

Über den dynamischen (1+1)-EA

Jens Jägersküpper

11. September 2001

Inhaltsverzeichnis

1	Einleitung	1
1.1	Motivation	1
1.2	Die Evolution als Vorbild	2
1.3	Der (1+1)-EA	3
1.4	Theoretische Analyse des (1+1)-EA	5
1.5	Sprachgebrauch und Notation	8
2	Die lineare Funktion BLIC_k	11
2.1	Lineare Funktionen	11
2.2	Motivation und Definition von BLIC_k	12
2.3	Laufzeit-Analyse für BLIC_k mit $k = \Theta(1)$	13
2.4	Laufzeit-Schranken für weitere lineare Funktionen	15
2.5	Erkenntnisse für BLIC_n	16
3	Strukturelle Erkenntnisse für lineare Funktionen	21
3.1	Warum große Mutationswahrscheinlichkeiten „eher ungünstig“ sind	21
3.2	Konkretisierung für BLIC_n	22
3.3	Intuitive Überlegungen zum dynamischen (1+1)-EA	24
3.4	Über die Verteilung der Einsen im Individuum	25
3.5	Die Wahrscheinlichkeit bestimmter Mutationen	33
4	Anwendung der strukturellen Erkenntnisse	43
4.1	Die erwartete Veränderung der Anzahl der Einsen	43
4.2	Die Laufzeit des Standard-(1+1)-EA auf linearen Funktionen	47
4.3	Überlegungen zum dynamischen (1+1)-EA auf linearen Funktionen	55
4.4	Zusammenfassung	60
A	Abschätzungen und Identitäten	63

Kapitel 1

Einleitung

1.1 Motivation

Nicht nur in sehr vielen Bereichen der Wirtschafts- und Ingenieurwissenschaften, sondern auch in vielen Naturwissenschaften und nicht zuletzt bei der Organisation des täglichen Lebens sind Optimierungsaufgaben zu lösen. Sei es die möglichst wirtschaftliche Nutzung der vorhandenen Ressourcen, wie z. B. die von LKWs und Fahrern in einem Speditionsunternehmen, oder aber die Größen- und Gewichtsoptimierung bei Fertigungsprozessen; unter den in der Regel sehr vielen Lösungen für das entsprechende Problem soll möglichst eine optimale Lösung gefunden werden.

Für viele Optimierungsaufgaben ist das Finden, d. h. das Berechnen, einer optimalen Lösung sehr schwierig (NP-hart). Das heißt, eine exakte Lösung zu berechnen, kann bei üblichen Problemgrößen sehr lange dauern – unter Umständen so lange, dass kein Mensch genug Zeit haben wird, auf die Lösung zu warten. Dies ist jedoch kein Grund, das Problem als „praktisch unlösbar“ abzustempeln und nicht weiter zu untersuchen. In vielen Fällen ist das exakte Lösen einer Optimierungsaufgabe gar nicht nötig, sondern eine „annähernd optimale“ Lösung reicht für den jeweiligen Zweck vollkommen aus. Der zeitliche Aspekt ist bei der Lösung von Optimierungsaufgaben also wichtig. Lieber eine mäßig gute Lösung nach fünf Minuten, als eine Woche auf eine optimale Lösung warten, wird sich beispielsweise der Routen-Planer in einem Speditionsunternehmen denken.

Bei der Lösung von Optimierungsaufgaben bietet sich unter dem Zeitaspekt als heuristische Vorgehensweise eine schrittweise Optimierung an. Ausgehend von einer trivialen, in der Regel unbefriedigenden, aber schnell zu berechnenden Lösung werden sukzessive bessere Lösungen entwickelt, so dass sich – hoffentlich – nach und nach dem Optimum genähert wird. Eine solche Optimierungsheuristik als Vorgehensweise erinnert an die Evolution in der Natur. Dort passen sich im Laufe der Generationen die Vertreter einer Spezies immer besser an ihre Umwelt an. Optimierungsverfahren bzw. -heuristiken, die sich an natürlichen Prozessen orientieren bzw.

versuchen, diese zu simulieren, werden „natural analog“ genannt.

Natural analoge Optimierungsverfahren werden heute in der Praxis tatsächlich angewendet. In manchen Fällen mit großem Erfolg, in machen eher ohne Erfolg. Die meisten Erkenntnisse über solche Verfahren und Heuristiken sind experimenteller Natur, d. h. es wird versucht, diese an ein konkretes Problem anzupassen, und dann ausprobiert, ob gute oder schlechte Ergebnisse resp. Lösungen erzielt werden. Theoretisch fundierte Erkenntnisse, warum welches natural analoge Verfahren welches Problem gut oder schlecht optimiert, sind immer noch Mangelware, da sich eine theoretische, mathematische Analyse der Abläufe bei einem solchen Verfahren meist als schwierig erweist.

1.2 Die Evolution als Vorbild

Wie schon erwähnt wurde, bietet sich für eine schrittweise Optimierung ein Vorgehen analog der Evolution in der Natur an. Diese Idee entstand in den 60er Jahren¹, nachdem schon zuvor „entdeckt“ worden war, dass in Algorithmen ein Zufallsmoment dazu genutzt werden kann, um Verbesserungen in der (erwarteten) Laufzeit erzielen zu können. Als Untergruppe der randomisierten Algorithmen – „randomisiert“ heißt, dass Zufall verwendet wird – bildete sich somit die Gruppe der evolutionären Algorithmen. Zu dieser Gruppe zählen alle Algorithmen, meist Optimierungsheuristiken, die die natürliche Evolution als Vorbild haben. Die anfänglichen Erfolge führten dazu, dass in sehr vielen Bereichen versucht wurde, passende evolutionäre Algorithmen zu entwerfen und zu verwenden. Unter anderem wurden genetische Algorithmen [Gol89] und die genetische Programmierung [BNKF98] entwickelt. Des weiteren zählen Evolutionsstrategien [Sch95] und die evolutionäre Programmierung zu den evolutionären Algorithmen. Ein Überblick über die verschiedenen Anwendungsmöglichkeiten und -gebiete von evolutionären Algorithmen ist in [Bäc94] zu finden. Im Bereich der Informatik beschäftigen sich vor allem die Robotik, die Muster-Erkennung und -Verarbeitung und die künstliche Intelligenz mit evolutionären Algorithmen. Häufig angewendet werden evolutionäre Algorithmen heute außerdem im Bereich des Operations Research und bei der Optimierung von Bauteilen.

Ebendiese Optimierung von Bauteilen, genauer die Minimierung von Strömungsverlusten, führte zu der Entwicklung der Evolutionsstrategien. Bienert, Rechenberg und Schwefel gingen dabei von der Erkenntnis aus, dass die Natur im Laufe der Zeit äußerst strömungsgünstige Formen „entwickelt“ hat, so dass ein evolutionärer Algorithmus möglicherweise bei der Verbesserung der Bauteil-Form, d. h. bei der Verringerung der Strömungsverluste, von Nutzen sein könnte (vgl. [Sch95] und [Rec94]). Tatsächlich wurde eine bessere Form gefunden, die überraschenderweise intuitiv bzw. „augenscheinlich“ nicht für besonders strömungsgünstig gehalten wurde. Ein

¹des vorigen, 20. Jahrhunderts

Bauteil-Designer wäre also wohl nie auf die Idee gekommen, eine solche Form zu wählen.²

Bei einer Evolutionsstrategie wird eine konkrete Lösung eines Optimierungsproblems als Individuum aufgefasst. Die Güte einer Lösung bestimmt dann die Fitness des entsprechenden Individuums. Eine Population besteht wiederum aus einer gewissen Anzahl von Individuen. Die für einen evolutionären Algorithmus typische schrittweise Optimierung besteht nun darin, dass aus einer Population eine nächste Population erzeugt wird. Die neue Generation sollte dabei möglichst bessere Individuen enthalten als diejenige, aus der sie hervorgegangen ist. Wie bei der natürlichen Evolution, werden bei der Erzeugung der nächsten Generation Mutation, Rekombination und Selektion verwendet. Eine Mutation erzeugt durch eine zufällige Veränderung eines Individuums ein neues Individuum. Bei der Rekombination wird aus zwei oder mehr Individuen mindestens ein neues Individuum erzeugt. Innerhalb der Selektion wird anhand der Fitness entschieden, welche der alten und welche der neu erzeugten Individuen die nächste Generation bilden. Hat in der Natur neben der Fitness der Individuen auch der Zufall Einfluss auf den Selektionsprozess, so werden bei Evolutionsstrategien die Selektionsentscheidungen (in der Regel) deterministisch, d. h. ausschließlich auf der Fitness basierend, getroffen.

Evolutionsstrategien wurden von Schwefel danach klassifiziert, wie viele Individuen eine Population bilden und ob „alte“ Individuen überhaupt an der Selektion teilnehmen, oder nur unter den neu erzeugten Individuen selektiert wird: Bei einer (μ, λ) -Strategie besteht eine Population aus μ Individuen. Aus den μ Individuen einer Population werden durch Mutation und Rekombination λ neue erzeugt und von diesen neuen werden wiederum μ selektiert, die dann die nächste Generation bilden. Eine $(\mu + \lambda)$ -Strategie unterscheidet sich von einer (μ, λ) -Strategie genau darin, dass neben den λ neu erzeugten auch die μ Individuen, aus denen die neuen erzeugt wurden, an der Selektion teilnehmen.

1.3 Der (1+1)-EA

Bei dem in dieser Arbeit betrachteten naturanalogen Optimierungsverfahren handelt es sich um eine (1+1)-Evolutionsstrategie. Da also eine Populationsgröße von eins verwendet wird, kann offensichtlich keine (sinnvolle) Rekombination stattfinden. Die simulierte Evolution beruht folglich ausschließlich auf Mutationen. Die Individuen sind Vektoren bzw. Bit-Sequenzen aus $\{0, 1\}^n$ und deren jeweilige Fitness wird durch eine Fitness-Funktion $f : \{0, 1\}^n \rightarrow \mathbb{R}$ bestimmt.

Da Evolutionsstrategien im Allgemeinen eine Fitness-Funktion $g : \mathbb{R}^n \rightarrow \mathbb{R}$ optimieren, d. h. die betrachteten Individuen sind nicht aus $\{0, 1\}^n$, sondern aus \mathbb{R}^n , wurde für den hier betrachteten Algorithmus die Bezeichnung „(1+1)-Evolutionärer-Algorithmus“ („(1+1)-EA“) gewählt.

²Der „Lotos-Effekt“ und die „Hai-Haut“ sind weitere Beispiele, bei denen sich „augenscheinlich optimale Lösungen“ als suboptimal herausgestellt haben.

Das Ziel beim (1+1)-EA ist natürlich, ein Individuum mit optimaler Fitness zu finden. Im Folgenden wird davon ausgegangen, dass die Fitness (-Funktion) maximiert werden soll, da anderenfalls die Funktion $-f$ betrachtet werden kann.

Definition 1.3.1. [Statischer (1+1)-EA] Sei $f : \{0, 1\}^n \rightarrow \mathbb{R}$ für $n \geq 3$ (damit $1/n < 1/2$ gilt) die zu maximierende Fitness-Funktion und $p(n) \in [1/n; 1/2)$ die zu verwendende Mutationswahrscheinlichkeit. Ein Ablauf des statischen (1+1)-EA mit der Mutationswahrscheinlichkeit $p(n)$ auf der Funktion f besteht dann in der Ausführung der folgenden Anweisungen:

1. Wähle $x \in \{0, 1\}^n$ zufällig mit Gleichverteilung.
2. Erzeuge x' , indem jedes Bit in x mit Wahrscheinlichkeit $p(n)$ negiert wird.
3. Setze $x := x'$ genau dann, wenn $f(x') \geq f(x)$ gilt.
4. Weiter bei 2.

Das in Zeile 1 erzeugte Individuum wird „Start-Individuum“ genannt. Eine Ausführung der Anweisung in Zeile 2 wird „Mutationsschritt“ genannt und eine Ausführung der Anweisung in Zeile 3 „Selektionsschritt“. Ein Mutationsschritt und der direkt folgende Selektionsschritt werden zusammen „Schritt“ bzw. „Mutation-Selektion-Zyklus“ genannt. Das Individuum, das im Mutationsschritt eines Mutation-Selektion-Zyklus mutiert wird, wird „das (in diesem Schritt) aktuelle Individuum“ genannt.

Der (1+1)-EA erzeugt eine unendliche Folge von Individuen aus $\{0, 1\}^n$, da in Zeile 4 ein Abbruch-Kriterium fehlt und er somit offensichtlich nicht terminiert. Es handelt sich streng genommen also nicht um einen Algorithmus. Für die theoretische Analyse stellt dies jedoch kein Problem dar. Als Laufzeit wird die Anzahl der Mutation-Selektion-Zyklen betrachtet, bis das aktuelle Individuum eine bestimmte Eigenschaft besitzt, z. B. die, bezüglich f optimal zu sein. In der praktischen Anwendung von naturanalogen Optimierungsverfahren spielt die Festlegung des Abbruch-Kriteriums hingegen eine wichtige Rolle. Ein zu früher Abbruch der Evolution wird zu einem unnötig schlechten Individuum führen und ein zu langes Warten wird nur die Laufzeit unnötig erhöhen, ohne dass eine weitere Verbesserung eintritt.

Da evolutionäre Algorithmen oft dann angewendet werden, wenn die zu optimierende Funktion nicht bekannt ist, ist a priori nicht klar, welche Mutationswahrscheinlichkeit $p(n)$ zu einer möglichst schnellen bzw. guten Optimierung führt. Droste, Jansen und Wegener haben in [DJW99] gezeigt, dass es Funktionen gibt, bei denen sogar die Variation der Mutationswahrscheinlichkeit während der Optimierung nötig ist, um eine akzeptable Laufzeit erzielen zu können. Dazu wurde der (1+1)-EA um eine einfache dynamische Kontrolle der Mutationswahrscheinlichkeit erweitert:

Definition 1.3.2. [Dynamischer (1+1)-EA] Sei $f : \{0, 1\}^n \rightarrow \mathbb{R}$ für $n \geq 3$ die zu maximierende Fitness-Funktion. Ein Ablauf des dynamischen (1+1)-EA auf der Funktion f besteht dann in der Ausführung der folgenden Anweisungen:

1. Wähle $x \in \{0, 1\}^n$ zufällig mit Gleichverteilung. Setze $p := 1/n$.
2. Erzeuge x' , indem jedes Bit in x mit Wahrscheinlichkeit p negiert wird.
3. Setze $x := x'$ genau dann, wenn $f(x') \geq f(x)$ gilt.
4. Falls $p < 1/4$ gilt, setze $p := 2 \cdot p$, sonst setze $p := 1/n$.
5. Weiter bei 2.

Die Begriffe „Mutationsschritt“, „Selektionsschritt“ usw. werden analog zur statischen Variante des (1+1)-EA verwendet.

Während eines Ablaufs des dynamischen (1+1)-EA wird mit der Mutationswahrscheinlichkeit $1/n$ begonnen und diese nach jedem Schritt verdoppelt, solange sie kleiner als $1/2$ bleibt. Anschließend wird die Mutationswahrscheinlichkeit auf $1/n$ zurückgesetzt und das Vorgehen wiederholt. Es werden also Mutationswahrscheinlichkeit aus $\{p = 2^i/n \mid 0 \leq i \leq \lceil \log n \rceil - 2\}$ verwendet. Auch der dynamische (1+1)-EA terminiert offensichtlich nicht, sodass für die Laufzeit dasselbe zutrifft wie bei der statischen Variante.

1.4 Theoretische Analyse des (1+1)-EA

Evolutionäre Algorithmen werden häufig dann angewendet, wenn zu einer gegebenen Lösung eines Optimierungsproblems leicht deren Güte bestimmt werden kann, jedoch der Zusammenhang zwischen einer Lösung und ihrer Güte entweder derart komplex, oder auch nur „unerforscht“ ist, dass eine exakte Optimierung inakzeptabel lange dauern würde, oder aber kein exaktes Optimierungsverfahren zur Verfügung steht. Das heißt, die einzige Möglichkeit, an Informationen über die zu optimierende „Güte-Funktion“ zu gelangen, ist das Auswerten dieser Funktion auf konkreten Lösungen des Optimierungsproblems. Nutzt ein Optimierungsverfahren ausschließlich die Möglichkeit, die zu optimierende Funktion auszuwerten, ohne die Funktion selbst zu kennen, so wird von einer „Black-Box-Optimierung“ gesprochen. Denn für das Optimierungsverfahren bleibt dann „im Dunklen“, wie der Zusammenhang zwischen einer Lösung und ihrer Güte ist.

Der (1+1)-EA ist ein „Black-Box-Optimierungsverfahren“ und so bietet sich als Maß für die Laufzeit die Anzahl der Auswertungen der zu optimierenden Funktion an. Da in einem Schritt des (1+1)-EA die zu optimierende Funktion genau einmal für den erzeugten Mutanten

ausgewertet wird – vorausgesetzt, neben dem jeweils aktuellen Individuum wird auch dessen jeweilige Fitness gespeichert –, wird die Laufzeit des (1+1)-EA „anschaulich“ in der Anzahl der durchlaufenen Mutation-Selektion-Zyklen gemessen. Eine Analyse der Laufzeit des (1+1)-EA ist nur dann (sinnvoll) möglich, wenn die Funktion bekannt ist, die maximiert werden soll. Die Optimierung ist beendet bzw. gilt als beendet, wenn das aktuelle Individuum erstmals optimal ist, d. h. ein Individuum mit maximaler Fitness erzeugt wurde. Da im Mutationsschritt Zufall verwendet wird, handelt es sich bei einem Ablauf des (1+1)-EA um einen stochastischen Prozess, sodass über den Erwartungswert der Laufzeit, kurz „die erwartete Laufzeit“ argumentiert werden muss. Die dafür notwendige Analyse dieses stochastischen Prozess gestaltet sich meist schwierig. Hilfreich dabei kann manchmal der folgende Sachverhalt sein. Da der (1+1)-EA ein Black-Box-Optimierungsverfahren ist und er anhand des Vergleiches der Fitness zweier Individuen selektiert, ist nicht die absolute Fitness, sondern die relative Fitness eines Individuums (in Relation zu den übrigen Individuen) entscheidend für der Verlauf der Optimierung.

Definition 1.4.1. *Zwei Funktionen $f, g : \{0, 1\}^n \rightarrow \mathbb{R}$ sind für den (1+1)-EA genau dann äquivalent, wenn*

$$(\forall x, y \in \{0, 1\}^n) f(x) \leq f(y) \Leftrightarrow g(x) \leq g(y)$$

gilt. Dies wird durch „ $f \equiv_{EA} g$ “ ausgedrückt.

Aus $f \equiv_{EA} g$ folgt für $x, y \in \{0, 1\}^n$:

$$\begin{aligned} f(x) = f(y) &\Leftrightarrow f(x) \leq f(y) \wedge f(x) \geq f(y) \\ &\Leftrightarrow g(x) \leq g(y) \wedge g(x) \geq g(y) \Leftrightarrow g(x) = g(y) \end{aligned}$$

Gilt $f \equiv_{EA} g$, so existiert also eine streng monoton wachsende Funktion $\psi : \mathbb{R} \rightarrow \mathbb{R}$, sodass $f = \psi \circ g$ gilt. Der zuvor angedeutete Sachverhalt kann nun wie folgt formuliert werden:

Proposition 1.4.2. *Ein³ (1+1)-EA besitzt für zwei Funktion $f, g : \{0, 1\}^n \rightarrow \mathbb{R}$, für die $f \equiv_{EA} g$ gilt, dieselbe erwartete Laufzeit.*

Soll die Laufzeit für eine Funktion f untersucht werden, so kann also statt dessen eine Funktion g , für die $f \equiv_{EA} g$ gilt, analysiert werde. Außerdem gelten die gewonnen Erkenntnisse dann für alle Funktionen der Äquivalenzklasse von g bezüglich \equiv_{EA} , zu der auch f gehört. Auch der folgende Sachverhalt kann eine Analyse unter Umständen erleichtern:

Lemma 1.4.3. *Seien $w \in \{0, 1\}^n$ und $f : \{0, 1\}^n \rightarrow \mathbb{R}$ fest gewählt. Ferner sei die Funktion $\oplus_w : \{0, 1\}^n \rightarrow \{0, 1\}^n$ definiert durch $\oplus_w(x) := x \oplus w$, durch die bitweise XOR-Verknüpfung*

³d. h. der statische, unabhängig von der gewählten Mutationswahrscheinlichkeit, und der dynamische

der Individuen x und w . Ein (1+1)-EA besitzt dann für die beiden Funktionen f und $f \circ \oplus_w$ die gleiche erwartete Laufzeit.

Beweis: Die XOR-Verknüpfung ist kommutativ und assoziativ. Für $x, y \in \{0, 1\}^n$ gilt

$$x \oplus y = w \oplus (x \oplus y) \oplus w = (w \oplus x) \oplus (y \oplus w),$$

d. h. x und y unterscheiden sich genau an den Positionen, an denen sich $\oplus_w(x)$ und $\oplus_w(y)$ unterscheiden. Dass $\oplus_w \circ \oplus_w$ die Identität auf $\{0, 1\}^n$ ist, hat ferner zu Folge, dass

$$f(x) \leq f(y) \Leftrightarrow f \circ \oplus_w(\oplus_w(x)) \leq f \circ \oplus_w(\oplus_w(y))$$

gilt. Wenn ein (1+1)-EA die Funktion f maximiert, dabei x^0 als Start-Individuum erzeugt und die zufälligen Entscheidungen im Mutationsschritt zu der Individuen-Folge x^0, x^1, \dots führen, so würden ebendiese Entscheidungen bei der Maximierung der Funktion $f \circ \oplus_w$ durch ebendiesen (1+1)-EA zu der Individuen-Folge $\oplus_w(x^0), \oplus_w(x^1), \dots$ führen, wenn $\oplus_w(x^0)$ als Start-Individuum erzeugt würde. Wenn x^0 gleichverteilt zufällig aus $\{0, 1\}^n$ gewählt wird, so ist auch $\oplus_w(x^0)$ gleichverteilt zufällig aus $\{0, 1\}^n$. Dass ein (1+1)-EA zur Maximierung von f die Individuen-Folge x^0, x^1, \dots erzeugt, ist also genauso wahrscheinlich, wie dass dieser (1+1)-EA zur Maximierung von $f \circ \oplus_w$ die Individuen-Folge $\oplus_w(x^0), \oplus_w(x^1), \dots$ erzeugt. \square

Soll eine Funktion f analysiert werden, so kann also statt dessen immer eine Funktion analysiert werden, für die das Individuum 1^n optimal ist. Sei y ein Individuum, sodass $f(y)$ maximal ist, und \bar{y} bezeichne das bitweise Komplement von y . Dann ist nämlich $1^n (= y \oplus \bar{y})$ ein optimales Individuum für die Funktion $f \circ \oplus_{\bar{y}}$.

In den Anfängen der theoretischen Analyse des (1+1)-EA wurde zunächst die statische Variante mit der Mutationswahrscheinlichkeit $p(n) = 1/n$ untersucht. Diese wird im Folgenden „Standard-(1+1)-EA“ genannt, da die „Standard-Mutationswahrscheinlichkeit“ $1/n$ verwendet wird. Zu den ersten Ergebnissen im Rahmen der theoretischen Untersuchung der Laufzeit des Standard-(1+1)-EA gehört die Analyse der Funktion ONEMAX : $\{0, 1\}^n \rightarrow \mathbb{N}$ mit $\text{ONEMAX}(x) = x_1 + \dots + x_n$ durch Bäck [Bäc92] und Mühlenbein [Müh92]. Diese Analyse wurde später von Rudolph in [Rud97] überarbeitet. Darin wird eine $O(n \log n)$ -Laufzeit-Schranke anhand der Analyse des zugrunde liegenden stochastischen Prozess, den der Standard-(1+1)-EA auf der Funktion ONEMAX darstellt, ausführlich bewiesen. Droste, Jansen und Wegener konnten in [DJW98] zeigen, dass die Laufzeit des Standard-(1+1)-EA für jede lineare Funktion – eine Definition von „linear“ folgt später – durch $O(n \log n)$ beschränkt ist. Außerdem werden Funktionen angegeben, die der Standard-(1+1)-EA nicht effizient optimiert, d. h. die erwartete Laufzeit ist für diese Funktionen exponentiell. Als ein Grund für eine schlechte, d. h. große, erwartete

Laufzeit wird die „kleine“ Mutationswahrscheinlichkeit von $1/n$ erkannt. Jansen und Wegener untersuchen in [JW00b] die Auswirkung der verwendeten Mutationswahrscheinlichkeit auf die Laufzeit des (1+1)-EA genauer. Für die Funktion PATHTOJUMP wird u. a. gezeigt, dass der statische (1+1)-EA diese für jede Mutationswahrscheinlichkeit mit „sehr hoher“ Wahrscheinlichkeit nicht effizient optimiert, d. h. superpolynomiell viele Schritte zur Optimierung nötig sind, der dynamische (1+1)-EA sie jedoch mit „überwältigender“ Wahrscheinlichkeit in $O(n^3 \log n)$ Schritten optimiert. In [JW00a] betrachten Jansen und Wegener den dynamischen (1+1)-EA und zeigen Laufzeit-Schranken für weitere Funktionen und Funktionenklassen. Insbesondere wird mit einer „anschaulichen“, „relativ einfachen“ Methode für die Klasse der linearen Funktionen gezeigt, dass die Laufzeit des dynamischen (1+1)-EA durch $O(n^2 \log n)$ beschränkt ist. Für die beiden linearen Funktionen ONEMAX und BINVALUE wird jeweils eine $O(n \log^2 n)$ -Laufzeit-Schranke bewiesen, wobei die Schranke für BINVALUE ungleich aufwändiger bzw. „Trick-reicher“ ist als die für ONEMAX.

Die Vermutung liegt somit nahe, dass der dynamische (1+1)-EA für jede lineare Funktion, die von allen Bits „essentiell“ abhängt, eine erwartete Laufzeit von $O(n \log^2 n)$ hat. Diese Vermutung „ein wenig zu erhärten“, ist das Thema dieser Arbeit.

1.5 Sprachgebrauch und Notation

„Offene Intervallgrenzen“ werden durch „(“ bzw. „)“ ausgedrückt und „abgeschlossene Intervallgrenzen“ durch „[“ bzw. „]“. Beispielsweise sind in dem Intervall $(1,1 ; 2,2]$ alle reellen Zahlen enthalten, die größer als 1,1, aber nicht größer als 2,2 sind.

Wird im Folgenden von einer „Phase“ im Ablauf des dynamischen (1+1)-EA gesprochen, so ist damit eine Folge von Mutation-Selektion-Zyklen gemeint, in der jede der verschiedenen Mutationswahrscheinlichkeiten aus $[1/n ; 1/2)$, beginnend mit $1/n$, genau einmal verwendet wird. Eine Phase besteht also aus $\lceil \log n \rceil - 2$ Schritten, wobei mit „log“ immer „log₂“ gemeint ist.

Soll ein Zusammenhang sowohl für den (1+1)-EA – womit der statische gemeint ist – als auch für den dynamischen (1+1)-EA gelten, so wird dies durch eckige Klammern wie im nachfolgenden Satz ausgedrückt.

Wird in einem Schritt des [dynamischen] (1+1)-EA der erzeugte Mutant selektiert, so wird genau dieser Sachverhalt auch durch „der Mutant wird akzeptiert“ oder „die Mutation wird akzeptiert“ ausgedrückt. Wird in einem Schritt der erzeugte Mutant nicht selektiert, so wird dieser Mutant „verworfen“. Genauso wird auch eine Mutation, die nicht akzeptiert wird, „verworfen“. In diesem Zusammenhang ist „verwerfen“ also die Negation von „akzeptieren“.

Ein Schritt des [dynamischen] (1+1)-EA ist ein „erfolgreicher Schritt“, wenn der darin erzeugte Mutant eine echt bessere, hier also größere Fitness besitzt als sein Eltern-Individuum.

Die Mutation innerhalb eines erfolgreichen Schritts wird „erfolgreiche Mutation“ genannt. Eine erfolgreiche Mutation wird also immer akzeptiert, aber eine akzeptierte Mutation muss nicht unbedingt erfolgreich gewesen sein – Mutanten gleicher Fitness werden ja akzeptiert. Eine Phase des dynamischen (1+1)-EA ist eine „erfolgreiche Phase“, wenn mindestens einer ihrer Schritte erfolgreich war.

Das „Fitness-Defizit“ eines betrachteten Individuums ist die Differenz zwischen der optimalen Fitness, also dem bestmöglichen, hier größtmöglichen Wert, den die Fitness-Funktion annehmen kann, und seiner eigenen Fitness.

Bei linearen Funktionen (vgl. Definition 2.1.2 (auf Seite 11)) hat das Flippen eines Bits auf die Änderung der Fitness unabhängig von der Belegung aller übrigen Bits immer die gleiche Wirkung, nämlich die Erhöhung bzw. Verminderung der Fitness um den Wert des zugehörigen Koeffizienten. In diesem Sinn kann bei linearen Funktionen von dem „Wert eines Bits“ gesprochen werden.

In diesem Text gilt $\mathbb{N} = \{1, 2, \dots\}$ und $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. Ferner gilt $\mathbb{Z}^+ = \mathbb{N}$ und $\mathbb{Z}^- = \mathbb{Z} \setminus \mathbb{N}_0$. Ebenso enthält \mathbb{R}^+ alle positiven reellen Zahlen usw. Die Null ist weder positiv, noch negativ. Soll die Null enthalten sein, so wird dies durch \mathbb{Z}_0^+ , \mathbb{R}_0^+ usw. ausgedrückt. Für eine endliche Menge M bezeichnet „ $\#M$ “ die Anzahl der enthaltenen Elemente.

Die Länge einer Bit-Sequenz $b \in \{0, 1\}^*$ wird durch die Anzahl der Bits gegeben und mit „ $|b|$ “ bezeichnet. Die Anzahl der Einsen in b wird mit „ $|b|_1$ “ und die der Nullen mit „ $|b|_0$ “ bezeichnet. Für zwei Individuen $a, b \in \{0, 1\}^n$ bezeichnet „ $H(a, b)$ “ die Hamming-Distanz zwischen diesen beiden Bit-Sequenzen, d. h. es gilt $H(a, b) = \#\{i \mid 1 \leq i \leq n \wedge a_i \neq b_i\} = |a \oplus b|_1$.

Kapitel 2

Die lineare Funktion BLIC_k

2.1 Lineare Funktionen

Da der (1+1)-EA eine Funktion $f : \{0, 1\}^n \rightarrow \mathbb{R}$ maximiert, müssen für solche Funktionen zunächst einige Begriffe definiert bzw. erläutert werden, bevor die spezielle Funktion BLIC_k definiert und untersucht werden kann. Der Definitionsbereich $\{0, 1\}^n$ einer solchen „pseudo-booleschen“ Funktion f ist diskret und endlich. Jede pseudo-boolesche Funktion kann also als Polynom dargestellt werden. Wie reelle Polynome können pseudo-boolesche Funktionen nach ihrem Grad unterschieden bzw. klassifiziert werden.

Definition 2.1.1. Eine Funktion $f : \{0, 1\}^n \rightarrow \mathbb{R}$ sei für $N \geq 0$ durch

$$f(x) = \sum_{1 \leq i \leq N} w_i \cdot \prod_{j \in J_i} x_j$$

dargestellt, wobei $w_i \in \mathbb{R} \setminus \{0\}$ sowie $J_i \subseteq \{1, \dots, n\}$ für $1 \leq i \leq N$ gelte. Diese Funktion besitzt dann den Grad $k := \max \bigcup_{1 \leq i \leq N} \{\#J_i\} \cup \{0\}$.¹

Für $J_i = \emptyset$ nimmt das entsprechende, leere Produkt den (bezüglich der Multiplikation neutralen) Wert 1 an, sodass dann ein konstanter, d. h. von x unabhängiger, Summand w_i vorliegt. Lineare Funktionen sind – wie üblich – Funktionen vom Grad eins:

Definition 2.1.2. Eine Funktion $f : \{0, 1\}^n \rightarrow \mathbb{R}$ heißt genau dann „linear“, wenn sie den Grad $k = 1$ besitzt, d. h. es gilt

$$f(x) = \tau + \sum_{i=1}^n w_i \cdot x_i$$

¹Für $N = 0$ liegt also die konstante Funktion $f(x) = 0$ mit dem Grad $k = \max\{0\} = 0$ vor.

für entsprechende $\tau, w_1, \dots, w_n \in \mathbb{R}$, wobei für mindestens ein $i \in \{1, \dots, n\}$ $w_i \neq 0$ gilt. Jedes Bit x_i „besitzt“ bei einer linearen Funktion „einen Wert“ („ein Gewicht“) in der Größe des ihm zugeordneten Koeffizienten w_i .

Angenommen, ein (1+1)-EA würde für eine fest gewählte Funktion $f : \{0, 1\}^n \rightarrow \mathbb{R}$ im Erwartungsfall $\theta(2^n)$ Schritte für deren Maximierung benötigen. Da insbesondere n fest gewählt ist, liegt also eine konstante erwartete Laufzeit vor. Genauso läge eine konstante erwartete Laufzeit vor, wenn die erwartete Anzahl von Schritten bis zur Optimierung von f durch $\theta(n^2)$ beschränkt wäre. Wenn also die Laufzeit bezüglich n , der Anzahl der Bits im Individuum, analysiert werden soll, so sollte n variabel sein, die Funktion f also „sinnvoll“ in n skalierbar sein. Zum Beispiel so wie die Funktion ONEMAX , die die Anzahl der Einsen im Individuum „zählt“. ONEMAX ist also für $n \in \mathbb{N}$ skalierbar. Denkbar ist auch, dass eine Funktion für ein konstantes $k \in \mathbb{N}$ nur für $n \in k\mathbb{N}$ skalierbar ist, oder Ähnliches. Um die Laufzeit bezüglich n (sinnvoll) abzuschätzen, ist also eine Funktionenfolge zu betrachten. Eine „wenig sinnvolle“ Skalierung würde beispielsweise die Funktionenfolge $g_n : \{0, 1\}^n \rightarrow \mathbb{R}$, $n \in \mathbb{N}$, mit $g_n(x) := x_1$ darstellen. Egal wie viele Bits das Individuum hat, die Fitness hängt nur von dem einen Bit x_1 ab. Wie an diesem Beispiel zu sehen ist, ist es (in der Regel) nicht sinnvoll, die Laufzeit bezüglich der Anzahl der Bits im Individuum abzuschätzen, wenn die Funktion gar nicht von allen Bits, d. h. von dieser Anzahl, abhängt. In der Praxis würden in einem Individuum, das eine Lösung eines Optimierungsproblems beschreibt, ja auch nur Informationen kodiert, die Einfluss auf die Ausprägung einer Lösung haben. Im Folgenden wird deshalb davon ausgegangen, dass eine lineare Funktion „essenziell“ von allen Bits abhängt, d. h. $w_1, \dots, w_n \in \mathbb{R} \setminus \{0\}$ gilt.

Für eine Funktion $f(x)$ und die Funktion $f(x) - \tau$, $\tau \in \mathbb{R}$, gilt offensichtlich $f(x) \equiv_{EA} f(x) - \tau$. Werden im Folgenden lineare Funktionen analysiert, so wird also angenommen, dass diese die Form $f(x) = \sum_{i=1}^n w_i \cdot x_i$ mit $w_1, \dots, w_n \in \mathbb{R} \setminus \{0\}$ haben. Für eine solche Funktion ist das optimale Individuum eindeutig. Im optimalen Individuum ist nämlich das Bit x_i auf 1 gesetzt, wenn $w_i > 0$ gilt, und anderenfalls auf 0. Da statt einer solchen Funktion wiederum eine Funktion betrachtet werden kann, für die das Individuum 1^n optimal ist, kann bei der Argumentation über lineare Funktion angenommen werde, dass eine solche die Form $f(x) = \sum_{i=1}^n w_i \cdot x_i$ mit $w_1, \dots, w_n \in \mathbb{R}^+$ hat. Sollen Aussagen über den (1+1)-EA auf linearen Funktionen – die essenziell von allen Bits abhängen – getroffen werden, so kann also o. B. d. A. über lineare Funktion mit ausschließlich positiven Koeffizienten argumentiert werden.

2.2 Motivation und Definition von BLIC_k

Im Folgenden wird der Ablauf des dynamischen (1+1)-EA auf linearen Funktionen betrachtet, da, wie schon oben erwähnt wurde, zwar eine relativ leicht zu zeigende $O(n^2 \cdot \log n)$ -Schranke

existiert, jedoch vermutet wird, dass die erwartete Laufzeit des dynamischen (1+1)-EA auf einer linearen Funktion sogar durch $O(n \cdot \log^2 n)$ beschränkt ist. Dass die erwartete Laufzeit des dynamischen (1+1)-EA sowohl auf der Funktion ONEMAX , die als Funktionswert jeweils die Anzahl der Einsen im aktuellen Individuum liefert, als auch auf der Funktion BINVALUE , die das Individuum als binär kodierte Zahl interpretiert, $\Theta(n \log^2 n)$ beträgt, wurde bereits von Jansen und Wegener in [JW00a] gezeigt. Diese zwei lineare Funktionen sind in gewissem Sinn „Extreme“: Bei ONEMAX sind die einzelnen Bits „extrem gleich“ gewichtet und bei BINVALUE „extrem ungleich“ gewichtet. In beiden Fällen erleichtert dies die Analyse der Laufzeit, wengleich die Analyse von BINVALUE sehr viel schwieriger bzw. „Trick-reicher“ ist als die von ONEMAX .

Um das Verhalten des dynamischen (1+1)-EA auf weiteren linearen Funktionen zu untersuchen, wird die nun folgende Funktionenfamilie betrachtet, die „weniger extreme“ lineare Funktionen enthält:

Definition 2.2.1. [BLIC_k] Für $x \in \{0, 1\}^n$ ist die Funktion BLIC_k („Block-ly Linear Increasing Coefficients“) folgendermaßen definiert, wenn $k|n$ gilt: Die n Bits werden in k gleich lange Blöcke der Länge $l = n/k$ geteilt und die Bits des i -ten Blocks bekommen jeweils den Koeffizienten i zugeordnet. Es gilt also

$$\text{BLIC}_k(x) = 1(x_1 + \dots + x_l) + 2(x_{l+1} + \dots + x_{2l}) + \dots + k(x_{n-l+1} + \dots + x_n).$$

Für $k = 1$ ergibt sich somit die Funktion ONEMAX . Was ändert sich für $k > 1$ gegenüber ONEMAX ? Der entscheidende Unterschied ist, dass mit einer Verbesserung der Fitness nicht mehr zwingend eine Erhöhung der Anzahl der Einsen im Individuum einhergehen muss. Für „kleine“ k sollte dieser Effekt aber noch ziemlich gering sein, sodass eine Argumentation ähnlich der bei ONEMAX möglich sein könnte.

2.3 Laufzeit-Analyse für BLIC_k mit $k = \Theta(1)$

Als nächstes soll deshalb – sozusagen zum Aufwärmen – eine obere Schranke für die erwartete Laufzeit des dynamischen (1+1)-EA auf BLIC_2 gezeigt werden.

Lemma 2.3.1. Die erwartete Anzahl Schritte, bis der dynamische (1+1)-EA die Funktion BLIC_2 optimiert hat, ist durch $O(n \log^2 n)$ beschränkt.

Beweis: Per Definition von BLIC_2 gilt $2|n$. Der Funktionswert des optimalen Individuums 1^n ist $n/2 + 2 \cdot n/2 = 3n/2$, sodass nur Fitness-Werte aus $\{0, 1, 2, \dots, 3n/2\}$ angenommen werden können.

Da ein Bit höchstes den Wert 2 hat, besitzt ein Individuum $x \in \{0, 1\}^n$ mit $\text{BLIC}_2(x) = f$, das folglich ein Fitness-Defizit von $3n/2 - f$ hat, mindestens $\lceil (3n/2 - f)/2 \rceil$ auf Null gesetzte

Bits. Bezeichne E das Ereignis, dass in einer Phase des dynamischen (1+1)-EA x zu einem x' mit $\text{BLIC}_2(x') > f$ mutiert, diese also erfolgreich ist. Dann gilt

$$\text{Prob}(E) \geq \lceil (3n/2 - f)/2 \rceil \cdot p \cdot (1 - p)^{n-1}$$

für jede vorkommende Mutationswahrscheinlichkeit $p \in \{1/n, 2/n, 4/n, \dots\}$ mit $p < 1/2$. Für $p = 1/n$ gilt demnach:

$$\text{Prob}(E) \geq \lceil (3n/2 - f)/2 \rceil \cdot \frac{1}{n} \left(1 - \frac{1}{n}\right)^{n-1} \geq \lceil (3n/2 - f)/2 \rceil \cdot n^{-1} \cdot e^{-1}$$

Die erwartete Anzahl von Phasen, bis E eintritt, ist also sicherlich durch $e \cdot n / \lceil (3n/2 - f)/2 \rceil$ nach oben beschränkt. Durch die pessimistische Annahme, dass mit dem „schlechtesten“ Individuum, dem mit Fitness 0, gestartet wird, lässt sich die erwartete Gesamtanzahl von Phasen, die zur Optimierung von BLIC_2 nötig sind, durch

$$\sum_{i=0}^{3n/2-1} e \cdot n / \lceil (3n/2 - i)/2 \rceil = e \cdot n \cdot \sum_{i=1}^{3n/2} \frac{1}{\lceil i/2 \rceil} \leq e \cdot n \cdot 2 \cdot \sum_{i=1}^{\lceil 3n/4 \rceil} \frac{1}{i} = O(n \log n)$$

nach oben abschätzen. Da eine Phase aus $\Theta(\log n)$ Schritten besteht, folgt die Behauptung. \square

Der obige Beweis lässt sich für beliebiges, fest gewähltes k anpassen:

Lemma 2.3.2. *Die erwartete Laufzeit des dynamischen (1+1)-EA auf der Funktion BLIC_k ist durch $O(k \cdot n \cdot \log^2 n)$ beschränkt.*

Beweis: Der optimale Funktionswert für BLIC_k ist, da $k|n$ gilt, gleich

$$\sum_{i=1}^k i \cdot n/k = n/k \cdot (k(k+1)/2) = n(k+1)/2.$$

Die Anzahl der Nullen in einem Individuum x mit $\text{BLIC}_k(x) = f$ beträgt also mindestens $\lceil (n(k+1)/2 - f)/k \rceil$. Die Wahrscheinlichkeit, dass der erste Schritt einer Phase, in dem die Mutationswahrscheinlichkeit $p = 1/n$ verwendet wird, erfolgreich ist, lässt sich somit durch

$$\lceil (n(k+1)/2 - f)/k \rceil \cdot p \cdot (1 - p)^{n-1} \geq \lceil (n(k+1)/2 - f)/k \rceil \cdot n^{-1} \cdot e^{-1}$$

nach unten abschätzen. Analog zum Beweis des vorigen Lemmas lässt sich die erwartete Ge-

samtanzahl von Phasen durch

$$\begin{aligned}
 e \cdot n \cdot \sum_{i=0}^{n(k+1)/2-1} \left\lceil \frac{n(k+1)/2 - i}{k} \right\rceil^{-1} &= e \cdot n \cdot \sum_{i=1}^{n(k+1)/2} 1/\lceil i/k \rceil \\
 &\leq e \cdot n \cdot k \cdot \sum_{i=1}^{\lceil n(k+1)/2k \rceil} 1/i \\
 &\leq e \cdot n \cdot k \cdot (\ln n + \Theta(1))
 \end{aligned}$$

nach oben abschätzen, woraus die Behauptung folgt. \square

Für einige „einfache“ lineare Funktionen konnte somit die vermutete $O(n \cdot \log^2 n)$ -Laufzeit-Schranke gezeigt werden:

Korollar 2.3.3. *Für jedes fest gewählte $k \in \mathbb{N}$, d. h. für $k = \Theta(1)$, hat der dynamische (1+1)-EA auf der Funktion BLIC_k eine erwartete Laufzeit von $O(n \cdot \log^2 n)$.*

2.4 Laufzeit-Schranken für weitere lineare Funktionen

Da die spezielle Struktur von BLIC_k nicht in die vorige Beweisführung eingeht, ist diese auch für weitere lineare Funktionen anwendbar. Mit ihr lässt sich zeigen, dass folgender Sachverhalt gilt:

Lemma 2.4.1. *Für eine lineare Funktion $F(x) = \sum_{i=1}^n a_i \cdot x_i$, für die $0 < a_1 \leq \dots \leq a_n$ gilt, ist die erwartete Laufzeit des dynamischen (1+1)-EA durch $O((a_n/a_1) \cdot n \log^2 n)$ beschränkt.*

Beweis: Es wird die Funktion $F^*(x) = \sum_{i=1}^n a_i^* \cdot x_i$ mit $a_i^* := a_i/a_1$ betrachtet. Für diese Funktion gilt offensichtlich

$$(\forall x, y \in \{0, 1\}^n) F(x) \leq F(y) \Leftrightarrow F^*(x) \leq F^*(y),$$

d. h. $F \equiv_{EA} F^*$, und $1 = a_1^* \leq \dots \leq a_n^*$. Da die Selektion im (1+1)-EA ausschließlich auf dem Vergleich der Fitness-Werte einzelner Individuen beruht, wird für beide Funktionen die gleiche Laufzeit erwartet. Es gilt $F_{max}^* := \max\{F^*(x) \mid x \in \{0, 1\}^n\} \leq a_n^* \cdot n$. Für die Anzahl der Nullen in einem Individuum $x \in \{0, 1\}^n$ gilt $|x|_0 \geq \lceil (F_{max}^* - F^*(x))/a_n^* \rceil$. Eine 1-Bit-Mutation, die genau eine Null flippt, erhöht die Fitness mindestens um $a_1^* = 1$. Eine solche Mutation tritt mit der Wahrscheinlichkeit $\binom{|x|_0}{1} \cdot p \cdot (1-p)^{n-1}$ auf. Für die Mutationswahrscheinlichkeit $p = 1/n$, die im ersten Schritt einer Phase verwendet wird, ist diese Wahrscheinlichkeit durch $|x|_0/(e \cdot n)$ nach unten beschränkt. Die erwartete Anzahl von Phasen, bis eine solche Mutation auftritt, die

die Fitness um mindestens eins erhöht, ist somit durch $(e \cdot n)/|x|_0$ nach oben beschränkt. Die erwartete Anzahl von Phasen, bis der dynamische (1+1)-EA das optimale Individuum erzeugt hat, lässt sich also nach oben durch

$$\begin{aligned} \sum_{i=0}^{F_{max}^* - 1} \frac{e \cdot n}{\lceil (F_{max} - i)/a_n^* \rceil} &= e \cdot n \cdot \sum_{i=1}^{F_{max}} 1/\lceil i/a_n^* \rceil \\ &\leq e \cdot n \cdot a_n^* \cdot \sum_{i=1}^{\lceil F_{max}/a_n^* \rceil} 1/i \\ &= e \cdot n \cdot a_n^* \cdot H(\lceil F_{max}/a_n^* \rceil) \\ &\leq e \cdot n \cdot (a_n/a_1) \cdot H(n) \end{aligned}$$

abschätzen. Da $H(n) = \sum_{i=1}^n 1/i = \ln(n) + \Theta(1)$ gilt, folgt die Behauptung. \square

Da eine $O(n^2 \log n)$ -Laufzeit-Schranke für lineare Funktionen existiert (vgl. [JW00a]), stellt dieses Ergebnis nur für einen kleinen Teil der Klasse linearer Funktionen eine Verbesserung der bisherigen Laufzeit-Schranken dar – nämlich für $a_n/a_1 = o(n/\log n)$. Für Funktionenfolgen folgt aus diesem Lemma

Korollar 2.4.2. *Für eine Folge $F_n(x) = \sum_{i=1}^n a_i(n) \cdot x_i$ von linearen Funktionen, bei der $a_i(n) > 0$ für $1 \leq i \leq n$ gilt, ist die erwartete Laufzeit des dynamischen (1+1)-EA für $k(n) = \max\{a_i(n)/a_j(n) \mid i, j \in \{1, \dots, n\}\}$ durch $O(k(n) \cdot n \cdot \log^2 n)$ beschränkt. Ist $k(n)$ für jedes n durch eine Konstante beschränkt, so ist die erwartete Laufzeit folglich durch $O(n \cdot \log^2 n)$ beschränkt.*

2.5 Erkenntnisse für BLIC_n

Nun aber zurück zu der Funktion BLIC_k . Für $k = \omega(1)$ sind die zuvor im Text gezeigten Schranken sicherlich nicht optimal. Um Ansatzpunkte für eine bessere Schranke zu finden, wird hier das bisherigen Vorgehen bei der Beweisführung nochmal zusammengefasst:

1. Die Erfolgswahrscheinlichkeit einer Phase wird durch die Wahrscheinlichkeit, dass der Schritt mit Mutationswahrscheinlichkeit $1/n$ erfolgreich ist, nach unten abgeschätzt.
2. Die Erfolgswahrscheinlichkeit eines Schritts, in dem die Mutationswahrscheinlichkeit $1/n$ verwendet wird, wird nach unten abgeschätzt mit Hilfe einer pessimistischen Schätzung der Anzahl der Nullen im jeweils aktuellen Individuum.
3. Die Abschätzung der Anzahl der Nullen hängt ausschließlich von der aktuellen Fitness und k bzw. dem größten vorkommenden Koeffizienten ab.

4. Bei jedem Erfolg wird von einer Erhöhung der Fitness um lediglich 1 bzw. um den kleinsten Wert eines Koeffizienten ausgegangen, obwohl ein von 0 nach 1 flippendes, „durchschnittliches“ Bit bei BLIC_k eine Fitness-Zunahme von $(k + 1)/2$ verursacht.

Was kann an dieser Vorgehensweise verbessert werden, z. B. wenn die Blöcke konstante Länge $l = n/k \in \mathbb{N}$ haben, also $k = \Theta(n)$ gilt? Um diese Frage zu beantworten, wird im Folgenden die Funktion $\text{BLIC}_n(x) = \sum_{i=1}^n i \cdot x_i$ genauer betrachtet.

Sei $k \in \{1, \dots, n\}$. Das Individuum $0^{n-k}1^k$, bei dem genau die k geringstwertigen Bits gesetzt sind, besitzt die Fitness $k(k + 1)/2$. Eine Fitness, die kleiner als $k(k + 1)/2$ ist, impliziert somit weniger als k Einsen bzw. mehr als $n - k$ Nullen im Individuum. Für Fitness-Werte aus² $[(k - 1)k/2; k(k + 1)/2)$ sind also $n - k + 1$ Nullen sichergestellt. Wenn $n - k + 1$ auf Null gesetzte Bits existieren, so haben für $1 \leq i \leq n - k + 1$ mindestens i dieser Bits einen Wert von mindestens $n - k + 2 - i$ bzw. mindestens $n - k + 2 - i$ Bits einen Wert von mindestens i . Sei deshalb für $k \in \{1, \dots, n\}$ die Menge von Individuen

$$A_k := \{x \in \{0, 1\}^n \mid (k - 1)k/2 \leq \text{BLIC}_n(x) < k(k + 1)/2\}$$

definiert. Einem Individuum $x \in A_k$ reicht ein Fitness-Zuwachs von k , um A_k „zu verlassen“. Genauso reichen aber auch $c \in \mathbb{N}$ Verbesserungen um $\lceil k/c \rceil$. Nach der obigen Argumentation besitzt x mindestens $n + 2 - (k + \lceil k/c \rceil)$ auf Null gesetzte Bits mit einem jeweiligen Wert von mindestens $\lceil k/c \rceil$. Aber Vorsicht: Aus $n + 2 - (k + \lceil k/c \rceil) \geq 1$ folgt, dass diese Argumentation nur für A_k mit $k \leq \frac{c}{c+1} \cdot n$ gilt.

Für $x \in A_k$ mit $k \leq \frac{c}{c+1} \cdot n$ sind mindestens $n + 2 - (k + \lceil k/c \rceil)$ 1-Bit-Mutationen möglich, von denen jede die Fitness um mindestens $\lceil k/c \rceil$ erhöht. Damit aus x ein Mutant aus $A_{k+1} \cup \dots \cup A_n$ entsteht, reichen also c derartige Mutationen aus. Die Wahrscheinlichkeit für eine solche 1-Bit-Mutation innerhalb einer Phase des dynamischen (1+1)-EA lässt sich für den Schritt, in dem die Mutationswahrscheinlichkeit $p = 1/n$ verwendet wird, durch $1/n \cdot 1/e \cdot (n + 2 - (k + \lceil k/c \rceil))$ nach unten abschätzen. Die erwartete Laufzeit für die benötigten c Verbesserungen lässt sich somit durch $nc/(n + 2 - (k + \lceil k/c \rceil))$ nach oben abschätzen. Für $(c + 1) \mid n$ gilt:

$$\begin{aligned} \sum_{k=1}^{nc/(c+1)} n \cdot e \cdot c / (n + 2 - (k + \lceil k/c \rceil)) &\leq n \cdot e \cdot c \cdot \sum_{k=1}^n 1/k \\ &= n \cdot e \cdot c \cdot (\ln n + \Theta(1)) \\ &= O(n \cdot \log n) \end{aligned}$$

²bzw. aus $\{(k - 1)k/2, \dots, k(k + 1)/2 - 1\} \subset \mathbb{N}_0$

Für jedes feste $\lambda < 1$ kann c passend festgesetzt werden, sodass nach $O(n \cdot \log n)$ Phasen eine Fitness von $\lambda \cdot F_{max}$ erreicht ist. Die Vermutung liegt nahe, dass auch zur „vollständigen“ Optimierung von BLIC_n $O(n \cdot \log^2 n)$ Schritte ausreichen. Warum kann dies mit obiger Argumentation nicht gezeigt werden?

Um diese Frage zu beantworten, wird im Folgenden der Fitness-Wert $F_{max} - n$, d. h. ein Fitness-Defizit von n , exemplarisch betrachtet. Da bei der Funktion BLIC_n ein Bit mit dem Wert n existiert, kann für diese Fitness nur die Existenz von einer Null sicher angenommen werden. Bei der obigen Vorgehensweise wird aus einer sichergestellten Existenz von i Nullen gefolgert, dass (mindestens) eine dieser Nullen im betrachteten Individuum mindestens den Wert i besitzt. Bei dem hier betrachteten Fitness-Defizit von n würde also für die eine „sichergestellte“ Null nur ein Wert von (mindestens) 1 angenommen. Für die Abschätzung des größten gesicherten Wertes eines auf Null gesetzten Bits wird die Fitness bzw. das Fitness-Defizit also nicht (noch einmal) berücksichtigt, was zu einer schlechten Abschätzung führt, wie die folgenden Überlegungen zeigen.

Schon weiter oben wurde erwähnt, dass bei der Funktion BLIC_n die i geringstwertigen Bits zusammen einen Wert von $i(i+1)/2$ besitzen. Da $i \geq 0$ gelten muss, besitzt die Gleichung $i(i+1)/2 = n$ nur die Lösung $i = \sqrt{2n+1/4} - 1/2$. Ein Individuum mit der Fitness n besitzt also – egal wie diese Fitness dargestellt wird – (mindestens) ein auf Eins gesetztes Bit, das mindestens den Wert $\lceil \sqrt{2n+1/4} - 1/2 \rceil$ hat. Auf komplementäre Weise besitzt demnach ein Individuum mit einem Fitness-Defizit von n (mindestens) ein auf Null gesetztes Bit mit einem Wert von ebenfalls mindestens $\lceil \sqrt{2n+1/4} - 1/2 \rceil$.

Bei einem Fitness-Defizit von n kann bei der Funktion BLIC_n für das höchstwertige 0-Bit also ein Wert von ungefähr $\sqrt{2n}$ – statt des bisherigen Wertes von 1 – sicher angenommen werden, was gegenüber der bisherigen Abschätzung eine deutliche Verbesserung darstellt. Um ein Individuum mit dem Fitness-Defizit n zu optimieren, mussten bei der bisherigen Vorgehensweise n 1-Bit-Mutationen mit einer jeweiligen Wahrscheinlichkeit des Auftretens von $1/(ne)$ veranschlagt werden. Dies würde wiederum zu einem Beitrag von $e \cdot n^2$ in der Abschätzung der Anzahl der Phasen bis zur (endgültigen) Optimierung führen. Aus diesem Grund konnte mit der oben verwendeten Argumentation die $O(n \cdot \log n)$ -Schranke für die erwartete Anzahl von Phasen, die zur Optimierung von BLIC_n nötig sind, nicht gezeigt werden. Hilft nun die bessere Abschätzung des größten sicheren Wertes eines auf Null gesetzten Bits? Leider nur in Maßen.

Bisher wurde bei einem Fitness-Defizit von n angenommen, dass eine erfolgreiche 1-Bit-Mutation dieses um eins auf $n-1$ reduziert. Mit der besseren Abschätzung reduziert eine solche Mutation das Fitness-Defizit jetzt auf ungefähr $n - \sqrt{2n}$. Die bisher zur Optimierung eines Individuums mit dem Fitness-Defizit von n für nötig gehaltene Anzahl erfolgreicher 1-Bit-Mutationen reduzieren sich also von n auf eine Anzahl in der Größenordnung $\Theta(\sqrt{n})$. Dies würde zwar zu

einer besseren Schranke von $O(n \cdot \sqrt{n})$ – statt bisher $O(n^2)$ – für die erwartete Anzahl von Phasen, die der dynamische (1+1)-EA für die Optimierung von BLIC_n benötigt, führen, nicht aber zur vermuteten und erhofften $O(n \cdot \log n)$ -Schranke. Der einfache Grund: Wird das betrachtete Fitness-Defizit von n tatsächlich nur von einer einzigen Null verursacht, so ist offensichtlich auch nur noch eine erfolgreiche 1-Bit-Mutation zur Optimierung nötig. Wie gezeigt wurde, führt auch das bessere Schätzen des gesicherten Höchstwertes eines 0-Bits in diesem Fall immer noch zu $\Theta(\sqrt{n})$ „nötigen“ Erfolgen. Die bisher zur Beweisführung benutzte Art der „sicheren“ Abschätzung ist also ausgereizt; es müssen andere, „stochastisch präzisere“ Wege gesucht werden.

Kapitel 3

Strukturelle Erkenntnisse für lineare Funktionen

3.1 Warum große Mutationswahrscheinlichkeiten „eher ungünstig“ sind

Betrachtet wird eine zu maximierende lineare Funktion $F : \{0, 1\}^n \rightarrow \mathbb{R}$ mit ausschließlich positiven Koeffizienten¹, wobei im Folgenden $F_{max} := \max\{F(x) \mid x \in \{0, 1\}^n\}$ gelte.

Lemma 3.1.1. *Das zufällige Start-Individuum hat eine erwartete Fitness von $F_{max}/2$.*

Beweis: Die n Bits des Individuums haben zusammen einen Wert von F_{max} . Da bei linearen Funktionen der Wert jedes einzelnen Bits von den übrigen Bits unabhängig ist und beim Erzeugen des Start-Individuums mit Wahrscheinlichkeit $1/2$ „aktiviert“ wird, ergibt sich die „erwartete Start-Fitness“. \square

In einem Schritt mit der Mutationswahrscheinlichkeit p wird für ein Individuum x von den flippenden Einsen eine Verminderung der Fitness um $p \cdot F(x)$ erwartet und von den flippenden Nullen ein Fitness-Zuwachs von $p \cdot (F_{max} - F(x))$. Zusammen wird also eine Veränderung der Fitness um $p \cdot (F_{max} - 2 \cdot F(x))$ erwartet. Ist einmal ein Fitness-Niveau größer als $F_{max}/2$, der erwarteten Fitness des Start-Individuums, erreicht, so werden für das jeweils aktuelle Individuum x in den folgenden Schritten ausschließlich Verschlechterungen erwartet. Für $F(x) \geq (1 + \varepsilon) \cdot F_{max}/2$ mit $\varepsilon > 0$ haben diese Verschlechterungen die Größenordnung $\Theta(F(x) \cdot p)$. Für ein festes Fitness-Niveau wächst die erwartete (absolute) Fitness-Wert-Abnahme linear in p . Beim dynamischen (1+1)-EA wird bei der Verdoppelung der Mutationswahrscheinlichkeit also auch der erwartete

¹folglich ist \mathbb{R}_0^+ der Bildbereich von F

(absolute) Fitness-Verlust verdoppelt – wenn eine lineare Funktion mit ausschließlich positiven Koeffizienten maximiert wird.

Für die erwartete relative Veränderung der Fitness ergibt sich ein erwarteter (relativer) Anteil von $p \cdot (F_{max}/F(x) - 2)$. Für $F(x) \rightarrow F_{max}$ strebt der Anteil folglich gegen $-p$. Für $F(x) \geq (1 + \varepsilon) \cdot F_{max}/2$ mit $\varepsilon > 0$ hat der erwartete relative Fitness-Verlust die Größenordnung $\Theta(p)$.

Mit wachsendem p nimmt die „Konzentration um den Erwartungswert“ zu, d. h. die Anzahl tatsächlich flippender Nullen und die Anzahl tatsächlich flippender Einsen werden „wahrscheinlich“ nahe der jeweils erwarteten Anzahl liegen. Also nimmt mit wachsendem p nicht nur die erwartete Verschlechterung der Fitness zu, sondern auch die Wahrscheinlichkeit, dass eine solche von einer Mutation auch wirklich realisiert wird.

Ist eine Phase erfolgreich, so besitzt der Schritt mit der (kleinsten) Mutationswahrscheinlichkeit $1/n$ demnach unter allen Schritten der Phase die größte Wahrscheinlichkeit, der Grund dafür zu sein. Natürlich würde eine noch kleinere Mutationswahrscheinlichkeit auch den erwarteten Verlust noch weiter verkleinern, jedoch würde dann von einer Mutation nicht einmal ein flippendes Bit erwartet. Keine Veränderung im Individuum verhindert zwar einen Verlust, jedoch auch eine (erwünschte) Verbesserung. In der Erwartung sollte also zumindest ein Bit flippen, was durch die Wahl von $1/n$ als kleinstmögliche Mutationswahrscheinlichkeit gesichert ist.

3.2 Konkretisierung für $BLIC_n$

Für die Funktion $BLIC_n$ lässt sich dieser Sachverhalt konkretisieren. Wenn z. B. für $2|n$ ein Individuum $x = x_n \dots x_1 \in \{0, 1\}^n$ mindestens $0,9n$ Einsen enthält, so befinden sich davon mindestens $0,4n$ 1-Bits in der linken Hälfte $x_n \dots x_{(n/2)+1}$, haben also jeweils einen Wert größer als $n/2$. Da höchstens $0,1n$ Nullen vorhanden sind, deren jeweiliger Wert höchstens n ist, beträgt der von den flippenden Einsen erwartete Fitness-Verlust mehr als das Doppelte des von den flippenden Nullen erwarteten Fitness-Gewinns. Insgesamt wird also eine Verschlechterung der Fitness erwartet, die betragsmäßig sogar größer als der erwartete Wert der flippenden Nullen ist.

Da die Werte der hier betrachteten Bits durch das Betrachten der linken Hälfte des Individuums um weniger als den Faktor 2 variieren, ist das Akzeptieren einer Mutation nur dann möglich, wenn die Anzahl der flippenden Einsen und/oder die Anzahl der flippenden Nullen in die jeweils „günstige“ Richtung von der Erwartung abweichen. Im Gegensatz zu der erwarteten Fitness-Wert-Änderung, kann auf diese die Tschernoff-Schranke angewendet werden.

Die Wahrscheinlichkeit dafür, dass eine Mutation in der linken Hälfte (mindestens) doppelt so viele Einsen flippt wie insgesamt Nullen, ist eine untere Schranke für die Wahrscheinlichkeit, dass sie verworfen wird. Dieser Sachverhalt wird im Beweis des folgenden Satzes benutzt.

Lemma 3.2.1. *Für die Funktion $BLIC_n$ existiert ein $\varepsilon > 0$, sodass für Individuen, die minde-*

stens $0,9n$ Einsen enthalten, die Wahrscheinlichkeit, dass in einem Mutation-Selektion-Zyklus der erzeugte Mutant akzeptiert wird, durch $(1 - \varepsilon)^{np}$ nach oben beschränkt ist. Dabei bezeichnet p die in diesem Mutation-Selektion-Zyklus verwendete Mutationswahrscheinlichkeit.

Beweis: Bei einem Individuum mit $0,9n$ Einsen gibt es mindestens $0,4n$ 1-Bits mit einem jeweiligen Wert größer als $n/2$. Sei B definiert als die Menge genau der Positionen, an denen solche Einsen stehen. Insgesamt gibt es höchstens $0,1n$ Nullen. Sei A definiert als die Menge genau der Positionen, an denen sich Nullen befinden. Es gilt also $\#A \leq \#B/4$.

Angenommen in einem Schritt mit der Mutationswahrscheinlichkeit p „landen“ i Mutationspositionen in $A \cup B$. Dann wird von diesen in B ein Anteil von mindestens $4/5$ erwartet. Damit in diesem Schritt überhaupt die Chance besteht, dass der erzeugte Mutant akzeptiert wird, muss dieser Anteil aber kleiner als $2/3$ sein. Es darf also nur weniger als $5/6 = (2/3)/(4/5)$ des Erwartungswertes erreicht werden. Eine Mutationsposition, die in $A \cup B$ „landet“, liegt mit einer Wahrscheinlichkeit von mindestens $4/5$ in B (und damit nicht in A). Die Wahrscheinlichkeit dafür, dass vom Erwartungswert von mindestens $(4/5)i$ Mutationspositionen in B anteilig mehr als $1/6$ nach unten abgewichen wird, ist nach Tschernoff exponentiell klein (vgl. Anhang A resp. Abschnitt 4.1 in [MR95]):

$$\text{Prob}(\text{relative Abweichung} > |-1/6|) < \left(\frac{e^{-\frac{1}{6}}}{\left(1 - \frac{1}{6}\right)^{\left(1 - \frac{1}{6}\right)}} \right)^{\frac{4}{5}i} < 0,9854^{(4/5)i}$$

Da die für B erwarteten $(4/5)i$ Positionen insgesamt mindestens $0,4pn$ Positionen entsprechen, gilt schließlich:

$$\text{Prob}(\text{Mutation wird akzeptiert}) < 0,9854^{0,4pn} < 0,995^{np}$$

Mit $\varepsilon = 1 - 0,995 = 0,005$ ist der behauptete Sachverhalt damit bewiesen. \square

Für Schritte mit einer Mutationswahrscheinlichkeit von $p = \omega(1/n)$, d. h. $np = \omega(1)$, strebt die Erfolgswahrscheinlichkeit mit $n \rightarrow \infty$ folglich gegen Null. Eine Verdoppelung der Mutationswahrscheinlichkeit beim dynamischen (1+1)-EA hat im betrachteten Fall ein Quadrieren der Erfolgswahrscheinlichkeit zur Folge.

Dass der dynamische (1+1)-EA bei der Maximierung der Funktion BLIC_n im Erwartungsfall $O(n \cdot \log^2 n)$ Schritte benötigt, um ein Individuum mit 90% Einsen zu erzeugen, wurde in Abschnitt 2.5 (auf Seite 16) gezeigt. In diesem Abschnitt wurde gezeigt, dass „anschließend“ Schritte, in denen eine große Mutationswahrscheinlichkeit $p = \omega(1/n)$ verwendet wird, keinen Schaden im Sinne eines Einsen-Verlusts anrichten können – zumindest asymptotisch für $n \rightarrow \infty$. Dies bestärkt, neben der Argumentation in Abschnitt 2.5 (auf Seite 16), die Vermutung, dass dem

dynamischen (1+1)-EA im Erwartungsfall $O(n \cdot \log n)$ Phasen reichen sollten, um BLIC_n auch vollständig zu optimieren.

3.3 Intuitive Überlegungen zum dynamischen (1+1)-EA

Im Start-Individuum werden $n/2$ Einsen erwartet, da jedes Bit in diesem mit Wahrscheinlichkeit $1/2$ auf Eins resp. Null gesetzt wird. Bei ungeradem n kann ein Individuum nicht gleich viele Einsen wie Nullen enthalten, sodass in diesem Fall schon das Start-Individuum (aus Symmetriegründen) mit Wahrscheinlichkeit $1/2$ mehr Einsen als Nullen enthält. Sei $d = |x|_1 - |x|_0$ die Differenz zwischen der Anzahl der Einsen und der Anzahl der Nullen in einem Individuum $x \in \{0, 1\}^n$. Von einer Mutation mit der Mutationswahrscheinlichkeit p wird dann erwartet, dass diese d um $-2dp$ ändert. Jede Mutation neigt also dazu, die Anzahl der Einsen und die der Nullen anzugleichen bzw. beide dem Wert $n/2$ anzunähern. Diese Tendenz ist um so stärker, je größer die verwendete Mutationswahrscheinlichkeit ist. Beispielsweise wird von einer Mutation eines Individuums mit $0,9n$ Einsen, d. h. einer Differenz von $d = 0,8n$, eine Verminderung dieser Differenz um $1,6np$ erwartet. Dieser Sachverhalt gilt unabhängig von der zu optimierenden Funktion. Im Folgenden wird nur noch über lineare Funktionen mit ausschließlich positiven Koeffizienten argumentiert.

Für lineare Funktionen, bei denen jedes Bit einen „unabhängigen“ Wert in der Größe des ihm zugeordneten Koeffizienten hat, drängt sich folgende Intuition auf. Werden verschiedene Individuen desselben Fitness-Niveaus betrachtet, so sind zwei Extreme möglich, wenn nicht alle Bits gleich gewichtet sind: Die Darstellung mit maximal vielen Einsen, die mit dem kleinstmöglichen Durchschnittswert der Einsen – d. h. der auf Eins gesetzten Bits – einhergeht, und andererseits die mit minimal vielen Einsen, die zum größtmöglichen Durchschnittswert für die Einsen führt. Besitzen dabei alle Individuen eines betrachteten Fitness-Niveaus jeweils mehr als $n/2$ Einsen, so erscheint eine Darstellung der Fitness mit relativ vielen, eher niederwertigen Einsen unwahrscheinlich, da Mutationen – von Individuen mit mehr als $n/2$ Einsen – eben nicht dazu neigen, „freiwillig“ die Anzahl der Einsen zu erhöhen.

Eine Erhöhung der Anzahl der Einsen durch eine Mutation bei einem bestimmten Fitness-Niveau ist um so wahrscheinlicher, je weniger Einsen zur Darstellung der Fitness benutzt werden. Im zweiten oben genannten Extremfall ist die Wahrscheinlichkeit für eine Erhöhung der Anzahl der Einsen am größten, unter Umständen muss sogar in einem erfolgreichen Schritt die Anzahl der Einsen wachsen.

Dies führt zu der Intuition, dass auf lange Sicht Einsen nicht gerade auf geringwertigen Positionen im Individuum zu finden sein werden. Eine formale Formulierung dieser Vermutung könnte wie folgt lauten:

Vermutung 3.3.1. Wenn der [dynamische] (1+1)-EA eine lineare Funktion $F(x) = \sum_{i=1}^n a_i \cdot x_i$ mit $a_n \geq \dots \geq a_1 > 0$ maximiert, so gilt im gesamten Verlauf, d. h. in jedem Schritt, für das jeweils aktuelle Individuum $x = x_n \dots x_1 \in \{0, 1\}^n$: $\text{Prob}(x_n = 1) \geq \dots \geq \text{Prob}(x_1 = 1)$.

Dass diese Vermutung über die Verteilung der Einsen im aktuellen Individuum tatsächlich bewiesen werden kann, wird im folgenden Abschnitt gezeigt.

3.4 Über die Verteilung der Einsen im Individuum

Die Aussage „ $\text{Prob}(x_h = 1) \geq \text{Prob}(x_l = 1)$ “ ist für $\{x_h, x_l\} \in \{0, 1\}$ äquivalent zu der Aussage „ $\text{Prob}(x_h x_l = 10) \geq \text{Prob}(x_h x_l = 01)$ “. Da die Ereignisse „ $x_h x_l = 10$ “, „ $x_h x_l = 01$ “ und „ $x_h x_l = 11$ “ paarweise disjunkt sind², gelten nämlich die beiden folgenden Gleichungen:

$$\begin{aligned} \text{Prob}(x_h = 1) &= \text{Prob}(x_h x_l = 10) + \text{Prob}(x_h x_l = 11) \\ \text{Prob}(x_l = 1) &= \text{Prob}(x_h x_l = 01) + \text{Prob}(x_h x_l = 11) \end{aligned}$$

Die Vermutung 3.3.1 kann also bewiesen werden, indem für das jeweils aktuelle Individuum $x \in \{0, 1\}^n$ gezeigt wird, dass für jedes Paar benachbarter Bits, d. h. für $i \in \{0, \dots, n-2\}$, gilt:

$$\sum_{\substack{X \in \{0,1\}^i, \\ Y \in \{0,1\}^{n-2-i}}} \text{Prob}(x = X10Y) \geq \sum_{\substack{X \in \{0,1\}^i, \\ Y \in \{0,1\}^{n-2-i}}} \text{Prob}(x = X01Y)$$

Äquivalent können diese $n-1$ Ungleichungen durch die Aussage

$$(\forall i \in \{0, \dots, n-2\}) \sum_{\substack{X \in \{0,1\}^i, \\ Y \in \{0,1\}^{n-2-i}}} (\text{Prob}(x = X10Y) - \text{Prob}(x = X01Y)) \geq 0$$

formuliert werden. Im Weiteren gelte die folgende, verkürzende Schreibweise:

Definition 3.4.1. $XY \in \{0, 1\}^j \iff X, Y \in \{0, 1\}^* \wedge |X| + |Y| = j$

Um die Nicht-Negativität jeder der $n-1$ Summen zu beweisen, ist es hinreichend zu zeigen, dass alle vorkommenden, aufsummierten Differenzen nicht-negativ sind, d. h.

$$(\forall XY \in \{0, 1\}^{n-2}) \text{Prob}(x = X10Y) - \text{Prob}(x = X01Y) \geq 0$$

gilt. Entsprechend der vorangegangenen Definition ist hierin „ $(\forall XY \in \{0, 1\}^{n-2})$ “ eine verkür-

²d. h. es gilt $\text{Prob}(x_h x_l = 01 \mid x_h x_l = 11) = 0$ usw.

zende, dennoch äquivalente Schreibweise für

$$(\forall i \in \{0, \dots, n-2\}) (\forall X \in \{0, 1\}^i) (\forall Y \in \{0, 1\}^{n-2-i})$$

und ist nicht mit „ $(\forall X, Y \in \{0, 1\}^{n-2})$ “ zu verwechseln³. Beide Ausdrücke erzeugen also exakt die gleichen Belegungen für X und Y . Der zuvor formulierte hinreichende Sachverhalt ist wiederum äquivalent zu der folgenden Aussage:

$$(\forall XY \in \{0, 1\}^{n-2}) \text{ Prob}(x = X10Y) \geq \text{Prob}(x = X01Y)$$

Bevor der formale Beweis dieser „hinreichenden Vermutung“ angetreten wird, wird zunächst ein Sachverhalt gezeigt, der dabei hilfreich sein wird.

Lemma 3.4.2. *In einem Mutationsschritt des [dynamischen] (1+1)-EA, in dem die Mutationswahrscheinlichkeit p verwendet wird (d. h. es gilt $1/n \leq p < 1/2$), ist für $0 \leq i < j \leq n$ die Wahrscheinlichkeit, dass eine bestimmte i -Bit-Mutation auftritt, größer als die, dass eine bestimmte j -Bit-Mutation auftritt.*

Beweis: $p^i(1-p)^{n-i} > p^j(1-p)^{n-j} \Leftrightarrow (1-p)^{j-i} > p^{j-i} \Leftrightarrow 0 < p < 1/2$ □

Um eine Aussage zu beweisen, die während des gesamten Ablaufs des [dynamischen] (1+1)-EA gültig sein soll, bietet sich eine Induktion über die Anzahl der (bis zu einem Zeitpunkt durchlaufenen) Mutation-Selektion-Zyklen an. Es wird also die Folge x^0, x^1, x^1, \dots von den in den Schritten $0, 1, 2, \dots$ selektierten Individuen betrachtet, wobei x^0 das Start-Individuum bezeichne. Als Start-Individuum wird jedes Individuum akzeptiert, da es noch keinen „Selektionskonkurrenten“ gibt. Da das Start-Individuum nach Gleichverteilung zufällig aus $\{0, 1\}^n$ gewählt wird, gilt

Proposition 3.4.3. *Für das Start-Individuum $x^0 \in \{0, 1\}^n$ beim [dynamischen] (1+1)-EA gelten die beiden folgenden Aussagen:*

$$\begin{aligned} (\forall y \in \{0, 1\}^n) \text{ Prob}(x^0 = y) &= 2^{-n} \\ (\forall y \in \{0, 1\}^n) (\forall z \in \{0, 1\}^n) \text{ Prob}(x^0 = y) &= \text{Prob}(x^0 = z) \end{aligned}$$

Dieser Sachverhalt wird später für den Induktionsanfang Verwendung finden. Wird der Ablauf des [dynamischen] (1+1)-EA als Markoff-Prozess betrachtet, so ergibt sich die Menge der möglichen Zustände als die Menge $\{0, 1\}^n$ der möglichen Individuen. Für die Übergangswahrscheinlichkeiten zwischen zwei aufeinander folgenden Zuständen/Individuen x^i und x^{i+1} gelten die folgenden Sachverhalte, wenn die Funktion $F(x)$ maximiert wird:

³was üblicherweise „ $(\forall X \in \{0, 1\}^{n-2}) (\forall Y \in \{0, 1\}^{n-2})$ “ abkürzt

- $F(x^{i+1}) < F(x^i)$

Da immer das bessere Individuum selektiert wird, ist ein solcher Übergang ausgeschlossen. In diesem Fall gilt also: $\text{Prob}(x^i \rightarrow x^{i+1}) = 0$.

- $F(x^{i+1}) \geq F(x^i) \wedge x^{i+1} \neq x^i$

Da ein Mutant gleicher oder besserer Fitness immer akzeptiert, d. h. selektiert, wird, gilt hier $\text{Prob}(x^i \rightarrow x^{i+1}) = \text{Prob}(H(x^i, x^{i+1})\text{-Bit-Mutation})$, wobei $H(\cdot, \cdot)$ die Hamming-Distanz zweier Bit-Sequenzen bezeichne.

- $x^{i+1} = x^i$

Dieser Übergang findet immer genau dann statt, wenn eine/die 0-Bit-Mutation erzeugt wird oder der aus x^i erzeugte Mutant eine geringere Fitness aufweist, sodass dieser verworfen wird bzw. x^i selektiert wird. Folglich gilt in diesem Fall:

$$\text{Prob}(x^i \rightarrow x^{i+1}) = \text{Prob}(0\text{-Bit-Mutation}) + \text{Prob}(F(\text{Mut}(x^i)) < F(x^i))$$

Dabei bezeichnet $\text{Mut}(x^i)$ den zufälligen, aus x^i erzeugten Mutanten.

Die gerade benutzte Wahrscheinlichkeit, (in einem bestimmten Schritt) eine Mutation zu erzeugen, die verworfen wird, wird nun näher betrachtet.

Proposition 3.4.4. *In einem Schritt des [dynamischen] (1+1)-EA, in dem das Individuum $y \in \{0, 1\}^n$ mutiert wird, um eine Funktion $F : \{0, 1\}^n \rightarrow \mathbb{R}$ zu maximieren, gilt für die Wahrscheinlichkeit, dass diese Mutation verworfen wird:*

$$\text{Prob}(F(\text{Mut}(y)) < F(y)) = \sum_{\substack{w \in \{0, 1\}^n \\ F(w) < F(y)}} \text{Prob}(\text{bestimmte } H(y, w)\text{-Bit-Mutation})$$

Dabei bezeichne wieder $H(\cdot, \cdot)$ die Hamming-Distanz zweier Bit-Sequenzen und $\text{Mut}(y)$ den zufälligen, in diesem Schritt aus y erzeugten Mutanten.⁴

Zusammenfassend ergibt sich für den betrachteten Markoff-Prozess, den der Ablauf des [dynamischen] (1+1)-EA darstellt, die folgende

Proposition 3.4.5. *Es wird der Ablauf des [dynamischen] (1+1)-EA auf einer zu maximierenden Funktion F betrachtet. Es bezeichnen x^0 das Start-Individuum, $H(\cdot, \cdot)$ die Hamming-Distanz zweier Bit-Sequenzen und $p(h) = p^h(1-p)^{n-h}$ die Wahrscheinlichkeit einer bestimmten h -Bit-Mutation, wenn die Mutationswahrscheinlichkeit p verwendet wird. Wenn x^i das im i -ten Schritt*

⁴Die Verteilung von $\text{Mut}(y)$ über $\{0, 1\}^n$ hängt von der verwendeten Mutationswahrscheinlichkeit ab.

aus $\{x^{i-1}, \text{Mut}(x^{i-1})\}$ selektierte Individuum bezeichnet, so gilt für $i \geq 1$ und $y \in \{0, 1\}^n$:

$$\begin{aligned} \text{Prob}(x^i = y) &= \text{Prob}(x^{i-1} = y) \cdot \sum_{\substack{w \in \{0,1\}^n, \\ F(y) > F(w)}} p(H(y, w)) \\ &+ \sum_{\substack{u \in \{0,1\}^n, \\ F(u) \leq F(y)}} \text{Prob}(x^{i-1} = u) \cdot p(H(u, y)) \end{aligned}$$

Diese Identität für die Wahrscheinlichkeit eines bestimmten Zustands zu einem bestimmten Zeitpunkt im Markoff-Prozess, den der Ablauf des [dynamischen] (1+1)-EA darstellt, wird beim Induktionsschritt verwendet werden. Zuvor soll aber die Erkenntnis über die Wahrscheinlichkeit, mit der eine Mutation auftritt, die verworfen wird, für lineare Funktionen konkretisiert werden. Werden zwei verschiedene Individuen betrachtet, so können nun die beiden entsprechenden Wahrscheinlichkeiten verglichen werden.

Lemma 3.4.6. *Eine lineare Funktion $F(x) = \sum_{i=1}^n a_i \cdot x_i$ mit $a_n \geq \dots \geq a_1 > 0$ soll für $x = x_n \dots x_1 \in \{0, 1\}^n$ maximiert werden. Der Operator $\text{Mut} : \{0, 1\}^n \rightarrow \{0, 1\}^n$ realisiere genau die Funktionalität eines Mutationsschrittes des [dynamischen] (1+1)-EA mit der Mutationswahrscheinlichkeit p . Dann gilt für $AB \in \{0, 1\}^{n-2}$:*

$$\text{Prob}(F(\text{Mut}(A10B)) < F(A10B)) \geq \text{Prob}(F(\text{Mut}(A01B)) < F(A01B))$$

Bei ein und derselben Mutationswahrscheinlichkeit ist die Wahrscheinlichkeit, dass eine Mutation des Individuums $A10B$ verworfen wird, also mindestens so groß wie die Wahrscheinlichkeit, dass eine Mutation des Individuums $A01B$ verworfen wird.

Beweis: Es muss gezeigt werden, dass

$$S_{10} := \sum_{\substack{u \in \{0,1\}^n, \\ F(u) < F(A10B)}} p(H(A10B, u)) \geq \sum_{\substack{w \in \{0,1\}^n, \\ F(w) < F(A01B)}} p(H(A01B, w)) =: S_{01}$$

gilt, wobei $H(\cdot, \cdot)$ die Hamming-Distanz zweier Bit-Sequenzen und $p(j)$ die Wahrscheinlichkeit für eine bestimmte j -Bit-Mutation in dem betrachteten Schritt bezeichnen. Seien im Folgenden $A' \in \{0, 1\}^{|A|}$ und $B' \in \{0, 1\}^{|B|}$. Die Wertigkeit der Bits nimmt nach Voraussetzung von links nach rechts monoton ab. Für die möglichen Fitness-Wert-Änderungen, die durch eine Mutation bzw. einen Zustandsübergang $Ax_i x_{i-1} B \rightarrow A'x'_i x'_{i-1} B'$ verursacht werden, gilt für die vier möglichen Belegungen des Bit-Paares $x'_i x'_{i-1}$:

$$\begin{aligned} F(A'01B') < F(\mathbf{A01B}) &\Rightarrow F(A'10B') < F(\mathbf{A10B}) \\ F(A'00B') < F(\mathbf{A01B}) &\Rightarrow F(A'00B') < F(\mathbf{A10B}) \end{aligned}$$

$$\begin{aligned}
F(A'10B') < F(\mathbf{A01B}) &\Rightarrow F(A'01B') < F(\mathbf{A10B}) \\
F(A'11B') < F(\mathbf{A01B}) &\Rightarrow F(A'11B') < F(\mathbf{A10B})
\end{aligned}$$

Der durch die Mutation von A und B verursachte Anteil der Fitness-Wert-Änderung ist jeweils für $A10B$ und $A01B$ gleich. Die erste Implikation gilt offensichtlich. Die zweite gilt, da der Anteil der Fitness-Wert-Änderung durch die Mutation des jeweiligen Bit-Paares für $A01B$ genau $-a_{i-1}$ beträgt und für $A10B$ genau $-a_i$ ($\leq -a_{i-1}$). Die dritte Implikation gilt, da der Anteil der Fitness-Wert-Änderung durch die Mutation des jeweiligen Bit-Paares für $A01B$ genau $+a_i - a_{i-1}$ beträgt und für $A10B$ genau $-a_i + a_{i-1}$ ($\leq +a_i - a_{i-1}$). Die vierte gilt schließlich, da der Anteil der Fitness-Wert-Änderung durch die Mutation des jeweiligen Bit-Paares für $A01B$ genau $+a_i$ beträgt und für $A10B$ genau $+a_{i-1}$ ($\leq +a_i$). Einer möglichen Mutation von $A01B$ wird also immer eine Mutation von $A10B$ gegenübergestellt, die eine nicht-größere Fitness-Wert-Änderung verursacht. Würde die Mutation von $A01B$ verworfen, d. h. sie verursacht eine negative Fitness-Wert-Änderung, so verursacht die gegenübergestellte Mutation von $A10B$ eine „mindestens so negative“ Fitness-Wert-Änderung, sie würde also ebenfalls verworfen.

Wird jedem w , über das in S_{01} summiert wird, genau das diesen Implikationen entsprechende u , über das in S_{10} summiert wird, durch

$$\begin{aligned}
w = A'01B' &\rightarrow A'10B' = u \\
w = A'00B' &\rightarrow A'00B' = u \\
w = A'10B' &\rightarrow A'01B' = u \\
w = A'11B' &\rightarrow A'11B' = u
\end{aligned}$$

zugeordnet, so handelt es sich dabei offensichtlich um eine umkehrbar eindeutige Abbildung. Jedem Summanden in S_{01} kann somit umkehrbar eindeutig ein gleich großer Summand in S_{10} zugeordnet werden, da die jeweils gegenübergestellten Mutationen jeweils gleich viele Einsen und gleich viele Nullen flippen. Die Summe S_{10} , die der Wahrscheinlichkeit gleicht, dass eine Mutation von $A10B$ verworfen wird, kann somit – bei gleicher Mutationswahrscheinlichkeit – nie einen kleineren Wert annehmen als die Summe S_{01} , die der Wahrscheinlichkeit gleicht, dass eine Mutation von $A01B$ verworfen wird. \square

Soweit die Vorarbeiten, nun zur angekündigten Induktion:

Theorem 3.4.7. *Wenn der [dynamische] (1+1)-EA eine lineare Funktion $F(x) = \sum_{i=1}^n a_i \cdot x_i$ mit $a_n \geq \dots \geq a_1 > 0$ maximiert, so gilt im gesamten Ablauf der Optimierung, d. h. in jedem Schritt, für das (im jeweiligen Schritt) aktuelle Individuum $x = x_n \dots x_1 \in \{0, 1\}^n$:*

$$(\forall XY \in \{0, 1\}^{n-2}) \text{ Prob}(x = X10Y) \geq \text{Prob}(x = X01Y)$$

Beweis: Durch Induktion über die Anzahl der (bis zu einem bestimmten Zeitpunkt) durchlaufenen Mutation-Selektion-Zyklen im Ablauf des [dynamischen] (1+1)-EA auf F . Für das Start-Individuum x^0 gilt:

$$(\forall XY \in \{0, 1\}^{n-2}) \text{ Prob}(x^0 = X10Y) = \text{Prob}(x^0 = X01Y)$$

Die Behauptung trifft also für x^0 zu und der Induktionsanfang ist damit gezeigt.

Die Induktionsannahme lautet jetzt, dass für $i \geq 1$ nach $i - 1$ Mutation-Selektion-Zyklen die Verteilung des Individuums x^{i-1} über $\{0, 1\}^n$ die behauptete Eigenschaft besitzt. Für den Induktionsschritt ist somit zu zeigen, dass unter dieser Voraussetzung die Behauptung auch für das Individuum x^i bzw. dessen Verteilung über $\{0, 1\}^n$ zutrifft, also die Eigenschaft durch den i -ten Mutation-Selektion-Zyklus nicht zerstört wird.

Für den Induktionsschritt seien $XY \in \{0, 1\}^{n-2}$ beliebig, aber im Folgenden fest gewählt. Im i -ten Schritt wird x^{i-1} entweder durch den erzeugten Mutanten $\text{Mut}(x^{i-1})$ ersetzt – nämlich genau dann, wenn dieser mindestens so fit wie x^{i-1} ist –, oder aber die Mutation wird verworfen und für x^i bleibt das Individuum x^{i-1} erhalten. Es gilt also:

$$\begin{aligned} \text{Prob}(x^i = X01Y) &= \text{Prob}(x^{i-1} = X01Y) \cdot \sum_{\substack{w \in \{0,1\}^n, \\ F(X01Y) > F(w)}} p(H(X01Y, w)) \\ &+ \sum_{\substack{u \in \{0,1\}^n, \\ F(u) \leq F(X01Y)}} \text{Prob}(x^{i-1} = u) \cdot p(H(u, X01Y)) \end{aligned}$$

Dabei bezeichnen wieder $p(h)$ die Wahrscheinlichkeit für eine bestimmte h -Bit-Mutation in dem betrachteten, i -ten Schritt und $H(\cdot, \cdot)$ die Hamming-Distanz zweier Individuen. Es gilt also $p(h) = p^h(1-p)^{n-h}$, wobei p die im i -ten Schritt verwendete Mutationswahrscheinlichkeit bezeichnet. Die Argumentation in diesem Induktionsschritt ist für jedes $p \in [1/n; 1/2)$ gültig. Der durch die Induktion gezeigte Sachverhalt gilt damit für jeden (1+1)-EA, der ausschließlich Mutationswahrscheinlichkeiten aus $[1/n; 1/2)$ verwendet, also (insbesondere) für den statischen (1+1)-EA – unabhängig von der fest gewählten Mutationswahrscheinlichkeit – und den dynamischen (1+1)-EA.⁵ Die gewichtete Summe von Mutationswahrscheinlichkeiten in der obigen Darstellung von $\text{Prob}(x^i = X01Y)$ deckt dabei genau die Mutationen des Individuums $x^{i-1} = X01Y$ ab, die verworfen werden und so zu $x^i = X01Y$ führen. Die Summe der gewichteten Mutationswahrscheinlichkeiten deckt andererseits genau die Mutation ab, bei denen ein Akzeptieren des Mutanten $X01Y$ zu $x^i = X01Y$ führt. Für das Individuum $X10Y$ bzw. die

⁵Der gezeigte Sachverhalt würde aber beispielsweise auch für einen (1+1)-EA gelten, der von Schritt zu Schritt zwischen den Mutationswahrscheinlichkeiten $1/n$ und $1/\sqrt{n}$ hin und her wechselt – wenn $\sqrt{n} > 2$ gilt.

Wahrscheinlichkeit, mit der es im i -ten Schritt erzeugt wird, ergibt sich analog:

$$\begin{aligned} \text{Prob}(x^i = X10Y) &= \text{Prob}(x^{i-1} = X10Y) \cdot \sum_{\substack{w \in \{0,1\}^n, \\ F(X10Y) > F(w)}} p(H(X10Y, w)) \\ &+ \sum_{\substack{u \in \{0,1\}^n, \\ F(u) \leq F(X10Y)}} \text{Prob}(x^{i-1} = u) \cdot p(H(u, X10Y)) \end{aligned}$$

In die für den Induktionsschritt zu zeigende Ungleichung

$$\text{Prob}(x^i = X10Y) \geq \text{Prob}(x^i = X01Y)$$

können nun diese beiden Identitäten eingesetzt werden. Nach Induktionsvoraussetzung gilt:

$$\text{Prob}(x^{i-1} = X10Y) \geq \text{Prob}(x^{i-1} = X01Y)$$

Ferner gilt nach Lemma 3.4.6 (auf Seite 28):

$$\sum_{\substack{w \in \{0,1\}^n, \\ F(X10Y) > F(w)}} p(H(X10Y, w)) \geq \sum_{\substack{w \in \{0,1\}^n, \\ F(X01Y) > F(w)}} p(H(X01Y, w))$$

Zusammen folgt also:

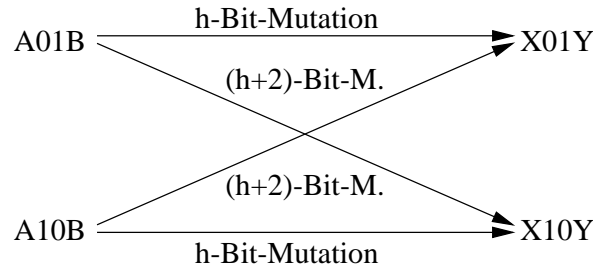
$$\begin{aligned} &\text{Prob}(x^{i-1} = X10Y) \cdot \sum_{\substack{w \in \{0,1\}^n, \\ F(X10Y) > F(w)}} p(H(X10Y, w)) \\ &\geq \text{Prob}(x^{i-1} = X01Y) \cdot \sum_{\substack{w \in \{0,1\}^n, \\ F(A01B) > F(w)}} p(H(X01Y, w)) \end{aligned}$$

Um „ $\text{Prob}(x^i = X10Y) \geq \text{Prob}(x^i = X01Y)$ “ zu zeigen, ist es folglich hinreichend zu zeigen, dass für die „restlichen“ Summanden in den Summen-Darstellungen dieser beiden verglichenen Wahrscheinlichkeiten die folgende Ungleichung gilt:

$$\begin{aligned} &\sum_{\substack{u \in \{0,1\}^n, \\ F(u) \leq F(X10Y)}} \text{Prob}(x^{i-1} = u) \cdot p(H(u, X10Y)) \quad =: S_{10} \\ &\geq \sum_{\substack{w \in \{0,1\}^n, \\ F(w) \leq F(X01Y)}} \text{Prob}(x^{i-1} = w) \cdot p(H(w, X01Y)) \quad =: S_{01} \end{aligned}$$

Um diese Ungleichung zu zeigen, werden nun die einzelnen Summanden dieser beiden Summen S_{10} und S_{01} verglichen.

Da $F(X01Y) \leq F(X10Y)$ gilt, wird über jedes Individuum, über das in S_{01} summiert wird, auch in S_{10} summiert. Wenn für $A \in \{0, 1\}^{|X|}$, $B \in \{0, 1\}^{|Y|}$ in S_{01} über das Individuum $w = A00B$ summiert wird, so verursacht $A00B = u$ einen gleich großen Summanden auch in S_{10} , da jeweils $H(AB, XY) + 1$ Bits flippen müssen. Gleiches gilt für $w = A11B = u$. Für die potenziellen Eltern-Individuen $A01B$ und $A10B$ verdeutlicht die folgende Abbildung die Zusammenhänge, wobei $h = H(AB, XY)$ gelte.



Im Fall „ $F(A01B), F(A10B) > F(X01Y)$ “ ist nichts zu zeigen. Da $F(A01B) \leq F(A10B)$ gilt, ist $F(A01B) > F(X01Y) \geq F(A10B)$ ausgeschlossen – dass in S_{01} über $w = A10B$ summiert wird, nicht aber über $w = A01B$, ist also unmöglich.

Für den „umgekehrten“ und möglichen Fall, dass $F(A10B) > F(X01Y) \geq F(A01B)$ gilt, wird in S_{01} über $w = A01B$, nicht aber über $w = A10B$ summiert. In diesem Fall gilt jedoch auch $F(A10B) \leq F(X10Y)$. Also verursacht $u = A10B$ in S_{10} einen mindestens so großen Summanden wie $w = A01B$ in S_{01} , da jeweils $h = H(AB, XY)$ Bits flippen müssen und nach Induktionsvoraussetzung $\text{Prob}(x^{i-1} = A10B) \geq \text{Prob}(x^{i-1} = A01B)$ gilt.

Gilt schließlich $F(A01B), F(A10B) \leq F(X01Y)$ ($\leq F(X10Y)$), so würden alle vier in der Grafik gezeigten Mutationen (bei einem Auftreten) akzeptiert. Da nach Lemma 3.4.2 (auf Seite 26) $p(h) - p(h+2) > 0$ gilt, gilt in diesem Fall:

$$\left. \begin{array}{l} \text{Prob}(x^{i-1} = A01B) \cdot p(h) + \\ \text{Prob}(x^{i-1} = A10B) \cdot p(h+2) \end{array} \right\} \leq \left\{ \begin{array}{l} \text{Prob}(x^{i-1} = A10B) \cdot p(h) + \\ \text{Prob}(x^{i-1} = A01B) \cdot p(h+2) \end{array} \right.$$

$$\Leftrightarrow \text{Prob}(x^{i-1} = A01B) \cdot (p(h) - p(h+2)) \leq \text{Prob}(x^{i-1} = A10B) \cdot (p(h) - p(h+2))$$

$$\Leftarrow \text{Prob}(x^{i-1} = A01B) \leq \text{Prob}(x^{i-1} = A10B)$$

Die unterste Zeile/Ungleichung gilt nach Induktionsvoraussetzung. Damit folgt in diesem Fall, dass die beiden zu $A01B$ und zu $A10B$ gehörenden Summanden in S_{10} zusammen einen mindestens so großen Wert haben, wie die zwei Summanden, die diese beiden Individuen in S_{01} verursachen.

In S_{10} gibt es also mindestens so viele Summanden mit einem mindestens so großen Gesamtwert wie in S_{01} . Damit bleibt die zu zeigende Eigenschaft während des i -ten Schritts erhalten. Dies war für den Induktionsschritt zu zeigen, da die Behauptung damit auch für x^i gilt. \square

Dieses Theorem lässt die Vermutung 3.3.1 (auf Seite 25) nun zur Gewissheit werden:

Korollar 3.4.8. *Wenn der [dynamische] (1+1)-EA eine lineare Funktion $F(x) = \sum_{i=1}^n a_i \cdot x_i$ mit $a_n \geq \dots \geq a_1 > 0$ maximiert, so gilt im gesamten Verlauf der Optimierung für das jeweils aktuelle Individuum $x = x_n \dots x_1 \in \{0, 1\}^n$: $\text{Prob}(x_n = 1) \geq \dots \geq \text{Prob}(x_1 = 1)$.*

3.5 Die Wahrscheinlichkeit bestimmter Mutationen

In der Formulierung des obigen Korollars eignet sich die errungene Struktur-Erkenntnis jedoch wenig zur weiteren Verwendung. Für die angestrebte Argumentation über den Ablauf des [dynamischen] (1+1)-EA auf linearen Funktionen ist das folgende Resultat besser geeignet.

Korollar 3.5.1. *Wenn der [dynamische] (1+1)-EA eine lineare Funktion $F(x) = \sum_{i=1}^n a_i \cdot x_i$ mit $a_n \geq \dots \geq a_1 > 0$ maximiert, so gilt im gesamten Ablauf für das jeweils aktuelle Individuum $x = x_n \dots x_1 \in \{0, 1\}^n$:*

$$(\forall FGH \in \{0, 1\}^{n-2}) \text{Prob}(x = F1G0H) \geq \text{Prob}(x = F0G1H)$$

Dieses Resultat folgt zwar nicht (unmittelbar) aus Theorem 3.4.7 (auf Seite 29), jedoch lassen sich dessen Beweis und auch Lemma 3.4.6 (auf Seite 28) und wiederum dessen Beweis entsprechend anpassen. Unter anderem wurde beispielsweise festgestellt, dass ein und derselbe Mutation-Selektion-Zyklus den Zustandsübergang „A00B → X10Y“ mindestens so wahrscheinlich realisiert wie den Übergang „A00B → X01Y“. Genauso ist leicht einzusehen, dass ein und derselbe Schritt für $C \in \{0, 1\}^{|F|}$, $D \in \{0, 1\}^{|G|}$, $E \in \{0, 1\}^{|H|}$ mindestens so wahrscheinlich den Zustandsübergang „C0D0E → F1G0H“ wie den Übergang „C0D0E → F0G1H“ realisiert usw.

Mit Hilfe dieser Erkenntnis über die Verteilung der Einsen im aktuellen Individuum kann nun eine Aussage darüber formuliert werden, wie wahrscheinlich das Auftreten von bestimmten Mutationen im Ablauf der Maximierung einer linearen Funktion durch den [dynamischen] (1+1)-EA ist. Diese könnte wiederum dabei helfen, die Wahrscheinlichkeiten für bestimmte Veränderungen der Anzahl der Einsen durch einen Mutation-Selektion-Zyklus besser als bisher abschätzen zu können. Tritt beispielsweise die 2-Bit-Mutation auf, die in einem Individuum x die beiden Bits an den Positionen i und j flippt, wobei $j > i$ gelte, so folgt nämlich aus dem vorangegangenen Korollar (da $a_j \geq a_i > 0$ vorausgesetzt ist):

$$\text{Prob}(x_j = 1 \wedge x_i = 0) \geq \text{Prob}(x_j = 0 \wedge x_i = 1)$$

Dies gilt für beliebige i, j (solange $n \geq j > i \geq 1$ gilt). Wenn also im Verlauf der Optimierung eine 2-Bit-Mutation auftritt, so ist sicher, dass der Fall „ $x_j x_i = 10$ “ mindestens so wahrscheinlich

auftritt wie der Fall „ $x_j x_i = 01$ “. Die Schreibweise „ $x_j x_i$ “ bedeutet dabei nicht, dass x_j und x_i zwangsläufig benachbarte Bits sind.

Tritt in einem Mutation-Selektion-Zyklus eine 2-Bit-Mutation auf, so wird die Anzahl der Einsen im Individuum durch diesen Schritt sicher nicht verringert, da dies nur dann möglich wäre, wenn die Mutation zwei Einsen flippt. Eine solche Mutation wird aber, da lineare Funktionen mit ausschließlich positiven Koeffizienten betrachtet werden, sicher verworfen. Flippt eine Mutation in einem Mutation-Selektion-Zyklus mehr als zwei Bits, so wird die Anzahl der Einsen durch diesen Schritt unter Umständen verringert: Flippt beispielsweise eine 3-Bit-Mutation eine Null und zwei Einsen, so kann der Fitness-Gewinn durch die flippende Null größer sein als der Fitness-Verlust durch die beiden flippenden Einsen. Die Mutation würde in diesem Fall akzeptiert und die Anzahl der Einsen um $1 - 2 = -1$ verändert. Dafür müsste der flippenden Null allerdings ein Koeffizient zugeordnet sein, der mindestens so groß ist wie die Summe der beiden Koeffizienten, die den zwei flippenden Einsen zugeordnet sind. Die flippende Null müsste also auf einer relativ höherwertigen Position im Individuum stehen als beide flippenden Einsen. Das Struktur-Resultat über die Verteilung der Einsen im aktuellen Individuum besagt nun jedoch, dass sich im Verlauf der Optimierung auf höherwertigen Positionen „eher“ Einsen als Nullen befinden, sodass eine 3-Bit-Mutation, die eine Null und zwei Einsen flippt, „eher unwahrscheinlich“ zu einer Verbesserung, d. h. Erhöhung, der Fitness führen wird. Tritt in einem Mutation-Selektion-Zyklus eine Mutation auf, die zu einer Verringerung der Anzahl der Einsen führen könnte, da sie mehr Einsen als Nullen flippt, ist also zu vermuten, dass diese „wahrscheinlich“ eine Verschlechterung der Fitness verursacht und im Selektionsschritt somit verworfen wird, sodass der Mutation-Selektion-Zyklus das Individuum nicht verändert, er also insbesondere die Anzahl der Einsen im Individuum nicht verringert.

Für die folgende Argumentation, die diese Vermutung formal konkretisieren soll, müssen zunächst einige Begriffe eingeführt und erläutert werden. Eine i -Bit-Mutation flippt genau i Bits des aktuellen Individuums $x \in \{0, 1\}^n$. Es sind also $\binom{n}{i}$ verschiedene i -Bit-Mutationen eines Individuums möglich. Bestimmt wird eine i -Bit-Mutation durch eine i -elementige Teilmenge $I \subseteq \{1, \dots, n\}$, die genau die Indizes der Bits enthält, die durch die Mutation negiert werden.

Wird ein bestimmter Schritt im Ablauf des [dynamischen] (1+1)-EA betrachtet, so wird auch ein konkretes aktuelles Individuum $x = x_n \dots x_1 \in \{0, 1\}^n$ betrachtet, das in diesem Schritt mutiert wird. Einer bestimmten i -Bit-Mutation, repräsentiert durch die entsprechende i -elementige Index-Menge $I \subseteq \{1, \dots, n\}$, kann dann die Bit-Sequenz $B = x_{j_i} \dots x_{j_1} \in \{0, 1\}^i$ mit $\{j_1, \dots, j_i\} = I$ und $j_i > \dots > j_1$ zugeordnet werden. Bei dieser Bit-Sequenz B handelt es sich also genau um die Sub-Bit-Sequenz von x , durch deren (bitweise) Negation der Mutant von x erzeugt wird. Werden z. B. in dem Individuum $1\bar{1}1\bar{0}00$ die beiden überstrichenen Bits geflippt, so liegt eine bestimmte 2-Bit-Mutation vor. In diesem Fall gilt $I = \{3, 5\}$, da das dritte und

das fünfte Bit (von rechts) geflippt werden. Da das konkrete Individuum $x = x_6 \dots x_1 = 1^3 0^3$ betrachtet wird, negiert diese bestimmte 2-Bit-Mutation die Sub-Bit-Sequenz $B = x_5 x_3 = 10$ in x , d. h. eine Sub-Bit-Sequenz der Form 10. Die gleiche (durch $I = \{3, 5\}$ bestimmte) 2-Bit-Mutation negiert in dem Individuum $0\overline{0}0\overline{0}0$ die beiden überstrichenen Nullen, also eine Sub-Bit-Sequenz der Form 00. Offensichtlich wird bei dem Individuum 0^6 durch jede 2-Bit-Mutation eine Sub-Bit-Sequenz der Form 00 negiert. Um derartige Sachverhalte kürzer ausdrücken und einfacher verwenden zu können, wird nun der Begriff „ B -Mutation“ eingeführt:

Definition 3.5.2. [B -Mutation, $B \in \{0, 1\}^*$] *Es wird eine bestimmte i -Bit-Mutation eines bestimmten Individuums x betrachtet. Sei $B \in \{0, 1\}^i$ die durch diese Mutation in x (bitweise) negierte Sub-Bit-Sequenz (vgl. oben). Die vorliegende i -Bit-Mutation des Individuums x wird dann (und nur dann) „ B -Mutation“ genannt.*

In dem obigen Beispiel, bei dem im Individuum $x = 1\overline{1}1\overline{0}0$ die beiden überstrichenen Bits geflippt werden, handelt es sich also um eine 10-Mutation. Flippt eine andere (durch $I = \{1, 6\}$) bestimmte 2-Bit-Mutation in diesem Individuum das ganz linke und das ganz rechte Bit ($\overline{1}1100\overline{0}$), so handelt es sich ebenfalls um eine 10-Mutation des Individuums $1^3 0^3$. Für das schon betrachtete Individuum 0^6 lässt sich der oben beschriebene, beispielhafte Sachverhalt nun wie folgt ausdrücken: „Jede 2-Bit-Mutation des Individuums 0^6 ist eine 00-Mutation.“

Insbesondere ermöglicht dieser neue Begriff, Aussagen über einen Mutation-Selektion-Zyklus deutlich kürzer formulieren zu können. Zum Beispiel kann die Aussage

„Wenn in einem Schritt eine 3-Bit-Mutation auftritt, die im aktuellen Individuum eine Null und zwei Einsen flippt, wobei die flippende Null relativ zu beiden flippenden Einsen links im Individuum steht, dann . . . “

nun deutlich kürzer formuliert werden:

„Wenn in einem Schritt eine 011-Mutation auftritt, dann . . . “

Weitere „Erweiterungen“ des Begriffes „ i -Bit-Mutation“, die die Argumentation vereinfachen bzw. die Formulierung von Aussagen verkürzen sollen, sind:

Definition 3.5.3. [z -Nullen- k -Einsen-Mutation] *Mit „ z -Nullen- k -Einsen-Mutation“ wird eine $(z + k)$ -Bit-Mutation eines bestimmten Individuums genau dann bezeichnet, wenn sie genau z Nullen und genau k Einsen flippt.*

Definition 3.5.4. [z -Nullen-Mutation] *Mit „ z -Nullen-Mutation“ wird eine i -Bit-Mutation eines bestimmten Individuums genau dann bezeichnet, wenn unter den $i \geq z$ flippenden Bits genau z Nullen sind.*

Definition 3.5.5. [k -Einsen-Mutation] Mit „ k -Einsen-Mutation“ wird eine i -Bit-Mutation eines bestimmten Individuums genau dann bezeichnet, wenn unter den $i \geq k$ flippenden Bits genau k Einsen sind.

Zum Beispiel ist eine 10-Mutation folglich auch eine 1-Eins-1-Null-Mutation, eine 1-Eins-Mutation und eine 1-Null-Mutation. Werden im Folgenden Individuen oder Bit-Sequenzen betrachtet, so wird weiterhin davon ausgegangen, dass die Wertigkeit der Bits von links nach rechts monoton abnimmt. Wenn für $B \in \{0, 1\}^+$ eine B -Mutation des Individuums $x \in \{0, 1\}^n$ vorliegt, so nimmt (wie in x) auch in B die Wertigkeit der Bits monoton von links nach rechts ab. Tritt beispielsweise eine 1-Null-1-Eins-Mutation auf, so handelt es sich entweder um eine 10-Mutation, oder um eine 01-Mutation. Im Fall der 10-Mutation ist der Wert der flippenden Eins mindestens so groß wie der Wert der flippenden Null. Im Fall der 01-Mutation ist der Wert der flippenden Eins hingegen höchstens so groß wie der Wert der flippenden Null. Eine 01-Mutation wird folglich immer akzeptiert, während eine 10-Mutation nur dann akzeptiert wird, wenn beide flippende Bits den gleichen Wert besitzen.

Werden zum Beispiel die möglichen 1-Null-2-Einsen-Mutationen eines bestimmten Individuums betrachtet, so kommen neben 011-Mutationen nur noch 101- und 110-Mutationen in Frage. Mutationen, die zu den beiden letzteren Kategorien gehören, werden in einem Mutation-Selektion-Zyklus mit Sicherheit verworfen (wenn eine lineare Funktion mit ausschließlich positiven Koeffizienten maximiert wird und die Wertigkeit der Bits von links nach rechts monoton abnimmt). Ein hinreichendes Kriterium dafür ist (in diesem Fall) nämlich bei Mutationen, die mehr Einsen als Nullen flippen, das folgende: Jeder flippenden Null kann umkehrbar eindeutig eine mindestens so wertige flippende Eins zugeordnet werden. Jede flippende Null bildet also mit einer flippenden Eins ein Paar mit nicht-positivem Anteil an der gesamten Fitness-Wert-Änderung, die eine solche Mutation verursacht. Da mehr Einsen als Nullen flippen, verursacht eine solche Mutation insgesamt mit Sicherheit eine negative Fitness-Wert-Änderung, was das Akzeptieren einer solchen Mutation ausschließt. Also gilt bei der Maximierung einer linearen Funktion mit ausschließlich positiven Koeffizienten

Proposition 3.5.6. Wenn bei einer Mutation, die mehr Einsen als Nullen flippt, jeder flippenden Null umkehrbar eindeutig eine mindestens gleichwertige flippende Eins zugeordnet werden kann, so ist ein Akzeptieren dieser Mutation ausgeschlossen.

Da der Übersichtlichkeit halber immer vorausgesetzt wird, dass die Wertigkeit der Bits von links nach rechts monoton abnimmt, wird hier jedoch eine andere Definition gewählt, um später einfacher bzw. anschaulicher argumentieren zu können:

Definition 3.5.7. [sicher inakzeptable Mutation] Eine Mutation eines bestimmten Individuums wird genau dann „sicher inakzeptabel“ genannt, wenn sie in diesem mehr Einsen als Nullen

flippt und jeder flippenden Null umkehrbar eindeutig eine relativ (zu der jeweiligen Null) links im Individuum stehende flippende Eins zugeordnet werden kann.

Damit auch nicht sicher inakzeptable Mutationen einen Namen haben:

Definition 3.5.8. [potenziell akzeptable Mutation] *Eine Mutation (eines bestimmten Individuums) wird genau dann „potenziell akzeptabel“ genannt, wenn sie nicht sicher inakzeptabel ist.*

Offensichtlich gilt mit diesen Definitionen

Proposition 3.5.9. *Entsteht im Ablauf des [dynamischen] (1+1)-EA, der eine lineare Funktion mit ausschließlich positiven Koeffizienten maximiert, wobei die Wertigkeit der Bits von links nach rechts monoton abnehme, ein Mutant durch eine sicher inakzeptable Mutation, so wird dieser nie selektiert/akzeptiert.*

Auch der folgende Sachverhalt gilt per Definition:

Proposition 3.5.10. *Für eine potenziell akzeptable Mutation (eines bestimmten Individuums) existiert keine Abbildung, die jeder flippenden Null umkehrbar eindeutig eine relativ (zu der jeweiligen Null) links stehende flippende Eins zuordnet.*

Sind alle Koeffizienten einer linearen Funktion paarweise verschieden, so sind diese „geographischen“ Definitionen äquivalent zu den oben angedeuteten „Koeffizienten-basierten“. Eine passende Bijektion muss in diesem Fall jede flippende Null auf eine höherwertige flippende Eins abbilden, welche sicher relativ links steht. Besitzen jedoch verschiedene Bits Koeffizienten gleichen Wertes, so kann sich eine gleichwertige flippende Eins auch rechts von einer Null befinden. Durch die geographische Definition werden somit u. U. Mutationen als potenziell akzeptabel angesehen, für die eine auf den tatsächlichen Koeffizienten(-werten) basierende Bijektion definiert werden könnte, deren Akzeptanz also mit Sicherheit ausgeschlossen ist. Da die Definitionen aber so gewählt wurden, dass sie für alle betrachteten linearen Funktionen sinnvoll sind, werden in der Regel für eine konkrete lineare Funktion aber sowieso Mutationen als potenziell akzeptabel eingestuft, die niemals akzeptiert werden würden. Außerdem liegt mein besonderes Interesse nicht auf linearen Funktionen mit gleich gewichteten Bits, sondern der Ausgangspunkt war mit $BLIC_n$ ja gerade eine mit paarweise verschiedenen Koeffizienten.

Die Hoffnung ist jetzt, dass das Ausschließen der Akzeptanz von „ein paar schlechten“ Mutationen nach dieser einfachen Regel ausreicht, um eine erwartete Zunahme der Anzahl der Einsen im Individuum durch einen Mutation-Selektion-Zyklus des [dynamischen] (1+1)-EA „einfach“ zeigen zu können. Über die erwartete Zunahme der Anzahl der Einsen ließe sich dann wiederum die erwartete Laufzeit abschätzen.

Zurück zum Beispiel, in dem 1-Null-2-Einsen-Mutationen betrachtet wurden: Auch wenn sicher ist, dass von 011-, 101- und 110-Mutationen alleine erstere überhaupt potenziell akzeptabel

sind, so sagt dies noch nichts darüber aus, mit welcher Wahrscheinlichkeit eine solche im Ablauf des [dynamischen] (1+1)-EA auftritt. Eine Aussage wie

$$\text{Prob}(011\text{-Mutation} \mid E) \leq \text{Prob}(101\text{-Mutation} \mid E) \leq \text{Prob}(110\text{-Mutation} \mid E),$$

wobei E das Ereignis bezeichne, dass es sich bei der betrachteten Mutation um eine 1-Null-2-Einsen-Mutation handelt, wäre hilfreich. Aus der Gültigkeit dieser Aussage würde nämlich

$$\text{Prob}(011\text{-Mutation} \mid E) > 1/3 \Rightarrow \text{Prob}(E \mid E) > 1$$

und im Umkehr-Schluss $\text{Prob}(011\text{-Mutation} \mid E) \leq 1/3$ folgen. Dies würde bedeuten, dass eine Mutation im Ablauf des [dynamischen] (1+1)-EA, deren Akzeptieren die Anzahl der Einsen im Individuum um eins verringern würde, da sie genau eine Null und genau zwei Einsen flippt, höchstens mit einer Wahrscheinlichkeit von $1/3$ akzeptiert würde.

Ein Zusammenhang, der für einen Beweis des angedeuteten Sachverhalts in allgemeinerer Form, d. h. für z -Nullen- k -Einsen-Mutationen und nicht nur für 1-Null-2-Einsen-Mutationen, verwendet werden kann, wird nun mit Hilfe der Struktur-Erkenntnis über die Verteilung der Einsen im Individuum bewiesen.

Lemma 3.5.11. *Es wird der i -te Mutation-Selektion-Zyklus in einem Ablauf des [dynamischen] (1+1)-EA auf einer linearen, zu maximierenden Funktion mit ausschließlich positiven Koeffizienten betrachtet, wobei die Wertigkeit der Bits im Individuum von links nach rechts monoton abnimmt. Bezeichne M die betrachtete Mutation. Dann gilt für $J, K, L \in \{0, 1\}^*$:*

$$\text{Prob}(M \text{ ist eine } J1K0L\text{-Mutation}) \geq \text{Prob}(M \text{ ist eine } J0K1L\text{-Mutation})$$

Beweis: Sei $b := |JKL| + 2$. Damit es sich bei M überhaupt um eine $J1K0L$ -Mutation oder eine $J0K1L$ -Mutation handeln kann, muss M offensichtlich eine b -Bit-Mutation sein. In ein und demselben Schritt treten alle b -Bit-Mutationen offensichtlich jeweils mit der gleichen Wahrscheinlichkeit auf. Sei $p(b)$ wieder die Wahrscheinlichkeit für eine bestimmte b -Bit-Mutation im betrachteten Mutation-Selektion-Zyklus, d. h. es gilt $p(b) = p^b(1-p)^{n-b}$, wenn die Mutationswahrscheinlichkeit p verwendet wird. Eine bestimmte b -Bit-Mutation wird repräsentiert durch eine b -elementige Teilmenge $I \subseteq \{1, \dots, n\}$, die genau die Indizes der flippenden Bits enthält. Sei Γ die Menge aller b -elementigen Teilmengen von $\{1, \dots, n\}$. Γ repräsentiert (für Individuen aus $\{0, 1\}^n$) dann genau die $\binom{n}{b}$ möglichen b -Bit-Mutationen.

Für ein bestimmtes Individuum $x \in \{0, 1\}^n$ und eine durch $I \in \Gamma$ bestimmte b -Bit-Mutation bezeichne $B_{x,I} \in \{0, 1\}^b$ die von dieser Mutation in x negierte Sub-Bit-Sequenz, sodass es sich um eine $B_{x,I}$ -Mutation handelt. Wird von M das bestimmte Individuum x mutiert – d. h. x wird

als das im i -ten Schritt aktuelle Individuum angenommen –, so ist die Wahrscheinlichkeit, dass es sich bei M um eine $J0K1L$ -Mutation handelt, gleich

$$\begin{aligned}
& \sum_{I \in \Gamma} \text{Prob}(M \text{ flippt in } x \text{ genau die Bits } x_j \text{ mit } j \in I \text{ und ist eine } J0K1L\text{-Mutation}) \\
&= \sum_{I \in \Gamma} \text{Prob}(M \text{ ist eine bestimmte } b\text{-Bit-Mutation}) \cdot [B_{x,I} = J0K1L] \\
&= \sum_{I \in \Gamma} p(b) \cdot [B_{x,I} = J0K1L].
\end{aligned}$$

Das Prädikat (in den eckigen Klammern) liefert dabei den Wert 1, wenn es erfüllt ist, und sonst den Wert 0. Für den betrachteten i -ten Mutation-Selektion-Zyklus bezeichnet x^{i-1} die Zufallsvariable, die die Verteilung des in diesem Schritt mutierten Individuums über $\{0, 1\}^n$ beschreibt. Da die zufällige Auswahl der Bits, die eine Mutation flippt, unabhängig vom mutierten Individuum ist, gilt somit:

$$\begin{aligned}
\text{Prob}(M \text{ ist eine } J0K1L\text{-Mut.}) &= \sum_{x \in \{0,1\}^n} \text{Prob}(x^{i-1} = x) \cdot \left(\sum_{I \in \Gamma} p(b) \cdot [B_{x,I} = J0K1L] \right) \\
&= \sum_{\substack{I \in \Gamma, \\ x \in \{0,1\}^n}} \text{Prob}(x^{i-1} = x) \cdot p(b) \cdot [B_{x,I} = J0K1L]
\end{aligned}$$

Es wird also über alle möglichen b -Bit-Mutationen aller möglichen Individuen summiert. „Gezählt“ wird eine Mutation aber nur dann, wenn es sich um eine $J0K1L$ -Mutation handelt – und zwar gewichtet mit der Wahrscheinlichkeit, mit der die Kombination aus bestimmtem Individuum und bestimmter Mutation auftritt, die dieser $J0K1L$ -Mutation zugrunde liegt. Analog gilt:

$$\text{Prob}(M \text{ ist eine } J1K0L\text{-Mutation}) = \sum_{\substack{I \in \Gamma, \\ x \in \{0,1\}^n}} \text{Prob}(x^{i-1} = x) \cdot p(b) \cdot [B_{x,I} = J1K0L]$$

Bezeichne P_{J0K1L} die erste und P_{J1K0L} die zweite der beiden oben betrachteten Wahrscheinlichkeiten. Mit Hilfe der in Korollar 3.5.1 (auf Seite 33) formulierten Eigenschaft der Zufallsvariable x^{i-1} wird nun gezeigt, dass jedem Summanden in der Darstellung von P_{J0K1L} umkehrbar eindeutig ein jeweils mindestens so großer Summand in der Darstellung von P_{J1K0L} zugeordnet werden kann.

$B_{x,I}$ ist die durch I bestimmte Sub-Bit-Sequenz des Individuums $x = x_n \dots x_1 \in \{0, 1\}^n$. Für $B_{x,I} = J0K1L \in \{0, 1\}^b$ gilt also $J0K1L = x_{j_b} \dots x_{j_1}$ mit $I = \{j_b, \dots, j_1\}$ und $j_b > \dots > j_1$. Insbesondere gilt $x_{j_{|L|+1}} = 1$ und $x_{j_{|KL|+2}} = 0$. Für jede Kombination von x und I , die zu $B_{x,I} = J0K1L$ führt, gilt demnach für eindeutige $X, Y, Z \in \{0, 1\}^*$: $x = Xx_{j_{|KL|+2}}Yx_{j_{|L|+1}}Z = X0Y1Z$. Bei dem in $X0Y1Z$ herausgestellten Null-Eins-Paar handelt es sich also um zwei Bits

in x , die von der durch I festgelegten Mutation geflippt werden. Ferner bilden genau diese beiden flippenden Bits das in $B_{x,I} = J0K1L$ herausgestellte Null-Eins-Paar. Für das Individuum $y := X1Y0Z \in \{0,1\}^n$ führt die durch I bestimmte b -Bit-Mutation dann zu $B_{y,I} = J1K0L$. Aus Korollar 3.5.1 (auf Seite 33) folgt nun $\text{Prob}(x^{i-1} = X1Y0Z) \geq \text{Prob}(x^{i-1} = X0Y1Z)$. Für $x = X0Y1Z$ und $y = X1Y0Z$ gilt somit:

$$\text{Prob}(x^{i-1} = y) \cdot p(b) \cdot [B_{y,I} = J1K0L] \geq \text{Prob}(x^{i-1} = x) \cdot p(b) \cdot [B_{x,I} = J0K1L]$$

Für beliebige, aber fest gewählte $J, K, L \in \{0,1\}^*$ und eine durch $I \in \Gamma$ festgelegte Mutation kann folglich jedem Individuum x , das zu $B_{x,I} = J0K1L$ führt, umkehrbar eindeutig ein Individuum y zugeordnet werden, das zu $B_{y,I} = J1K0L$ führt und außerdem mindestens so wahrscheinlich das im i -ten Schritt aktuelle Individuum ist wie x . Jedem Summanden in der Darstellung von P_{J0K1L} , in dem das Prädikat erfüllt ist, lässt sich somit umkehrbar eindeutig ein mindestens so großer Summand in der Darstellung von P_{J1K0L} zuordnen. Da ein Summand, bei dem das Prädikat nicht erfüllt ist, offensichtlich gleich Null ist, nimmt P_{J1K0L} schließlich einen mindestens so großen Wert an wie P_{J0K1L} . \square

Der gerade bewiesene Sachverhalt lässt sich auch wie folgt ausdrücken:

Korollar 3.5.12. *Im gesamten Ablauf des [dynamischen] $(1+1)$ -EA auf einer linearen, zu maximierenden Funktion mit ausschließlich positiven Koeffizienten, wobei die Wertigkeit der Bits im Individuum von links nach rechts monoton abnimmt, gilt für $J, K, L \in \{0,1\}^*$: In ein und demselben Mutation-Selektion-Zyklus findet mindestens so wahrscheinlich eine $J1K0L$ -Mutation statt wie eine $J0K1L$ -Mutation.*

Um diese neue Erkenntnis über die Wahrscheinlichkeit des Auftretens bestimmter Mutationen einfacher verwenden zu können, wird nun auf der Menge der Mutationen, die in einem bestimmten Schritt auftreten können, die Relation R definiert:

Definition 3.5.13. [die Relation R] Für $A, B \in \{0,1\}^i$, $2 \leq i \leq n$, sei die Relation R wie folgt definiert:

$$R(A\text{-Mutation}, B\text{-Mutation}) \quad :\Leftrightarrow \quad (\exists J, K, L \in \{0,1\}^*) \quad A = J0K1L \wedge B = J1K0L$$

„ $R(A\text{-Mutation}, B\text{-Mutation})$ “ wird auch durch „ $A\text{-Mutation} \leq_R B\text{-Mutation}$ “ ausgedrückt.

Für diese Relation R gilt nämlich mit dem vorangegangenen Korollar 3.5.12 bzw. mit Lemma 3.5.11 (auf Seite 38):

Proposition 3.5.14. *Für $A, B \in \{0,1\}^i$, $2 \leq i \leq n$, gilt in jedem Schritt S im Ablauf des [dynamischen] $(1+1)$ -EA auf einer linearen Funktion, die zu maximieren ist und deren Koeffizienten*

ausschließlich positiv sind, wobei die Wertigkeit der Bits im Individuum von links nach rechts monoton abnimmt, die folgende Implikation:

$$\begin{aligned} & A\text{-Mutation} \leq_R B\text{-Mutation} \\ \Rightarrow & \text{Prob}(\text{in } S \text{ tritt eine } A\text{-Mutation auf}) \leq \text{Prob}(\text{in } S \text{ tritt eine } B\text{-Mutation auf}) \end{aligned}$$

Für das obige Beispiel, in dem 1-Null-2-Einsen-Mutationen betrachtet wurden, bzw. für Mutationen, die genau eine Null und mindestens zwei Einsen flippen, kann nun mit Hilfe von R der zu Beginn dieses Abschnitts vermutete Sachverhalt gezeigt werden. Vermutet wurde, dass eine solche „schlechte Mutation“, die die Anzahl der Einsen verringern könnte (da sie mehr Einsen als Nullen flippt), im Ablauf des [dynamischen] (1+1)-EA auf einer linearen Funktion mit ausschließlich positiven Koeffizienten „wahrscheinlich“ verworfen wird, der drohende „Einsen-Verlust“ also im Selektionsschritt „eher nicht“ realisiert wird.

Lemma 3.5.15. *Es wird der Ablauf des [dynamischen] (1+1)-EA auf einer linearen, zu maximierenden Funktion mit ausschließlich positiven Koeffizienten betrachtet. Dann gilt für $k \geq 2$: Tritt im Mutationsschritt eines Mutation-Selektion-Zyklus eine 1-Null- k -Einsen-Mutation auf, so wird der entstehende Mutant im anschließenden Selektionsschritt höchstens mit einer Wahrscheinlichkeit von $1/(k+1)$ selektiert. Die 1-Null- k -Einsen-Mutation wird also mit einer Wahrscheinlichkeit von mindestens $k/(k+1)$ verworfen.*

Beweis: Die Wertigkeit der Bits im Individuum nehme o. B. d. A. von links nach rechts monoton ab. Für $k \geq 1$ ist R^* , die transitive Hülle der oben definierten Relation R , auf der Menge der 1-Null- k -Einsen-Mutationen eine totale Ordnung, denn es gilt:

$$01^k\text{-Mutation} \leq_R 101^{k-1}\text{-Mutation} \leq_R 1101^{k-2}\text{-Mutation} \leq_R \dots \leq_R 1^k0\text{-Mutation}$$

Eine 1-Null- k -Einsen-Mutation ist für $k \geq 2$ nur dann potenziell akzeptabel, wenn es sich um eine 01^k -Mutation handelt. Da die übrigen k der $\binom{k+1}{1} = k+1$ möglichen Mutationstypen ausschließlich sicher inakzeptable Mutationen charakterisieren und jede dieser k , wie zuvor mit Hilfe von R^* resp. Lemma 3.5.11 (auf Seite 38) bewiesen, mindestens so wahrscheinlich auftritt wie eine 01^k -Mutation, folgt:

$$\text{Prob}(01^k\text{-Mutation} \mid 1\text{-Null-}k\text{-Einsen-Mutation}) \leq \frac{1}{k+1}$$

Daraus folgt der behauptete Sachverhalt. □

Die Struktur-Erkenntnis über die Verteilung der Einsen im jeweils aktuellen Individuum wurde dazu verwendet, die Wahrscheinlichkeit des Auftretens von bestimmten Mutationen in Relation

setzen zu können. Mit Hilfe einer Aussage darüber, für welche Mutationstypen ein Akzeptieren ausgeschlossen werden kann, konnte dies wiederum beispielhaft dazu genutzt werden, die Wahrscheinlichkeit für den „Verlust von $k - 1$ Einsen“ beim Auftreten einer 1-Null- k -Einsen-Mutation im Ablauf des [dynamischen] (1+1)-EA auf einer linearen Funktion mit ausschließlich positiven Koeffizienten nach oben durch $1/(k + 1)$ zu beschränken. Wie diese neue strukturelle Erkenntnis über die Wahrscheinlichkeit des Auftretens von bestimmten Mutationen wiederum angewendet werden kann, um (obere) Laufzeit-Schranken zu zeigen, wird im nächsten Kapitel behandelt.

Kapitel 4

Anwendung der strukturellen Erkenntnisse

4.1 Die erwartete Veränderung der Anzahl der Einsen

Eine b -Bit-Mutation wird durch eine b -elementige Teilmenge von $\{1, \dots, n\}$ bestimmt, die genau die Indizes der flippenden Bits enthält. Sei Γ die Potenzmenge von $\{1, \dots, n\}$. Wird das Individuum $x = x_n \dots x_1 \in \{0, 1\}^n$ mutiert, so ist die für den aus x erzeugten Mutanten erwartete Veränderung der Anzahl der Einsen gleich

$$\sum_{I \in \Gamma} \text{Prob}(\text{genau die } x_i \text{ mit } i \in I \text{ flippen}) \cdot (\#\{i \in I \mid x_i = 0\} - \#\{i \in I \mid x_i = 1\}).$$

Soll nun die erwartete Veränderung der Anzahl der Einsen im Individuum für einen Mutation-Selektion-Zyklus bestimmt werden, so muss beachtet werden, dass ein aus x erzeugter Mutant nur dann akzeptiert/Selektiert wird, wenn er mindestens so fit ist wie x . Es bezeichne $p(b)$ wieder die Wahrscheinlichkeit für eine bestimmte b -Bit-Mutation in einem betrachteten Mutation-Selektion-Zyklus, d. h. es gilt $p(b) = p^b(1-p)^{n-b}$, wobei p die in diesem Schritt verwendete Mutationswahrscheinlichkeit ist. Ferner sei $F : \{0, 1\}^n \rightarrow \mathbb{R}$ die zu maximierende Fitnessfunktion und $\text{Mut}(x, I)$ bezeichne den Mutanten von x , der durch die Mutation entsteht, die durch I bestimmt ist. Wird in einem Mutation-Selektion-Zyklus während der Maximierung von F durch den [dynamischen] (1+1)-EA das Individuum x mutiert, so ist die von diesem Schritt erwartete Veränderung der Anzahl der Einsen im Individuum gleich

$$\sum_{I \in \Gamma} p(\#I) \cdot (\#\{i \in I \mid x_i = 0\} - \#\{i \in I \mid x_i = 1\}) \cdot [F(\text{Mut}(x, I)) \geq F(x)]. \quad (*)$$

Besitzt nämlich der im betrachteten Schritt aus x erzeugte Mutant eine geringere Fitness als x , so wird er verworfen (d. h. x wird selektiert) und die Anzahl der Einsen im Individuum wird in diesem Fall nicht (bzw. um Null) verändert.

Die $\#\Gamma = 2^n$ Summanden in der Summen-Darstellung der durch einen Mutation-Selektion-Zyklus erwarteten Veränderung der Anzahl der Einsen im Individuum können nun (beliebig) gruppiert werden, sodass dieser Erwartungswert möglichst gut abgeschätzt werden kann. Im Folgenden werden die Summanden danach gruppiert, wie viele Nullen eine betrachtete Mutation im aktuellen Individuum flippt. Das heißt für $0 \leq z \leq n$ werden jeweils alle Summanden in (*), bei denen die jeweils durch I bestimmte Mutation des Individuums x eine z -Nullen-Mutation ist, in einer Gruppe zusammengefasst und gemeinsam abgeschätzt.

Das Akzeptieren einer z -Nullen- k -Einsen-Mutation in einem Mutation-Selektion-Zyklus verändert die Anzahl der Einsen um $z - k$, da z Nullen zu Einsen und k Einsen zu Nullen flippen. Für die Wahrscheinlichkeit des Auftretens einer solchen Mutation gilt unabhängig von der betrachteten Funktion

Proposition 4.1.1. *Die Wahrscheinlichkeit, dass eine Mutation des Individuums x mit der Mutationswahrscheinlichkeit p genau z Nullen und genau k Einsen flippt – es sich also um eine z -Nullen- k -Einsen-Mutation handelt – beträgt*

$$P_{p,x}(z, k) := \binom{|x|_0}{z} \cdot \binom{|x|_1}{k} \cdot p^{z+k} \cdot (1-p)^{n-(z+k)},$$

wobei die Indizes p und x bei Eindeutigkeit auch entfallen können.

Wenn nun die erwartete Veränderung der Anzahl der Einsen durch einen Mutation-Selektion-Zyklus nach unten abgeschätzt werden soll, so müssen alle „schlechten“ Mutationen, durch deren Akzeptieren die Anzahl der Einsen verringert würde, beachtet werden. Von den „guten“ Mutationen, durch deren Akzeptieren die Anzahl der Einsen erhöht würde, können dagegen welche unbeachtet bleiben. Neben der oben bestimmten Wahrscheinlichkeit des Auftretens ist die Akzeptanzwahrscheinlichkeit einer z -Nullen- k -Einsen-Mutation entscheidend dafür, welchen Einfluss eine solche auf die erwartete Veränderung der Anzahl der Einsen durch einen Mutation-Selektion-Zyklus hat. Es gelte deshalb

Notation 4.1.2. $[A_{x,p}(z, k)]$ *Es wird der Ablauf des [dynamischen] (1+1)-EA auf einer bestimmten Funktion betrachtet. Für $k > z \geq 1$ bezeichne $A_{p,x}(z, k)$ eine obere Schranke für die Akzeptanzwahrscheinlichkeit einer z -Nullen- k -Einsen-Mutation, die in einem Mutation-Selektion-Zyklus auftritt, in dem die Mutationswahrscheinlichkeit p verwendet wird und das Individuum x mutiert wird. Bei Eindeutigkeit können die Indizes p und x auch entfallen.*

Im Folgenden wird ein Mutation-Selektion-Zyklus im Ablauf des [dynamischen] (1+1)-EA auf einer linearen Funktion mit ausschließlich positiven Koeffizienten betrachtet, in dem die Mutationswahrscheinlichkeit p verwendet wird und das Individuum x mutiert wird.

Sei $z \geq 1$ fest gewählt. Für $k > z$ tritt eine z -Nullen- k -Einsen-Mutation mit der Wahrscheinlichkeit $P(z, k)$ auf, wird höchstens mit der Wahrscheinlichkeit $A(z, k)$ akzeptiert und verändert dann die Anzahl der Einsen um $z - k$, d. h. er verringert diese dann um $k - z \geq 1$. Wird sie nicht akzeptiert, so bleibt das Individuum (und damit auch die Anzahl der Einsen) unverändert. Alle Summanden in der Summen-Darstellung der erwarteten Veränderung der Anzahl der Einsen (*) (auf Seite 43), bei denen die jeweils durch I bestimmte Mutation des Individuums x eine z -Nullen- k -Einsen-Mutation ist, haben also zusammen einen nicht-positiven Wert von mindestens $P(z, k) \cdot A(z, k) \cdot (z - k)$.

Eine z -Nullen-0-Einsen-Mutation tritt mit der Wahrscheinlichkeit $P(z, 0)$ auf, wird, da keine Einsen flippen, mit Sicherheit akzeptiert und erhöht die Anzahl der Einsen um z . Alle Summanden in der Summen-Darstellung der erwarteten Veränderung der Anzahl der Einsen (*), bei denen die jeweils durch I bestimmte Mutation des Individuums x eine z -Nullen-0-Einsen-Mutation ist, haben folglich zusammen einen Wert von genau $P(z, 0) \cdot z$.

Alle zuvor betrachteten Summanden haben (für das fest gewählte z) zusammen einen Wert von mindestens

$$\Delta_{x,p}(z) := P_{x,p}(z, 0) \cdot z + \sum_{z < k \leq |x|_1} P_{x,p}(z, k) \cdot A_{x,p}(z, k) \cdot (z - k).$$

Bei Eindeutigkeit von x und p werden diese Indizes wieder fortgelassen. Bei dieser Abschätzung wird (für das fest gewählte z) die Akzeptanzwahrscheinlichkeit einer z -Nullen- k -Einsen-Mutation für $1 \leq k \leq z$ nach unten durch Null abgeschätzt, da eine solche (bei einem Auftreten) zwar die Anzahl der Einsen möglicherweise erhöht, jedoch die bisherigen, in diesem Text gezeigten Erkenntnisse die Abschätzung der Akzeptanzwahrscheinlichkeit einer solchen Mutation nach unten nicht ermöglichen. Da aber zum Ende der Optimierung – wenn es interessant wird – das Individuum sehr viel mehr Einsen als Nullen enthält, ist dann schon die Wahrscheinlichkeit des Auftretens einer Mutation, die mehr Nullen als Einsen flippt, sehr gering, sodass das Abschätzen dieser Akzeptanzwahrscheinlichkeiten durch Null dann nicht zu einer schlechten Abschätzung von (*) führt. $\Delta(z)$ ist also eine untere Schranke für die Summe aller Summanden in (*), in der Summen-Darstellung der erwarteten Veränderung der Anzahl der Einsen im Individuum durch den betrachteten Mutation-Selektion-Zyklus, bei denen die jeweils durch I festgelegte Mutation des Individuums x eine z -Nullen-Mutation ist.

Da eine Mutation potenziell $z \in \{0, \dots, |x|_0\}$ Nullen flippen kann, wird vom betrachteten Mutation-Selektion-Zyklus insgesamt eine Veränderung der Anzahl der Einsen um mindestens

$$\Delta_{x,p} := \sum_{z=0}^{|x|_0} \Delta_{x,p}(z)$$

erwartet. Bei Eindeutigkeit von x und p entfallen diese Indizes wieder.

Nun wird die Summe

$$\Delta_x(z) = P(z, 0) \cdot z + \sum_{z < k \leq |x|_1} P(z, k) \cdot A(z, k) \cdot (z - k)$$

genauer betrachtet. Für $z = 0$ gilt folglich:

$$\Delta_x(0) = - \sum_{k=1}^{|x|_1} P(z, k) \cdot A(z, k) \cdot k \leq 0$$

Da bei den hier betrachteten linearen Funktionen mit ausschließlich positiven Koeffizienten ein Mutation-Selektion-Zyklus, in dessen Mutationsschritt keine Null flippt, zu einem unveränderten Individuum führt¹, gilt bei einer solchen Funktion $\Delta(0) = 0$. Für $z > 0$ lässt sich die Formel für $\Delta(z)$ folgendermaßen umformen. Das Ausklammern von z ergibt:

$$\Delta_x(z) = z \cdot \left(P(z, 0) + \sum_{z < k \leq |x|_1} P(z, k) \cdot A(z, k) \cdot \frac{z - k}{z} \right)$$

Weiterhin kann $\binom{|x|_0}{z}$ ausgeklammert werden, da es in allen $P(z, \cdot)$ vorkommt:

$$\Delta(z) = z \cdot \binom{|x|_0}{z} \cdot \left(p^z \cdot (1 - p)^{n-z} + \sum_{z < k \leq |x|_1} \binom{|x|_1}{k} \cdot p^{z+k} \cdot (1 - p)^{n-(z+k)} \cdot A(z, k) \cdot \frac{z - k}{z} \right)$$

Schließlich kann noch $p^z(1 - p)^{n-z}$ „heraus gezogen“ werden, sodass für $z > 0$ gilt:

$$\Delta(z) = z \cdot \binom{|x|_0}{z} \cdot p^z(1 - p)^{n-z} \cdot \left(1 + \sum_{z < k \leq |x|_1} \binom{|x|_1}{k} \cdot \left(\frac{p}{1 - p} \right)^k \cdot A(z, k) \cdot \frac{z - k}{z} \right)$$

Für $1 \leq z \leq |x|_0$ nimmt dieses Produkt das gleiche Vorzeichen an wie der rechteste, in der großen Klammer stehende Faktor.

Es bezeichne y das in dem betrachteten Mutation-Selektion-Zyklus selektierte Individuum. In diesem Mutation-Selektion-Zyklus, in dem das Individuum x mutiert wird und dafür die Mutationswahrscheinlichkeit p verwendet wird, gilt folglich für die erwartete Veränderung der Anzahl der Einsen:

$$E[|y|_1 - |x|_1] \geq \Delta_{x,p} = \sum_{z=0}^{|x|_0} \Delta_{x,p}(z)$$

¹entweder wird die Mutation verworfen, oder die 0-Bit-Mutation wird akzeptiert

Diese Abschätzung gilt nicht nur bei linearen Funktionen mit ausschließlich positiven Koeffizienten, sondern für alle Funktionen, bei denen Mutationen, die ausschließlich Nullen flippen, mit Sicherheit akzeptiert werden. Zur Abschätzung von Laufzeiten kann sie dann (sinnvoll) eingesetzt werden, wenn der erwartete Fortschritt anhand der Potential-Funktion ONEMAX bestimmt werden kann bzw. soll. Für lineare Funktionen mit ausschließlich positiven Koeffizienten drängt sich diese Vorgehensweise förmlich auf. Um gute, d. h. möglichst kleine, obere Laufzeit-Schranken zu erhalten, sollte $\Delta_{x,p}$ nicht nur positiv, sondern auch noch möglichst groß sein. $\Delta_{x,p}$ muss also möglichst gut nach unten abgeschätzt werden. Dass dies trotz der kompliziert aussehenden Formel für $\Delta_{x,p}(z)$ tatsächlich bewerkstelligt werden kann, wird im nächsten Abschnitt gezeigt.

4.2 Die Laufzeit des Standard-(1+1)-EA auf linearen Funktionen

Wie schon erwähnt wurde, gilt für lineare Funktionen mit ausschließlich positiven Koeffizienten $\Delta(0) = 0$, da – unabhängig vom mutierten Individuum x und der verwendeten Mutationswahrscheinlichkeit p – eine Mutation, die keine Null flippt, nur dann akzeptiert wird, wenn sie auch keine Eins flippt. Entweder wird also ein „Clone“ akzeptiert, oder aber der Mutant wird verworfen. In beiden Fällen bleibt das Individuum – und damit auch die Anzahl der Einsen – unverändert. Um für $z \geq 1$

$$\Delta_{x,p}(z) = z \cdot \binom{|x|_0}{z} \cdot p^z (1-p)^{n-z} \cdot \left(1 + \sum_{z < k \leq |x|_1} \binom{|x|_1}{k} \cdot \left(\frac{p}{1-p} \right)^k \cdot A(z, k) \cdot \frac{z-k}{z} \right)$$

nach unten abzuschätzen, wird im Folgenden zunächst die Summe in der großen Klammer für die beim Standard-(1+1)-EA verwendete Mutationswahrscheinlichkeit $p = 1/n$ nach unten abgeschätzt.

Lemma 4.2.1. *Unabhängig vom betrachteten Individuum $x \in \{0, 1\}^n$ gilt für $p = 1/n$ und $1 \leq z \leq |x|_0$ bei linearen Funktionen mit ausschließlich positiven Koeffizienten: $\Delta_{x,p}(z) \geq 0$*

Beweis: Offensichtlich gilt $z \cdot \binom{|x|_0}{z} \cdot p^z (1-p)^{n-z} \geq 0$. Es reicht also zu zeigen, dass

$$1 + \sum_{z < k \leq |x|_1} \binom{|x|_1}{k} \cdot \left(\frac{p}{1-p} \right)^k \cdot A(z, k) \cdot \frac{z-k}{z} \geq 0$$

gilt. $A(z, k) := 1$ ist sicher eine obere Schranke für die Akzeptanzwahrscheinlichkeit einer jeden

Mutation, sodass die folgende Ungleichung zu zeigen ausreicht (Anm.: $-(z - k) = k - z$):

$$\sum_{z < k \leq |x|_1} \binom{|x|_1}{k} \cdot \left(\frac{p}{1-p}\right)^k \cdot \frac{k-z}{z} \leq 1$$

Die Summe auf der linken Seite wird im Folgenden genauer betrachtet. Es gilt

$$\left(\frac{p}{1-p}\right)^k = \left(\frac{1}{n \cdot \left(1 - \frac{1}{n}\right)}\right)^k = (n-1)^{-k}$$

und, da $|x|_1 = n - |x|_0 = n - z \leq n - 1$ gilt, ferner

$$\binom{|x|_1}{k} \leq \frac{(n-1)^k}{k!}.$$

Die betrachtete Summe lässt sich somit durch

$$\begin{aligned} \sum_{z < k \leq |x|_1} \binom{|x|_1}{k} \cdot \left(\frac{p}{1-p}\right)^k \cdot \frac{k-z}{z} &\leq \sum_{z < k \leq |x|_1} \frac{(n-1)^k}{k!} \cdot (n-1)^{-k} \cdot \frac{k-z}{z} \\ &= \sum_{z < k \leq |x|_1} \frac{k-z}{k! \cdot z} \end{aligned}$$

nach oben abschätzen. Da $z \geq 1$ vorausgesetzt ist, ist der Wert der rechten Summe hier für $z = 1$ maximal, sodass diese weiter abgeschätzt werden kann:

$$\sum_{z < k \leq |x|_1} \frac{k-z}{k! \cdot z} \leq \sum_{k \geq 2} \frac{k-1}{k!} = \sum_{k \geq 2} \left(\frac{1}{(k-1)!} - \frac{1}{k!} \right) = \frac{1}{(2-1)!} = 1$$

Daraus folgt der behauptete Sachverhalt. \square

Da, wie bereits mehrfach erwähnt, für lineare Funktionen mit ausschließlich positiven Koeffizienten $\Delta(0) = 0$ gilt, wurde damit gezeigt, dass für jede mögliche Anzahl flippender Nullen z , d. h. für $z \in \{0, \dots, |x|_0\}$, $\Delta(z) \geq 0$ gilt. Es gilt also insbesondere $\Delta_x = \sum_{z=0}^{|x|_0} \Delta(z) \geq 0$. Um aber überhaupt eine Laufzeit-Schranke mit Hilfe der Potential-Funktion ONEMAX zeigen zu können, muss $\Delta > 0$ gelten und für kleine obere Laufzeit-Schranken sollte Δ darüberhinaus möglichst groß sein. Da bei einer Mutationswahrscheinlichkeit von $1/n$ im Erwartungsfall eine Mutation nur genau ein Bit flippt und Mutation-Selektion-Zyklen, in denen keine Null flippt, die Anzahl der Einsen nicht verändern, ist zu vermuten, dass 1-Null-0-Einsen-Mutationen am häufigsten zu einer Erhöhung der Anzahl der Einsen durch einen Mutation-Selektion-Zyklus führen. Aus diesem Grund wird nun die Abschätzung von $\Delta(1)$ noch einmal genauer betrachtet und

verbessert. In der Formel für

$$\Delta_{x,p}(z) = z \cdot \binom{|x|_0}{z} \cdot p^z (1-p)^{n-z} \cdot \left(1 + \sum_{z < k \leq |x|_1} \binom{|x|_1}{k} \cdot \left(\frac{p}{1-p} \right)^k \cdot A(z, k) \cdot \frac{z-k}{z} \right)$$

wurde die Summe, in der über k summiert wird, im vorangegangenen Lemma für $z = 1$ mit der trivialen, sicheren Annahme $A(z, k) = 1$ durch der Wert -1 nach unten abgeschätzt. Es wurde also lediglich ausgeschlossen, dass der rechteste, in großen Klammern stehende Faktor einen negativen Wert annimmt. Mit Hilfe der Struktur-Erkenntnisse über den Ablauf des [dynamischen] (1+1)-EA auf linearen Funktionen mit ausschließlich positiven Koeffizienten konnte in Lemma 3.5.15 (auf Seite 41) gezeigt werden, dass eine 1-Null- k -Einsen-Mutation für $k \geq 2$ nur mit einer Wahrscheinlichkeit von höchstens $1/(k+1)$ akzeptiert wird. Für $k \geq 2$ kann also $A(1, k) = 1/(k+1)$ als obere Schranke für die Akzeptanzwahrscheinlichkeit einer 1-Null- k -Einsen-Mutation gewählt werden. Die besagte Summe kann also besser abgeschätzt werden:

Lemma 4.2.2. Für $x \in \{0, 1\}^n$, $z = 1$, $p = 1/n$ und $A_{x,p}(1, k) = 1/(k+1)$ gilt:

$$\sum_{z < k \leq |x|_1} \binom{|x|_1}{k} \cdot \left(\frac{p}{1-p} \right)^k \cdot A(z, k) \cdot \frac{k-z}{z} < 0,282$$

Beweis: Wie schon im Beweis des vorigen Lemmas gezeigt wurde, kann die Summe auf der linken Seite nach oben durch

$$\sum_{z < k \leq |x|_1} \frac{1}{k!} \cdot A(z, k) \cdot \frac{k-z}{z}$$

abgeschätzt werden. Das Einsetzen der festgelegten Werte ergibt dann:

$$\sum_{z < k \leq |x|_1} \frac{1}{k!} \cdot A(z, k) \cdot \frac{k-z}{z} \leq \sum_{k \geq 2} \frac{1}{k!} \cdot \frac{1}{k+1} \cdot (k-1) = \sum_{k \geq 2} \frac{k-1}{(k+1)!}$$

Die rechte Reihe kann weiter umgeformt werden:

$$\sum_{k \geq 2} \frac{k-1}{(k+1)!} = \sum_{k \geq 3} \frac{k-2}{k!} = \sum_{k \geq 2} \frac{1}{k!} - \sum_{k \geq 3} \frac{2}{k!} = (e-2) - 2(e-2,5)$$

Da $(e-2) - 2(e-2,5) = 3-e < 0,282$ gilt, folgt die Behauptung. \square

Dies verbessert die Abschätzung von $\Delta(1)$ entscheidend. Aus den beiden vorangegangenen Lemmata folgt nämlich nun

Korollar 4.2.3. *Maximiert der Standard-(1+1)-EA eine lineare Funktion mit ausschließlich positiven Koeffizienten, so gilt für die von einem Mutation-Selektion-Zyklus erwartete Veränderung der Anzahl der Einsen:*

$$\Delta_{x,1/n} = \sum_{z=0}^{|x|_0} \Delta_{x,1/n}(z) \geq \Delta_{x,1/n}(1) \geq \binom{|x|_0}{1} \cdot \frac{1}{n} \left(1 - \frac{1}{n}\right)^{n-1} \cdot (1 - 0,282)$$

Dabei bezeichnet x das (zu Beginn dieses Mutation-Selektion-Zyklus) aktuelle Individuum.

Aus diesem Sachverhalt folgt wiederum direkt

Korollar 4.2.4. *Maximiert der statischen (1+1)-EA mit der Standard-Mutationswahrscheinlichkeit $p = 1/n$ eine lineare Funktion, deren Koeffizienten ausschließlich positiv sind, so gilt: Wird im i -ten Schritt das suboptimale Individuum $x = x^{i-1} \in \{0, 1\}^n \setminus \{1^n\}$ mutiert und das selektierte Individuum mit x^i bezeichnet, so lässt sich $E[|x^i|_1 - |x^{i-1}|_1]$, die erwartete Änderung der Anzahl der Einsen durch diesen Schritt, durch*

$$\binom{|x|_0}{1} \cdot \frac{1}{n} \left(1 - \frac{1}{n}\right)^{n-1} \cdot (1 - 0,282) \geq \frac{|x|_0}{n} \cdot \frac{0,718}{e} > \frac{|x|_0}{n} \cdot 0,264 = \Omega(|x|_0/n)$$

nach unten abschätzen.

In jedem Mutation-Selektion-Zyklus wird also (mindestens) eine Erhöhung der Anzahl der Einsen in der Größenordnung des (relativen) Nullen-Anteils im Eltern-Individuum erwartet. Um dieses Ergebnis in eine Laufzeit-Schranke „ummünzen“ zu können, muss gezeigt werden, dass aus der erwarteten Erhöhung der Anzahl der Einsen um $\Delta = \Omega(|x|_0/n)$ folgt, dass eine solche im Erwartungsfall nach $O(1/\Delta) = O(n/|x|_0)$ Schritten eintritt. Die Anwendung der Wald'schen Identität, die genau eine solche Aussage beinhaltet, ist hier aus zwei Gründen nicht möglich: Erstens ist der erwartete Fortschritt nicht unabhängig vom Zeitpunkt, da die erwartete Erhöhung der Anzahl der Einsen von x , dem aktuellen Individuum abhängt. Zweitens ist ein Fortschritt, d. h. eine Erhöhung der Anzahl der Einsen, von mehr als eins möglich. Gleichwohl ist natürlich zu vermuten, dass der obige Zusammenhang dennoch gilt.

Theorem 4.2.5. *Die erwartete Anzahl von Schritten, bis der Standard-(1+1)-EA eine lineare Funktion $F = \sum_{i=1}^n a_i \cdot x_i$ mit $a_1, \dots, a_n \in \mathbb{R} \setminus \{0\}$ maximiert hat, ist durch $3,8 \cdot n \cdot H(n)$ beschränkt. Die erwartete Laufzeit ist also durch $O(n \cdot \log n)$ beschränkt.*

Beweis: Ohne Beschränkung der Allgemeinheit kann $a_n \geq \dots \geq a_1 > 0$ angenommen werden. Sei also F eine Funktion, deren Koeffizienten ausschließlich positiv sind. Im Verlauf der Optimierung von F wird bei einem Individuum mit i Einsen erwartet, dass ein Mutation-Selektion-Zyklus die Anzahl der Einsen um mindestens $0,264 \cdot (n - i)/n$ erhöht. Wäre die erwartete

Anzahl von Schritten, bis aus einem Individuum mit i Einsen erstmals ein aktuelles Individuum mit mehr als i Einsen erzeugt wird, nach oben durch den Kehrwert dieser erwarteten Einsenzunahme beschränkt, so würde für die erwartete Laufzeit

$$E[\text{Laufzeit}] \leq \sum_{n>i\geq 0} \frac{1}{0,264 \cdot (n-i)/n} \leq 3,8 \cdot n \cdot H(n)$$

gelten. Hierbei wird pessimistischerweise angenommen, dass das schlechteste Individuum 0^n als Start-Individuum erzeugt wird. Da für die n -te harmonische Zahl $H(n) = \sum_{i=1}^n 1/i = \ln n + \Theta(1)$ gilt, würde somit die Behauptung folgen.

Um diese Laufzeit-Schranke zu zeigen, wird nun ein stochastischer Prozess, genauer ein „random walk“, definiert und analysiert, der (im Erwartungsfall) nicht schneller fortschreitet als die Optimierung von F durch den Standard-(1+1)-EA. Bei dem neuen Prozess sind die Zustände $0, 1, \dots, n$ möglich, da ein Individuum genau diese Anzahlen von Einsen besitzen kann. Abhängig vom Zustand i tritt in einem Schritt des neuen Prozesses eine (Zustands-) Veränderung um $\delta(i) \in \mathbb{Z}$ mit einer bestimmten, noch zu definierenden Wahrscheinlichkeit auf. Wichtig für eine spätere Argumentation analog des Beweises der Wald'schen Identität ist, dass hierbei $\text{Prob}(\delta(i) > 1) = 0$ gilt – im Gegensatz zum Ablauf des (1+1)-EA auf F , bei dem eine Erhöhung der Anzahl der Einsen durch einen Mutation-Selektion-Zyklus um mehr als eins nicht ausgeschlossen ist. Deshalb wird beim neuen Prozess für $d \geq 2$: $\text{Prob}(\delta(i) = d) := 0$ gewählt und $\text{Prob}(\delta(i) = 1)$ gleich der Wahrscheinlichkeit gesetzt, mit der bei der Optimierung von F in einem Mutation-Selektion-Zyklus, in dem ein Individuum mit i Einsen mutiert wird, die Anzahl der Einsen um eins oder mehr wächst. Für $d \leq 0$ wird $\text{Prob}(\delta(i) = d)$ gleich der Wahrscheinlichkeit gesetzt, mit der in einem solchen Mutation-Selektion-Zyklus eine Veränderung der Anzahl der Einsen um d , d. h. eine Verringerung um $-d \geq 0$, eintritt.

Im Folgenden wird verdeutlicht, warum der so definierte Prozess im Erwartungsfall nicht schneller fortschreitet als die Optimierung von F durch den Standard-(1+1)-EA. Bezeichne S_i^{EA} die erwartete Anzahl Schritte, die beim (1+1)-EA einem Schritt, in dem aus einem Individuum, das nicht genau i Einsen enthält, eines mit genau i Einsen erzeugt wurde, folgen, bis erstmals nach diesem Schritt ein Individuum mit mehr als i Einsen erzeugt wird. S_i^P bezeichne die erwartete Anzahl Schritte, die beim neuen Prozess (P) einem Schritt, in dem der Zustand i von einem Zustand ungleich i erreicht wurde, folgen, bis erstmals nach diesem Schritt der Zustand $i+1$ (der nicht „übersprungen“ werden kann) erreicht wird. Offensichtlich gilt $S_0^{EA} = S_0^P$. Wird für eine Induktion angenommen, dass für $i \in \{0, \dots, k-1\}$ mit $0 < k < n$ die Ungleichung $S_i^{EA} \leq S_i^P$ gilt, so kann $S_k^{EA} \leq S_k^P$ gefolgert werden: $\text{Prob}(\delta(k) = 1)$ ist so groß wie die Wahrscheinlichkeit, dass in einem Mutation-Selektion-Zyklus aus einem Individuum mit genau k Einsen eines mit mehr als k Einsen erzeugt wird. Ferner ist $\text{Prob}(\delta(k) = d)$ für $d < 0$ gleich

der Wahrscheinlichkeit, dass in einem Mutation-Selektion-Zyklus aus einem Individuum mit genau k Einsen eines mit genau $k + d < k$ Einsen erzeugt wird. Der neue Prozess benötigt im Zustand $k + d$ im Erwartungsfall $\sum_{k+d \leq i < k} S_i^P$ Schritte, um wieder den Zustand k zu erreichen. Der (1+1)-EA benötigt im Erwartungsfall höchstens $\sum_{k+d \leq i < k} S_i^{EA}$ Schritte, um aus einem Individuum mit $k + d$ Einsen wieder eine Individuum mit k oder mehr Einsen zu erzeugen. Da mit der Induktionsvoraussetzung $\sum_{k+d \leq i < k} S_i^{EA} \leq \sum_{k+d \leq i < k} S_i^P$ gilt, folgt für den Induktionsschritt somit $S_k^{EA} \leq S_k^P$. Folglich gilt $S_i^{EA} \leq S_i^P$ für $0 \leq i < n$. Die erwartete Anzahl Schritte, bis der neue Prozess, gestartet im Zustand 0, erstmals den Zustand $k \in \{1, \dots, n\}$ erreicht, ist gleich $T_k^P := \sum_{0 \leq i < k} S_i^P$. Die erwartete Anzahl Schritte, bis der (1+1)-EA erstmals aus dem Individuum 0^n ein Individuum mit mindestens k Einsen erzeugt hat, ist durch $T_k^{EA} := \sum_{0 \leq i < k} S_i^{EA}$ nach oben beschränkt. Da, wie oben gezeigt wurde, $S_i^{EA} \leq S_i^P$ für $0 \leq i < n$ gilt, gilt folglich $T_i^{EA} \leq T_i^P$ für $1 \leq i \leq n$. Werden die beiden Prozesse im Zustand 0 bzw. mit dem Individuum 0^n gestartet, so ist also zu erwarten, dass während der Optimierung von F mindestens so schnell erstmals ein Individuum mit i oder mehr Einsen erzeugt wird, wie der neue Prozess erstmals den Zustand i erreicht.

Interessant wird es nun bei der Abschätzung von $E[\delta(i)]$. Durch einen Mutation-Selektion-Zyklus, indem ein Individuum mit i Einsen bzw. $n - i$ Nullen mutiert wird, wird bei der Optimierung von F eine Erhöhung der Anzahl der Einsen um mindestens $0,264 \cdot (n - i)/n$ erwartet. Damit die behauptete Laufzeit-Schranke zutrifft, muss demnach für den neuen Prozess gezeigt werden, dass auch $E[\delta(i)] \geq 0,264 \cdot (n - i)/n$ gilt. Die erwartete Zunahme der Anzahl der Einsen durch eine Mutation-Selektion-Zyklus wurde für ein Individuum x durch $\Delta_x = \sum_{0 \leq z \leq |x|_0} \Delta_x(z)$ nach unten abgeschätzt (vgl. Seite 45), wobei durch $\Delta_x(z)$ genau die Summanden zusammengefasst und abgeschätzt werden, die auf z -Nullen-Mutationen zurückzuführen sind. Eine Erhöhung der Anzahl der Einsen durch einen Mutation-Selektion-Zyklus um mehr als eins ist nur dann möglich, wenn in dessen Mutationsschritt mindestens zwei Nullen flippen. Für $z \geq 2$ gelte deshalb:

$$\Delta_x^*(z) := P_x(z, 0) \cdot (+1) + \sum_{z < k \leq |x|_1} P_x(z, k) \cdot A_x(z, k) \cdot (z - k)$$

Denn während beim zugrunde liegenden (1+1)-EA für $z \geq 2$ das Auftreten einer z -Nullen-0-Einsen-Mutation die Anzahl der Einsen sicher um z erhöht, kann beim neuen Prozess nur ein Fortschritt von „+1“ mit der entsprechenden Wahrscheinlichkeit verzeichnet werden. Sei $u_i(z)$ für $i = |x|_1$ eine untere Schranke für $\Delta_x^*(z)$. Da bei der Abschätzung der erwarteten Zunahme der Anzahl der Einsen durch einen Schritt $\Delta_x(2), \dots, \Delta_x(|x|_0)$ für alle $x \in \{0, 1\}^n$ jeweils durch Null abgeschätzt wurden (vgl. Lemma 4.2.1 (auf Seite 47)), gilt dann für $0 \leq i \leq n$:

$$E[\delta(i)] \geq 0,264 \cdot (n - i)/n + \sum_{z=2}^{n-i} u_i(z)$$

Könnte also $\Delta_x^*(z)$, genau wie $\Delta_x(z)$, für $z \geq 2$ und $x \in \{0, 1\}^n$ nach unten durch Null abgeschätzt werden (d. h. $u_i(z) := 0$), so würde $E[\delta(i)] \geq 0,264 \cdot (n - i)/n$ folgen. Die Formel für $\Delta_x^*(z)$ lässt sich umformen (vgl. Seite 46), sodass

$$\Delta_{x,p}^*(z) = \binom{|x|_0}{z} \cdot p^z (1-p)^{n-z} \cdot \left(1 + \sum_{z < k \leq |x|_1} \binom{|x|_1}{k} \cdot \left(\frac{p}{1-p} \right)^k \cdot A(z, k) \cdot (z - k) \right)$$

gilt. Die enthaltene Summe, in der über k summiert wird, kann für die beim Standard-(1+1)-EA verwendete Mutationswahrscheinlichkeit $p = 1/n$ mit den Abschätzungen aus dem Beweis zu Lemma 4.2.1 (auf Seite 47) wie folgt abgeschätzt werden:

$$\begin{aligned} \sum_{z < k \leq |x|_1} \binom{|x|_1}{k} \cdot \left(\frac{p}{1-p} \right)^k \cdot (k - z) &\leq \sum_{z < k \leq |x|_1} \frac{(n-1)^k}{k!} \cdot (n-1)^{-k} \cdot (k - z) \\ &\leq \sum_{k > z} \frac{k - z}{k!} \end{aligned}$$

Der Wert der rechten Reihe fällt für $z \geq 2$ streng monoton in z , sodass $\sum_{k > 2} \frac{k-2}{k!} \leq 1$ zu zeigen ausreicht, um $\Delta_x^*(z) \geq 0$ für $z \geq 2$ zu zeigen. Dass $\sum_{k \geq 3} \frac{k-2}{k!} = 3 - e < 1$ gilt, wurde bereits im Beweis zu Lemma 4.2.2 (auf Seite 49) gezeigt. Für den neuen Prozess gilt also tatsächlich, dass im Zustand $i \in \{0, \dots, n\}$ der erwartete Fortschritt $E[\delta(i)]$ ebenfalls mindestens $\delta_E(i) := 0,264 \cdot (n - i)/n$ beträgt und außerdem $\text{Prob}(\delta(i) > 1) = 0$ gilt.

Nun muss „nur noch“ gezeigt werden, dass beim neuen Prozess S_i^P (die erwartete Anzahl Schritte, bis vom Zustand i der Zustand $i+1$ erreicht wird, vgl. oben) nach oben durch $\delta_E(i)^{-1}$ beschränkt ist. Da der erwartete Fortschritt vom aktuellen Zustand abhängt, lässt sich die Wald'sche Identität nicht anwenden, jedoch lässt sich deren Beweis nun anpassen (vgl. [DJW98]). Der Beweis wird durch eine Induktion über die Zustände $i \in \{0, \dots, n-1\}$ geführt.

Für den Induktionsanfang wird der Zustand $i = 0$ betrachtet. Im Zustand 0 ist die Wahrscheinlichkeit, in einem Schritt den Zustand 1 zu erreichen, per Definition von $\delta(0)$ mindestens so groß wie die Wahrscheinlichkeit, dass auf dem Individuum 0^n eine 1-Null-0-Einsen-Mutation stattfindet. Sie ist also durch $n \cdot 1/n \cdot (1 - 1/n)^{n-1} \geq e^{-1}$ nach unten beschränkt, da per Definition des (1+1)-EA $n \geq 3$ gilt. S_0^P ist somit durch e nach oben beschränkt. Behauptet wurde für diesen Fall, dass im Erwartungsfall höchstens $\delta_E(0)^{-1} > 3,7 > e$ Schritte nötig sind. Der Induktionsanfang ist somit gezeigt. Die Induktionsvoraussetzung lautet nun wie folgt: Für $i \in \{0, \dots, k-1\}$ und $0 < k < n$ ist S_i^P nach oben durch $\delta_E(i)^{-1}$ beschränkt. Für den Induktionsschritt ist zu zeigen, dass dann auch $S_k^P \leq \delta(k)^{-1}$ gilt. Im Folgenden wird also der Zustand k genauer betrachtet.

Ein Schritt im Zustand k hat eine Zustandsänderung um d zur Folge, sodass der Zustand $k+d$ erreicht wird. Falls $d \leq 0$ gilt, so reichen im Erwartungsfall $\sum_{k-d \leq i \leq k-1} S_i^P$ Schritte, um erneut

den Zustand k zu erreichen. Da hierbei nach Induktionsvoraussetzung $S_i^P \leq \delta_E(i)^{-1}$ gilt und $\delta_E(i)$ in i monoton abnimmt, ist diese Summe (da $d \leq 0$ gilt) nicht größer als $-d \cdot \delta_E(k)^{-1}$. Nachdem der Zustand k wieder erreicht ist, werden weitere S_k^P Schritte erwartet, um schließlich den Zustand $k + 1$ zu erreichen. Es gilt also:

$$\begin{aligned}
S_k^P &\leq 1 + \text{Prob}(\delta(k) \leq 0) \cdot S_k^P + \sum_{d \leq 0} \text{Prob}(\delta(k) = d) \cdot (-d) \cdot \delta_E(k)^{-1} \\
&= 1 + \text{Prob}(\delta(k) \leq 0) \cdot S_k^P - \delta_E(k)^{-1} \cdot \left(\sum_{d \leq 0} \text{Prob}(\delta(k) = d) \cdot d \right) \\
&= 1 + \text{Prob}(\delta(k) \leq 0) \cdot S_k^P - \delta_E(k)^{-1} \cdot \left(E[\delta(k)] - \sum_{d \geq 1} \text{Prob}(\delta(k) = d) \cdot d \right) \\
&= 1 + \text{Prob}(\delta(k) \leq 0) \cdot S_k^P - \frac{E[\delta(k)]}{\delta_E(k)} + \delta_E(k)^{-1} \cdot \left(\sum_{d \geq 1} \text{Prob}(\delta(k) = d) \cdot d \right) \\
&\leq \text{Prob}(\delta(k) \leq 0) \cdot S_k^P + \delta_E(k)^{-1} \cdot \text{Prob}(\delta(k) = 1)
\end{aligned}$$

Die letzte Abschätzung gilt, da $E[\delta(k)] \geq \delta_E(k) > 0$ und $\text{Prob}(\delta(k) > 1) = 0$ gilt. Ferner gilt

$$S_k^P - \text{Prob}(\delta(k) \leq 0) \cdot S_k^P = (1 - \text{Prob}(\delta(k) \leq 0)) \cdot S_k^P = \text{Prob}(\delta(k) = 1) \cdot S_k^P,$$

sodass insgesamt folgt:

$$\text{Prob}(\delta(k) = 1) \cdot S_k^P \leq \delta_E(k)^{-1} \cdot \text{Prob}(\delta(k) = 1)$$

Für den Zustand $k < n$ gilt offensichtlich $\text{Prob}(\delta(k) = 1) > 0$, sodass $S_k^P \leq \delta_E(k)^{-1}$ schließlich sichergestellt ist. Dies war für den Induktionsschritt zu zeigen. \square

Diese Laufzeit-Schranke stellt eine Verbesserung der bisher bekannten, von Droste, Jansen und Wegener in [DJW98] gezeigten Schranke für die Klasse von linearen Funktionen dar. Im bisherigen Beweis der $O(n \log n)$ -Schranke wird nämlich über ein n , das „groß genug“ ist, argumentiert. Außerdem wird der durch die O -Notation „versteckte“ konstante Faktor von Droste, Jansen und Wegener erst gar nicht quantifiziert – er wäre deutlich größer als 3,8. Ermöglicht wurde diese Verbesserung durch die in dieser Arbeit gewonnene strukturelle Erkenntnis über die Verteilung der Einsen im Individuum, während ein (1+1)-EA eine lineare Funktion optimiert. Mit Hilfe der gezeigten Eigenschaft der Verteilung der Individuen über $\{0, 1\}^n$ (für $n \geq 1$), konnte die von einem Mutation-Selektion-Zyklus erwartete Veränderung der Anzahl der Einsen im Individuum deutlich besser als bisher abgeschätzt werden. Dies ermöglichte schließlich die gezeigte Verbesserung der Laufzeit-Schranke.

4.3 Überlegungen zum dynamischen (1+1)-EA auf linearen Funktionen

Im Folgenden wird über lineare Funktionen mit ausschließlich positiven Koeffizienten argumentiert. Die erwartete Veränderung der Anzahl der Einsen konnte für den Standard-(1+1)-EA deshalb so gut abgeschätzt werden, weil die (in jedem Schritt) verwendete Mutationswahrscheinlichkeit $p = 1/n$ dazu führt, dass in einem Mutationsschritt „höchstwahrscheinlich“ „sehr wenige“ Bits flippen. Die Anzahl der flippenden Bits ist nämlich binomialverteilt zu den Parametern n , der Anzahl der Versuche, und $1/n$, der Erfolgswahrscheinlichkeit eines Versuchs. Von einer Mutation wird folglich genau ein flippendes Bit erwartet. Für Mutationen, die sehr wenige Bits flippen, konnte sehr gut abgeschätzt werden, welchen Einfluss diese auf die erwartete Veränderung der Anzahl der Einsen durch einen Mutation-Selektion-Zyklus haben. So war – und ist natürlich weiterhin – klar, dass Mutationen, die ausschließlich Einsen flippen, immer verworfen werden und somit die Anzahl der Einsen nicht verändern. Außerdem konnte mit der Struktur-Erkenntnis über die Verteilung der Einsen im (jeweils) aktuellen Individuum die Akzeptanzwahrscheinlichkeit einer im Ablauf eines (1+1)-EA auftretenden 1-Null- k -Einsen-Mutationen für $k \geq 2$ durch $1/(k+1)$ nach oben beschränkt werden. Gerade zum Ende der Optimierung einer linearen Funktion mit ausschließlich positiven Koeffizienten, wenn der relative Nullen-Anteil im Individuum sehr klein ist, ist es sehr unwahrscheinlich, dass eine Mutation, die im Erwartungsfall nur ein Bit flippt, mehr als eine Null flippt. Genau die entscheidenden, wahrscheinlich auftretenden Mutationen konnten demnach bezüglich ihres Einflusses auf die erwartete Veränderung der Anzahl der Einsen sehr gut abgeschätzt werden.

In einem Schritt des dynamischen (1+1)-EA ist die Anzahl flippender Bits binomialverteilt zu den Parametern n und p , wobei p die im jeweiligen Mutationsschritt verwendete Mutationswahrscheinlichkeit bezeichnet. Insbesondere ist die Anzahl der flippenden Nullen bzw. die der flippenden Einsen binomialverteilt zu den Parametern $|x|_0$ bzw. $|x|_1$ und p . Von einer Mutation werden $|x|_0 \cdot p$ flippende Nullen und $|x|_1 \cdot p$ flippende Einsen erwartet. Für die Analyse des dynamischen (1+1)-EA auf linearen Funktionen ist es sicherlich wichtig, die Auswirkung der „erwarteten Mutationen“ gut abzuschätzen.

Tritt im Ablauf des dynamischen (1+1)-EA auf einer linearen Funktion mit ausschließlich positiven Koeffizienten eine 1-Null- k -Einsen-Mutation auf, so gilt – unabhängig von der Mutationswahrscheinlichkeit, die zum Auftreten dieser Mutation geführt hat (vgl. Lemma 3.5.15 (auf Seite 41) und für dessen Beweis Lemma 3.5.11 (auf Seite 38)) – natürlich weiterhin, dass diese für $k \geq 2$ höchstens mit einer Wahrscheinlichkeit von $1/(k+1)$ akzeptiert wird. Ebenso gilt weiterhin, dass von einem Mutation-Selektion-Zyklus, in dem die Mutationswahrscheinlichkeit $p = 1/n$ verwendet wird, eine Zunahme der Anzahl der Einsen um mindestens $0,264 \cdot |x|_0/n$

zu erwartet ist, wobei x das in diesem Schritt aktuelle, mutierte Individuum bezeichnet (vgl. Lemma 4.2.2 (auf Seite 49)). Bleiben die restlichen Schritte einer Phase.

Im Resultat für die 1-Null- k -Einsen-Mutationen (Lemma 3.5.15 (auf Seite 41)) wurde ausgenutzt, dass unter diesen nur 01^k -Mutationen potenziell akzeptabel sind und alle übrigen k der $\binom{k+1}{1}$ Mutationstypen sicher inakzeptable Mutationen charakterisieren. Von den $k+1$ möglichen B -Mutationen, $B \in \{0, 1\}^{k+1}$ und $|B|_0 = 1$, die 1-Null- k -Einsen-Mutationen charakterisieren, charakterisiert also nur ein (relativer) Anteil von $1/(k+1)$ potenziell akzeptable Mutationen. Allgemeiner kann für $k > z \geq 1$ gezeigt werden, dass die potenziell akzeptablen B -Mutationen, $B \in \{0, 1\}^n$ und $|B|_0 = z$, unter allen z -Nullen- k -Einsen-Mutationen höchstens einen (relativen) Anteil von $z/(k+1)$ stellen:

Lemma 4.3.1. *Unter allen Bit-Sequenzen mit genau $z \geq 1$ Nullen und genau $k > z$ Einsen beträgt der (relative) Anteil der Sequenzen, für die keine Abbildung existiert, die jeder Null umkehrbar eindeutig eine relativ weiter links stehende Eins zuordnet, höchstens $z/(k+1)$.*

Beweis: Eine Bit-Sequenz habe genau dann die Eigenschaft E, wenn nicht jede darin vorkommende Null umkehrbar eindeutig auf eine relativ weiter links stehende Eins abgebildet werden kann. Es gilt:

$$\frac{\binom{z+k}{z-1}}{\binom{z+k}{z}} = \frac{z}{(z+k) - (z-1)} = \frac{z}{k+1}$$

Es gibt $\binom{z+k}{z}$ Bit-Sequenzen mit genau z Nullen und genau k Einsen. Es reicht somit zu zeigen aus, dass höchstens $\binom{z+k}{z-1}$ dieser Bit-Sequenzen die Eigenschaft E besitzen. Um dies zu zeigen, reicht es wiederum aus, jede solche Bit-Sequenz, die die Eigenschaft E besitzt, umkehrbar eindeutig auf eine Bit-Sequenz der (gleichen) Länge $z+k$ abzubilden, die genau $z-1$ Nullen (und genau $k+1$ Einsen) enthält. Eine solche Abbildung ι wird nun angegeben.

Da $z \geq 1$ vorausgesetzt ist, gilt für jede Bit-Sequenz $m \in \{0, 1\}^{z+k}$, die die Eigenschaft E besitzt, $m = a0b$, sodass a gleich viele Nullen wie Einsen enthält und außerdem eine Bijektion existiert, die jede Null in a auf eine relativ links stehende Eins in a abbildet. Wird in m – von links nach rechts – der ersten Null die erste Eins zugeordnet usw., so ist leicht einzusehen, dass durch die Eigenschaft E das Präfix $a0$ von m eindeutig bestimmt ist. Die von a und b eingerahmte Null ist nämlich genau die Null, der bei obiger Vorgehensweise als erste eine relativ rechts stehende Eins zugeordnet wird, sie ist damit eindeutig „die entscheidende Null“ für die Eigenschaft E.

Die Abbildung $\iota : a0b \rightarrow \bar{a}1b$ (mit $a0b$ wie im vorigen Absatz) wird die geforderte Eigenschaft haben, wobei \bar{a} das (bitweise) Komplement von a bezeichnet. Da \bar{a} , genau wie a , gleich viele Nullen wie Einsen enthält, enthält $\bar{a}1b$ folglich genau eine Null weniger als $a0b$ (nämlich genau $z-1$), wie von ι gefordert wurde. Genau wie in $a0b$ die entscheidende Null, bzw. das Präfix $a0$, eindeutig bestimmt ist, ist auf „komplementäre Weise“ die Eins in $\bar{a}1b$, bzw. das Präfix

$\bar{a}1$, eindeutig bestimmt. Jede Eins in \bar{a} kann (analog der auf a existierenden Bijektion) bijektiv auf eine relativ links stehende Null in \bar{a} abgebildet werden. Aus dem Bild $\bar{a}1b = \iota(a0b)$ kann somit eindeutig das Urbild $a0b$ rekonstruiert werden, womit ι eindeutig umkehrbar ist.

Folglich können höchstens $\binom{z+k}{z-1}$ der $\binom{z+k}{z}$ Bit-Sequenzen die Eigenschaft E besitzen; deren (relativer) Anteil ist also nach oben durch $z/(k+1)$ beschränkt. \square

Analog zum Resultat über die 1-Null- k -Einsen-Mutationen (Lemma 3.5.15 (auf Seite 41)), wäre dann entsprechend zu vermuten, dass der folgende Sachverhalt gilt:

Vermutung 4.3.2. *Für $k > z \geq 1$ gilt: Tritt im Ablauf des [dynamischen] (1+1)-EA auf einer linearen Funktion mit ausschließlich positiven Koeffizienten eine z -Nullen- k -Einsen-Mutation auf, so wird der durch sie entstandene Mutant mit einer Wahrscheinlichkeit nicht größer als $z/(k+1)$ selektiert/akzeptiert.*

„Leider“ ist R^* für $z \geq 2$ und $k > z$ auf der Menge der z -Nullen- k -Einsen-Mutation nur eine partielle Ordnung: Zum Beispiel sind

10 1 01 -Mutation und
01 1 10 -Mutation

unvergleichbar, wobei 10101-Mutationen sicher inakzeptabel und 01110-Mutationen potenziell akzeptabel sind. Um die eine Bit-Sequenz in die andere zu überführen, sind nämlich sowohl „10 \rightarrow 01-Operationen“ als auch „01 \rightarrow 10-Operationen“ nötig, sodass R^* bzw. Lemma 3.5.11 (auf Seite 38) keine Aussage über das Verhältnis der beiden entsprechenden Wahrscheinlichkeiten des Auftretens in einem (bestimmten) Schritt erlaubt. Der Beweis zu Lemma 3.5.15 (auf Seite 41) lässt sich also nicht so einfach anpassen. Dennoch könnte – und sollte meiner Meinung nach – R bzw. die partielle Ordnung R^* auf der Menge der z -Nullen- k -Einsen-Mutationen ausreichen, um den vermuteten Sachverhalt zeigen zu können.

Damit der behauptete Zusammenhang für $z > 1$ gilt – für $z = 1$ ist er ja bewiesen –, ist es nicht (unbedingt) notwendig, dass (in einem bestimmten Schritt) jede sicher inakzeptable Mutation mindestens so wahrscheinlich auftritt wie jede potenziell akzeptable. Für das oben genannte Beispiel kann dem Auftreten einer (potenziell akzeptablen) 01110-Mutation beispielsweise das Auftreten einer (sicher inakzeptablen) 10110-Mutation gegenübergestellt werden. Da die beiden Bit-Sequenzen durch das Flippen der zwei vordersten Bits ineinander überführt werden können, gilt für diese nämlich 01110-Mutation \leq_R 10110-Mutation. Für $z = 2$ und $k = 3$ muss gezeigt werden, dass beim Auftreten einer 2-Nullen-3-Einsen-Mutation diese höchstens mit einer Wahrscheinlichkeit von $\frac{z}{k+1} = \frac{2}{3+1} = \frac{1}{2}$ akzeptiert wird. Lemma 4.3.1 (auf der vorherigen Seite) sagt zumindest aus, dass von den $\binom{2+3}{2} = 10$ möglichen Mutationstypen nur $10 \cdot \frac{z}{k+1} = 10 \cdot \frac{1}{2} = 5$

potenziell akzeptabel sind. In diesem Fall gilt beispielsweise:

$$\begin{aligned}
 01110\text{-Mutation} &\leq_{R^*} 10110\text{-Mutation} \\
 01101\text{-Mutation} &\leq_{R^*} 10101\text{-Mutation} \\
 01011\text{-Mutation} &\leq_{R^*} 11001\text{-Mutation} \\
 00111\text{-Mutation} &\leq_{R^*} 11100\text{-Mutation} \\
 10011\text{-Mutation} &\leq_{R^*} 11010\text{-Mutation}
 \end{aligned}$$

Jedem der fünf möglichen potenziell akzeptablen Mutationstypen (jeweils auf der linken Seite) kann also umkehrbar eindeutig ein mindestens so wahrscheinlich auftretender sicher inakzeptabler Mutationstyp (jeweils auf der rechten Seite) zugeordnet werden. Insgesamt ergibt sich so die Behauptung, dass nämlich eine auftretende 2-Nullen-3-Einsen-Mutation mit einer Wahrscheinlichkeit von höchstens $1/2$ akzeptiert wird.

Bei einem Beweis der obigen Vermutung könnten die folgenden Sachverhalte nützlich sein. Dabei wird wieder davon ausgegangen, dass eine lineare Funktion mit ausschließlich positiven Koeffizienten vorliegt, deren Wertigkeit von links nach rechts monoton abnimmt.

Proposition 4.3.3. *Eine B -Mutation, $B \in \{0, 1\}^+$, ist genau dann sicher inakzeptabel, wenn für B gilt: Links der ersten Null (von links) steht mindestens eine Eins, links der zweiten Null stehen mindestens zwei Einsen usw.*

Proposition 4.3.4. *Wenn eine B -Mutation, $B \in \{0, 1\}^+$ und $|B|_0 = z$, eine potenziell akzeptable z -Nullen-Mutation ist, kann die entscheidende Null (vgl. „entscheidende Null“ im Beweis zu Lemma 4.3.1 (auf Seite 56)) in B nur an den z Positionen $1, 3, \dots, 2z - 1$ (von links) stehen. Steht die entscheidende Null in B an Position $2i - 1$ (von links), so befinden sich genau je $i - 1$ Nullen und Einsen links von dieser und genau $z - i$ Nullen (und genau $k - i + 1$ Einsen) rechts von dieser.*

Proposition 4.3.5. *Ist für $A \in \{0, 1\}^+$ eine A -Mutation sicher inakzeptabel und gilt für $B \in \{0, 1\}^{|A|}$ A -Mutation $\leq_R B$ -Mutation, so ist auch eine B -Mutation sicher inakzeptabel. Gleiches gilt dann auch für „ \leq_{R^*} “ statt „ \leq_R “.*

Unter der Voraussetzung, dass Vermutung 4.3.2 (auf der vorherigen Seite) zutrifft, lässt sich

$$\Delta_{x,p}(z) = P_{x,p}(z, 0) \cdot z + \sum_{z < k \leq |x|_1} P_{x,p}(z, k) \cdot A_{x,p}(z, k) \cdot (z - k)$$

für $z \geq 1$ mit $A_{x,p}(z, k) := z/(k + 1)$ abschätzen (vgl. Seite 45f).

Für $k > z > 0$ gilt dann:

$$A(z, k) \cdot (z - k) = \frac{z}{k+1} \cdot (z - k) = -z \cdot \frac{k - z}{k+1}$$

Für $z = 1$ und $k = 2$ gilt also $-z \cdot \frac{k-z}{k+1} = -\frac{z}{3}$ und außerdem gilt für $z \in \mathbb{N}$:

$$\lim_{k \rightarrow \infty} -z \cdot \frac{k - z}{k + 1} = -z$$

Es könnte also wie folgt abgeschätzt werden:

$$\Delta_{x,p}(z) \geq z \cdot \left(P_{x,p}(z, 0) - \sum_{z < k \leq |x|_1} P_{x,p}(z, k) \right)$$

Die Summe $\sum_{z < k \leq |x|_1} P_{x,p}(z, k)$ ist dabei gleich der Wahrscheinlichkeit, dass eine Mutation mit der Mutationswahrscheinlichkeit p in dem Individuum x genau z Nullen und mehr Einsen als Nullen flippt. Zum Ende der Optimierung – wenn es interessant wird – besitzt das aktuelle Individuum deutlich mehr Einsen als Nullen, d. h. es gilt beispielsweise $|x|_1/|x|_0 = \omega(1)$. Vor allem für große p , d. h. beispielsweise für $p = \omega(1/n)$, nimmt die Wahrscheinlichkeit, dass dann mehr Einsen als Nullen flippen stark zu, sodass die Abschätzung

$$\begin{aligned} \Delta_{x,p}(z) &\geq z \cdot (P_{x,p}(z, 0) - \text{Prob}(z\text{-Nullen-Mutation})) \\ &= z \cdot \left(\binom{|x|_0}{z} p^z (1-p)^{n-z} - \binom{|x|_0}{z} p^z (1-p)^{|x|_0-z} \right) \\ &= z \cdot \binom{|x|_0}{z} \cdot p^z (1-p)^{|x|_0-z} \cdot ((1-p)^{n-|x|_0} - 1) \\ &= z \cdot \text{Prob}(z\text{-Nullen-Mutation}) \cdot ((1-p)^{|x|_1} - 1) \\ &\geq z \cdot \text{Prob}(z\text{-Nullen-Mutation}) \cdot (-1) \end{aligned}$$

eine gute Näherung darstellt. Damit würde dann die folgende Abschätzung gelten:

$$\begin{aligned} \Delta_{x,p} &= \sum_{z=0}^{|x|_0} \Delta_{x,p}(z) \\ &\geq \sum_{z=0}^{|x|_0} (-z) \cdot \text{Prob}(z\text{-Nullen-Mutation}) \\ &= -E[\text{\#flippende Nullen}] \\ &= -p \cdot |x|_0 \end{aligned}$$

Die von einer Mutation des Individuums x erwartete Veränderung der Anzahl der Einsen ist gleich $p \cdot |x|_0 - p \cdot |x|_1 = -p(|x|_1 - |x|_0)$. Für $|x|_1 - |x|_0 > |x|_0 \Leftrightarrow |x|_0 < n/3$ ist also die von einem Mutation-Selektion-Zyklus erwartete Veränderung der Anzahl der Einsen größer, d. h. die erwartete Verringerung kleiner, als die in dessen Mutationsschritt – größer kann sie ja nie sein. Interessant ist dieses Ergebnis für den Fall, dass die Optimierung schon so weit fortgeschritten ist, dass $|x|_0 = o(n)$ gilt. Denn obwohl in einem Mutationsschritt dann eine Veränderung der Anzahl der Einsen von $-p \cdot (n - o(n))$ erwartet wird, ist die von dem zugehörigen Mutation-Selektion-Zyklus erwartete Veränderung der Anzahl der Einsen dennoch durch $-p \cdot |x|_0 = -p \cdot o(n)$ beschränkt.

Dies ist darauf zurückzuführen, dass für $z = o(k)$ bzw. $k = \omega(z)$ die durch $z/(k+1)$ nach oben abgeschätzte Akzeptanzwahrscheinlichkeit einer z -Nullen- k -Einsen-Mutation durch $o(1)$ beschränkt ist. Die Wahrscheinlichkeit, dass eine „ z -Nullen- $\omega(z)$ -Einsen-Mutation“ auftritt, wenn $|x|_0 = o(n)$ gilt, nimmt mit wachsendem p stark zu. Es werden dann ja $p \cdot |x|_0 = p \cdot o(n)$ flippende Nullen und $p \cdot |x|_1 = p \cdot (n - o(n))$ flippenden Einsen erwartet und dass von diesen Erwartungswerten jeweils um einen konstanten Faktor abgewichen wird, ist exponentiell klein im (jeweiligen) Erwartungswert.

Für einen Schritt, in dem $p = \omega(1/n)$ und $|x|_0 = o(n)$ gilt, würde die Wahrscheinlichkeit, mit der der erzeugte Mutant akzeptiert wird, folglich mit $n \rightarrow \infty$ gegen Null streben – wenn Vermutung 4.3.2 (auf Seite 57) zutrifft. Zum Ende der Optimierung einer linearen Funktion mit ausschließlich positiven Koeffizienten, wenn das aktuelle Individuum deutlich mehr Einsen als Nullen enthält, würden dann die Schritte einer Phase, in denen eine große Mutationswahrscheinlichkeit $p = \omega(1/n)$ verwendet wird, für $n \rightarrow \infty$ nicht (mehr) zu einem Verlust von Einsen führen können. Dies bestärkt wiederum die Vermutung, dass die „dynamische Erweiterung“ des (1+1)-EA bei linearen Funktionen zwar nicht hilft, aber auch nicht schadet. Das heißt, sie verursacht lediglich einen „Verlustfaktor“ von $\theta(\log n)$, einen „Verlustfaktor“ in der Größenordnung der Anzahl von Schritten in einer Phase des dynamischen (1+1)-EA.

4.4 Zusammenfassung

In Kapitel 2 wurde für einige „leichte“ lineare Funktionen die bereits vermutete Schranke von $O(n \cdot \log^2 n)$ für die erwartete Laufzeit des dynamischen (1+1)-EA gezeigt. Dazu wurde eine einfache Methode angewendet, die sichere Annahmen über die Anzahl der Einsen (bzw. die der Nullen) im aktuellen Individuum bei einer gegebenen Fitness verwendet. Außerdem wurde gezeigt, dass diese Methode nicht dazu geeignet ist, für jede lineare Funktion die vermutete Laufzeit-Schranke zu zeigen. Dazu wurde beispielhaft die Funktion $\text{BLIC}_n(x) = \sum_{i=1}^n i \cdot x_i$ betrachtet.

Für die Funktion BLIC_n wurde dann in Kapitel 3 gezeigt, dass zum Ende der Optimierung

durch den dynamischen (1+1)-EA – hier „versagte“ die Methode aus Kapitel 2 – Mutation-Selektion-Zyklen, in denen eine große Mutationswahrscheinlichkeit $p = \omega(1/n)$ verwendet wird, die Anzahl der Einsen im aktuellen Individuum mit „überwältigender“ Wahrscheinlichkeit nicht verringern werden. Dies stützte die Vermutung, dass die Dynamik zwar bei linearen Funktionen nicht hilft, aber – viel wichtiger – vor allem nicht schadet.

Um diesen bzw. einen ähnlichen Sachverhalt für jede lineare Funktion zeigen zu können, wurde dann ein Theorem über die Verteilung der Einsen im aktuellen Individuum bei der Maximierung einer jeden linearen Funktion² durch den dynamischen (1+1)-EA bewiesen. Ferner wurde aufgezeigt, wie diese strukturelle Erkenntnis angewendet werden kann, um die Wahrscheinlichkeit für das Auftreten von bestimmten Mutationen in einem Schritt der Optimierung besser als bisher möglich abzuschätzen.

In Kapitel 4 wurden diese Erkenntnisse weiter ausgebaut und gezeigt, dass sie eine gute Abschätzung der erwarteten Veränderung der Anzahl der Einsen durch eine Mutation-Selektion-Zyklus erlauben. Für den Standard-(1+1)-EA konnte sogar eine alternative $O(n \cdot \log n)$ -Laufzeit-Schranke für die Klasse von linearen Funktionen gezeigt werden, die in gewissem Sinn besser ist als die bisherige. Schließlich wurde angedeutet, teilweise gezeigt, wie mit den in dieser Arbeit errungenen Erkenntnissen, eine Verbesserung der bisherigen $O(n^2 \cdot \log n)$ -Laufzeit-Schranke für den dynamischen (1+1)-EA auf linearen Funktionen gezeigt werden könnte. Ein möglicher Weg zu einer $O(n \cdot \log^2 n)$ -Laufzeit-Schranke wurde durch diese Arbeit also „ein wenig“ „geebnet“.

²mit o. B. d. A. ausschließlich positiven Koeffizienten

Anhang A

Abschätzungen und Identitäten

Die verwendeten Schreibweisen und Notationen von Intervallen und Ähnlichem werden in Abschnitt 1.5 (auf Seite 8) erläutert. Die folgende Kollektion von Abschätzungen und Identitäten ist an den Anhang von „Randomized Algorithms“ [MR95] angelehnt.

- Mit $0! := 1$ gilt: $\sum_{k=0}^{\infty} \frac{1}{k!} = e$
- Für $n \in \mathbb{N}$ gilt: $\left(1 - \frac{1}{n}\right)^n \leq \frac{1}{e}$
- Für $n \in \mathbb{N} \setminus \{1\}$ gilt: $\left(1 - \frac{1}{n}\right)^{n-1} \geq \frac{1}{e}$
- Für die n -te harmonische Zahl, $n \in \mathbb{N}$, gilt: $H(n) = \sum_{k=1}^n \frac{1}{k} = \ln n + \Theta(1)$
- Für $n, k \in \mathbb{N}_0$ und $n \geq k$ gilt:

$$\binom{n}{k} \leq \frac{n^k}{k!}$$

$$\binom{n}{k} \approx \frac{n^k}{k!} \quad \text{für } n \gg k$$

$$\binom{n}{k} \leq \left(\frac{n \cdot e}{k}\right)^k$$

$$\binom{n}{k} \geq \left(\frac{n}{k}\right)^k$$

$$\binom{n}{k} = \binom{n}{n-k}$$

Die beiden folgenden wahrscheinlichkeitstheoretischen Lemmata werden in Abschnitt 4.1 von „Randomized Algorithms“ [MR95] bewiesen, wobei die Beweisidee Tschernoff zugeschrieben wird („Tschernoff-Schranke“).

Lemma. Seien X_1, \dots, X_n unabhängige Zufallsvariablen mit dem Wertebereich $\{0, 1\}$, sodass für $i \in \{1, \dots, n\}$ gilt: $p_i = \text{Prob}(X_i = 1) = 1 - \text{Prob}(X_i = 0)$ mit $p_i \in (0; 1)$. Dann gilt für die Zufallsvariable $X = \sum_{i=1}^n X_i$, wobei $\mu = E[X] = \sum_{i=1}^n p_i$ deren Erwartungswert bezeichnet, und $\delta \in \mathbb{R}^+$:

$$\text{Prob}(X > (1 + \delta) \cdot \mu) < \left(\frac{e^\delta}{(1 + \delta)^{(1+\delta)}} \right)^\mu$$

Für die Abschätzung der Wahrscheinlichkeit einer anteiligen Abweichung vom Erwartungswert nach unten gilt der folgende Sachverhalt:

Lemma. Seien X_1, \dots, X_n unabhängige Zufallsvariablen mit dem Wertebereich $\{0, 1\}$, sodass für $i \in \{1, \dots, n\}$ gilt: $p_i = \text{Prob}(X_i = 1) = 1 - \text{Prob}(X_i = 0)$ mit $p_i \in (0; 1)$. Dann gilt für die Zufallsvariable $X = \sum_{i=1}^n X_i$, wobei $\mu = E[X] = \sum_{i=1}^n p_i$ deren Erwartungswert bezeichnet, und $\delta \in (0; 1)$:

$$\text{Prob}(X < (1 - \delta) \cdot \mu) < \left(\frac{e^{-\delta}}{(1 - \delta)^{(1-\delta)}} \right)^\mu < e^{-\mu \cdot \delta^2 / 2}$$

Literaturverzeichnis

- [Bäc92] T. Bäck. The interaction of mutation rate, selection, and self-adaption within a genetic algorithm. In R. Männer und B. Manderick, Hrsg., *Parallel Problem Solving from Nature II*, Seiten 85–94. North Holland, Amsterdam, 1992.
- [Bäc94] T. Bäck. *Evolutionary Algorithms in Theory and Practice*. Dissertation, Universität Dortmund, 1994.
- [BNKF98] W. Banzhaf, P. Nordin, R. Keller und F. Francone. *Genetic Programming*. Morgan Kaufmann Publishers, San Francisco, CA, 1998.
- [DJW98] S. Droste, T. Jansen und I. Wegener. On the analysis of the (1+1) Evolutionary Algorithm. Technischer Bericht, SFB 531 CI-21/98, Universität Dortmund, 1998.
- [DJW99] S. Droste, T. Jansen und I. Wegener. Dynamic parameter control in evolutionary algorithms. Technischer Bericht, SFB 531, Universität Dortmund, 1999.
- [Eri96] M. J. Erickson. *Introduction to Combinatorics*. John Wiley & Sons, 1996.
- [Gol89] D. E. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison Wesley, 1989.
- [JW00a] T. Jansen und I. Wegener. On the analysis of a dynamic evolutionary algorithm. Technischer Bericht, SFB 531, Universität Dortmund, 2000.
- [JW00b] T. Jansen und I. Wegener. On the choice of the mutation probability for the (1+1) EA. In M. Schoenauer, K. Deb, G. Rudolph, X. Yao, E. Lutton, J. J. Merelo und H.-P. Schwefel, Hrsg., *Parallel Problem Solving from Nature VI*, Seiten 89–98. Springer, 2000.
- [Kre00] U. Krengel. *Einführung in die Wahrscheinlichkeitstheorie und Statistik*. Vieweg, Wiesbaden, 5. Auflage, 2000.

- [Müh92] H. Mühlenbein. How genetic algorithms really work: Mutation and hillclimbing. In R. Männer und B. Manderick, Hrsg., *Parallel Problem Solving from Nature II*, Seiten 15–25. North Holland, Amsterdam, 1992.
- [MR95] R. Motwani und P. Raghavan. *Randomized Algorithms*. Cambridge University Press, 1995.
- [Rec94] I. Rechenberg. *Evolutionstrategie '94*. Frommann-Holzboog, Stuttgart, 1994.
- [Rud97] G. Rudolph. *Convergence Properties of Evolutionary Algorithms*. Dissertation, Universität Dortmund, 1997. Verlag Dr. Kovač, Hamburg.
- [Sch95] H.-P. Schwefel. *Evolution and Optimum Seeking*. John Wiley & Sons, 1995.
- [Wit00] C. Witt. Die Analyse des (1+1)-EA für fast lineare und spezielle quadratische Funktionen. Diplomarbeit, Universität Dortmund, 2000.