Kapitel 9

Optimierung

9.1 Motivation

Optimierungsprobleme treten nicht nur in der Physik an vielen Stellen auf. Die sogenannte lineare Optimierung (Suche der optimalen Lösung eines Problems, dessen Nebenbedingungen durch eine grosse Anzahl linearer Ungleichungen gegeben sind) hat lange in den Wirtschaftswissenschaften und deren Teilgebiet *Operations Research* eine herausragende Rolle gespielt. Man kann hier zeigen, daß es ausreicht, die Ecken des durch die Nebenbedingungen in einem hochdimensionalen Raum gebildeten erlaubten Gebietes (Simplex) systematisch nach immer besseren Lösungen abzusuchen.

Eine ähnlich grundlegende Rolle für nichtlineare Optimierungsprobleme spielte das *traveling salesman* Problem, bei dem die insgesamt kürzeste Verbindung zwischen fest vorgegebenen Orten gesucht wird. Dieses letzte Problem ist eine "schwierige" Optimierungsaufgabe in dem Sinne als daß kein Algortihmus polynomialer Ausführungszeit in der Anzahl der Städte bekannt ist, der mit Sicherheit die optimale Lösung liefert.

Wer kennt darüberhinaus nicht die einfacheren Optimierungsfragen vom Typus *Gegeben sei die Form einer Konservendose, wie läßt sich dann bei gegebenem Inhalt die Oberfäche minimieren?*, die sich durch eine Kurvendiskussion einer Funktion einer Variablen bestimmen lassen. Relativ wenige Variablen kennzeichen auch Ausgleichsprobleme der Form $\sum_i |y_i - f(x_i)|^2 \stackrel{!}{=} \min$, bei denen eine vorgegebene Datengesamtheit (x_i, y_i) möglichst genau durch einen durch Fitparameter vorgegebenen, ansonsten analytischen Funktionsverlauf $f(x)$ wiedergegeben werden soll.

Im folgenden geben wir zur weiteren Motivation eine Auflistung der Bereiche, in denen in der Physik weitere Optmierungsprobleme auftreten:

- Suche nach dem quantenmechanischen Grundzustand (Wellenfunktion mit minimaler Energie) eines Systems: Variationsrechnung.
- Bestimmung der Dynamik eines Systems aus Extremalprinzipien:
  - Lagrange-Funktion(al) (Klassische Mechanik),
  - Fermatsches Prinzip (Strahlenoptik: Lichtlaufzeit ist extremal),
  - Thermodynamische Gleichgewichtszustände sind durch extremale freie Energien nach Gibbs oder Helmholtz gekennzeichnet, Nichtgleichgewichtsvorgänge zeichnen sich häufig durch maximal Entropieproduktion aus, usw.

Die aufgelisteten Probleme unterscheiden sich im wesentlichen durch die Anzahl der auftretenden Freiheitsgrade. Während das Konservendosenproblem und die Kurvenanpassung nur die Betrachtung einer kleinen Anzahl von Parametern erfordern, für die "fertige" Algorithmen existieren (siehe z.B. im bereits mehrfach erwähnten *Numerical Recipes*), so ist die Bestimmung eines quantenmechanischen Grundzustandes ein Problem mit vielen Parametern, denn die unbekannte Wellenfunktion nimmt ja prinzipiell unendlich viele Werte (einen an jedem Raumpunkt) an. Um die Probleme zu verdeutlichen, die bei derart hochdimensionalen Systemen eine Rolle spielen, wollen wir zunächst ein Standardmodell der statistischen Mechanik betrachten.

9.2 Das Ising-Modell und Verwandte

Der Versuch, das Phasenverhalten eines Systemes diskreter, klassischer, magnetischer "Spins" zu beschreiben, wurde 1924 durch Ernst Ising (geb. 10.05.1900) in seiner Doktorarbeit unternommen.

Stellen wir uns magnetische Domänen in einem dünnen Materialfilm vor und nehmen wir weiterhin an, daß die Orientierung der Domänen entweder aus der Ebene heraus oder in sie hinein gerichtet ist. Wir wollen uns diese Domänen etwa wie die zwei Spinorientierungen eines Spin-1/2 quantenmechanischen Teilchens vorstellen. Jeder "Spin" s_i soll durch einen Gitterplatz i in einem 2D Gitter und durch einen Wert von $+1$ oder -1 entsprechend "up" oder "down" repräsentiert sein.

Die Energie eines solchen Systems ist durch die Einstellung der Spins festgelegt. Im klassischen Fall (und näherungsweise meist auch im quantenmechanischen) wechselwirken Spins paarweise miteinander, zum Beispiel durch magnetische Dipol-Dipol-Wechselwirkungen, die das magnetische Moment eines Spins s_j an einem anderen Ort i als ein Feld einer Stärke $J_{ij}s_j$ zur Wirkung kommen läßt. Durch Ausrichten in Feldrichtung des Spins am Ort i kann die Energie des Systemes um den Wert $-J_{ij}s_is_j$ abnehmen. Positive Kopplungen J_ij bevorzugen energetisch gleichausgerichtete Spins (ferromagnetische Ordnung), negative können dagegen eine antiferromagnetische Ordnung (abwechselnd Spin "up" und "down") bewirken. In magnetischen Systemen können beide Fälle auftreten, da die Spins häufig durch quantenmechanische Korrelationen (Pauli-Prinzip, Austauschwechselwirkung) gekoppelt sind.

Wir wollen wie oben angedeutet annehmen, daß sich die Energiefunktion $H\{s_i\}$ des Systemes schreiben läßt als die Summe aller Paar-Spin-Energien

$$H\{s_i\} = -\frac{1}{2} \sum_{i \neq j} J_{ij} s_i s_j, \quad s_i = \pm 1. \tag{9.1}$$

Je nach der Wahl der J_{ij} resultiert ein sehr unterschiedliches Systemverhalten.

In Isings ursprünglichem Modell gilt

$$J_{ij} = \begin{cases} 1 & \text{für direkte Nachbarn} \\ 0 & \text{sonst} \end{cases} \tag{9.2}$$

Der Grundzustand ist der Zustand minimaler Energie H. Dieser liefert bei $T \to 0$ in der *Zustandssumme* $Z = \sum_{\{s_i\}} \exp(-\beta H\{s_i\})$ den dominanten Beitrag. Hier haben wir mit β die inverse Temperatur $\beta = 1/kT$ bezeichnet und die Summation über die Menge $\{s_i\}$ aller möglichen 2^N Spinkonfigurationen asugeführt, wobei N die Zahl der Gitterplätze ist. Es gibt zwei solcher Grundzustände, in denen jeweils ausnahmslos alle Spins nach oben bzw. unten zeigen.

Nebenbemerkung: In der Folgezeit zeigten Experimente, daß die Eigenschaften seines Modelles in der Nähe des Para-/Ferromagnetüberganges in 2D geeigneter war, um Flüssigkeitsverhalten in der Nähe eines kritischen Punktes zu beschreiben als die Magnetisierungeigenschaften eines Magneten. Trotzdem ist dieses Modell eines der wichtigsten in der statistischen Mechanik, weil es sich aufgrund seiner Einfachheit häufig noch als Ausgangspunkt analytischer Betrachtungen eignet. Für Isings ursprüngliches Modell gelang es Lars Onsager, die Zustandssumme und damit alle thermodynamischen Eigenschaften in 2D analytisch zu bestimmen.

9.2.1 Spingläser

Sherrington und Kirkpatrick haben Spin*gläser* betrachtet, indem sie die Kopplung J_{ij} als eine Zufallszahl mit

$$J_{ij} \in [-1, 1] \quad \forall\, i, j \tag{9.3}$$

gewählt haben. Das besondere bei diesem Spinmodell ist eine mögliche Entartung des Grundzustandes. Das heißt, daß viele (bei unendlich großen Systemen auch unendlich viele) sehr unterschiedliche Zustände mit großen Abständen im Konfigurationsraum aber kleinen (bzw. gar keinen) Energieunterschieden auftreten. Ein sehr vereinfachtes Spinglas ist in der Abbildung 9.1 gezeigt,

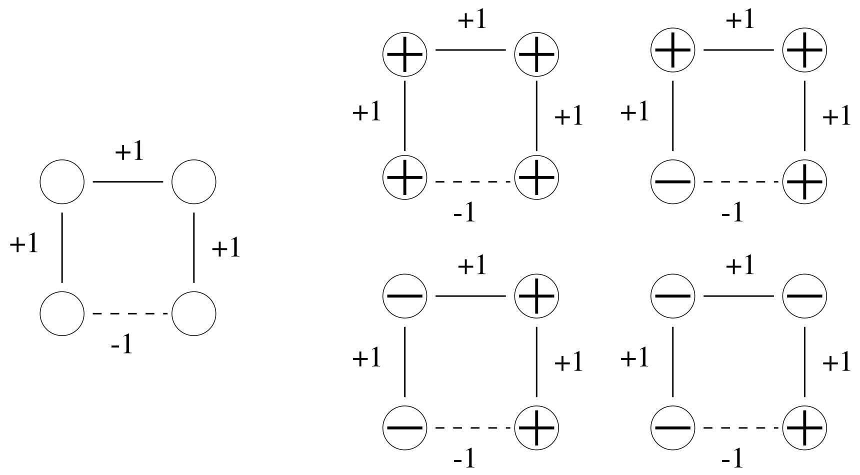

Abbildung 9.1: Spingläser: links ist eine mögliche Anordnung von Kopplungsstärken J_{ij} gezeigt, die nur nächste Nachbarn verbindet. Rechts sind die Hälfte aller möglichen Zustände geringster Energie gezeichnet. Die Konstruktion geschah dadurch (oben links), daß in der rechten unteren Ecke mit einem *up*-Spin begonnen wurde und dann zunächst entgegen dem Uhrzeigersinn soviele Bindungen wie möglich befriedigt, d.h. $J_{ij}s_is_j > 0$ gewählt wurde. Dies scheitert an der unteren Kopplung, die sich nicht mehr befriedigen läßt. Danach wurde einfach die Lage der nicht befriedigten Bindung variiert, was zu vier energetisch gleichwertigen Zuständen führt. Zu jeder gezeigten Konfiguration existiert auch noch diejenige mit invertierten Spins, die die gleiche Energie aufweist. In diesem Beispiel gibt es also bereits 8 Zustände gleicher (minimaler) Energie.

In der Hochtemperaturphase zeigen Spingläsern wie auch Ferromagnete keine spontane Magnetisierung, d.h. keine Ausrichtung der Spins überwiegend in eine Richtung. Sie weisen den üblichen Diamagnetismus auf, ggf. überlagert von paramagnetischen (das üßere Magnetfeld verstärkende) Effekten. In der Tieftemperaturphase zeigen Spingläser im Unterschied zum einfachen Isingmodel (und zu Ferromagneten) jedoch keine spontane Magnetisierung sondern "frieren" in einem der möglichen energetisch günstigen Zustände ein. Ein Übergang in einen anderen, energetisch ebenso günstigen Nebenzustand wird häufig durch Energiebarrieren kinetisch unterdrückt, so wie das auch in amorphen Phasen von Mischungen von SiO_2 mit ionischen Beimischungen auftritt, die gemeinhin als *Gläser* bezeichnet werden.

Wir haben uns hier den Ising-ähnlichen Modellen so ausführlich gewidmet, weil sie in der statischen Mechanik sehr wichtig sind und daher auch Gegenstand einer Übung sein werden.

9.2.2 Das Handelsreisendenproblem

Ein weiteres, vielleicht anschaulicheres Beispiel ist das sogenannte *traveling salesman* Problem (Handelsreisendenproblem). Die Aufgabe besteht darin, daß ein Händler auf einer

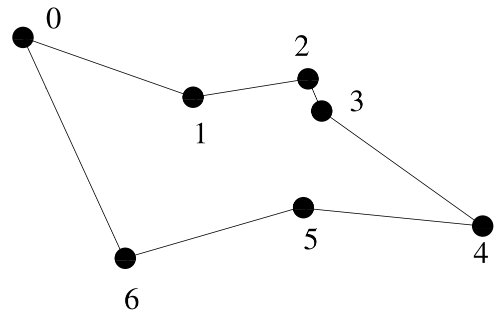

Abbildung 9.2: Handelsreisenden-Problem: Welches ist der kürzeste Weg einer Rundreise, die alle gezeigten Punkte genau einmal aufsucht?

Rundreise verschiedene verstreut liegende Städte besuchen muß und dabei den kürzesten Weg finden will. Die praktische Bedeutung dieser Fragestellung für logistische Probleme liegt auf der Hand. Etwas abstrakter formuliert suchen wir wie in den obigen Fällen wieder das Minimum einer reellwertigen Funktion, die hier allerdings nicht die Energie, sondern die Länge einer Rundreise ist. Wie im Beipiel mit den magnetischen Spins ist der Zustandsraum diskret: er enthält alle $N-1$-tupel mit Permutationen der Zahlen $(1,\ldots,N)$, hat also $(N-1)!$ Elemente. Hier zählen wir ab 1, weil wir o.B.d.A. den Anfangsort der Reise nach 0 verlegt haben.

Wir wollen die Eigenschaften zusammenfassen, die uns bei den beiden geschilderten Problemen begegnet sind und die wir bei der Ermittlung eines "energetisch günstigsten" oder anderweitig optimalen Zustandes berücksichtigen müssen:

- man arbeitet in einem extrem hochdimensionalen Suchraum. Wenn wir in den obigen Beispielen einfach alle Zustände absuchten, so müßten alle 2^N Möglichkeiten für Spinsysteme mit Spins von 2 Zuständen oder $(N-1)!$ im Handelsreisendenproblem berücksichtigt werden, einen Zustandsvektor mit N Elementen (den Positionen der Spins/Städte) zu füllen.

- die gesuchten Minima sind u.U. sehr zahlreich und liegen energetisch nahe beisammen. Selbst wenn ein eindeutiger energetisch günstigster Zustand existiert, so existieren möglicherweise viele Nebenminima fast gleicher Energie, die ggfs. wichtige physikalische Information tragen.

- die Energiefunktion ist nichtlinear. Eine quadratische Energiefunktion ist zwar im Prinzip für eine Optimierung geeignet (siehe auch das Beipiel der Bestimmung der Gewichte in mehrschichtigen Netzwerken im folgenden Kapitel) und führt in der Regel auf ein Problem aus der linearen Algebra, jedoch sind in unserem Falle zudem noch

- die Zustände diskret, so daß Methoden der Differentialrechung (Gradientenabstieg) nicht anwendbar sind.

Die einfachste Technik zur Lösung des Problems ist natürlich das Durchprobieren aller möglichen N-tupel. Allerdings wächst die Anzahl dieser N-tupel wie $N!$ und führt daher schnell zu astronomisch hohen Rechenzeitbedürfnissen für eine solche Strategie. Daher benötigen wir einen intelligenteren Algorithmus. In Ermangelung einer allgemeinen Lösungsvorschrift, die in kürzerer Zeit zum Erfolg führt, wählen wir an dieser Stelle sogenannte *heuristische* Verfahren, die im wesentlichen auf ein *gezieltes Raten* der Lösung hinauslaufen. Um solche Algorithmen zu finden, orientiert man sich häufig an natürlichen Vorgängen. Man findet in der Literatur u.a.

1. genetische (Evolutions-)Algorithmen
2. das *simulated annealing* ("Simulierte Abkühlung")
3. Neuronale Netze (Mustererkennung im Hopfield-Modell als Bestimmung eines der Grundzustände eines Spinglases)

Diese Verfahren werden nun im folgenden behandelt.

9.3 Genetische Algorithmen

Hier ist die Grundidee, die Wirkungsweise der Evolution zu kopieren. Eine Gegenüberstellung von Biologie und informatischem Vorgehen kann in diesem Falle so aussehen:

Biologie	**Informatik**
Problem der "bestmöglichen" Lebensbewältigung	konkretes "genetisch codiertes" Problem
"Evolutionsziel"	problemspezifische, zu optimierende *Fitness*-Funktion
Individuum (Phänotyp)/Genotyp	Lösungsvorschlag/dessen abstrakte Formulierung
Population	Pool von Lösungsvorschlägen
Mutation, Rekombination,	problemspezifische zufällige Änderungen an Individuen, Kombination des Informationsgehaltes verschiedener Individuen
Selektion	Verwerfen "schlechter" Lösungsvorschläge

Man sollte diesen Vergleich natürlich nicht zu weit treiben, denn es gibt kein Ziel oder keinen Zweck der Evolution (Die Natur ist kausal, aber nicht zweckbestimmt). Was die

Evolution hervorbringt, wird vielmehr durch die intrinsischen Regeln der Natur selbst bestimmt und nicht durch eine von außen festgelegte Gütefunktion.

Im Handelsreisendenproblem auf dem Computer wird entsprechend ein Genom als eine der möglichen Permutationen der Indizes $1...N$ der anzusteuernden Punkte angesehen (hier haben wir wieder den Ausgangsort 0 der Reise festgehalten). Die Genotypen haben dann folgende Form (hier ein Beispiel mit $N = 5$):

$$(1\ 2\ 5\ 3\ 4\ )$$

Die Fahrtroute (der Phänotyp), der diesem Genom entspricht, wäre die Rundreise

$$\mathbf{x}_0 \to \mathbf{x}_1 \to \mathbf{x}_2 \to \mathbf{x}_5 \to \mathbf{x}_3 \to \mathbf{x}_4 \to \mathbf{x}_0.$$

Die Fitness-Funktion eines Phänotyps ist die Länge der Fahrtstrecke,

$$\sum_{i=1}^{N-1} ||\mathbf{x}_{i+1} - \mathbf{x}_i||. \tag{9.4}$$

Mutationen bestehen in einer punktuellen, lokalisierten Änderung des Genoms, in unserem Falle etwa dem zufälligen Vertauschen zweier Positionen im N-Tupel. Dabei kann es allein schon durch durch diese Vertauschung zu einer Verbesserung des Lösungsvorschlages kommen (s. Abb. 9.3), aber die eigentliche Bedeutung der Mutation zeigt sich erst in

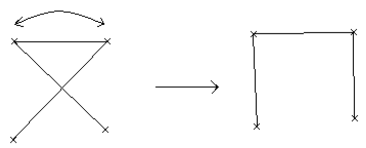

Abbildung 9.3: Durch Vertauschen (Mutation) erhält man eine bessere Lösung

Kombination mit einen Selektionsvorgang.

Eine Population bestehe etwa aus M Individuen. Selektion kann dann dadurch implementiert werden, daß zwei zufällige Genotypen ausgewählt werden, deren Fitness-Funktion berechnet wird, worauf dann (i) das schlechtere Genom verworfen, das bessere jedoch in der Population verdoppelt wird. Damit werden die schlechten Lösungsvorschläge aussortiert und die guten entsprechend vermehrt. So verbessert sich die Qualität der Lösungsvorschläge langsam.

Eine große Beschleunigung des Optimierungsprozesses kann man erreichen, wenn man das Genom zweier Individuen miteinander kombiniert (crossover, Rekombination, sexuelle

Fortpflanzung). Dabei sollen ganze Gensequenzen aus einem Individuum entnommen und unter Beibehaltung möglichst vieler Information mit dem Genom des anderen Individuums verschmolzen werden. In unserem Falle sei etwa das Elternpaar

$$
\begin{array}{cccccccccc}
(& 1 & 5 & | & 4 & 3 & 2 & | & 6 & 7 &) \\
(& 3 & 4 & | & 7 & 6 & 5 & | & 1 & 2 &)
\end{array}
$$

gegeben, in dem wir bereits durch senkrechte Striche zufällig bestimmte Schnittstellen gekennzeichnet haben.

Wir möchten hieraus zwei Nachfahren mit ausgetauschten Genen erhalten, etwa

$$
\begin{array}{lccccccccccc}
\text{1. Nachfahr:} & (& . & . & | & 7 & 6 & 5 & | & . & . &) \\
\text{2. Nachfahr:} & (& . & . & | & 4 & 3 & 2 & | & . & . &).
\end{array}
$$

Die ursprüngliche Sequenz wird am besten erhalten, wenn man die Leerstellen sukzessive mit ursprünglichem Genom auffüllt, wobei die bereits festliegende, übernommene Sequenz *nicht* erneut berücksichtigt wird (eingeklammerte Zahlen) und die anderen "Gene" der Reihe nach eingesetzt werden:

- 1. Nachfahr:

$$
\begin{array}{lccccccccccc}
\text{1. Elternteil:} & (& 1 & (5) & | & 4 & 3 & 2 & | & (6) & (7) &) \\
\Rightarrow \text{Kind:} & (& 1 & 4 & | & 7 & 6 & 5 & | & 3 & 2 &)
\end{array}
$$

- 2. Nachfahr:

$$
\begin{array}{lccccccccccc}
\text{2. Elternteil:} & (& (3) & (4) & | & 7 & 6 & 5 & | & 1 & (2) &) \\
\Rightarrow \text{Kind:} & (& 7 & 6 & | & 4 & 3 & 2 & | & 5 & 1 &)
\end{array}
$$

Die hinter diesem Verfahren steckende Idee ist, daß durch diese Rekombination günstige globale Konstellationen zwischen 2 Individuen ausgetauscht werden können und nicht nur lokale Änderungen wie bei Mutationen vorgenommen werden. Zusammenfassend läßt sich sagen, daß Evolutionsalgorithmen Beispiele sind für mehrskalige Verfahren, die Manipulationen auf lokalem Niveau wie die Mutation, die nur die Reihenfolge zweier anzufahrender Städte ändert, mit grobskaligen Manipulationen wie dem betrachteten crossover Mechanismus kombiniert. Eine solche Mehrskaligkeit oder auch ein hierarchischer Aufbau ist Kennzeichen vieler guter Algorithmen. Ein aktuelles Beispiel sind Mehrgitterverfahren zur Lösung großer linearer Gleichungssysteme oder auch der `quicksort` Sortieralgorithmus (vgl. Übungen in der ersten Kurshälfte).

Im Bild 9.3 zeigen wir noch als Beispiel einen Programmlauf eines genetischen Algorithmus für das Handelsreisendenproblem. Es fällt auf, daß auch der letzte Lösungsvorschlag

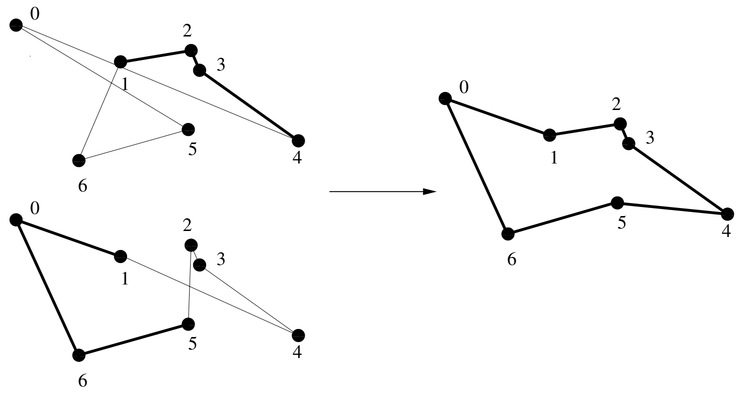

Abbildung 9.4: Durch Rekombination von (Teil-) Lösungsvorschlägen kann man zu günstigeren Lösungen gelangen. Diese Darstellung ist nur schematisch, die Ziffern beziehen sich nicht auf das Beispiel im Text

noch suboptimal ist, denn man erkennt durch näheres Betrachten leicht, daß sich der Lösungsvorschlag mehrfach um das Bildzentrum herumwindet. Die verwendete Implemetierung des genetischen Algorithmus ist nicht (oder nur nach sehr langer Zeit) in der Lage, aus diesem "Nebenminimum" zu entweichen. Aus der Entwicklung der Fitness-Funktion mit der Zeit sieht man, wie das "beste Individuum" immer wieder Mutationen oder crossover-Vorgängen zum Opfer fällt, der Vorrat an guten Genen in der Population jedoch schnell wieder das alte Niveau der Fitness Funktion erreicht.

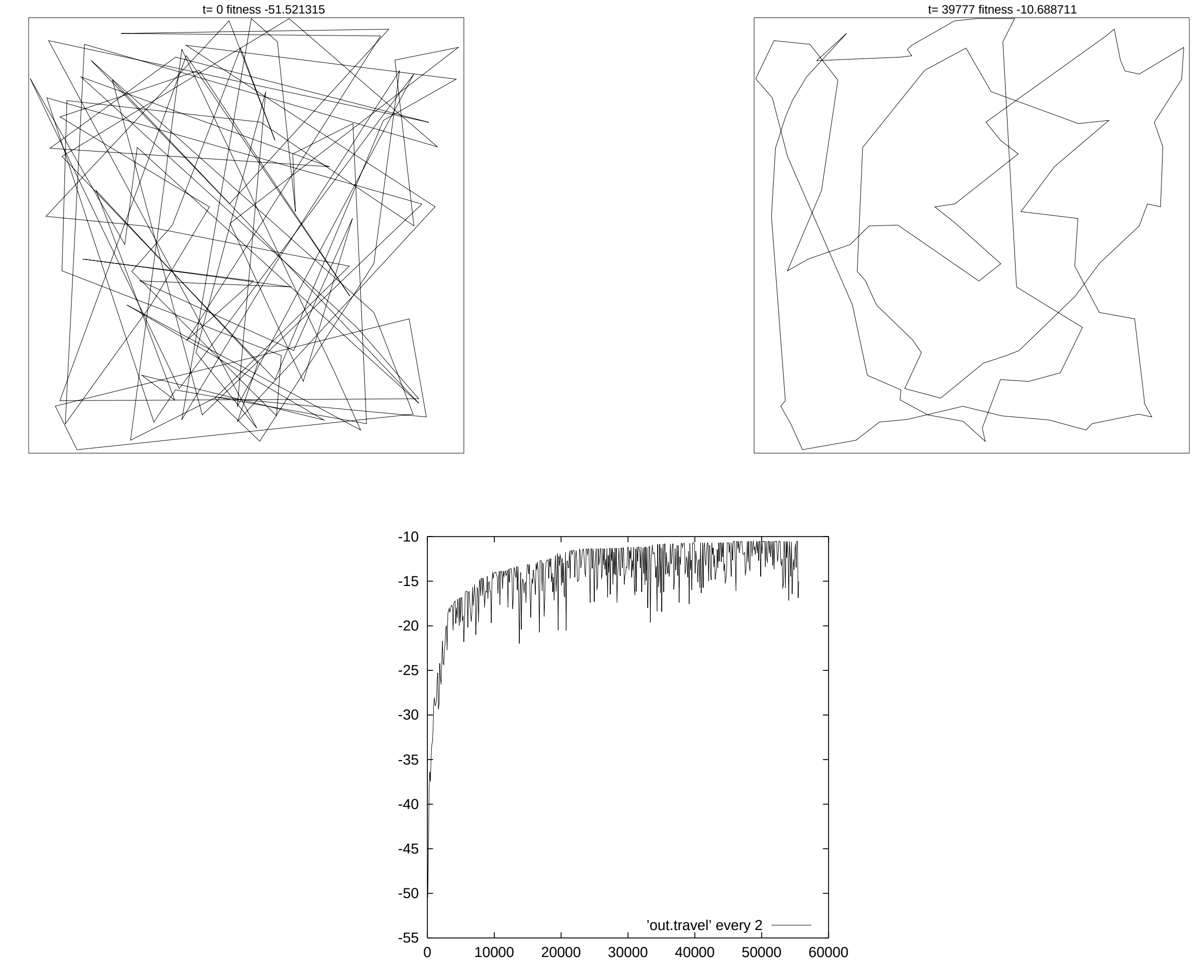

Abbildung 9.5: Entwicklung des "besten Individuums" einer Population von 50 und seiner Fitness als Funktion der Zeit. Ein Zeitschritt wurde mit der Sequenz `MCSMCSMCSSS` durchgeführt, wobei `M` für eine Mutation, `C` für eine Rekombination (crossover) und `S` für ein Selektionsereignis steht. Die Fitness wird maximiert, so daß sie zur negativen Länge der Rundreise gewählt wurde.

9.4 Simulated Annealing

Der englische Begriff *annealing* bedeutet aushärten, wie es etwa ein Schmied mit einem glühenden Stück Metall durch plötzliches Abkühlen am Ende eines Arbeitsganges durchführt.

Wie bereits oben angedeutet wurde, werden bei sinkender Temperatur die statistischen Gewichte der in der Zustandssumme des Systems auftauchenden Grundzustände immer bedeutender. Das System geht bei Abkühlung auf $T = 0$ schließlich in einen der Grundzustände über. Anschaulich statten wir dazu ein "Teilchen," das den Systemzustand (Zustandsvektor) repräsentieren soll, mit einer bestimmten kinetischen "Wärme"energie (charakterisiert durch die Temperatur T) aus. Das Teilchen/System ist in der Lage, je nach der Größe von T in Nachbarzustände zu wechseln, die energetisch nicht zu weit entfernt liegen. Nachbarzustände sind dabei solche, die wir durch geeignete Variationen aus dem Ausgangzustandsvektor erzeugen können.

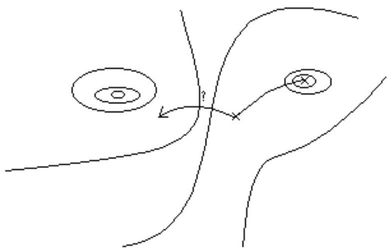

Abbildung 9.6: Energie*landschaft*: wie kann man das globale Minimum finden?

Durch Senken der Temperatur verliert das Teilchen Bewegungsenergie und gelangt tiefer und tiefer in die vorhandenen Potentialmulden. Es hat jedoch immer eine kleine Wahrscheinlichkeit, aus einer solchen Mulde wieder zu entweichen. So ist es möglich, die Minima der Energielandschaft abzusuchen und im tiefsten zu verbleiben. Wir wollen auch diesen Vorgang genauer verstehen und beginnen bei einer stark vereinfachten Vorversion des *Simulated Annealing.*

9.4.1 Hill Climbing

Diese Vorstufe ist das sogenannte *hill climbing.* Dazu

- nehme man einen Ausgangszustand s aus der Gesamtheit $Z(s)$ von Zustandsvektoren, etwa ein N-tupel des Handlungsreisendenproblems oder einen beliebigen Spinzustand des Isingmodelles.
- dann versuche man durch Anwendungen eines Satzes von Manipulationsvorschriften (etwa Vertauschen zweier Positionen bzw. Umklappen eines Spins) einen neuen Zustand $s' \in G(s)$ mit energetisch günstigerer Konfiguration zu erreichen. Dabei ist $G(s)$ die durch diese Manipulationsvorschrift definierte Nachbarschaft von s.
- Der Algorithmus endet, wenn keine günstigere Permutation gefunden werden kann (d.h. wenn $E(s') > E(s)$, $\forall s' \in G(s)$).

Der Name *hill climbing* kommt daher, daß sich der Zustand bei diesem Verfahren immer zu niedrigeren Energien hin bewegt (also immer den "Hügel hinabsteigt"). Das Problem dieses Verfahrens sind allerdings die lokalen Minima, da es einmal in einem solchen gefangen, nicht mehr "herausklettern" kann. Somit ist also nicht gewährleistet, daß die bestmögliche Konstellation gefunden wird.

9.4.2 Stochastic Hill Climbing

Um auch globale Minima aufzufinden, müssen auch scheinbar schlechte Schritte gegen die Abstiegsrichtung mit einer bestimmten Wahrscheinlichkeit zugelassen werden. Dies geschieht im *Probabilistic Hill Climbing.* Dabei soll die Wahrscheinlichkeit p für einen solchen Schritt abnehmen, je größer die Verschlechterung ist. Ein Beispiel dafür ist:

$$p = \begin{cases} e^{-\beta(E(s')-E(s))} & \text{falls } E(s') > E(s), \\ 1 & \text{falls } E(s') < E(s). \end{cases} \tag{9.5}$$

Allerdings treten auch hier Probleme auf, denn

- ein einmal erreichtes Minimum wird auch wieder verlassen, und
- die Umgebungen $G(s)$ müssen hinreichend groß gewählt werden, um ein "Steckenbleiben" zu verhindern.

Beim *Simulated Annealing* verfolgt man daher die Strategie, langsam vom *Probabilistic Hill Climbing* zum normalen *Hill Climbing* überzugehen.

9.4.3 Metropolis Monte Carlo und Simulated Annealing

Es läßt sich zeigen (Metropolis, hier ohne Beweis), daß die oben für das *Stochastic Hill Climbing* eingeführten Zugregeln (9.5) im Grenzwert langer Simulationszeiten zu einer Besuchswahrscheinlichkeit der einzelnen Zustände s führen, die der Boltzmannstatistik in der Thermodynamik und Statistischen Physik entspricht. Der Beweis benötigt als Voraussetzung, daß unsere Manipulationsvorschrift $s \to s'$ es erlaubt, alle Systemzustände zu erreichen und daß der Zugvorschlag von s nach s' mit gleicher Wahrscheinlichkeit wie der Vorschlag s' nach s gemacht wird.

$$p(s_i) = \frac{1}{Z} e^{-\beta E(s_i)} \tag{9.6}$$

wobei $\beta = \frac{1}{kT}$ die inverse Temperatur und Z der Normierungsfaktor $Z = \sum e^{-\beta E(s_i)}$ ist, der die Summe der Besuchswahrscheinlichkeiten aller Zustände zu 1 ergibt.

Was passiert hier für absinkende Temperaturen $T \to 0$? Für die Besuchswahrscheinlichkeit eines beliebigen Zustandes s_i gilt

$$p(s_i) = \frac{e^{-\beta E(s_i)}}{\sum e^{-\beta E(s_i)}} \tag{9.7}$$
$$= \frac{e^{-\beta E(s_i)}}{e^{-\beta E(s_i)} + \sum_{i \neq j} e^{-\beta E(s_j)}} \tag{9.8}$$
$$= \frac{1}{1 + \sum_{i \neq j} e^{-\beta (E(s_j) - E(s_i))}}. \tag{9.9}$$

Falls also nur ein einziges globales Minimum s_0 existiert mit $E_0 < E_j \; \forall j \neq 0$, dann finden wir

$$\Rightarrow \lim_{\substack{\beta \to \infty \\ (T \to 0)}} p(s_0) = \frac{1}{1+0} = 1, \tag{9.10}$$

d.h., der Zustand s_0 wird mit Wahrscheinlichkeit 1 angenommen. Bei Entartung gilt in analoger Weise für die energetisch gleichliegenden Zustände s_i,

$$p(s_i) = \frac{1}{1 + \sum\limits_{i \neq j} 1} = \frac{1}{\text{Anzahl der Zustände}}, \tag{9.11}$$

also Gleichverteilung über die entarteten Grundzustände.

Wir sehen daher, daß wir mit einer geeigneten Abkühlstrategie auf $T = 0$, bzw. $\beta \to \infty$ die Minima finden können. Leider ist nicht nur der Grenzübergang $T \to 0$ zu machen, sondern in den obigen Gleichungen ist noch der Grenzübergang versteckt, der uns überhaupt die Definition der Wahrscheinlichkeit als die relative Häufigkeit des Besuchs von s_i erlaubt. Die obigen Argumentation bedingt also, daß dem System bei jeder Temperatur ausreichend Zeit zur Verfügung steht, verschiedene Zustände zu besuchen. Die Abkühlstrategie spielt also eine wesentliche Rolle, und sie bestimmt wie gut das System es vermeiden kann, in Nebenminima steckenzubleiben.

Wie sieht der zugehörige Algorithmus aus?

1. Wähle einen beliebigen Ausgangszustand: $s \leftarrow s_1$, und $T > 0$
2. Suche ein $s' \in G(s)$, so daß alle s_i prinzipiell nach endlicher Zugzahl zugänglich sind
3. Setze $p = e^{-\beta(E(s')-E(s))}$ Falls $p > 1 \quad (E(s') < E(s)) \quad s \leftarrow s'$
   Falls $p < 1$ wähle E(s') mit Wahrscheinlichkeit p.
4. Wähle eine neue (geringfügig kleinere) Temperatur $T' < T$ oder verbleibe für eine gewisse Zeit bei der alten.

Als Abkühlungsstrategien kann man z.B. die folgenden benutzen:

1. exponentielles Abkühlen $T' = (1 - \epsilon)T$. Hier bietet sich das Abbrechen des Algorithmus nach Erreichen einer minimalen Temperatur an.
2. lineares Abkühlen $T' = (1 - \epsilon n)T$, wobei n die Anzahl der bisherigen Durchläufe des Algorithmus ist. Beim linearen Abkühlen terminiert der Algorithmus nach $n = \frac{1}{\epsilon}$ Versuchen.

Zum Abschluß zeigen wir auch für diesen Fall die Zeitentwicklung der Weglänge des Handelsreisendenproblems mit den gleichen Städten wie im Abschnitt über genetische Programmierung. Auch hier sieht man noch, wie der Algorithmus in einem Nebenminimum gefangen bleibt und die Güte der Lösung etwa vergleichbar mit der des genetischen Algorithmus ist.

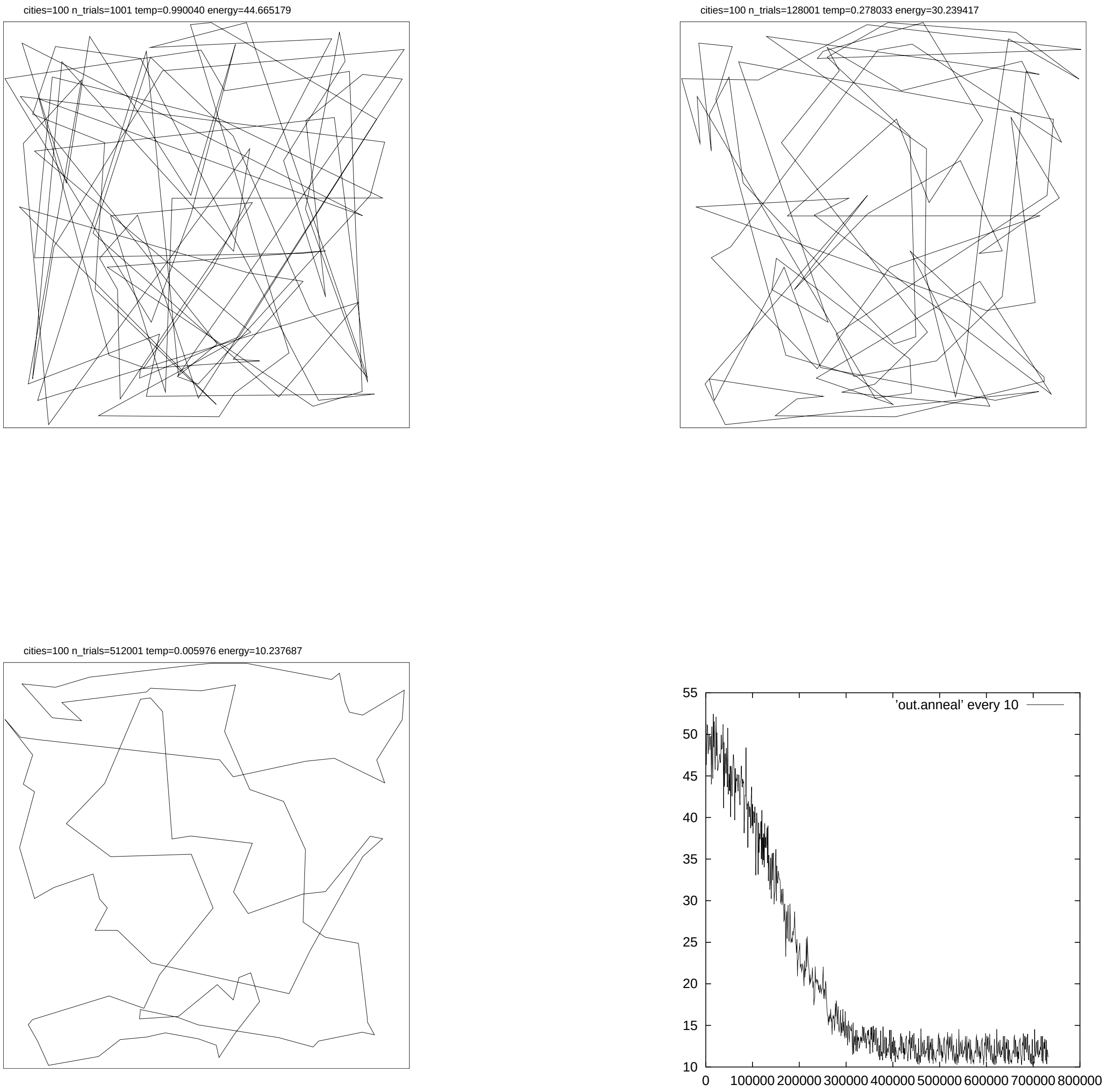

Abbildung 9.7: Entwicklung der Reiselänge bei Simulated Annealing. Ein Zeitschritt entspricht einem Monte-Carlo Schritt, ist also deutlich weniger aufwendig als ein Zeitschritt in der Vergleichssimulation für den genetischen Algorithmus. Dies erklärt die größeren Zahlen an der Zeitachse der Darstellung der Länge der Route gegen die Zeit (rechts unten).