Konvergenzanalyse für die Partikelschwarmoptimierung[1]

Manuel Schmitt [2]

Abstract: Partikelschwarmoptimierung (PSO) ist eine in der Praxis immer wieder sehr erfolgreich eingesetzte Metaheuristik zum Lösen von Black-Box-Optimierungsproblemen und wird speziell im Fall eines kontinuierlichen Suchraums verwendet. Dazu wird das in der Natur häufig auftretende Schwarmverhalten von miteinander kooperierenden Individuen imitiert. Die Dissertation liefert einen Beitrag zum besseren Verständnis des PSO-Algorithmus, basierend auf einer formalen mathematischen Analyse. Der Fokus liegt dabei auf der Untersuchung des Phänomens der *Konvergenz*. Es ist bekannt, dass die Individuen langfristig gegen einen Punkt im Suchraum streben. In der Dissertation wird detailliert untersucht, welche Eigenschaften dieser Punkt hat. Das Hauptergebnis bildet der formale Beweis, dass die Partikel unter relativ moderaten Voraussetzungen an die zu optimierende Funktion ein lokales Optimum finden, wobei der PSO-Algorithmus im Mehrdimensionalen dafür geringfügig modifiziert werden muss. Im Eindimensionalen wird zusätzlich ein allgemeines Laufzeitresultat bewiesen, nach dem bei der Bearbeitung einer beliebigen unimodalen Funktion die erwartete Laufzeit zur Ermittlung des Optimums mit einem Fehler von höchstens 2^{-k} linear in k ist.

Keywords: Partikelschwarmoptimierung, Konvergenzanalyse, Laufzeit

1 Einführung

Partikelschwarmoptimierung (PSO) wurde 1995 von Kennedy und Eberhart ([KE95, EK95]) entwickelt und zum Lösen von kontinuierlichen Black-Box-Optimierungsproblemen eingesetzt. Dieser Abschnitt liefert einen Überblick über Black-Box-Optimierungsprobleme, den PSO-Algorithmus und die Hauptergebnisse der Dissertation.

1.1 Black-Box-Optimierung

In vielen verschiedenen Gebieten, darunter Mineralogie, medizinische Bildverarbeitung sowie beim Erschließen von Gas- und Ölfeldern, treten Optimierungsprobleme auf, bei denen das (o. B. d. A.) Minimum x^* einer Zielfunktion f über einem Suchraum S gesucht ist. Man spricht von einem *Black-Box-Optimierungsproblem*, wenn die Zielfunktion f nicht explizit angegeben ist und keine Information über ihre Struktur, wie etwa über ihre Ableitungen, vorliegt. Stattdessen kann f nur punktweise ausgewertet werden. Abb. 1 gibt einen Überblick über die beschriebene Situation.

Da insbesondere keine Informationen über den Gradienten der Zielfunktion zur Verfügung stehen, können klassische Ansätze zur Optimierung nicht verwendet werden. Stattdessen

[1] Englischer Titel der Dissertation: “Convergence Analysis for Particle Swarm Optimization”

[2] Friedrich–Alexander–Universität Erlangen–Nürnberg (FAU), Department Informatik, Lehrstuhl für Informatik 12, manuel.schmitt@fau.de

greift man auf heuristische, naturinspirierte Verfahren zurück. Diese werten f wiederholt an verschiedenen Punkten aus und „lernen“ dabei etwas über die Lage der im Hinblick auf die Zielfunktion günstigen Regionen des Suchraums, wodurch die Suche in vielversprechendere Regionen gelenkt werden kann.

Da sich jedes Maximierungsproblem mit Zielfunktion f in ein Minimierungsproblem mit Zielfunktion $-f$ transformieren lässt, werden im Folgenden nur Minimierungsprobleme betrachtet. Ist S ein kontinuierlicher Suchraum, z. B. $S \subset \mathbb{R}^D$, so spricht man von einem *kontinuierlichen Black-Box-Optimierungsproblem*. Für die Behandlung solcher Probleme stellt PSO eine wichtige Lösungsmethode dar.

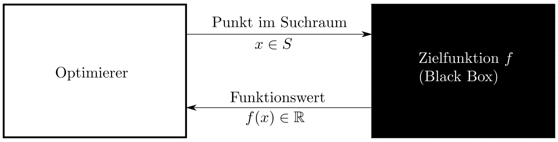


Abb. 1: Black-Box-Optimierung

1.2 Der PSO-Algorithmus

Die erste Version eines Partikelschwarmoptimierungsverfahrens wurde 1995 von Kennedy und Eberhart ([KE95, EK95]) veröffentlicht. Der Algorithmus sollte eine Population von Individuen, zum Beispiel Vogelschwärme oder Fischschwärme, simulieren, die nach einer in Hinblick auf ein bestimmtes Ziel optimalen Region suchen. Das könnte zum Beispiel die Stelle mit dem besten oder dem meisten Futter sein. Im Gegensatz zu anderen bekannten naturinspirierten Metaheuristiken, wie den prominenten Evolutionären Algorithmen, arbeiten die Partikel eines Schwarms zusammen und tauschen miteinander Informationen aus, anstatt gegeneinander zu konkurrieren. Seit der Erfindung der Partikelschwarmoptimierung wurden zahlreiche Varianten entwickelt und experimentell evaluiert. Übersichten dazu findet man unter anderem in der Dissertation ([Sc15]) und in [He10].

Der PSO-Algorithmus lässt sich wie folgt beschreiben. Zu einem gegebenen Optimierungsproblem mit Zielfunktion $f : S \subset \mathbb{R}^D \to \mathbb{R}$ wird eine Population aus N Partikeln eingesetzt. Dieser Schwarm bewegt sich durch den Suchraum S. Zum Zeitpunkt t befindet sich Partikel n an der *Position* X_t^n, die einem Punkt im Suchraum entspricht, und hat eine *Geschwindigkeit* V_t^n, welche durch einen Vektor aus $\mathbb{R}^D$ dargestellt ist. Zusätzlich speichert jedes Partikel die beste Position, also die mit dem niedrigsten Wert für f, die es bisher besucht hat. Diese Position wird *lokaler Attraktor* genannt und mit L_t^n bezeichnet. Der beste lokale Attraktor des gesamten Schwarms heißt *globaler Attraktor*. Dieser spezielle Punkt im Suchraum wird mit G_t bezeichnet und ist allen Partikeln bekannt. Das heißt, dass die Partikel Informationen austauschen können, indem sie für den gesamten Schwarm sichtbar den globale Attraktor aktualisieren. Die eigentliche Bewegung des Schwarms erfolgt durch Anwendung folgender *Bewegungsgleichungen*:

$$\begin{aligned}
V_{t+1}^{n,d} &= \chi \cdot V_t^{n,d} + c_1 \cdot r_t^{n,d} \cdot (L_t^{n,d} - X_t^{n,d}) + c_2 \cdot s_t^{n,d} \cdot (G_t^d - X_t^{n,d}), \\
X_{t+1}^{n,d} &= X_t^{n,d} + V_{t+1}^{n,d},
\end{aligned}$$

wobei t die Iteration, n die Nummer des sich bewegenden Partikels und d die Dimension bezeichnet. Die Konstanten c_1 und c_2 regulieren den Einfluss des persönlichen Gedächtnisses eines Partikels und des gemeinsamen Schwarmgedächtnisses. Durch $r_t^{n,d}$ und $s_t^{n,d}$, die zufällig unabhängig und gleichverteilt aus $[0,1]$ gewählt werden, wird dem Schwarm ein zufälliger Einfluss hinzugefügt. Dieser Bewegungsablauf wird bis zum Erreichen einer festgelegten Abbruchbedingung wiederholt. Abb. 2 gibt einen Überblick über die Bewegung der Partikel. Algorithmus 1 zeigt das PSO-Verfahren in seiner algorithmischen Darstellung. In Experimenten zeigt sich, dass für geeignete Wahl der Parameter der Schwarm nach einer gewissen Zeit gegen einen einzigen Punkt im Suchraum konvergiert. Dies wurde auch unter zusätzlichen Annahmen formal nachgewiesen ([JLY07]), ohne dabei jedoch Aussagen über die Qualität des Punktes zu machen, gegen den der Schwarm konvergiert. Der größte Teil der Dissertation beschäftigt sich mit der Untersuchung der Qualität dieses Grenzwertes im Hinblick auf die Zielfunktion f. Für die Analyse wird in der Dissertation zunächst ein geschlossenes mathematisches Modell beschrieben, innerhalb dessen die Positionen und Geschwindigkeiten als reellwertige stochastische Prozesse betrachtet werden.

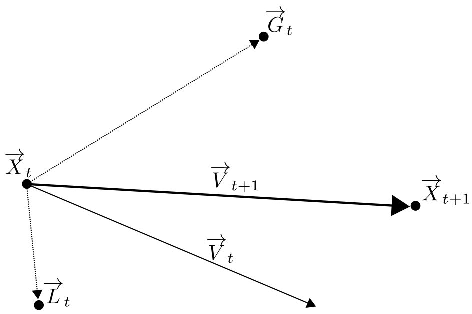


Abb. 2: Bewegung eines Partikels.

Ein wichtiges Hilfsmittel für die Analyse ist die Einführung des sogenannten *Potentials*, das eine Erweiterung der physikalischen Interpretation des Partikelschwarms darstellt. Haben die Partikel eine hohe Geschwindigkeit und befinden sie sich weit weg von ihrem globalen Attraktor, so ist ihr Potential hoch. Wenn der Schwarm andererseits konvergiert, so konvergiert auch das Potential gegen 0. Die folgende Definition beschreibt die beiden am häufigsten verwendeten Formulierungen für ein Maß des Potentials.

Definition 1 (Potential)**.** *Für $a > 0$ ist das Maß $\Phi_t^{n,d}$ des Potentials von Partikel n in Dimension d zum Zeitpunkt t definiert als*

$$Y_t^{n,d} := \sqrt{|V_t^{n,d}|} + \sqrt{|G_t^d - X_t^{n,d}|}.$$

Das Maß für das Potential des gesamten Schwarms in Dimension d zum Zeitpunkt t ist definiert als

$$\Phi_t^d := \sqrt{\sum_{n=1}^{N} \left(a \cdot |V_t^{n,d}| + |G_t^d - X_t^{n,d}| \right)}.$$

Input : Zu minimierende Funktion $f : S \to \mathbb{R}$
Output : $G \in \mathbb{R}^D$
`// Initialisierung`
for $n = 1 \to N$ **do**
  Initialisiere Position $X^n \in \mathbb{R}^D$ zufällig;
  Initialisiere Geschwindigkeit $V^n \in \mathbb{R}^D$;
  Initialisiere lokalen Attraktor $L^n := X^n$;
end
Initialisiere $G := \operatorname{argmin}_{\{L^1,\ldots,L^n\}} f$;
`// Bewegung`
repeat
  for $n = 1 \to N$ **do**
    for $d = 1 \to D$ **do**
      $V^{n,d} := \chi \cdot V^{n,d} + c_1 \cdot \mathrm{rand}() \cdot (L^{n,d} - X^{n,d}) + c_2 \cdot \mathrm{rand}() \cdot (G^d - X^{n,d})$;
      $X^{n,d} := X^{n,d} + V^{n,d}$;
    end
    if $f(X^n) \leq f(L^n)$ **then** $L^n := X^n$;
    if $f(X^n) \leq f(G)$ **then** $G := X^n$;
  end
until *Abbruchbedingung erfüllt*;
return G;

Algorithmus 1 : Klassische PSO.

Die verschiedenen Maße, die auftretenden Quadratwurzeln und der zusätzliche Parameter a werden aus technischen Gründen für die Analyse benötigt.

1.3 Hauptergebnisse

Das ultimative Ziel jedes Optimierungsverfahrens ist es, das globale Optimum einer gegebenen Funktion f zu finden. Allerdings ist dieser Anspruch im Fall eines kontinuierlichen Suchraums im Allgemeinen zu hoch, da sich mit endlich vielen Auswertungen der Zielfunktion nicht einmal entscheiden lässt, ob ein bestimmter Punkt das globale Optimum ist oder nicht. Daher wird in der Dissertation das Ziel verfolgt, zu zeigen, dass PSO zumindest ein lokales Optimum findet, das heißt, dass der Algorithmus einem lokalen Optimum beliebig nahe kommt.

PSO ist für beliebige Zielfunktionen konzipiert, aber um der Analysierbarkeit willen wird der Raum der Zielfunktionen auf die wie folgt beschriebene Menge $\mathbb{F}$ eingeschränkt.

Definition 2. *Sei $f : \mathbb{R}^D \to \mathbb{R}$ eine Funktion. $f \in \mathbb{F}$ genau dann, wenn*

(i) es eine kompakte Menge $K \subset \mathbb{R}^D$ mit positivem Lebesgue-Maß gibt, so dass $P(X_0^n \in K) = 1$ für jedes n und die Menge $I_K = \left\{ y \in \mathbb{R}^D \;\middle|\; f(y) \le \sup_K f \right\}$ beschränkt ist;
(ii) f stetig ist.

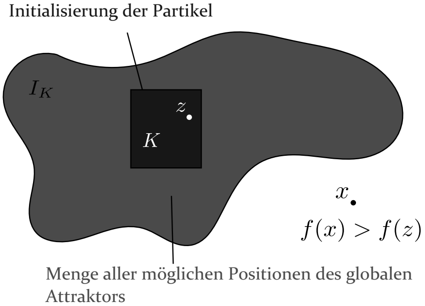


Abb. 3: Zulässige Zielfunktionen.

Gemäß (i) gibt es eine kompakte Menge K, so dass für jedes $x \in K$ nur eine beschränkte Menge I_K von Punkten y mindestens so gut wie x ist. Siehe Abb. 3 für eine Illustration.

Bei einer Initialisierung der Partikel innerhalb von K stellt (i) sicher, dass der globale Attraktor eine gewisse beschränkte Menge in keiner Iteration verlässt und der Schwarm sich nicht „unendlich weit" bewegt. Insbesondere ist damit auch die Existenz eines lokalen Optimums gewährleistet.

Insgesamt sind die Einschränkungen an $\mathbb{F}$ moderat und werden von üblichen Benchmark-Funktionen erfüllt.

Die Hauptergebnisse der Dissertation gelten für Zielfunktionen $f \in \mathbb{F}$ und lauten wie folgt:

- Im 1-dimensionalen findet PSO fast sicher ein lokales Optimum.
- Für unimodale, 1-dimensionale Zielfunktionen liegt die erwartete Laufzeit zur Annäherung an das Optimum bis auf einen Fehler von 2^{-k} in $\mathcal{O}(k)$.
- Im D-dimensionalen findet eine leicht modifizierte PSO für Zielfunktionen aus $\mathbb{F}$, die zusätzlich differenzierbar sind, fast sicher ein lokales Optimum.

In den folgenden Abschnitten werden diese Resultate und insbesondere, was unter „finden" zu verstehen ist, präzisiert und die wesentlichen Beweisideen dargestellt.

2 1-dimensionale PSO - (fast) sicheres schnelles Finden des Optimums

Das erste Hauptergebnis der Dissertation ist der Beweis dafür, dass der globale Attraktor fast sicher[2] ein lokales Optimum findet. Genauer bedeutet das:

Theorem 1. *Falls $f \in \mathbb{F}$, dann gibt es Parameter für die Bewegungsgleichungen des PSO-Algorithmus, so dass fast sicher jeder Häufungspunkt von $(g_t)_{t\in\mathbb{N}}$ ein lokales Minimum von f ist.*

Das heißt, dass jeder Punkt, der kein lokales Optimum ist, eine Umgebung hat, die vom globalen Attraktor nur endlich oft besucht wird. Andernfalls gäbe es einen Häufungspunkt $z \in S$, in dessen Umgebung f streng monoton ist. Das heißt, für jedes $\varepsilon > 0$ würde G_t die ε-Umgebung von z unendlich oft besuchen. Damit dürfte aber kein Partikel einen Punkt x mit $f(x) < f(z)$ treffen, denn da f stetig ist, hat mit z auch eine ganze Umgebung von z kleinere Funktionswerte als x. Da der globale Attraktor nicht durch einen schlechteren Punkt ersetzt werden kann, würde der globale Attraktor in dieser Umgebung von z nie mehr (und damit insgesamt nur endlich oft) liegen.

Zusammengefasst heißt das, dass der Schwarm zwar einerseits jede ε-Umgebung von z unendlich oft besuchen müsste, aber andererseits für hinreichend kleines ε nur die Halbumgebung treffen könnte, in der $f(x) > f(z)$ ist. Dies wird in der Dissertation mit Hilfe umfangreicher Rechnungen widerlegt.

Aus dem Theorem folgt, dass für eine unimodale Zielfunktion das einzige lokale und damit globale Minimum der einzige Häufungspunkt und damit der Grenzwert des globalen Attraktors ist. Gleichzeitig kann das Verhalten der PSO bei der Bearbeitung einer unimodalen Funktion als Modell für ihr Verhalten bei allgemeinen Funktionen gesehen werden, wenn in der Endphase die Konvergenz schon so weit fortgeschritten ist, dass die Partikel sich in der Umgebung eines lokalen Minimums gesammelt haben und die anderen lokalen Minima keinen Einfluss mehr auf den Schwarm haben. Daher erhält die Frage nach der Laufzeit bei der Optimierung einer unimodalen Funktion besondere Aufmerksamkeit.

[2] Gemeint ist hier der wohldefinierte mathematische Begriff, nach dem ein Ereignis *fast sicher* eintritt, wenn es eine Wahrscheinlichkeit von 1 besitzt. Man beachte, dass es im Kontinuierlichen nicht-triviale Ereignisse mit Wahrscheinlichkeit 0 gibt, beispielsweise das in der Literatur gelegentlich erwähnte Ereignis, bei dem die Partikel zufällig alle mit derselben Startposition initialisiert werden.

Die Dissertation führt eine Analysemethode ein, die auf Drifttheorie ([Ha82]) basiert und in dieser Form vor der Dissertation noch nicht verwendet wurde. Dabei wird ein Maß Ψ_t für den „Fortschritt“ des Partikelschwarms beziehungsweise den Abstand zwischen der derzeitigen Konfiguration aller Partikel und Attraktoren und dem angestrebten Zustand eingeführt, in dem alle Positionen gleich dem Optimum sind und alle Geschwindigkeiten 0 betragen. Dieser Zustand kann zwar in endlicher Zeit nicht erreicht werden, aber der Schwarm kann sich ihm beliebig annähern. In der Dissertation wird bewiesen, dass Ψ_t sich erwartungsgemäß permanent verringert und somit gegen 0 konvergiert. Die Konvergenzgeschwindigkeit von Ψ_t gegen 0 erlaubt Schlussfolgerungen auf die Konvergenzgeschwindigkeit des Partikelschwarms.

Aufgrund ihrer hohen Bedeutung für die zukünftige Forschung wird diese Analysetechnik im Folgenden detaillierter beschrieben.

Grundsätzlich eignet sich als Maß für die Qualität des globalen Attraktors G_t der Wert $|A(G_t)|$, wobei $A(z)$ die Menge aller Punkte bezeichnet, deren Funktionswert höchstens so groß wie $f(z)$ ist. Das heißt, die Qualität des globalen Attraktors wird umso höher eingeschätzt, je kleiner die Teilmenge des Suchraums ist, die ebensogute oder bessere Punkte enthält. Da f unimodal ist, ist diese Teilmenge stets ein Interval, das das Optimum enthält.

Es genügt jedoch nicht, lediglich die Qualität des globalen Attraktors (oder auch aller Attraktoren) zu messen, denn es können pathologische Situationen auftreten, in denen eine unmittelbare, signifikante Verbesserung des globalen Attraktors während der nächsten Iterationen sehr unwahrscheinlich oder sogar unmöglich ist. Wenn beispielsweise das Potential des Schwarms erheblich zu groß ist, so suchen die Partikel ein Areal ab, von dem $A(G_t)$ nur einen (beliebig) kleinen Teil ausmacht und das folglich mit sehr geringer Wahrscheinlichkeit besucht wird. In diesem Fall muss eine Verringerung des Potentials als eine Verbesserung der Konfiguration angesehen werden, auch wenn alle Attraktoren unverändert bleiben.

Andererseits könnte das Potential auch um Größenordnungen zu klein sein, so dass die Partikel in einem Schritt nur einen sehr kleiner Bruchteil der Strecke bis zum Optimum zurücklegen können. In diesem Fall könnten qualitative Verbesserungen des globalen Attraktors zwar stattfinden, wären aber quantitativ aufgrund des zu kleinen Abstandes zwischen altem und neuem globalen Attraktor insignifikant.

Diese Betrachtungen führen zu folgender Definition von Ψ_t.

Definition 3 (Optimalitätsmaß). *Für $C_\Psi > 0$, wird das Optimalitätsmaß Ψ_t wie folgt definiert:*

$$\Psi_t := C_\Psi \cdot \Psi_t^{(0)} + \Psi_t^H + \Psi_t^L \; mit \; \Psi_t^{(0)} := \sum_{n=1}^{N} \sqrt{|A(L_t^n)|}, \Psi_t^H := \sum_{n=1}^{N} Y_t^n \; und \; \Psi_t^L := \frac{|A(G_t)|}{\Phi_{t+1}}.$$

Die auftretende Quadratwurzel wird für die technischen Aspekte des Beweises benötigt. $\Psi_t^{(0)}$ ist das Grundmaß und steht unmittelbar mit der Qualität der Attraktoren in Verbindung. Ψ_t^H und Ψ_t^L werden in der Dissertation als Sekundärmaße bezeichnet. Die Sekundärmaße sind mit bestimmten Konfigurationen des Partikelschwarms verknüpft. Ist das Potential deutlich zu hoch, so macht Ψ_t^H den wesentlichen Anteil von Ψ_t aus und eine Verringerung des Potentials verbessert Ψ_t deutlich. Ist andererseits das Potential erheblich zu niedrig, so übertrifft Ψ_t^L die anderen Summanden innerhalb von Ψ_t bei weitem. Durch hinreichend große Wahl von C_Ψ kann sichergestellt werden, dass sich bei einem Potential, dessen Größenordnung für die Optimierung günstig ist, die Verbesserungen von $C_\Psi \cdot \Psi_t^{(0)}$ signifikant auf Ψ_t auswirken.

Der Hauptteil des Beweises besteht im Nachweis folgender Eigenschaften.

1. Der Partikelschwarm wird so initialisiert, dass $\mathrm{E}[\Psi_0] < \infty$.
2. Ist das Potential deutlich zu hoch, so verringert sich in den folgenden Iterationen Ψ_t^H erwartungsgemäß um einen konstanten Faktor. Gleichzeitig sind die Verschlechterung von Ψ_t^L von oben und der Anteil von Ψ_t^H an Ψ_t von unten derart beschränkt, dass die erwartete Verbesserung von Ψ_t^H auch zu einer erwarteten Verbesserung von Ψ_t um einen konstanten Faktor führt.
3. Dasselbe gilt mit vertauschten Rollen von Ψ_t^H und Ψ_t^L, falls das Potential deutlich zu niedrig ist.
4. Ist das Potential weder deutlich zu hoch noch deutlich zu niedrig, so verringert sich $\Psi_t^{(0)}$ erwartungsgemäß um einen konstanten Faktor. Gleichzeitig sind die Verschlechterungen von Ψ_t^H und Ψ_t^L von oben und die erwartete Verbesserung von $\Psi_t^{(0)}$ von unten so beschränkt, dass Ψ_t sich während der folgenden Iterationen erwartungsgemäß um einen konstanten Faktor verbessert.

Der Nachweis dieser Eigenschaften führt zur Verifikation einer Driftbedingung an den Prozess $(\Psi_t)_{t\in\mathbb{N}}$ im Sinne von [Ha82]. Ein im Rahmen der Dissertation für die vorliegenden Erfordernisse angepasstes Drifttheorem ermöglicht es, aus den nachgewiesenen Schranken an die Entwicklung des Erwartungswertes von Ψ_t Schlussfolgerungen auf die erwartete Zeit zu ziehen, die Ψ_t benötigt, um einen festen Wert zu unterschreiten, und liefert schließlich das zweite Hauptergebnis.

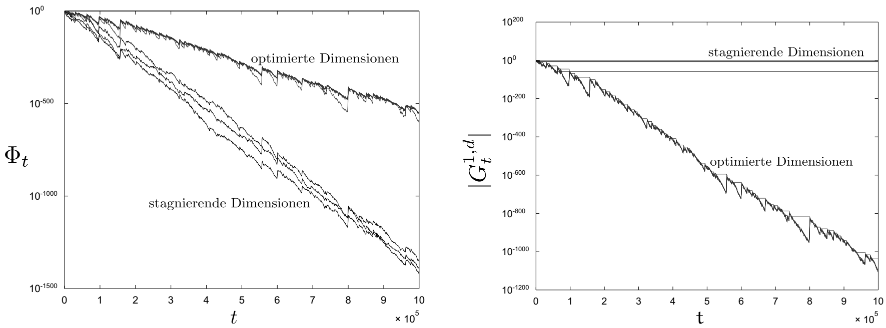
Theorem 2. *Sei b der Durchmesser des Suchraums. Wenn die Partikel unabhängig und uniform über den Suchraum initialisiert werden, und die Initialgeschwindigkeit einen endlichen Erwartungswert hat, dann gilt für $\tau := \min\{t \geq 0 \mid \Psi_t \leq 2^{-k}\}$:*

$$\mathrm{E}[\tau] \in \mathcal{O}(k + \log(b+1)).$$

Daraus folgt, dass die (zufällige) Zeit zur Annäherung des globalen Attraktors an das Optimum x^* bis auf einen Fehler von 2^{-k} nur linear von k abhängt. Man spricht von linearer Konvergenzordnung. Es ist bekannt, dass kein Verfahren existiert, dass für eine unimodale Funktion eine schnellere Konvergenzordnung besitzt. Lediglich die Konstante innerhalb der $\mathcal{O}$-Notation kann verbessert werden.

3 Mehrdimensionale PSO - Modifikation des Algorithmus

Es stellt sich die Frage, ob vergleichbare Resultate auch im allgemeinen, D-dimensionalen Fall bewiesen werden können. Jedoch stellt sich heraus, dass nun eine weitere Art von pathologischen Konfigurationen auftritt, die bei der Optimierung 1-dimensionaler Zielfunktionen nicht vorgekommen ist. Um dies zu verdeutlichen, zeigt Abb. 4 den Verlauf des Potentials Φ_t und des globalen Attraktors G_t bei der Bearbeitung der 10-dimensionalen Funktion $f(\vec{x}) = \sum_{d=1}^{10}(x_d)^2$, der sogenannten SPHERE-Funktion, durch einen Schwarm aus 3 Partikeln.



(a) Verlauf des Potentials. (b) Verlauf des globalen Attraktors.

Abb. 4: Partikelschwarm mit unbalancierten Potentialen beim Bearbeiten der 10-dimensionalen Zielfunktion SPHERE mit $N = 3$ Partikeln. Dimensionen 2, 4 und 10 stagnieren und werden nicht optimiert.

Man erkennt eine klare Separation der Dimensionen. Während 7 Dimensionen ordnungsgemäß optimiert werden und die entsprechenden Einträge sowohl im globalen Attraktor als auch im Potential gegen 0 streben, gibt es drei andere Dimensionen, deren zugehörige Einträge ein vollkommen anderes und unerwünschtes Verhalten zeigen. Hier konvergiert das Potential erheblich schneller gegen 0 als in den 7 optimierten Dimensionen. Gleichzeitig erfolgt keine erkennbare Verbesserung der entsprechenden Einträge des globalen Attraktors. Daher werden diese Dimensionen auch als stagnierende Dimensionen bezeichnet.

Die pathologische Konfiguration, in der das Potential unbalanciert ist und in einigen Dimensionen eine geeignete Größenordnung hat, in anderen aber deutlich zu gering ist, ist offenbar fatal und kann zum Konvergieren gegen einen Punkt führen, der kein lokales Optimum ist. Während sich der Schwarm von den anderen pathologischen Konfigurationen selbstständig heilen konnte, verstärkt sich hier das Ungleichgewicht zwischen den Potentialen in den verschiedenen Dimensionen fortwährend.

Weitere Experimente legen nahe, dass sich diesem Phänomen durch eine Erhöhung der Schwarmgröße beikommen lässt. Allerdings ist die Schwarmgröße, die für die Konvergenz gegen das Optimum benötigt wird, sowohl von der Dimension als auch von der Zielfunktion abhängig. In einer reinen Black-Box-Problemstellung ist also nicht klar, wie groß die

Schwarmgröße gewählt werden muss. Damit kann ein starkes Konvergenzresultat wie im 1-dimensionalen Fall nicht nachgewiesen werden.

In der Dissertation wird daher der PSO-Algorithmus mit dem Ziel modifiziert, auch diese pathologische Konfiguration überwindbar zu machen, gleichzeitig aber das Verhalten des Schwarms und damit seine Stärken möglichst wenig zu verändern. Das führt zu folgendem, modifiziertem PSO-Algorithmus.

Definition 4 (Modifizierte PSO). *Für beliebig kleines, aber festes $\delta > 0$ wird die modifizierte PSO durch dieselben Bewegungsgleichungen wie die klassische PSO (siehe Abschnitt 1.2) definiert, wobei lediglich das Geschwindigkeitsupdate wie folgt geändert wird.*

$$V_{t+1}^{n,d} = \begin{cases} (2 \cdot r_t^{n,d} - 1) \cdot \delta, \\ \textit{falls } \forall n' \in \{1,...,N\} : |V_t^{n',d}| + |G_{t+1}^{n',d} - X_t^{n',d}| < \delta, \\ \\ \chi \cdot V_t^{n,d} + c_1 \cdot r_t^{n,d} \cdot (L_t^{n,d} - X_t^{n,d}) + c_2 \cdot s_t^{n,d} \cdot (G_{t+1}^{n,d} - X_t^{n,d}), \\ \textit{sonst.} \end{cases}$$

In Worten: Sobald in einer Dimension die Summe aus Geschwindigkeit und Abstand zwischen Position und globalem Attraktor für jedes Partikel unterhalb einer festen Grenze δ liegt, wird die neue Geschwindigkeit zufällig gleichverteilt aus dem Intervall $[-\delta, \delta]$ ermittelt. Dadurch wird das oben angesprochene Phänomen vermieden, denn nun ist das Potential stochastisch nach unten beschränkt. Dadurch wird allerdings auch die Möglichkeit einer Konvergenz des Schwarms geopfert, lediglich die Attraktoren können noch konvergieren. Experimente legen jedoch den Schluss nahe, dass die Modifikation nur sehr selten zur Anwendung kommt, solange der globale Attraktor noch weiter als δ vom Optimum entfernt ist. Folglich bleibt das Schwarmverhalten erhalten, die Modifikation führt nur zur Vermeidung vorzeitig stagnierender Dimensionen, ohne die gesamte Schwarmbewegung zu bestimmen.

Mit ähnlicher Vorgehensweise wie im 1-dimensionalen Fall wird in der Dissertation nun folgendes Resultat bewiesen.

Theorem 3. *Falls $f \in \mathbb{F}$ und zusätzlich stetig differenzierbar, dann gibt es Parameter für die Bewegungsgleichungen des modifizierten PSO-Algorithmus, so dass fast sicher jeder Häufungspunkt von $(g_t)_{t\in\mathbb{N}}$ ein lokales Minimum von f ist.*

4 Zusammenfassung und Ausblick

Die Dissertation erweitert das Verständnis der Partikelschwarmoptimierung, indem sie Einblicke liefert, unter welchen Bedingungen die Partikel ein lokales Optimum finden. Sie beweist, dass Partikelschwarmoptimierung bei der Bearbeitung einer 1-dimensionalen Zielfunktion fast sicher ein lokales Optimum findet. Außerdem wird gezeigt, dass die Konvergenzordnung linear ist, ein im Bereich der Black-Box-Optimierung bis auf einen konstanten Faktor optimales Resultat. Schließlich wird eine Modifikation des Algorithmus vorgestellt, die es erlaubt, dass der Schwarm auch im allgemeinen D-dimensionalen Fall fast sicher ein lokales Optimum findet.

In der Literatur werden häufig zur Vereinfachung von Beweisen unbewiesene, aber entweder durch Experimente oder durch Anschauung plausible Annahmen gemacht und verwendet. Auch ist es üblich, eine Analyse auf einzelne, konkrete Zielfunktionen zu beschränken. In dieser Dissertation wurde jedoch keine derartige Vereinfachung benötigt.

Insbesondere die Anwendbarkeit der Beweistechnik, die der Laufzeitanalyse zugrunde liegt, muss nicht auf die hier vorliegende Analyse beschränkt bleiben.

Literaturverzeichnis

[EK95] Eberhart, Russell C.; Kennedy, James: A New Optimizer Using Particle Swarm Theory. In: Proceedings of the 6th International Symposium on Micro Machine and Human Science. S. 39–43, 1995. doi:10.1109/MHS.1995.494215 .

[Ha82] Hajek, Bruce: Hitting-Time and Occupation-Time Bounds Implied by Drift Analysis with Applications. Advances in Applied Probability, 14(3):502–525, 1982.

[He10] Helwig, Sabine: Particle Swarms for Constrained Optimization. Dissertation, Department of Computer Science, University of Erlangen-Nuremberg, 2010. urn:nbn:de:bvb:29-opus-19334.

[JLY07] Jiang, Ming; Luo, Yupin P.; Yang, Shiyuan Y.: Particle Swarm Optimization – Stochastic Trajectory Analysis and Parameter Selection. In (Chan, Felix T. S.; Tiwari, Manoj Kumar, Hrsg.): Swarm Intelligence – Focus on Ant and Particle Swarm Optimization, S. 179–198. I-TECH Education and Publishing, 2007.

[KE95] Kennedy, James; Eberhart, Russell C.: Particle Swarm Optimization. In: Proceedings of the IEEE International Conference on Neural Networks. Jgg. 4, S. 1942–1948, 1995. doi:10.1109/ICNN.1995.488968 .

[Sc15] Schmitt, Berthold Immanuel: Convergence Analysis for Particle Swarm Optimization. Dissertation, Department of Computer Science, University of Erlangen-Nuremberg, 2015. urn:nbn:de:bvb:29-opus4-61621.

[SW13] Schmitt, Manuel; Wanka, Rolf: Particle Swarm Optimization Almost Surely Finds Local Optima. In: Proc. 15th Genetic and Evolutionary Computation Conference (GECCO). S. 1629–1636, 2013. doi:10.1145/2463372.2463563.

Manuel Schmitt hat Mathematik und Informatik an der Friedrich-Alexander-Universität Erlangen studiert und 2012 mit dem Diplom in Mathematik und der ersten Staatsprüfung für das Lehramt in Mathematik und Informatik, jeweils mit Auszeichnung, abgeschlossen. Nach dem Studium war er dort als wissenschaftlicher Mitarbeiter tätig, wo er im Bereich der theoretischen Informatik forschte und lehrte. Im Rahmen dieser Tätigkeit entstand unter anderem gemeinsam mit Rolf Wanka die Veröffentlichung [SW13], wofür den Autoren der Best Paper Award verliehen wurde.