Lipson, Llorca, Mancoridis, Pelikan, Raidl, Soule, Tyrrell, Watson, und Zitzler (2005), S. 849–856.
- C. L. Karr (1999). An architecture for adaptive process control systems. In Banzhaf, Daida, Eiben, Garzon, Honavar, Jakiela, und Smith (1999), S. 146–148.
- S. Kirkpatrick, C. D. Gelatt und M. P. Vecchi (1983). Optimization by simulated annealing. *Science* (220), 671–680.
- M. Kleefeld (2004). Testfunktionen für die dynamische Optimierung. Technischer Bericht SYS-1/04, Universität Dortmund.
- U. Kockelkorn (2000). *Lineare statistische Methoden*. Oldenbourg, München.
- L. Kreienbrock (1993). *Einführung in die Stichprobenverfahren*. Oldenbourg, München.
- F. Kursawe (1999). *Grundlegende empirische Untersuchungen der Parameter von Evolutionsstrategien — Metastrategien*. Dissertation, Universität Dortmund.
- R. Männer und B. Manderick (Hrsg.) (1992). *Proc. Parallel Problem Solving from Nature (PPSN II)*. North-Holland, Amsterdam.
- S. Markon, D. V. Arnold, T. Bäck, T. Beielstein und H.-G. Beyer (2001). Thresholding - a selection operator for noisy ES. In J.-H. Kim, B.-T. Zhang, G. Fogel und I. Kuscü (Hrsg.), *Proc. of the 2001 Congress on Evolutionary Computation (CEC01)*, S. 465–472. IEEE Service Center, Piscataway, NJ.
- D. C. Mattfeld und C. Bierwirth (1999). Adaptation and dynamic optimization problems: A view from general system theory. In Banzhaf, Daida, Eiben, Garzon, Honavar, Jakiela, und Smith (1999), S. 138–141.
- N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller und E. Teller (1953). Equation of state calculations by fast computing machines. *J. Chem. Phys.* 6(21), 1087–1092.

- Z. Michalewicz, K. Deb, M. Schmidt und T. Stidsen (2000). Testcase generator for nonlinear continuous parameter optimization techniques. *IEEE Transactions on Evolutionary Computation* 3(4), 197–215.
- Z. Michalewicz und D. B. Fogel (2004). *How to Solve It: Modern Heuristics* (2. Auflage). Springer, Berlin.
- N. Mori, H. Kita und Y. Nishikawa (1996). Adaptation to a changing environment by means of the thermodynamical genetic algorithm. In W. Ebeling, I. Rechenberg, H.-P. Schwefel und H.-M. Voigt (Hrsg.), *Proc. Parallel Problem Solving from Nature (PPSN IV)*, S. 513–522. Springer, Berlin.
- N. Mori, H. Kita und Y. Nishikawa (1998). Adaptation to a changing environment by means of the feedback thermodynamical genetic algorithm. In Schoenauer, Schwefel, Eiben und Bäck (1998), S. 149–158.
- R. W. Morrison (2003). Performance measures in dynamic environments. In Barry (2003), S. 99–102.
- R. W. Morrison (2004). *Designing Evolutionary Algorithms for Dynamic Environments*. Springer, Berlin.
- A. Mück (1989). Einfluß verschiedener Wahrscheinlichkeitsverteilungen auf das Konvergenzverhalten von Evolutionsstrategien. Diplomarbeit, Universität Dortmund.
- R. Nakano, Y. Davidor und T. Yamada (1994). Optimal population size under constant computation cost. In Y. Davidor, H.-P. Schwefel und R. Männer (Hrsg.), *Proc. Parallel Problem Solving from Nature (PPSN III)*, S. 130–138. Springer, Berlin.
- F. Oppacher und M. Wineberg (1999). The shifting balance genetic algorithm: Improving the ga in a dynamic environment. In Banzhaf, Daida, Eiben, Garzon, Honavar, Jakiela, und Smith (1999), S. 504–510.
- A. Ostermeier, A. Gawelczyk und N. Hansen (1994). A derandomized approach to self-adaptation of evolution strategies. *Evolutionary Computation* 2(4), 369–380.
- A. Ostermeier und N. Hansen (1999). An evolutionary strategy with coordinate system invariant adaptation of arbitrary normal mutation distributions with the concept of mutative strategy parameter control. In Banzhaf, Daida, Eiben, Garzon, Honavar, Jakiela, und Smith (1999), S. 902–909.
- K. Ott (1993). Einfluß stochastischer Störungen auf das Konvergenzverhalten von Evolutionsstrategien. Diplomarbeit, Universität Dortmund.
- Z. Ras und A. Skowron (Hrsg.) (1999). *Foundations of Intelligent Systems. 11th Int'l Symposium (ISMIS'99)*. Springer, Berlin.
- A. Ratle (1998). Accelerating the convergence of evolutionary algorithms by fitness landscape approximation. In Schoenauer, Schwefel, Eiben und Bäck (1998), S. 87–96.
- I. Rechenberg (1965). *Cybernetic solution path of an experimental problem*. Royal Aircraft Establishment, Farnborough, Library Translation no. 1122.
- I. Rechenberg (1994). *Evolutionstrategie '94*. Frommann-Holzboog, Stuttgart.
- C. R. Reeves (1993). Using genetic algorithms with small populations. In S. Forrest (Hrsg.), *Proc. of the Fifth Int'l Conf. on Genetic Algorithms (ICGA'93)*, San Mateo, CA, S. 92–99. Morgan Kaufmann.

- G. Rudolph (1990). Globale Optimierung mit parallelen Evolutionsstrategien. Diplomarbeit, Universität Dortmund.
- G. Rudolph (1992). On correlated mutations in evolution strategies. In Männer und Manderick (1992), S. 105–114.
- G. Rudolph (1997). Local convergence rates of simple evolutionary algorithms with cauchy mutations. *IEEE Transactions on Evolutionary Computation* 1(4), 249–258.
- G. Rudolph (1998). Asymptotical convergence rates of simple evolutionary algorithms under factorizing mutation distributions. In J.-K. Hao, E. Lutton, E. Ronald, M. Schoenauer und D. Snyers (Hrsg.), *Artificial Evolution: Third European Conf. (AE'97)*, S. 223–233. Springer, Berlin.
- G. Rudolph (1999). Self-adaptation and global convergence: A counter-example. In Angeline (1999), S. 646–651. (Revised version).
- G. Rudolph (2001). Self-adaptive mutations may lead to premature convergence. *IEEE Transactions on Evolutionary Computation* 5(4), 410–414.
- T. P. Runarsson (2000). *Evolutionary Problem Solving*. Dissertation, University of Iceland, Faculty of Engineering, Reykjavik.
- L. Sachs (1997). *Angewandte Statistik: Anwendung statistischer Methoden* (8. Auflage). Springer, Berlin.
- J. Sarma und K. A. De Jong (1999). The behavior of spatially distributed evolutionary algorithms in non-stationary environments. In Banzhaf, Daida, Eiben, Garzon, Honavar, Jakiela, und Smith (1999), S. 572–578.
- J. D. Schaffer und A. Morishima (1987). An adaptive crossover distribution mechanism for genetic algorithms. In Grefenstette (1987), S. 36–40.
- D. Schlierkamp-Voosen (1999). *Populationsbasierte Wettbewerbsmodelle zur Strategieveränderung in Evolutionären Algorithmen*. Dissertation, Universität Dortmund.
- R. Schlittgen (2001). *Angewandte Zeitreihenanalyse*. Oldenbourg, München.
- R. Schlittgen und B. Streitberg (1989). *Zeitreihenanalyse* (3. Auflage). Oldenbourg, München.
- M. Schoenauer, H.-P. Schwefel, A. E. Eiben und T. Bäck (Hrsg.) (1998). *Proc. Parallel Problem Solving from Nature (PPSN V)*. Springer, Berlin.
- L. Schönemann (1999). Einfluß stochastischer Selektionsverfahren auf das Konvergenzverhalten von Evolutionsstrategien. Technischer Bericht SYS-1/99, Universität Dortmund.
- L. Schönemann (2003). On the influence of population sizes in evolution strategies in dynamic environments. In Barry (2003), S. 123–127.
- L. Schönemann (2004a). The impact of population sizes and diversity on the adaptability of evolution strategies in dynamic environments. In G. W. Greenwood, G. Fogel, S. Louis und T. English (Hrsg.), *Proc. of the 2004 Congress on Evolutionary Computation (CEC 2004)*, S. 1270–1277. IEEE Press, Piscataway, NJ.
- L. Schönemann (2004b). Maximal life span in evolutionary algorithms. In B. Filipič und J. Šilc (Hrsg.), *Proc. Int'l Conf. Bioinspired Optimization Methods and Their Applications (BIOMA'04)*, Ljubljana, Slowenien, S. 21–30. Jožef Stefan Institute.
- H.-P. Schwefel (1965). *Kybernetische Evolution als Strategie der experimentellen Forschung in der Strömungstechnik*. Diplomarbeit, Technische Universität Berlin.

- H.-P. Schwefel (1975). *Evolutionsstrategie und numerische Optimierung*. Dissertation, TU Berlin, Fachbereich Verfahrenstechnik.
- H.-P. Schwefel (1987). Collective phenomena in evolutionary systems. In P. Checkland und I. Kiss (Hrsg.), *Problems of Constancy and Change - The Complementarity of Systems Approaches to Complexity, Papers presented at the 31st Annual Meeting Int'l Soc. General System Research*, Band 2, Budapest, S. 1025–1033. Int'l Soc. for General System Research.
- H.-P. Schwefel (1988). Evolutionary learning optimum-seeking on parallel computer architectures. In A. Sydow, S. G. Tzafestas und R. Vichnevetsky (Hrsg.), *Proceedings of the Int'l Symposium on Systems Analysis and Simulation 1988*, Band I: Theory and Foundations, S. 217–225. Akademie-Verlag, Berlin.
- H.-P. Schwefel (1995). *Evolution and Optimum Seeking*. Wiley, New York.
- H.-P. Schwefel und F. Kursawe (1998). On natural life's tricks to survive and evolve. In D. B. Fogel, H.-P. Schwefel, T. Bäck und X. Yao (Hrsg.), *Proceedings of the Second IEEE World Congress on Computational Intelligence (WCCI '98) with Fifth IEEE Conference on Evolutionary Computation (ICEC '98)*, Band 1, S. 1–8. IEEE Press, Piscataway, NJ.
- H.-P. Schwefel und G. Rudolph (1995). Contemporary evolution strategies. In F. Morán, A. Moreno, J. J. Merelo und P. Chacón (Hrsg.), *Advances in Artificial Life, Proc. of the Third European Conference on Artificial Life*, S. 893–907. Springer, Berlin.
- R. E. Smith (1997). Population size. In Bäck, Fogel und Michalewicz (1997), S. E1.1:1–5.
- J. Sprave (1999). *Ein einheitliches Modell für Populationsstrukturen in Evolutionären Algorithmen*. Dissertation, Universität Dortmund.
- P. N. Suganthan (2005, [cited: April 1, 2005]). Test functions [online]. Available from Internet: [//staffx.webstore.ntu.edu.sg/personal/epnsugan/SharedDocuments/CEC2005/Intro-2-funs-09-Mar-05.pdf](http://staffx.webstore.ntu.edu.sg/personal/epnsugan/SharedDocuments/CEC2005/Intro-2-funs-09-Mar-05.pdf).
- A. Törn und A. Žilinskas (1989). *Global Optimization*. Berlin: Springer.
- K. Trojanowski und Z. Michalewicz (1999a). Evolutionary approach to non-stationary optimisation tasks. In Ras und Skowron (1999), S. 538–546.
- K. Trojanowski und Z. Michalewicz (1999b). Searching for optima in non-stationary environments. In Angeline (1999), S. 1843–1850.
- R. K. Ursem (2000). Multinational GAs: Multimodal optimization techniques in dynamic environments. In D. Whitley, D. Goldberg, E. Cantú-Paz, L. Spector, I. Parmee und H.-G. Beyer (Hrsg.), *Proc. of the Genetic and Evolutionary Computation Conf. GECCO 2000*, S. 19–26. Morgan Kaufmann, San Francisco, CA.
- K. Weicker (2000). An analysis of dynamic severity and population size. In M. Schoenauer, K. Deb, G. Rudolph, X. Yao, E. Lutton, J. J. Merelo und H.-P. Schwefel (Hrsg.), *Proc. Parallel Problem Solving from Nature (PPSN VI)*, S. 159–168. Springer, Berlin.
- K. Weicker (2002). Performance measures for dynamic environments. In J. J. M. Guervós, P. Adamiadis, H.-G. Beyer, J. L. Fernández-Villacañas und H.-P. Schwefel (Hrsg.), *Proc. Parallel Problem Solving from Nature (PPSN VII)*, S. 64–73. Springer, Berlin.
- K. Weicker und N. Weicker (1999). On evolution strategy optimization in dynamic environments. In Angeline (1999), S. 2039–2046.

- K. Weicker und N. Weicker (2000). Dynamic rotation and partial visibility. In A. Zalzala und R. Eberhart (Hrsg.), *Proc. of the 2000 Congress on Evolutionary Computation (CEC'00)*, Band 2, S. 1125–1131. IEEE Press, Piscataway, NJ.
- M. Wineberg und F. Oppacher (2003). Distance between populations. In Cantú-Paz (2003), S. 1481–1490.
- C. Witt (2004). *Über die Analyse randomisierter Suchheuristiken und den Entwurf spezialisierter Algorithmen im Bereich der kombinatorischen Optimierung*. Dissertation, Universität Dortmund.
- M. Wörmann (2000). Untersuchungen zum Einfluss der maximalen Lebensdauer auf das Konvergenzverhalten evolutionärer Algorithmen. Diplomarbeit, Universität Dortmund.