Grundlegende empirische Untersuchungen der Parameter von Evolutionsstrategien — Metastrategien

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

von
Frank Kursawe

Dortmund
1999

Tag der mündlichen Prüfung: 20. September 1999

Dekan: Prof. Dr. Bernd Reusch

Gutachter: Prof. Dr. Hans–Paul Schwefel, Prof. Dr. Ingo Wegener

Danksagung

Mein Dank gilt allen, die mich während der Entstehungszeit dieser Arbeit geduldig begleitet haben.

Sabine danke ich für ihre Unterstützung in der Schreibphase.

Dortmund, Juni 1999

Frank Kursawe

Inhaltsverzeichnis

Kapitel 1

Einleitung

Always act so as to increase your options[1].

In praktisch allen Bereichen des Alltags findet man Optimierprobleme, deren Lösung aus unterschiedlichen Gründen interessiert. Entsprechend existiert eine Fülle von Optimierverfahren, die mehr oder weniger auf einzelne Problemklassen zugeschnitten sind.

In Kapitel 2 werden zentrale Begriffe aus der biologischen Evolution sowie die z. T. widersprüchlichen Theorien vorgestellt. Bevor in Kapitel 4 die Hauptrichtungen des Gebietes der Evolutionären Algorithmen und deren, in dieser Arbeit im Mittelpunkt stehende, Unterklasse der Evolutionsstrategien genauer vorgestellt werden, liefert Kapitel 3 eine kurze Vorstellung der traditionellen Optimierverfahren und ihrer Schwächen.

Kapitel 5 untersucht die Leistungsfähigkeit in der Literatur empfohlener Standard–Parametrisierungen für Evolutionsstrategien durch gezielte Variation. Aufgrund der ersten festgestellten Abhängigkeiten der Parameter untereinander wird in Kapitel 6 schließlich eine Möglichkeit vorgestellt, zu besseren Parametrisierungen zu gelangen.

Bevor in Kapitel 8 Wege aufgezeigt werden, in welcher Richtung die Entwicklung der Evolutionsstrategien fortgesetzt werden könnte, gibt Kapitel 7 einen Überblick über die Möglichkeiten, die das Studium dieser Verfahren jenseits der Optimierung bietet.

[1] Alle Zitate zu Beginn eines Kapitels stammen aus (Gall, 1986).

Kapitel 2

Natürliche Evolution

If a system can be exploited, it will be.

Aufgrund des Vorbildes enthält die Terminologie der Evolutionären Algorithmen (s. Kapitel 4) viele Begriffe aus der natürlichen Evolution, die z. T. metaphorisch verwendet werden. Daher erscheint eine kurze biologische Einführung geboten.

Der Begriff *Evolution* stammt nicht — wie vielfach angenommen — von Charles Darwin, sondern von seinem Promoter Herbert Spencer. Darwin verwendete dieses Wort erst in der sechsten Auflage der *Origin of Species*. Ebenso stammt die Redewendung *Survival of the Fittest* von Spencer und nicht von Darwin. Dieser Slogan hat sich wegen seiner Prägnanz und Kürze durchgesetzt und wird bis heute verwendet, obwohl einerseits „Überleben" zu dramatisch formuliert ist und andererseits die Fitneß von Lebewesen in einem absoluten Sinne gar nicht gemessen werden kann und natürlich auch von, u. a. durch diese Lebewesen selbst, veränderbaren Umweltbedingungen abhängt (s. Kapitel 7).

Die moderne (neodarwinistische) *synthetische Theorie* beschreibt Evolution als Kombination aus zufälliger Veränderung und natürlicher Selektion. Jacques Monod sprach von „Zufall und Notwendigkeit" (Monod, 1971). Oft wird die Rolle des Zufalls auf den Mutationsoperator beschränkt; die Selektion läuft vermeintlich deterministisch ab.

2.1 Mikrobiologische Grundlagen

Nur 20 *Aminosäuren* bilden die Grundbausteine aller Eiweiße oder *Proteine*. Durch unterschiedliche Reihenfolge, Länge und Faltung übernehmen sie verschiedene Funktionen. Eine Kette aus zwei bis mehreren hundert Aminosäuren nennt man *Peptid*. Der Übergang zu den Proteinen ist unscharf. Diese können aus Millionen Aminosäuren in aneinandergelagerten Ketten bestehen. Proteine, die andere Stoffwechselprozesse katalysieren, werden *Enzyme* genannt.

Der Bauplan für die Proteine ist in der *Desoxyribonukleinsäure (DNS)* festgeschrieben, die durch die Aufreihung von Tripletts aus nur vier Bausteinen, den Nukleinsäuren oder Nukleotiden Adenin, Cytosin, Guanin und Thymin, alle Proteine kodiert. DNS–Abschnitte, die eine Informationseinheit bilden, nennt man *Gen*, alle Erbanlagen eines Organismus sein *Genom*, angeordnet auf *Chromosomen*. Beim Menschen sind etwa 4 000 Gene auf einem Chromosom zusammengefaßt. Aber nicht alle DNS–Abschnitte werden zur Proteinsynthese herangezogen: Sogenannte repetitive Sequenzen machen mehr als 95% unserer DNS aus. Ob diese, neben der Redundanz des genetischen Codes, der nur 20 von 64 Kodierungsmöglichkeiten realisiert, einen weiteren Puffer gegen Mutationen bilden oder nur Erbgutreste urzeitlicher Viren sind, ist ungeklärt.

In der *Ribonukleinsäure (RNS)* ist gegenüber der DNS die Nukleinsäure Thymin durch Uracil ersetzt. Ferner kommt sie im Gegensatz zur doppelsträngigen DNS nur als Einzelstrang vor. Drei Formen der RNS sorgen für den Aufbau der in der DNS kodierten Proteine:

- Die *Boten–RNS* liest die Information der DNS ab.
- Die *Transfer–DNS* holt die entsprechenden Aminosäuren aus dem Zellinnern, die dann an den Ribosomen zu Proteinen zusammengesetzt werden.
- Die *Ribosomen* enthalten eine dritte Sorte RNS, die für die passende Auswahl der Boten–RNS und die Beendigung des Transfers verantwortlich ist.

Sowohl DNS als auch RNS hängen bei ihrer Bildung von Enzymen ab, die aber erst durch Ablesen der Information auf den Nukleinsäuren gebildet werden können. Dies ist ein molekulares Henne–Ei–Problem, das wohl ungelöst bleiben wird, weil die ersten vermehrungfähigen Proteine keine fossilen Spuren hinterlassen haben können. Eine schlagartige, zufällige „Erfindung“ dieses Mechanismus’ ist äußerst unwahrscheinlich: Schon einfache Proteine bestehen aus etwa 100 Aminosäuren, von denen es 20 gibt. Also sind $20^{100} \approx 10^{130}$ Kombinationen möglich. Einfache Bakterien benötigen 400 solcher Proteine, höhere Lebewesen bis zu 100 000.

Von Leben spricht man üblicherweise erst ab dem Organisationsstadium der Zelle, daher rechnet man Viroide, die nur aus einem, allerdings vermehrungsfähigen RNS–Strang bestehen, und Viren, die entweder aus RNS oder DNS bestehen und einen Wirtsorganismus zur Vermehrung benötigen, nicht zur belebten Welt.

Die heute wohl überzeugendste Theorie zur Entstehung des Lebens auf unserem Planeten basiert lediglich auf der Gültigkeit der Naturgesetze und einer Art „Molekulardarwinismus“, der für immer komplexere Strukturen gesorgt hat (Eigen, 1971; Küppers, 1986).

Diese entstanden nicht um der größeren Komplexität willen, sondern um immer das Überleben durch immer ausgefeilteres Verhalten wahrscheinlicher zu machen: Bleibt die Umwelt eine gewisse Zeit konstant, trifft die Selektion bei immer anspruchsvollerer Konkurrenz durch die parallele Entwicklung der Räuber immer feinere Unterscheidungen des Anpassungsgrades von Individuen. Diese Entwicklung begann nach heutigem Wissen vor etwa 4 Milliarden Jahren, als ein Teil der Materie sich in reproduktiven Strukturen organisierte.

Eine *Mutation* auf molekularer Ebene bedeutet oft, daß ein bestimmtes Protein nicht mehr produziert wird. Hier bieten *diploide* oder sogar *polyploide* Genome, neben den ihrerseits wieder in DNS kodierten Reparaturenzymen, einen weiteren Schutz vor Mutationen: Bleibt das entsprechende Gen im zweiten Chromosomensatz unmutiert, kann das Protein noch synthetisiert werden. Zusätzlich hängt die Wirkungsweise eines Gens auch noch von seiner Umgebung auf dem Chromosom ab. Neben der Genmutation gibt es noch Chromosomen– und Ploidiemutationen, bei denen sich entweder die Struktur eines Chromosoms verändert oder aber deren Anzahl. Eine Chromosomenmutation kann als Deletion, Inversion, Duplikation oder Translokation auftreten. Zu letzterer gehört auch das im übernächsten Abschnitt erwähnte Crossover.

Ging man vor einigen Jahren noch davon aus, daß die Augen von Insekt, Tintenfisch und Säugetieren unabhängig voneinander entstanden sind, also eine Mutation hin zu lichtempfindlichen Zellen mehrfach aufgetreten ist, so gilt neuerdings als gesicherte Erkenntnis, daß diese von den gleichen (Struktur–)Genen hevorgebracht wurden (Grolle und Klein, 1998). Eine plötzliche Vermehrung dieser sogenannten *Hox–Gene* könnte auch die Artenexplosion im Kambrium erklären. Seitdem hat sich wenig getan: Zu den damals entstandenen 37 Grundbauplänen — diese sind in den Hox–Genen kodiert — kam keiner mehr hinzu. Somit können Mutationen wegen dieser Konstanz des genetischen Codes seit „Urzeiten“ nur innerhalb gewisser Grenzen Varianten erzeugen.

Die Erfindung der Diploidie führt zwangsläufig zu den Begriffen *Dominanz* und *Rezessivität*. Mutationen sorgen oft für die Entstehung einer neuen, i. d. R. rezessiven Variante (*Allel*), die sich phänotypisch (s. Kapitel 2.2) erst auswirken kann, wenn sich zwei Individuen mit der gleichen Mutation paaren (*Rekombination*). Auf chromosomaler Ebene heißt dieser Vorgang *Crossover*.

Ohne sexuelle Rekombination vermehren sich 27 Reptilienarten, ferner Wasserflöhe und Sackmotten (Miersch, 1996), d. h. Mütter und Töchter sind identische Klone. Aber 99.9% aller Lebewesen vermehren sich sexuell. Welchen Vorteil bietet diese vergleichsweise aufwendige Methode (Balz, Paarung, Brutpflege ...)? Einerseits sorgt sie für eine weitere „milde“ Variantenquelle, da die Rekombinanten sich ja bereits als lebensfähig erwiesen haben. Andererseits bewirkt die Rekombination auch eine Immunisierung gegen Parasiten: Bestimmte Arten, die eine sexuelle und asexuelle Vermehrung im Repertoire haben, schalten auf die energiesparende und schnellere asexuelle Komponente um, wenn ihnen keine Gefahr durch Parasiten droht. Mit Parasitenbefall wird sofort auf die sexuelle Variante gewechselt. Männliche Lebewesen dienen also nur als eine Art biologische „Krankenversicherung“ (Miersch, 1996). Zur Unterstützung dieser These gibt es auch eine Rechnersimulation (Koch, 1994). Bei Pantoffeltierchen geht der Austausch genetischen Materials nicht einmal mit einer Vermehrung einher (Wickler und Seibt, 1983; Wickler und Seibt, 1998). Mehr als zwei Eltern pro Nachkommen gibt es in der Natur nur bei Viren und Bakterien, wobei letzteren sogar ein horizontaler Genaustausch zwischen verschiedenen Arten möglich ist (Stockinger, 1999).

Jenseits einer gewissen Komplexität kann eine Art auf den Mechanismus der Rekombination gar nicht mehr verzichten: Wird die Zahl der aktiven Gene, multipliziert mit der Mutationsrate, erheblich größer als die Ploidiezahl (Grad der Redundanz), kann keine noch so harte Selektion die Fehlerausbreitung mehr verhindern, d. h. das Überleben der Art sichern. Dieser Gefahr begegnet die Natur durch folgende Maßnahmen bzw. Kombinationen davon (Atmar, 1994):

- Begrenzung der Komplexität,
- Senkung der effektiven Mutationsrate, z. B. durch Reparaturenzyme,
- Erhöhung der Redundanz (Polyploidie).

Die ersten Verständnisschwierigkeiten setzen allerdings schon auf der molekularen Ebene ein: Von den etwa 100 000 in Genen kodierten Eiweißmolekülen ist erst eine geringe Anzahl bekannt. Dies allein wäre ein rein analytisches Problem. Aber es gibt Wechselwirkungen zwischen den kodierten Proteinen, zwischen der DNS und den kodierten Eiweißen, und auch noch zwischen DNS und DNS. Vielleicht verbirgt sich aber gerade hinter diesem „Wechselwirkungschaos“ der große Unterschied zwischen Primaten und Menschen — bei 98% Übereinstimmung der Genome. Deutlich größer fällt auch unser genetischer Abstand zu Maus, Fruchtfliege oder auch nur zur Bierhefe nicht aus.

Eine Erklärung, wie sich z. B. bei genetisch identischer DNS–Ausstattung 254 verschiedene Zelltypen beim Menschen bilden (Epigenese), gibt es erst in Ansätzen. Zur Zellalterung gibt es zwei Theorien, die sich jeweils auch experimentell belegen lassen:

- Verschleißtheorie: Mitochondrien sondern freie Radikale ab, diese schädigen die Erbsubstanz und greifen die Zellwände an, bis die Zelle stirbt.
- Programmtheorie: Die geschützten Bereiche an den Chromosomenenden (Telomere) werden bei jeder Zellteilung kürzer, bis sie aufgebraucht sind und die Zelle stirbt.

Das Aussehen der Nachkommen wurde lange allein mit Dominanz und Rezessivität der von den Eltern geerbten Gene erklärt. Der Evolutionsbiologe Haig leitet jedoch u. a. aus Experimenten mit Mäusen die Hypothese einiger rein männlich bzw. weiblich geprägter Gene („Genomic Imprinting“) her (Anonymus, 1998): Bei allen Menschen ist z. B. auf

Chromosom 15 ein mütterliches Gen abgeschaltet. Fällt dann das väterliche aus, entwickelt sich eine erbliche Form von Übergewicht, das sogenannte Prader–Willi–Syndrom. Klont man Lebewesen aus erwachsenen Körperzellen, macht sich das Fehlen dieser genetischen „Flaggen“, die beim entwickelten Organismus wieder eingeholt werden, negativ bei der embryonalen Entwicklung des Klons bemerkbar (Berndt, 1999).

Soviel zu den weitgehend unbestrittenen molekularen Grundlagen. Für weitere Details sei auf die originelle Darstellung in (Gonick und Wheels, 1986) verwiesen. Geht es um darauf aufsetzende Theorien der Evolution, sind wir offensichtlich mit einer Erklärung der Vielfalt des Lebens vollends überfordert.

2.2 Evolutionstheorien

Bei der wichtigen Frage, wie stark das Erbbild (Genom) eines Lebewesens dessen Erscheinungsbild (*Phänotyp*) determiniert, gibt es erst recht kein einheitliches Bild, weil die Antwort von der Art der Fragestellung abzuhängen scheint. In den zwanziger Jahren lieferte vor allem in den USA der *Behaviorismus* die theoretische Grundlage: Allein durch Belohnung und Bestrafung sollte (menschliches) Verhalten fast beliebig beeinflußbar sein. Etwa zur gleichen Zeit beobachtete Konrad Lorenz an Graugänseküken die Möglichkeit der Prägung — ein Indiz dafür, daß komplexe Verhaltensmuster genetisch determiniert sein können. Folglich schlug das Pendel zur anderen Seite aus: Die Suche nach Intelligenz– oder Gewaltgenen wurde hoffähig.

Bei einigen Erbkrankheiten kann man eindeutig sagen: Nur bei Auftreten von Allel X an der Stelle Y tritt Krankheit Z auf. Diese $1:1$–Entsprechung ist aber eher die Ausnahme: Ein Gen kann mehrere Merkmale beeinflussen (*Pleiotropie*), mehrere Gene können aber auch gemeinsam ein Merkmal hervorrufen (*Polygenie*). Veränderungen des Phänotyps lassen sich auch z. B. durch Nahrung, Klima oder Licht hervorrufen (adaptive Modifikabilität). Daß aber bei genetisch identischen und unter gleichen Laborbedingungen aufgewachsenen Versuchstieren (entstanden durch Inzucht oder Klonierung) eine Varianz phänotypischer Merkmale (Gewicht, Größe, Immunsystemstatus, ...) fast wie beim entsprechenden Wildtyp festgestellt wird (Gärtner, Bube, Flamme, Peters und Pfaff, 1976), ist überraschend. Dies schränkt — nebenbei bemerkt — die Aussagekraft von Tierversuchen, wenn sie artübergreifend interpretiert werden, zusätzlich ein. Diese Varianz scheint (onto–)genetisch bedingt zu sein und sorgt für eine Unschärfe des Selektionszugriffes auf den Genotyp und damit für eine zusätzliche, nicht beeinflußbare Diversität bei gleicher Populationsgröße. In der Terminologie der Evolutionsstrategien (s. Kapitel 4.5) könnte man auch von versteckten Strategieparametern sprechen, die die Kausalität zwischen Geno– und Phänotyp abschwächen. Auf der anderen Seite macht diese Erfindung der Natur die vollkommene Selektion gegen ein rezessives Merkmal praktisch unmöglich: Die Herabsetzung der Wahrscheinlichkeit für eine Erbkrankheit von $q_0 = 10^{-2}$ auf $q_n = 10^{-3}$ dauerte $n = 900$ Generationen.

Die *Selektion* sorgt auf verschiedenen Ebenen dafür, daß besser an ihre Umwelt angepaßte Individuen mehr Nachkommen erzeugen, die dann die günstigen Eigenschaften weitertragen. Selektion ohne eine Variantenquelle oder in einer konstanten Umwelt wirkt stabilisierend, d. h. die Variabilität wird mangels Notwendigkeit reduziert. Eine dann eintretende Umweltänderung kann das Aus für die Art bedeuten, sofern sich nicht in den ggf. vorhandenen rezessiven Allelen noch genug Varianten befinden. Für Darwin war der Nachkommenüberschuß und die daraus resultierende Konkurrenz der Evolutionsmotor, während heute auch der Mangel an Ressourcen als Schrittmacher angesehen wird, sofern die Umweltbedingungen lange genug konstant bleiben, um Spezialistentum in einer Nische zu honorieren (Reichholf, 1992).

So „lästig" die oben erwähnte Restvarianz für die Arzneimittelforschung ist, so beruhigend ist doch auf der anderen Seite, daß wir der Natur nicht beliebig ins Handwerk pfuschen können. Denn der Machbarkeitswahn läßt eine nüchterne Betrachtung der tatsächlichen Möglichkeiten der Genom–Analyse häufig auf der Strecke bleiben. Quantität schlägt eben nicht automatisch in Qualität um, sondern erschlägt zuweilen die Forscherinnen und Forscher: Zu Beginn des Genom–Projektes galt die Erbkrankheit Cystische Fibrose (auch: Mukoviszidose) als erster Kandidat für die Gentherapie, da nur ein einziges Gen auf Chromosom 7 beteiligt ist. Acht Jahre später gibt es immer noch keine Therapie, aber irritierende Erkenntnisse: Inzwischen sind 700 Mutationen dieses Gens bekannt, von denen jede zum Ausbruch der Krankheit, einer Verschleimung der Lunge, führt. Menschen, die dieselbe Mutation tragen, sterben aber manchmal bereits als Kind oder genießen bei leicht verschleimter Lunge etwa dieselbe Lebenserwartung wie Menschen ohne diese Krankheit. Von Hunderten weiterer bekannter Erbkrankheiten läßt sich erst genau eine durch die Gentherapie beseitigen.

Nach einem Dutzend Einzellern ist in 8 Jahren Arbeit das Erbgut des ersten Vielzellers entschlüsselt worden: Der etwa 1 cm lange Fadenwurm verfügt über 19 100 Gene, von denen bei 12 000 die Funktion noch unbekannt ist. Folgendes Szenario ist also gar nicht so unwahrscheinlich: Nach Beendigung des Genom–Projektes im Jahr 2003 gibt es einen Berg neuer Informationen und einen ebenso hohen Berg an neuen Rätseln. Das komplexe Wechselspiel zwischen Organismus und Umwelt wird nicht einmal ansatzweise bekannt sein.

Einige Biologen stellen die Evolution als Baum dar, andere ziehen ein eher buschartiges Gewächs als Veranschaulichung vor. Strittig ist sogar, ob die Evolution überhaupt Fortschritt hervorgebracht hat oder „nur" Vielfalt: Richard Dawkins glaubt, immer besser angepaßte Lebewesen zu beobachten, während Stephen Jay Gould den reinen Zufall ohne irgendeine Art von Fortschritt walten sieht (Klein, 1998). Alle Theorien zur Entstehung des heutigen Lebens müssen sich auf Fossilienfunde stützen und sind wegen der Größe der Meßintervalle mit Vorsicht zu genießen. Vielleicht ist die plötzliche Artenvielfalt im Kambrium (vor 570 bis 510 Millionen Jahren) auch nur eine Täuschung: Nur Panzer und Hartschalen werden im Gestein konserviert. Was, wenn es präkambrisch nur skelettlose Wesen gab?

Auf der Verhaltensebene nimmt mit der Vielfalt der Beobachtungen die der Theorien noch weiter zu. Je nachdem, wohin man blickt, sieht man zwischen Kooperation bis hin zur Selbstaufgabe und Tötung von Artgenossen inklusive Kannibalismus alle Varianten. Bleiben wie z. B. bei den Präriehunden die Weibchen über Jahre in einer Kolonie, wogegen die Männchen migrieren, dann ist es mit einem „Egoismus der Gene“ naher Verwandter erklärbar, daß die Weibchen trotz des erhöhten Risikos, dadurch die Aufmerksamkeit des Jägers auf sich zu ziehen, viel häufiger vor Feinden warnen als die zugewanderten Männchen. Dawkins hat darauf hingewiesen, daß jene Eigenschaften, die ein Gen kodiert, wesentlich langlebiger sind als die eines Individuums, einer Gruppe oder einer Population. Daraus leitet er das Gen als Angriffspunkt der Selektion ab (Dawkins, 1976).

Derselbe Autor hat den Begriff „Meme“ geprägt, um die Fortpflanzung von Ideen und Verhaltensmustern zu beschreiben. In gewisser Weise stellt dies einen Rückgriff auf den schon lange überholten Lamarckismus dar, der noch vor Darwin von einer Vererbungsmöglichkeit erworbener Eigenschaften ausging. Von Dawkins ursprünglich wohl nur als Anregung gemeint, gibt es heute eine ganze Theorie, die den Menschen auf eine „Mem–Maschine“ reduziert und praktisch alle Bereiche des täglichen Lebens umfaßt. Das menschliche Gehirn hat sich nur deshalb so weit entwickelt, um Meme schneller bilden und verbreiten zu können. Medien sind folglich nur Mem–Vehikel, eigenständig handelnde Subjekte nur eine Illusion (Blackmore, 1999).

Mayr sieht dagegen die Diversitätserhaltung einer Art als Triebfeder der Evolution[1]. Andere (Sozio–)Biologen ziehen spieltheoretische Verhaltenserklärungen aus der Ökonomie heran (Zell, 1990). Sie betrachten dabei allerdings nicht tierisches Verhalten durch eine menschliche Brille, sondern suchen nach formalen Verhaltensmodellen, die z. B. die Entstehung von Kooperation unter eigennützigen Individuen erklären sollen. Und stärker physikalisch orientierte Evolutionstheoretiker schlagen „co–evolutionäre Lawinen“ als Modell vor: Über lange Zeiträume hinweg findet wenig bis keine Entwicklung innerhalb einer Art statt, aber wenn dann etwas passiert, werden auch die mit ihr interagierenden Arten angestoßen (Kauffman, 1990; Bak, Flyvbjerg und Sneppen, 1994). Dieses Modell erlaubt zwar eine einheitliche Beschreibung so unterschiedlicher Phänomene wie (Co–)Evolution, Erdbeben, Blitzen, Lawinen und Frequenzschwankungen bei Pulsaren, dürfte aber nur sehr schwer paläontologisch zu belegen sein.

[1]Komplexität muß von Diversität unterschieden werden: Komplexität setzt Diversität voraus und beinhaltet eine Interaktion zwischen den variierten Elementen.

Für eine gute Darstellung sei auf (Jacob, 1997) verwiesen. Am Ende des Buches findet sich ein Glossar, das die wesentlichen Begriffe kurz erklärt.

2.3 Metaevolution, Quasiorganismen, ...

Da die Menschen sich durch ihre Technik dem Selektionszugriff der Natur weitgehend entzogen haben, wurde der Begriff der Metaevolution geprägt, der alle technischen, sprachlichen und kulturellen Weiterentwicklungen zusammenfaßt. Die Terminologie der Evolution wird auch auf noch entlegenere Gebiete übertragen: Multinationale Konzerne weisen einige Ähnlichkeiten mit Bioorganismen auf: Sie benötigen Ressourcen aus ihrer Umwelt, stoßen Stoffwechselprodukte aus, konkurrieren mit ähnlichen Organismen; sie können kooperieren, wachsen, sind vermehrungsfähig und verändern ihre Umwelt. Allerdings sind diese Quasiorganismen sehr viel wandlungsfähiger, weil Neuerungen keiner genetischen Verankerung bedürfen. Und: Sie sind potentiell unsterblich. Da sie nach dem Prinzip der opportunistischen Fitneßmaximierung funktionieren (müssen), ist dieses Phänomen aus globaler Sicht nicht nur beruhigend.

Die fundamentalen Erkenntnisse der Physik dieses Jahrhunderts laufen darauf hinaus, daß es nicht *das* Bild der Welt gibt. Einander ausschließende Bilder können dennoch (oder gerade deshalb?) einzelne Phänomene veranschaulichen wie z. B. beim Welle–Teilchen–Dualismus des Lichts. Manchmal scheint es, als träfe die Heisenbergsche Unbestimmtheitsrelation auch auf die Biologie zu: Je genauer man hinschaut, desto mehr verschwimmt der beobachtete Gegenstand. Das reine Sammeln weiterer Daten müßte ergänzt werden durch eine ähnlich bahnbrechende Theorie wie sie die Raum und Zeit verbindende Relativitätstheorie in der Physik darstellt.

Eine geschlossene Theorie wird die Biologie in absehbarer Zeit wohl nicht liefern, vielleicht auch nie bieten können: Denn bei der Untersuchung genetischer Phänomene bewegt man sich im Mikrokosmos; bemüht man sich, den Fortpflanzungserfolg und das Verhalten unterschiedlicher Lebewesen und Arten zu verstehen, dagegen im Makrokosmos. Bei der Spezialisierung in der heutigen Wissenschaft stellt dies einen fast unüberbrückbaren Gegensatz dar. Ein weiteres Problem: Um überhaupt zu Aussagen zu kommen, müssen einzelne Komponenten (Gene, Neuronen, ...) oder Prozesse (Rekombination, Mutation, ...) isoliert betrachtet werden, obwohl sie in der Natur nicht isoliert auftreten. „Eine systematische Darstellung reißt Ideen aus dem Boden, der sie wachsen ließ, und ordnet sie einem künstlichen Schema ein." (Paul Feyerabend)

In gewisser Weise erinnert das an die Probleme, beginnend mit z. B. einem Rekombinationsoperator ein theoretisches Verständnis evolutionärer Algorithmen zu entwickeln (s. Kapitel 4.5.1). Im Falle der Biologie kommt als Hemmschuh die schleichende Mutation von einer analytischen hin zu einer synthetischen Wissenschaft dazu, die Leben verändern und neue Organismen hervorbringen kann: Da derselbe genetische Code allem Leben auf diesem Planeten zugrunde liegt, lassen sich isolierte Gene artübergreifend verpflanzen, so daß z. B. sich schnell vermehrende Organismen wie Bakterien menschliche Proteine produzieren können.

Kapitel 3

Optimierung — Traditionelle Algorithmen

The army is now fully prepared to fight the previous war.

Bevor in Kapitel 4 die in dieser Arbeit zentralen naturanalogen Optimierstrategien vorgestellt werden, soll kurz definiert[1] werden, worum es bei der Optimierung überhaupt geht, welche Methoden bereits existieren und wo deren Schwachstellen liegen.

3.1 Problemformulierung

Die Suche nach der bestmöglichen Lösung eines Problems wird als Optimierung, das Ergebnis der Suche als Optimum bezeichnet. Oft sind die Aufgaben so schwierig, daß man nicht sicher sein kann, wirklich die bestmögliche Lösung gefunden zu haben. Bei technischen Problemen reicht es daher oft schon aus, eine bessere als die bisher bekannte beste Lösung zu finden.

[1] Dieses Kapitel verzichtet auf eine längere, formal exakte Einführung der Begriffe, Sätze und Verfahren zugunsten einer eher intuitiven, da keine der Definitionen im Rest der Arbeit in ihrer exakten Form benötigt wird, jeder weitergehende Formalismus m. E. also eher Ballast denn hilfreich wäre. Die existierenden Sätze über notwendige und hinreichende Bedingungen für lokale und globale Optima sind für die in dieser Arbeit untersuchten Verfahren nicht relevant.

Mathematisch formuliert, stellt sich ein nicht–diskretes, n–dimensionales Parameteroptimierungsproblem wie folgt dar, wobei man sich wegen $\max f = -\min(-f)$ auf Minimierungsprobleme beschränken kann:

$$
\left.
\begin{array}{ll}
\text{minimiere} & f : X \subseteq \mathbb{R}^n \to \mathbb{R} \\
\text{ggf. mit den Restriktionen} & g_j \left\{ \begin{array}{c} \geq \\ = \end{array} \right\} 0, \; j = 1, \ldots, m \\
\text{gesucht:} & \vec{x}^* \in X \subseteq \mathbb{R}^n \\
\text{so daß gilt:} & \forall \vec{x} \in X \subseteq \mathbb{R}^n : f(\vec{x}^*) \leq f(\vec{x})
\end{array}
\right\}
\begin{array}{l}
\text{oft verkürzt auf:} \\
f(\vec{x}) \to \min \quad \text{ggf. mit} \\
g_j(\vec{x}) \left\{ \begin{array}{c} \geq \\ = \end{array} \right\} 0, \\
j = 1, \ldots, m
\end{array}
$$

Ein Vektor $\vec{x}$, bestehend aus n Objektvariablen x_i, heißt *zulässig*, falls er die Restriktionen erfüllt. Der Vektor $\vec{x}^*$ wird *globale Minimalstelle* (allgemein: *Optimalstelle*) genannt, $f(\vec{x}^*)$ *globales Minimum (Optimum)*. Falls $f(\vec{x}^*) \leq f(\vec{x})$ nur in einer ε–Umgebung um $\vec{x}^*$ gilt, liegt lediglich ein *lokales Minimum* vor. Jedes globale ist auch ein lokales Optimum. Eine Funktion heißt *unimodal*, falls sie genau ein lokales Extremum aufweist, das dann auch ein globales ist, andernfalls *multimodal*.

Die Analysis bietet Sätze über notwendige und hinreichende Bedingungen für lokale Optima, die aber ein– bzw. zweimalige stetige partielle Differenzierbarkeit voraussetzen. Bei praktischen Optimierproblemen kann man oft nicht einmal Stetigkeit garantieren, daher wird auf den analytischen Ansatz hier nicht weiter eingegangen.

3.2 Iterative Optimierverfahren ...

Unter dem Begriff *Optimierverfahren* faßt man jene Algorithmen zusammen, die in der Lage sind, ausgehend von einer Startkonfiguration einen besonders vorteilhaften Zustand — das globale Optimum — eines Systems aufzufinden. Diese Aufgabe ist im allgemeinen nicht in endlich vielen Schritten zu bewältigen, d. h. praktisch *unlösbar* (Kan und Timmer, 1989). Führt man eine iterative Suche auf einem Rechner durch, wird man sich aufgrund der begrenzten Genauigkeit dem Optimum nur nähern können, was aber für praktische Probleme i. d. R. keine Einschränkung darstellt.

Die nächstliegende Methode, den Suchraum mittels einer *Rastersuche* zu durchmustern und sich den besten aufgetretenen Wert zu merken, scheitert schon bei relativ kleinen Problemen am exponentiell wachsenden Suchaufwand, oder wie Bellman es formuliert, am „Fluch der Dimension“.

Daher versucht man, das allgemeine Problem auf Spezialfälle einzuschränken, für diese — falls möglich — eine Theorie zu entwickeln und daraus ein Optimierverfahren abzuleiten. So ist eine Fülle von mehr oder weniger spezialisierten Optimierverfahren entstanden, die gut funktionieren, falls das System alle Voraussetzungen erfüllt, die aber versagen können oder sich zumindest als ineffizient erweisen, falls erwartetes „internes Modell“ und System nicht übereinstimmen.

In der Literatur gibt es verschiedene Klassifizierungen, orientiert an Eigenschaften des Problems oder des Verfahrens. Diese baumartigen Aufteilungen werden aber schnell unübersichtlich, will man sie erschöpfend machen. Daher wird hier ein informeller Ansatz verfolgt.

Die für numerische Optimierverfahren übliche Iterationsvorschrift lautet:

$$\vec{x}^{(g+1)} := \vec{x}^{(g)} + s^{(g)} \cdot \vec{v}^{(g)}$$

Der neue Variablenvektor $\vec{x}^{(g+1)}$ ergibt sich also aus dem alten $\vec{x}^{(g)}$ durch Addition eines Schrittes der Länge $s^{(g)}$ in Richtung $\vec{v}^{(g)}$ (normiert auf die Länge 1). Die verschiedenen Optimierverfahren unterscheiden sich durch die Schrittweitensteuerung und die Wahl des Richtungsvektors. Im einfachsten Fall hält man s konstant und setzt für $\vec{v}$ zyklisch die n Einheitsvektoren ein. Allgemein erfordert die Bestimmung der Schrittweitenlänge in der aktuellen Suchrichtung ihrerseits die Lösung eines eindimensionalen Optimierproblems.

Optimierverfahren müssen sich im allgemeinen aufgrund fehlender Informationen über die inneren Zusammenhänge des zu optimierenden Systems („Black Box“) mit Heuristiken behelfen, d. h. sie nutzen extern vorgegebene Informationen über erfolgreiche Suchrichtungen, oder sie orientieren sich am Erfolg und Mißerfolg bereits durchgeführter Suchschritte, um weiter in Richtung Optimum voranzukommen.

Hill Climbing–Verfahren ersetzen einen willkürlichen Startpunkt solange durch einen besseren Nachbarn, bis dies nicht mehr möglich ist. Diese Verfahren können nur das dem Startpunkt nächstgelegene lokale Optimum finden.

Der bekannteste Vertreter mehrdimensionaler Optimierverfahren ist wohl die *Simplex–Methode* (inklusive ihrer Modifikationen), die für einen durch lineare Restriktionen eingeschränkten Suchraum und eine lineare (bzw. nichtlineare, aber konvexe) Zielfunktion in der Praxis relativ schnell und sicher (Zeitaufwand: $\sim n^2 \cdot m$ bzw. $\sim m^2 \cdot n$, je nachdem, ob das primale oder das duale Problem gelöst wird) arbeitet (Dantzig, 1966). Im schlimmsten Fall muß man aber mit $(m+n)!\ /\ (n! \cdot m!)$ Schritten rechnen. Mathematisch gesprochen, optimiert man eine höchstens quadratische (und damit konvexe) Zielfunktion über einem ebenfalls konvexen Suchraum. Diese strengen Voraussetzungen stellen sicher, daß ein gefundenes lokales immer auch das globale Optimum ist; Zielfunktionen in dieser extrem eingeschränkten Welt verhalten sich „gutmütig“.

Vertreter aus der Klasse der *Gradientenverfahren*, zu der auch das Simplex–Verfahren gehört, sind an diese lineare Welt zwar nicht mehr gebunden, doch sie „glätten“ die rauhe Welt durch eine andere mathematische Voraussetzung: Benötigt werden vor dem Start des Verfahrens explizit die ersten partiellen Ableitungen der Zielfunktion, also Informationen über lokale Steigungen. Diese Strategien legen also ein lokal lineares internes Modell des

Systems zugrunde. Die Suchrichtung orientiert sich am steilsten An– bzw. Abstieg der Funktion.

Das *Newton–Verfahren* benötigt zusätzlich sogar die zweiten partiellen Ableitungen in Form der Hesse–Matrix, also Informationen über lokale Krümmungen, verfügt aber damit über ein internes quadratisches Modell seiner Umwelt. Die *Quasi–Newton–Verfahren* beschaffen sich die nötigen Informationen approximativ im Laufe der Suche selbst, ebenso die Verfahren der *konjugierten Gradienten* und der *variablen Metrik*.

In die Welt der bisher vorgestellten Verfahren passen keine Verschlechterungen, so daß die Suche nach dem Optimum abgebrochen werden kann, falls antizipierte Verbesserungen ausbleiben. Bei einer unimodalen Zielfunktionstopologie führt dies auch zum gewünschten Ergebnis, bei einer multimodalen Topologie wird man aber nur das dem Startpunkt nächstgelegene lokale Optimum finden und dort „steckenbleiben". Das *Newton–Raphson–Verfahren* kann sogar divergieren, falls eine Diskrepanz zwischen internem Modell und der Zielfunktion auftritt.

Die bisher vorgestellten deterministischen Strategien sind natürlich anderen Verfahren überlegen, wenn die Aufgabenstellung in das Schema ihrer Modellannahmen paßt. So sind Newton–Raphson–Verfahren in der Lage, das Optimum einer quadratischen Zielfunktion innerhalb einer Iteration zu finden, wenn auch dieser eine Schritt eine Matrixinversion enthält und somit aufwendig ist. Für den allgemeinen Fall sind einige unkonventionelle Strategien entworfen worden, darunter auch stochastische.

Polyeder–Verfahren (*Simplex–* oder *Complex–Strategie*) für die nichtlineare Optimierung haben sich in einem umfangreichen Vergleichstest als robust in schwierigen Situationen erwiesen, ebenso wie das *Pattern Search–Verfahren* oder die Methode der *rotierenden Koordinaten* (Schwefel, 1977; Schwefel, 1981). Interessanterweise suchen viele Strategien in „pathologischen" Situationen Zuflucht beim Zufall, d. h. wenn ihr Vorrat an deterministischen Regeln kein Weiterkommen mehr ermöglicht.

Speziell für multimodale Probleme gibt es *Clustering–Methoden*, die wiederholt eine endliche Anzahl von Testpunkten im Suchraum verteilen, diese zu Gruppen um lokale Optima zusammenfassen und ggf. lokale Suchen ausführen. Cluster nennt man die Attraktionsgebiete der lokalen Minima. Angefangen mit der Anzahl der Testpunkte, der Art und Weise der Clusterbildung, der Generation der Punkte für die nächste Iteration bis zu den Abbruchkriterien gibt es eine Fülle notwendiger Einstellungen, die den Einsatz dieser Verfahren erschweren.

Für weitere Verfahrensdetails sei auf die Literatur zum Thema Optimierung verwiesen, z. B. (Hoffmann und Hofmann, 1971; Schwefel, 1977; Schwefel, 1995), für Verfahrensklassifikationen z. B. auf (Törn und Žilinskas, 1989; Rudolph, 1990), für Konvergenzmaße z. B.

auf (Dennis und Schnabel, 1989), für die Behandlung von Restriktionen z. B. auf (Bazaraa und Shetty, 1979).

3.3 ... und ihre Unzulänglichkeiten

Bei technischen Optimierproblemen wird man relativ selten das System mittels einer geschlossenen, analytischen Zielfunktion der Form $\vec{x} \rightarrow f(\vec{x})$ beschreiben können. Oft benötigt man ein *Simulationsmodell*, um die Wirklichkeit hinreichend genau zu erfassen. Und wenn ein solches Modell einmal erstellt und validiert ist, liegt der Wunsch nahe, über die reine Simulation hinaus nach „besseren" Parametrisierungen zu suchen.

Dies kann aber mit Hilfe der bisher beschriebenen konventionellen Optimierverfahren nur dann geschehen, falls man sich bereits in der Modellierungsphase an den Erfordernissen eines oder mehrerer der bisher vorgestellten Verfahren orientiert hat. Bei einer komplexen Realität wird man in einen Zwiespalt zwischen der Urbildtreue des Modells und den notwendigen Voraussetzungen für eine spätere Optimierung geraten. Die Kategorien linear, quadratisch und konvex für die Zielfunktion werden oft der Realität nicht gerecht, so z. B. bei Klima– oder Verbrennungsprozessen sowie Roboter– oder Fahrzeugsteuerungen. Die entsprechenden Modelle werden in der Regel nicht einmal die elementare Voraussetzung der Stetigkeit — kleine Änderungen an den Eingängen bewirken auch nur kleine Änderungen der Ausgänge — erfüllen, von einer stetigen Differenzierbarkeit ganz zu schweigen. In folgenden Fällen muß man mit dem Versagen der klassischen Optimierverfahren rechnen:

- bei nicht–metrischer Bewertungsfunktion — im Extremfall kann lediglich „besser" bzw. „schlechter" als qualitative Antwort zurückkommen
- bei stochastisch gestörten Modellgrößen, die in der experimentellen Optimierung den Normalfall darstellen
- bei nichtlinear–dynamischen Modellen, die sich scheinbar stochastisch verändernde Modellgrößen, hervorbringen können
- bei „schwierigen" Computermodellen, die nicht stetig und damit auch nicht differenzierbar sind
- bei über der Zeit wandernden Optima, wie sie in Prozeßsteuerungs– und –kontrollaufgaben auftreten
- bei Problemen aus dem Bereich der diskreten Optimierung. Bei praktisch relevanten diskreten Entscheidungsproblemen (z. B. Maschinenbelegungsprobleme, Routingprobleme in Netzen, Packungsprobleme, Anlagenauslegungsprobleme, ...) ist man an der Realisierung der optimalen Lösung aus Kostengründen interessiert. Um sie zu finden, benötigen deterministische Algorithmen eine exponentiell von der Anzahl

der Entscheidungsvariablen abhängige Rechenzeit, d. h. diese Entscheidungsprobleme liegen in der Komplexitätsklasse NP.
Im allgemeinen Fall hat man nicht mehr Teilmengen des $\mathbb{R}^n$ als Definitionsbereich von f, sondern einen inhomogenen Suchraum: $f : M \subseteq M_1 \times \ldots \times M_n \to Q$. Entsprechend sucht man $\vec{x}^*$ dann in M.

Damit man sich nicht schon in der Modellierungsphase der normativen Kraft des leicht Optimierbaren unterwerfen und somit auf Realitätsnähe verzichten muß, wurden Optimierverfahren entwickelt, die universell einsetzbar, also „anspruchslos“ sind, diese Eigenschaft aber mit einem Verzicht auf die Garantie des Auffindens eines globalen Optimums und einer langsameren Konvergenzgeschwindigkeit — verglichen mit jeweils für den Spezialfall entwickelten Verfahren — bezahlen. Diese Verfahren stellen einen Kompromiß dar zwischen einer Volumenorientierung wie bei der Rastersuche und einer Pfadorientierung wie z. B. bei Gradientenstrategien. Diese „Generalisten“ werden im folgenden Kapitel vorgestellt.

Kapitel 4

Optimierung — Evolutionäre Algorithmen (EA)

New systems mean new problems.

Metaphern — in diesem Kapitel biologische — können Ideen sehr schnell veranschaulichen oder in die Irre führen. Sie können aber niemals eine vollständige oder gar präzise Abbildung leisten. Entsprechend kann die Überlagerung von Ähnlichkeit und Verschiedenheit sowohl kreative als auch schädliche Auswirkungen haben: Das Bild vom Flügelschlag eines Insekts, der sich weit entfernt zu einem Sturm auswächst (*Schmetterlingseffekt*), hat sicher eine erhöhte Sensibilisierung für die mehrheitlich nichtlinearen Zusammenhänge unseres Alltags einerseits und eine intellektuelle Bescheidenheit bei der Folgenabschätzung unseres Handelns andererseits erreicht, führt aber vollständig in die Irre, wenn daraus Befürchtungen erwachsen, jeder Sprühstoß aus einer Dose könne das Weltklima kippen lassen.

Leider steht aufgrund einiger bizarr anmutender Kontroversen zwischen den historisch gewachsenen „Schulen" evolutionärer Algorithmen nicht immer der Ideenaustausch und die Erschließung weiterer Problemdomänen im Mittelpunkt des Interesses[1]. Entweder führen die unterschiedlichen Modellierungsebenen der einzelnen Ansätze zu Mißverständnissen, oder es wird die Komplexität des natürlichen Vorbildes ignoriert, das — wie in Kapitel 2 beschrieben — eine Vielzahl z. T. widersprüchlicher Mechanismen bereithält und damit fast jeden Modellierungsansatz rechtfertigt.

In den letzten Jahren haben Algorithmen, die evolutionäre Prinzipien imitieren, immer mehr Aufmerksamkeit auf sich gezogen. Dies hat einerseits damit zu tun, daß die enorm

[1] Eine Initiative aus dem Frühjahr 1999 bemüht sich z. B. um eine Namensänderung des Gebietes vom etablierten „Evolutionary Computation" in „Genetic and Evolutionary Computation"

gewachsene Rechnerleistung sehr viel aufwendigere Simulationen erlaubt, andererseits aber auch mit der Erkenntnis, daß reale Optimierprobleme sich eben nicht immer auf geschlossen darstellbare und dazu noch lineare oder stetig differenzierbare Probleme abbilden lassen, so daß die in Kapitel 3 erwähnten Verfahren benutzt werden können. Und wenn man im Sinne einer urbildtreueren Modellierung Nichtlinearitäten und Unstetigkeiten der Zielfunktion und ggf. der Restriktionen zuläßt, warum sollte man sich dann nicht auch zur Optimierung jener Strategien bedienen, die die Natur zur Beherrschung dieser Probleme erfunden und offensichtlich erfolgreich verwendet hat?

4.1 Basisprinzipien Evolutionärer Algorithmen

Ziel dieses Kapitels ist es, kurz die wesentlichen Untergliederungen jenes Gebietes zu erwähnen, das heute unter dem Namen *Evolutionäre Algorithmen (EA)* firmiert. Auf frühe Algorithmen, die mit der heutigen Terminologie auch als evolutionäre dargestellt werden können, wie z. B. jene von Friedberg (Friedberg, 1958) und Bremermann (Bremermann, 1962), wird in dieser Arbeit nicht weiter eingegangen.

Auch andere, ebenfalls naturanaloge, aber eben nicht evolutionäre Verfahren wie *Zelluläre Automaten*, *Fuzzy Logik*, *Neuronale Netze*, *Classifier–Systeme* oder *Immunsystemmodelle* werden nicht näher vorgestellt. Ebenso wird nicht genauer auf das *Simulated Annealing* eingegangen, obwohl die dort wirkende Boltzmann–Thermodynamik nichts anderes realisiert als die Selektion auf einer Population von Lösungsvarianten. Da diese Varianten aber über keinen Genotyp verfügen, kann keine Anpassung auf einer anderen Ebene als der phänotypischen stattfinden. Notwendig ist daher ein problemspezifisches, extern vorgegebenes Abkühlschema, dessen Güte allein über den Erfolg des Verfahrens entscheidet. Das Gebiet des *Artificial Life* verwendet zwar evolutionäre Regeln, hat aber m. E. keinerlei Ambitionen in Richtung Problemlösung. Daher wird es im weiteren ebenso außer acht gelassen.

Hat natürliche Evolution überhaupt etwas mit Optimierung zu tun, geht es nicht in erster Linie um Anpassung? Auf einer sehr abstrakten Ebene kann man vielleicht Einigung darüber erzielen, daß der „Gesamtfehler“ einer Art minimiert wird. Engt man die Sichtweise etwas weiter ein, sieht man durchaus *auch* einige Resultate eines echten Optimierprozesses: Bleiben die Rahmenbedingungen lange genug stabil, werden auch komplexe (Teil–)Probleme nachweisbar optimal durch die Natur „gelöst“ (Rosen, 1967): Die Durchmesserverhältnisse an den Verzweigungsstellen der Blutgefäßsysteme von Säugetieren liegen sehr nahe am theoretischen, durch die Naturgesetze vorgegebenen Optimum von $2^{-1/3}$. Systematisch befaßt sich die *Bionik* (s. Kapitel 7.2) mit der Nutzung natürlicher „Erfindungen“.

Die Idee, Prinzipien der biologischen Evolution unter Optimierungsgesichtspunkten zu interpretieren, wurde — trotz der auch technisch bemerkenswerten Erfindungen der Natur — lange Zeit belächelt oder sogar rigoros abgelehnt, weil

- gemessen an der Lebenszeit eines Menschen die Evolution sehr langsam und nicht direkt beobachtbar verläuft,
- keine globale „Zielfunktion" identifizierbar ist,
- ein auf stochastisch beschriebenen Störungen basierendes Mutations–Selektions–Verfahren aufwendig und verschwenderisch gegenüber den für einen Spezialfall entwickelten Strategien wirkt,
- bei ersten Simulationsversuchen der Slogan „Survival of the Fittest" (Singular!) zu wörtlich genommen wurde oder eine evolutionäre Suchstrategie auf ein Problem der Linearen Programmierung angesetzt wurde und versagen oder zumindest schlecht abschneiden mußte (Bremermann, 1962),
- das der Evolution inhärente Zufallselement oft falsch interpretiert wird, denn einfaches „Würfeln" kann die heutige Vielfalt an komplexen Lebewesen nicht hervorgebracht haben: Menschliche Keimzellen beherbergen etwa 50 000 Gene, von denen jedes im Mittel aus 300 Nukleotidbasentripletts besteht. Unter Berücksichtigung der Redundanz des genetischen Codes, der statt der 64 möglichen nur 20 verschiedene Aminosäuren kodiert, verbleiben noch $20^{300 \cdot 50\,000} \approx 10^{19\,500\,000}$ mögliche Genotyp–Kombinationen. Geht man von einer konstanten Population von 5 Milliarden Menschen aus, so wären $10^{19\,499\,990}$ Generationen nötig, um jede Möglichkeit einmal auszuprobieren. Zur Orientierung: Angenommen, jedes der etwa 10^{80} Elementarteilchen im Universum wäre in der Lage gewesen, seit dem Urknall pro Elementarzeitschritt (Annahme: Picosekunden, $10^{-12}s$) einen solchen Genotyp zu erzeugen und zu bewerten, dann hätten erst 10^{120} Möglichkeiten durchgespielt werden können. Die Würfel–„Evolutionstheorie" kann also die Existenz des Menschen nicht erklären oder auch nur halbwegs wahrscheinlich erscheinen lassen.
  Entsprechend erfolglos muß das gleichverteilte Raten im Parameterraum einer Zielfunktion sein (s. Kapitel 3). Man kann zeigen, daß eine vollständige Enumeration durch Vermeidung von Wiederholungen besser abschneiden muß als eine Monte–Carlo–Suche.

Interpretiert man Evolution dagegen als einen kumulativen, hochgradig parallelen Siebprozeß, dessen Resultate leicht modifiziert in das nächste Sieb weitergegeben werden, erscheint die entstandene Vielfalt und Effizienz nicht mehr so wundersam. Bei der Nachahmung geht es nicht darum, diese Vielfalt durch ein Rechnermodell im Zeitraffer nachzuerfinden, sondern darum, jene Prinzipien und Mechanismen zu isolieren und auszunutzen, die auch ihrerseits einer Evolution unterlagen: So ist mit der Rekombination fast zwangsläufig ein Mechanismus erfunden worden, der es artverwandten Individuen erlaubt, Teile ihres genetischen Materials auszutauschen, um wechselnden Umweltbedingungen besser begegnen zu können.

Die beiden heute identifizierbaren Hauptrichtungen *Genetische Algorithmen (GA)* jenseits und *Evolutionsstrategien (ES)* diesseits des Atlantiks entstanden etwa gleichzeitig und unabhängig voneinander Mitte der 60er Jahre als globale Optimierverfahren. In allen hier vorgestellten Verfahren „lernt“ eine Population kollektiv durch Nachahmung folgender natürlicher Prinzipien:

- Trennung von Geno– und Phänotyp[2]
- Vererbung (Rekombination bzw. Crossover, nicht bei EP)
- Mutation
- Selektion
- Population (konstanter Größe)

Während man Mutation und Rekombination als Motor der Evolution im Rechner bezeichnen kann, sorgt die Selektion für dessen Steuerung. Sie kann zwar nur den Phänotyp — den äußerlich sichtbaren Anpassungsgrad — bewerten, damit indirekt aber auch den zugehörigen Genotyp. Aus der Auto–Metapher mag auch deutlich werden, wie wichtig eine ständige Quelle neuer Varianten ist, denn ein stehendes Fahrzeug benötigt keine Steuerung. Angepaßtheit meint einen immer wieder neu zu beurteilenden Zustand, da die selektierende Umwelt selbst auch Veränderungen unterworfen sein kann. Im Gegensatz zu den traditionellen Optimierverfahren können evolutionäre auch ein zeitabhängig wanderndes Optimum finden und ihm folgen.

Spricht man im Zusammenhang mit EA von Individuen, sind im biologischen Sinne haploide Einzeller gemeint. Übersetzt in eine Pseudo–Sprache, stellt sich der Grundalgorithmus wie folgt dar, wobei die genetischen Operatoren nicht alle bzw. nicht notwendigerweise in dieser Reihenfolge angewendet werden:

setze $t := 0$
initialisiere Population P_0
bewerte die Individuen mittels f
`while` Terminierungskriterium nicht erfüllt `do`
  rekombiniere Individuen aus P_t zu Nachkommen
  mutiere die Nachkommen

[2] Da es bei EA i. d. R. eine 1 : 1–Entsprechung zwischen Gen und Merkmal gibt, wirkt diese Trennung bei separablen Zielfunktionen eher künstlich. Vorteilhaft ist dabei, daß ein optimaler Zielfunktionswert gleichzeitig dafür sorgt, daß alle Objektvariablen des Individuums optimal sind. Bei Optimierproblemen mit Interdependenzen zwischen den Objektvariablen (*Epistasis*) bekommen Standard–EA Schwierigkeiten.

bewerte die Nachkommen mittels f
selektiere die Individuen für P_{t+1}
setze $t := t + 1$
`od`

Abbildung 4.1 veranschaulicht den Grundalgorithmus aller EA:

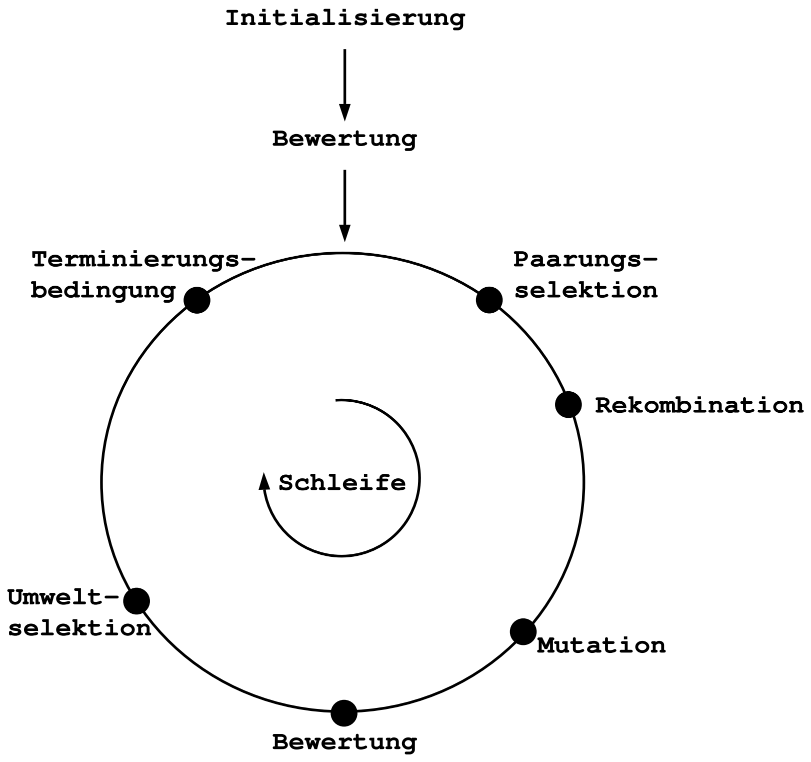


Abbildung 4.1: Iterationsschema eines Standard–EA, aus (Schwefel und Kursawe, 1998)

Hinter den einzelnen genetischen Operatoren können sich allerdings (fast) beliebig komplexe Berechnungen verbergen. Im weiteren wird nun auf die einzelnen Hauptrichtungen des Gebietes näher eingegangen. Den Stand der Dinge bezüglich aller Untergliederungen der Evolutionären Algorithmen faßt ein Handbuch zusammen (Bäck, Fogel und Michalewicz, 1997).

4.2 Genetische Algorithmen (GA)

Weltweit sind die auf Holland (Holland, 1975) zurückgehenden *Genetischen Algorithmen* am weitesten verbreitet. Wichtiger für ihre Verbreitung und das Zusammenwachsen mit den anderen Gebieten evolutionären Rechnens ist De Jong (De Jong, 1975). Obwohl Holland eher zu einer Theorie adaptiver und reproduktiver Systeme gelangen wollte, deckt sich der GA–Einsatzbereich heute prinzipiell mit dem der Evolutionsstrategien (s. Kapitel 4.5), und auch die grundlegenden Ideen entsprechen den in der Einleitung dieses Kapitels erwähnten natürlichen Prinzipien. Für eine Darstellung ähnlicher Ideen vor Holland sei auf (Bäck, 1996) verwiesen. Holland erwähnte auch schon die Möglichkeit, ganze Computerprogramme als Individuen zu repräsentieren und diese zu „züchten", um komplexeres Verhalten generieren zu können (s. Kapitel 4.3). Als erster Schritt in diese Richtung können die *Classifier Systems* angesehen werden, die mittels eines GA zu Regeln gelangen, deren Aktionsteil von einer definierten Umgebung bewertet wird.

Verwendet man l_x Bits zur Kodierung einer Objektvariablen, sieht der Algorithmus eines GA zur Behandlung eines n–dimensionalen Parameteroptimierungproblems wie folgt aus:

setze $t := 0$
initialisiere Population P_0 mit Individuen $\in \mathbb{B}^l$, $(l = l_x \cdot n)$
bewerte die Individuen mittels f
`while` Terminierungskriterium nicht erfüllt `do`
  selektiere Individuen zur Rekombination gemäß ihrem Anteil an der Gesamtfitneß
  rekombiniere mit Wahrscheinlichkeit $p_{\,Crossover}$
  mutiere mit Wahrscheinlichkeit $p_{\,Mutation}$
  bewerte die Individuen mittels f und ersetze ggf. alte durch neue Individuen
  setze $t := t + 1$
`od`

Heute werden GA mehrheitlich zur Optimierung eingesetzt, wenn auch selten in der gerade dargestellten Form. Ist durch das Problem ein komplizierterer als der binäre Suchraum vorgegeben (Permutationen, Matrizen, Graphen, ...), muß entweder eine Abbildung auf Bitstrings gefunden werden, oder man benutzt eine problemnähere Datenstruktur und paßt die genetischen Operatoren an. Michalewicz' Buch bietet viele Beispiele für dieses Vorgehen (Michalewicz, 1992; Michalewicz, 1994), das allerdings den Boden für „Mogelpackungen" bereitet: Versteckt man in den modifizierten genetischen Operatoren problemspezifische Heuristiken, löst nicht der GA das Problem, sondern die Heuristik.

Als theoretische Grundlage für das Funktionieren von GA wird immer noch das *Schema–Theorem* herangezogen (Holland, 1975; Goldberg, 1989). Es postuliert, daß gute Teil–Bitstrings (*Building Blocks*, genauer: kurze, an wenigen Positionen festgelegte, überdurchschnittlich gute Schemata) sich im Laufe der Generationen exponentiell vermehren. Das *Implicit Parallelism*–Postulat behauptet sogar, daß eine Population aus μ Individuen ohne zusätzliches „Gedächtnis" in der Größenordnung μ^3 Schemata verarbeitet (Goldberg, 1989). Für eine kritische Einschätzung sei auf Menke verwiesen (Menke, 1997).

Obwohl das Schema–Theorem eine Aussage über einen GA mit Binärkodierung, Mutation, 1–Punkt–Crossover und proportionaler Selektion macht, hat es nur wenig mit dem realen Verhalten des Algorithmus zu tun:

- In einer endlichen Population dominiert ein überdurchschnittlich gutes Schema sehr schnell die ganze Population.
- Bei unendlich großer Population wäre dagegen jeder (außer einem Sortier–) Algorithmus überflüssig, weil bereits die Startpopulation das gesuchte Optimum unendlich oft enthält.

Auch bei der Bewertung der Operatoren spielt das Schema–Theorem eine unglückliche Rolle. Der Mutation wird eine ausschließlich zerstörerische Wirkung unterstellt, bei der Rekombination kommt es darauf an, ob der Crossover–Punkt in einem Building Block liegt oder nicht. Viele Anwender glauben sich durch die theoretische Fundierung auf sicherem Terrain, wenn sie ihr Problem mit einem oben erwähnten GA zu lösen versuchen. Das eventuelle Scheitern muß aber nicht die prinzipielle Unfähigkeit des GA demonstrieren, sondern kann z. B. an der Wahl der proportionalen Selektionsvariante liegen. Nach Beyer bedeutet es, Ursache und Wirkung zu vertauschen, wenn man eine Hypothese über Beobachtungen an einem Optimierprozeß — das Schema–Theorem — als Erklärung für dessen Funktionieren benutzt (Beyer, 1997).

Ferner gibt es einen globalen Konvergenzbeweis für GA (Hartl, 1990), der im Prinzip folgende Aussage macht: Wenn jeder Punkt des Suchraumes (mittels Rekombination und Mutation) erreichbar bleibt und die Selektion keine Verschlechterungen akzeptiert, findet man das globale Optimum garantiert. Es gibt auch Varianten dieses Satzes mit leicht geänderten bzw. abgeschwächten Voraussetzungen. Alle diese Sätze sind wahr, elegant beweisbar und völlig nutzlos, weil praktisch einsetzbare Optimierverfahren in *endlicher* Zeit ein Ergebnis liefern, d. h. konvergieren müssen.

Trotz der gemeinsamen Wurzeln gibt es jedoch einige Unterschiede zu den Evolutionsstrategien, die im letzten Abschnitt dieses Kapitels genauer vorgestellt werden (Hoffmeister und Bäck, 1991; Bäck und Schwefel, 1993).

- Der größte Unterschied liegt in der Kodierung der zu optimierenden Parameter durch ein (in der Regel) binäres Alphabet. Da die genetische Information eines Individuums als Bitstring vorliegt, muß zur Berechnung der Fitneß also erst dekodiert werden. In diese Dekodierung ist auch die Behandlung von Restriktionen integriert, da diese im allgemeinen zunächst einen ganzzahligen Wert ergibt, der dann in einem zweiten Schritt in ein reellwertiges, notwendigerweise beschränktes Intervall abgebildet wird. Allgemeinere Nebenbedingungen werden z. B. über Straffunktionen realisiert, die bei Restriktionsverletzung(–en) die Fitneß eines Individuums vermindern.

- Die genetischen Operatoren Mutation und Rekombination bzw. Crossover haben eine andere Rollenverteilung als in Evolutionsstrategien: Das Crossover, das mit einer festen Wahrscheinlichkeit angewendet wird (typischerweise $p_{Crossover} \approx 0.6$), spielt die Hauptrolle bei der Erzeugung von Individuen, während Mutationen, die wie in der Natur an einzelnen Genen relativ selten auftreten (Wahrscheinlichkeit $p_{Mutation} \approx 0.001$), nur die Aufgabe haben, den Verlust eines Allels in der Population an einer Bitposition zu verhindern. Semantische Grenzen innerhalb des Bitstrings werden bei der Rekombination nicht berücksichtigt. Vergleicht man die Abbildungen 4.2 und 4.5 miteinander, sieht man, daß das Crossover im Gegensatz zur Rekombination in ES nicht unbedingt für eine Ähnlichkeit zwischen Eltern und Nachkommen sorgt:

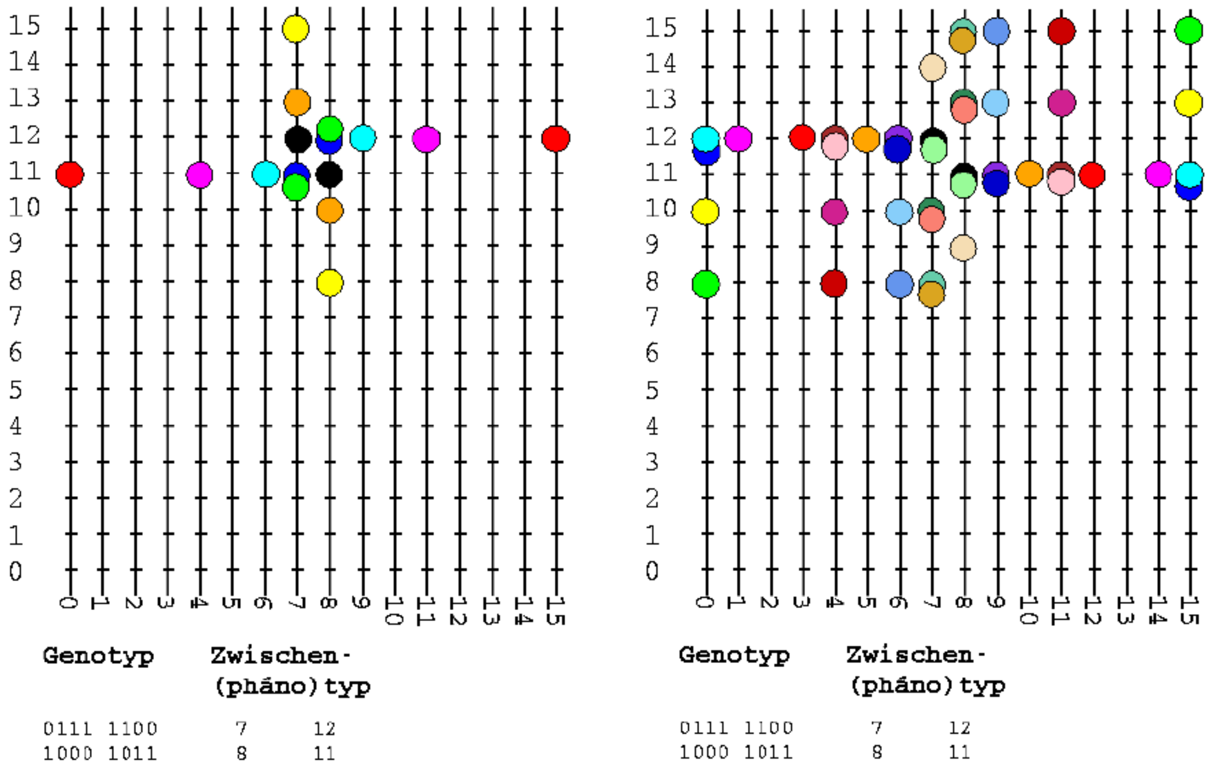


Abbildung 4.2: Mögliche Nachkommen bei 1– (links) und 2–Punkt–Crossover (rechts), $l_x = 4$, zwei Variablen, Eltern jeweils schwarz markiert

Beide Wahrscheinlichkeiten werden exogen vorgegeben und nicht im Laufe der Generationen adaptiert. Auch die Wahl des Gray–Code als Abbildung zwischen Bitstring und ganzer Zahl stellt nicht sicher, daß eine Bitmutation nur eine Veränderung der ganzen Zahl um ± 1 bewirken kann. Man kann nur umgekehrt garantieren, daß sich zur Veränderung der ganzen Zahl um ± 1 nur genau ein Bit ändern muß.

- Es werden keine überschüssigen Nachkommen wie bei den Evolutionsstrategien (und in der Natur) erzeugt.

- Ein weiterer wesentlicher Unterschied liegt im Selektionsmechanismus. Reproduktionswahrscheinlichkeiten sind dem Anteil des jeweiligen Individuums an der Gesamtfitneß der Population proportional, können also bei Vorliegen eines sehr guten Individuums zu dessen Dominanz in den Folgegenerationen führen. Ein solches Verhalten kann sogar schon bei einfachen, unimodalen Zielfunktionen auftreten und den weiteren Fortschritt der Suche hemmen. Diese Schwäche läßt sich aber durch andere Selektionsmechanismen teilweise beheben, wie z. B. durch das Ranking, bei dem einem Individuum in Abhängigkeit von seinem Rang in der Population, also unabhängig von seiner relativen Fitneß, eine Selektionswahrscheinlichkeit zugeteilt wird. Ein weiteres Konzept zur Diversitätserhaltung stellt das Sharing–Konzept dar, bei dem die Fitneß eines Individuums reduziert wird, falls sich weitere Individuen in seiner Nähe befinden. Dies kann dazu führen, daß eine Population um das globale Optimum herumwandert, anstatt es zu finden. Abbildung 4.3 zeigt schematisch die mögliche Dominanz eines Individuums bei der proportionalen Selektion, für eine umfassende Untersuchung aller gängigen GA–Selektionsmechanismen sei auf (Blickle und Thiele, 1995) verwiesen:

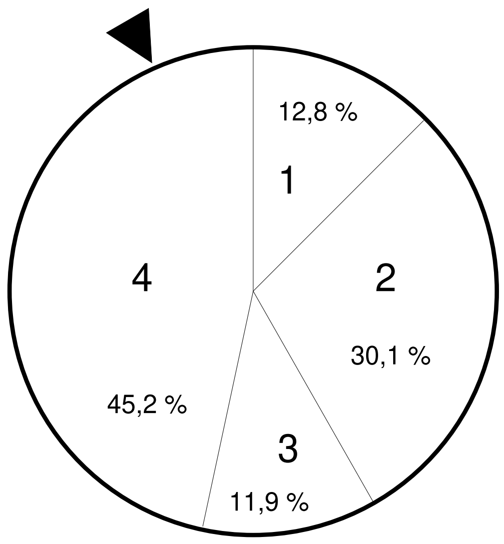


Abbildung 4.3: Proportionale Selektion, aus (Schwefel und Kursawe, 1998)

4.3 Genetic Programming (GP)

Als geistiger Vater des *Genetic Programming (GP)* wird häufig Koza genannt, obwohl die Idee, Computerprogramme automatisch zu „züchten“, sicher nicht von ihm stammt. Ihm gebührt aber das Verdienst, dieses Gebiet u. a. durch ein grundlegendes Buch vorangetrieben zu haben (Koza, 1992). Sein Ansatz verwendet Lisp-Bäume zur Repräsentation von Programmen, es gibt jedoch auch lineare und Graph–Repräsentationen. Die Individuen in diesem Ansatz sind Computerprogramme, die danach bewertet werden, wie gut sie eine gegebene Aufgabe lösen.

Der Einsatzbereich reicht von Datenanalyse–Aufgaben, bei denen aber a priori kein Modell unterstellt werden muß, sondern mit evolviert wird, bis hin zu Teilgebieten der Künstlichen Intelligenz wie Robotik, Mustererkennung und Data Mining.

Das Gebiet des GP wird nur aus Gründen der Vollständigkeit erwähnt, es spielt in dieser Arbeit keine weitere Rolle.

4.4 Evolutionary Programming (EP)

Die Idee des *Evolutionary Programming (EP)* entstand wie die der anderen Verfahren dieses Kapitels ebenfalls Mitte der 60er Jahre in den USA (Fogel, Owens und Walsh, 1966). Evolviert wurden endliche Automaten, die einfache Zeichenfolgen vorhersagen sollten. Mutationen bestanden aus einem gleichverteilten Ziehen aus einem endlichen Alphabet. Die Fitneß wurde dann anhand der korrekt prognostizierten Zeichen berechnet. Jedes Elternindividuum produziert genau einen Nachkommen, die jeweils beste Hälfte der Individuen überlebt, in der Terminologie der ES also eine Art $(\mu + \mu)$–Strategie. David Fogel erweiterte das ursprüngliche Verfahren auf reellwertige Variablen mit normalverteilten Mutationen (Fogel, 1992). Das, was dort *Meta–EP* genannt wird, hat mit der Meta–Strategie dieser Arbeit (s. Kapitel 6) nichts zu tun, sondern bezeichnet lediglich einen Schrittweitenanpassungsmechanismus. Hält man sich an die Vorgaben aus (Fogel, 1992) und setzt die Parameter β_i, die die Mutation der Schrittweiten steuern, gleich 1, wird man nicht einmal ein $n = 30$–dimensionales Kugelmodell optimieren können. Eine weitere EP–Modifikation besteht darin, n selbstadaptive Varianzen pro Individuum zu erlauben. Allerdings muß man negative Varianzen verhindern, und der additive Mutationsprozeß der Varianzen ist im Gegensatz zum log–normalverteilten Schema der ES nicht driftfrei. Die Mutation von Objekt– und Strategievariablen — im Gegensatz zu ES in dieser Reihenfolge — ist die einzige Variantenquelle in EP, auf die Rekombination wird ganz verzichtet. Die ursprüngliche $(\mu + \mu)$– wird nun durch eine stochastische q–Turnierselektion abgelöst, die aber das Überleben des besten Individuums garantiert. Je größer q wird, desto mehr nähert man sich wieder der ursprünglichen deterministischen Variante.

In David Fogels Dissertation gibt es einen Beweis für die globale Konvergenz von EP, der aber mit dem realen Verhalten des Verfahrens für $n > 6$ (s.o.) nichts zu tun hat (Bäck, 1996). Wegen der immer größer werdenden Nähe zu Evolutionsstrategien wird der EP–Ansatz in dieser Arbeit nicht weiter berücksichtigt.

4.5 Evolutionsstrategien (ES)

Die *Evolutionsstrategien (ES)* wurden von Rechenberg und Schwefel Mitte der sechziger Jahre entwickelt (Rechenberg, 1973; Schwefel, 1977), zunächst als Ersatz für versagende etablierte Verfahren der experimentellen Optimierung, später dann auch für numerische Optimierprobleme. Ihr Einsatz lohnt immer dann, wenn bekannte Methoden versagen oder aber gar keine anderen Methoden existieren. Diese Generalisten–Eigenschaft bezahlt man natürlich mit einem erhöhten Zeitaufwand.

Die grundlegende Iterationsformel sequentieller Optimum–Suchmethoden (s. Kapitel 3) findet sich auch in den ES wieder, nur daß auf Ableitungen verzichtet wird und die Suchrichtung stochastisch bestimmt wird. Ein Individuum repräsentiert eine Lösung der gestellten Aufgabe, i. d. R. einen Parametervektor. Jedes x_i entspricht einem Gen, dessen Mutabilität (oder Schrittweite) über ein oder mehrere σ_i gesteuert wird. Als Phänotyp wird der zuhehörige Zielfunktionswert benutzt, eventuell versehen mit einem ggf. tödlichen Malus im Falle von Restriktionsverletzungen. Abbildung 4.4 verdeutlicht die verschiedenen Ebenen in ES: Man unterscheidet Geno– und Phänotyp sowie innerhalb des Genotyps Objekt– und Strategieparameter x_i bzw. σ_i. Diese Art der Trennung würde von Mitgliedern der EP–Gemeinde abgelehnt. Als Genotyp bezeichnet man sinnvollerweise jene Informationen, die Mutation und Rekombination ausgesetzt sind, und als Phänotyp das, was die Selektion als Entscheidungsgrundlage benutzt.

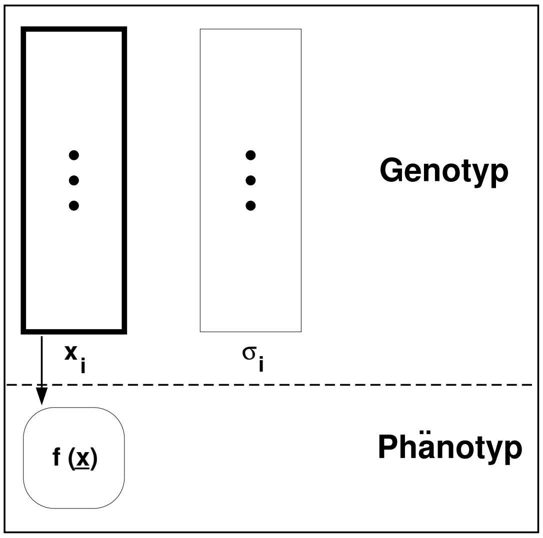


Abbildung 4.4: Geno– und Phänotyp bei ES, aus (Schwefel und Kursawe, 1998)

Die erste ES war eine $(1 + 1)$– oder zweigliedrige Evolutionsstrategie. Ein Elter erzeugt pro Generation einen Nachkommen, der seinen Erzeuger ersetzt, falls er mindestens ebenso gut ist. Um kleine Änderungen wahrscheinlicher als große Sprünge zu machen und in Analogie zum biologischen Vorbild eine gewisse Ähnlichkeit zwischen Elter und Nachkommen sicherzustellen, erfolgt die Mutation im Falle kontiuierlicher Variablen durch Addition eines normalverteilten Zufallsvektors mit Erwartungswert Null und Standardabweichung σ (bzw. Varianz σ^2), die im weiteren synonym zum intuitiveren Begriff der Schrittweite(–n) verwendet wird. Im Falle ganzzahliger Variablen wird aus gleichen Gründen eine Binomialverteilung verwendet. Nach dem zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung konvergiert die Summe unendlich vieler unabhängiger Zufallszahlen gleicher Verteilungsfunktion gegen die Normalverteilung. Da sich im Laufe der Iterationen einer ES viele Zufallszahlen unabhängig voneinander realisieren, erscheint die Verwendung der Normalverteilung im Mutationsoperator plausibel.

Zur Schrittweitenanpassung wird die sogenannte 1/5–Erfolgsregel verwendet, die Rechenberg aus zwei ganz unterschiedlichen Zielfunktionstopologien abgeleitet hat, nämlich dem Korridor– und dem Kugelmodell, die typische Topologien in Optimumnähe bzw. weit entfernt davon repräsentieren sollen (Rechenberg, 1973). Die deterministische Vorschrift zur Anpassung der Schrittweite lautet dann: Prüfe alle n Generationen, wieviele Verbesserungen durch die letzten $10 \cdot n$ Mutationen erzielt wurden. Ist diese Anzahl kleiner als $2 \cdot n$, dann multipliziere σ mit 0.85, andernfalls dividiere durch 0.85. Auf diese Weise hofft man, im „Evolutionsfenster“ zu bleiben, d. h. einem σ–Bereich, der sich über zwei Zehnerpotenzen erstreckt und in dem man mindestens die Hälfte des maximal möglichen Fortschritts (s. Kapitel 4.5.1) erzielt. Außerhalb dieses Bereichs gibt es sowohl Stagnation durch zu kleine als auch Rückschritt durch zu große Schrittweiten. Aus den hergeleiteten Formeln erkennt man ferner, daß die Fortschrittsrate umgekehrt proportional zur Problemgröße n und proportional zu einem topologiespezifischen Parameter ist. Bei einer linearen Funktion wie dem (offenen) Korridormodell ist dies trotz der nichtlinearen Restriktionen eine Konstante, d. h. man muß die Schrittweite nur einmal optimal einstellen, während sich beim Kugelmodell der Zielabstand ständig ändert und daher die Schrittweite laufend angepaßt werden muß.

Obwohl sie eine neue Idee realisiert, gehört die zweigliedrige Evolutionsstrategie zu den stochastischen Gradientenverfahren, besitzt also auch eher lokale Konvergenzeigenschaften. Aber schon für $n > 3$ konvergiert die zweigliedrige ES schneller als die einfachste Gradientenstrategie, die die Umgebung des aktuellen Punktes in n Tastschritten absucht, um dann den besten Versuch als Arbeitsschritt zu realisieren.

Die erste mehrgliedrige Strategie von Rechenberg war eine $(\mu + 1)$–ES. Wie bei der zweigliedrigen Strategie können keine Verschlechterungen auftreten, der erzeugte Nachkomme ersetzt ggf. das schlechteste Elternindividuum. Bei dieser Strategievariante kann für $\mu > 1$ zwar erstmals die Rekombination als Operator genutzt werden, sie setzte sich aber wegen der weiteren Verwendung der 1/5–Erfolgsregel zur Schrittweitensteuerung nicht durch. Denn diese Regel führt zur Stagnation des Verfahrens durch ständige Verkleinerung der Schrittweite, falls die maximal mögliche Erfolgswahrscheinlichkeit geringer als 1/5 wird. Um kein problemspezifisches Wissen in den Mechanismus zur Schrittweitensteuerung einfließen lassen zu müssen, entwickelte Schwefel die mehrgliedrigen oder $(\mu \overset{+}{,} \lambda)$–ES mit (selbst–)adaptiver Schrittweitensteuerung (Schwefel, 1977). Die Mutation besteht jetzt aus zwei Prozessen:

$$\begin{aligned}
\sigma_i^{(g+1)} &= \sigma_i^{(g)} \cdot \exp\left(\mathcal{NV}\ (0,\, \tau_0\,)\right) \cdot \exp\left(\mathcal{NV}\ (0,\, \tau_i\,)\right) \\
x_i^{(g+1)} &= \mathcal{NV}\left(x_i^{(g)},\, \sigma_i^{(g+1)}\right)
\end{aligned}$$

$$\text{mit:} \quad i = 1, \ldots, n, \quad \tau_0 \sim \frac{1}{\sqrt{n}}\,, \qquad \tau_i = 0 \quad \text{für } n_\sigma = 1$$

$$\tau_0 \sim \frac{1}{\sqrt{2 \cdot n}}\,, \qquad \tau_i \sim \frac{1}{\sqrt{2 \cdot \sqrt{n}}}\,, \qquad \tau_i > \tau_0 \quad \text{für } n_\sigma = n$$

$$\mathcal{NV}\ (E,\, s\,) : \text{Realisierung einer Normalverteilung mit Erwartungswert } E \text{ und Standardabweichung } s$$

In die Heuristiken für τ_0 und τ_i geht eigentlich jeweils als Proportionalitätsfaktor φ^*_{max} ein, also der normierte erwartete maximale Fortschritt. Da man diesen aber für beliebige Zielfunktionen nicht kennt, setzt man diesen bei mehrgliedrigen Strategien oft auf Eins, was eine eher konservative Wahl darstellt.

Die beiden Parameter τ_0 und τ_i kann man sich als Lernraten der Schrittweitenmutation vorstellen. Je größer diese Werte gewählt werden, desto größer ist die Mutabilität der Schrittweiten. Die Zufallsvariable $\exp\left(\mathcal{NV}\ (0,\, \tau_0\,)\right)$ realisiert sich einmal pro Nachkomme, die Zufallsvariable $\exp\left(\mathcal{NV}\ (0,\, \tau_i\,)\right)$ komponentenweise. Schwefel empfiehlt daher, die Relation $\tau_0 < \tau_i$ einzuhalten, damit erstere die Ausdifferenzierung unterschiedlicher Einzelschrittweiten nicht behindert. In der ersten Version der mehrgliedrigen Strategien gab es auch bei $n_\sigma = n$ zunächst nur ein τ_0 (Schwefel, 1977). Die ggf. nötige individuelle Schrittweiteneinstellung sollte dann durch die Rekombination hergestellt werden. Bei einer Funktion, die skalierte Schrittweiten für eine gute Konvergenzgeschwindigkeit verlangt (s. Kapitel 5.1), aber nicht zum Katalog der von Schwefel untersuchten Zielfunktionen gehörte, reicht dies jedoch nicht aus, um die Schrittweiten passend zueinander einzustellen. Die oben beschriebene Heuristik zur Aufteilung der Schrittweitenmutation in einen globalen (τ_0) und einen individuellen Anteil (τ_i) führt zwar bei entsprechender Zielfunktion zu einer Beschleunigung der Konvergenz etwa um den Faktor 4.5 ($n = 100$, 2 500 Generationen). Dennoch realisiert diese Aufteilung lediglich *eine* Möglichkeit und ist bisher noch nicht theoretisch untermauert worden.

Die Objektvariablen werden weiter additiv normalverteilt mutiert, die Strategieparameter dagegen lognormalverteilt, damit keine negativen Schrittweiten auftreten können und Vergrößerungen und Verkleinerungen gleichwahrscheinlich auftreten und der Prozeß ohne Selektion driftfrei verläuft. Der Median der Lognormalverteilung beträgt Eins, womit auch bei den Schrittweiten die Ähnlichkeit zwischen Eltern und Nachkommen erhalten bleibt. Es gibt auch andere Möglichkeiten zur Schrittweitensteuerung, die aber in dieser Arbeit nicht berücksichtigt werden (Ostermeier, 1992; Ostermeier, Gawelczyk und Hansen, 1994).

Da jetzt mit einer Population gearbeitet wird, falls $\mu > 1$ gilt, ergeben sich für die Rekombination folgende Möglichkeiten, jeweils getrennt für Objekt– und Strategieparameter einstellbar. Die Ziffern stellen den Code dar, der in allen Diagrammen dieser Arbeit, die mit Rekombination zu tun haben, verwendet wird (in Klammern die entsprechende Notation aus (Bäck, 1996)). So kodiert z. B. der Rekombinationstyp 521 (Id-) eine global intermediäre Rekombination[3] für die Objektvariablen, eine lokal diskrete Rekombination für die Schrittweiten und keine Rekombination auf den Rotationswinkeln, sofern diese überhaupt benutzt werden (s. Seite 41).

1 keine Rekombination (-)
Ein Nachkomme wird durch Klonen eines Elters erzeugt.
Diese Variante wird nur der Vollständigkeit halber und zu Vergleichszwecken mit aufgeführt, i. d. R. sind ES mit Rekombination sowohl robuster als auch effizienter.

2 diskrete Rekombination zweier Eltern (d)
Zwei verschiedene Eltern[4] werden zufällig ausgewählt, und der Nachkomme erbt die jeweilige Information zufällig von Elter 1 oder Elter 2.

3 diskrete Rekombination aller μ Eltern (D)
Für jedes „Gen“ wird ein Elternindividuum zufällig ausgewählt, das die entsprechende Information vererbt.
Die Population wird in dieser Variante als Genpool betrachtet, aus dem einzelne Gene gezogen werden.

4 intermediäre Rekombination zweier Eltern (i)
Zwei verschiedene Eltern werden zufällig ausgewählt, und der Nachkomme erbt die Mittelwerte der beiden Eltern.

---

[3]Rechenberg und Beyer bezeichnen dagegen die Mittelwertbildung aus allen Eltern als global intermediäre Rekombination. Rechenberg begründet diesen Operator mit der Notwendigkeit des „Gradientenzielens“ in hochdimensionalen Suchräumen, Beyer spricht von „genetischer Reparatur“. Diese Schwerpunkt–Rekombination wird in dieser Arbeit nicht untersucht.

[4]Manche ES–Varianten verzichten auf den Test auf Ungleichheit, in vielen Veröffentlichungen bleibt offen, ob dieser Test durchgeführt wird oder nicht.

5 intermediäre Rekombination aller μ Eltern (I)
Für jedes „Gen“ werden je zwei verschiedene Eltern zufällig ausgewählt, deren Mittelwert dem Nachkommen zugewiesen wird.

6 1–Punkt–Crossover (c_1)
Je zwei verschiedene Eltern werden zufällig ausgewählt. Der Nachkomme erbt die „Gene“ bis zur zufälligen Bruchstelle vom ersten Elter und den zweiten Teil entsprechend vom zweiten Elternindividuum.

7 2–Punkt–Crossover (c_2)
Analog zum 1–Punkt–Crossover, nur mit einer zweiten zusätzlichen Zufallsbruchstelle.

Ohne Winkel gibt es $7^2 = 49$ Rekombinationsmöglichkeiten, mit Winkeln $7^3 = 343$. Benutzt man nur eine Schrittweite, reduziert sich die Anzahl der Möglichkeiten auf 35. Schwefel empfiehlt eine diskrete Rekombination auf den Objekt– sowie eine globale intermediäre Rekombination auf den Strategievariablen (251 bzw. dI-) als Standard (Schwefel, 1987).

Wie in Kapitel 2 erwähnt, lassen sich alle diese Varianten biologisch rechtfertigen. Als weitere Variante der intermediären Rekombination kann man statt einer Gleichgewichtung der Elterninformation (Faktor 0.5) auch einen Zufallsfaktor aus dem Intervall $[0, 1]$ verwenden, der sich bei jeder Rekombination neu realisiert. In Abgrenzung zu den GA sei hervorgehoben, daß in ES alle Elternindividuen gleiche Reproduktionswahrscheinlichkeiten haben (s. Abbildung 4.6).

Abbildung 4.5 veranschaulicht die Unterschiede zwischen intermediärer und diskreter Rekombination und zeigt ferner, daß die möglichen Nachkommen ihren Eltern deutlich näher sind als beim GA–Crossover (s. Abbildung 4.2).

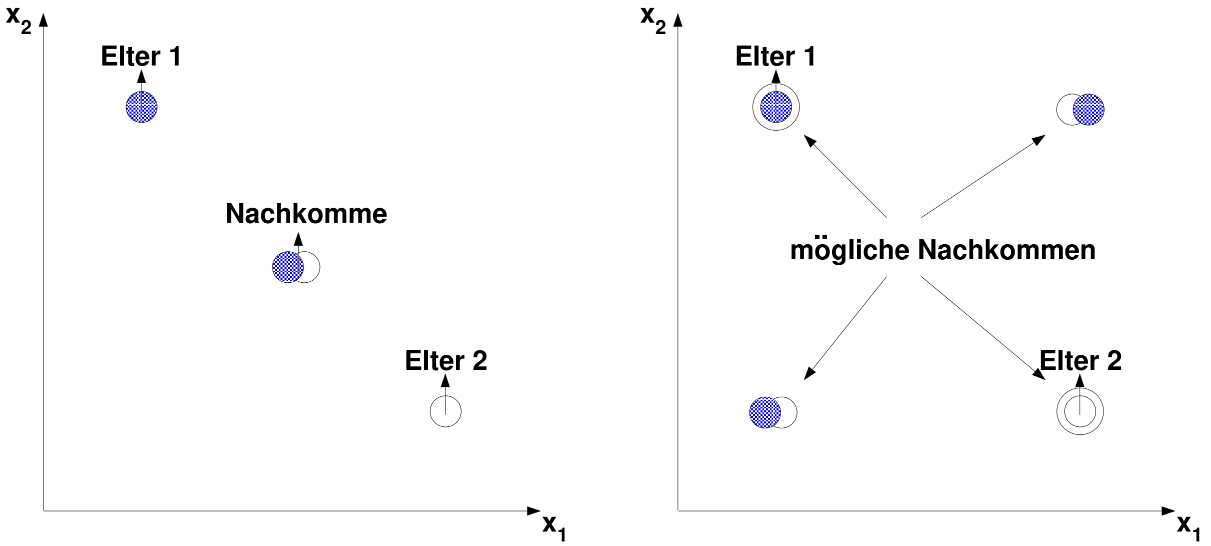


Abbildung 4.5: ES–Rekombinationstypen: gleichgewichtet intermediär (links) und diskret (rechts)

Die Bewertung eines Individuums erfolgt üblicherweise über den Zielfunktionswert. Dieser kann im Gegensatz zu den in dieser Arbeit verwendeten einfachen Funktionen aber auch das Resultat einer komplexen Simulation sein. Da Optimierprobleme aus dem technischen Bereich häufig restriktionsbehaftet sind, soll an dieser Stelle noch erwähnt werden, wie Nebenbedingungen in EA berücksichtigt werden können:

- Unzulässige Individuen werden schon durch die Wahl der Kodierung — sofern vorhanden — prinzipiell ausgeschlossen.
- Unzulässige Individuen werden sofort verworfen und der Rekombinations– und Mutationsprozeß solange wiederholt, bis λ gültige Individuen erzeugt worden sind.
- Unzulässige Individuen werden mit einem Fitneßmalus bestraft.
- Unzulässige Individuen werden genetisch „repariert“.

In der älteren Literatur findet man durchgehend die $(\mu \overset{+}{,} \lambda)$–Notation, neuerdings werden Plus– und Komma–Variante durch folgende Notation zusammengeführt: $(\mu,\ \kappa,\ \lambda,\ \rho)$–ES (Schwefel und Rudolph, 1995). Mittels κ kann man das maximale Alter der Individuen steuern, für $\kappa = 1$ erhält man eine Komma–, für $\kappa = \infty$ eine Plus–Strategie. Der Parameter ρ gibt die an der Erzeugung eines Nachkommen beteiligten Eltern an. Der Grundalgorithmus einer $(\mu \overset{+}{,} \lambda)$–ES sieht wie folgt aus:

initialisiere Population $P^{(g:=0)} \in \mathbb{R}^n$
bewerte die Individuen mittels f
`while` Terminierungskriterium nicht erfüllt `do`
  rekombiniere λ Nachkommen aus 1 bis μ Elternindividuen aus $P^{(g)}$
  mutiere die λ Nachkommen
  bewerte die λ Nachkommen mittels f
  selektiere μ Individuen für $P^{(g+1)}$:
    `if` $(\mu\,,\ \lambda)$–Selektion `then` selektiere μ aus λ Nachkommen
      `else` selektiere μ aus μ Eltern und λ Nachkommen
    `fi`
  setze $g := g + 1$
`od`

Als Terminierungskriterium kann die absolute oder relative Zielfunktionsdifferenz zwischen momentanen Individuen oder über der Zeit, eine maximale Anzahl an Generationen, eine maximale Rechenzeit oder eine Kombination dieser Kriterien benutzt werden.

Das Verhältnis von λ zu μ wird als *Selektionsdruck* bezeichnet. Über ihn kann man den Charakter der ES in Richtung einer eher lokalen Suche (harte Selektion, Pfadorientierung) oder eher globalen Suche (sanfte Selektion, Volumenorientierung) einstellen. Die Komma–Selektionsvariante erscheint nur auf den ersten Blick verschwenderisch, weil sie auch Verschlechterungen von einer Generation zur nächsten in Kauf nimmt. Aber die Komma–Strategien erleichtern eine ständige Selbstadaptation der Strategieparameter und eine Überwindung lokaler Optima. Die Schrittweiten gehen zwar nicht direkt in die Fitneßberechnung ein (s. Abbildung 4.4), beeinflussen aber natürlich die Chancen der erzeugten Nachkommen. Die Strategieparameter werden also zeitverzögert über die Nachkommen indirekt mitbewertet. Ferner erlaubt es nur diese Variante, wandernden Optima zu folgen. Die Plus–Selektionsvariante dagegen erhält zwar Individuen mit guter Fitneß, jedoch kann diese auf einem besonders guten Zufallstreffer basieren und für den weiteren Fortschritt ungeeignete Schrittweiten „verstecken“, die dann prinzipiell unendlich lange an die Nachkommen vererbt werden. Dies ist kein theoretisches Szenario, sondern eine reale Gefahr, da sich die Fortschrittsgeschwindigkeit umgekehrt proportional zur Problemgröße n verhält (s. Kapitel 4.5.1), die Suche in Unterräumen sich also temporär immer auszahlt.

Abbildung 4.6 zeigt die identische Reproduktionswahrscheinlichkeit der Eltern einer ES–Generation (vgl. Abbildung 4.3). Ferner verdeutlicht sie das Prinzip der extinktiven Selektion: Die $(\lambda - \mu)$ schlechtesten Individuen sind von der Reproduktion ausgeschlossen.

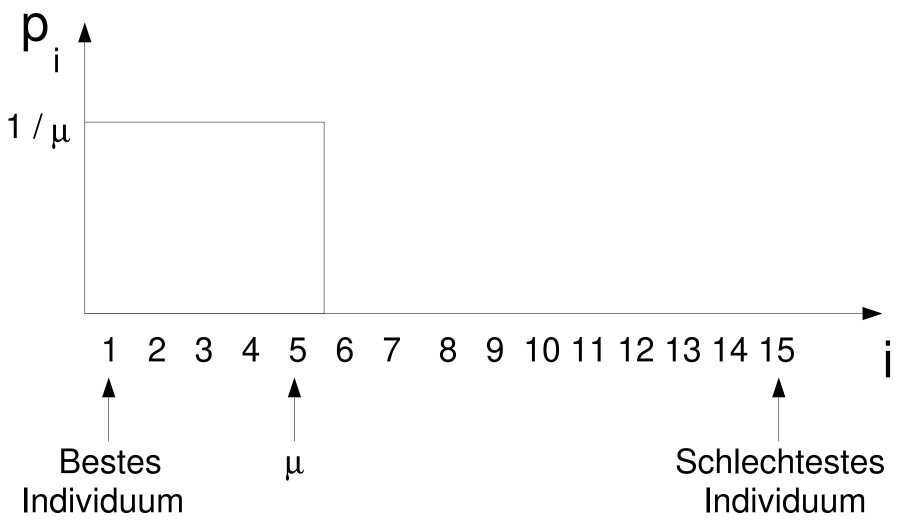


Abbildung 4.6: Reproduktionswahrscheinlichkeit p_i bei einer $(5 \overset{+}{,} 15)$–ES, aus (Schwefel und Kursawe, 1998)

Der Vollständigkeit halber soll noch eine weitere ES–Variante erwähnt werden, die eine noch größere Flexibilität durch Korrelationen zwischen den Änderungen der Variablen bietet. Das Genom eines Individuums verfügt dann zusätzlich über bis zu $n \cdot (n-1)/2$ Rotationswinkel, die dafür sorgen, daß sich die Hyperellipsoide gleicher Mutationswahrscheinlichkeitsdichte nicht nur achsenparallel, sondern frei im Raum orientieren können.

Diese Winkel werden wie die anderen Parameter rekombiniert und additiv normalverteilt mutiert, allerdings mit einer festen Standardabweichung (typischerweise $5° = 0.0873$). Während man jedoch durch Experimente mit springenden Optima (Bäck und Hoffmeister, 1994) oder durch die Wahl einer entsprechenden Zielfunktion (s. Abbildung 5.6) die Anpassungsfähigkeit der Schrittweiten beobachten kann, gelingt dies für die Doppelsumme (s. Kapitel 5.1) bei den Winkeln nicht mehr.

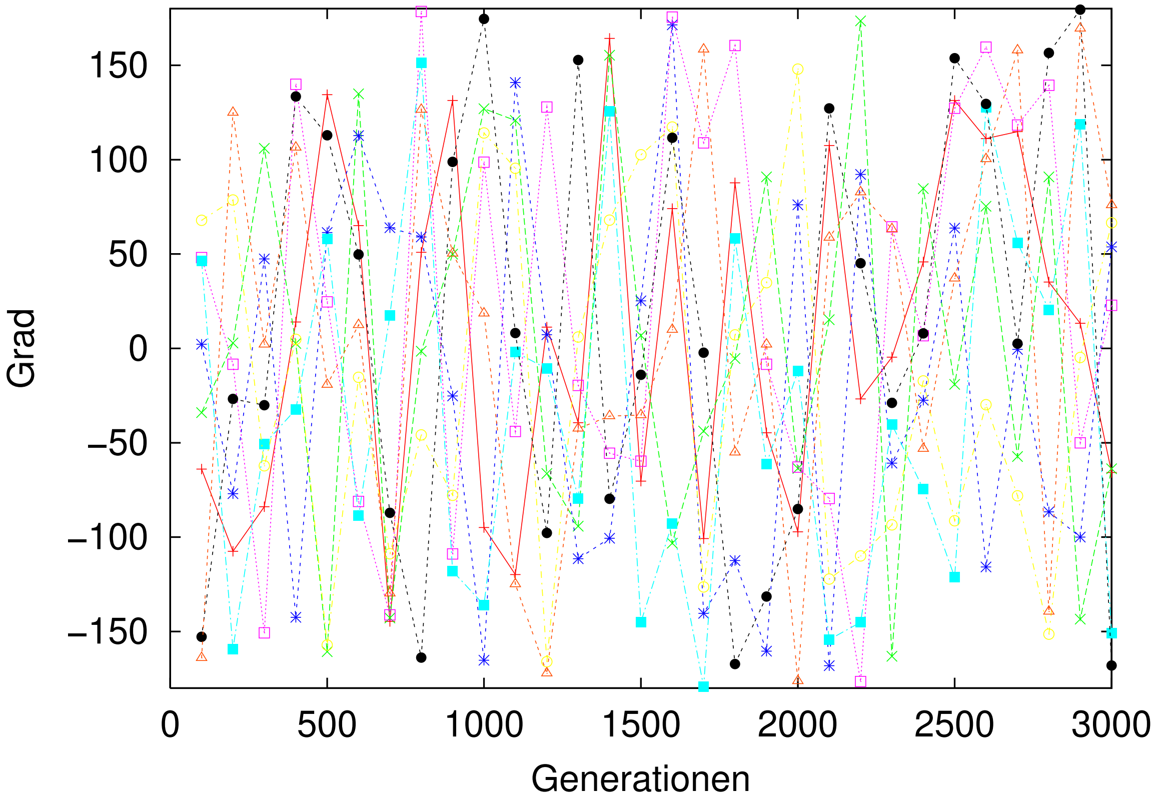


Abbildung 4.7: Entwicklung einiger Winkel über der Zeit

Anstatt sich im Laufe der Iterationen an den elliptischen Höhenlinien auszurichten, entlang derer der größte Fortschritt zu erzielen ist, scheinen sie beliebig im Raum zu streuen. Somit können sie als Reparaturmechanismus eine durchaus positive Wirkung auf die Konvergenzgeschwindigkeit haben, eine Selbstadaptation wie bei den Schrittweiten findet aber nicht statt, obwohl die Selbstähnlichkeit der Funktion dies in jeder Phase des Laufes mit einer erhöhten Konvergenzgeschwindigkeit honorieren müßte. Abbildung 4.7 zeigt die Werte von 8 Winkeln einer $(15\,,\,100)$–ES für die $n = 15$–dimensionale Doppelsumme. Der Lauf erstreckte sich über 3 000 Generationen, was bei den Fitneßwerten des Generationsbesten einer Entwicklung über 103 Zehnerpotenzen entspricht.

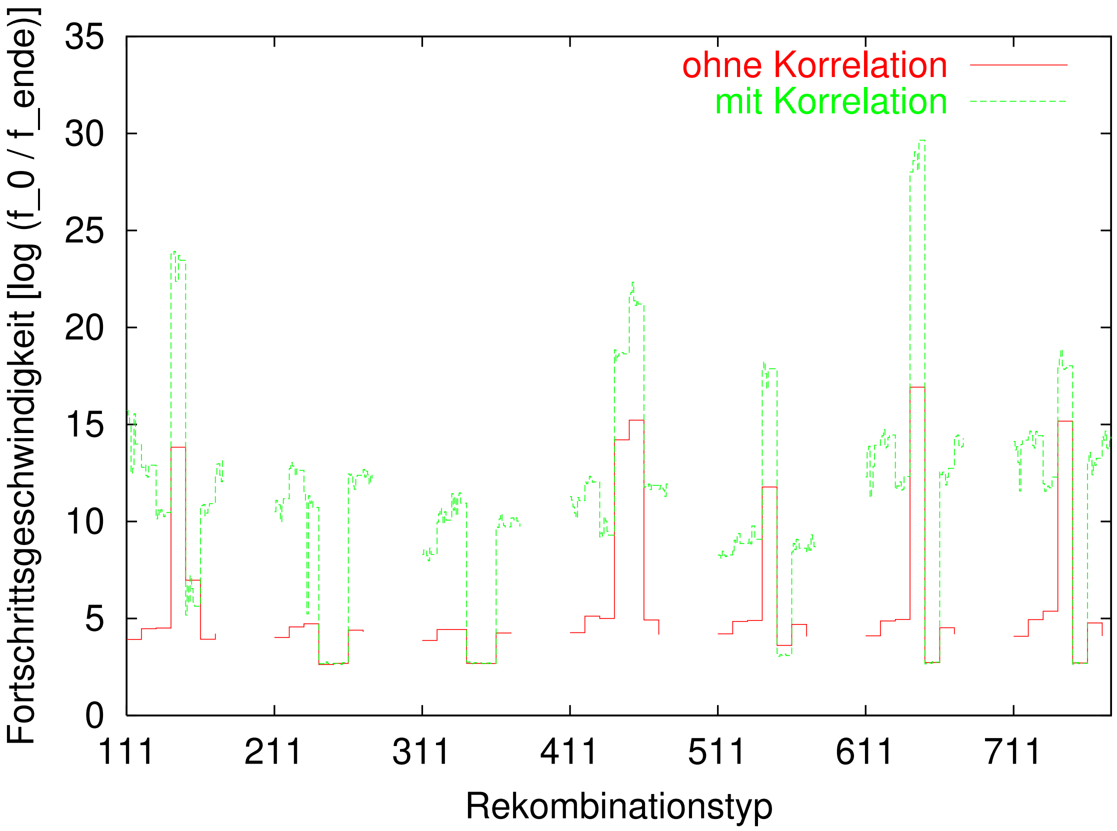


Abbildung 4.8: (15 , 100)–ES, Doppelsumme, alle Rekombinationsmöglichkeiten, korrelierte und unkorrelierte Mutationen, $n = n_\sigma = 30$, Mittel von je 25 Läufen über 2 500 Generationen, $x_i^{Start} = -1.0 \cdot 10^9, \quad (i = 1, \ldots, n)$, $\sigma_i^{Start} = 2.0 \cdot 10^7 \quad (i = 1, \ldots, n_\sigma)$, $\tau_0 = 0.1291$, $\tau_i = 0.3021$

Wie Abbildung 4.8 demonstriert, ist die Einstellung des richtigen Rekombinationstyps wichtiger als die Frage, ob mit oder ohne Korrelationen mutiert wird. Berücksichtigt man ferner, daß die maximale Anzahl der Winkel quadratisch mit n steigt, λ also sinnvollerweise auch quadratisch wachsen sollte[5], und ihre Erzeugung sehr rechenintensiv ist ($n = 30 \rightarrow$ siebenfache Laufzeit, $n = 50 \rightarrow$ vierzehnfache Laufzeit, $n = 100 \rightarrow$ siebzehn– bis vierzigfache Laufzeit, bei jeweils gleicher Annäherung an das Optimum, mit einer bzw. n Schrittweiten bei der Strategie ohne korrelierte Mutationen), dann erscheint es

[5]Empfohlen werden folgende Populationsgrößen für ES mit korrelierten Mutationen (Rudolph, 1992):

n	5	10	20	25	50	75	100
μ	2	7	25	39	147	325	572
λ	20	60	207	313	1 180	2 602	4 580

m. E. gerechtfertigt, auf diese ES–Variante in den folgenden Untersuchungen nicht näher einzugehen.

Fairerweise sollte erwähnt werden, daß man mit einer härteren Selektion ($\mu = 10$ bzw. $\mu = 5$) die Divergenz[6] (tiefe „Plateaus“ bei z. B. $[2,3][4,5]*$) für $n = 30$ bei einigen Rekombinationstypen verhindern kann, für $n = 100$ aber nicht mehr. Ferner sind alle Einstellungen, die nach 2 500 Generationen nicht mehr als 5 Zehnerpotenzen überwunden haben und die auch nach 25 000 Generationen nicht weitergekommen sind, nicht unbedingt als konvergent zu bezeichnen — gemessen an dem, was die besseren Rekombinationstypen leisten. Außerdem wird die Vermeidung der Divergenz z. T. mit einer verminderten Konvergenzgeschwindigkeit bei den besten Rekombinationstypen für $\mu = 15$ bezahlt, ein erster Hinweis darauf, daß die isolierte Betrachtung von Operatoren gefährlich ist (s. Kapitel 5.2).

Für Implementierungsdetails von ES mit korrelierten Mutationen sei auf (Schwefel, 1977) oder (Bäck, 1996) verwiesen, für eine theoretische Untersuchung auf (Rudolph, 1992).

4.5.1 Theorie

Alle theoretischen Resultate basieren auf Vereinfachungen. Im Falle der ES–Theorie greifen diese sowohl bei der Wahl der unterstellten Zielfunktion als auch beim eigentlichen Algorithmus. Es klafft nach wie vor eine große Lücke zwischen dem Schwierigkeitsgrad jener realen Probleme, die ES erfolgreich gelöst haben, und den einfachen Strukturen, die in der Theorie unterstellt werden. Globale Konvergenz in beliebigen Zielfunktionstopologien (*Robustheit*) mit einer möglichst niedrigen Anzahl an Zielfunktionsauswertungen (*Effizienz*) werden allerdings immer widersprüchliche Ziele bleiben.

Bisher gibt es noch keine umfassende Theorie für alle $(\mu, \kappa, \lambda, \rho)$–ES. Einige Grenzfälle sind detailliert untersucht worden, so jene mit $\kappa = 1$ oder $\kappa = \infty$, $\rho = 1$ oder $\rho = \mu$ und $\mu = 1$, auf die im weiteren nur kurz eingegangen wird.

Ein früher Konvergenzbeweis für eine $(1 + 1)$–ES verzichtet zwar auf Annahmen über die Zielfunktion, garantiert die globale Konvergenz aber nur nach unendlicher Zeit und falls die Schrittweite immer größer als Null bleibt (Born, 1978). Dieses Resultat läßt sich natürlich auf $(\mu + \lambda)$–Strategien verallgemeinern. Etwas schwieriger ist ein Konvergenzbeweis für nicht–elitäre, d. h. Komma–Strategien. Rudolph gibt die Bedingungen an, unter denen die Konvergenzsicherheit erhalten bleibt (Rudolph, 1994).

[6] Daß divergente Läufe nicht durch negative Werte in den Diagrammen gekennzeichnet sind, hängt mit der Art der Messung zusammen: Es wird der während eines Laufes beste aufgetretene Funktionswert benutzt, und auch divergente Läufe erzielen anfangs Verbesserungen gegenüber dem Startwert (s. Kapitel 5.1).

Beyer hat die Theorie am weitesten in Richtung der wirklich benutzten ES entwickelt. Er macht Aussagen über allgemeine $(\mu\,,\,\lambda)$–ES mit global mittelnder Rekombination (Beyer, 1995b; Beyer, 1995a). Die Theorie gilt — streng genommen — allerdings nur für das Kugelmodell und für λ, λ/μ und $n \gg 1$. Die Fortschrittsgeschwindigkeit φ mißt den zurückgelegten Weg im Parameterraum, beim Kugelmodell also einfach den euklidischen Abstand zum Optimum.

$$\varphi^* \simeq \overbrace{c_1 \cdot \sigma^*}^{\text{linearer Gewinn}} - \underbrace{\frac{1}{2 \cdot c_2} \cdot \sigma^{*\,2}}_{\text{quadratischer Verlust}}$$

$$\text{mit} \quad \varphi^* = \frac{n}{r} \cdot \varphi, \quad \sigma^* = \frac{n}{r} \cdot \sigma \quad \text{und} \quad \varphi = E\{r^{(g)} - r^{(g+1)}\} = \varphi(n, r, \sigma)$$

$$\text{sowie} \quad r: \text{Abstand zum Optimum}, \quad n: \text{Anzahl Dimensionen}, \quad \sigma: \text{„Schrittweite“}$$

Optimal sind dann folgende Einstellungen: $\hat{\sigma}^* = c_1 \cdot c_2$ mit zugehörigem $\hat{\varphi}^* = 1/2 \cdot c_1^2 \cdot c_2$. Daß der erwartete Fortschritt nur von einer Generation g zur nächsten maximal werden soll, macht bei einer unimodalen Funktion wie dem Kugelmodell keinen Unterschied zu einer globalen Betrachtungsweise über alle Generationen. Bei multimodalen Funktionen wäre dieses Maß allerdings sinnlos, weil man temporär auch Rückschritte in Kauf nehmen muß, um das globale Optimum zu erreichen. In Abhängigkeit vom Rekombinationstyp berechnen sich die sogenannten Fortschrittsbeiwerte c_1 und c_2 wie folgt:

- ohne Rekombination:

$$c_1 \simeq \sqrt{2 \cdot \ln \frac{\lambda}{\mu}}, \quad c_2 = 1 \qquad \Longrightarrow \qquad \hat{\varphi}^* \simeq \ln \frac{\lambda}{\mu}, \quad \hat{\sigma}^* \simeq \sqrt{2 \cdot \ln \frac{\lambda}{\mu}}$$

- mit global mittelnder Rekombination:

$$c_1 \simeq \sqrt{2 \cdot \ln \frac{\lambda}{\mu}}, \quad c_2 = \mu \qquad \Longrightarrow \qquad \hat{\varphi}^* \simeq \mu \cdot \ln \frac{\lambda}{\mu}, \quad \hat{\sigma}^* \simeq \mu \cdot \sqrt{2 \cdot \ln \frac{\lambda}{\mu}}$$

- mit global diskreter Rekombination:

$$c_1 \simeq \sqrt{2 \cdot \mu \cdot \ln \frac{\lambda}{\mu}}, \quad c_2 = 1 \qquad \Longrightarrow \qquad \hat{\varphi}^* \simeq \mu \cdot \ln \frac{\lambda}{\mu}, \quad \hat{\sigma}^* \simeq \sqrt{\mu} \cdot \sqrt{2 \cdot \ln \frac{\lambda}{\mu}}$$

Abbildung 4.9 zeigt die gute Übereinstimmung zwischen Theorie und Empirie für den Fall ohne Rekombination.

Bei den heute gebräuchlichen mehrgliedrigen ES mit Rekombination darf man also für große n und λ sowie nicht zu kleine μ erwarten, daß sich die Konvergenzrate proportional

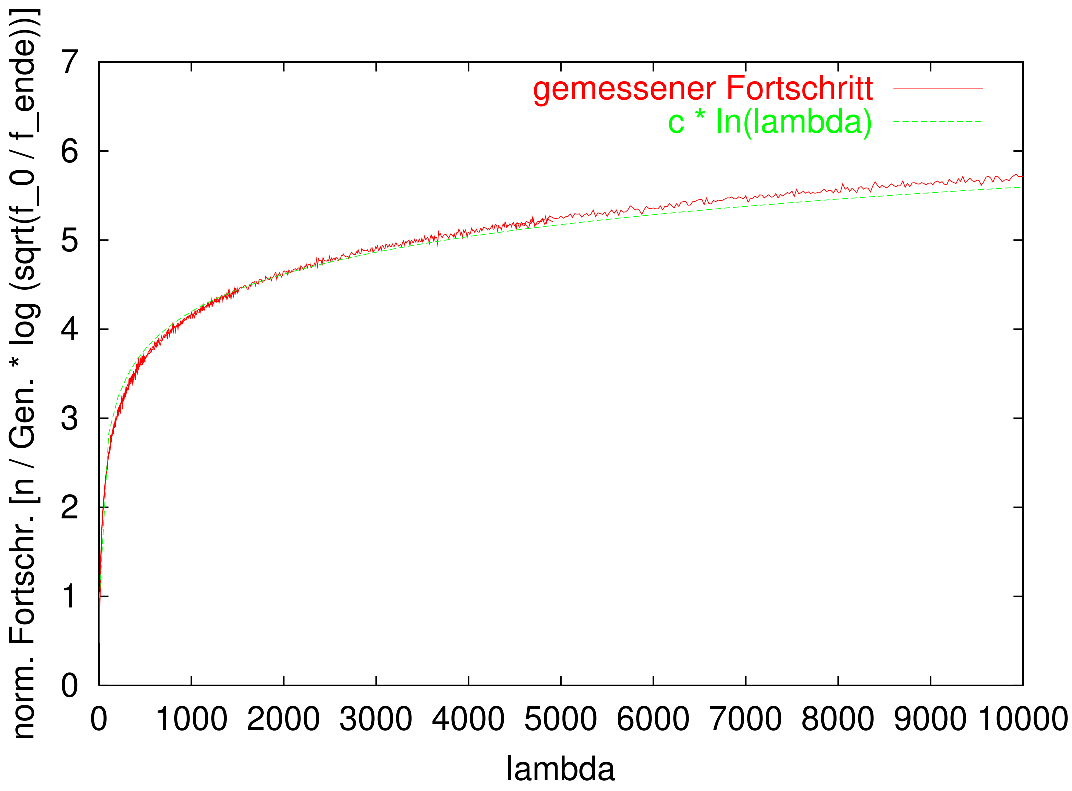


Abbildung 4.9: $(1\,,\,\lambda)$–ES, Kugelmodell, $n = 30, n_\sigma = 1$, Vergleich mit der Theorie, $c = 0.6075$

zum Selektionsdruck $\ln(\lambda/\mu)$ erhöht, ferner zusätzlich linear mit der Diversität in μ, die von der Rekombination genutzt wird, also insgesamt: $\hat{\varphi}^* \sim \mu \cdot \ln(\lambda/\mu)$.

Bei multimodalen Problemen wächst mit μ nur die Robustheit des Verfahrens, die Konvergenzgeschwindigkeit ist bei diesen schwierigen Problemen nachrangig.

Untersuchungen der $(1\,,\,\lambda)$–ES haben zur Empfehlung von $\lambda = 5$ oder $\lambda = 6$ als optimal geführt. Dieses Verhältnis wurde später auf die Strategien mit $\mu > 1$ übertragen, so daß ein Verhältnis von μ/λ von 1/5 bis 1/7 bis heute beibehalten wird (Schwefel, 1987; Schwefel, 1995). Beyer kommt für $n < \infty$ zu einer Empfehlung von $\mu^* < 0.27 \cdot \lambda$, d. h. μ/λ sollte sich zwischen 1/3 und 1/4 bewegen, allerdings mit einer ES, die global mittelnd rekombiniert (Beyer, 1995a).

Ebenfalls von Beyer stammt die Bestätigung der Empfehlung, τ_0 proportional zu $1/\sqrt{n}$ einzustellen bzw. für $\lambda > 10$: $\tau_0 \sim c_{1,\lambda}/\sqrt{n}$ (Beyer, 1995c), wobei $c_{1,\lambda}$ den sogenannten Fortschrittsbeiwert (Rechenberg) oder –koeffizienten (Beyer) repräsentiert, der sich für die praktisch interessanten λ–Werte zwischen 1 und 3 bewegt und durch die Näherung $c_{1,\lambda} \approx \sqrt{2 \cdot \ln \lambda}$ ohne numerische Integration berechnen läßt.

Ostermeier (Ostermeier, 1997) bestreitet, daß in Unterräumen des $\mathbb{R}^n$ eine höhere Fortschrittsgeschwindigkeit erzielbar ist. Stattdessen spricht er von „... zufälligen Schrittweitenschwankungen im Zusammenhang mit einer verrauschten Schrittweitenselektion ...“ (Seite 66), wenn es um die Degeneration in Unterräume geht. Eine korrekte Beobachtung, aber keine Erklärung. Für kleine Populationen (typischerweise eine (1 , 10)–ES) schlägt er eine deterministische Schrittweitensteuerung vor, die durch zwei Dämpfungsfaktoren und eine Mittelung über mehrere Generationen die Degeneration des Suchraumes verhindern soll. Vermeiden läßt sich dieses Verhalten allerdings auch durch eine andere Einstellung der externen ES–Parameter: Man kann z. B. eine $(10 \overset{+}{,} 30)$–ES mit intermediärer Rekombination verwenden, die den Weg zum Optimum eines 30–dimensionalen Kugelmodells auch mit eigentlich für dieses symmetrische Problem überflüssigen $n_\sigma = 30$ Einzelschrittweiten ohne Stagnation findet.

4.5.2 Praxis

Trotz der noch nicht voll entwickelten Theorie gibt es eine Fülle von ES–Anwendungen, weil diese Algorithmen folgende Vorteile bieten[7]:

- Sie sind leicht implementierbar.
- Es werden nicht unbedingt Kenntnisse über die Problemstruktur benötigt.
- Sie sind aber auch leicht mit problemspezifischen Heuristiken kombinierbar, sofern diese verfügbar sind[8] (Michalewicz, 1994).
- Sie sind geeignet für hochdimensionale reelle Suchräume.
- Sie sind geeignet für diskret–reelle Suchräume variabler Dimension (Lohmann, 1992; Schütz, 1994; Bäck und Schütz, 1995).
- Sie sind beliebig parallelisierbar als Migrations– oder Diffusionsmodelle (Hoffmeister, Rudolph und Sprave, 1991; Rudolph, 1991; Lohmann, 1991; Hammel und Bäck, 1994).
- Sie sind auf Probleme mit mehrfacher Zielsetzung erweiterbar (Kursawe, 1990; Laumanns, Rudolph und Schwefel, 1998).
- Sie sind auf Problemstellungen mit zeitabhängig wandernden Optima einsetzbar (Kursawe, 1996).
- Sie sind geeignet für stochastisch gestörte Zielfunktionen (Hammel und Bäck, 1994).

[7] Die angegebenen Quellen erheben keinen Anspruch auf Vollständigkeit.

[8] Dies gilt nicht nur für ES, sondern für alle EA. Integriert man eine Heuristik erfolgreich, sollte man sich allerdings die Frage ehrlich beantworten, welchen Erfolgsanteil der EA dann noch hat. In Tagungsbänden findet man immer wieder EA, die sich dem *Travelling Salesman Problem (TSP)* widmen und nur deshalb gute Lösungen finden, weil sie die Lin–Kernighan–Heuristik benutzen, diese aber „Mutation“ oder „lokale Suche“ nennen.

- Man kann sie auch mit subjektiver Selektion betreiben (Herdy, 1996; Herdy, 1997).

Die folgende Aufstellung soll lediglich einen Eindruck von der breiten Anwendungsmöglichkeit von ES geben, auch hier gibt es umfassendere Sammlungen (Bäck, Hoffmeister und Schwefel, 1992; Alander, 1994).

- Parameteroptimierung, d. h. auch Optimierung externer Simulationsmodellparameter
- Optimierung (künstlicher) neuronaler Netze in Struktur und Parametern
- Nährmedienoptimierung für biotechnologische Prozesse,
- Tragwerkoptimierung
- Standortplanung
- Produktionsplanung
- Einsatzplanung von Kraftwerken
- Routing in Kommunikationsnetzwerken
- Fahrplanoptimierung
- Schaltkreisentwurf
- Medizintechnik
- Optik

Kapitel 5

Parameterstudien an Evolutionsstrategien

Complex systems exhibit unexpected behaviour.

Evolutionäre Algorithmen im allgemeinen und Evolutionsstrategien im besonderen müssen bzgl. ihrer Effizienz problemspezifisch entworfenen Algorithmen unterlegen sein. Als Ausgleich werden jedoch ihre Anspruchslosigkeit an die Zielfunktion, ihre Robustheit und ihre Adaptierbarkeit als Argumente ins Feld geführt. In diesem Kapitel wird dieser Anspruch anhand einiger Zielfunktionen überprüft. Ferner enthalten ES einige Heuristiken, deren Gültigkeit hinterfragt werden soll.

Bei einigen der in diesem Kapitel vielleicht noch willkürlich anmutenden Untersuchungen gab es bereits die Idee, die aufgetretenen Probleme mittels einer Meta–Strategie zu lösen. Im Hinblick auf die Validierung dieses Gedankens und das spätere Operatoren–Design stellten sich diese Untersuchungen als durchaus nützlich heraus. Schließlich liefern die Resultate dieses Kapitels einen Maßstab für die Qualität der mit dem Meta–Algorithmus gefundenen Parameter (s. Kapitel 6).

5.1 Verwendete Maße und Zielfunktionen

Alle folgenden Resultate wurden mit einer (Standard)–$(15\,,\,100)$–Evolutionsstrategie erzielt. Dargestellt ist jeweils der Mittelwert der normierten Konvergenzgeschwindigkeit von 25 Läufen. Die Anzahl der Generationen variiert in Abhängigkeit von n und n_σ, wurde aber immer so gewählt, daß die Fitneßwerte sich bei gut konvergierenden Läufen über viele Zehnerpotenzen erstrecken, d. h. die ES sich in einem „eingeschwungenen" Zustand befindet. Ein Lauf wird vor Erreichen des jeweiligen Generationslimits abgebrochen, falls sich die

Fitneßwerte von bestem und schlechtestem Individuum um weniger als 10^{-10} unterscheiden. Gestartet wird immer an der Stelle $x_i^{Start} = -1.0 \cdot 10^9, \quad (i = 1, \ldots, n)$. Bei praktischen Anwendungen gibt man typischerweise ein Intervall an, aus dem die x_i^{Start} zufällig gezogen werden. Die Startschrittweiten betragen $\sigma_i^{Start} = 2.0 \cdot 10^7 \quad (i = 1, \ldots, n_\sigma)$. Anhand von Vergleichsläufen wurde sichergestellt, daß diese Einstellungen keinen entscheidenden Einfluß auf das beobachtete Verhalten haben. Die Einstellung von τ_0 und ggf. τ_i erfolgt gemäß der Heuristik (s. Seite 37). Die Werte für einige n in der Variante $n_\sigma = n$ bzw. $n_\sigma = 1$ zeigt die folgende Tabelle:

	$n_\sigma = n$		$n_\sigma = 1$
n	$\tau_0 = \frac{1}{\sqrt{2 \cdot n}}$	$\tau_i = \frac{1}{\sqrt{2 \cdot \sqrt{n}}}$	$\tau_0 = \frac{1}{\sqrt{n}}, \tau_i = 0$
10	0.2236	0.3976	0.3162
30	0.1291	0.3021	0.1826
50	0.1000	0.2659	0.1414
75	0.0816	0.2403	0.1155
90	0.0745	0.2296	0.1054
100	0.0707	0.2236	0.1000

Tabelle 5.1: Standardeinstellungen für τ_0 und τ_i

Der Rekombinationstyp wird in dieser Arbeit durchgehend dreistellig angegeben, obwohl mit Ausnahme von Abbildung 4.8 keine korrelierten Mutationen benutzt werden, die letzte Stelle also keine Bedeutung hat. Die Kodierung erfolgt gemäß der Aufzählung im vorherigen Kapitel (s. Seite 38). Die Notation $[4,5][6,7]1$ bedeutet z. B. intermediäre Rekombination der Objektvariablen aus 2 bzw. μ Eltern und 1– bzw. 2–Punkt–Crossover der Strategievariablen. Alle erwähnten Standardwerte gelten, sofern bei einzelnen Untersuchungen nicht explizit andere angegeben werden oder diese externen ES–Parameter selbst Untersuchungsgegenstand sind.

Die normierte Konvergenzgeschwindigkeit Kg_1 wird wie folgt definiert:

$$Kg_1 \quad := \quad \frac{n}{Gen.} \cdot \log \sqrt{\frac{f_{Startwert}}{f_{Bestwert}}}$$

Da in dieses Maß die Generationenzahl eingeht, werden bei multimodalen Problemen Strategievarianten bevorzugt, die lokal konvergieren. Aus diesem Grund wird noch ein weiteres Maß Kg_2 für die Konvergenzgeschwindigkeit verwendet, an dem man direkt die Anzahl an zurückgelegten Zehnerpotenzen ablesen kann:

$$Kg_2 \quad := \quad \log \frac{f_{Startwert}}{f_{Bestwert}}$$

Natürlich lassen sich die beiden Geschwindigkeitsmaße ineinander umrechnen:

$$\frac{Kg_1}{Kg_2} = \frac{n \cdot \ln 10}{2 \cdot Gen.} = 1.1513 \cdot \frac{n}{Gen.}$$

Für unimodale Zielfunktionen liefern beide Maße identische Rangfolgen, bei multimodalen Problemen ist das zweite Kriterium vorzuziehen, weil es (auch langsame) globale Konvergenz „belohnt". Da Kg_1 Vergleiche mit der Theorie erlaubt, wird es generell in den Abbildungen dieses Kapitels verwendet, Kg_2 wird in den Tabellen des Anhangs zusätzlich angegeben (s. Anhang A). Da in beiden Maßen der beste während des gesamten Optimierlaufes erreichte Funktionswert benutzt wird, wird die Divergenz von Strategievarianten sich nicht durch negative, sondern nur durch sehr kleine positive Werte verraten, die die kurze Periode des Fortschritts am Anfang der Läufe widerspiegeln. Diese Art der Messung stellt m. E. keine wesentliche Einschränkung dar, da aus den Abbildungen die *Unterschiede* zwischen gut und schlecht bzw. gar nicht konvergierenden Läufen auch so erkennbar sind.

Folgende Zielfunktionen werden in dieser Arbeit verwendet, in Klammern findet man die Numerierung aus (Schwefel, 1977), sofern sie dort vertreten sind:

$$\begin{aligned}
\text{Kugelmodell (1.1)} &: \sum_{i=1}^{n} x_i^2 \\
\text{Skalierungstest, gewichtetes Kugelmodell} &: \sum_{i=1}^{n} \left(i \cdot x_i^2\right) \\
\text{Doppelsumme (1.2)} &: \sum_{i=1}^{n} \left(\sum_{j=1}^{i} x_j\right)^2 \\
\text{(eigene) multimodale Funktion} &: \sum_{i=1}^{n} \left(|x_i|^{0.8} + 5 \cdot \sin(x_i)^3 + 3.5828\right) \\
\text{Griewank} &: 1 + \sum_{i=1}^{n} \left(\frac{x_i^2}{4\,000} - \prod_{i=1}^{n} \cos \frac{x_i}{\sqrt{i}}\right) \\
\text{Rastrigin} &: 3 \cdot n + \sum_{i=1}^{n} \left(x_i^2 - 3 \cdot \cos(2 \cdot \pi \cdot x_i)\right)
\end{aligned}$$

Mit einer Ausnahme befindet sich das globales Optimum jeweils im Koordinatenursprung, bei der (eigenen) multimodalen Funktionen müssen alle x_i^* den Wert -1.5216 annehmen. Der Funktionswert $f(\vec{x}^*)$ ist jeweils Null.

Nicht alle Zielfunktionen spielen schon bei der nun folgenden Untersuchung eine Rolle, weil sie dem Kugelmodell sehr ähnlich (Rastrigin) oder einfach zu optimieren sind (Griewank). Spätestens in Kapitel 6 werden sie jedoch untersucht.

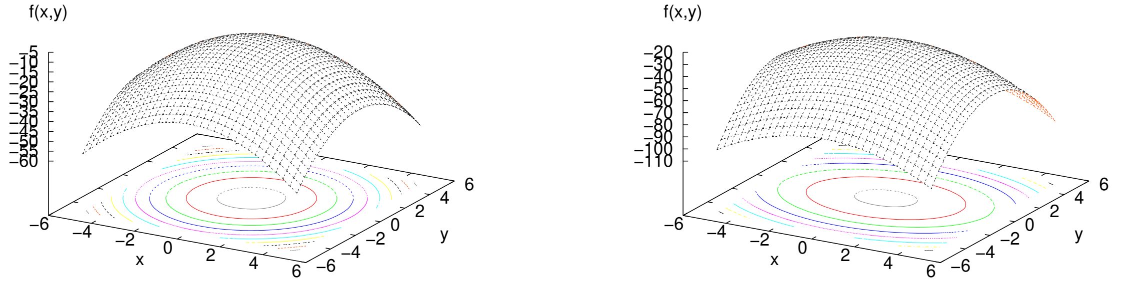


Abbildung 5.1: Kugelmodell (links) und Skalierungstest (rechts)

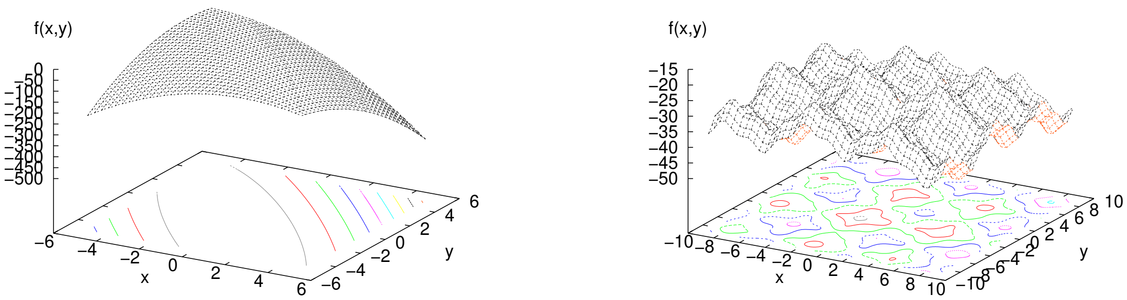


Abbildung 5.2: Doppelsumme (links) und multimodale Funktion (rechts)

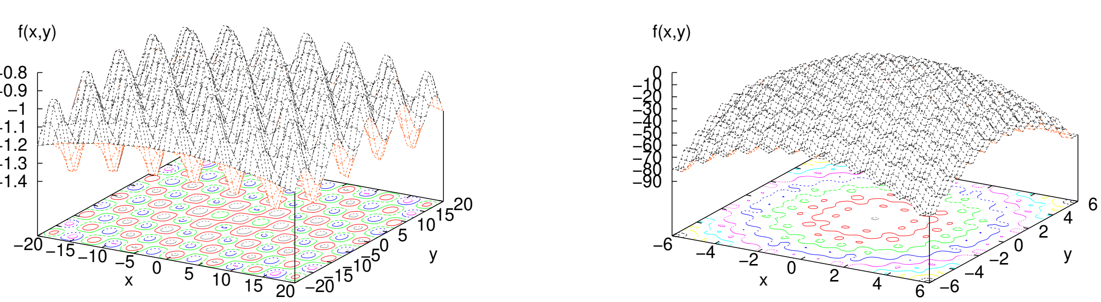


Abbildung 5.3: Griewank– (links) und Rastrigin–Funktion (rechts)

5.2 Untersuchung des Rekombinationstyps

Wie Abbildung 5.4 zeigt, kann man eine Standard–ES nur durch die Wahl eines falschen Rekombinationstyps bei einfachen unimodalen Funktionen und bereits für nicht zu große n zur Divergenz bringen, während die gleiche Strategie, versehen mit einem angemessenen Rekombinationsoperator, gut konvergiert. In diesem Abschnitt soll der Frage nachgegangen werden, welche Rekombinationstypen auf einer bestimmten Topologie besser als andere sind, und ob es nicht vielleicht doch *eine* Einstellung gibt, die zumindest auf allen hier getesteten Zielfunktionen gute Ergebnisse liefert.

Das unerwünschte divergente Verhalten mag im Fall der Doppelsumme noch keine große Überraschung bieten, weil eine diskrete Rekombination der Objektvariablen unangemessen erscheint, will man einem Grat folgen. Aber daß die empfohlene Standardeinstellung für die Rekombination beim 100–dimensionalen Kugelmodell zu Divergenz führt, widerspricht dem Robustheitsanspruch (Bäck, Hoffmeister und Schwefel, 1991; Bäck und Hoffmeister, 1994). Natürlich ist die Einstellung $n_\sigma = n$ im Falle des symmetrischen Kugelmodells unsinnig, für die Doppelsumme vermeidet man für $n = 100$ mit dem Rekombinationstyp 351 und nur einer Schrittweite bei einer $(10\,,\,100)$–ES die Divergenz nur knapp, für $(15\,,\,100)$ aber schon nicht mehr. Diese Einstellungen kann man allerdings nur deshalb als bösartig bezeichnen, weil die beiden Zielfunktionen schon bei genauem Hinsehen ihre Charakteristika preisgeben. Was aber soll man bei echten Optimierproblemen tun, denen man nicht so leicht auf die Schliche kommt? Und was, wenn die Zielfunktion gar nicht in geschlossener Form darstellbar ist, sondern nur in Form eines Simulationsmodells vorliegt?

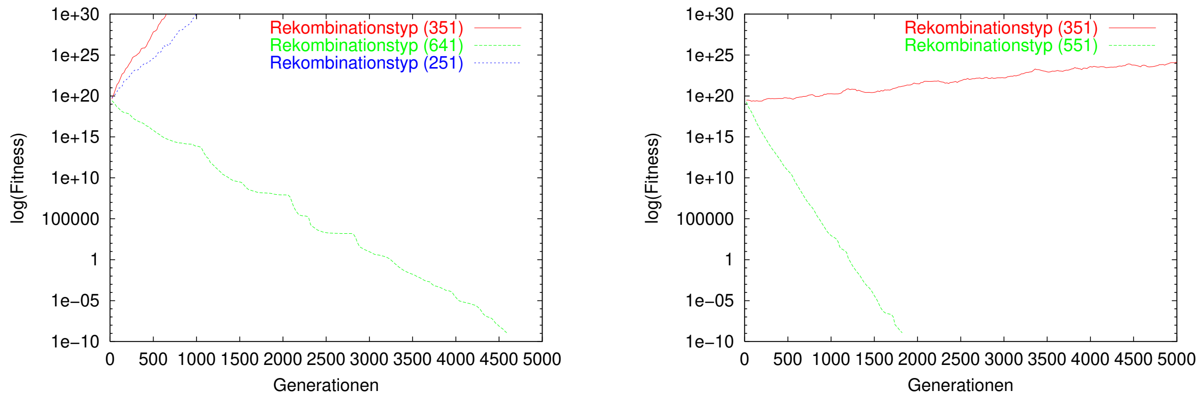


Abbildung 5.4: $(10\,,\,100)$–ES, Doppelsumme, $n = n_\sigma = 30$ (links) und $(15\,,\,100)$–ES, Kugelmodell, $n = n_\sigma = 100$ (rechts), jeweils typische Einzelläufe

Diese ersten Beobachtungen legen bereits die Vermutung nahe, daß die genetischen Operatoren nicht isoliert betrachtet werden dürfen, da sich falsche bzw. nicht optimale Einstellungen eines Parameters manchmal durch andere kompensieren lassen. Divergenz läßt sich natürlich immer durch eine Plus–Strategie verhindern, dies bedeutet aber z. B. bei der Doppelsumme nicht automatisch Konvergenz, und im Falle multimodaler Zielfunktionstopologien verbietet sich laut ES–Literatur diese Selektionsvariante grundsätzlich (Bäck und Schwefel, 1993).

Im folgenden werden alle jeweils möglichen Rekombinationstypen für einige typische Zielfunktionen in der Hoffnung auf generelle Empfehlungen bzw. Warnungen untersucht.

In Abbildung 5.5 erkennt man, daß sich beim Kugelmodell alle ES für $n = 30$ noch so verhalten, wie man es von einem robusten Optimierverfahren erwarten darf: Die nicht–optimale Wahl des Rekombinationstyps wird lediglich mit einer reduzierten Konvergenzgeschwindigkeit bestraft. Die deutlich besten vier Varianten sind die möglichen intermediären $[4,5][4,5]1$, und die langsamste Variante ist jene, die ganz auf die Rekombination verzichtet. Für $n = 100$ führen dagegen alle $[1,2,3,6,7]51$–Rekombinationstypen zur Divergenz, sie sind noch schlechter als ES ohne Rekombination. Die Einstellungen $[4,5]51$ sorgen dagegen für den größten Fortschritt. In keinem Fall lohnt die Benutzung der neuen, von ihrer Natur her eher diskreten Crossover–Varianten.

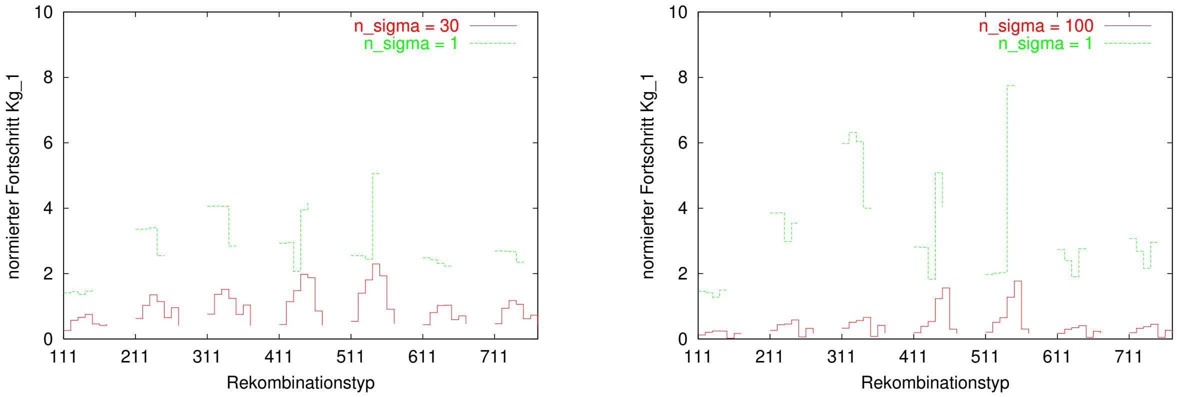


Abbildung 5.5: (15 , 100)–ES, Kugelmodell, alle Rekombinationsmöglichkeiten, $n = 30$ (links), $n = 100$ (rechts), $n_\sigma = \{n, 1\}$

Das Versagen der diskreten Rekombination auf den Objektvariablen in Kombination mit einer intermediären Schrittweitenvererbung liegt an der Entkopplung von Objekt– und Strategievariablen. Durch die diskrete Rekombination können gute Punkte im Suchraum entstehen, gleichzeitig wird aber die zugehörige Schrittweiteninformation durch Mittelung „verschmiert“, sie wird zu groß für weitere Fortschritte. Dieses Phänomen tritt genauso bei kleinen Startschrittweiten auf, nur entsprechend später. Statt die Schrittweiten der aktuellen Position im Suchraum anzupassen, versucht die Strategie erfolglos genau das Umgekehrte, nämlich die Lage den Schrittweiten anzupassen. Mit steigender Zahl der Dimensionen und wachsender Mutationsschrittweite wird die Erfolgswahrscheinlichkeit beliebig klein (Rechenberg, 1994), so daß das Evolutionsfenster derjenigen Schrittweiten, die nicht zu klein und nicht zu groß für weiteren Fortschritt sind, nicht wiedergefunden wird, wenn es einmal verlassen wurde.

Bei der Vorstellung der mehrgliedrigen ES (s. Seite 37) wurden die externen Parameter τ_0 und τ_i als Lernraten der Schrittweitenmutation bezeichnet. Werden ihre Werte gemäß der üblichen Heuristik eingestellt, reicht die Mutabilität der Schrittweiten offensichtlich nicht. Überprüft man diese Hypothese, indem man den schlechtesten Lauf für $n = 100$ (151) z. B. mit den τ–Werten für $n = 30$ versieht, sieht man Fortschrittsgeschwindigkeiten, die gleich hinter den besten vier Varianten aus Abbildung 5.5 einzuordnen sind. Verdoppelt man aber einfach die von der Heuristik vorgegebenen τ–Werte, erhält man wieder Divergenz. Interdependenzen der externen Parameter gibt es also nicht nur zwischen Rekombinationstyp und Selektionsdruck (s. Seite 44), sondern auch zwischen Rekombinationstyp und Mutabilität der Schrittweiten sowie auch innerhalb des Rekombinationsoperators.

Benutzt man nur eine Schrittweite, wie es der Symmetrie des Kugelmodells angemessen ist, konvergieren alle Rekombinationstypen robust, wobei es aber natürlich Geschwindigkeitsunterschiede bis zum Faktor 6 (3.5) zwischen der jeweils schlechtesten Variante (131) und den besten (5[4, 5]1) für $n = 100\,(30)$ gibt.

Bei der Skalierungstestfunktion reicht eine Schrittweite nicht aus, sondern es werden immer n Schrittweiten benötigt, um die richtige Skalierung $c/\sqrt{i}$ der Schrittweiten für optimalen Fortschritt erlernen zu können. Wie aus Abbildung 5.6 ersichtlich ist, funktioniert das Erlernen einer Skalierung der σ_i zueinander wirklich, wenn man n Schrittweiten benutzt.

Wie Abbildung 5.7 allerdings zeigt, kann man durch Wahl eines schlechten Rekombinationstyps bei n Schrittweiten ebensoviel an Konvergenzgeschwindigkeit einbüßen wie durch die bei dieser Funktion unsinnige Entscheidung, mit nur einer Schrittweite zu arbeiten. Die am besten konvergierenden Rekombinationstypen sind für $n = 30$ wie beim Kugelmodell [4, 5][4, 5]1. Für $n = 100$ divergieren erneut [1, 2, 3, 6, 7]51, während [4, 5]51 die beste Konvergenz liefern. Auch bei dieser Funktion läßt sich die schlechteste Variante für $n = 100$ (151) mit den τ–Werten für $n = 30$ „retten“, während die Verdopplung auch bei dieser Funktion Divergenz zur Folge hat.

Benutzt man beim Skalierungstest nur eine Schrittweite, vermeidet man zwar immer die Divergenz, erhält aber in keinem Fall gut konvergierende Strategien.

Eine ES–Variante, die die Schrittweitenzahl während des Laufes variieren kann, liegt wie erwartet in der Konvergenzgeschwindigkeit zwischen bester ($n_\sigma = 1$) und schlechtester ($n_\sigma = n$) Variante für das Kugelmodell, löst aber das Problem der Divergenz bei dieser Zielfunktion ebensowenig wie beim Skalierungstest. Dafür konvergiert die neue Variante aber bei einigen Rekombinationstypen ([4, 5][4, 5]1) in beiden Topologien schneller als die

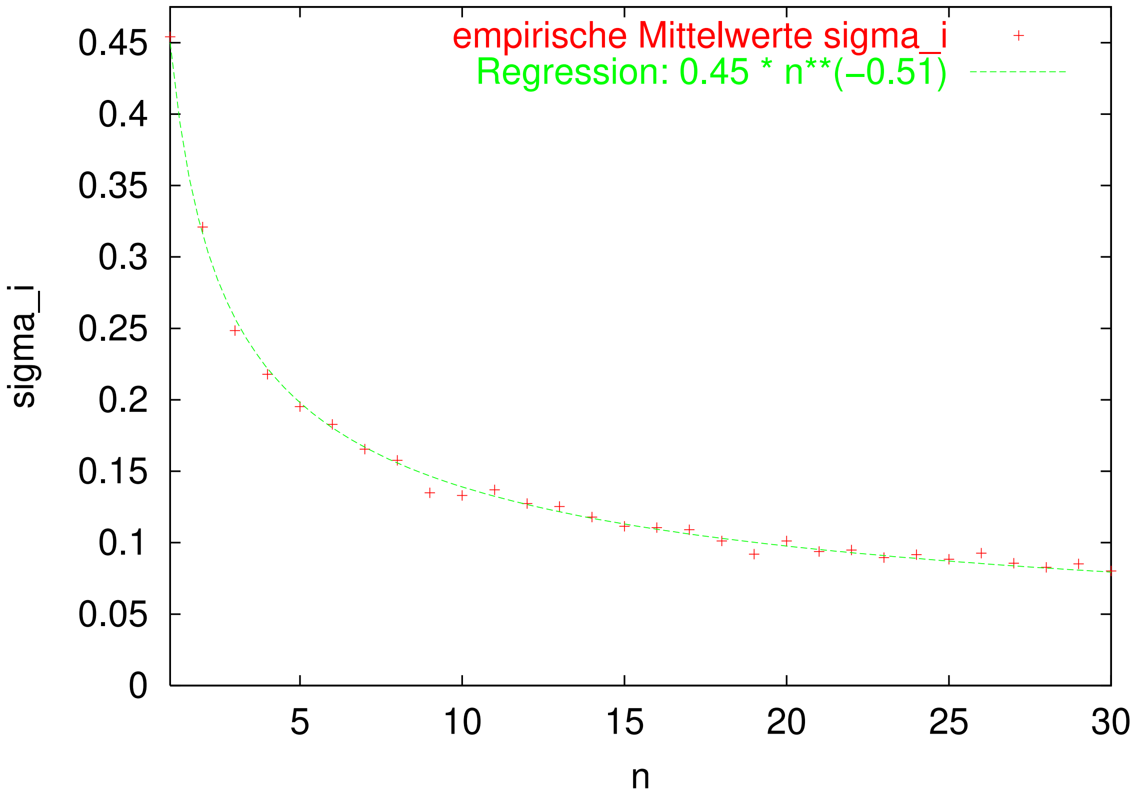


Abbildung 5.6: (15 , 100)–ES, Skalierungstest, $n = 30$, $n_\sigma = n$, Mittelung über 100 Läufe

Standard–ES mit fester Schrittweitenzahl 1 bzw. n (Kursawe, 1995) — ein weiteres Indiz dafür, daß die isolierte Betrachtung einzelner Parameter nicht ausreicht.

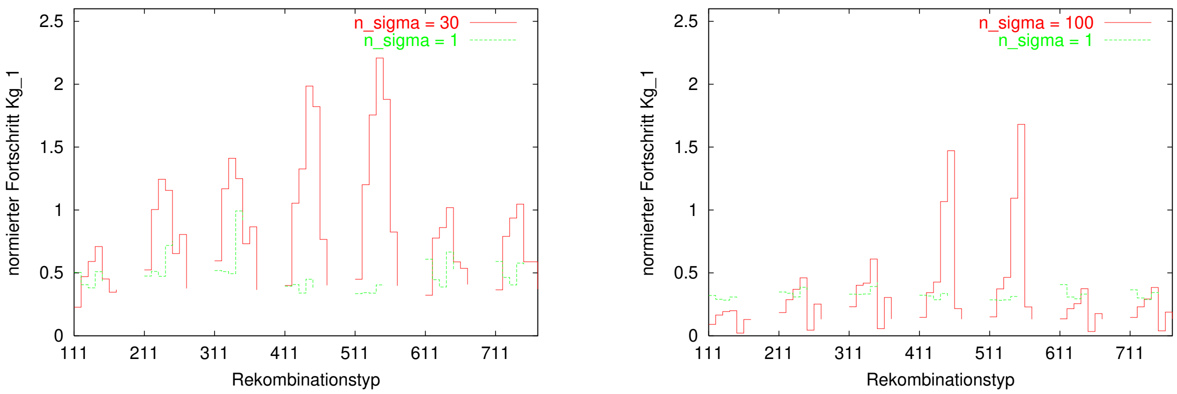


Abbildung 5.7: (15 , 100)–ES, Skalierungstest, alle Rekombinationsmöglichkeiten, $n = 30$ (links), $n = 100$ (rechts), $n_\sigma = \{n, 1\}$

Die Doppelsumme belohnt kurzfristig schnellere Individuen auf zweifache Art: Zusätzlich zu der sich bei allen Problemen temporär auszahlenden Reduktion des Suchraumes auf Unterräume werden hier neben dem Grat liegende Individuen belohnt, die nur auf diesen zurückstreben, sich aber dabei ggf. sogar vom Optimum entfernen. Lokaler und globaler Gradient fallen nicht mehr wie beim Kugelmodell zusammen.

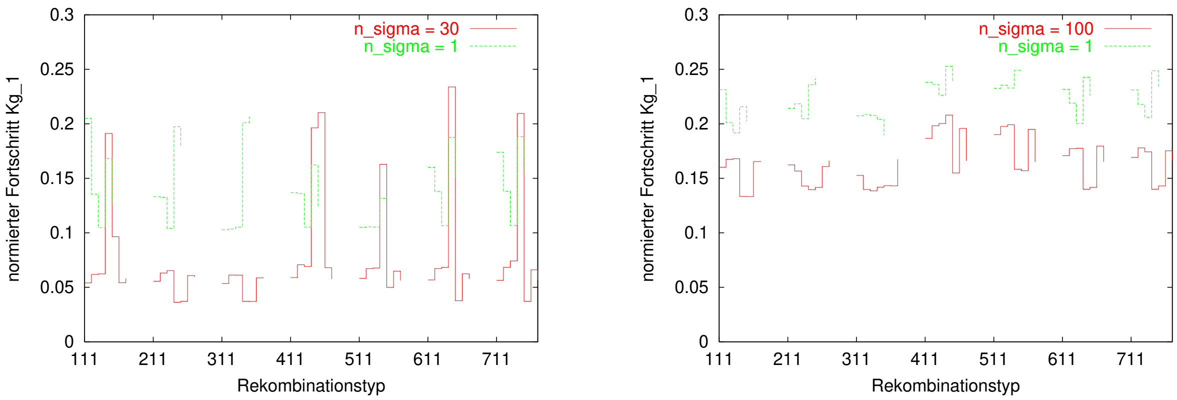
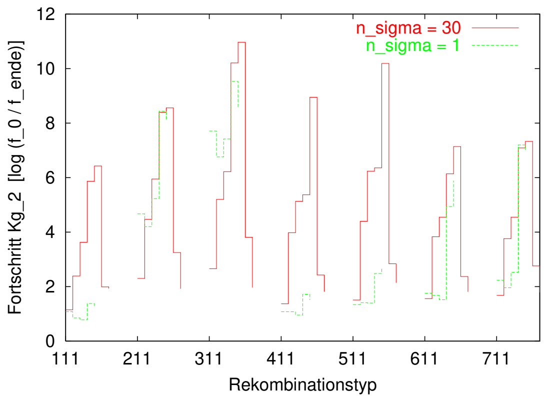


Abbildung 5.8: (15 , 100)–ES, Doppelsumme, alle Rekombinationsmöglichkeiten, $n = 30$ (links), $n = 100$ (rechts), $n_\sigma = \{n, 1\}$

Bei der Doppelsumme ergibt die Untersuchung der Rekombination (s. Abbildung 5.8) ein uneinheitliches Bild. Weder kann man für $n = 30$ eindeutig sagen, ob eine oder n Schrittweiten benutzt werden sollen, noch sind die besten Rekombinationstypen identisch mit denen von Kugelmodell und Skalierungstest. Auffallend ist ferner, daß bei Benutzung nur einer Schrittweite eine ES ohne Rekombination die zweitbeste Variante hinter Rekombinationstyp 351 darstellt. Bei $n_\sigma = 30$ Schrittweiten führen die Rekombinationsmöglichkeiten $[2, 3][4, 5]1$ zur Divergenz (s. auch Abbildung 4.8), weil eine diskrete Rekombination der Objekt– in Kombination mit einer intermediären der Strategievariablen die Entwicklung dieser Variablen voneinander entkoppelt. Diese These wird unterstützt durch den Erfolg der Varianten, die Crossover ($[6, 7]41$) benutzen, da die beiden beteiligten Elternindividuen dieselben sind, somit also eine Kopplung zwischen Objekt– und Strategievariablen hergestellt wird, bei der die Schrittweitenmittelung nichts mehr zerstören kann. Durchbricht man diese Kopplung bewußt durch eine Schrittweitenmittelung jeweils aller Eltern ($[6, 7]51$), beobachtet man Divergenz.

Für $n = 100$ ist das Bild eindeutiger: Eine Schrittweite ist n Schrittweiten vorzuziehen, und in beiden Fällen kommt die Rekombinationsvariante 441 am schnellsten voran. Auch hier wirkt erneut die Kopplung zwischen Objekt– und Strategievariablen. Zusätzlich verhindert eine intermediäre Schrittweitenrekombination die Reduktion des Suchraumes auf Unterräume, die sich kurzfristig immer durch eine schnellere Konvergenz auszahlt. Intermediäre Rekombinanten wirken dieser Gefahr insofern besser entgegen, als sie kleine Schrittweiten durch die Mittelung vergrößern können, während im diskreten Fall nur verschiedene Unterräume ausgetauscht werden. Auch bei dieser Funktion kann man die positive Wirkung der passenden Schrittweitenrekombination durch eine schlechte Einstellung der Objektvariablenrekombination wieder zunichte machen.

In Abbildung 5.9 wird das Konvergenzmaß Kg_2 verwendet, weil der normierte Fortschritt Kg_1 schnelle und daher nur lokal konvergierende Strategien hervorhebt, was bei einer multimodalen Funktion i. d. R. nicht erwünscht ist.



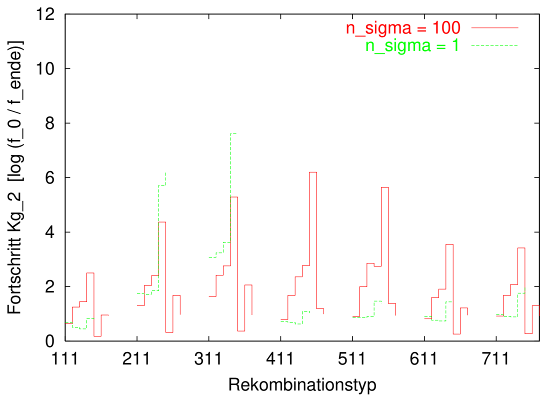


Abbildung 5.9: (15 , 100)–ES, multimodale Funktion, alle Rekombinationsmöglichkeiten, $n = 30$ (links), $n = 100$ (rechts), $n_\sigma = \{n, 1\}$

Für $n = 30$ stellt sich heraus, daß eine intermediäre Rekombination der Schrittweiten vorteilhaft ist, und in Kombination mit einer diskreten Rekombination der Objektvariablen hat man die größte Chance, das globale Optimum zu finden (3[4, 5]1). Realisiert wird also einerseits der Austausch guter Punkte im Suchraum bei gleichzeitiger Diversitätserhaltung der Schrittweiten, sowohl bei einer als auch bei n Schrittweiten. Aufgrund der großen Unterschiede in der Anzahl der Generationen (300 zu 1 800), erscheint es unter Effizienzaspekten sinnvoller, mit nur einer Schrittweite zu arbeiten, die ES bis in die Nähe des globalen Optimums konvergieren zu lassen und mit einer größeren Schrittweite die Optimierung weiterlaufen zu lassen. Ohne einen solchen Eingriff findet die beste Strategie bei n Schrittweiten das globale Optimum in 66 von 100 Läufen, bei einer Schrittweite dagegen nur in 20 Fällen.

Die Ergebnisse für $n = 100$ unterstützen die Idee, mit nur einer Schrittweite zu arbeiten (750 zu 3 250 Generationen), wobei allerdings zu beachten ist, daß keiner der $(49 + 35) \cdot 25 = 2\,100$ Läufe das globale Optimum gefunden hat, alle konvergierten also vorzeitig. Ferner zeigen erneut alle $[1, 2, 3, 6, 7]51$–Rekombinationsvarianten Divergenz. Die Varianten $[4, 5]51$ kommen dem globalen Optimum bei n Schrittweiten am nächsten, bei einer Schrittweite sind es erneut die Rekombinationstypen $[2, 3][4, 5]1$, die auch im Mittel eine bis zwei Zehnerpotenzen weiter kommen.

Die Untersuchung der Griewank–Funktion bestätigte dieses Ergebnis. Getestet wurde hier aber nur der $n = 100$–dimensionale Fall mit einer Schrittweite. Die besten Rekombinationstypen $[2, 3][4, 5]1$ fanden das globale Optimum in allen Läufen, wobei die global diskrete Rekombination der Objektvariablen im Mittel 100 Generationen weniger benötigt als jene, die nur zwei Eltern zur Rekombination der x_i heranzieht.

5.3 Untersuchung von τ_0 und τ_i

In diesem Abschnitt wird die übliche Heuristik für τ_0 und τ_i untersucht (Schwefel, 1977). Die im folgenden dargestellten Fortschrittsgebirge beschränken sich auf einen Definitionsbereich von $[0, 1.2]$ für die beiden variierten Parameter, da zuvor durch Tests mit größeren rechten Intervallgrenzen sichergestellt wurde, daß jenseits dieses Intervalls keine Überraschungen mehr warten (s. Anhang A, s. Abbildung A.6). Durch weitere Tests wurde festgestellt, daß die τ_0–τ_i–Gebirge ihre Struktur bei den unimodalen Zielfunktionen unter verändertem Selektionsdruck beibehalten, obwohl sich natürlich die Konvergenzgeschwindigkeiten verändern. Um den Einfluß beider τ–Parameter testen zu können, gilt natürlich stets $n_\sigma = n$. Benutzt wurde immer der vorher ermittelte beste Rekombinationstyp für die jeweilige Zielfunktion und das untersuchte n, sofern nicht explizit ein anderer angegeben wird. Die Rastergröße beträgt 0.025 für $0 <= \tau_0, \tau_i <= 0.75$ und 0.05 für $0.75 < \tau_0, \tau_i <= 1.2$. Die anderen Parameter sind wie in Abschnitt 5.1 eingestellt.

In einer Einzeldarstellung kann man zwar leichter Details erkennen, weil das Gebirge komplett sichtbar ist, will man aber Unterschiede zwischen den Gebirgen für verschiedene n oder den Einfluß der Rekombination erkennen, wird das mit einer über mehrere Seiten verteilten Einzeldarstellung mühsam. In diesem Kapitel befinden sich daher nur zusammenfassende Darstellungen der Fortschrittsgebirge, die Einzeldarstellung befindet sich im Anhang (s. Anhang A).

Beim Kugelmodell stellt sich in allen Tests ($n = 10,\ 30,\ 50,\ 100$) heraus, daß der Parameter für die individuelle Mutabilität τ_i bei den besten Fortschrittswerten so gut wie ausgeschaltet wird, so daß trotz der $n_\sigma = n$ Schrittweiten praktisch mit nur einer Gesamtschrittweite[1] gearbeitet wird, deren Mutabilität allein τ_0 steuert. Dieses Ergebnis überrascht nicht, da

[1] Von einer Gesamtschrittweite trotz $n_\sigma = n$ zu sprechen, ist natürlich nicht ganz richtig. Wie man anhand der Skalierungstestfunktion am besten sehen kann, gibt es einen Unterschied zwischen ES mit $n_\sigma = 1$ und ES mit $n_\sigma = n$ sowie $\tau_i = 0$. Allerdings reicht die kombinierte Wirkung von τ_0 und der (Schrittweiten–)Rekombination nicht, die nötige Skalierung der Schrittweiten zu leisten (quantitativ für $n = 100$ nach 1 500 Generationen: eine Standard–ES überwindet 22 Zehnerpotenzen, mit der Parametrisierung aus Kapitel 6 sind sogar 30 möglich, eine Standard–ES mit $n_\sigma = 1$ schafft 4, jene mit $n_\sigma = n$ und $\tau_i = 0$ nur 6 Zehnerpotenzen). Deshalb wird hier und im weiteren Verlauf der Arbeit verkürzend von einer Gesamtschrittweite gesprochen.

aufgrund der Symmetrie des Kugelmodells eine Schrittweite ausreicht. Mit zunehmendem n wird der „Gipfel“, bei dem der Einfluß von τ_i praktisch ausgeschaltet ist, immer schmaler, so daß sich die Gebirge für Kugelmodell und Skalierungstest nur noch an genau dieser Gipfelstelle unterscheiden. Abbildung 5.10 zeigt die resultierenden Fortschrittsdaten für drei verschiedene n, auf die zusätzliche Darstellung der Fortschrittswerte für $n = 30$ wurde aus Gründen der Übersichtlichkeit an dieser Stelle verzichtet.

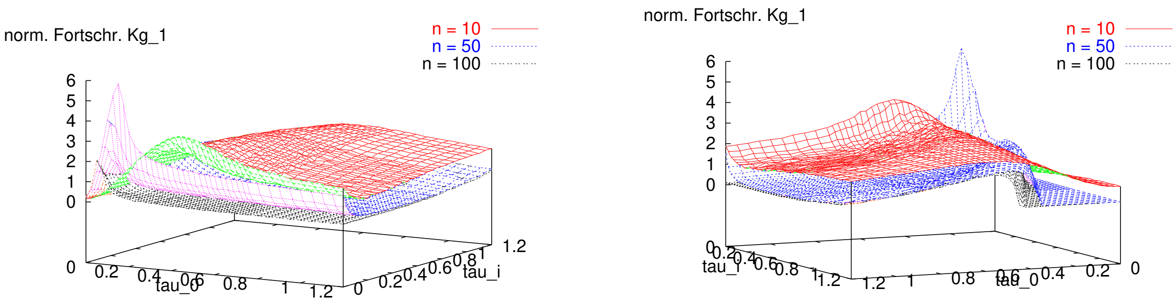


Abbildung 5.10: (15 , 100)–ES, Kugelmodell, $n = 10,\ 50,\ 100$, $n_\sigma = n$

Abbildung 5.11 veranschaulicht den geschwindigkeitssteigernden Effekt der (richtig parametrisierten) Rekombination. Ohne Rekombination ist das τ_0–τ_i–Gebirge in weiten Teilen eine Ebene, kleinen Fortschritt gibt es nur für die beiden kleinsten untersuchten τ_i–Werte (0.0 und 0.025), die die Möglichkeit einer individuellen Schrittweitenentwicklung praktisch ausschalten bzw. extrem dämpfen. Erneut sieht man, daß die genetischen Operatoren nur in Kombination ihre Wirkung voll entfalten.

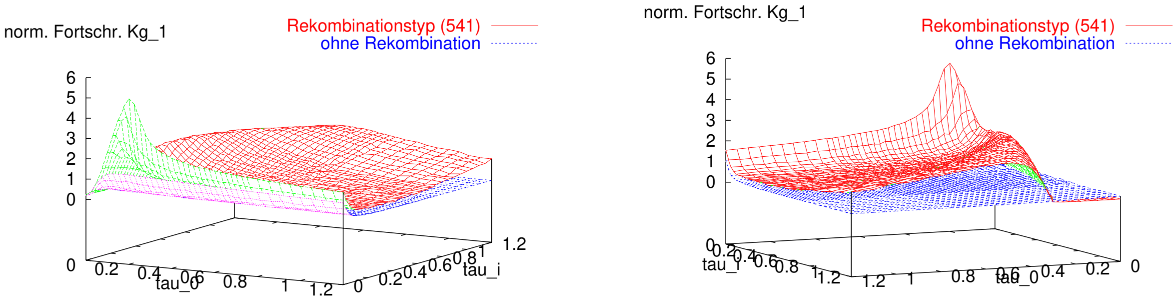


Abbildung 5.11: (15 , 100)–ES, Kugelmodell, $n = 30$, $n_\sigma = n$, mit und ohne Rekombination

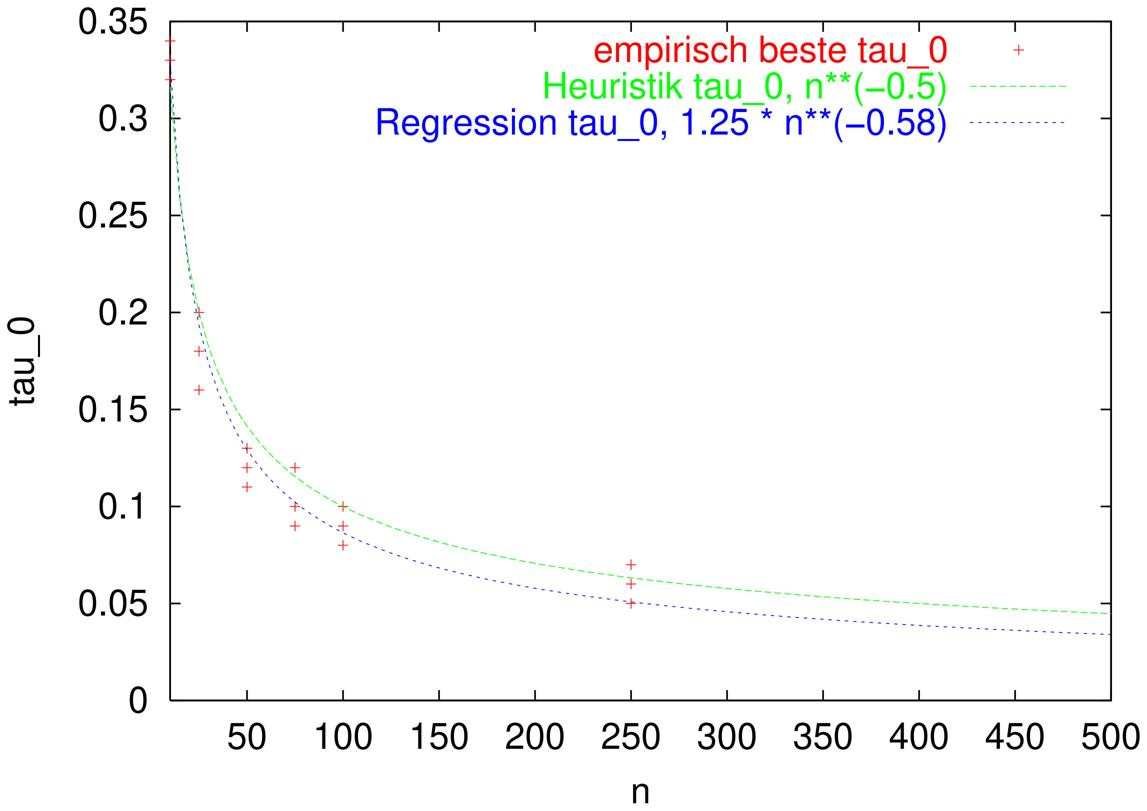


Abbildung 5.12: (15 , 100)–ES, Kugelmodell, $n = 10, 25, 50, 75, 100, 250$, $n_\sigma = 1$, Vergleich von empfohlenen und gemessenen besten 3 Werten für τ_0

Da beim Kugelmodell Evolutionsstrategien mit nur einer Schrittweite am schnellsten konvergieren, läßt sich an ihm die Heuristik für τ_0 überprüfen (s. Seite 37). Abbildung 5.12 veranschaulicht deren Güte. Die Abweichungen zwischen Heuristik und Regressionskurve sind vernachlässigbar.

Die interessantere Zielfunktion für die Untersuchung des Zusammenspiels beider τ–Parameter ist jedoch die Skalierungstestfunktion. Auch sie verhält sich insofern erwartungskonform, als die am schnellsten konvergierenden Läufe wirklich beide Lernraten benutzen, um die Skalierung der Schrittweiten untereinander zu erlernen (s. Abbildung 5.13). Der für $n = 10$ noch relativ breite „Gebirgsrücken“ mit gutem Fortschritt wird mit zunehmendem n schmaler und entwickelt eine steile Flanke für zu große τ_i. Bei $n = 100$, $\tau_0 = 0.25$, $\tau_i = 0.75$ beobachtet man noch fast optimalen Fortschritt, erhöht man τ_i auf 0.85, divergiert die ES. Dies passiert mit der erwähnten Parametersetzung auch, wenn man mit erheblich kleineren Schrittweiten und / oder zufälligen x_i startet, nur setzt dann die Divergenz etwas später ein.

Ohne Rekombination erhält man wie beim Kugelmodell fast eine Ebene, wobei die besten Fortschrittswerte erneut bei praktisch abgeschaltetem τ_i erzielt werden: Da der Rekombinationsoperator fehlt, um die Schrittweiten passend zueinander einzustellen, bleibt der Strategie nur die Beschränkung auf das Machbare, nämlich das Anpassen einer Gesamtschrittweite. In Abbildung 5.14 sind zusätzlich zum besten (541) und schlechtesten (111) Rekombinationstyp für diese Funktion mit $n = 30$ noch zwei weitere (351 und 141) untersucht worden, um nicht nur die Extreme darzustellen und um auch einmal zu untersuchen, was eine Rekombination, die nur auf den Schrittweiten arbeitet, zu leisten imstande ist.

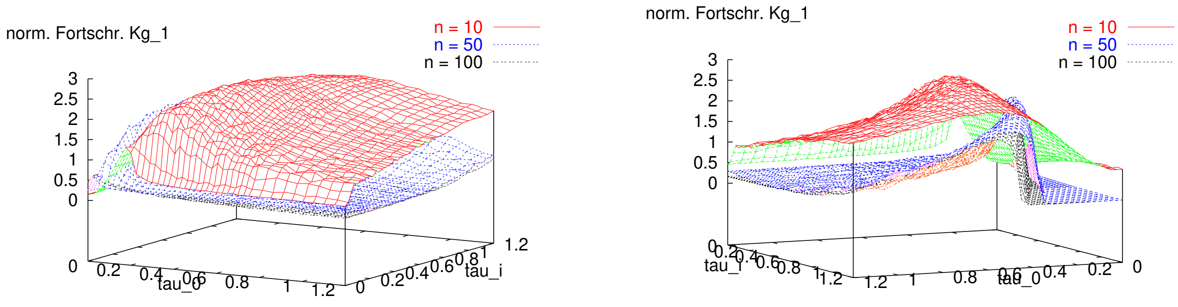


Abbildung 5.13: (15 , 100)–ES, Skalierungstest, $n = 10, 50, 100$, $n_\sigma = n$

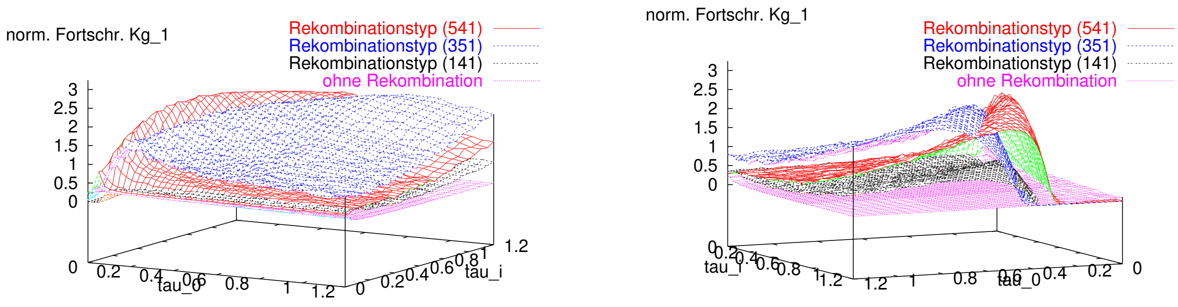


Abbildung 5.14: (15 , 100)–ES, Skalierungstest, $n = 30$, $n_\sigma = n$, Rekombinationstypen 541, 351, 141 und 111

Weitere Untersuchungen ergaben, daß die τ_0–τ_i–Gebirge sich nicht dramatisch verändern, falls man mit einer Plus–Strategie oder, bezogen auf den Startpunkt, sehr kleinen Schrittweiten startet. Dabei ist bemerkenswert, daß diese beiden Gebirge sich fast nicht voneinander unterscheiden, sieht man einmal davon ab, daß die Plus–Selektion die oben erwähnte

Flanke etwas flacher gestaltet und natürlich die Divergenz überall vermeidet. Die Einstellung zu kleiner Startschrittweiten verändert das Gebirge an den Punkten mit kleinen τ_0 und τ_i. Bevor der Anstieg auf den Grat der guten Lernraten einsetzt, gibt es einen kleinen Bereich, in dem der Fortschritt nahe bei Null bleibt. Das aus der Plus–Variante resultierende Gebirge liegt wie erwartet an den Stellen mit den guten Fortschrittswerten unterhalb des Gebirges der Komma–ES (in Abbildung 5.15 nicht dargestellt, dafür einzeln im Vergleich mit einer Standard–ES im Anhang A zu finden), die durch das Verwerfen der Eltern nach nur einem Reproduktionszyklus eine bessere Schrittweitenadaptation leistet.

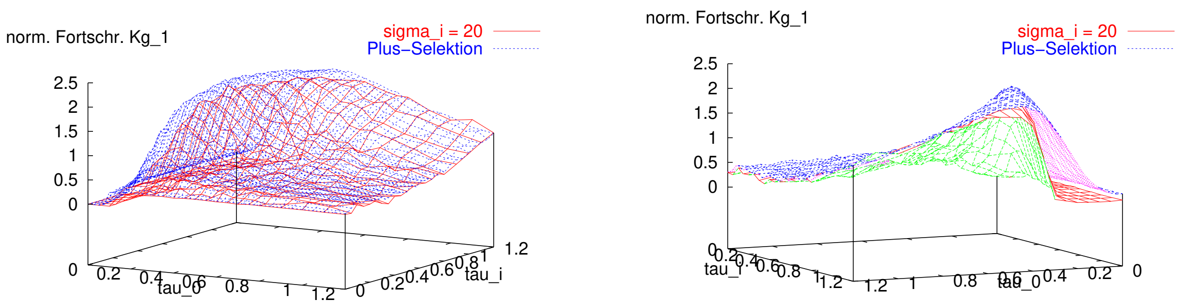


Abbildung 5.15: $(15\,,\,100)$–ES mit $\sigma_i^{Start} = 20$ und $(15+100)$–ES, Skalierungstest, $n = 30$, $n_\sigma = n$

Beide neuen Gebirge bieten also keine Überraschungen. Im Hinblick darauf, daß bei der Meta–ES auch der ES–Selektionsoperator zu den variablen Parametern zählt (s. Kapitel 6), ist es aber beruhigend, daß die guten τ–Werte relativ unabhängig von der Art der Selektion und der Größe der Startschrittweite(n) sind. Bei allen Untersuchungen des Skalierungstests wurde die Empfehlung $\tau_i > \tau_0$ bestätigt, sofern nicht auf die Rekombination verzichtet wurde.

In Abbildung 5.16 sieht man, daß die Heuristik für τ_i nicht mehr angemessen ist. Für den Parameter τ_0 stimmt die Charakteristik von heuristischer und empirischer Kurve noch überein, für τ_i kann man das allerdings nicht mehr sagen. Die Abweichung zwischen empfohlenen und gemessenen τ–Werten führt bei $n = 100\,(30)$ aber nur zu einer Beschleunigung um etwa 20% (8%), d. h. die zu kleinen Werte aus der Heuristik haben keinen dominierenden Einfluß, sofern man mit halbwegs passenden Schrittweiten startet. Der Geschwindigkeitsunterschied läßt sich natürlich bei unimodalen Funktionen durch zu kleine Startschrittweiten fast beliebig vergrößern. Allgemein gilt: Man sollte lieber mit großen Schrittweiten und kleinen τ–Werten starten als umgekehrt, um eine Reduzierung auf Unterräume bzw. die Stagnation in einem lokalen Optimum zu vermeiden. So gesehen, bewegt man sich auf der sicheren Seite, wenn man τ_0 und τ_i gemäß Heuristik einstellt, sollte sich aber bewußt machen, daß man ggf. auf Konvergenzgeschwindigkeit verzichtet. Im

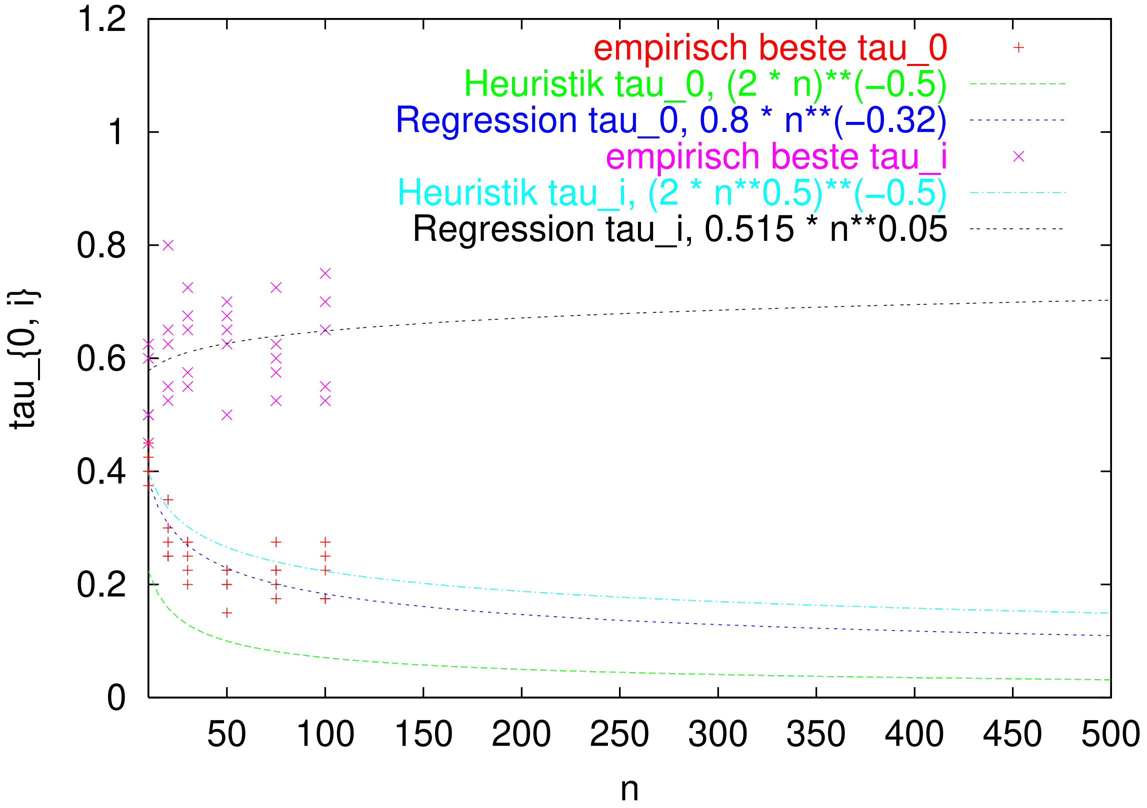


Abbildung 5.16: (15 , 100)–ES, Skalierungstest, $n = 10, 20, 30, 50, 75, 100$, $n_\sigma = n$, Vergleich der empfohlenen mit den gemessenen 4 fortschrittsbesten Werten für τ_0 und τ_i

Anhang findet man dazu für die jeweilige Zielfunktion und die untersuchten n quantitative Aussagen (s. Kapitel A).

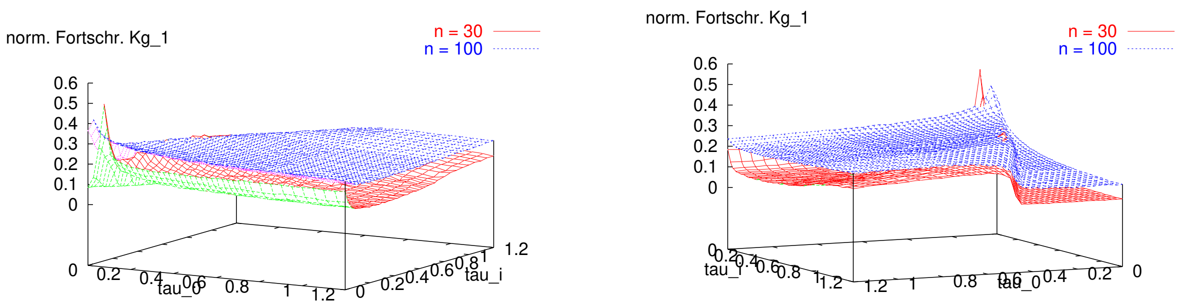


Abbildung 5.17: (15 , 100)–ES, Doppelsumme, $n = 30, 100$, $n_\sigma = n$

Bei der Doppelsumme wird wie beim Kugelmodell die Einzelanpassung der Schrittweiten bei den guten Läufen mittels $\tau_i = 0.0$ bzw. $\tau_i = 0.025$ ausgeschaltet. Das Fortschrittsgebirge für $n = 100$ liegt fast überall oberhalb von dem für $n = 30$, weil in beiden Untersuchungen die Messung über 2 000 Generationen lief und Kg_1 aufgetragen wurde. Beim 30–dimensionalen Problem reicht diese Generationenzahl auch aus, um im besten Fall 29 Zehnerpotenzen zu überwinden, bei 100 Dimensionen ist dafür je nach Selektionsdruck mit Standard–ES mindestens die zehnfache Anzahl nötig. An der Struktur des Gebirges verändert eine größere Generationenzahl allerdings nichts, weil einer ES 2 000 Generationen reichen, um sich auf die Topologie dieser Funktion derart einzustellen, daß die Konvergenz bis zum globalen Optimum lediglich eine Geduldsfrage ist.

Abbildung 5.18 demonstriert für $n = 30$ und $n = 100$, daß die Griewank–Funktion, obwohl multimodal, den bereits bekannten Gebirgen für Kugelmodell und Skalierungstest sehr ähnliche liefert. Daher wird hier bei der Darstellung auch weiter Kg_1 verwendet, für eine Darstellung mit Kg_2 sei auf Abbildung 5.19 verwiesen. Die Ergebnisse für $n = 30$ deuten noch darauf hin, daß man bei dieser Funktion mit $n_\sigma = n$ arbeiten sollte, die guten τ_0–τ_i–Kombinationen sind praktisch identisch mit denen des Skalierungstests. Für $n = 100$ fand die beste τ_0–τ_i–Kombination (0.05 und 0.0) in allen Läufen das globale Optimum, und diese Werte sind identisch mit denen beim Kugelmodell gleicher Dimension. Weitere gute τ_0–τ_i–Kombinationen liegen in der Nähe jener guten Kombinationen bei Kugelmodell und Skalierungstest, die alle Schrittweiten benutzen ($\tau_0 \sim 0.2$, $\tau_i \sim 0.6$).

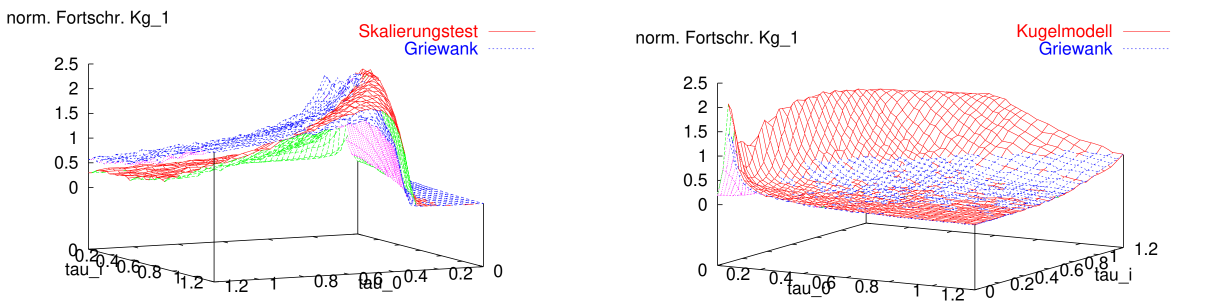


Abbildung 5.18: (15 , 100)–ES, $n_\sigma = n$, Vergleich von Griewank mit Skalierungstest, $n = 30$ (links) und Kugelmodell, $n = 100$ (rechts)

Bei der Untersuchung der multimodalen Funktion wiesen erstmals ungewöhnliche τ_0–τ_i–Kombinationen, d. h. solche, die nicht mehr der Empfehlung $\tau_0 < \tau_i$ folgen, gute Konvergenz auf. Der Grat guter τ–Werte ist deutlich schroffer und schmaler als bei den anderen Funktionen, weil das globale Optimum nicht immer gefunden wird. Auch hier gibt es einen kleinen Bereich relativ sicherer Konvergenz, in dem praktisch nur eine Gesamtschrittweite benutzt wird. Noch sicherer konvergieren allerdings Strategien mit n Schrittweiten, nur müssen τ_0 und τ_i dann sehr genau eingestellt werden.

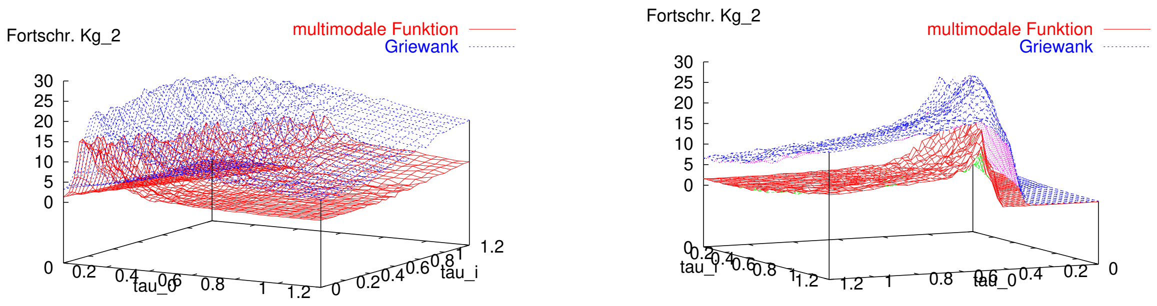


Abbildung 5.19: (15 , 100)–ES, $n = 30$, $n_\sigma = n$, Vergleich von multimodaler Funktion und Griewank

5.4 Schlußfolgerungen

Die in älteren Artikeln ausgesprochene Empfehlung, Objektvariablen diskret und Strategievariablen intermediär zu rekombinieren, kommt zwar der Intuition entgegen — gute Positionen im Suchraum werden zu noch besseren „zusammengewürfelt“, das in den Schrittweiten angesammelte Wissen über das Problem wird als Kompromiß vererbt —, kann aber nicht länger aufrechterhalten werden. Ignoriert man die schlechte Konvergenz bei der Doppelsumme mit $n = n_\sigma = 100$, macht man mit den intermediären Rekombinationsparametern $[4, 5]51$ zumindest nichts falsch. Die Empfehlung nach der Untersuchung aller Rekombinationstypen sollte also „intermediär“ lauten, und je größer die Problemdimension n ist, desto eindringlicher sollte sie ausgesprochen werden. In der Tendenz befindet sich dieses Ergebnis auch in Übereinstimmung mit der Theorie über die globale, schwerpunktbildende Rekombination nach Rechenberg und Beyer. Bei einigen Zielfunktionen und Suchräumen kleinerer Dimension ($n \leq 30$) sollte man aber darauf eingestellt sein, daß einige der diskreten Rekombinationsvarianten besser funktionieren können. Wirklich eindeutig sind lediglich die Resultate für die nicht zu einer Standard–ES gehörenden Crossover–Rekombinationstypen. Ihr Einsatz ragt nur in genau einem Test heraus, und selbst da ist die erzielte Geschwindigkeit nicht viel besser als die beste, die mit den Standard–Einstellungen erzielbar ist. Ferner hängt der mit einem Rekombinationsoperator erzielbare Fortschritt vom Selektionsdruck ab (s. Seite 44 sowie Seite 53). Generell liegen die Varianten mit Beteiligung des 1– oder 2–Punkt–Crossovers eher im hinteren Mittelfeld.

Aus den dargestellten τ_0–τ_i–Fortschrittsgebirgen kann man zwar leicht die Bedeutung der Rekombination ablesen, aber in einigen Fällen ist die Strategievariante ohne Rekombination nicht, wie eigentlich erwartet, die schlechteste: Durch eine isolierte Schrittweitenrekombination wird die Divergenz noch angefacht, es sei denn, man benutzt bei τ_0 und τ_i andere als die vorgeschlagenen Werte, d. h. man erhöht die Mutabilität der Schrittweiten. Auch bei der Untersuchung dieser beiden Lernraten τ_0 und τ_i findet mit zunehmendem n eine Veränderung statt: Der Grat, auf dem guter Fortschritt erzielt wird, wird immer schmaler. Die Heuristik für τ_0 steht nicht im Widerspruch zu den Messungen, bei τ_i scheint ein ganz anderes Gesetz als das in der Heuristik unterstellte zu gelten. So liefert die Funktion $\ln(n \cdot 10^6) \cdot 0.035$ einen ähnlichen Verlauf wie die in Abbildung 5.16 für τ_i eingezeichnete. Sieht man von den Gipfeln ab, die durch $\tau_i = 0$ zustande kommen, bei denen also praktisch nur mit einer Gesamtschrittweite gearbeitet wird, werden sich die Fortschrittsgebirge der hier untersuchten Zielfunktionen in ihrer Struktur mit größerem n immer ähnlicher.

Bei einigen Funktionen kann man eindeutig sagen, ob eine Schrittweite ausreicht oder n nötig sind, bei der Doppelsumme ändert sich die Empfehlung mit zunehmender Dimension n. Ohne eine Rekombination auf den Schrittweiten lohnt sich die Einstellung $n_\sigma = n$ in keinem Fall, da sich nur über die Mutation der Strategievariablen kein topologiespezifisches Wissen bilden kann. Es kann auch durchaus sein, daß noch bessere ES-Varianten zwischen den beiden Extremen 1 bzw. n liegen[2].

Aus den hier untersuchten Zielfunktionen kann man den Schluß ziehen, daß die ES — wie alle anderen Optimierverfahren auch — eher mit dem „Fluch der Dimension“ als mit der eigentlichen Zielfunktionstopologie zu kämpfen haben: Das Volumen des $\mathbb{R}^n$ explodiert mit steigendem n, und offensichtlich reichen Tests mit $n = 30$ noch nicht aus, um einem Problem seine Charakteristika zu entlocken. Rechenberg spricht von einer der menschlichen Einsicht nur schwer zugänglichen Distanz–Volumenexplosion beim Übergang zu hochdimensionalen Variablenräumen, die die Metapher einer im Nebel nach dem höchsten Gipfel suchenden Bergsteigergruppe, welche gern zur Veranschaulichung des ES–Funktionsprinzips herangezogen wird, gefährlich macht, weil sie zur Assoziation einer eben nur dreidimensionalen Welt führt (Rechenberg, 1994). Die Grate im $\mathbb{R}^n$ — und es gilt für alle Optimierverfahren, den Grat des Gradienten möglichst gut zu treffen — werden nicht einfach nur immer etwas schmaler, sondern sie werden unendlich schmal.

Die in diesem Kapitel immer wieder aufgezeigte Interdependenz zwischen den Operatoren macht deren richtige Einstellung a priori schwierig. So kann man alle bis auf einen Parameter richtig justieren, wählt man den falschen Rekombinationstyp für die Objektvariablen oder ein etwas zu großes τ_i, versagt die ES. Weit schlimmer an den wechselseitigen Abhängigkeiten der externen ES-Parameter im Hinblick auf unbekannte Zielfunktionstopologien ist aber, daß es nicht *den* besten Rekombinationstyp für ein Problem zu geben

[2] Bei n Objektvariablen und $m < n$ Schrittweiten benutzt man σ_m zur Mutation von $x_m, \ldots, x_n$.

scheint, ebensowenig wie *den* optimalen Selektionsdruck, nur in Ausnahmefällen *die* richtige Anzahl an Schrittweiten oder *die* optimale τ_0–τ_i–Parametrisierung. Einerseits ändern sich diese Werte in Abhängigkeit von n, was allerdings bei praktischen Problemen mit üblicherweise fester Größe nicht relevant ist, andererseits ändern sich diese Werte bei festem n in Abhängigkeit voneinander: Beim Wechsel der Rekombinationsart verschiebt sich der interessante Teil des τ_0–τ_i–Gebirges mit den guten Konvergenzgeschwindigkeiten, bei verändertem Selektionsdruck kann sich eine andere als die zuvor ermittelte Variante als beste Rekombinationsart herausstellen.

Die in diesem Kapitel beschriebenen Verflechtungen der Operatoren lassen es nicht ratsam erscheinen, noch weitere Parameter zu rastern. Im folgenden Kapitel 6 wird ein Algorithmus vorgestellt, gute ES–Parametrisierungen stattdessen zu „züchten“.

Kapitel 6

Die Meta–Evolutionsstrategie

If a problem seems unsolvable, consider that you may have a meta–problem.

Sich zur Problemlösung ausgerechnet jener Methode bedienen zu wollen, die das Problem hervorgerufen hat, klingt nach einem Tausch Teufel gegen Beelzebub. Aber wie sehen mögliche Alternativen aus? Die Theorie liefert bestenfalls asymptotische Aussagen, bezieht aber (noch) nicht alle Parameter ein, und eine Rasterung aller 8 relevanten Parameter erscheint undurchführbar[1]. Die Untersuchung der Leistungsfähigkeit neuer Operatoren mit ggf. zusätzlichen Parametern verbietet sich auf diese Weise ganz. Und neben der fehlenden Eleganz der Enumerationsmethode birgt sie auch die Gefahr, daß man nur genau jene Wertebereiche untersucht, die einem aufgrund eigener Erfahrungen und der Literatur sinnvoll erscheinen.

Die in diesem Kapitel vorgestellte Meta–ES verfügt zwar ebenfalls über diesen „blinden Fleck“, wenn man sich die Initialisierung und z. T. die Operatoren ansieht, sie verfügt aber dennoch über genügend Freiheitsgrade, „ungewöhnliche“ Parameterwerte einzustellen, wenn diese das durch gute Konvergenz rechtfertigen.

Natürlich erscheint es verschwenderisch, bessere Strategien durch Simulationen züchten zu wollen, andererseits bietet die Simulation den letzten Zufluchtsort, solange die Theorie noch keine Antworten liefern kann. Um z. B. zu zeigen, daß eine beiderseitig korrekt gespielte Partie Mühle unentschieden endet, wurde der Zustandsraum drei Jahre lang (mit Unterbrechungen) erschöpfend durchsucht (Gasser und Nievergelt, 1994).

[1] Abschätzung: 2 Selektionsarten, je 10 Werte für μ und λ, je 30 Werte für τ_0 und τ_i, 5 Schrittweitenzahlen, je 5 Rekombinationstypen für x_{reko} und σ_{reko} und nur je 10 Läufe pro Parametrisierung zur Absicherung, d. h. $(2 \cdot 10^2 \cdot 30^2 \cdot 5 \cdot 5^2) \cdot 10 = 2.25 \cdot 10^8$ Läufe pro untersuchtem n und pro untersuchter Funktion, d. h. bei Parallelisierung auf 50 Rechnern und nur einer Minute pro Lauf wartet man ein knappes Jahrzehnt auf das Ergebnis für genau eine Zielfunktion und ein n.

Das Aufsetzen einer Meta–Ebene mangels Alternativen ist kein bloßer Trick, sondern läßt sich auch biologisch rechtfertigen. Elementare Prinzipien der natürlichen Evolution wie z. B. das der Rekombination sind erst im Laufe der Zeit entstanden, d. h. sie unterlagen selbst einer Evolution. Wie auch Kapitel 2 zu zeigen versucht, ist die Natur hierarchisch organisiert, wenn sich auch die Ebenen dort nicht immer so leicht voneinander trennen lassen wie bei dem Modell in diesem Kapitel.

6.1 Verwandte Arbeiten

In der Berliner Gruppe um Rechenberg wurde der Begriff der geschachtelten Evolutionsstrategien samt zugehöriger Notation geprägt (Rechenberg, 1994). Eine $[m, l\,(\mu\,,\,\lambda)^{\gamma}]$–ES bezeichnet eine Strategie, bei der in der äußeren Klammer die l Populationen nach ihrer Konvergenzgeschwindigkeit selektiert werden. Diese waren zuvor γ isolierten $(\mu\,,\,\lambda)$–Selektionen unterworfen. Wie der Meta–Ansatz in dieser Arbeit (s. Kapitel 6) wurde er entworfen, um strategiebezogene Parameter adaptieren zu können. Eine Darstellung der Resultate inklusive Fortschrittstheorie gibt es aber nur für eine 10–dimensionale lineare Zielfunktion $\Sigma_{i=1}^{n} x_i$ mit einer $[1, 2\,(1\,,\,\lambda)^{\gamma}]$–ES (Santibáñez-Koref, 1995). Es fehlt eine systematische Darstellung der Resultate, und die angegebenen Isolationszeiten ($\gamma \sim 30$) lassen es m. E. zweifelhaft erscheinen, daß sich andere als nur lokal schnelle Strategien durchsetzen. Geyer et al. benutzen für ein multimodales Regressionsproblem ebenfalls geschachtelte ES, allerdings steht dort im Vordergrund, überhaupt ein Verfahren zur Verfügung zu haben, das das globale Optimum zuverlässig findet. Ferner ist die Dimension des Suchraumes mit $n = 2$ auch für eine Verallgemeinerung viel zu klein (Geyer, Ulbig, Schulz und Bräuer, 1999).

Bäck verwendet einen Meta–Ansatz, um die externen Parameter eines um ES–Operatoren erweiterten GA auf maximale Konvergenzgeschwindigkeit hin zu optimieren. Benutzt wird dabei das Kugelmodell mit $n = 20$ (Bäck, 1996). Ein noch früherer Meta–GA–Ansatz stammt von Grefenstette, dessen Algorithmus aber sogar die kontinuierlichen Parameter diskretisiert, für die er nur eine kleine Menge vorgegebener Einstellungen zur Auswahl hat, und die Fitneß über gleitende Durschnitte mißt, was bei statischen Optimierproblemen nicht unbedingt angemessen erscheint (Grefenstette, 1986).

Schaffer et al. haben die von De Jong (De Jong, 1975) empfohlenen Kontrollparameter eines GA (Populationsgröße, Crossover–Wahrscheinlichkeit, Mutationsrate) systematisch überprüft und erhebliche Abweichungen gefunden (Schaffer, Caruana, Eshelman und Das, 1989). Allerdings haben sie eine Menge ihnen sinnvoll erscheinende Kontrollparameter (6 Populationsgrößen, 10 Crossover–Wahrscheinlichkeiten, 7 Mutationsraten und ein oder zwei Crossover–Punkte) vorgegeben und anhand von 10 Zielfunktionen alle sich ergebenden 8 400 Möglichkeiten untersucht.

Schlierkamp–Voosen hat in seiner Dissertation ebenfalls die Parametrisierung des sogenannten Breeder Genetic Algorithm — ein hybrider Ansatz zwischen GA und ES — mittels eines Wettbewerbs auf der Ebene darüber untersucht. Im Gegensatz zu dem im folgenden Abschnitt beschriebenen Meta–Algorithmus arbeitet er, der GA–Tradition folgend, mit einer proportionalen Selektion, um die Parametrisierung laufend der momentanen Situation anzupassen. Einen Mutationsoperator aufgrund von Voruntersuchungen aus einer endlichen Alternativenmenge wählen zu lassen, stellt m. E. eine unnötige Einschränkung dar (Schlierkamp-Voosen, 1998).

6.2 Beschreibung des Algorithmus

Die Meta–ES soll für die folgenden Parameter möglichst gute Werte finden. Diese Parameter sind dann Einstellungen für eine ES auf der Ebene darunter, deren mittlere Fortschrittsgeschwindigkeit gemessen und als Selektionskriterium auf der Meta–Ebene benutzt wird.

- die Anzahl der Eltern μ und die die Anzahl der Nachkommen λ
- die Selektionsart sel: $1 \mathrel{\hat{=}} (\mu + \lambda)$, $2 \mathrel{\hat{=}} (\mu , \lambda)$, $3 \mathrel{\hat{=}} (\mu \overset{(+)}{,} \lambda)$[2]
- die Schrittweitenzahl n_σ
- die Mutabilität der Schrittweiten τ_0 und τ_i
- die Rekombinationsart der Objekt– und Strategievariablen x_{reko} und σ_{reko}

Folgende Parameter müssen zusätzlich vorgegeben werden, bleiben aber sinnvollerweise für die Dauer der Meta–Evolution konstant.

- die Zielfunktion
- die Dimension n des Problems
- die Anzahl an Generationen, die ein Lauf maximal dauern darf[3]
- die Anzahl der Läufe, deren mittlerer Fortschritt gemessen wird
- die Startwerte von Objekt– und Strategievariablen

Der Grundalgorithmus der Meta–ES entspricht weitgehend dem einer Standard–ES (s. Seite 40). Als Terminierungskriterium dient eine maximale Generationenzahl der Meta–Ebene, zusätzlich kann man nach jeder aktuellen Generation unterbrechen.

[2] Diese Variante ist aus historischen Gründen vertreten und bedeutet, daß das beste Individuum einer Generation immer in die nächste übernommen wird, während auf den anderen Individuen eine Komma–Selektion ausgeführt wird.

[3] Das Generationslimit der einzelnen Läufe sollte der Problemdimension n angemessen sein. Läßt man bei einem ggf. multimodalen 100–dimensionalen Problem nur 200 Generationen zu, sollte man auf überraschende Ergebnisse gefaßt sein.

initialisiere Population $P_{meta}^{(g:=0)}$
`while` Terminierungskriterium nicht erfüllt `do`
  rekombiniere λ_{meta} Nachkommen aus je 2 der μ_{meta} Elternindividuen aus $P_{meta}^{(g)}$
  mutiere die λ_{meta} Nachkommen
  bewerte die λ_{meta} Nachkommen
  selektiere μ_{meta} Individuen für $P_{meta}^{(g+1)}$:
    `if` $(\mu_{meta}\,,\,\lambda_{meta})$–Selektion `then` selektiere μ_{meta} aus λ_{meta} Nachkommen
      `else` selektiere μ_{meta} aus μ_{meta} Eltern und λ_{meta} Nachkommen
    `fi`
  setze $g := g + 1$
`od`

Der Startwert für λ wird von außen vorgegeben, μ wird gleichverteilt aus dem Intervall $[1, \ldots, [\lambda/2]]$ gezogen, die Art der Selektion sel gleichverteilt aus dem Intervall $[1, \ldots, 3]$, der Parameter n_σ gleichverteilt dem Intervall $[1, \ldots, n]$. Die Werte für τ_0 und τ_i werden gemäß Heuristik initialisiert, und die beiden Parameter, die die Rekombinationsart enthalten, werden zunächst gleichverteilt aus dem Intervall $[1, \ldots, 7]$ gezogen, dann aber, wie vor jeder Bewertung, auf ihre Plausibilität hin überprüft. So kann mit $\mu = 1$ keine Rekombination stattfinden. Und die beiden Crossover–Varianten stehen nicht zur Verfügung, falls die Werte von n bzw. n_σ die erforderlichen Bruchstellen nicht hergeben.

Um das Problem der Einstellung der externen Parameter nicht einfach nur eine Ebene nach oben zu verschieben, wird der Meta–Algorithmus aller Parametrisierungsmöglichkeiten außer μ^{meta}, λ^{meta} und der Selektionsart beraubt. Die Art der Rekombination wird wie folgt festgelegt, wobei die beiden Elternindividuen für jeden Parameter neu und zufällig bestimmt werden: Die Parameter μ, λ, n_σ eines Nachkommen werden aus dem gerundeten arithmetischen Mittel der jeweiligen Information zweier Elternindividuen berechnet, für τ_0 und τ_i gilt dasselbe ohne Rundung. Die beiden Rekombinationsparameter x_{reko} und σ_{reko} werden mit Wahrscheinlichkeit $1/2$ diskret von einem der beiden zufällig gewählten Eltern vererbt, ebenso wie der Parameter sel, der die Art der Selektion kodiert.

Ferner wird auf eine Selbstadaptation der Meta–Schrittweiten verzichtet, so daß zwar ein ständiges Grundrauschen der Parameter eine echte Konvergenz auf der Meta–Ebene verhindert, aber der Algorithmus neben den 8 Parametern nicht noch 8 zugehörige Mutationsraten einstellen muß. Die Resultate (s. Kapitel 6.3) rechtfertigen dieses Vorgehen. Die einzelnen Parameter werden wie folgt mutiert:

$$\begin{aligned}
\mu^{(g+1)} &= \left|\left[\mathcal{NV}\left(\mu^{(g)},\, 4.0\right)\right]\right| \\
\lambda^{(g+1)} &= \left|\left[\mathcal{NV}\left(\lambda^{(g)},\, 5.0\right)\right]\right| \\
sel^{(g+1)} &= \left|sel^{(g)} + [\mathcal{DNV}\,(1,\, 1/3\,)]\right|
\end{aligned}$$

$$
\begin{aligned}
n_\sigma^{(g+1)} &= \left|\left[\mathcal{NV}\left(n_\sigma^{(g)},\, 5.0\right)\right]\right| \\
\tau_0'^{(g+1)} &= \left|\mathcal{NV}\left(\tau_0'^{(g)},\, 0.2\right)\right| \\
\tau_0^{(g+1)} &= \tau_0'^{(g+1)}/\sqrt{n} \\
\tau_i'^{(g+1)} &= \left|\mathcal{NV}\left(\tau_i'^{(g)},\, 0.2\right)\right| \\
\tau_i^{(g+1)} &= \tau_i'^{(g+1)}/\sqrt{2\cdot\sqrt{n_\sigma^{(g+1)}}} \\
x_{reko}^{(g+1)} &= \left|x_{reko}^{(g)} + [\mathcal{DNV}\,(1,\, 1/3\,)]\right| \\
\sigma_{reko}^{(g+1)} &= \left|\sigma_{reko}^{(g)} + [\mathcal{DNV}\,(1,\, 1/3\,)]\right|
\end{aligned}
$$

mit: $\mathcal{NV}\,(E,\, s\,)$: Realisierung einer Normalverteilung mit Erwartungswert E und Standardabweichung s

$[\mathcal{DNV}\,(n,\, p\,)]$: Realisierung einer diskretisierten Normalverteilung, deren Parameter an eine $(n,\, p)$–Binomialverteilung angelehnt sind, $n \in I\!N$, $0 < p < 1$, Erwartungswert $E = n \cdot p$ und Standardabweichung $\sqrt{n \cdot p \cdot (1-p)}$

Alle beschriebenen Mutationsarten erhalten eine gewisse Ähnlichkeit zwischen Eltern– und Nachkommeninformation. Je nach Natur der Variablen wird dies über die Normalverteilung oder deren Diskretisierung realisiert. Um die Darstellung oben nicht zu überfrachten, wurden folgende Plausibilitätstests der mutierten Parameter zunächst weggelassen:

- μ darf nur Werte aus dem Intervall $[1, \ldots, \lambda]$ annehmen.
- λ bleibt auf das Intervall $[\mu, \ldots, 300]$ beschränkt, wobei 300 eine obere Grenze ist, die bei 100 als der maximalen Problemgröße n angemessen erscheint. Bei einigen Läufen der Meta–ES ging es nur um den Selektionsdruck, bezogen auf das vertraute $\lambda = 100$. In diesen Läufen blieb λ dann unmutiert.
- *sel* bleibt mittels *modulo* 3 auf das Intervall $[1, \ldots, 3]$ beschränkt. Bei zunächst nur drei zur Verfügung stehenden Selektionsvarianten hätte man diesen Parameter auch gleichverteilt ziehen können. Die hier realisierte Mutationsart soll spätere Erweiterungen erleichtern.
- n_σ bleibt mittels *modulo* n auf das Intervall $[1, \ldots, n]$ beschränkt.
- τ_0 und τ_i werden mittels einer Hilfsgröße mutiert. Dieses Vorgehen erhält einerseits die Proportionalität zu n bzw. der Schrittweitenzahl n_σ des Meta–Individuums und ermöglicht andererseits über die Mutation von τ_0' und τ_i' die Variation. Der Parameter τ_i wird auf Null gesetzt, falls $n_\sigma = 1$ gilt, τ_i' wird aber unmutiert vererbt, um für

andere n_σ–Werte auf einen bereits erprobten Wert von τ_i' zurückgreifen zu können. Eventuell auftretende negative Werte werden auf Null gesetzt.

- x_{reko} und σ_{reko} bleiben mittels *modulo* 7 auf das Intervall $[1, \ldots, 7]$ beschränkt. Weitere Plausibilitätstests werden analog zu denen in der Initialisierung durchgeführt.

Die Bewertung der λ_{meta} Nachkommen als Mittelwert von 5 Läufen erfolgt i. d. R. synchron parallel, einstellbar ist aber auch die sequentielle Variante. Die Fitneßfunktion auf der Meta–Ebene unterliegt damit trotz der Mittelung stochastischen Störungen. Verwendet wird üblicherweise Kg_2 zur Beurteilung eines Parametersatzes, möchte man nur möglichst schnelle Strategien finden, kann man die Meta–Selektion auf Kg_1 umstellen.

6.3 Resultate der Meta–ES

Um in den Abbildungen dieses Abschnitts alle variierten ES–Parameter gleichzeitig darstellen zu können, werden μ, die Selektionsart und n_σ ohne Transformation verwendet, τ_0 und τ_i werden jeweils mit 100 multipliziert, der Rekombinationstyp wird nur mit seinen beiden relevanten Stellen berücksichtigt, und λ wird ggf. durch 3 dividiert, um die einheitliche Skalierung beibehalten zu können. Dies ist dann in den Abbildungen entsprechend angegeben. Zusätzlich zur Parametrisierung ist immer die erreichte Konvergenzgeschwindigkeit Kg_2 eingezeichnet. Dadurch kann man den Abbildungen entnehmen, wie schnell sich gute Strategien einstellen. Die Meta–Evolution lief immer länger als in den Abbildungen dargestellt. Unabhängig von der Generationenzahl, nach der die Meta–Evolution abgebrochen wurde, ist immer nur jener Teil eines Laufes dargestellt, in dem noch erkennbar Veränderungen stattfinden. Bei den vergleichenden Aussagen wurde die Fortschrittsgeschwindigkeit der aus der Meta–Strategie resultierenden ES mit einer bzgl. μ, λ und n_σ angepaßten Standard–ES in je 100 Läufen separat verglichen.

Ein erster Test, der die Parameter einer möglichst schnell konvergierenden ES auf dem $n = 10$–dimensionalen Kugelmodell herausfinden sollte, bestätigte die Funktionsfähigkeit des Meta–Algorithmus: Eine $(5\,,\,50)$–Meta-ES, gestartet mit einem anfänglichen λ von 20 sollte in 150 Generationen mit einer ES möglichst weit kommen. Als Ergebnis erhält man eine $(15\,,\,300)$–ES, die mit nur einer Schrittweite operiert, $\tau_0 = 0.4$ einstellt (Heuristik: 0.32) und als Rekombinationstyp 531. Nach den erlaubten 150 Generationen überwindet sie fast doppelt so viele Zehnerpotenzen wie eine $(15\,,\,300)$–Standard–ES. Dieser erhöhte Selektionsdruck bestätigte sich in einem weiteren Test, bei dem $\lambda = 100$ konstant gehalten wurde und eine $(5\,,\,100)$–ES den besten Fortschritt erzielte. Weitere Tests mit niedrig–dimensionalen Kugelmodellen bestätigten die Validität des gewählten Meta–Ansatzes. Immer wurde $n_\sigma = 1$ eingestellt, λ lief gegen die vorgegebene Obergrenze von 300, und der beste Selektionsdruck war deutlich härter als 1/6 bis 1/7. Die für das Kugelmodell zu erwartenden sinnvollen Einstellungen wurden also gefunden.

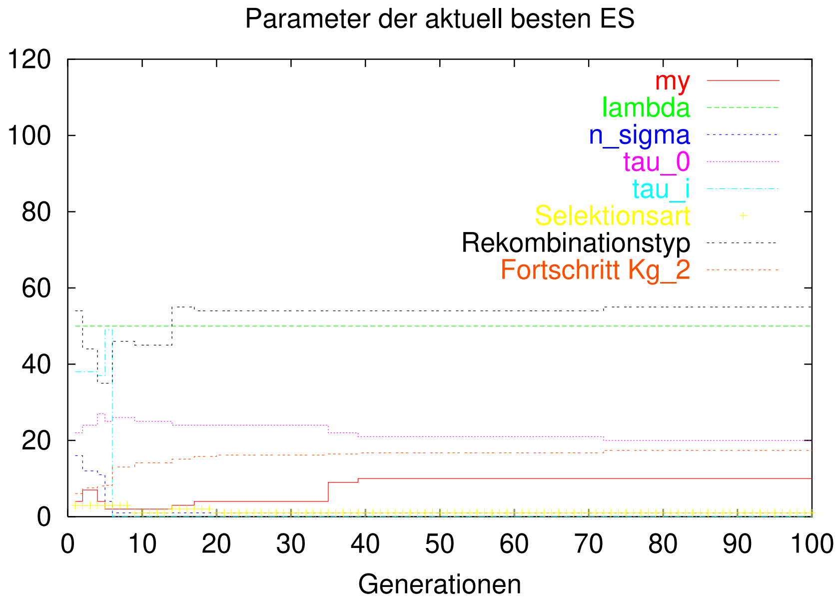


Abbildung 6.1: $(1\,,\,5)$–Meta–ES, Kugelmodell, $n = 30$

Für das $n = 30$–dimensionale Kugelmodell stellte sich der erwartete Rekombinationstyp $5[4,5]1$ ein, $n_\sigma = 1$ wird eingestellt, und bei fixiertem $\lambda = 100$ kommt eine $(7 - 10\,,\,100)$–ES mit $\tau_0 = 0.21$ am schnellsten voran, ist aber nur etwa 10% schneller als die vergleichbare Standard–ES. Fixiert man $\lambda = 50$, stellt sich $\mu = 5$ dazu passend ein, die anderen Parameter bleiben unverändert (s. Abbildung 6.1). Bei variablem λ ergeben sich $(15\,,\,300)$–ES mit $\tau_0 = 0.23$ und ansonsten gleichen Parametern. Sowohl Rekombinationstyp als auch τ_0 passen zu den besten Werten aus Kapitel 5.

Auch beim $n = 100$–dimensionalen Kugelmodell entsprachen die Ergebnisse den Erwartungen, was $n_\sigma = 1$, den Rekombinationstyp $[4,5]51$ und $\tau_0 = 0.1$ betrifft. Bei festem $\lambda = 100$ stellt sich μ zwischen 6 und 10 ohne große Auswirkungen auf die Konvergenzgeschwindigkeit ein, eine bevorzugte Selektionsart kann man ebensowenig erkennen (s. Abbildung 6.2) wie einen nennenswerten Geschwindigkeitsunterschied zur Standard–ES. Fixiert man zusätzlich $n_\sigma = 100$, kommen Plus–Strategien mit μ zwischen 25 und 35 sowie $\tau_0 = 0.22$ und $\tau_i = 0.55$ am schnellsten voran (s. Abbildung 6.3). Diese Werte entsprechen ebenfalls jenen aus Kapitel 5. Im Vergleich mit der Standard–ES sind die neuen Strategien etwa 30% schneller, allerdings nur, wenn man den Beobachtungszeitraum groß genug wählt, denn anfänglich konvergiert die Standard–ES schneller.

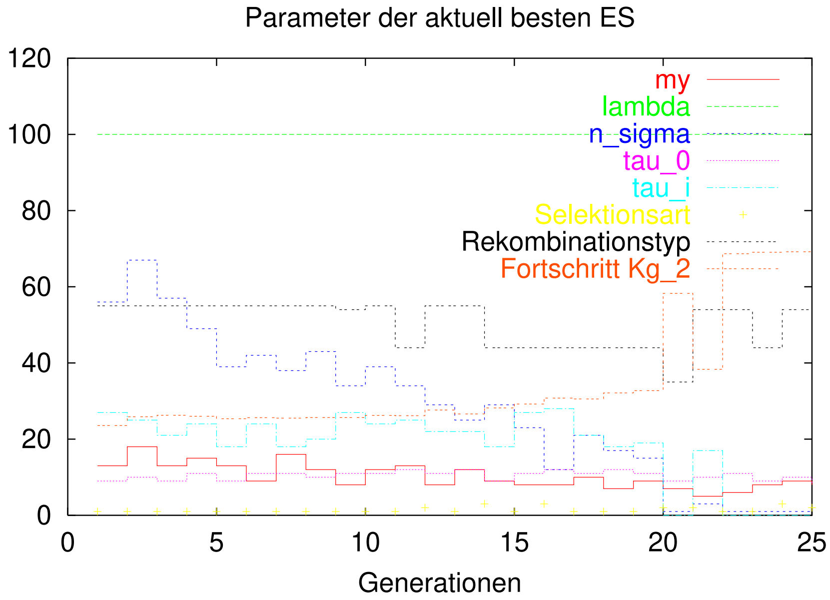


Abbildung 6.2: (8 , 50)–Meta–ES, Kugelmodell, $n = 100$

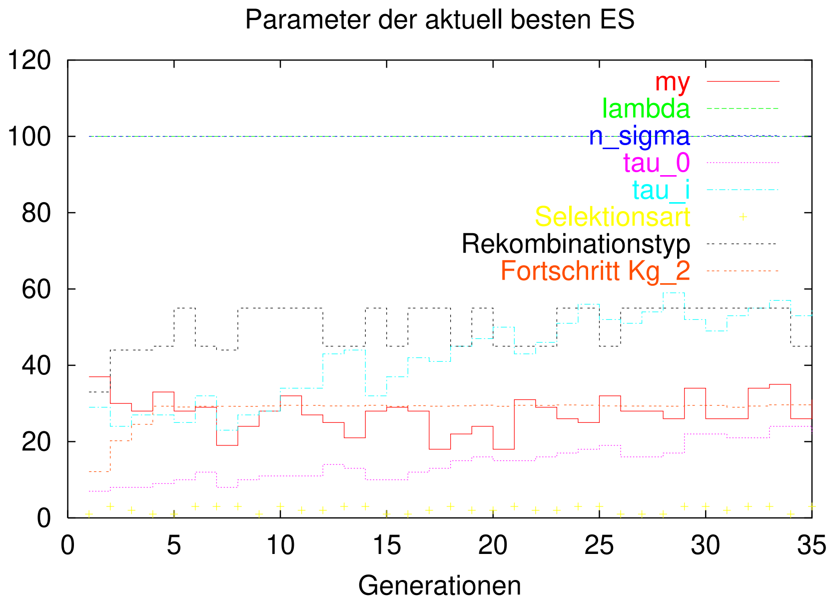


Abbildung 6.3: (8 , 50)–Meta–ES, Kugelmodell, $n = 100$, $n_\sigma = n$

Beim Skalierungstest mit $n = 30$ setzten sich bei voller Variabilität der $(8\,,\,50)$–Meta–ES ausschließlich Komma–Strategien durch, die n_σ nahe n einstellen und den Rekombinationstyp 551 benutzen. Diese drei Parameter werden zuerst eingestellt und bleiben dann praktisch konstant. Der weitere Zuwachs an Konvergenzgeschwindigkeit resultiert aus dem gegen 300 wachsenden λ, die Werte für τ_0 und τ_i schwanken zwar, bleiben aber nahe 0.25 bzw. 0.4, μ sollte dann nahe 30 eingestellt werden (s. Abbildung 6.4). Die gezüchtete Strategie ist 4.5 mal schneller als eine $(30\,,\,300)$–Standard–ES. Stellt man deren Rekombinationstyp auf 551 ein, reduziert sich der Vorteil auf den Faktor 2.

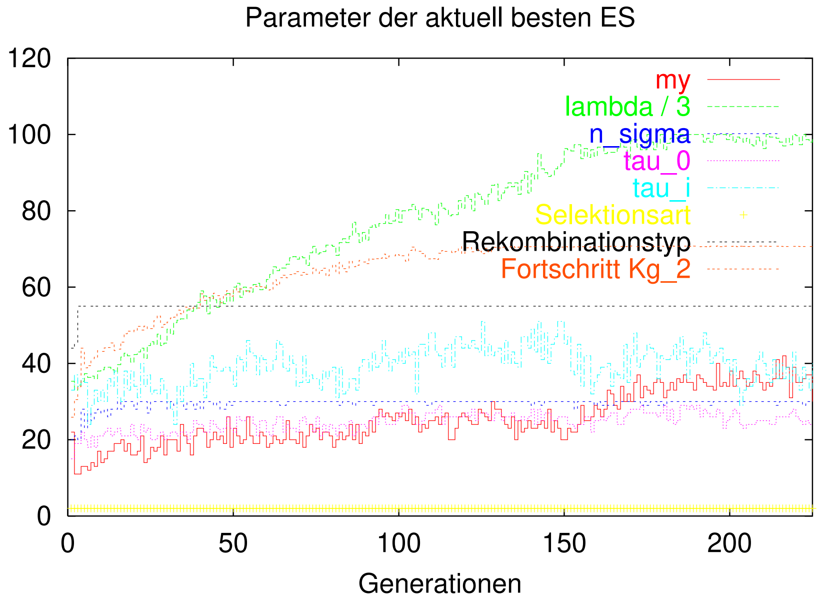


Abbildung 6.4: $(8\,,\,50)$–Meta–ES, Skalierungstest, $n = 30$

Für $n = 100$ setzten sich jene ES durch, die mit $n_\sigma \geq 90$ und Rekombinationstyp 551 arbeiten (s. Abbildung 6.5). Bei fixiertem λ gewannen überraschend eindeutig Plus–Strategien mit zwischen 25 und 40 schwankendem μ. Die Parameter τ_0 und τ_i entwickeln sich in Richtung der besten empirisch ermittelten Werte aus Kapitel 5 (0.25 und 0.6, Heuristik: 0.07 und 0.22). Entgegen den Aussagen in der Literatur sind also auch Plus–Strategien bei hinreichender Diversität in der Lage, eine Schrittweitenskalierung zu erlernen. Mit $n_\sigma = 90$ Schrittweiten ist eine $(35 + 100)$ etwa 25% schneller als die Standard–ES mit Rekombinationstyp 551, bei voller Schrittweitenzahl erhöht sich der Vorteil auf 35%.

Eine große Abweichung vom erwarteten Rekombinationstyp lieferten die Meta–ES–Läufe mit der $n = 30$–dimensionalen Doppelsumme (s. Abbildung 6.6): Sehr schnell stellte sich ein Wert von 651 ein, der bei den Tests in Kapitel 5 noch gezielt gewählt wurde, um

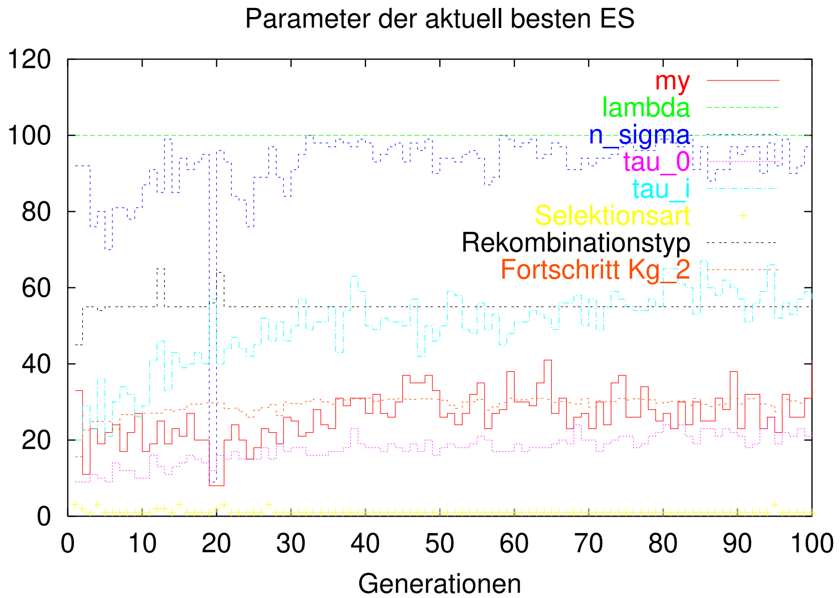


Abbildung 6.5: (8 , 50)–Meta–ES, Skalierungstest, $n = 100$

divergentes Verhalten zu provozieren. Allerdings ergaben sich hier deutlich härter selektierende (13 – 17 , 300)–ES. Das Ergebnis, bei dieser Funktion eher mit der vollen Anzahl an Schrittweiten zu arbeiten, bestätigte sich ($n_\sigma \geq 26$), entsprechend sind die τ–Parameter erneut deutlich größer als in der Heuristik ($\tau_0 = 0.27$, $\tau_i = 0.65$ gegen 0.13 und 0.3). Der Geschwindigkeitsvorteil gegenüber einer (5 , 100)–ES mit Rekombinationstyp 641 beträgt 66%.

Bleiben λ und n_σ bei der $n = 100$–dimensionalen Doppelsumme jeweils auf 100 fixiert, geht τ_i schnell gegen Null, um auf diesem Wege nur eine Gesamtschrittweite erlernen zu müssen. Die besten resultierenden Strategien sind (3 – 5 , 100)–ES mit sehr kleinem $\tau_0 = 0.02$ bis 0.04. Eine dauerhafte Einstellung des passenden Rekombinationstyps wird dadurch verhindert, daß μ immer wieder auf 1 eingestellt wird, somit die Rekombination abgeschaltet wird und bei Werten von $\mu > 1$ der Lernvorgang erneut bei 111 beginnen muß. Vor dem ersten Auftreten von $\mu = 1$ gab es eine klare Tendenz zu $\sigma_{reko} = 4$, beim Parameter x_{reko} traten die bei entsprechendem Selektionsdruck gut konvergierenden Werte $[4, 5, 6, 7]$ auf. Abbildung 6.7 veranschaulicht den Werteverlauf, solange $\mu > 1$ gilt. Gegenüber einer (3 , 100)–Strategie mit $n_\sigma = 1$, Rekombinationstyp 441 und τ_0 gemäß Heuristik ist die gezüchtete ES mit $n_\sigma = 100$, aber $\tau_0 = 0.03$ etwa 80%, die Variante mit $n_\sigma = 1$ immerhin noch 33% schneller.

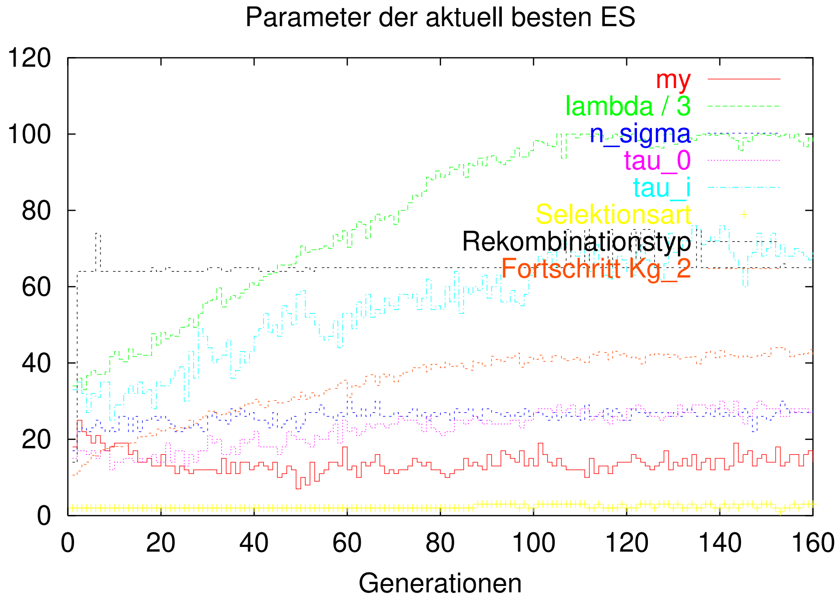


Abbildung 6.6: (8 , 50)–Meta–ES, Doppelsumme, $n = 30$

Die Doppelsumme gilt als die Testfunktion, bei der korrelierte Mutationen für eine gute Konvergenz nötig sind. Die oben beschriebenen Parameter für $n = 30$ können dann auch mit einer ES, die korrelierte Mutationen verwendet, noch nicht konkurrieren, sieht man einmal von der eingesetzten Rechenzeit ab, denn sie konvergieren nur etwa halb so schnell.

Überträgt man aber die für $n = 100$ gezüchteten Parameter ((3 , 100)–ES, Rekombinationstyp 441, $n_\sigma = 1$ oder $n_\sigma = n$ und $\tau_i = 0$) auf den 30–dimensionalen Fall, ist man etwa 50% und bzgl. CPU–Zeit um den Faktor 20 schneller als die entsprechende ES mit korrelierten Mutationen. Daß die Meta–ES diese guten Einstellungen nicht gefunden hat, liegt an einem in Unkenntnis der Möglichkeiten zu großzügig eingestellten Generationslimit auf ES–Ebene und Kg_2 als Selektionskriterium auf der Meta–Ebene. Es gab somit für die Meta–ES keinen Druck, noch schnellere Strategien zu finden.

Für $n = 100$ weist die beste durch Züchtung gefundene Strategie eine um 10% höhere Konvergenzgeschwindigkeit auf als eine ansonsten gleich parametrisierte (3 , 100)–ES mit korrelierten Mutationen und τ_0, τ_i gemäß Heuristik. Durch den Verzicht auf die korrelierten Mutationen ist die gezüchtete Strategie bzgl. CPU–Zeit nach Überwindung derselben Anzahl an Zehnerpotenzen mehr als 50–mal schneller.

Bei dieser Zielfunktion muß man allerdings etwas Geduld aufbringen: Die in ihren Parametern stark abgeänderte (Standard–)Vergleichs–ES ist bis zur Überwindung der ersten beiden Zehnerpotenzen noch deutlich besser, erst danach setzen sich die gezüchteten Pa-

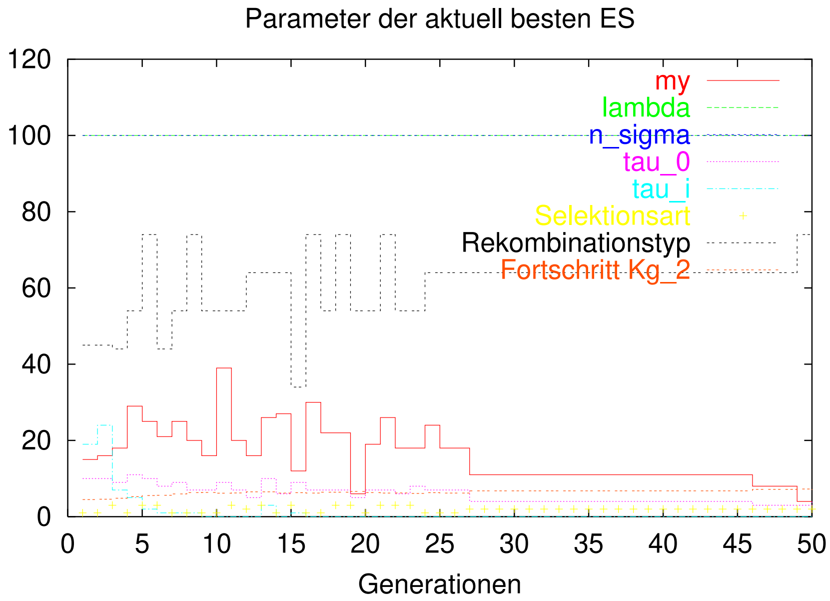


Abbildung 6.7: (8 , 50)–Meta–ES, Doppelsumme, $n = 100$, $n_\sigma = n$

rametrisierungen durch. Ferner weisen die schnellen Varianten immer wieder Phasen auf, die nach Divergenz aussehen. Dies ist insofern beruhigend, als das in Kapitel 5 Gesagte über die Entkopplung von Objekt– und Strategievariablen nicht falsch ist, sondern lediglich durch eine sehr harte Selektion in Kombination mit einer reduzierten Mutabilität der Schrittweiten nicht dauerhaft zum Tragen kommt.

Diese über ein sehr kleines τ_0 reduzierte Schrittweitenmutabilität ist m. E. auch der Schlüssel, um zu verstehen, warum ES mit korrelierten Mutationen bei der Doppelsumme doch nicht nötig sind: Wie ab Seite 41 ja bereits dargelegt, kann man für $n = 100$ von einer ES mit $\lambda = 100$ nicht erwarten, daß sie 4 950 Winkel auch nur halbwegs passend zueinander einstellt. Was stattdessen passiert, ist wohl nur eine „Nachbesserung“ eines durch die Objektvariablenrekombination vom Grat abgekommenen Individuums, indem es dieses idealerweise zurück auf den Grat setzt. Statt der erwünschten Ausrichtung entlang der elliptischen Höhenlinien für optimalen Fortschritt begrenzen oder reparieren sie jenen Schaden, den die Variabilität von $n_\sigma = n$ Schrittweiten angerichtet hat. Begrenzt man diese Variationsmöglichkeiten durch $n_\sigma = 1$, ein sehr kleines τ_0 und eine Rekombination, die Objekt– und Strategievariablen gekoppelt, d. h. von denselben beiden Eltern vererbt, gelingt es der Strategie bei hohem Selektionsdruck, den Gradienten in $\lambda = 100$ Versuchen hinreichend oft zu treffen. Eine Kombination der besten gezüchteten ES, allerdings dann wieder mit $n_\sigma = n$ und $\tau_i = 0$, mit korrelierten Mutationen bringt dann auch keinen weiteren Zuwachs an Konvergenzgeschwindigkeit.

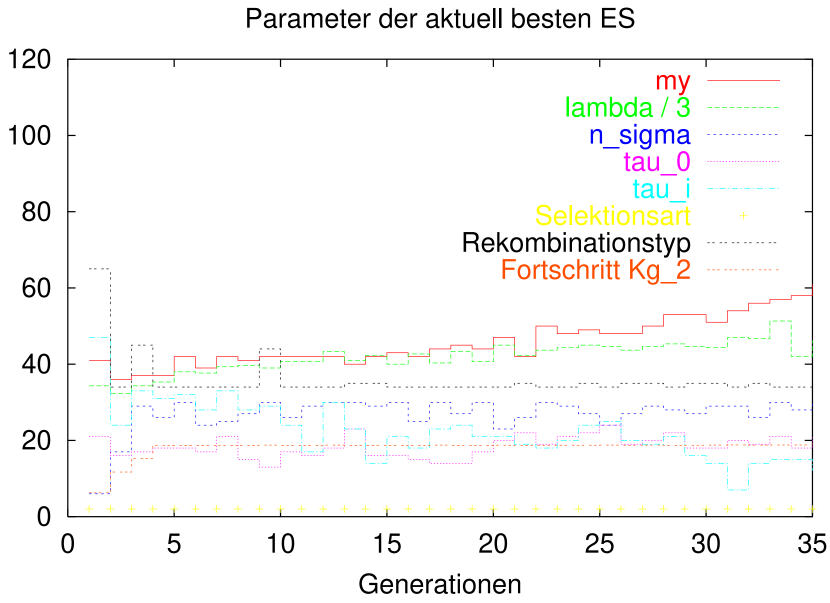


Abbildung 6.8: (8 , 50)–Meta–ES, multimodale Funktion, $n = 30$

Für die 30–dimensionale multimodale Funktion erwiesen sich wie erwartet $n_\sigma = n$ Schrittweiten als nötig, und auch die besten Rekombinationstypen $3[4, 5]1$ wurden eingestellt. Bei fixiertem $\lambda = 100$ stellt sich μ um 40 herum mit Komma–Selektion ein. Gibt man λ zusätzlich frei, wachsen μ und λ gemeinsam, ohne daß λ allerdings gegen die Grenze von 300 läuft. Wichtiger als die Erhöhung der Nachkommenzahl ist offenbar die Reduktion des Selektionsdruckes durch Vergrößerung von μ (s. Abbildung 6.8). Zunächst entsprechen auch die τ–Werte den Erwartungen, d. h. sie reduzieren die Mutabilität über Werte von 0.15 bzw. 0.2, um möglichst lange den explorativen Charakter der Suche aufrechterhalten zu können. Im Verlaufe der Meta–Evolution setzten sich aber jene Kombinationen durch, die die Empfehlung $\tau_0 < \tau_i$ der Heuristik ignorieren (0.2 bzw. 0.1). Daß mit diesen Parametern versehene ES schneller als der Standard sind, ist bei multimodalen Funktionen nur ein Aspekt. Zusätzlich sollte die Robustheit mindestens gleich gut sein. Der Vergleich mit einer Standard–ES läßt sich nicht durchführen, da diese bei derart geringem Selektionsdruck divergiert. Als Maßstab werden also die in Kapitel 5 erwähnten Ergebnisse herangezogen: 66% Trefferquote nach im Mittel 1 800 Generationen. Die gemäß Meta–ES parametrisierten Strategien liefern 83% Treffer nach 1 220 Generationen mit Komma–Selektion, mit Plus–Selektion sogar 95% Treffer nach nur 900 Generationen. Daß sich die Plus–Selektion nicht eindeutig durchsetzen konnte, liegt einerseits an der zu geringen Anzahl an Läufen, die mit einem Parametersatz zur Bestimmung der mittleren Fortschrittsgeschwindigkeit

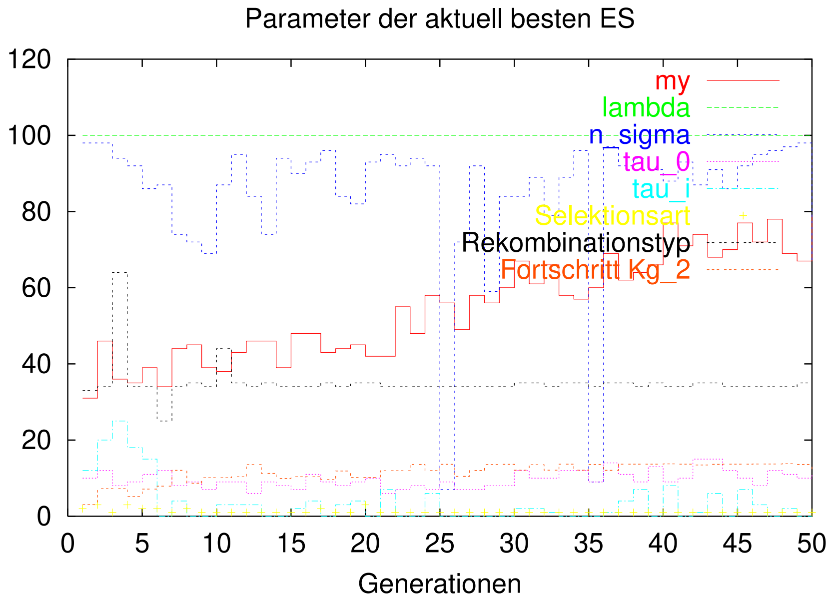


Abbildung 6.9: (8 , 50)–Meta–ES, multimodale Funktion, $n = 100$

durchgeführt wurden und andererseits an der Verwendung von Konvergenzmaß Kg_2, das die Anzahl der Generationen nicht berücksichtigt.

Bei $n = 100$ wächst μ gegen das fixierte λ, dies macht natürlich nur noch in Kombination mit einer Plus–Selektion Sinn. Die vorgegebene Generationsgrenze von 5 000 auf der ES– bzw. die Verwendung von Kg_2 als Selektionskriterium auf der Meta–Ebene bevorzugen langsame, aber sehr sicher konvergierende Parametrisierungen. Die erfolgreichen Rekombinationstypen sind nach wie vor 3[4, 5]1, n_σ bleibt zwar relativ groß, über ein sehr kleines τ_i (0.02) wird aber dennoch die Möglichkeit einer individuellen Schrittweitenentwicklung stark gedämpft (s. Abbildung 6.9). Der Parameter τ_0 stellte sich in der Nähe des von der Heuristik vorgesehenen Wertes 0.1 ein. Das globale Optimum wird mit diesen Einstellungen von einer (70 + 100)–ES in 92% der Läufe bei Inanspruchnahme der maximal erlaubten 5 000 Generationen gefunden, was bei den systematisch variierten Parametern aus Kapitel 5 in keinem Fall gelang. Mit obigen Einstellungen, aber nur einer Schrittweite, ist man zwar doppelt so schnell, findet allerdings das globale in Optimum nur noch in 62% der Läufe.

Weitere Läufe mit der Meta–ES wurden für die in Kapitel 5 nur erwähnten, aber nicht näher untersuchten Zielfunktionen gemacht.

Für die $n = 30$–dimensionale Griewank–Funktion ergaben sich Werte, die denen des Kugelmodells eher ähneln als denen „echter“ multimodaler Funktionen: Am besten konvergiert eine ES mit relativ harter Komma–Selektion (7 , 100) bzw. (15 , 300) mit intermediärer Rekombination 5[4, 5]1, $n_\sigma = 1$ und $\tau_0 = 0.16$ (s. Abbildung 6.10). Das globale Optimum wird in 99% der Läufe gefunden, und man ist dabei 60% schneller als eine Standard–ES, die allerdings ebenfalls keine Probleme zeigt, das globale Optimum zu finden.

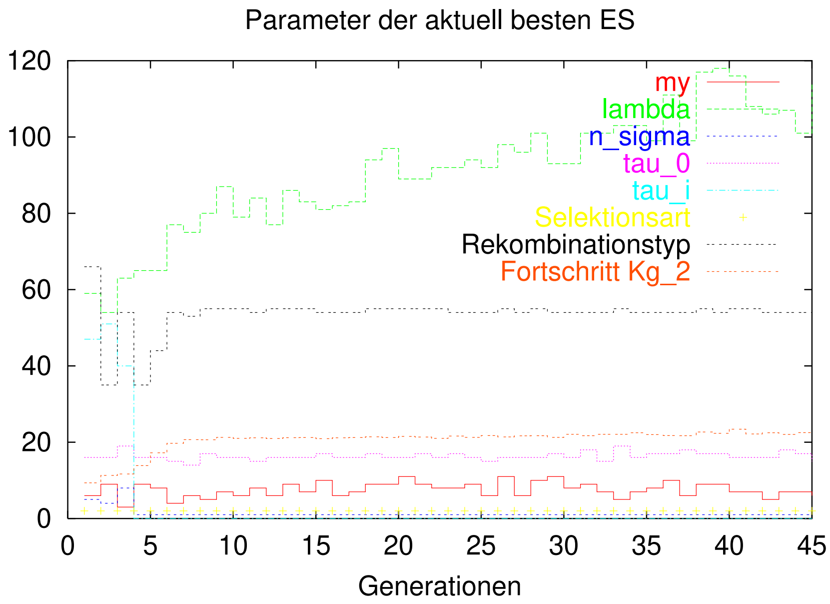


Abbildung 6.10: (6 , 40)–Meta–ES, Griewank, $n = 30$

Bei der Rastrigin–Funktion bestätigten sich die Ergebnisse für die multimodale Funktion. Für $n = 30$ ist bei n_σ kein eindeutiger Trend erkennbar, aber erneut stellt sich τ_0 bei 0.2 und τ_i bei 0.1 ein. Als Rekombinationstyp sollte 3[4, 5]1 bei einer (50 + 100) eingestellt werden (s. Abbildung 6.11). Für $n = 100$ bestätigte sich dieses Resultat, allerdings mit fixiertem $n_\sigma = 100$: Plus–Strategien mit $\mu \sim \lambda/2$, Rekombinationstyp 3[4, 5]1 und $\tau_0 = \tau_i = 0.1$ konvergieren am schnellsten und sichersten.

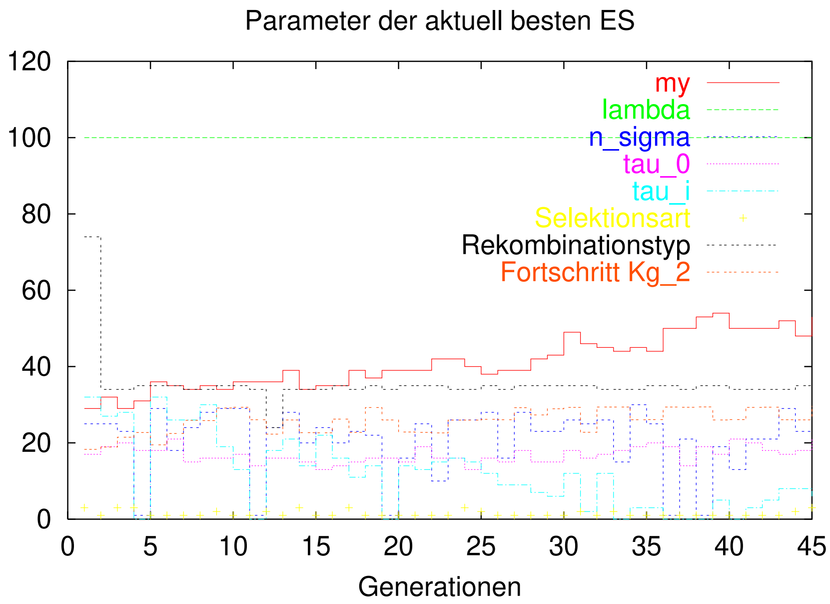


Abbildung 6.11: $(6\,,\,40)$–Meta–ES, Rastrigin, $n = 30$

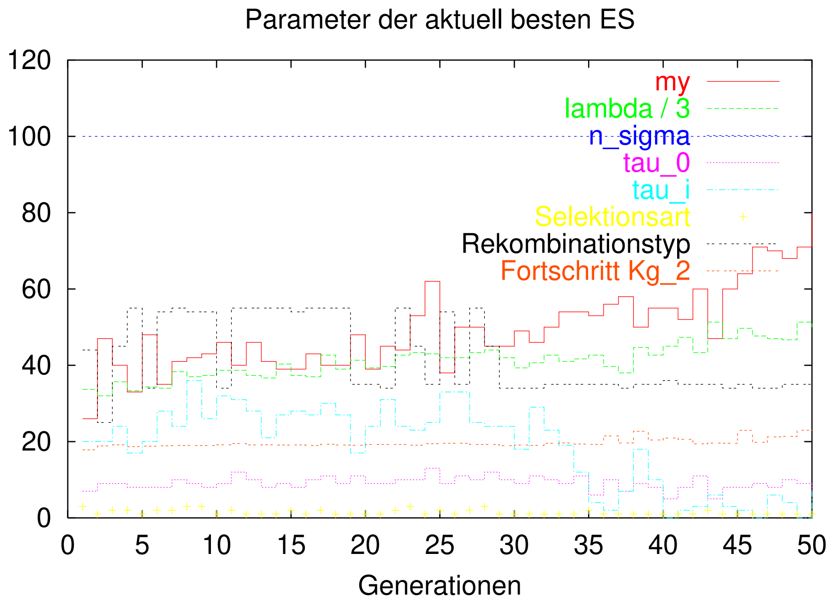


Abbildung 6.12: $(10\,,\,50)$–Meta–ES, Rastrigin, $n = 100$, $n_\sigma = n$

6.4 Zusammenfassung

Die Meta–ES zeigte sich in der Lage, jene bis zu 8 externen Parameter einer ES, die deren Konvergenzverhalten beeinflussen, so einzustellen, daß immer z. T. erheblich schnellere und — bei den multimodalen Funktionen — sogar schnellere *und* robustere Parametrisierungen schon nach relativ wenigen Generationen auf der Meta–Ebene entdeckt wurden. Falls man ES als Optimierverfahren verwendet, kann man also ein Optimierproblem beliebiger Dimension gegen ein maximal 8–dimensionales eintauschen, sofern die Zielfunktionsauswertung nicht zu aufwendig ist. *Der* universelle Optimieralgorithmus, der beliebige Probleme sicher und schnell löst, existiert damit natürlich immer noch nicht. Das „No–Free–Lunch–Theorem“ besagt, daß sich zwei beliebige Optimierverfahren gemittelt über alle Probleme in ihrer Leistungsfähigkeit nicht unterscheiden (Wolpert und Macready, 1997)[4].

In der Tendenz überraschen einzelne Aspekte der „gezüchteten“ Parametrisierungen sicher nicht (Schrittweitenzahl, Selektionsdruck bei den einzelnen Zielfunktionen). Aber daß das Zusammenspiel der einzelnen Operatoren derart komplex ist, daß ein Rekombinationsoperator, der bei allen üblichen Selektionsdrücken divergiert, sich bei reduzierter Mutabilität der Schrittweiten als bester herausstellt, ist nicht vorherzusehen und wäre daher bei systematischen Tests eines „Evolutionsstrategen“ sicher ebenso verborgen geblieben wie die Tatsache, daß die Plus–Selektion sowohl beim Skalierungstest als auch bei multimodalen Problemen durchaus eine Existenzberechtigung hat, sofern man ihr über die anderen Parameter einen geeigneten Rahmen verschafft. Die eindeutige Bevorzugung der Komma–Selektion in der Literatur konnte nicht bestätigt werden.

Ebenso sollte die zukünftige Erwähnung der Heuristik für die Parameter τ_0 und τ_i mit dem Hinweis versehen werden, daß es bei Verwendung von n Schrittweiten Einstellungen gibt, die deutlich besser konvergieren. Falls die Bemühungen um eine Derandomisierung der ES Erfolg haben, könnte die klassische Schrittweitenmutation allerdings abgelöst und die τ–Parameter dadurch obsolet werden.

Die ES–Variante mit Rotationswinkeln kann nicht mehr als letzte Zuflucht gelten, will man Korrelationen zwischen den Objektvariablen etablieren. Die in Kapitel 4 geäußerten Zweifel an ihrer Funktionsweise, die dazu führten, diese Variante bei den folgenden Untersuchungen unberücksichtigt zu lassen, bestätigten sich durch die Ergebnisse der Meta–ES.

Generelle Empfehlungen für die ES–Parametrisierung lassen sich leider aus den Ergebnissen dieses und des vorherigen Kapitels trotz der sehr einfachen Zielfunktionen nicht ableiten. Nicht einmal die intermediäre Schrittweitenrekombination, die sich in allen Tests

[4] Für eine kritische Würdigung des unterstellten Optimierungszenarios sei auf (Droste, Jansen und Wegener, 1998) verwiesen.

durchsetzte, kann aufgrund der Voruntersuchung uneingeschränkt empfohlen werden (s. Abbildung 5.4). Bei den τ–Parametern erhält man manchmal durch ihre Vergrößerung bessere Konvergenzgeschwindigkeiten, bei der Doppelsumme aber durch eine drastische Verkleinerung.

Natürlich wäre es schöner gewesen, allgemein gültige Einstellungen zu finden, aber daß es diese nicht gibt, ist immerhin auch ein Ergebnis. Die Leistungsreserven, die in diesen Verfahren stecken und nur über ihre Parametrisierung, also ohne eine algorithmische Änderung an den ES selbst, abgerufen werden können, sind so beträchtlich, daß bei einer neuen Zielfunktion vor einer Modifikation des Algorithmus zunächst untersucht werden sollte, ob nicht doch eine ES–Parametrisierung existiert, die das Problem hinreichend schnell löst.

Hat man sich aber entschieden, den Algorithmus zu modifizieren, und möchte man die Leistungsfähigkeit neuer genetischer Operatoren nicht durch eine Gegenüberstellung mit einzelnen anderen Parametrisierungen testen, bietet die Meta–ES einen geeigneten Rahmen, in dem sich neue Ideen bewerten lassen.

Von der Detailkritik bleiben grundlegender Ansatz und generelle Funktionsfähigkeit der Evolutionsstrategien natürlich unberührt. Im Gegenteil, sie bieten sogar selbst den Mechanismus zur Überwindung der Probleme an — wenn auch durch eine aufgesetzte Ebene.

Kapitel 7

Evolutionsstrategien jenseits der Optimierung

Large complex systems are beyond human capacity to evaluate.

Verwendet man Evolutionsstrategien nicht als Problemlöser, sondern betrachtet sie als einfache Modelle natürlicher Prozesse, können sie auch in anderen Wissenschaftsgebieten nützlich sein. Gemeint ist dabei kein „Lernen von der Natur", sondern eher ein disziplinübergreifendes „Lernen von Theorien über die Natur". Formale Modelle der Naturwissenschaften auf ökologische und sozioökonomische Fragestellungen zu übertragen, bedarf natürlich einer Methodenreflexion, die allerdings aus dem Rahmen dieser Arbeit herausfällt. Ein Vorschlag zur Gründung eines trans– und interdisziplinären „Instituts für Evolutionswissenschaft" existiert bereits (Stichweh, Reyer und Uszkoreit, 1999). Vorbild war dabei das 1984 gegründete Santa Fe Institute, das sich mit der Erforschung komplexer Systeme beschäftigt.

7.1 Biologie

Aufgrund des biologischen Vorbildes liegt es nahe, ES als einfaches Evolutionsmodell zu betrachten und am Modell gewonnene Erkenntnisse am Vorbild zu überprüfen.

Auch in der Natur gibt es Strategieparameter, die die Lebensfähigkeit eines Individuums beeinflussen. Auf molekularer Ebene sind dies Mutatorgene und Reparaturenzyme, die Punktmutationen vermehrt hervorrufen oder rückgängig machen können, auf einer höheren Ebene sorgt die Streßmodulation des Immunsystems für eine erhöhte Infektdisposition z. B. bei Überbevölkerung und somit für eine Anpassung der „Schrittweite".

Die überall beobachtbare endliche Lebensdauer stellt kein Versagen der Natur dar, sondern einen Mechanismus, um einer Art die Anpassung an wechselnde Umweltbedingungen zu ermöglichen. Nur eine Komma–Strategie (s. Kapitel 4.5), die die Elterngeneration nach einem Reproduktionszyklus verwirft, kann einem zeitabhängig wandernden Optimum folgen, während eine Plus–Strategie versagt, die an einmal gefundenen „guten“ Phänotypen festhält (Kursawe, 1996).

Polyploidie (mehrfache Chromosomensätze) lohnt sich in einer konstanten Umwelt nicht, während sie in einer sich ändernden Umwelt sogar nötig ist. Für bestimmte Pflanzen gilt im nacheiszeitlichen Europa sogar: Polyploidiegrad $\hat{=}$ Breitengrad. Dieses Phänomen läßt sich in Evolutionsstrategien wie folgt reproduzieren: Die Optimierung mit nur einer Zielfunktion entspricht aus der Sicht eines Individuums einer konstanten Umwelt, in der jeder zusätzliche Chromosomensatz Ballast wäre. Modifiziert man die Selektionskomponente der Strategie jedoch geringfügig, kann man auch Probleme mit mehrfacher Zielsetzung lösen. Dabei unterliegt dann allerdings die Umwelt aus der Sicht eines Individuums recht drastischen Änderungen, da jeweils eine andere Zielfunktion als Selektionskriterium „ausgewürfelt“ wird, so daß hier polyploide Wesen vorteilhaft sind (Kursawe, 1990; Kursawe, 1991).

Die „Erfindung“ der sexuellen Rekombination kann mittels ES–Simulation nachvollzogen werden: Insbesondere in Gebirgen wie in Abbildung 5.1 (rechts) und 5.3 erweist sich die Rekombination des vorhandenen genetischen Materials, d. h. der Informationsaustausch zwischen Individuen als notwendig, um eines der besseren lokalen oder sogar das globale Optimum zu finden.

Phänomene wie das Massensterben vieler Arten in einer Epoche müssen nicht unbedingt exogene Ursachen wie einen Meteoriteneinschlag oder Vulkanausbrüche haben, sie können auch durch das System selbst hervorgebracht werden, das die auch zu Zeiten scheinbaren Gleichgewichts ständig auftretenden kleinen Frequenzänderungen plötzlich anfacht. Dies kann zu neuen quasi–stabilen Zuständen führen, aber auch zum Verschwinden des Systems. Zum Begriff *Punctuated Equilibrium* und dem Bezug zu ES siehe auch Seite 93.

7.2 Bionik

Es zeugt von der Hybris der Menschen, sich einerseits als Krone der Schöpfung oder auch nur als Homo Sapiens zu bezeichnen, aber die Prinzipien und Produkte jener Natur, die auch uns schließlich hervorgebracht hat, nicht konsequenter zu nutzen. Die *Bionik* macht sich zwar schon seit Leonardo da Vincis Zeiten „Produkte“ der Natur zunutze, den Begriff gibt es allerdings erst seit 1960, und gemessen an den etablierten Ingenieurdisziplinen fristet sie eher ein Randdasein.

Natur und Technik sind nur auf den ersten, flüchtigen Blick Gegensätze. Schließlich unterliegen alle Lebewesen und die später dazugekommenen technischen Konstrukte denselben Naturgesetzen. Ferner gibt es auf der Erde nur eine relativ kleine Menge an chemischen Elementen, die in nennenswerter Menge, aber natürlich begrenzt, verfügbar ist. Der Mensch hat die Vielfalt an Substanzen zwar beträchtlich erweitert, dabei allerdings nicht auf kreislauffähige Produkte geachtet, die nach ihrer Verwendung nicht auf Deponien endgelagert und somit dem natürlichen Kreislauf entzogen werden müssen, sondern in Stoffe zerfallen, die unter Zufuhr von Energie erneut verarbeitet und genutzt werden können.

Die Nachhaltigkeitsdiskussion wird m. E. zu sehr auf Ressourcen– und Schadstoffprobleme reduziert, die Rolle des Produktivfaktors Humankapital, d. h. des Wissens wird vernachlässigt. Denn im *stofflich* geschlossenen System Erde gibt es kein Rohstoffproblem, sondern „nur" das Problem, verbrauchte Stoffe mit Hilfe von Information und Energie wieder in einen nutzbaren Ordnungszustand zu bringen. Dies geschieht allerdings zu 88% aus dissipierenden fossilen Energieträgern. Die Sonne liefert Tag für Tag etwa 100–mal mehr Energie als auf der Welt verbraucht wird, nutzbar in Form von Wind–, Wellen–, Biomasse– oder direkten Solarkraftwerken. Die Natur hatte keine Alternative zur Solarenergie, die Bionik kann also einen wichtigen Beitrag zur Wissensvermehrung leisten.

Als inhärent multidisziplinäre Wissenschaft hat die Bionik natürlich mit der zunehmenden Spezialisierung in der heutigen Wissenschaft zu kämpfen. Zerbst wirft die berechtigte Frage auf, „wie eine solche interdisziplinäre Forschung vorstrukturiert und systematisiert werden kann, damit sie noch effektiver zu Ergebnissen kommt, als dies bei zufälligen Interessenverknüpfungen der bislang in diesem Felde tätigen Forscher der Fall ist" (Zerbst, 1987). Dazu kommt die Schwierigkeit, daß die Natur keine Blaupausen liefert, sondern nur Anregungen.

Nach Nachtigall lassen sich folgende Teilgebiete der Bionik unterscheiden (Nachtigall, 1992a):

In der *Historischen Bionik* werden alle jene Ideen bis zum 19. Jahrhundert zusammengefaßt, die sich als Versuche einer Übertragung Natur – Technik interpretieren lassen, so insbesondere die vielen Nachahmungsversuche des Vogelfluges mittels Schlagflügeln.

Angefangen mit traditionellen Baumaterialien (Ton), über Leichtbaukonstruktionen (Kieselalgenschalen), Seilkonstrukte (Spinnennetze), ideale Flächenbedeckungen (Blattüberlagerungen), optimale Flächennutzung (Wabenprinzip) bis hin zu passiver Lüftung, Heizung und Kühlung (Präriehundbauten) finden sich in der *Struktur–, Bau– und Klimabionik* viele Möglichkeiten, Biologie und Architektur einander zu nähern.

Die *Konstruktionsbionik* versucht, Elemente und Mechanismen der Natur herauszuarbeiten, die zu integrierten Lösungen geführt haben, bei denen das einzelne Element oft mehr als eine Aufgabe übernimmt. Das bekannteste Beispiel aus diesem Bereich dürften Klettverschlüsse sein, die heute an vielen Produkten zu finden sind. Hat man das Prinzip des Baum– und Knochenwachstums verstanden, so lassen sich dann nach dem gleichen Schema auch technische Bauteile unter Last im Rechner „züchten“, indem man Material nur dort beläßt oder anlagert, wo die auftretenden Kräfte dies auch wirklich erfordern. Umgekehrt wird dort Material weggenommen, wo es nicht (mehr) benötigt wird (Mattheck, 1992). Auf diese Weise ist es gelungen, eine orthopädische Schraube 105–fach belastbarer zu machen oder die Lebensdauer einer zylindrischen Welle auf das 40–fache zu steigern.

Die *Bewegungsbionik* studiert das Laufen, Schwimmen und Fliegen in der Natur, um aus den Anpassungen der Bewegungsorgane an das jeweilige Medium Fragen der Strömungsanpassung, des Antriebs und der Effizienz von Organen zu beantworten. Klebt man eine künstliche Haihaut auf Rumpf und Tragflächen von Flugzeugen, so lassen sich durch die Widerstandsreduktion Kerosineinsparungen von 1 bis 3% erzielen, ohne an der Flugzeugkonstruktion etwas ändern zu müssen. Orientierte man sich allerdings schon bei der (Rumpf–)Konstruktion an der Natur, wären Passagierkapazitäten von 150% bei gleichem Treibstoffeinsatz möglich. Die entscheidende Idee dabei ist, nicht den Vogelflug zu kopieren, sondern Lebewesen als Vorbild zu nehmen, die sich zwar in einem anderen Medium, aber in einem ähnlichen Reynoldszahlbereich wie Flugzeuge bewegen wie z. B. Haie und Schwertfische (Videler, 1992).

Die Untersuchung der aufgespreizten Flügelenden großer (Raub–)Vögel hat dagegen zum Bau eines „Windkonzentrators“ geführt, der eine kleine Turbine in schnelle Rotation versetzt und so die im Wind enthaltene Energie besser nutzen kann als die heute üblichen großen Anlagen mit mehreren Tonnen bewegter Masse (Göpfert, 1988).

Auch im Bereich der Bootsantriebe gibt es Untersuchungen, ob Schiffe nicht effizienter über eine Art Schwanzflosse angetrieben werden können. Ein Patent auf einen Flossenantrieb wurde bereits 1903 erteilt. Dieser Antrieb wies bereits den vierfachen Schub gegenüber einer konventionellen Schiffsschraube auf, so daß er insbesondere für Schlepper geeignet wäre. Jüngere Arbeiten auf diesem Gebiet legen die Vermutung nahe, daß ein solcher Antrieb für langsame Schiffe erheblich effizienter ist als die heute verwendeten Schraubenantriebe (Blickhan, 1992). Zusätzlich könnte ein Flossenantrieb die Wellenenergie zum Vortrieb nutzen, falls der Antrieb nicht mehr unabhängig von der Rumpfkonstruktion betrachtet wird, sondern beide als Einheit entworfen werden. Eine imitierte horizontal schwingende Schwanzflosse führt umgekehrt zu einer Flossenpumpe, die unempfindlich gegen Lufteinschlüsse oder zähflüssige Stoffe ist.

Obwohl schon seit über hundert Jahren im Prinzip verstanden, ist es erst kürzlich gelungen, die Flugtechnik des Albatros, mit der er über 5 000 km pro Woche durch Ausnutzung des Windgradienten fast ohne einen Flügelschlag zurücklegen kann, per Software nachzubilden und für Segelflugzeuge verfügbar zu machen (Frauendienst, 1994).

Die *Sensorbionik* sucht nach Möglichkeiten, Umweltreize physikalischer und chemischer Art möglichst in Echtzeit zu messen. So basieren die heute erhältlichen elektronischen Zollstöcke auf der Nachahmung der Ultraschall–Beuteortung von Fledermäusen, die aus der Laufzeitverschiebung ihrer Signal–Echos Beute sowie Hindernisse lokalisieren können.

Das Gebiet der *Anthropobionik* beschäftigt sich mit dem weiten Feld der Mensch–Maschine–Interaktion. Das Ziel besteht darin, die Maschinenschnittstellen den sensorischen Eigenschaften von Menschen anzupassen. In dieses Gebiet fallen ferner einige Aspekte aus der Robotik. So kommt man z. B. zu einer ruckfreien Roboterbewegung, indem man pneumatisch angetriebene Stellglieder verwendet, die der Funktionsweise von Muskeln nachempfunden sind.

Unter *Neurobionik* faßt man diejenige Informationsverarbeitung zusammen, die intelligente Verschaltungen verwendet. In den Gehirnen vieler Lebewesen gibt es einen Mechanismus („laterale Inhibition“), der die Kontraste an Hell–Dunkel–Grenzen verstärkt. Diaprojektoren, die selbständig fokussieren, imitieren dieses biologische Prinzip (Rechenberg, 1973).

Um die Erforschung chemisch–physikalischer Verfahren kümmert sich die *Verfahrensbionik*. Besonders interessant sind heutzutage die Rezyklierbarkeit aller verwendeten Stoffe sowie zunehmend die Photosynthese als Einstieg in die Wasserstofftechnologie. Warum erst die Energie des Sonnenlichtes in Strom umwandeln, der dann mangels effizienter Speichermöglichkeiten Wasser in seine Bestandteile Sauer– und Wasserstoff zerlegt? Solche Anlagen dürften prinzipiell kaum über einen Gesamtwirkungsgrad von 30% hinauskommen. Sinnvoller erscheint auch hier, sich direkt an der Photosynthese zu orientieren, einem Verfahren, das seit einigen Milliarden Jahren auf der Erde verwendet und optimiert wird. Oder man sucht wie z. B. Rechenberg (Rechenberg, 1994) nach Bakterien, die unter Sonneneinstrahlung direkt Wasserstoff produzieren.

Während bei den bisher erwähnten Teildisziplinen die Erforschung und Übertragung von Verfahren und „Produkten“ aus der Natur im Vordergrund steht, beschäftigt sich die *Evolutionsbionik* mit jenen Prinzipien, die zu unserer heutigen Welt geführt haben. In dieses Gebiet fallen die Evolutionären Algorithmen (s. Kapitel 4).

7.3 Evolutorische Ökonomik

Die Beschränktheit von Ressourcen und damit der Zwang, bestimmte Optionen zu selektieren, ist universell. Selbstorganisationsfähigkeit kann daher als das konstitutive Merkmal ökonomischer und ökologischer Systeme angesehen werden. Die Lehre von der *Synergetik* (syn – zusammen, ergon – Werk) als formaler Modellrahmen für Selbstorganisationsprozesse versucht, entstehende Ordnung aus dem jeweiligen System selbst heraus (endogen) zu erklären[1]. In der Ökonomie hat Adam Smith mit dem Bild der *Invisible Hand* bereits sehr früh die Idee sich selbst organisierender Märkte vermittelt, in denen individuelle Entscheidungen ein geordnetes System von Märkten entstehen lassen können. Bei dieser Idee handelt es sich allerdings noch nicht um eine Theorie darüber, wie der Prozeß der Selbstorganisation abläuft.

Auch hat sich das Interesse der Ökonomen während der Entwicklung der *Neoklassik* auf die Analyse von Marktgleichgewichten konzentriert bzw. auf die statisch–komparative Analyse der Änderung von Systemvariablen, wenn sich exogene Einflüsse ändern. Der Koordinationsprozeß individueller Handlungen wurde zugunsten der Existenz– und Stabilitätsanalyse von Marktgleichgewichten vernachlässigt. Im Zuge der Differenzierung der Neoklassik wurden zwar zahlreiche Modifikationen in ihren Theorierahmen integriert wie z. B. Unsicherheit, institutionelle Regelungen, Marktversagen, . . . , der abstrakte Referenzpunkt sich unendlich schnell einstellender und prinzipiell reversibler Tauschgleichgewichte sowie der Glaube an rationales, wissensbasiertes Handeln blieb jedoch erhalten.

Seit einigen Jahren versuchen die Anhänger der *Evolutorischen Ökonomik*, der Neoklassik mit ihren individualistischen Nutzenmaximierungsansätzen und den an der klassischen Mechanik orientierten Gleichgewichtskonzepten eine konzeptionelle Alternative mit der Biologie und der Synergetik als Leitbildern entgegenzustellen, in der die Zeit erstmals irreversibel ist.

Dieser neue Ansatz konzentriert sich zunächst auf eine Theorie der Entstehung und Diffusion von Innovationen, da dies eine Schwachstelle der Neoklassik ist: Einem Wirtschaftssubjekt sind nicht mehr alle für die Entscheidung relevanten Alternativen vorgegeben, sondern es generiert sich mittels seiner Vorstellungskraft viele dieser Möglichkeiten selbst. Die der evolutorischen Ökonomik zugrundeliegende Annahme, daß alles vernetzt ist und ständig Verhaltensrevisionen im Zuge von Lernprozessen stattfinden — Lernen jetzt nicht mehr als Anpassung an ein Gleichgewicht, sondern als Entdecken neuer Handlungsmöglichkeiten — ermöglicht gerade eine weitgehende Endogenisierung der entsprechenden Modelle. Für diesen Ansatz spricht auch die Beobachtung, daß es in der Startphase neuer Industrien keineswegs einer Fülle von „Geistesblitzen“ bedarf. Es reicht aus, „alte“ Lösungen auf neue Weise zu neuen Entitäten zu kombinieren — genau das findet in ES mittels der Rekombination statt.

Hodgson (Hodgson, 1993) bietet einen guten Überblick über die Entwicklung von mechanistisch inspirierten Modellen hin zu evolutionären, wobei die biologische Evolution nicht

[1] Jene „Synergie“, die nach Firmenfusionen so häufig bemüht wird, hat damit wenig zu tun und bezeichnet oft nur quantitative Aspekte wie die Kostenreduktion durch den gemeinsamen Einkauf, Zusammenlegung von Teilen der Produktion oder den Zuwachs an Marktanteilen. Die Fusion der verschiedenen Unternehmenskulturen dürfte dagegen schwieriger zu bewerkstelligen sein.

als Optimier– sondern als Adaptationsverfahren[2] interpretiert wird; Adaptation an eine (Um-)Welt, die nicht zuletzt durch die in ihr handelnden Agenten verändert wird. Handlungsweisen, die profitabel für wenige sind, müssen das nicht auch für alle sein. Ferner kann eine kritische Masse analog der Sandhaufenlawine (Bak, Flyvbjerg und Sneppen, 1994) auftreten, ein Phänomen, das Witt (Witt, 1989) für die Emergenz sozialer Institutionen untersucht hat. Nelson und Winter interpretieren Arbeitsabläufe in Firmen als deren Gene (Nelson und Winter, 1982), Holland und Miller versuchen dagegen, GA bzw. Classifier–Systeme als ökonomische Agenten zu interpretieren (Holland und Miller, 1991).

Statt ein Konzept wie das der *Bounded Rationality* in die Neoklassik einzuführen, das die idealistischen Annahmen intentionaler Individuen zu überwinden sucht, kann man individuelles Verhalten auch (rein deskriptiv!) stochastisch[3] modellieren. Akteure können sich irren, situative Faktoren können eine Abweichung vom im Mittel „vernünftigen“ Verhalten bewirken. Existiert zusätzlich ein Selektionsschritt (eine Externalität), kann trotz der Stochastik auf der Mikroebene eine Ordnung auf der Makroebene entstehen.

Dieses Verhalten läßt sich auch an ES beobachten: Stagnationsphasen und Phasen schnellen Wandels entstehen ohne äußere Einflußnahme aus dem Zusammenspiel von zufälligen und deterministischen Operatoren — ein Phänomen, für das die Biologen den Begriff *Punctuated Equilibria* verwenden. Aus der Sicht eines ES–Individuums geht es darum, gute neue Punkte im Suchraum aufzufinden, d. h. „innovativ“ zu sein, und diese zusammen mit der Information, wie diese Innovation zustande gekommen ist, an Nachkommen weiterzugeben. In ES verändern die Individuen ihre Umwelt zwar nicht, betrachtet man aber das Sharing–Konzept aus dem Bereich der GA (s. Kapitel 4.2), dann gibt es dort einen Mechanismus, der das Phänomen begrenzter Ressourcen innerhalb einer Nische modelliert. Aus der Sicht eines Individuums stellt die Übervölkerung einer solchen Nische eine Umweltänderung dar, die es selbst (mit–)hervorgerufen hat.

Ohne Innovationen kommt es in einem Markt zu einer Angleichung der Fähigkeiten, mit Innovationen beobachtet man einen eigendynamischen Wettbewerb unter Beibehaltung des Angebotsgefälles. Zu einem funktionierenden Markt gehören Sanktionen, Innovationen und die Diffusion von Wissen. Obsiegt nur ein Wettbewerber (oder wenige), degeneriert der Wettbewerb, und das Wachstum verlangsamt sich. Je größer die Diversität, desto größer das Wachstum. Ersetzt man die ökonomischen Begriffe Sanktion, Innovation und Wissensdiffusion durch die passenden biologischen (Selektion, Mutation, Rekombination), kann man diesen Absatz in jeden Artikel über ES einfügen.

[2]Diese Unterscheidung stellt keinen Gegensatz dar, sondern nur unterschiedliche Sichtweisen: Bei Optimierverfahren steht üblicherweise das Ergebnis im Vordergrund, betont man die Adaptationsfähigkeit, ist der Weg das Ziel.

[3]Zufallsgeneratoren auf Digitalrechnern erzeugen natürlich keinen „reinen“ Zufall, sondern nur deterministisches Chaos. Diese Tatsache ändert aber nichts an der Gegensätzlichkeit von Zufall und Determiniertheit in der Argumentation.

Bei einer neoklassischen Interpretationsweise ist der Startpunkt relativ egal, das System wird sich immer auf dasselbe Gleichgewicht einstellen. Aber mit dem Eindringen von Ideen aus der nichtlinearen Dynamik wird auch in der Ökonomie das Konzept der *Pfadabhängigkeit* diskutiert. Als ein klassisches Beispiel wird häufiger der Sieg des VHS–Videosystems über das technisch überlegene Betamax–System angeführt (Arthur, 1988). Neben den Startwerten spielt auch eine Rolle, ob ein bestimmtes Verhalten eine kritische Masse in der Population erreicht. Das Scheitern einer Strategie oder eines Produktes muß also nicht aus einer inhärenten Schwäche resultieren, sondern kann auch mit fehlender Masse (Durchsetzungsfähigkeit) zu tun haben (Axelrod und Dion, 1988). Diese Idee der *Frequenzabhängigkeit* stellt eine weitere Neuerung der evolutorischen Ökonomik gegenüber der Neoklassik dar. In Axelrods Kunstwelt des iterierten Gefangenen–Dilemmas ist eine Population, die Tit–for–Tat (zuerst kooperieren, dann den letzten Zug des Gegners kopieren) spielt, verwundbar durch eine Invasion von Kooperateuren, die ihrerseits wieder anfällig für eine Invasion defektierender Strategien ist. Wenn die Immigrationsrate hoch genug ist, kann sich danach wieder Tit–for–Tat durchsetzen. Dieses (Spielwelt–)Beispiel mag die Relativität der Begriffe „Survival“ und „Fitneß“ illustrieren (s. Kapitel 2).

Auch dieses Verhalten läßt sich in ES beobachten: Bei Komma–Selektion haben Revolutionäre nur dann eine Chance zur Weitergabe ihres Wissens, wenn sie sich mit anderen Revolutionären paaren können.

Die Wahrscheinlichkeit eines Zustandswechsels wird mittels Übergangsraten modelliert, wobei diese i. d. R. Funktionen mit weiteren Parametern sind, hinter denen Verhaltenshypothesen stecken. Eine Übergangsrate zwischen zwei Zuständen kann z. B. auch von der Häufigkeit abhängen, mit der die beiden Alternativen bisher von den anderen Akteuren gewählt wurden. Diese Beziehung der Akteure untereinander bringt ein neues Element in die Modellierung ein: Auf diese Weise lassen sich Phänomene wie die Orientierung an Trendsettern, sozialer Anpassungsdruck, Zeit, ... abbilden sowie auch die in der Neoklassik häufig verwendeten Parameter Preis und Nutzen. Die jeweilige Lösung der Master– (für diskrete Zustände) oder Fokker–Planck–Gleichung (kontinuierlich) macht dann eine Aussage über die Entwicklung von Wahrscheinlichkeitsverteilungen, stellt also keine Trajektorie im üblichen Sinne mehr dar (Haag, 1990; Weidlich und Braun, 1992).

Modelliert man komplexe nichtlineare Systeme entsprechend, geht damit ein Verlust der exakten Vorhersagbarkeit einher, und stabile Zustände können instabil werden. Dieses „Chaos“ wird oft als Bedrohung empfunden. Aber sollte das Gewicht beim Studium natürlicher Systeme nicht eher auf qualitativer Erklärung denn auf quantitativer Prognose liegen? Und legen die weltweiten Turbulenzen an den Börsen nicht den Gedanken nahe, daß hochgradig geordnete Systeme mit starren Regeln auf Störungen sehr viel empfindlicher reagieren? Die Natur lehrt uns, daß Chaos kreativ, strukturbildend und –erhaltend sein kann. Chaotische Strukturen finden sich z. B. in der Herzfrequenz, in den Schwankungen der Menge weißer Blutkörperchen, bei Stoffwechselprozessen und Hormonkonzentrationen. Ökonomen, die mit dem Bild rationaler Individuen groß geworden sind, haben Schwierigkeiten mit der Vorstellung, daß Entscheidungsprozesse auch nichtdeterministisch beschrieben werden können. Wie Simulationen von Huberman zeigen, reicht eine gewisse Verhaltensdiversität — erzeugt durch einen Pseudo–Zufallszahlengenerator — schon aus, um kollektive Phänomene hervorzubringen (Huberman und Glance, 1993).

Bei der Übernahme von Begriffen in die Ökonomie gibt es besonders mit der Synergetik Schwierigkeiten: Physiker haben keine Probleme damit, von einer „Versklavung“ der schnell veränderlichen durch langsamer reagierende Größen eines Systems zu sprechen — bei einem Laser versklavt die ausgestrahlte Lichtwelle die angeregten Elektronen. Übertragen auf soziologische Prozesse wie z. B. das Verhältnis Minister/in zur Ministerialbürokratie, dürfte der Boden für Mißverständnisse bereitet sein. Hier spräche man wohl besser von Konsensualisierung oder Konsensbildung, was den Vorgängen in einem Laser durchaus auch angemessen, Physikern aber wohl zu anthropomorph wäre.

Einen zusätzlichen Hemmschuh bei der weiteren Verbreitung der evolutorischen Ökonomik kann ferner die Historie darstellen: Gerade die Biologie ist in der Vergangenheit von den Sozialwissenschaften mißbraucht worden. Erinnert sei an den Sozialdarwinismus. Wegen dieser schlechten Erfahrungen sind viele Menschen heutzutage sehr zurückhaltend, was den Einsatz biologischer Metaphern bei der Ablösung der mechanistischen angeht. Erschwerend kommt natürlich hinzu, daß die Biologie gar kein einheitliches Theoriegebäude präsentieren kann (s. Kapitel 2) und ihrerseits auch keineswegs frei ist von mechanistischen oder reduktionistischen Methoden.

Außerdem gibt es diametral entgegengesetzte Konkurrenz aus der Soziologie: Individuelles Verhalten ist durch Erziehung, Normen und Rollenerwartungen geprägt oder sogar determiniert. Die geringe Bedeutung des Individuums ist der Preis für die Komplexität. In Luhmanns autopoietischen Systemen kommen Individuen sogar nur noch als Umweltressourcen vor, von einer Theorie als Letztbegründung hat er sich verabschiedet. Für ihn besteht sie aus einem ständigen Umgehen mit Differenzen, auf ein Fundament wird verzichtet: Man beschreibt, was vorgeht. Die Gesellschaft als Ganzes bildet ein System, dessen Teile sich wechselseitig hervorbringen. Aus diesem Theorieprozeß (nicht –gebäude) folgt, daß es keine universellen Rezepte gibt — für einige ein schwer zu akzeptierender Gedanke. Theorie als externes Werkzeug zur Reduktion von Komplexität und damit zum Verständnis eines Systems gibt es bei Luhmann nicht mehr, sie gehört zu dem System, das sie zu verstehen sucht. Theorie ist nur noch eine Anleitung zum Sehen (Luhmann, 1997).

7.4 Organisationspsychologie

Darwins Annahme, daß eine Art nur überlebt, falls sie sich ihrer Umwelt anpassen kann, hat auch Psychologen inspiriert, die sich mit der Organisationsstruktur von Unternehmen beschäftigen. Warum sollte genau *eine* Organisationsstruktur, die noch dazu über der Zeit konstant gehalten wird, hinreichend für die diversen Anforderungen einer sich ständig ändernden Umwelt sein? Aus Studien über Kommunikationsnetze weiß man, daß zentralistisch organisierte Netze (ein „Führer“, viele „Geführte“) bei einfachen Aufgaben leistungsfähiger sind als dezentrale. Bei komplexen Aufgaben verhält es sich jedoch genau umgekehrt: Hier sind dezentrale Netze angemessener, weil sie fehlerfreundlicher und eher rückholbar als zentrale sind.

Formale Hierarchien in Firmen stehen der problemangepaßten Gruppenbildung aber normalerweise im Wege. Dieser situative oder Kontingenzansatz hat ferner festgestellt, daß bürokratische Organisationsformen in statischen Umgebungen florieren, während in dynamischen Umwelten häufig die „organisch“ gewachsenen Organisationsstrukturen innewohnende Flexibilität benötigt wird, die sich auf gut ausgebildete und selbständig agierende Mitarbeiter/innen stützt. Hier treffen sich also die Ideen der Organisationspsychologen mit denen des *evolutionären Managements*, das ein Unternehmen als lernendes System betrachtet, in dem die Information das Kapital als Schlüsselressource abzulösen beginnt, in dem eine dezentrale Selbstorganisation die zentrale Weisungskultur ersetzt und die Zukunft eines Unternehmens auch davon abhängt, wie gut man innerbetrieblich und mit der Konkurrenz zusammenarbeitet. Das „Wissen“ einer Firma ist ein soziales Produkt, das durch die Koevolution von Gruppenstruktur und sozialer Intelligenz hervorgebracht und durch sozialen Austausch idealerweise ständig erweitert und korrigiert wird.

In diesen Rahmen passen auch die Begriffe *Lean Management* und *Lean Production*, deren Grundidee ebenfalls flachere Hierarchien sowie die Integration von Aufgaben und Verantwortung ist. Allerdings wurden diese Begriffe speziell in Deutschland als Rechtfertigung für Entlassungen und „Gesundschrumpfen“ verwendet. Dies kann kurzfristig sicher die Effizienz erhöhen, wird aber auf längere Sicht die Anpassungsfähigkeit eines Unternehmens einengen, wenn man dem alten kybernetischen Postulat der *Requisite Variety* vertraut, welches verlangt, daß die interne Diversität jedes Systems der Komplexität seiner Umwelt entsprechen muß, um deren Veränderungen adäquat verarbeiten zu können.

In ES läßt sich dieses Verhalten wie folgt nachbilden: In einer anspruchslosen Umwelt wie z. B. der des Kugelmodells (s. Abbildung 5.1) reicht es, wenn man eine kleine Population mit harter Selektion auf die Suche schickt. Will man jedoch das globale Optimum in einer zerklüfteten Topologie finden, muß man mit einer großen Population und weicher Selektion arbeiten. Fehlt diese Redundanz, konvergiert das Verfahren weit ab vom globalen Optimum. Dies passiert auch, wenn man zwar eine hinreichend große Population mit weicher Selektion benutzt, aber die Rekombination, d. h. Kooperation ausschaltet, so daß die Individuen als Einzelkämpfer unterwegs sind. Erst die Rekombination als Austauschoperator für Wissen ermöglicht die Beobachtung von synergistischen Lerneffekten: In der künstlichen Umwelt auf dem Rechner ist es möglich, für bestimmte Zielfunktionen den schnellsten Weg zum „Gipfel“ analytisch zu berechnen und eine Population mit diesem perfekten Wissen über ihre Umwelt auszustatten. Bei einer solchen Vorgehensweise verliert die Rekombination natürlich ihre Bedeutung, weil alle Individuen ständig über vollständiges Wissen verfügen. Interessant ist dann der Vergleich mit einer Strategie, die ohne dieses Vorwissen startet und sich die Informationen über ihre Umwelt selbst beschaffen muß. Diese Population von „Dummen" erreicht als Kollektiv mit Rekombination fast genauso schnell das Ziel wie die „Experten" (Schwefel, 1987).

7.5 ... und der Rest

Auch in der Psychologie wird unter dem Begriff *Systemische Familientherapie* an einem evolutionär motivierten Handlungsmodell gearbeitet. Die Familie wird dabei als ein sich von der Umwelt abgrenzendes System interpretiert, in das der Therapeut/die Therapeutin nicht direkt eingreifen kann.

Zwischen Soziologie und Künstlicher Intelligenz gibt es ein weiteres interdisziplinäres Forschungsfeld — die *Sozionik* (Malsch, 1997). In diesem Fall dienen menschliche Gesellschaften bzw. die von Soziologen entworfenen Kommunikationsmodelle als natürliches Vorbild. Durch Nachahmung gesellschaftlicher Mechanismen sollen bessere Computerprogramme in bezug auf Robustheit, Anpassungsfähigkeit und sogar Kreativität entstehen. In dieses Gebiet fallen somit auch alle Anstrengungen, der Informationsflut im Internet durch (Software–)Agenten zu begegnen.

In gewisser Hinsicht paradox scheint es, von einer *evolutionären Software–Entwicklung* zu sprechen. Natürlich sind gerade Computerprogramme einer sich ändernden Umwelt ausgesetzt, sie „altern" schnell. Aber kann man bei ihrer Erstellung geplant evolutionär vorgehen (Floyd, Krabbel, Ratuski und Wetzel, 1997)? Nein, gemeint ist hier nur eine erhöhte situative Anpassungsfähigkeit, der Einbau von Rückkopplungen bereits in frühen Projektstadien gegenüber dem traditionellen Vorgehen in sequentiellen, linearen Phasen.

Coy führt die Software–Krise weniger auf logische oder Entwurfsmängel zurück als vielmehr auf eine „unzureichende Reflexion des Wechselspiels von technischer Gestaltung und sozialer Wirkung informationstechnischer Systeme" (Coy, 1989). Jedes komplexe System sollte langsam und folgemäßig aufgebaut werden, indem man stets einen Schritt (oder wenige Schritte) dem Bestehenden hinzufügt. Und man sollte sich stets bewußt sein, daß dasselbe System in einem Bruchteil der zu seiner Errichtung notwendigen Zeit verändert oder sogar zerstört werden kann.

Um zu neuen Erkenntnissen zu gelangen, reicht es oft schon aus, bereits vorhandenes Wissen in einen neuen Bezugsrahmen zu setzen, und vielleicht können Evolutionsstrategien in ganz anderen Bereichen als jenen, für die sie ursprünglich entwickelt wurden, auch einen Beitrag leisten.

Kapitel 8

Ausblick

Perfection can be achieved on the day after the final deadline.

Ziel dieser Arbeit war einerseits die Untersuchung der heuristischen Anteile in Evolutionsstrategien und, aus dieser Untersuchung resultierend, andererseits die Entwicklung eines Verfahrens, das mit Evolution statt Enumeration zu guten Parametrisierungen gelangt. Dies ist gelungen, auch wenn der Meta–Algorithmus sicher noch Raum für Verbesserungen bietet. So erscheint es wünschenswert, die Anzahl der Generationen mit in die Bewertung einzubeziehen. Dies kann über einen mehrkriteriellen Ansatz erfolgen, der Kg_1 und Kg_2 einzeln berücksichtigt, über ein aus diesen beiden kombiniertes oder ein ganz anderes Konvergenzmaß. Auf jeden Fall sollte man nicht durch ein zu groß eingestelltes Generationslimit noch bessere Parametrisierungen ausschließen. Um den erlernten Rekombinationstyp nicht durch $\mu = 1$ wieder zu zerstören, erscheint ferner eine untere Grenze von 2 für die Elternzahl geboten. Interessant wäre auch, wie weit man die Meta–Population verkleinern kann, um mit weniger Simulationszeit zu Ergebnissen gleicher Qualität zu gelangen.

Wünschenswert ist natürlich langfristig eine Vereinigung der Ebenen, weil der Meta–Ansatz für aufwendig zu berechnende praktische Probleme sicher keine Alternative mehr darstellt. Zu diesem Zwecke muß allerdings für jeden oder zumindest einige ES–Parameter eine Konkurrenzsituation etabliert werden. Schlechte Parameter frühzeitig zu erkennen, stellt kein Problem dar, aber gute Parametrisierungen geben sich manchmal erst nach einiger Zeit als solche zu erkennen, so daß die Integration der Ebenen einiges an Arbeit erfordern wird.

Die Auswahl der Testfunktionen bietet naturgemäß immer Anlaß zur Kritik. Schittkowski beschreibt einen Zufalls–Testproblemgenerator, dem man die Lösung $\vec{x}^*$ sowie die Anzahl an Restriktionen vorgeben kann. Auch wenn die Probleme dort quadratischer Natur sind, läßt sich diese Idee vielleicht zu einem allgemeineren Generator weiterentwickeln (Schittkowski, 1980), der dann die Basis zukünftiger empirischer Arbeiten bilden kann.

Unabhängig davon, ob die oben erwähnten Ideen sich realisieren lassen oder nicht, wird die Entwicklung der ES und EA in verschiedenen Richtungen (Repräsentationen, Operatoren, Derandomisierung, ...) natürlich weitergehen. Und bei dieser Entwicklung können die Ergebnisse dieser Arbeit vielleicht als Bestandsaufnahme in dem Sinne dienen,

- daß man nicht leichtfertig eine Weiterentwicklung, die bis $n = 30$ getestet worden ist, auf größere n überträgt.
- daß die isolierte Betrachtung eines Operators dem komplexen Zusammenspiel innerhalb von ES / EA nicht gerecht wird.
- daß auch vermeintlich uninteressante Gebiete des Suchraumes positive Überraschungen bieten können.
- daß es sich lohnen kann, vermeintlich gesichertes Wissen zu überprüfen, auch wenn dieses plausibel klingt und der Intuition leicht zugänglich ist.
- daß „intelligentes“ Verhalten eines gewissen Masses an Unvorhersagbarkeit bedarf.
- daß die Evolution, ob natürlich oder im Rechner, einem Seiltanz gleicht, Abstürze also unvermeidlich sind.

Literaturverzeichnis

Alander, J. T.: An Indexed Bibliography of Genetic Algorithms: Years 1957–1993, Art of CAD Ltd., Espoo, Finland, 1994

Anonymus: Kalter Krieg der Gene, DER SPIEGEL, (17), S. 174–177, 1998

Arthur, W. B.: Self–Reinforcing Mechanisms in Economics, In: *The Economy as an Evolving Complex System*, herausgegeben von P. W. Anderson, K. J. Arrows und D. Pines, S. 9–31, Addison–Wesley, Reading, MA, 1988

Atmar, W.: Notes on the Simulation of Evolution, *IEEE Transactions on Neural Networks*, **5** (1), S. 130–147, 1994

Axelrod, R. M. und Dion, D.: The Further Evolution of Cooperation, *Science*, **242**, S. 1385–1390, 1988

Bäck, T.: *Evolutionary Algorithms in Theory and Practice*, Oxford University Press, New York, 1996

Bäck, T., Fogel, D. B. und Michalewicz, Z. (Hrsg.): *Handbook of Evolutionary Computation*, Oxford University Press, New York, and Institute of Physics Publishing, Bristol, 1997

Bäck, T. und Hoffmeister, F.: Basic Aspects of Evolution Strategies, *Statistics and Computing*, **4** (2), S. 51–63, 1994

Bäck, T., Hoffmeister, F. und Schwefel, H.-P.: A Survey of Evolution Strategies, In: *Proceedings of the Fourth International Conference on Genetic Algorithms*, herausgegeben von R. K. Belew und L. B. Booker, S. 2–9, Morgan Kaufmann, San Mateo, CA, 1991

Bäck, T., Hoffmeister, F. und Schwefel, H.-P.: Applications of Evolutionary Algorithms, Report of the Systems Analysis Research Group SYS–2/92, University of Dortmund, Department of Computer Science, 1992

Bäck, T. und Schütz, M.: Evolution Strategies for Mixed–Integer Optimization of Optical Multilayer Systems, In: *Evolutionary Programming IV: Proceedings of the Fourth Annual Conference on Evolutionary Programming*, herausgegeben von J. R. McDonnell, R. G. Reynolds und D. B. Fogel, S. 33–51, MIT Press, Cambridge, MA, 1995

Bäck, T. und Schwefel, H.-P.: An Overview of Evolutionary Algorithms for Parameter Optimization, *Evolutionary Computation*, **1** (1), S. 1–23, 1993

Bak, P., Flyvbjerg, H. und Sneppen, K.: Can we model Darwin?, *New Scientist*, **141** (1916), S. 36–39, 1994

Bazaraa, M. S. und Shetty, C. M.: *Nonlinear Programming Theory and Algorithms*, Wiley & Sons, New York, 1979

Berndt, C.: Dollys Erblast, DER SPIEGEL, (23), S. 220–224, 1999

Beyer, H.-G.: Toward a Theory of Evolution Strategies: On the Benefits of Sex — the $(\mu/\mu,\lambda)$–Theory, *Evolutionary Computation*, **3** (1), S. 81–111, 1995a

Beyer, H.-G.: Toward a Theory of Evolution Strategies: The (μ,λ)–Theory, *Evolutionary Computation*, **2** (4), S. 381–407, 1995b

Beyer, H.-G.: Towards a Theory of 'Evolution Strategies': The $(1,\lambda)$–Self–Adaptation, Report of the Systems Analysis Research Group SYS–1/95, University of Dortmund, Department of Computer Science, 1995c

Beyer, H.-G.: An alternative explanation for the manner in which genetic algorithms operate, *BioSystems*, **41**, S. 1–15, 1997

Blackmore, S.: *The Meme Machine*, Oxford University Press, Oxford, 1999

Blickhan, R.: Bionische Perspektiven der aquatischen und terrestrischen Lokomotion, In: (Nachtigall, 1992b), S. 135–154

Blickle, T. und Thiele, L.: A Comparison of Selection Schemes Used in Genetic Algorithms, TIK-Report 11, Computer Engineering and Communication Networks Lab (TIK), Swiss Federal Institute of Technology (ETH), Zürich, 1995

Born, J.: *Evolutionsstrategien zur numerischen Lösung von Adaptationsaufgaben*, Dissertation A, Humboldt–Universität, Berlin, 1978

Bremermann, H. J.: Optimization through Evolution and Recombination, In: *Self–Organizing Systems*, herausgegeben von M. C. Yovits, G. T. Jacobi und D. G. Goldstein, S. 93–106, Spartan Books, Washington, D.C., 1962

Coy, W.: Brauchen wir eine Theorie der Informatik?, *Informatik Spektrum*, **12**, S. 256–266, 1989

Dantzig, G. B.: *Lineare Programmierung und Erweiterungen*, Springer, Berlin, 1966

Davidor, Y., Schwefel, H.-P. und Männer, R. (Hrsg.): *Parallel Problem Solving from Nature — PPSN III*, Band 866 der Reihe *Lecture Notes in Computer Science*, Springer, Berlin, 1994

Dawkins, R.: *The Selfish Gene*, Oxford University Press, Oxford, 1976

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

Dennis, J. E. und Schnabel, R. B.: *A View of Unconstrained Optimization*, Kapitel I, Band 1 der Reihe (Nemhauser, Kan und Todd, 1989), 1989

Droste, S., Jansen, T. und Wegener, I.: Perhaps Not a Free Lunch But At Least a Free Appetizer, Technical Report CI–45/98, Sonderforschungsbereich 531 „Computational Intelligence“, Universität Dortmund, 1998

Eigen, M.: Self–Organization of Matter and the Evolution of Biological Macromolecules, *Naturwissenschaften*, **58** (10), S. 465–523, 1971

Floyd, C., Krabbel, A., Ratuski, S. und Wetzel, I.: Zur Evolution der evolutionären Systementwicklung: Erfahrungen aus einem Krankenhausprojekt, *Informatik Spektrum*, **20**, S. 13–20, 1997

Fogel, D. B.: *Evolving Artificial Intelligence*, Dissertation, University of California, San Diego, CA, 1992

Fogel, L. J., Owens, A. J. und Walsh, M. J.: *Artificial Intelligence through Simulated Evolution*, Wiley, New York, 1966

Frauendienst, B.: Konkurrenz für den Albatros, *Bild der Wissenschaft*, (11), S. 120, 1994

Friedberg, R. M.: A Learning Machine: Part I, *IBM Journal*, **2** (1), S. 2–13, 1958

Gall, J.: *Systemantics*, The General Systemantics Press, Ann Arbor, MI, zweite Auflage, 1986

Gärtner, K., Bube, P., Flamme, A., Peters, K. und Pfaff, J.: Komponenten biologischer Variabilität und die Grenzen ihrer Manipulierbarkeit, *Tierlaboratorium der medizinischen Hochschule Hannover*, (18), S. 146–158, 1976

Gasser, R. und Nievergelt, J.: Es ist entschieden: Das Mühlespiel ist unentschieden, *Informatik Spektrum*, **17**, S. 314–317, 1994

Geyer, H., Ulbig, P., Schulz, S. und Bräuer, P.: Vergleich zwischen klassischen und verschachtelten Evolutionsstrategien am Beispiel einer nichtlinearen Regression an Oberflächenspannungen in $\mathbb{R}^2$, Technical Report CI–66/99, Sonderforschungsbereich 531 „Computational Intelligence“, Universität Dortmund, 1999

Goldberg, D. E.: *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison–Wesley, Reading, MA, 1989

Gonick, L. und Wheels, M.: *Genetik in Cartoons*, Paul Parey, Berlin und Hamburg, 1986

Göpfert, W.: Bionik – Lernen von der Natur, *Bild der Wissenschaft*, (4), S. 50–58, 1988

Grefenstette, J. J.: Optimization of control parameters for genetic algorithms, *IEEE Transactions on Systems, Man, and Cybernetics*, **SMC–16** (1), S. 122–128, 1986

Grolle, J. und Klein, S.: Interview mit Stephen Jay Gould, DER SPIEGEL, (10), S. 190–196, 1998

Haag, G.: Die Beschreibung sozialwissenschaftlicher Systeme mit der Master–Gleichung, In: *Individuelles Verhalten und kollektive Phänomene*, herausgegeben von Peter Weise, Nummer 8, Jahrbuch Ökonomie und Gesellschaft, Frankfurt, 1990

Hammel, U. und Bäck, T.: Evolution Strategies on Noisy Functions: How to Improve Convergence Properties, In: (Davidor, Schwefel und Männer, 1994), S. 159–168

Hartl, R. F.: A Global Convergence Proof for a Class of Genetic Algorithms, Technische Universität Wien, 1990

Herdy, M.: Evolution Strategies with Subjective Selection, In: *Parallel Problem Solving from Nature — PPSN IV*, herausgegeben von Hans-Michael Voigt, Werner Ebeling, Ingo Rechenberg und Hans-Paul Schwefel, Band 1141 der Reihe *Lecture Notes in Computer Science*, S. 22–31, Springer, Berlin, 1996

Herdy, M.: Evolutionsstrategie mit subjektiver Selektion zur Optimierung von Kaffeemischungen, http://www.vdivde-it.de/forschungsforum97/153/hy/Kaffee/kaffee.htm, 1997

Hodgson, G. M.: *Economics ans Evolution: Bringing Life Back into Economics*, University of Michigan Press, Ann Arbor, MI, 1993

Hoffmann, U. und Hofmann, H.: *Einführung in die Optimierung mit Anwendungsbeispielen aus dem Chemie–Ingenieur–Wesen*, Verlag Chemie, Weinheim / Bergstraße, 1971

Hoffmeister, F. und Bäck, T.: Genetic Algorithms and Evolution Strategies: Similarities and Differences, In: (Schwefel und Männer, 1991), S. 445–469

Hoffmeister, F., Rudolph, G. und Sprave, J.: DIOGENES: A fine-grained parallel algorithm with coarse-grain communication requirements, ESPRIT Parallel Computing Action, Bonn Workshop, 1991

Holland, J. H.: *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, MI, 1975

Holland, J. H. und Miller, J. H.: Artificial Adaptive Agents in Economic Theory, *American Economic Review Papers and Proceeding*, **81**, S. 365–370, 1991

Huberman, B. A. und Glance, N. S.: Diversity and Collective Action, In: *Interdisciplinary Approaches to Nonlinear Systems*, herausgegeben von M. Haken und M. Mikhailov, Springer, New York, 1993

Jacob, C.: *Principia Evolvica*, dpunkt–Verlag, Heidelberg, 1997

Kan, A. H. G. R. und Timmer, G. T.: *Global Optimization*, Kapitel IX, Band 1 der Reihe (Nemhauser, Kan und Todd, 1989), 1989

Kauffman, S. A.: *Origins of Order: Self–Organization and Selection in Evolution*, Oxford University Press, Oxford, 1990

Klein, S.: Lotterie im Garten Eden, DER SPIEGEL, (10), S. 184–189, 1998

Koch, K.: Sex gewinnt, *Bild der Wissenschaft*, (8), S. 105, 1994

Koza, J. R.: *Genetic Programming: On the Programming of Computers by Means of Natural Selection*, Complex Adaptive Systems, MIT Press, Cambridge, MA, 1992

Küppers, B.-O.: *Der Ursprung biologischer Information*, Piper, München, 1986

Kursawe, F.: *Evolutionsstrategien für die Vektoroptimierung*, Diplomarbeit, Fachbereich Informatik, Universität Dortmund, 1990

Kursawe, F.: A Variant of Evolution Strategies for Vector Optimization, In: *Parallel Problem Solving from Nature — Proc. 1st Workshop PPSN*, herausgegeben von H.-P. Schwefel und R. Männer, Band 496 der Reihe *Lecture Notes in Computer Science*, S. 193–197, Springer, Berlin, 1991

Kursawe, F.: Towards Self–Adapting Evolution Strategies, In: *Proc. of the 1995 IEEE Int'l Conference on Evolutionary Computation (ICEC '95)*, herausgegeben von IEEE Inc., S. 283–288, IEEE Press, Piscataway, NJ, 1995

Kursawe, F.: Following a Moving Optimum with an ES, http://ls11-www.cs.uni-dortmund.de/people/kursawe/Demos/MovingOpt/mov-optGIFE.html, 1996

Laumanns, M., Rudolph, G. und Schwefel, H.-P.: A spatial predator–prey approach to multi–objective optimization, In: *Parallel Problem Solving from Nature – PPSN V*, herausgegeben von Agoston E. Eiben, Thomas Bäck, Marc Schoenauer und Hans-Paul Schwefel, Band 1498 der Reihe *Lecture Notes in Computer Science*, S. 241–249, Springer, Berlin, 1998

Lohmann, R.: Application of Evolution Strategy in Parallel Populations, In: (Schwefel und Männer, 1991), S. 198–208

Lohmann, R.: Structure Evolution and Incomplete Induction, In: (Männer und Manderick, 1992), S. 175–185

Luhmann, N.: *Die Gesellschaft der Gesellschaft (2 Teilbände)*, Suhrkamp, Frankfurt/M., 1997

Malsch, T.: Sozionik, *Informatik Spektrum*, **20**, S. 229, 1997

Männer, R. und Manderick, B. (Hrsg.): *Parallel Problem Solving from Nature 2*, Elsevier, Amsterdam, 1992

Mattheck, C.: Bauteile wachsen wie Bäume, In: (Nachtigall, 1992b), S. 93–100

Menke, R.: A Revision of the Schema Theorem, Technischer Bericht, Universität Dortmund, 1997

Michalewicz, Z.: *Genetic Algorithms + Data Structures = Evolution Programs*, Springer, Berlin, 1992

Michalewicz, Z.: *Genetic Algorithms + Data Structures = Evolution Programs*, Springer, Berlin, zweite Auflage, 1994

Miersch, M.: Wozu taugt der Mann?, DIE ZEIT, (41), S. 35, 1996

Monod, J.: *Chance and Necessity*, Knopf, New York, 1971

Nachtigall, W.: Technische Biologie und Bionik – was ist das?, In: (Nachtigall, 1992b), S. 1–12

Nachtigall, W. (Hrsg.): *Technische Biologie und Bionik 1, 1. Bionik–Kongress, BIONA–report 8*, Gustav Fischer, Stuttgart, 1992b

Nelson, R. R. und Winter, S. G.: *An Evolutionary Theory of Economic Change*, Harvard University Press, Cambridge, MA, 1982

Nemhauser, G. L., Kan, A. H. G. R. und Todd, M. J. (Hrsg.): *Optimization*, Band 1 der Reihe *Handbooks in Operations Research and Management Science*, Elsevier, Amsterdam, 1989

Ostermeier, A.: An Evolution Strategy with Momentum Adaptation of the Random Number Distribution, In: (Männer und Manderick, 1992), S. 197–206

Ostermeier, A., Gawelczyk, A. und Hansen, N.: Step–size adaptation based on non–local use of selection information, In: (Davidor, Schwefel und Männer, 1994), S. 189–198

Ostermeier, A.: *Schrittweitenadaptation in der Evolutionsstrategie mit einem entstochastisierten Ansatz*, Dissertation, TU Berlin, Fachbereich Verfahrenstechnik, Umwelttechnik, Werkstoffwissenschaften, 1997

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

Rechenberg, I.: *Evolutionsstrategie '94*, Frommann–Holzboog, Stuttgart, 1994

Reichholf, J. H.: Reichtum durch Mangel, *Bild der Wissenschaft*, (3), S. 62–64, 1992

Rosen, R.: *Optimality Principles in Biology*, Butterworth, London, 1967

Rudolph, G.: *Globale Optimierung mit parallelen Evolutionsstrategien*, Diplomarbeit, Fachbereich Informatik, Universität Dortmund, 1990

Rudolph, G.: Global Optimization by means of Distributed Evolution Strategies, In: (Schwefel und Männer, 1991), S. 209–213

Rudolph, G.: On Correlated Mutations in Evolution Strategies, In: (Männer und Manderick, 1992), S. 105–114

Rudolph, G.: Convergence of Non–Elitist Strategies, In: *Proceedings of the First IEEE Conference on Evolutionary Computation, Orlando, FL*, S. 63–66, IEEE Press, Piscataway, NJ, 1994

Santibáñez-Koref, I.: Parallele Implementation von Multipopulationsstrategien, Interner Bericht TR–03–95, Technische Universität Berlin, 1995

Schaffer, J. D., Caruana, R. A., Eshelman, L. J. und Das, R.: A study of control parameters affecting online performance of genetic algorithms for function optimization, In: *Proceedings of the Third International Conference on Genetic Algorithms*, herausgegeben von J. D. Schaffer, S. 51–60, Morgan Kaufmann, San Mateo, CA, 1989

Schittkowski, K.: A Numerical Comparison of 13 Nonlinear Programming Codes with Randomly Generated Test Problems, In: *Numerical Optimisation of Dynamic Systems*, herausgegeben von L. C. W. Dixon und G. P. Szegö, S. 213–234, North–Holland, Amsterdam, 1980

Schlierkamp-Voosen, D.: *Populationsbasierte Wettbewerbsmodelle zur Strategieanpassung in Evolutionären Algorithmen*, Dissertation, Universität Dortmund, Fachbereich Informatik, 1998

Schütz, M.: *Eine Evolutionsstrategie für gemischt-ganzzahlige Optimierungsprobleme mit variabler Dimension*, Diplomarbeit, Fachbereich Informatik, Universität Dortmund, 1994

Schwefel, H.-P.: *Numerische Optimierung von Computer–Modellen mittels der Evolutionsstrategie*, Birkhäuser, Basel, 1977

Schwefel, H.-P.: *Numerical Optimization of Computer Models*, Wiley & Sons, Chichester, 1981

Schwefel, H.-P.: Collective Phenomena in Evolutionary Systems, In: *Problems of Constancy and Change — The Complementarity of Systems Approaches to Complexity*, herausgegeben von P. Checkland und I. Kiss, Band 2, S. 1025–1033, International Society for General System Research, Budapest, 1987

Schwefel, H.-P.: *Evolution and Optimum Seeking*, Wiley & Sons, Chichester, 1995

Schwefel, H.-P. und Kursawe, F.: On Natural Life's Tricks to Survive and Evolve, In: *Proc. Second IEEE World Congress on Computational Intelligence (WCCI'98) with Fifth IEEE Conf. Evolutionary Computation (IEEE/ICEC'98)*, herausgegeben von David B. Fogel, Hans-Paul Schwefel, Thomas Bäck und X. Yao, Band 1, S. 1–8, IEEE Press, Piscataway, NJ, 1998

Schwefel, H.-P. und Männer, R. (Hrsg.): *Parallel Problem Solving from Nature, Proceedings of the* 1^{st} *PPSN–Workshop, Dortmund, 1990*, Band 496 der Reihe *Lecture Notes in Computer Science*, Springer, Berlin, 1991

Schwefel, H.-P. und Rudolph, G.: Contemporary Evolution Strategies, In: *Advances in Artificial Life. Third International Conference on Artificial Life*, herausgegeben von F. Morán, A. Moreno, J. J. Merelo und P. Chacón, Band 929 der Reihe *Lecture Notes in Artificial Intelligence*, S. 893–907, Springer, Berlin, 1995

Stichweh, R., Reyer, H.-U. und Uszkoreit, H.: Memorandom zu einem Institut für Evolutionswissenschaft, Schriftenreihe *Suchprozesse für innovative Fragestellungen in der Wissenschaft* 1, Werner Reimers Stiftung, Bad Homburg, 1999

Stockinger, G.: Die Rückkehr der Killer, DER SPIEGEL, (16), S. 164–180, 1999

Törn, A. und Žilinskas, A.: *Global Optimization*, Band 350 der Reihe *Lecture Notes in Computer Science*, Springer, Berlin, 1989

Videler, J.: Comparing the cost of flight: Aircraft designers can still learn from nature, In: (Nachtigall, 1992b), S. 53–72

Weidlich, W. und Braun, M.: The master equation approach to nonlinear economics, *Evolutionary Economics*, **2**, S. 233–265, 1992

Wickler, W. und Seibt, U.: Die Evolution der Sexualität: Es mußte so kommen, *Bild der Wissenschaft*, (10), S. 142–152, 1983

Wickler, W. und Seibt, U.: *Männlich Weiblich. Ein Naturgesetz und seine Folgen*, Spektrum Akademischer Verlag, Heidelberg Berlin, 1998

Witt, U.: The Evolution of Economic Institutions as a Propagation Process, *Public Choice*, S. 155–172, 1989

Wolpert, D. H. und Macready, W. G.: No free lunch theorems for optimization, *IEEE Transactions on Evolutionary Computation*, **1** (1), S. 67–82, 1997

Zell, R. A.: Die Gen–Automaten, *Bild der Wissenschaft*, (4), S. 64–72, 1990

Zerbst, E. W.: *Bionik*, Teubner, Stuttgart, 1987

Anhang A

Untersuchung von τ_0 und τ_i

Jedes Fortschrittsgebirge wird aus zwei Perspektiven dargestellt, die 10 Parameterkombinationen mit den besten Konvergenzgeschwindigkeiten werden zusätzlich als Tabelle dargestellt. Sortierkriterium ist dabei die zuerst angegebene Konvergenzgeschwindigkeit, die jeweils andere wird aber zusätzlich angegeben. Verwendet wurde immer der beste Rekombinationstyp für das jeweilige n.

Die τ_0–τ_i–Untersuchungen ohne Rekombination unterstreichen die Wichtigkeit dieses Operators. Die mutative Schrittweitensteuerung kann die Anpassung der Schrittweiten nicht allein leisten, ein Informationsaustausch zwischen den Individuen ist zusätzlich nötig.

A.1 Kugelmodell

Beim Kugelmodell ist der Vergleich mit der „reinen“ τ_0–τ_i–Heuristik nicht fair, da man statt der Eins im Zähler φ^*_{max} für eine $(15\,,\,100)$ kennt und folglich benutzen kann (≈ 2). Auf diesen zusätzlichen Vergleich mit verdoppelten τ–Werten wird hier verzichtet, einerseits aus Konsistenzgründen, andererseits, weil auch diese Werte nicht an jene in der Tabelle heranreichen.

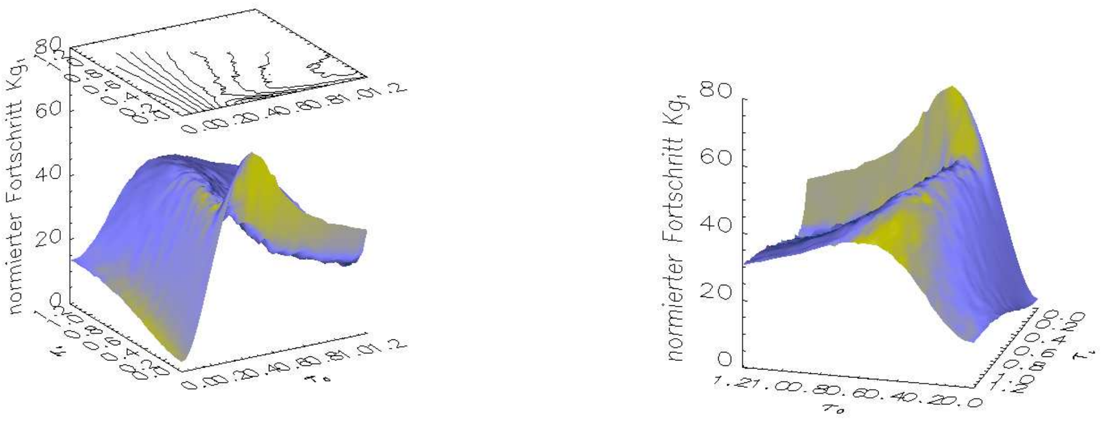


Abbildung A.1: $(15\,,\,100)$, Kugelmodell, $n = 10$, $n_\sigma = n$

τ_0–τ_i–Kombinationen sortiert nach der größten Konvergenzgeschwindigkeit $Kg_1 = \frac{n}{Gen.} \cdot \log\sqrt{\frac{f_{Startwert}}{f_{Bestwert}}}$						
n	Rekotyp	Gen.	τ_0	τ_i	Kg_1	Kg_2
10	541	200.00	0.4500	0.0000	3.6818	63.96
10	541	200.00	0.4250	0.0000	3.6431	63.29
10	541	200.00	0.4750	0.0000	3.6308	63.07
10	541	200.00	0.4250	0.0250	3.5847	62.27
10	541	200.00	0.5000	0.0000	3.5827	62.24
10	541	200.00	0.4000	0.0000	3.5560	61.77
10	541	200.00	0.4000	0.0250	3.5256	61.25
10	541	200.00	0.4500	0.0250	3.5200	61.15
10	541	200.00	0.5250	0.0000	3.5161	61.08
10	541	200.00	0.4750	0.0250	3.5114	61.00

Die der τ_0–τ_i–Heuristik nächsten untersuchten Werte $\tau_0 = 0.225$ und $\tau_i = 0.4$ liefern $Kg_1 = 1.8582$ bzw. $Kg_2 = 32.28$.

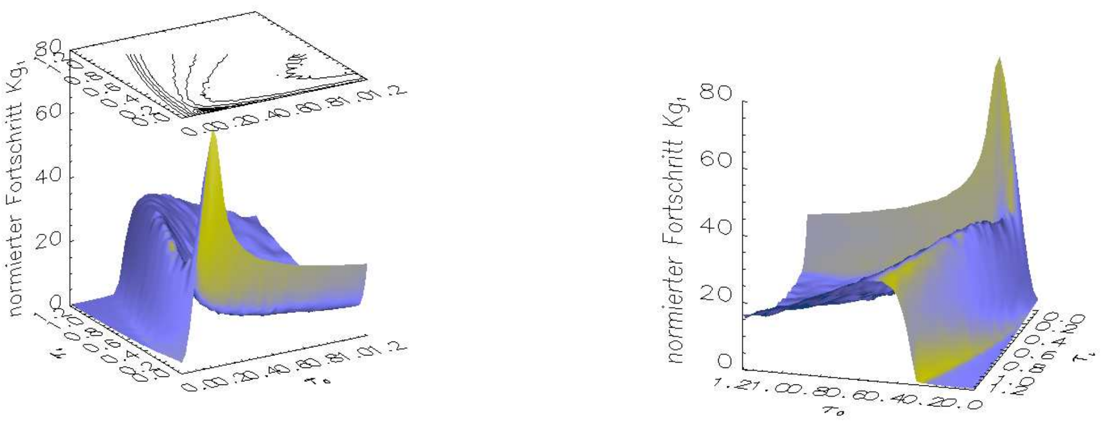


Abbildung A.2: (15 , 100), Kugelmodell, $n = 30$, $n_\sigma = n$

τ_0–τ_i–Kombinationen sortiert nach der größten Konvergenzgeschwindigkeit $Kg_1 = \frac{n}{Gen.} \cdot \log \sqrt{\frac{f_{Startwert}}{f_{Bestwert}}}$						
n	Rekotyp	Gen.	τ_0	τ_i	Kg_1	Kg_2
30	541	500.00	0.2000	0.0000	5.1573	74.66
30	541	500.00	0.2250	0.0000	4.9114	71.10
30	541	500.00	0.1750	0.0000	4.7228	68.37
30	541	500.00	0.2500	0.0000	4.2130	60.99
30	541	500.00	0.1750	0.0250	4.2078	60.91
30	541	500.00	0.1500	0.0000	4.0518	58.66
30	541	500.00	0.1500	0.0250	3.9627	57.37
30	541	500.00	0.2750	0.0000	3.5031	50.71
30	541	500.00	0.2000	0.0250	3.4103	49.37
30	541	500.00	0.1250	0.0000	3.2631	47.24

Die der τ_0–τ_i–Heuristik nächsten untersuchten Werte $\tau_0 = 0.125$ und $\tau_i = 0.3$ liefern $Kg_1 = 2.0901$ bzw. $Kg_2 = 30.26$.

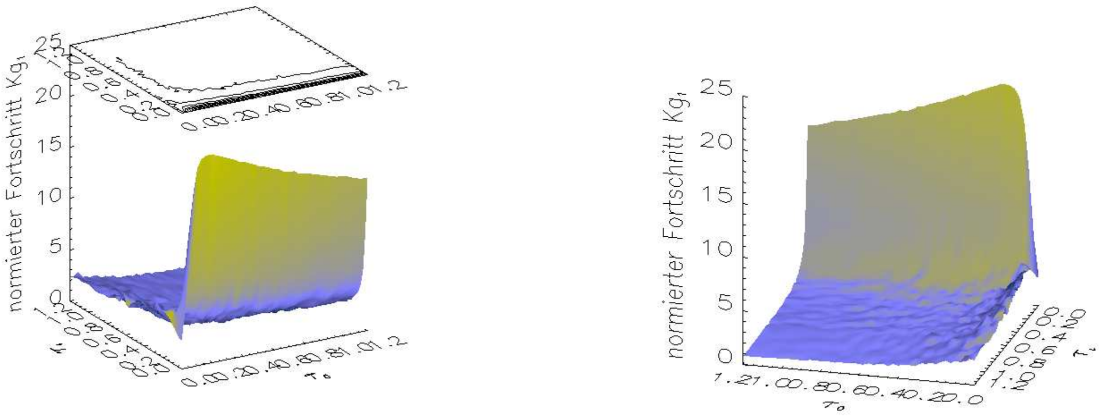


Abbildung A.3: (15 , 100), Kugelmodell, $n = 30$, $n_\sigma = n$, ohne Rekombination

τ_0–τ_i–Kombinationen sortiert nach der größten Konvergenzgeschwindigkeit $Kg_1 = \frac{n}{Gen.} \cdot \log \sqrt{\frac{f_{Startwert}}{f_{Bestwert}}}$						
n	Rekotyp	Gen.	τ_0	τ_i	Kg_1	Kg_2
30	111	500.00	0.1750	0.0000	1.4037	20.32
30	111	500.00	0.2000	0.0000	1.4017	20.29
30	111	500.00	0.1500	0.0000	1.3980	20.24
30	111	500.00	0.2250	0.0000	1.3869	20.08
30	111	500.00	0.1250	0.0000	1.3823	20.01
30	111	500.00	0.2500	0.0000	1.3771	19.94
30	111	500.00	0.2750	0.0000	1.3639	19.74
30	111	500.00	0.3000	0.0000	1.3529	19.59
30	111	500.00	0.3250	0.0000	1.3412	19.42
30	111	500.00	0.3500	0.0000	1.3287	19.23

Die der τ_0–τ_i–Heuristik nächsten untersuchten Werte $\tau_0 = 0.125$ und $\tau_i = 0.3$ liefern $Kg_1 = 0.1615$ bzw. $Kg_2 = 2.34$.

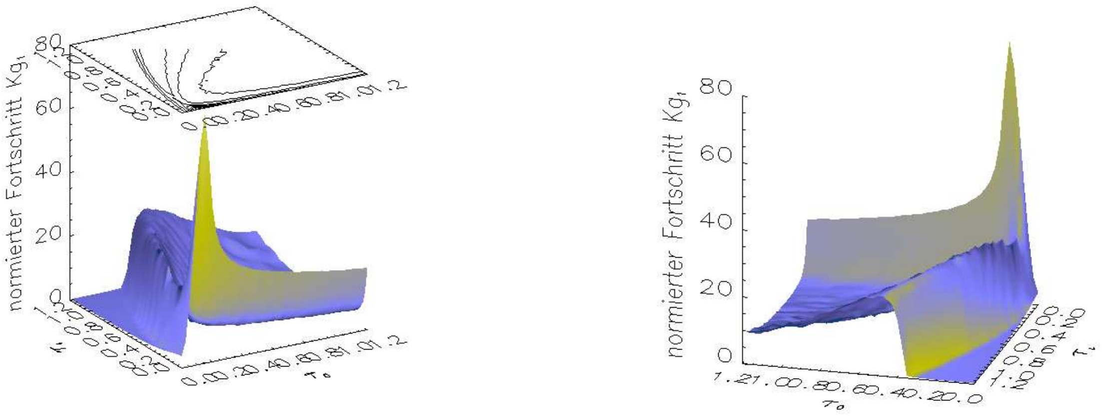


Abbildung A.4: (15 , 100), Kugelmodell, $n = 50$, $n_\sigma = n$

τ_0–τ_i–Kombinationen sortiert nach der größten Konvergenzgeschwindigkeit $Kg_1 = \frac{n}{Gen.} \cdot \log \sqrt{\frac{f_{Startwert}}{f_{Bestwert}}}$						
n	Rekotyp	Gen.	τ_0	τ_i	Kg_1	Kg_2
50	541	749.24	0.1500	0.0000	5.9946	78.02
50	541	750.00	0.1250	0.0000	5.3650	69.90
50	541	750.00	0.1750	0.0000	4.9735	64.80
50	541	750.00	0.1000	0.0000	4.1532	54.11
50	541	750.00	0.1000	0.0250	3.9615	51.61
50	541	750.00	0.1250	0.0250	3.7196	48.46
50	541	750.00	0.2000	0.0000	3.6664	47.77
50	541	750.00	0.2250	0.0000	3.0480	39.71
50	541	750.00	0.1500	0.0250	2.8472	37.10
50	541	750.00	0.0750	0.0000	2.7886	36.33

Die der τ_0–τ_i–Heuristik nächsten untersuchten Werte $\tau_0 = 0.1$ und $\tau_i = 0.275$ liefern $Kg_1 = 1.7506$ bzw. $Kg_2 = 22.81$.

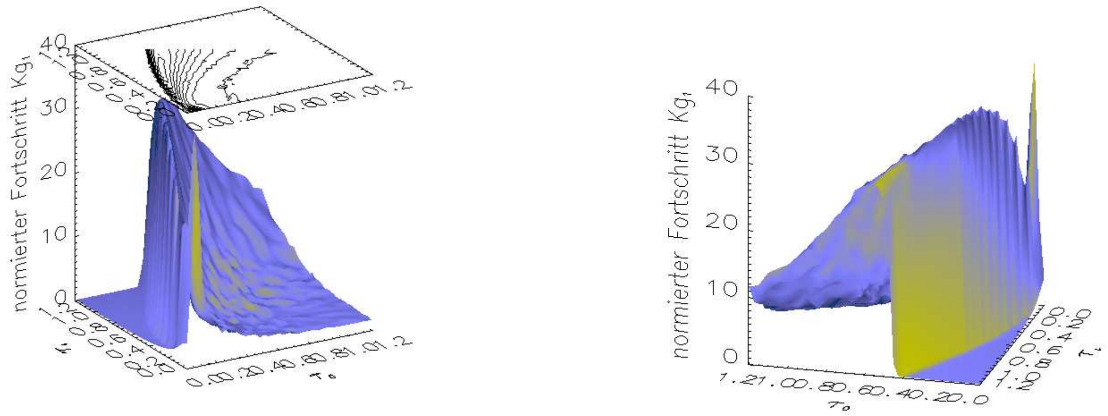


Abbildung A.5: (15 , 100), Kugelmodell, $n = 100$, $n_\sigma = n$

τ_0–τ_i–Kombinationen sortiert nach der größten Konvergenzgeschwindigkeit $Kg_1 = \frac{n}{Gen.} \cdot \log \sqrt{\frac{f_{Startwert}}{f_{Bestwert}}}$						
n	Rekotyp	Gen.	τ_0	τ_i	Kg_1	Kg_2
100	551	2000.00	0.0500	0.0000	2.0935	36.37
100	551	2000.00	0.1750	0.5500	1.9989	34.72
100	551	2000.00	0.2000	0.6000	1.9954	34.66
100	551	2000.00	0.2250	0.6500	1.9803	34.40
100	551	2000.00	0.2500	0.7000	1.9767	34.34
100	551	2000.00	0.1750	0.5250	1.9652	34.14
100	551	2000.00	0.2250	0.6250	1.9519	33.91
100	551	2000.00	0.2500	0.6750	1.9476	33.83
100	551	2000.00	0.2000	0.6250	1.9422	33.74
100	551	2000.00	0.2000	0.5750	1.9418	33.73

Die der τ_0–τ_i–Heuristik nächsten untersuchten Werte $\tau_0 = 0.075$ und $\tau_i = 0.225$ liefern $Kg_1 = 1.4503$ bzw. $Kg_2 = 25.19$.

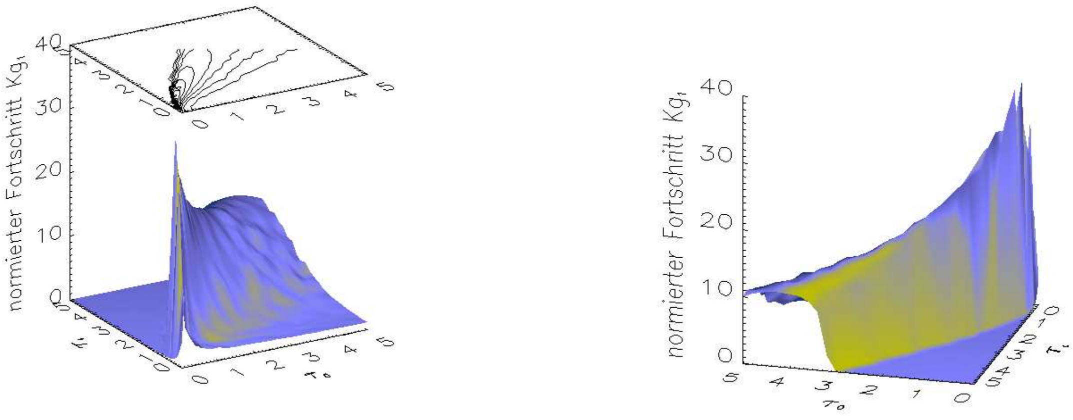


Abbildung A.6: (15 , 100), Kugelmodell, $n = 100$, $n_\sigma = n$, τ_0, $\tau_i \in [0, 5]$

τ_0–τ_i–Kombinationen sortiert nach der größten Konvergenzgeschwindigkeit $Kg_1 = \frac{n}{Gen.} \cdot \log \sqrt{\frac{f_{Startwert}}{f_{Bestwert}}}$						
n	Rekotyp	Gen.	$\tau_\mathbf{0}$	$\tau_\mathbf{i}$	Kg_1	Kg_2
100	551	2000.00	0.2000	0.6000	1.9875	34.53
100	551	2000.00	0.3000	0.8000	1.9435	33.76
100	551	2000.00	0.4000	1.0000	1.7673	30.70
100	551	2000.00	0.3000	0.7000	1.7363	30.16
100	551	2000.00	0.1000	0.3000	1.6471	28.61
100	551	2000.00	0.2000	0.5000	1.6321	28.35
100	551	2000.00	0.6000	1.2000	1.6260	28.25
100	551	2000.00	0.5000	1.0000	1.6155	28.06
100	551	2000.00	0.4000	0.8000	1.5235	26.47
100	551	2000.00	0.7000	1.4000	1.5205	26.41

Abweichungen zu Abbildung A.5 erklären sich aus einem vierfach gröberen Raster.

A.2 Skalierungstest

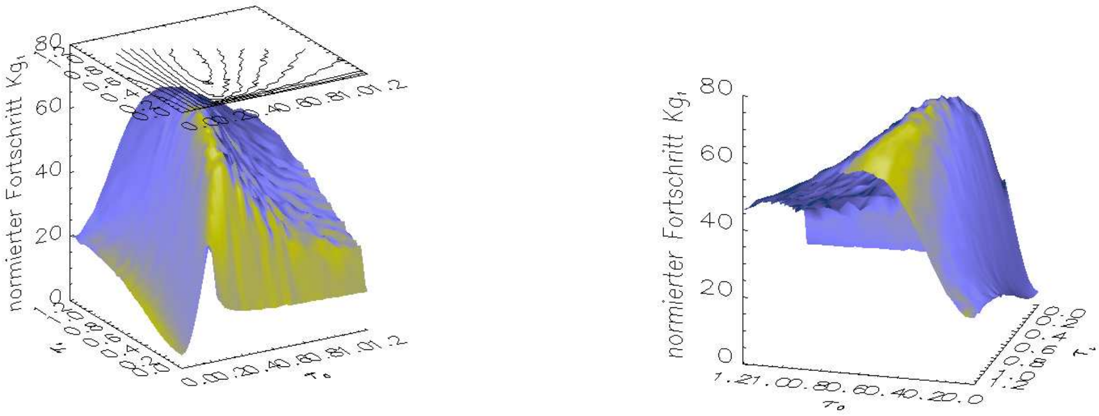


Abbildung A.7: (15 , 100), Skalierungstest, $n = 10$, $n_\sigma = n$

τ_0–τ_i–Kombinationen sortiert nach der größten Konvergenzgeschwindigkeit $Kg_1 = \frac{n}{Gen.} \cdot \log \sqrt{\frac{f_{Startwert}}{f_{Bestwert}}}$						
n	Rekotyp	Gen.	τ_0	τ_i	Kg_1	Kg_2
10	541	300.00	0.4000	0.6000	2.6655	69.46
10	541	300.00	0.4000	0.5000	2.6511	69.08
10	541	300.00	0.4250	0.5000	2.6500	69.05
10	541	300.00	0.4500	0.6250	2.6478	69.00
10	541	300.00	0.3750	0.4500	2.6472	68.98
10	541	300.00	0.4250	0.4750	2.6365	68.70
10	541	300.00	0.4250	0.5750	2.6336	68.62
10	541	300.00	0.4500	0.5000	2.6334	68.62
10	541	300.00	0.4000	0.5250	2.6323	68.59
10	541	300.00	0.4250	0.5500	2.6323	68.59

Die der τ_0–τ_i–Heuristik nächsten untersuchten Werte $\tau_0 = 0.225$ und $\tau_i = 0.4$ liefern $Kg_1 = 1.7967$ bzw. $Kg_2 = 46.82$.

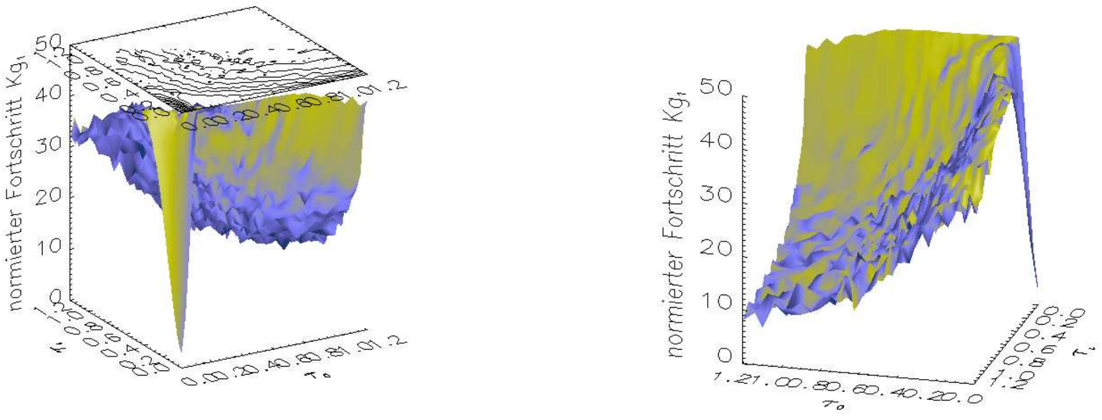


Abbildung A.8: (15 , 100), Skalierungstest, $n = 10$, $n_\sigma = n$, ohne Rekombination

τ_0–τ_i–Kombinationen sortiert nach der größten Konvergenzgeschwindigkeit $Kg_1 = \frac{n}{Gen.} \cdot \log \sqrt{\frac{f_{Startwert}}{f_{Bestwert}}}$						
n	Rekotyp	Gen.	τ_0	τ_i	Kg_1	Kg_2
10	111	513.20	0.2500	0.0000	1.1142	49.56
10	111	524.20	0.2000	0.0000	1.0922	49.64
10	111	526.12	0.2250	0.0000	1.0881	49.58
10	111	529.92	0.2750	0.0000	1.0797	49.49
10	111	532.44	0.3000	0.0000	1.0749	49.52
10	111	537.28	0.3250	0.0000	1.0650	49.51
10	111	543.12	0.1750	0.0000	1.0556	49.74
10	111	559.48	0.1500	0.0000	1.0255	49.79
10	111	558.48	0.3500	0.0000	1.0231	49.44
10	111	564.20	0.2000	0.0250	1.0229	49.51

Die der τ_0–τ_i–Heuristik nächsten untersuchten Werte $\tau_0 = 0.225$ und $\tau_i = 0.4$ liefern $Kg_1 = 0.5776$ bzw. $Kg_2 = 37.63$.

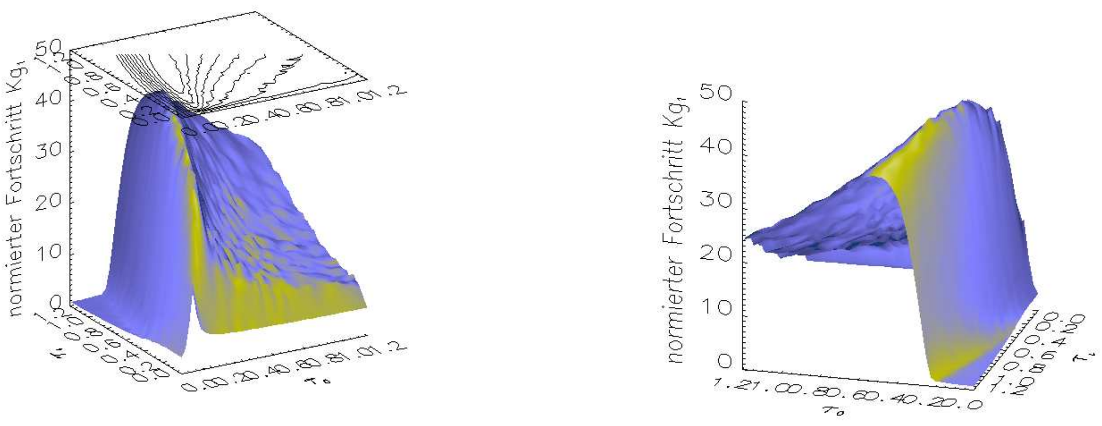


Abbildung A.9: (15 , 100), Skalierungstest, $n = 20$, $n_\sigma = n$

τ_0–τ_i–Kombinationen sortiert nach der größten Konvergenzgeschwindigkeit $Kg_1 = \frac{n}{Gen.} \cdot \log \sqrt{\frac{f_{Startwert}}{f_{Bestwert}}}$						
n	Rekotyp	Gen.	τ_0	τ_i	Kg_1	Kg_2
20	541	400.00	0.2500	0.5250	2.5867	44.94
20	541	400.00	0.2500	0.5500	2.5748	44.73
20	541	400.00	0.2750	0.6250	2.5726	44.69
20	541	400.00	0.3500	0.8000	2.5616	44.50
20	541	400.00	0.3000	0.6500	2.5602	44.48
20	541	400.00	0.2500	0.5000	2.5593	44.46
20	541	400.00	0.3000	0.7000	2.5555	44.39
20	541	400.00	0.3250	0.6750	2.5515	44.32
20	541	400.00	0.2500	0.5750	2.5509	44.31
20	541	400.00	0.2750	0.5750	2.5453	44.22

Die der τ_0–τ_i–Heuristik nächsten untersuchten Werte $\tau_0 = 0.15$ und $\tau_i = 0.325$ liefern $Kg_1 = 2.0282$ bzw. $Kg_2 = 35.23$.

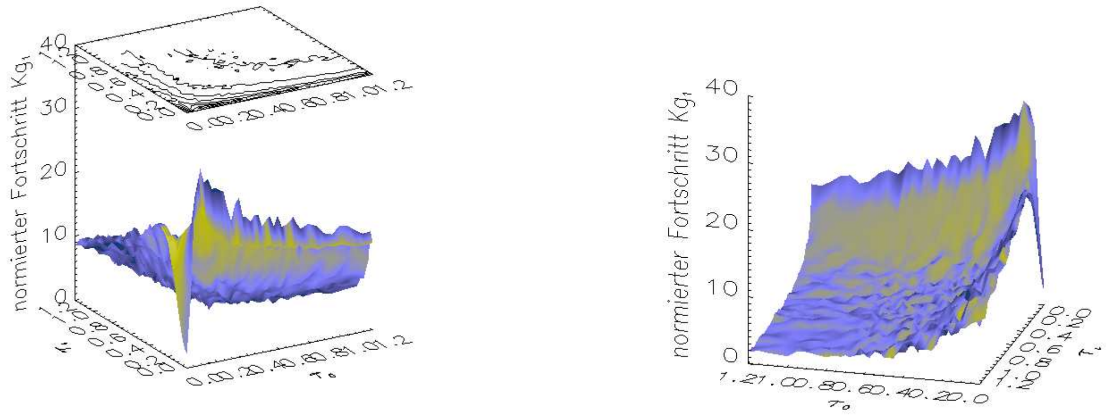


Abbildung A.10: (15 , 100), Skalierungstest, $n = 20$, $n_\sigma = n$, ohne Rekombination

τ_0–τ_i–Kombinationen sortiert nach der größten Konvergenzgeschwindigkeit $Kg_1 = \frac{n}{Gen.} \cdot \log \sqrt{\frac{f_{Startwert}}{f_{Bestwert}}}$						
n	Rekotyp	Gen.	τ_0	τ_i	Kg_1	Kg_2
20	111	1000.00	0.1000	0.0250	0.7000	30.40
20	111	1000.00	0.0750	0.0000	0.6742	29.28
20	111	1000.00	0.1500	0.0250	0.6577	28.57
20	111	1000.00	0.1250	0.0250	0.6466	28.08
20	111	1000.00	0.0750	0.0250	0.6215	26.99
20	111	1000.00	0.2000	0.0250	0.6184	26.86
20	111	1000.00	0.1750	0.0250	0.6179	26.84
20	111	1000.00	0.0500	0.0000	0.6090	26.45
20	111	1000.00	0.2250	0.0250	0.6082	26.41
20	111	1000.00	0.2500	0.0250	0.6004	26.07

Die der τ_0–τ_i–Heuristik nächsten untersuchten Werte $\tau_0 = 0.15$ und $\tau_i = 0.325$ liefern $Kg_1 = 0.2320$ bzw. $Kg_2 = 10.08$.

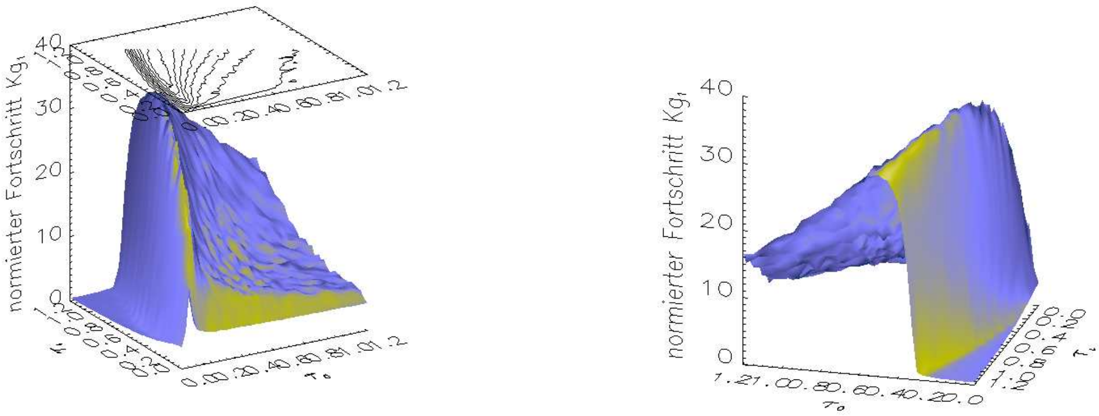


Abbildung A.11: (15 , 100), Skalierungstest, $n = 30$, $n_\sigma = n$

τ_0–τ_i–Kombinationen sortiert nach der größten Konvergenzgeschwindigkeit $Kg_1 = \frac{n}{Gen.} \cdot \log\sqrt{\frac{f_{Startwert}}{f_{Bestwert}}}$						
n	Rekotyp	Gen.	τ_0	τ_i	Kg_1	Kg_2
30	541	500.00	0.2500	0.6500	2.4336	35.23
30	541	500.00	0.2000	0.5500	2.4332	35.22
30	541	500.00	0.2750	0.6750	2.4211	35.05
30	541	500.00	0.2250	0.5750	2.4197	35.03
30	541	500.00	0.2750	0.7250	2.4049	34.81
30	541	500.00	0.2750	0.7000	2.3989	34.73
30	541	500.00	0.2500	0.7000	2.3938	34.65
30	541	500.00	0.2250	0.6000	2.3935	34.65
30	541	500.00	0.2250	0.6250	2.3930	34.64
30	541	500.00	0.2500	0.6750	2.3930	34.64

Die der τ_0–τ_i–Heuristik nächsten untersuchten Werte $\tau_0 = 0.125$ und $\tau_i = 0.3$ liefern $Kg_1 = 2.1132$ bzw. $Kg_2 = 30.59$.

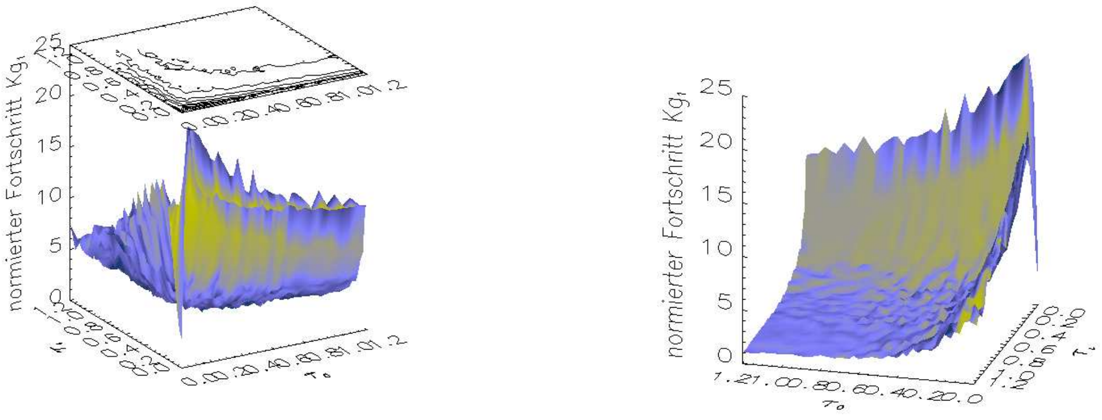


Abbildung A.12: (15 , 100), Skalierungstest, $n = 30$, $n_\sigma = n$, ohne Rekombination

τ_0–τ_i–Kombinationen sortiert nach der größten Konvergenzgeschwindigkeit $Kg_1 = \frac{n}{Gen.} \cdot \log \sqrt{\frac{f_{Startwert}}{f_{Bestwert}}}$						
n	Rekotyp	Gen.	$\tau_\mathbf{0}$	$\tau_\mathbf{i}$	Kg_1	Kg_2
30	111	1962.08	0.0500	0.0250	0.4143	23.29
30	111	1972.48	0.0750	0.0250	0.4065	23.05
30	111	2000.00	0.0500	0.0000	0.4032	23.35
30	111	1947.44	0.1250	0.0250	0.4007	22.21
30	111	1995.88	0.1000	0.0250	0.3906	22.56
30	111	1993.00	0.1500	0.0250	0.3696	21.30
30	111	1993.84	0.1750	0.0250	0.3587	20.67
30	111	1993.96	0.2500	0.0250	0.3506	20.21
30	111	2000.00	0.0250	0.0000	0.3436	19.90
30	111	2000.00	0.2000	0.0250	0.3407	19.73

Die der τ_0–τ_i–Heuristik nächsten untersuchten Werte $\tau_0 = 0.125$ und $\tau_i = 0.3$ liefern $Kg_1 = 0.0922$ bzw. $Kg_2 = 5.34$.

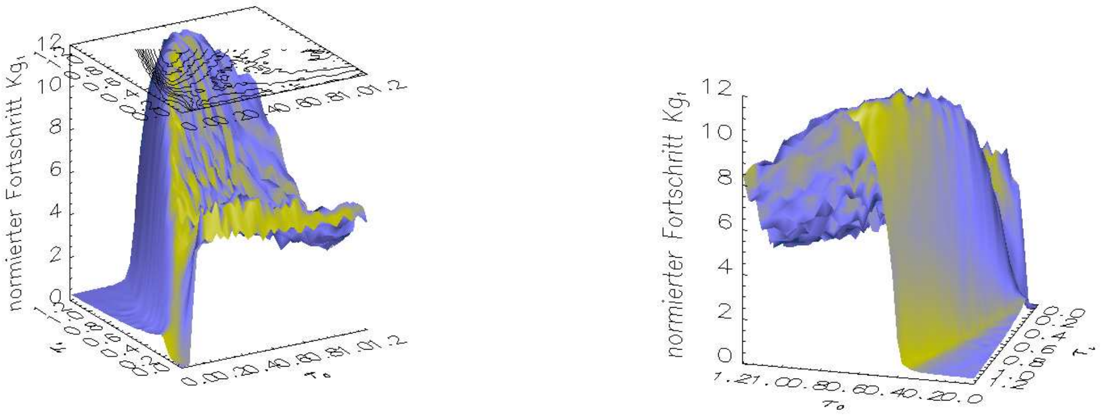


Abbildung A.13: (15 , 100), Skalierungstest, $n = 30$, $n_\sigma = n$, Rekombinationstyp 141

τ_0–τ_i–Kombinationen sortiert nach der größten Konvergenzgeschwindigkeit $Kg_1 = \frac{n}{Gen.} \cdot \log \sqrt{\frac{f_{Startwert}}{f_{Bestwert}}}$						
n	Rekotyp	Gen.	τ_0	τ_i	Kg_1	Kg_2
30	141	500.00	0.4750	0.8500	0.8208	11.88
30	141	500.00	0.6000	1.0500	0.8170	11.83
30	141	500.00	0.6000	1.1000	0.8163	11.82
30	141	500.00	0.5500	1.0000	0.8162	11.82
30	141	500.00	0.4750	0.9000	0.8154	11.80
30	141	500.00	0.5000	0.9000	0.8131	11.77
30	141	500.00	0.6500	1.1500	0.8129	11.77
30	141	500.00	0.5250	1.0000	0.8126	11.76
30	141	500.00	0.5750	1.0000	0.8116	11.75
30	141	500.00	0.4500	0.9000	0.8115	11.75

Die der τ_0–τ_i–Heuristik nächsten untersuchten Werte $\tau_0 = 0.125$ und $\tau_i = 0.3$ liefern $Kg_1 = 0.3669$ bzw. $Kg_2 = 5.31$.

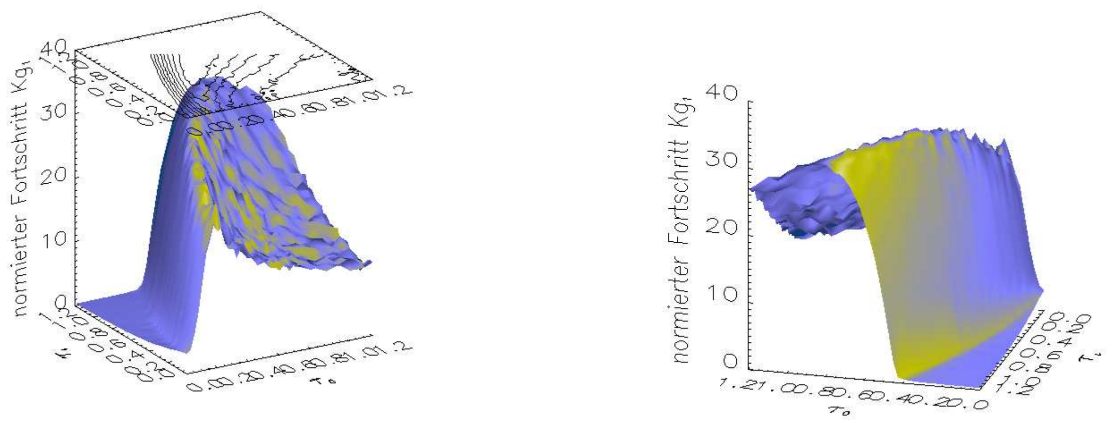


Abbildung A.14: (15 , 100), Skalierungstest, $n = 30$, $n_\sigma = n$, Rekombinationstyp 351

τ_0–τ_i–Kombinationen sortiert nach der größten Konvergenzgeschwindigkeit $Kg_1 = \frac{n}{Gen.} \cdot \log \sqrt{\frac{f_{Startwert}}{f_{Bestwert}}}$						
n	Rekotyp	Gen.	τ_0	τ_i	Kg_1	Kg_2
30	351	500.00	0.7250	1.0500	2.2632	32.76
30	351	500.00	0.6000	0.8500	2.2578	32.68
30	351	500.00	0.4750	0.6750	2.2569	32.67
30	351	500.00	0.6750	0.9500	2.2564	32.66
30	351	500.00	0.5000	0.7000	2.2502	32.57
30	351	500.00	0.6000	0.9000	2.2447	32.49
30	351	500.00	0.5500	0.8000	2.2443	32.49
30	351	500.00	0.6750	1.0000	2.2428	32.47
30	351	500.00	0.5250	0.7250	2.2413	32.45
30	351	500.00	0.5000	0.7250	2.2411	32.44

Die der τ_0–τ_i–Heuristik nächsten untersuchten Werte $\tau_0 = 0.125$ und $\tau_i = 0.3$ liefern $Kg_1 = 0.6870$ bzw. $Kg_2 = 9.95$.

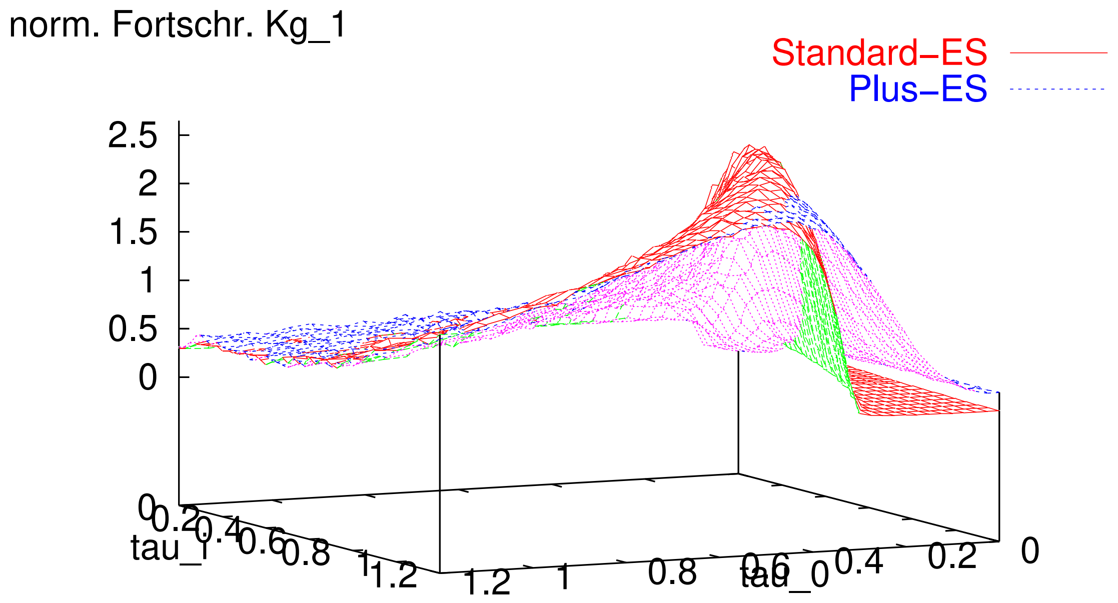


Abbildung A.15: Vergleich einer Standard–ES mit einer (15 + 100)–ES, Skalierungstest, $n = 30$, $n_\sigma = n$

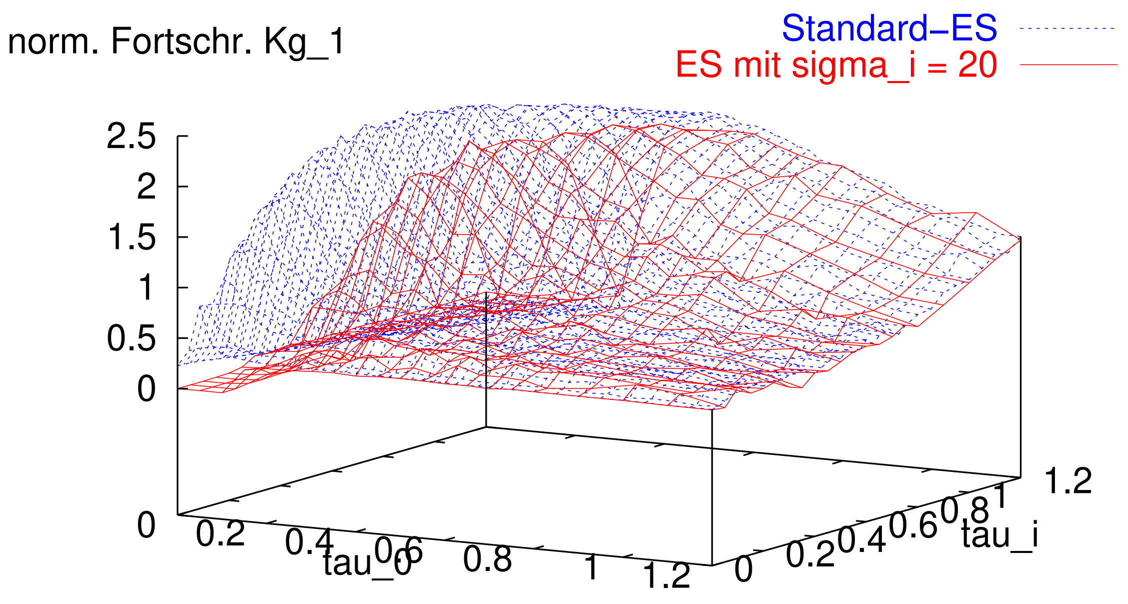


Abbildung A.16: Vergleich einer Standard–ES mit einer (15 , 100)–ES mit $\sigma_i^{Start} = 20$, Skalierungstest, $n = 30$, $n_\sigma = n$

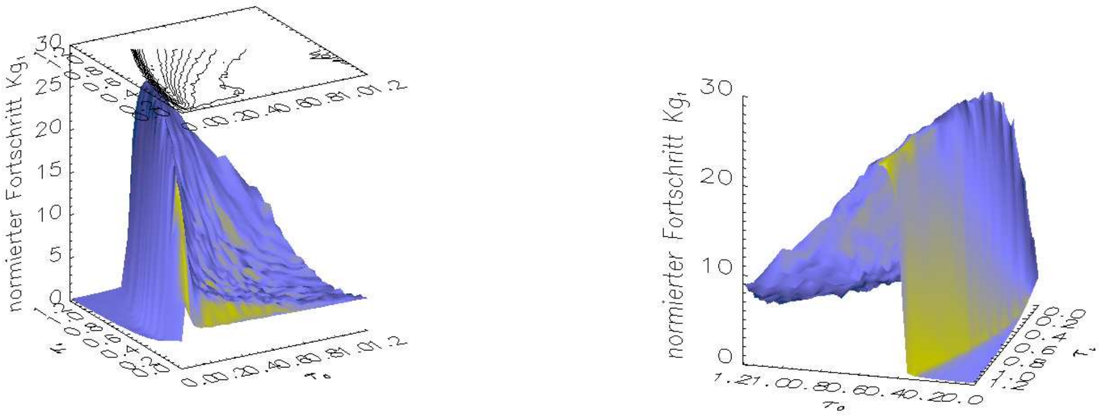


Abbildung A.17: (15 , 100), Skalierungstest, $n = 50$, $n_\sigma = n$

τ_0–τ_i–Kombinationen sortiert nach der größten Konvergenzgeschwindigkeit $Kg_1 = \frac{n}{Gen.} \cdot \log \sqrt{\frac{f_{Startwert}}{f_{Bestwert}}}$						
n	Rekotyp	Gen.	τ_0	τ_i	Kg_1	Kg_2
50	541	750.00	0.2000	0.6500	2.1075	27.46
50	541	750.00	0.2250	0.7000	2.1020	27.39
50	541	750.00	0.1500	0.5000	2.1014	27.38
50	541	750.00	0.2000	0.6250	2.0871	27.19
50	541	750.00	0.2250	0.6750	2.0835	27.15
50	541	750.00	0.2500	0.7250	2.0787	27.08
50	541	750.00	0.1750	0.5500	2.0754	27.04
50	541	750.00	0.1750	0.5750	2.0746	27.03
50	541	750.00	0.2500	0.7500	2.0667	26.93
50	541	750.00	0.1500	0.4500	2.0595	26.83

Die der τ_0–τ_i–Heuristik nächsten untersuchten Werte $\tau_0 = 0.1$ und $\tau_i = 0.275$ liefern $Kg_1 = 1.8090$ bzw. $Kg_2 = 23.57$.

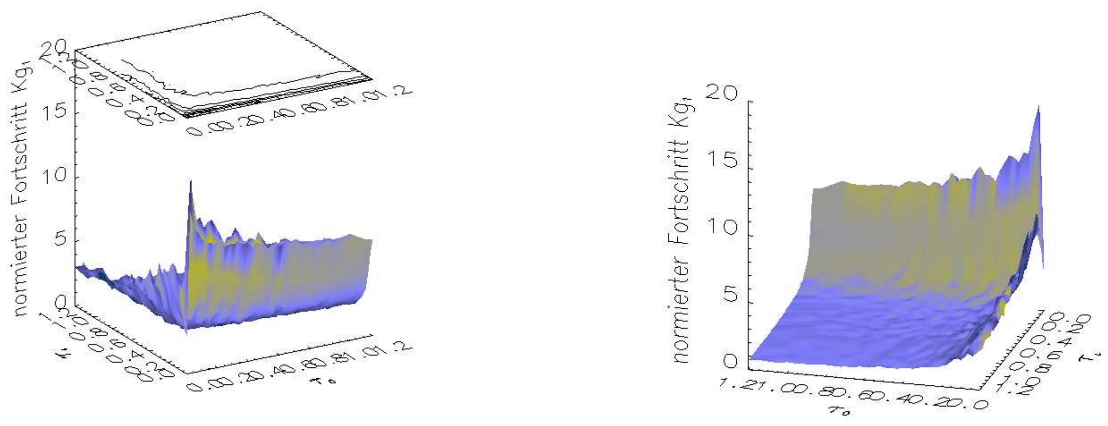


Abbildung A.18: (15 , 100), Skalierungstest, $n = 50$, $n_\sigma = n$, ohne Rekombination

τ_0–τ_i–Kombinationen sortiert nach der größten Konvergenzgeschwindigkeit $Kg_1 = \frac{n}{Gen.} \cdot \log\sqrt{\frac{f_{Startwert}}{f_{Bestwert}}}$						
n	Rekotyp	Gen.	τ_0	τ_i	Kg_1	Kg_2
50	111	3000.00	0.0250	0.0000	0.2965	15.45
50	111	3000.00	0.0500	0.0250	0.2639	13.75
50	111	3000.00	0.0250	0.0250	0.2481	12.93
50	111	3000.00	0.1000	0.0250	0.2262	11.79
50	111	3000.00	0.0750	0.0250	0.2244	11.69
50	111	3000.00	0.1750	0.0250	0.2197	11.45
50	111	3000.00	0.1500	0.0250	0.2175	11.34
50	111	3000.00	0.0500	0.0000	0.2129	11.10
50	111	3000.00	0.1250	0.0250	0.2105	10.97
50	111	3000.00	0.2000	0.0250	0.2083	10.86

Die der τ_0–τ_i–Heuristik nächsten untersuchten Werte $\tau_0 = 0.1$ und $\tau_i = 0.275$ liefern $Kg_1 = 0.0509$ bzw. $Kg_2 = 2.65$.

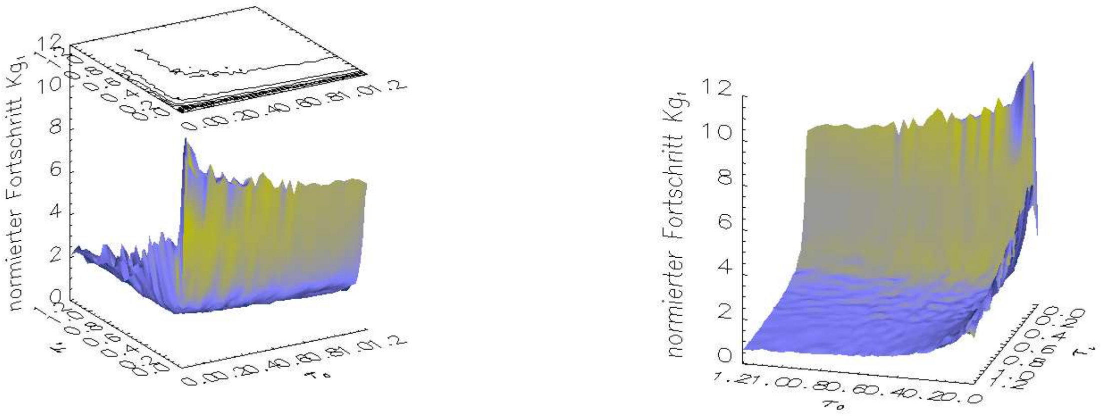


Abbildung A.19: (15 , 100), Skalierungstest, $n = 60$, $n_\sigma = n$, ohne Rekombination

τ_0–τ_i–Kombinationen sortiert nach der größten Konvergenzgeschwindigkeit $Kg_1 = \frac{n}{Gen.} \cdot \log\sqrt{\frac{f_{Startwert}}{f_{Bestwert}}}$						
n	Rekotyp	Gen.	τ_0	τ_i	Kg_1	Kg_2
60	111	3500.00	0.0250	0.0000	0.2157	10.93
60	111	3500.00	0.0500	0.0250	0.2071	10.50
60	111	3500.00	0.0750	0.0250	0.2012	10.19
60	111	3500.00	0.0250	0.0250	0.1972	9.99
60	111	3500.00	0.1000	0.0250	0.1941	9.83
60	111	3500.00	0.0500	0.0000	0.1857	9.41
60	111	3500.00	0.2000	0.0250	0.1822	9.23
60	111	3500.00	0.1250	0.0250	0.1794	9.09
60	111	3500.00	0.2000	0.0000	0.1783	9.03
60	111	3500.00	0.0750	0.0000	0.1778	9.01

Die der τ_0–τ_i–Heuristik nächsten untersuchten Werte $\tau_0 = 0.1$ und $\tau_i = 0.25$ liefern $Kg_1 = 0.0355$ bzw. $Kg_2 = 1.80$.

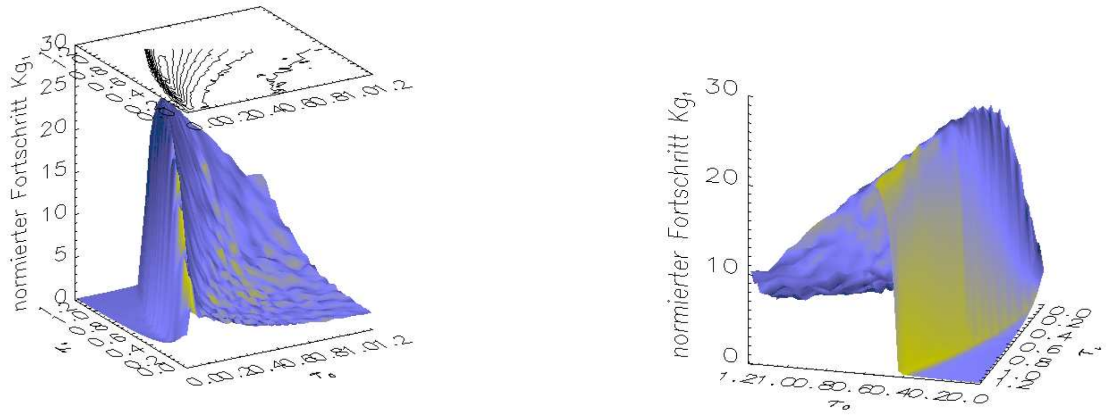


Abbildung A.20: (15 , 100), Skalierungstest, $n = 75$, $n_\sigma = n$

τ_0–τ_i–Kombinationen sortiert nach der größten Konvergenzgeschwindigkeit $Kg_1 = \frac{n}{Gen.} \cdot \log \sqrt{\frac{f_{Startwert}}{f_{Bestwert}}}$						
n	Rekotyp	Gen.	τ_0	τ_i	Kg_1	Kg_2
75	551	1000.00	0.2000	0.5750	2.2560	26.13
75	551	1000.00	0.2250	0.6000	2.2257	25.78
75	551	1000.00	0.2250	0.6250	2.2231	25.75
75	551	1000.00	0.1750	0.5250	2.2222	25.74
75	551	1000.00	0.2750	0.7250	2.2159	25.66
75	551	1000.00	0.1750	0.5000	2.2126	25.62
75	551	1000.00	0.1500	0.4500	2.2009	25.49
75	551	1000.00	0.2000	0.5500	2.1988	25.46
75	551	1000.00	0.2000	0.6000	2.1976	25.45
75	551	1000.00	0.2250	0.6500	2.1973	25.45

Die der τ_0–τ_i–Heuristik nächsten untersuchten Werte $\tau_0 = 0.075$ und $\tau_i = 0.25$ liefern $Kg_1 = 1.4482$ bzw. $Kg_2 = 16.77$.

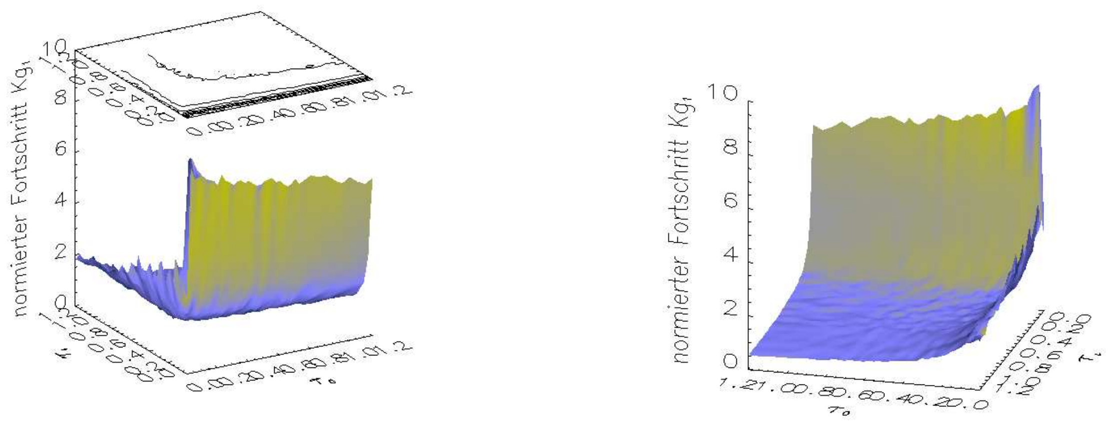


Abbildung A.21: (15 , 100), Skalierungstest, $n = 75$, $n_\sigma = n$, ohne Rekombination

τ_0–τ_i–Kombinationen sortiert nach der größten Konvergenzgeschwindigkeit $Kg_1 = \frac{n}{Gen.} \cdot \log \sqrt{\frac{f_{Startwert}}{f_{Bestwert}}}$						
n	Rekotyp	Gen.	τ_0	τ_i	Kg_1	Kg_2
75	111	4000.00	0.0250	0.0000	0.1819	8.43
75	111	4000.00	0.0500	0.0250	0.1785	8.27
75	111	4000.00	0.0250	0.0250	0.1773	8.22
75	111	4000.00	0.0750	0.0250	0.1695	7.85
75	111	4000.00	0.1000	0.0000	0.1636	7.58
75	111	4000.00	0.0750	0.0000	0.1632	7.56
75	111	4000.00	0.0500	0.0000	0.1622	7.51
75	111	4000.00	0.2250	0.0000	0.1615	7.48
75	111	4000.00	0.1250	0.0000	0.1604	7.43
75	111	4000.00	0.1500	0.0000	0.1589	7.36

Die der τ_0–τ_i–Heuristik nächsten untersuchten Werte $\tau_0 = 0.075$ und $\tau_i = 0.25$ liefern $Kg_1 = 0.0328$ bzw. $Kg_2 = 1.52$.

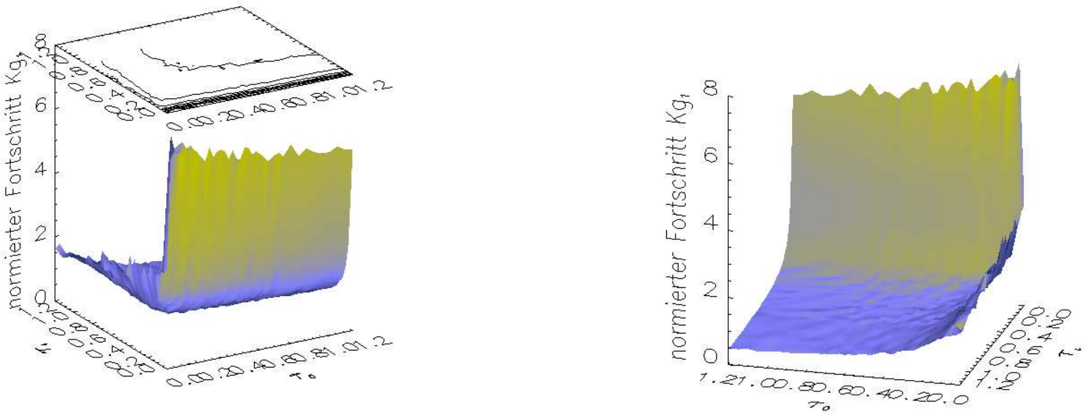


Abbildung A.22: (15 , 100), Skalierungstest, $n = 90$, $n_\sigma = n$, ohne Rekombination

τ_0–τ_i–Kombinationen sortiert nach der größten Konvergenzgeschwindigkeit $Kg_1 = \frac{n}{Gen.} \cdot \log\sqrt{\frac{f_{Startwert}}{f_{Bestwert}}}$						
n	Rekotyp	Gen.	τ_0	τ_i	Kg_1	Kg_2
90	111	4500.00	0.0250	0.0000	0.1672	7.26
90	111	4500.00	0.1000	0.0000	0.1604	6.97
90	111	4500.00	0.1500	0.0000	0.1571	6.82
90	111	4500.00	0.0750	0.0250	0.1568	6.81
90	111	4500.00	0.0500	0.0000	0.1551	6.74
90	111	4500.00	0.1250	0.0000	0.1550	6.73
90	111	4500.00	0.0500	0.0250	0.1533	6.66
90	111	4500.00	0.1750	0.0000	0.1533	6.66
90	111	4500.00	0.2750	0.0000	0.1527	6.63
90	111	4500.00	0.0750	0.0000	0.1527	6.63

Die der τ_0–τ_i–Heuristik nächsten untersuchten Werte $\tau_0 = 0.075$ und $\tau_i = 0.225$ liefern $Kg_1 = 0.0326$ bzw. $Kg_2 = 1.42$.

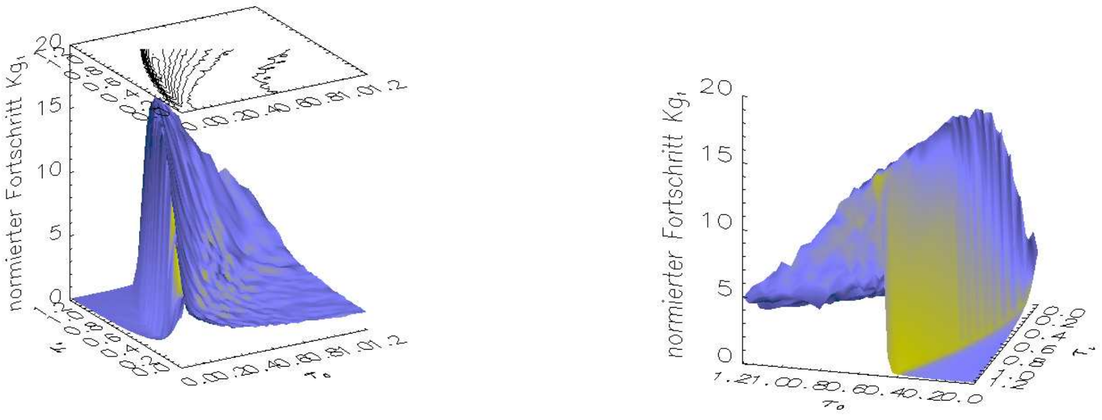


Abbildung A.23: (15 , 100), Skalierungstest, $n = 100$, $n_\sigma = n$

τ_0–τ_i–Kombinationen sortiert nach der größten Konvergenzgeschwindigkeit $Kg_1 = \frac{n}{Gen.} \cdot \log \sqrt{\frac{f_{Startwert}}{f_{Bestwert}}}$						
n	Rekotyp	Gen.	τ_0	τ_i	Kg_1	Kg_2
100	551	1000.00	0.2500	0.7000	1.9875	17.26
100	551	1000.00	0.2250	0.6500	1.9688	17.10
100	551	1000.00	0.1750	0.5500	1.9672	17.09
100	551	1000.00	0.1750	0.5250	1.9575	17.00
100	551	1000.00	0.2750	0.7500	1.9567	17.00
100	551	1000.00	0.2250	0.6750	1.9560	16.99
100	551	1000.00	0.2250	0.6250	1.9502	16.94
100	551	1000.00	0.2750	0.7250	1.9433	16.88
100	551	1000.00	0.2000	0.6250	1.9423	16.87
100	551	1000.00	0.2500	0.6750	1.9396	16.85

Die der τ_0–τ_i–Heuristik nächsten untersuchten Werte $\tau_0 = 0.075$ und $\tau_i = 0.225$ liefern $Kg_1 = 1.3195$ bzw. $Kg_2 = 11.46$.

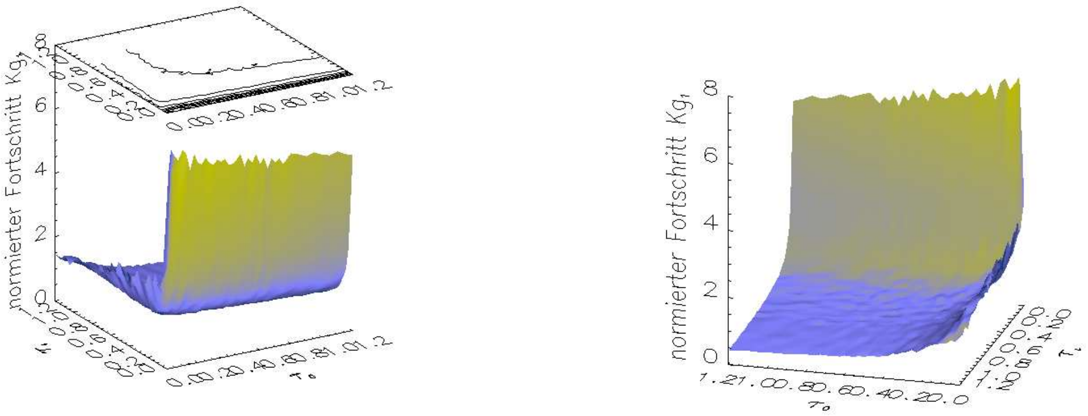


Abbildung A.24: (15 , 100), Skalierungstest, $n = 100$, $n_\sigma = n$, ohne Rekombination

τ_0–τ_i–Kombinationen sortiert nach der größten Konvergenzgeschwindigkeit $Kg_1 = \frac{n}{Gen.} \cdot \log \sqrt{\frac{f_{Startwert}}{f_{Bestwert}}}$

n	Rekotyp	Gen.	τ_0	τ_i	Kg_1	Kg_2
100	111	5000.00	0.0250	0.0000	0.1566	6.80
100	111	5000.00	0.1000	0.0000	0.1547	6.72
100	111	5000.00	0.1250	0.0000	0.1513	6.57
100	111	5000.00	0.0500	0.0000	0.1513	6.57
100	111	5000.00	0.1750	0.0000	0.1502	6.52
100	111	5000.00	0.0750	0.0000	0.1473	6.40
100	111	5000.00	0.2500	0.0000	0.1457	6.33
100	111	5000.00	0.2000	0.0000	0.1436	6.24
100	111	5000.00	0.4750	0.0000	0.1435	6.23
100	111	5000.00	0.0750	0.0250	0.1433	6.23

Die der τ_0–τ_i–Heuristik nächsten untersuchten Werte $\tau_0 = 0.075$ und $\tau_i = 0.225$ liefern $Kg_1 = 0.0305$ bzw. $Kg_2 = 1.32$.

A.3 Doppelsumme

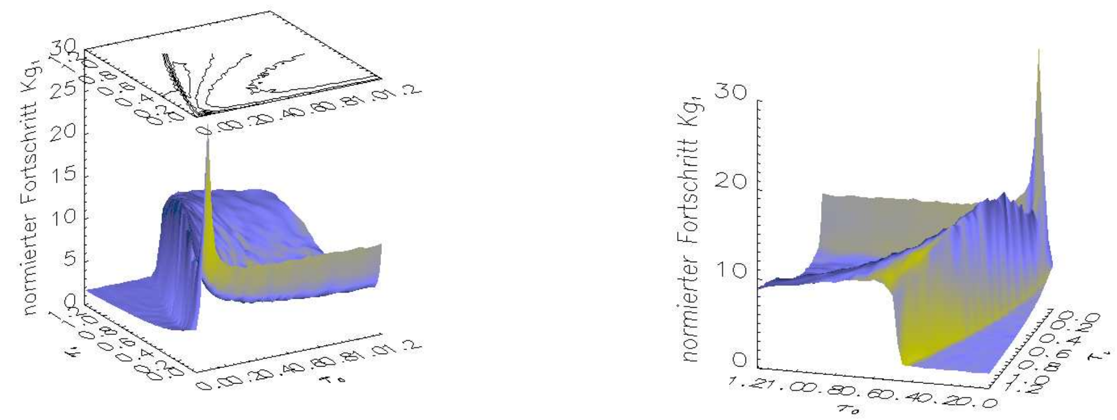


Abbildung A.25: (15 , 100), Doppelsumme, $n = 30$, $n_\sigma = n$

τ_0–τ_i–Kombinationen sortiert nach der größten Konvergenzgeschwindigkeit $Kg_1 = \frac{n}{Gen.} \cdot \log \sqrt{\frac{f_{Startwert}}{f_{Bestwert}}}$						
n	Rekotyp	Gen.	τ_0	τ_i	Kg_1	Kg_2
30	641	2000.00	0.0750	0.0000	0.5041	29.19
30	641	2000.00	0.0750	0.0250	0.3828	22.17
30	641	2000.00	0.1000	0.0000	0.3132	18.13
30	641	2000.00	0.0500	0.0000	0.2835	16.42
30	641	2000.00	0.0500	0.0250	0.2739	15.86
30	641	2000.00	0.2000	0.5000	0.2726	15.78
30	641	2000.00	0.2250	0.5500	0.2711	15.70
30	641	2000.00	0.1500	0.3750	0.2707	15.67
30	641	2000.00	0.2250	0.5250	0.2675	15.49
30	641	2000.00	0.1750	0.4500	0.2648	15.33

Die der τ_0–τ_i–Heuristik nächsten untersuchten Werte $\tau_0 = 0.125$ und $\tau_i = 0.3$ liefern $Kg_1 = 0.2466$ bzw. $Kg_2 = 14.28$.

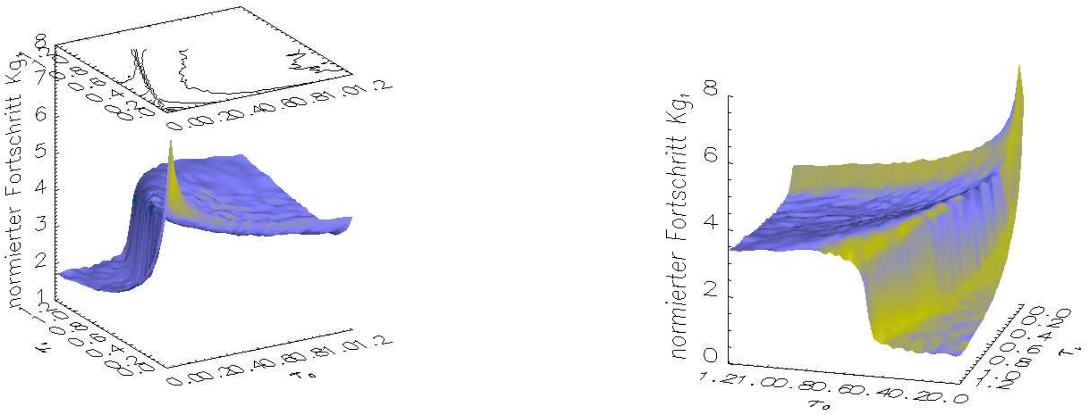


Abbildung A.26: (15 , 100), Doppelsumme, $n = 100$, $n_\sigma = n$

τ_0–τ_i–Kombinationen sortiert nach der größten Konvergenzgeschwindigkeit $Kg_1 = \frac{n}{Gen.} \cdot \log \sqrt{\frac{f_{Startwert}}{f_{Bestwert}}}$						
n	Rekotyp	Gen.	τ_0	τ_i	Kg_1	Kg_2
100	441	2000.00	0.0250	0.0000	0.4214	7.32
100	441	2000.00	0.0250	0.0250	0.3789	6.58
100	441	2000.00	0.0500	0.0000	0.3660	6.36
100	441	2000.00	0.0000	0.0000	0.3633	6.31
100	441	2000.00	0.0500	0.0250	0.3469	6.03
100	441	2000.00	0.0750	0.0000	0.3395	5.90
100	441	2000.00	0.0000	0.0250	0.3329	5.78
100	441	2000.00	0.0250	0.0500	0.3314	5.76
100	441	2000.00	0.0750	0.0250	0.3313	5.75
100	441	2000.00	0.1000	0.0000	0.3259	5.66

Die der τ_0–τ_i–Heuristik nächsten untersuchten Werte $\tau_0 = 0.075$ und $\tau_i = 0.225$ liefern $Kg_1 = 0.2720$ bzw. $Kg_2 = 4.73$.

A.4 Griewank

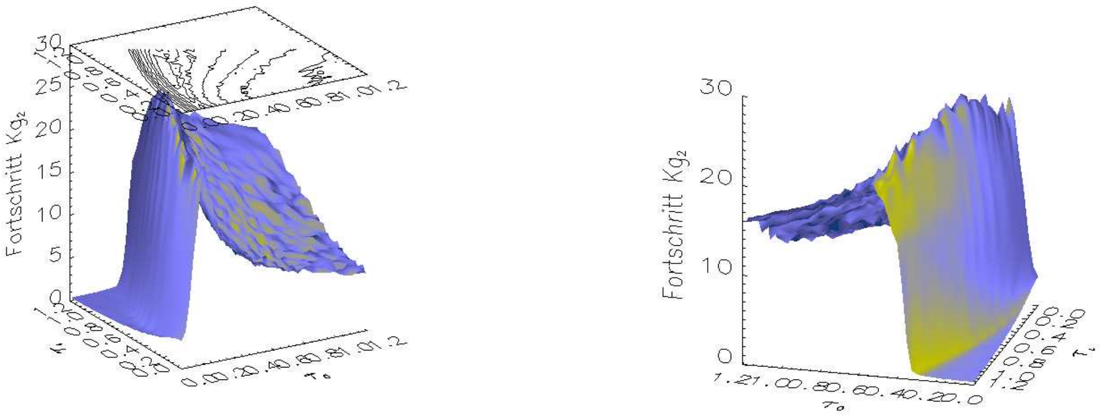


Abbildung A.27: (15 , 100), Griewank, $n = 30$, $n_\sigma = n$

τ_0–τ_i–Kombinationen sortiert nach der größten Konvergenzgeschwindigkeit $Kg_2 = \log \frac{f_{Startwert}}{f_{Bestwert}}$						
n	Rekotyp	Gen.	τ_0	τ_i	Kg_2	Kg_1
30	551	400.00	0.2500	0.5500	27.28	2.3556
30	551	399.56	0.2500	0.5250	27.04	2.3369
30	551	400.00	0.3000	0.6500	26.90	2.3230
30	551	400.00	0.2250	0.5250	26.80	2.3138
30	551	400.00	0.3000	0.7000	26.71	2.3064
30	551	400.00	0.2500	0.5750	26.45	2.2842
30	551	400.00	0.2750	0.5750	26.32	2.2727
30	551	400.00	0.2750	0.6000	26.27	2.2679
30	551	400.00	0.3250	0.8000	26.24	2.2657
30	551	400.00	0.2250	0.5500	26.15	2.2577

Die der τ_0–τ_i–Heuristik nächsten untersuchten Werte $\tau_0 = 0.125$ und $\tau_i = 0.3$ liefern $Kg_1 = 1.5142$ bzw. $Kg_2 = 17.54$.

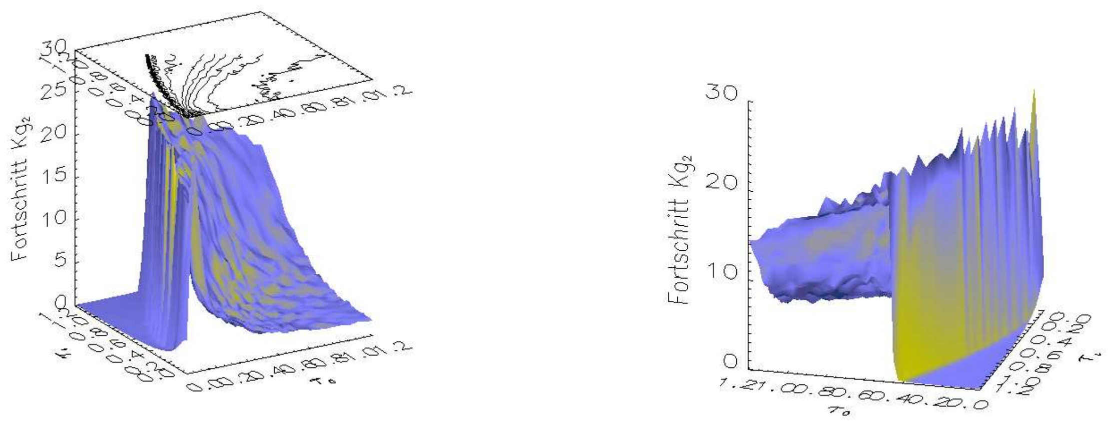


Abbildung A.28: (15 , 100), Griewank, $n = 100$, $n_\sigma = n$

τ_0–τ_i–Kombinationen sortiert nach der größten Konvergenzgeschwindigkeit $Kg_2 = \log \frac{f_{Startwert}}{f_{Bestwert}}$						
n	Rekotyp	Gen.	τ_0	τ_i	Kg_2	Kg_1
100	551	2069.48	0.2250	0.7250	25.45	1.4168
100	551	2255.96	0.1500	0.5500	24.88	1.2710
100	551	2100.76	0.0750	0.2750	24.86	1.3634
100	551	1948.80	0.0500	0.0000	24.81	1.4827
100	551	2406.28	0.1000	0.4000	24.59	1.1779
100	551	2176.92	0.1250	0.4750	24.54	1.3024
100	551	1792.60	0.1250	0.4500	24.42	1.5709
100	551	2071.12	0.0500	0.0250	24.27	1.3857
100	551	1790.48	0.1500	0.5250	24.24	1.5598
100	551	2438.92	0.0750	0.3000	24.19	1.1419

Die der τ_0–τ_i–Heuristik nächsten untersuchten Werte $\tau_0 = 0.075$ und $\tau_i = 0.225$ liefern $Kg_1 = 1.0860$ bzw. $Kg_2 = 19.56$.

A.5 multimodale Funktion

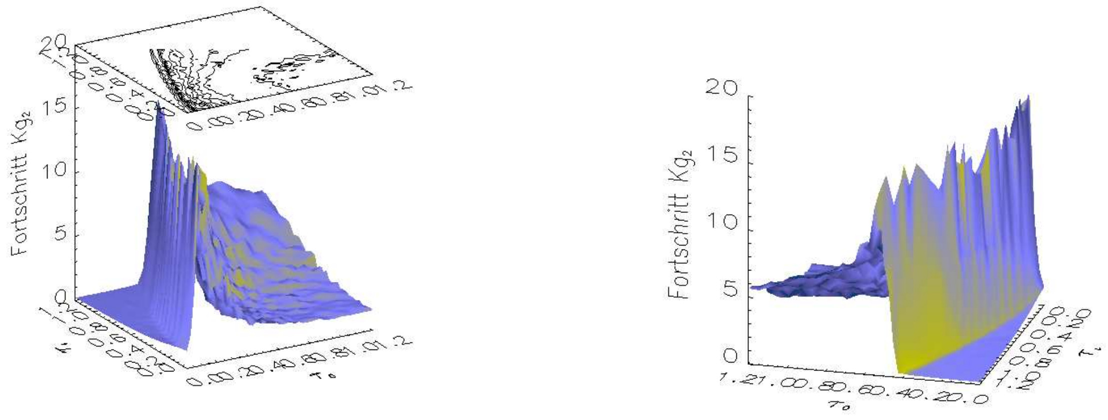


Abbildung A.29: (15 , 100), multimodale Funktion, $n = 30$, $n_\sigma = n$

τ_0–τ_i–Kombinationen sortiert nach der größten Konvergenzgeschwindigkeit $Kg_2 = \log \frac{f_{Startwert}}{f_{Bestwert}}$						
n	Rekotyp	Gen.	$\tau_\mathbf{0}$	$\tau_\mathbf{i}$	Kg_2	Kg_1
30	351	1397.84	0.1000	0.1000	15.91	0.3934
30	351	1997.76	0.0750	0.0250	15.80	0.2732
30	351	1615.76	0.1000	0.1500	15.55	0.3328
30	351	1988.24	0.0750	0.0000	15.32	0.2660
30	351	1464.28	0.2750	0.6750	15.21	0.3731
30	351	1319.24	0.3000	0.7250	15.14	0.4121
30	351	1252.28	0.1000	0.0000	15.14	0.4174
30	351	1517.92	0.1000	0.1250	15.09	0.3437
30	351	1214.88	0.2500	0.6000	14.77	0.4244
30	351	1161.88	0.5000	1.1500	14.77	0.4387

Die der τ_0–τ_i–Heuristik nächsten untersuchten Werte $\tau_0 = 0.125$ und $\tau_i = 0.3$ liefern $Kg_1 = 0.2055$ bzw. $Kg_2 = 11.90$.

Anhang B

Verzeichnisse

Abbildungsverzeichnis

Tabellenverzeichnis