

Joerg Uhlig

Bayessches Lernen in der Quantentheorie

2000

Theoretische Physik

Bayessches Lernen in der Quantentheorie

Inaugural-Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften im Fachbereich Physik
der Mathematisch-Naturwissenschaftlichen Fakultät
der Westfälischen Wilhelms-Universität Münster

vorgelegt von
Joerg Uhlig
aus Münster (Westf.)
–2000–

Dekan:

Prof. Dr. W. Lange

Erster Gutachter:

Prof. Dr. A. Weiguny

Zweiter Gutachter:

Prof. Dr. M. Stingl

Tag der mündlichen Prüfung(en):

Tag der Promotion:

Theoretische Physiker wie ich
verbringen den größten Teil
ihrer Zeit damit,
sie zu verschwenden.
(Steven Weinberg im Spiegel-Interview 30/99)

Inhaltsverzeichnis

Einleitung	1
1 Prinzipien des Bayesschen Lernens	3
1.1 Notation und Formalismus	3
1.2 Unabhängige Messungen	5
1.3 Energien	6
1.4 Lernstrategien	6
1.5 Hyperparameter	8
1.6 Algorithmen	10
2 Lernen von zeitabhängigen Wahrscheinlichkeitsdichten	11
2.1 Formulierung in Wahrscheinlichkeitsdichten	11
2.2 Formulierung in logarithmischen Wahrscheinlichkeiten	14
2.3 Behandlung von Mustern und Korrelationen	15
2.4 Anwendungen für die Dichteapproximation	18
3 Lernen von statistischen Operatoren	25
3.1 Statistische Operatoren	25
3.2 Lernen unter Maximierung der Entropie	26
3.3 Spezialfall gleicher Messungen	28
3.4 Lernen unter Minimierung der Entropie	30
4 Lernen von parametrisierten Potentialen	34
4.1 Doppelmuldenpotentiale	34
4.2 Anharmonischer Oszillator	37
5 Lernen von Potentialen (Energiedarstellung)	45
5.1 Klassische Näherung	46
5.2 Spektraler Zugang	47
5.3 Hoch- und Tieftemperaturgrenzfälle	49
5.4 Numerische Implementierung	50
5.5 Beispiele für Apriori-Wissen	51
5.6 Rekonstruktion von Potentialen	54
6 Lernen von Potentialen (Pfadintegraldarstellung)	61
6.1 Erste Variante	61
6.2 Zweite Variante	64
6.3 Näherungen für die Matrixelemente $\langle x e^{-\beta H} x \rangle$	67
6.4 Näherungen für $\int_0^\beta d\beta' \langle x e^{-(\beta-\beta')H} x' \rangle \langle x' e^{-\beta' H} x \rangle$	69

6.5 Numerische Implementierung	72
7 Lernen von Zweiteilchenwechselwirkungen	76
7.1 Kanonisches Ensemble bei $T = 0$	76
7.2 Großkanonisches Ensemble (Energiedarstellung)	82
7.3 Großkanonisches Ensemble (Pfadintegraldarstellung)	86
Zusammenfassung	93
A Datenerzeugung aus vorgegebener Wahrscheinlichkeitsverteilung	95
B Operatorenableitung	96
C Pseudoinversion von $H - E$	98
D Hesse-Matrix im Tieftemperaturfall	100
E N-fache Ableitung der Zustandssumme	103
Literaturverzeichnis	105

Einleitung

Das Hauptanliegen der Naturwissenschaften besteht darin, die beobachtbaren Phänomene der Natur zu verstehen und erklären zu können. Um dies zu erreichen, gibt es zwei Parteien, die Hand in Hand arbeiten: Der Experimentator liefert durch technischen Fortschritt und neue Meßideen immer genauere Meßdaten für die Beschreibung von Naturerscheinungen und versucht, durch seine Experimente theoretische Berechnungen empirisch zu verifizieren oder zu widerlegen. Der Theoretiker versucht, aus experimentellen Erkenntnissen und mit gesundem Menschenverstand die Natur in einem Modell vereinfacht darzustellen und hiermit (für weitere Experimente) vorhersagbar zu machen. Das Manko des Experimentators liegt dabei in der begrenzten Meßgenauigkeit seiner Daten – das des Theoretikers in der Notwendigkeit von analytischen und numerischen Näherungen. Lernen – als eine Verbesserung des generellen Wissenstands im Laufe der Zeit – hat die Aufgabe, die Ergebnisse beider Parteien zusammenzuführen, um immer präzisere Aussagen über den Zustand der Natur zu machen. Es kann somit als eine Art Schnittstelle von Theorie und Experiment angesehen werden.

Das Lernen aus Bayesscher Sicht wird dabei im wesentlichen von drei Faktoren beeinflußt: dem Vorwissen, der Lernstrategie und dem Zufall. Unter Vorwissen versteht man dabei zusammenfassend alle Erwartungen, die an den Zustand des betrachteten Systems vor Erhebung der Daten geknüpft sind. Hierunter fallen beispielsweise Meinungen und Erfahrungen von Experten, unterstützt von plausiblen Argumenten von Laien, die Existenz von Mustern für die Problemstellung oder das Wissen über zu erwartende Meßgenauigkeiten. Die Lernstrategie hängt im wesentlichen von der Zielsetzung und der Komplexität des Problems ab. Sie regelt die gesamte Organisation für die Entscheidungsfindung. Zu guter Letzt muß man sich darüber im klaren sein, daß man auf den Umgang mit empirischen Daten angewiesen und somit zu probabilistischen Aussagen gezwungen ist. Die Lerntheorie wird also von den Gesetzen der Wahrscheinlichkeitsrechnung getragen.

Diese Arbeit befaßt sich mit dem Lernen an quantenmechanischen Systemen. Dabei nehmen die Gesetze der Quantenmechanik durch Festlegung von Meßwahrscheinlichkeiten und durch das quantenmechanische Meßprinzip Einfluß auf den Formalismus. Möchte man durch Messungen Aussagen über das Potential eines Hamilton–Operators gewinnen, so handelt es sich dabei um ein inverses Problem, da die quantenmechanische Theorie Auskunft über Meßwahrscheinlichkeiten bei gegebenem Potential macht – und nicht umgekehrt. Es gibt bis dato mehrere Teilbereiche der Physik, die sich mit dieser Thematik beschäftigen: In der Spektraltheorie der mathematischen Physik wird untersucht, welche Mindestinformation über Wellenfunktionen und Eigenwerte vorliegen müssen, um einen Hamilton–Operator eindeutig zu spezifizieren. Die inverse Streutheorie versucht, mittels Wirkungsquerschnitten und Streuphasen Rückschlüsse auf das zugrundeliegende Streupotential zu ziehen, und erhält eine experimentelle Verwirklichung in Kern–, Atom– und Molekülphysik sowie in der Elektronenmikroskopie. Als letzter Punkt sei hier die Rastertunnelmikroskopie erwähnt, bei der der Tunnelstrom Aussagen über das Potential im Oberflächenbereich des untersuchten Festkörpers zuläßt.

Der entscheidende Zugang für das Lernen in der Quantenmechanik gelingt in dieser Arbeit durch die Anwendung des Bayes–Theorems für bedingte Wahrscheinlichkeiten. Hierdurch ergibt sich die

Möglichkeit, das inverse Lernproblem unter Kontrolle zu bringen, indem die logische Abhängigkeit der "Ereignisse" vertauscht wird. Die Einbindung von Vorwissen erfolgt hier auf natürliche Weise. Sie ist essentiell, da eine endliche Menge von Meßdaten keine beliebig genauen Aussagen über die unendlichdimensionale Potentialfunktion zuläßt. Durch Betrachtung von statistischen Ensembles, die sich im thermischen Gleichgewicht mit ihrer Umgebung befinden, können voneinander unabhängige Messungen vorgenommen werden. Die durch den Bayesschen Ansatz resultierenden Lernalgorithmen haben das iterative, selbstkonsistente Lösen von nichtlinearen Differentialgleichungen zum Ziel.

Insgesamt kann man sagen, daß in dieser Arbeit die Theorien der klassischen Wahrscheinlichkeitsrechnung und der Quantenstatistik miteinander verbunden werden. Ziel der Arbeit ist es, anhand verschiedener Anwendungen einen Einblick in das Bayessche Lernen in der Quantentheorie zu liefern und dabei aufzuzeigen, worin die Vor- und Nachteile dieses Ansatzes liegen.

Die Arbeit ist folgendermaßen gegliedert: Im ersten Kapitel wird die Theorie des Bayesschen Lernens vorgestellt. Dabei werden die Grundlagen für das Verständnis und Motivationen für das Vorgehen in den speziellen Anwendungen der weiteren Kapitel geliefert.

Das zweite Kapitel befaßt sich mit dem Regressions- und dem Dichteapproximationsproblem. Es wird das Bayessche Lernen für sich zeitlich ändernde Wahrscheinlichkeitsdichteverteilungen angewandt. Dabei könnte es sich um die "Trajektorie" eines quantenmechanischen Teilchens handeln, wobei das Teilchen nur zu bestimmten Zeiten lokalisiert werden kann. Das Vorwissen verbindet sich mit der Erwartung, daß eine in Orts- und Zeitkoordinaten glatte Dichteverteilung auftritt. Das dritte Kapitel beinhaltet das Lernen statistischer Operatoren anhand von Messungen verschiedener Observablen. Dabei werden zwei entgegengesetzte Fälle behandelt: Der statistische Operator soll das Prinzip der Maximierung der Informationsentropie erfüllen oder einen reinen Zustand beschreiben.

Im vierten Kapitel werden Potentiale gelernt, welche durch ein- oder zweidimensionale Variablen parametrisiert sind. So werden Aussagen über ein Doppelmuldenpotential gemacht, indem Ortsmessungen im Tieftemperaturfall vorgenommen werden. Dabei werden genäherte Grundzustandswellenfunktionen betrachtet. In einem weiteren Beispiel wird für einen anharmonischen Oszillator die unbekannte Kopplungskonstante gelernt. Die Ortsmeßwahrscheinlichkeiten werden durch ein Pfadintegral dargestellt und hierfür Näherungen für die beiden Temperaturgrenzfälle vorgenommen.

Der Generalisierungsschritt auf nichtparametrische Potentiale erfolgt im fünften und sechsten Kapitel. Für das Lernen von Einteilchenpotentialen mittels Ortsmessung am kanonischen Ensemble wird eine Theorie einer inversen Quantenstatistik konstruiert [1]. Dabei werden unterschiedliche Darstellungen diskutiert: So können die Ergebnisse durch das Arbeiten in der Energiedarstellung und der Pfadintegraldarstellung miteinander verglichen werden.

Die Situation aus den vorherigen Kapiteln wird im siebten Kapitel auf den fermionischen Vielteilchenfall verallgemeinert. So wird für den Tieftemperaturfall eine Zweiteilchenwechselwirkung gelernt, indem näherungsweise der Hartree-Fock-Grundzustand betrachtet wird [2]. In einem weiteren Schritt werden auf zwei verschiedene Arten Hartree-Fock-Theorien für das großkanonische Ensemble ausgearbeitet: Diese unterscheiden sich in der Reihenfolge, in der eine Mean-Field-Näherung für das Vielteilchenproblem und eine Maximum-Aposteriori-Approximation durchgeführt werden. So können beide Approximationen gleichzeitig oder nacheinander angewandt werden. Die jeweilige Näherungsstrategie hat Konsequenzen für die Struktur der zu lösenden Gleichungen.

Die analytischen Ergebnisse einzelner Kapitel werden durch numerische Beispiele veranschaulicht und damit die Anwendbarkeit der Ansätze verdeutlicht beziehungsweise deren Grenzen aufgezeigt.

Kapitel 1

Prinzipien des Bayesschen Lernens

In diesem Kapitel werden die Grundzüge der Theorie des Lernens aus Sicht der Bayesschen Wahrscheinlichkeitsrechnung vorgestellt. Es wird darauf hingewiesen, daß versucht wurde, Formeln und Begriffe möglichst einfach, abstrakt und allgemein zu halten, um die Aufmerksamkeit des Lesers auf die wesentlichen Merkmale der Theorie zu lenken. Detailliertere Darstellungen erfolgen dann in den Anwendungen der nächsten Kapitel. Es wird davon ausgegangen, daß der Leser mit den Grundregeln der Wahrscheinlichkeitsrechnung vertraut ist. Als fundamentale Einführungen in den Formalismus sind die Bücher von Bishop [3] und Berger [4] zu empfehlen. Weiterführende Anwendungen und Varianten werden bei Lemm vorgeführt [5, 6].

1.1 Notation und Formalismus

Das zugrundeliegende – meist physikalische – System wird über eine feste Anzahl von Variablen kontrolliert, welche den Zustand des Systems charakterisieren. Diese Größen sind in der Regel nicht direkt meßbar und werden kollektiv mit $f \in \mathcal{F}$ bezeichnet – wobei \mathcal{F} die Menge aller möglichen f kennzeichnet – und können unter Umständen sehr hochdimensional sein. Der Begriff "Zustand" kann dabei verallgemeinert aufgefaßt werden. So legt der statistische Operator den Zustand in einem quantenmechanischen System eindeutig fest, während ein Hamilton–Operator die Dynamik eines Systems bestimmt, ohne den Zustand durch Angabe von Anfangsbedingungen eindeutig zu charakterisieren. Bei physikalischen Anwendungen kann – je nach Fragestellung – "Zustand" eine Wellenfunktion, einen statistischen Operator oder das Potential im Hamilton–Operator beziehungsweise deren klassisches Analogon bedeuten.

Ziel ist es, probabilistische Aussagen über den wahren Zustand des Systems \tilde{f} zu machen, indem man sich Informationen durch Messungen von Observablen verschafft, welche direkt oder indirekt von \tilde{f} abhängen. Wichtig ist dabei, daß vor jeder einzelnen Messung jeweils der gleiche Zustand vorliegt. Jedes Meßergebnis wird dabei durch zwei Größen beschrieben: erstens durch die Observable, die man mißt und die vom Experimentator vorgegeben oder kontrolliert wird – repräsentiert durch $x \in \mathcal{X}$ – und zweitens durch den eigentlichen Meßwert $y(x) \in \mathcal{Y}_x$. Die Gesamtheit der empirischen Daten aus N Messungen wird mit der Menge $D = \{(x_i, y_i) \mid i = 1, \dots, N\}$ abgekürzt. Die erwähnte Unsicherheit bei der Festlegung des wahren – aber unbekannten – Zustands \tilde{f} nach Erhebung der Daten wird über die bedingte Wahrscheinlichkeit $p[f|D]$ ausgedrückt. $p[f|D]$ repräsentiert den aktuellen Stand des Wissens. Ziel des Lernens ist es, durch immer mehr Daten immer genauere Aussagen über \tilde{f} machen zu können. Das heißt, eine immer schärfere konzentrierte Wahrscheinlichkeitsverteilung zu erhalten, bis möglicherweise am Ende eine deterministische

Aussage steht:

$$p[f|D] \longrightarrow \delta(f - \tilde{f}) \quad \text{für } N \rightarrow \infty. \quad (1.1)$$

Der formelle Ausdruck $p[f|D]$ ist von der logischen Reihenfolge der Ereignisse D und f her konträr zur gewöhnlich anzutreffenden Ordnung. So ist in den meisten Fällen das Augenmerk auf die Verteilung $p[D|f]$ gerichtet – also die Messung von Daten bei gegebenem Systemzustand. Somit kann man sagen, daß es sich beim Lernen – also der Bestimmung von der Wahrscheinlichkeit $p[f|D]$ und ihrer zeitlichen Entwicklung – um ein inverses Problem handelt. Oft hat man bei einfachen theoretischen Modellen die Größe $p[D|f]$ analytisch unter Kontrolle, während die inverse Verteilung $p[f|D]$ schwierig zu handhaben ist. Hierin liegt die Motivation, mittels der Identität für verbundene und bedingte Wahrscheinlichkeiten

$$p[f, D] = p[f|D] p[D] = p[D|f] p[f] \quad (1.2)$$

die logische Abhängigkeit nach dem Bayes–Theorem umzudrehen:

$$p[f|D] = \frac{p[D|f] p[f]}{p[D]}. \quad (1.3)$$

Die in der Identität auftretenden Faktoren haben folgende Bedeutung:

- Die Verteilung $p[f]$ gibt das Vorwissen und die Erwartungen, die man mit f verknüpft, vor Erhebung der Daten wieder. Sie beschreibt den Startpunkt des Lernens und wird deshalb auch Apriori–Wahrscheinlichkeitsverteilung genannt. Bei Nichvorhandensein von Vorwissen wird $p[f]$ als Gleichverteilung gewählt.
- $p[f|D]$ ist der Stand des Wissens nach Erhebung der Daten. Der Vergleich mit $p[f]$ beschreibt den Lernfortschritt. $p[f|D]$ bezeichnet man als Aposteriori–Verteilung.
- $p[D|f]$ wird in der Literatur Likelihood für f genannt, weil für fixierte Daten die Funktion $g_D(f) = p[D|f]$ große Funktionswerte für solche f hat, welche dem wahren, datenerinnernden \tilde{f} sehr ähnlich sind. Man beachte, daß bei $p[f] = \text{const}$ die Proportionalität $p[D|f] \sim p[f|D]$ gilt.
- Die unbedingte Datenwahrscheinlichkeit $p[D]$ ist eine f –unabhängige Normierungskonstante. Sie ist über die Normierungsbedingung von (1.3) festgelegt: $p[D] = \int df p[D|f] p[f]$.

Oft ist gar nicht von Interesse, in welchem Zustand sich das System nach den Messungen von Trainingsdaten D_0 befindet, sondern welche Daten D (Testdaten) bei weiteren Messungen zu erwarten sind. Es gilt nach künstlicher Einführung eines Integrals über alle möglichen Zustände f :

$$\begin{aligned} p[D|D_0] &= \int df p[D, f|D_0] = \int df p[D|f, D_0] p[f|D_0] = \int df p[D|f] p[f|D_0] \\ &= \frac{1}{p[D_0]} \int df p[D|f] p[D_0|f] p[f] = \frac{1}{p[D_0]} \langle p[D|f] p[D_0|f] \rangle_{p[f]}, \end{aligned} \quad (1.4)$$

wobei angenommen wurde, daß die Testdaten nur direkt von f abhängen ($p[D|f, D_0] = p[D|f]$). $\langle \dots \rangle_{p[f]}$ steht für die Erwartungswertbildung unter der Verteilung $p[f]$.

Man sieht, daß die Möglichkeit, Aussagen über zukünftige Daten zu treffen, effektiv von der Korrelation der beiden betreffenden Likelihood–Wahrscheinlichkeiten unter dem Vorwissen $p[f]$ abhängt. Dieses verdeutlicht den engen Zusammenhang von Apriori–Wissen und Korrelationen für die Fähigkeit zu generalisieren. Bei der Abwesenheit von Vorwissen ($p[f] = \text{const}$) gilt:

$$p[D|D_0] \sim \int df p[D|f] p[D_0|f]. \quad (1.5)$$

Das Integral wird von demjenigen \hat{f} dominiert, für das das Produkt der beiden Likelihood-Verteilungen maximal ist. Im Falle, daß genügend große Datenmengen D, D_0 vorliegen, sind diese Verteilungen in der Variablen f lokalisiert, und es kann genähert werden:

$$\operatorname{argmax}_D p[D|D_0] = \operatorname{argmax}_D \int df p[D|f] p[D_0|f] \approx \operatorname{argmax}_D p[D|\hat{f}] p[D_0|\hat{f}]. \quad (1.6)$$

Wie nicht anders zu erwarten, sind die wahrscheinlichsten Testdaten D solche, für die die Likelihood-Dichte nahe derjenigen der Trainingsdaten ist, d.h. für Daten D , die D_0 sehr ähnlich sind. Generalisieren ist also nur unmittelbar in der Nähe von D_0 möglich, wenn kein Vorwissen vorhanden ist.

1.2 Unabhängige Messungen

Es wird meistens bei den Daten davon ausgegangen, daß sie unabhängig voneinander gewonnen wurden ($p[x_i, y_i|x_j, y_j] = p[x_i, y_i], \forall i, j$), so daß die verbundenen Wahrscheinlichkeiten faktorisieren:

$$p[D|f] = p[(x_1, y_1), \dots, (x_N, y_N)|f] = \prod_{i=1}^N p[x_i, y_i|f] = \prod_{i=1}^N p[y_i|x_i, f] p[x_i|f] = \prod_{i=1}^N p[y_i|x_i, f] p[x_i], \quad (1.7)$$

wobei die Unabhängigkeit der Wahl der Messung x von f angenommen wurde. Es wird davon ausgegangen, daß der Experimentator die Verteilung $p[x]$ kennt oder kontrolliert, er aber möglicherweise auswürfelt, welche Messung er jeweils vornimmt. Oft wird man nur eine Observable \tilde{x} zulassen und deren Messung wiederholt durchführen ($p[x] = \delta(x - \tilde{x})$).

Im quantenmechanischen Fall sei zum Meßprozeß folgendes angemerkt: Durch eine sortierende Messung geht das System von einem Ausgangszustand in einen neuen Zustand – einen Eigenzustand der Observablen zum gemessenen Eigenwert – über. Befindet sich beispielsweise ein System in einem reinen Zustand $f = |\Psi\rangle\langle\Psi|$ und möchte man die verbundene Wahrscheinlichkeit für die Messung zweier nichtkommutierender Observablen \hat{A} und \hat{B} mit den jeweiligen Meßwerten a und b angeben, so muß die Reihenfolge der Messungen festgelegt werden. Man bekommt im Fall von "erst \hat{B} dann \hat{A} " die bedingten Wahrscheinlichkeiten

$$p[(a, \hat{A}), (b, \hat{B})|f] = p[(a, \hat{A})|(b, \hat{B}), f] p[(b, \hat{B})|f] \quad \text{mit } f = |\Psi\rangle\langle\Psi|. \quad (1.8)$$

Unter der Voraussetzung, daß \hat{B} keine entarteten Eigenwerte besitzt, gilt $p[(a, \hat{A})|(b, \hat{B}), f] = p[(a, \hat{A})|(b, \hat{B})] \neq p[(a, \hat{A})|f]$. Die Messungen faktorisieren also nicht analog zu (1.7), und die explizite Abhängigkeit von $|\Psi\rangle$ ist nur für die erste Messung gegeben. Insgesamt kann man dann sagen, daß durch die Messung die Möglichkeit verloren geht, den unbekannten Zustand $|\Psi\rangle$ eindeutig zu bestimmen, da durch den Meßprozeß nur die Komponente von $|\Psi\rangle$ auf dem Eigenraum des Meßwerts überlebt. Alle folgenden Messungen erhöhen im Falle von Nichtentartung die Information über $|\Psi\rangle$ nicht. In den Büchern von Ballentine und Balian [7, 8] wird eine Diskussion zur Berechnung von bedingten Wahrscheinlichkeiten und deren Interpretation anhand eines quantenmechanischen Spinsystems vorgeführt.

Die Unabhängigkeit der Messungen kann nur erreicht werden, indem gewährleistet wird, daß vor jeder einzelnen Messung der quantenmechanische Zustand durch den gleichen statistischen Operator beschrieben wird. Diese Bedingung kann durch die Betrachtung von kanonischen Gesamtheiten erfüllt werden. Eine Alternative dazu ist das wiederholte Messen an einem einzelnen Ensemblemitglied, wobei nach jeder Messung genügend Zeit gelassen wird, bis das System abermals in das thermische Gleichgewicht gelangt. Eine andere Möglichkeit ist eine jeweilige Präparation von identischen Ausgangszuständen durch ein vorangegangenes Experiment.

1.3 Energien

Es ist aus verschiedenen Gründen sehr bequem, Wahrscheinlichkeitsverteilungen $p[z]$ durch "Energiefunktionen" $\phi(z)$ auszudrücken und dabei künstlich eine Skalierung λ einzuführen:

$$p[z] \equiv \frac{e^{-\lambda\phi(z)}}{Z(\lambda)} \quad , \quad Z(\lambda) = \int dz e^{-\lambda\phi(z)} \quad (1.9)$$

mit Umkehrabbildung

$$\phi(z) = -\frac{1}{\lambda} \ln p[z] \quad , \quad \int dz p[z] = 1. \quad (1.10)$$

Die Motivation hierfür ist das bequeme Arbeiten mit den unnormierten Energien $\phi(z)$. Die notwendige Bedingung $p[z] \geq 0 \forall z$ ist schon explizit durch Parametrisierung gewährleistet. Die Monotonie der e -Funktion lässt die Schlussfolgerung zu, daß kleine $\phi(z)$ zu großen Wahrscheinlichkeiten $p[z]$ führen. In der Parametrisierung (1.9) lassen sich bequem Wahrscheinlichkeitsverteilungen mit gewünschten Eigenschaften konstruieren. So führt die eindimensionale Energiefunktion $\phi(z) = (1/2)(z - z_0)^2$, welche die Abweichung von z zu einer Referenz z_0 "bestraft", bekannterweise zur Gauß-Verteilung

$$p[z] = \sqrt{\frac{\lambda}{2\pi}} e^{-\frac{\lambda}{2}(z - z_0)^2}. \quad (1.11)$$

$p[z]$ könnte zum Beispiel der Apriori-Wahrscheinlichkeit $p[f]$ aus (1.3) entsprechen, wobei dann das Vorwissen der Erwartung entsprechen würde, daß f nahe einem vorgegebenen Zustand $f_0 = z_0$ liegt.

Es wird darauf hingewiesen, daß eine formale Analogie von (1.9) zur Boltzmann-Verteilung vorliegt, beide Ansätze aber unter verschiedenen Gesichtspunkten gesehen werden müssen. Während die Boltzmann-Verteilung diejenige Lösung $p[z]$ ist, welche bei physikalisch vorgegebener Hamilton-Funktion die Entropie eines kanonischen Ensembles bei konstanter Temperatur $\beta = 1/(kT)$ maximiert, wird hier die Gleichung (1.9) als Abbildung von ϕ nach p angesehen, weil es bequemer ist, ϕ zu kontrollieren. Die Näherungsformalismen der statistischen Mechanik (Hoch- und Tieftemperaturnäherungen, Sattelpunktsnäherungen, etc.) lassen sich auf die Theorie übertragen, wenn man die Skalierung λ als inverse Temperatur auffaßt.

Bei faktorisierenden Wahrscheinlichkeitsverteilungen – beispielsweise bei vorliegenden unabhängigen Ereignissen z_i – für die die jeweiligen Energiefunktionen $\phi_i(z_i)$ und Zustandssummen Z_i bekannt sind, können die Energiefunktionen addiert und die Zustandssummen multipliziert werden:

$$p[z_1, \dots, z_M] = \prod_{i=1}^M p[z_i] = \frac{e^{-\lambda \sum_{i=1}^M \phi_i(z_i)}}{\prod_{i=1}^M Z_i} = \frac{e^{-\lambda \Gamma(z_1, \dots, z_M)}}{Z} \quad (1.12)$$

mit $Z = \prod_i Z_i$ und $\Gamma = \sum_i \phi_i$.

1.4 Lernstrategien

Wenn das eigentliche Ziel des Lernens nicht die Angabe von $p[f|D]$ ist, sondern man sich bei der Kenntnis von $p[f|D]$ für eine Hypothese $f^* \in \mathcal{F}$ für den wahren Zustand \tilde{f} entscheiden muß, gibt es verschiedene Vorgehensweisen.

Man wird in der Regel dasjenige f^* wählen, welches $p[f|D]$ maximiert:

$$f^* = \operatorname{argmax}_{f \in \mathcal{F}} p[f|D], \quad (1.13)$$

also insbesondere nach Stellen suchen, für die

$$\frac{\partial p[f|D]}{\partial f} = 0 \quad \text{oder} \quad \frac{\partial}{\partial f} (p[D|f] p[f]) = 0 \quad , \quad \frac{\partial^2 p[f|D]}{\partial f^2} < 0 \quad (1.14)$$

erfüllt ist. Diese Entscheidung für ein festes f^* und deren weitere Konsequenzen für den Formalismus wird Maximum–Aposteriori–Approximation genannt. Falls die Verteilung $p[f|D_0]$ in (1.4) im Falle einer großen Datenmenge D_0 scharf um ein f^* lokalisiert ist, kann für die Wahrscheinlichkeit weiterer Messungen mit $p[D|f, D_0] = p[D|f]$ mittels

$$p[D|D_0] = \int df p[D|f] p[f|D_0] \approx p[D|f^*] \int df p[f|D_0] = p[D|f^*] \quad (1.15)$$

genähert werden. Sie entspricht der Likelihood–Verteilung bezüglich f^* .

Im Gegensatz dazu ist es auch möglich, die Likelihood–Verteilung $p[D|f]$, aufgefaßt als Funktion von f für fixierte Daten D , bezüglich f zu maximieren:

$$f^* = \operatorname{argmax}_{f \in \mathcal{F}} p[D|f]. \quad (1.16)$$

Diese Vorgehensweise wird Maximum–Likelihood–Methode genannt. Sie ist dann sinnvoll, wenn das Apriori–Wissen schwach von f abhängt, denn für $p[f] = \text{const}$ fallen die beiden Entscheidungen zusammen. Natürlich ist es auch möglich, den Erwartungswert $f^* = \int df f p[f|D]$ zu wählen, falls dieser in \mathcal{F} enthalten ist. Dieses setzt voraus, daß $p[f|D]$ analytisch für alle f kontrolliert und das möglicherweise hochdimensionale f –Integral ausgeführt werden kann.

Manchmal ist es gar nicht so sinnvoll, sich für die wahrscheinlichste Hypothese f^* zu entscheiden, da zusätzlich zu dem Apriori–Wissen und Meßdaten noch andere Faktoren eine Rolle spielen. Falls ein Experimentator beispielsweise weiß, daß sein Meßgerät für einige Meßbereiche unterschiedliche Meßgenauigkeiten liefert, dann wird er einzelnen Meßwerten mehr Vertrauen schenken als anderen. Deshalb führt man eine Kostenfunktion $l = l(x, y, \hat{f})$ ein, welche die Wahl einer möglichen Hypothese $\hat{f} \in \mathcal{F}$ unter allen möglichen Observablenkombinationen (x, y) bestraft. Diese Funktion ist oft schwer mathematisch zu fassen, da sie unter anderem Entscheidungsmotivationen und psychologische Aspekte enthält.

Für fixierten Systemzustand f , welcher die möglichen Datenpaare (x, y) erzeugt, ist der Erwartungswert von l als Funktion einer möglichen Hypothese \hat{f} – das sogenannte Risiko – gegeben durch:

$$r_f(\hat{f}) = \int_{\mathcal{X}} dx \int_{\mathcal{Y}_x} dy l(x, y, \hat{f}) p[y, x|f]. \quad (1.17)$$

Falls der Stand des Wissens durch $p[f|D]$ gegeben wird, betrachtet man das mittlere Risiko R , den Erwartungswert von r_f unter dieser Verteilung,

$$R(\hat{f}, D) = \int df r_f(\hat{f}) p[f|D] = \int df dx dy l(x, y, \hat{f}) p[f|D] p[y|x, f] p[x], \quad (1.18)$$

unter Ausnutzung von $p[x|f] = p[x]$. Die Entscheidung für eine Hypothese entspricht dann der Minimierung von $R(\hat{f}, D)$ bezüglich \hat{f} :

$$f^* = \operatorname{argmin}_{\hat{f} \in \mathcal{F}} R(\hat{f}, D). \quad (1.19)$$

Man beachte, daß für $l = \text{const}$ sofort $R = \text{const}$ folgt, also keine Entscheidung möglich ist (auch für $p[f|D] \neq \text{const}$). Wird aber zum Beispiel als Bestrafungsfunktion die negative Aposteriori–Wahrscheinlichkeit für \hat{f} nach Messung des einzelnen Datenpaares (x, y) angenommen ($l(x, y, \hat{f}) =$

$-p[\hat{f}|x, y]$), so erhält man unter der Annahme, daß die Verteilung für \hat{f} nicht direkt von f , sondern nur indirekt über (x, y) abhängt ($p[\hat{f}|x, y] = p[\hat{f}|x, y, f]$), und unter Ausnutzung von $p[\hat{f}|f] = p[\hat{f}|f, D]$ letztlich die Entscheidung:

$$\begin{aligned} f^* = \operatorname{argmin}_{\hat{f}} R(\hat{f}, D) &= \operatorname{argmin}_{\hat{f}} \left[- \int df dx dy p[\hat{f}|x, y] p[x, y|f] p[f|D] \right] \\ &= \operatorname{argmin}_{\hat{f}} \left[- \int df p[\hat{f}|f] p[f|D] \right] = \operatorname{argmin}_{\hat{f}} \left[-p[\hat{f}|D] \right] \\ &= \operatorname{argmax}_{\hat{f}} p[\hat{f}|D], \end{aligned} \quad (1.20)$$

was genau der Maximum–Aposteriori–Approximation (1.13) entspricht.

1.5 Hyperparameter

Oft kommt es vor, daß man nur an bestimmten Anteilen h vom Gesamtzustand f interessiert ist, während man über die übrigen Komponenten keine Aussagen machen will. Deshalb teilt man f in zwei Komponenten $f = (h, \beta)$ auf und behandelt diese probabilistisch separat. Zum Beispiel wird der Zustand eines kanonischen Ensembles durch den statistischen Operator $\rho = \exp(-\beta h)/Z$ mit dem Hamilton–Operator h , der inversen Temperatur β und der Zustandssumme Z festgelegt. Möchte man durch Messung am Ensemble Aussagen über $p[h|D]$ gewinnen, so muß man β entweder genau kennen oder zumindest $p[\beta]$ kontrollieren. Für diese Konstellation erhält man statt (1.3) für die Verteilung $p[h|D]$ eine über $p[\beta]$ gemittelte Version:

$$p[h|D] = \int d\beta p[h, \beta|D] = \frac{1}{p[D]} \int d\beta p[D|h, \beta] p[h] p[\beta], \quad (1.21)$$

wobei die Unabhängigkeit $p[h, \beta] = p[h] p[\beta]$ angenommen wurde. Andererseits kann es sein, daß die Apriori–Verteilung für h nur bis auf eine Parametrisierung in α bestimmt ist. Beispielsweise würde bei dem Vorwissen

$$p[h] \Rightarrow p[h|\alpha] \sim e^{-\alpha(h-h_0)^2} \quad (1.22)$$

der Parameter α die mögliche Abweichung von h bezüglich einer Referenz h_0 beschreiben. α hat aber keinen direkten Einfluß auf die Datenerzeugung: $p[D|\alpha, h] = p[D|h]$. Ist $p[\alpha]$ bekannt, so kann die Abhängigkeit von α ohne Kenntnis der Likelihood weggemittelt werden:

$$p[h|D] = \int d\alpha p[h, \alpha|D] = \frac{1}{p[D]} \int d\alpha p[D|\alpha, h] p[\alpha, h] = \frac{1}{p[D]} \int d\alpha p[D|h] p[h|\alpha] p[\alpha]. \quad (1.23)$$

Man beachte, daß im Gegensatz zu (1.21) die zu integrierende Variable im Prior–Term und nicht im Likelihood–Term auftritt. Variablen wie α und β werden Hyperparameter, die Verteilungen $p[\alpha]$ und $p[\beta]$ Hyperprior genannt.

Sind α, β deterministisch bekannt, so wird die Verteilung $p[h|\alpha, \beta, D]$ betrachtet, und man erhält unter Anwendung des Bayesschen Theorems zur Vertauschung der logischen Reihenfolge von h und D :

$$p[h|\alpha, \beta, D] = \frac{p[D|\alpha, \beta, h] p[h|\alpha, \beta]}{p[D|\alpha, \beta]} = \frac{p[D|\beta, h] p[h|\alpha]}{p[D|\alpha, \beta]}, \quad (1.24)$$

wobei die oben erwähnten Abhängigkeiten von den Hyperparametern ausgenutzt wurden.

Möchte man den Einfluß von α und β probabilistisch behandeln, so müssen (1.21) und (1.23)

miteinander verknüpft werden. Mit $p[h, \alpha, \beta] = p[h|\alpha] p[\alpha] p[\beta]$ folgt:

$$\begin{aligned} p[h|D] &= \int d\alpha d\beta p[h, \alpha, \beta|D] = \frac{1}{p[D]} \int d\alpha d\beta p[D|\alpha, \beta, h] p[h, \beta, \alpha] \\ &= \frac{1}{p[D]} \int d\alpha d\beta p[D|\beta, h] p[h|\alpha] p[\alpha] p[\beta]. \end{aligned} \quad (1.25)$$

Als Maximum-Aposteriori-Approximation für Gleichung (1.25) ergibt sich folgende Bedingung:

$$0 = \frac{\partial p[h|D]}{\partial h} \quad \text{und damit} \quad 0 = \int d\alpha d\beta \left[\frac{\partial}{\partial h} p[D|\beta, h] p[h|\alpha] \right] p[\alpha] p[\beta], \quad (1.26)$$

also eine bezüglich $p[\alpha]$ und $p[\beta]$ gemittelte Version von (1.14).

Es ist durch vorherige Messungen von D_0 auch möglich, eine Wahrscheinlichkeitsverteilung für $p[\alpha, \beta|D_0]$ zu konstruieren. So erhält man mit $p[\alpha, \beta] = p[\alpha] p[\beta]$ nach Bayes:

$$\begin{aligned} p[\alpha, \beta|D_0] &= \frac{p[D_0|\alpha, \beta] p[\alpha, \beta]}{p[D_0]} = \frac{p[\alpha] p[\beta]}{p[D_0]} \int dh p[D_0, h|\alpha, \beta] \\ &= \frac{p[\alpha] p[\beta]}{p[D_0]} \int dh p[D_0|h, \alpha, \beta] p[h|\alpha, \beta] = \frac{p[\alpha] p[\beta]}{p[D_0]} \int dh p[D_0|h, \beta] p[h|\alpha]. \end{aligned} \quad (1.27)$$

Gleichung (1.27) enthält ein meist schwer zu handhabendes Integral über den gesamten Zustandsraum von h , und man ist in der Regel auf Näherungen angewiesen. Eine Näherungsstrategie könnte sein, daß h -Integral durch den maximalen Integranden zu nähern. Diese Stellen sind gegeben durch

$$0 = \frac{\partial}{\partial h} (p[D_0|\beta, h] p[h|\alpha]). \quad (1.28)$$

Diese Approximation wird umso besser, je größer die Datenmenge D_0 ist, da die Verteilung $p[D_0|h, \beta]$ umso schärfer lokalisiert ist. Weiter ist zu beachten, daß ein nach (1.28) ausgewählter Zustand α - und β -abhängig wäre. Führt man gleichzeitig zu (1.28) in (1.27) eine partielle Ableitung nach den Hyperparametern durch, so erhält man die zusätzlichen Bedingungen:

$$0 = \frac{\partial}{\partial \alpha} (p[\alpha] p[h|\alpha]) \quad , \quad 0 = \frac{\partial}{\partial \beta} (p[\beta] p[D_0|\beta, h]), \quad (1.29)$$

welche simultan mit (1.28) gelöst werden müssen. Die stationären Werte $\tilde{\alpha}$ und $\tilde{\beta}$ sind so näherungsweise an das Vorwissen und gleichzeitig an D_0 angepaßt. Am Ende dieser Näherungen für (1.27) steht dann

$$p[\alpha, \beta|D_0] \approx \delta(\alpha - \tilde{\alpha}) \delta(\beta - \tilde{\beta}). \quad (1.30)$$

Faßt man (1.25) und (1.27) zusammen, wobei nur D_0 direkten Einfluß auf die Hyperparameter hat, so faktorisieren die Wahrscheinlichkeiten folgendermaßen:

$$\begin{aligned} p[h|D, D_0] &= \int d\alpha d\beta p[h, \alpha, \beta|D, D_0] = \int d\alpha d\beta p[h|\alpha, \beta, D, D_0] p[\alpha, \beta|D, D_0] \\ &= \int d\alpha d\beta p[h|\alpha, \beta, D] p[\alpha, \beta|D_0]. \end{aligned} \quad (1.31)$$

Indem man sich gemäß (1.30) für ein Paar $\tilde{\alpha}, \tilde{\beta}$ festlegt, bekommt man ähnlich zu (1.24)

$$p[h|D, D_0] \approx p[h|\tilde{\alpha}, \tilde{\beta}, D]. \quad (1.32)$$

1.6 Algorithmen

Die Maximum–Aposteriori–Approximation entspricht nach (1.13) dem Maximieren der Aposteriori–Wahrscheinlichkeit $p[f|D]$ beziehungsweise dem Minimieren der zugehörigen Energiefunktion $F[f] = -\ln p[f|D]$ bezüglich aller betrachteten Zustände f , wobei f ein möglicherweise hochdimensionaler Vektor ist. Das Problem lautet also

$$f^* = \operatorname{argmin}_f F[f]. \quad (1.33)$$

In der Regel wird die Abhängigkeit der Funktion F von f nicht so einfach sein, daß sich die Stellen, an denen der Gradient $\nabla_f F[f]$ verschwindet, analytisch angeben lassen. Versucht man numerisch Lösungen von (1.33) zu finden, so wird man mit einer Vermutung $f^{(0)}$ für den gesuchten Zustand starten, in der Hoffnung, daß dieser so nahe bei f^* liegt, daß zwischen diesen beiden Zuständen die Energiefunktion kein lokales Maximum besitzt. Berechnet man den Gradienten $\nabla_f F[f^{(0)}]$, so zeigt dieser Vektor in Richtung des steilsten Anstiegs der Energiefunktion F an dieser Stelle. Es kann durch Taylor–Entwicklung gezeigt werden, daß dann eine hinreichend kleine reelle Zahl $\eta > 0$ existiert, so daß ein Schritt zu einem neuen Zustand in die entgegengesetzte Richtung $-\nabla_f F[f^{(0)}]$ zu einem kleineren Energiewert führt. Wiederholt man dieses Vorgehen, so wandert man auf der Energieoberfläche abwärts zum nächsten Minimum. Der Algorithmus für den i -ten Iterationsschritt besitzt dann die Form

$$f^{(i+1)} = f^{(i)} - \eta^{(i)} \nabla_f F[f^{(i)}] \quad \text{mit } \eta^{(i)} > 0 \quad \Rightarrow \quad F[f^{(i+1)}] \leq F[f^{(i)}] \quad (1.34)$$

und wird bis zur Konvergenz $f^{(i+1)} = f^{(i)}$ wiederholt. Die Schrittweite $\eta^{(i)}$ kann für jeden Iterationsschritt nach der Regel

$$\eta^{(i)} = \operatorname{argmin}_{\eta > 0} F[f^{(i)} - \eta \nabla_f F[f^{(i)}]] \quad (1.35)$$

angepaßt werden. (1.34) wird Gradientenabstiegsverfahren genannt. Die Vorteile liegen in seiner Einfachheit, die Nachteile in der meist schlechten Konvergenz im Vergleich zu aufwendigeren Verfahren.

Das Gradientenabstiegsverfahren kann verallgemeinert werden, indem der Vektor $\nabla_f F[f]$ durch einen neuen Vektor $\mathcal{O} \nabla_f F[f]$ mit einer beliebigen positiv–definiten, symmetrischen Matrix \mathcal{O} ersetzt wird. Die an \mathcal{O} gestellten Bedingungen gewährleisten, daß der Vektor $-\mathcal{O} \nabla_f F[f]$ weiterhin in Richtung des Abstiegs von F zeigt. Der abgeänderte Algorithmus besitzt dann die Struktur

$$f^{(i+1)} = f^{(i)} - \eta^{(i)} \mathcal{O}^{(i)} \nabla_f F[f^{(i)}] \quad \text{mit } \eta^{(i)} > 0 \quad \Rightarrow \quad F[f^{(i+1)}] \leq F[f^{(i)}]. \quad (1.36)$$

Bei der Iteration nach der Newton–Methode wählt man für $\mathcal{O}^{(i)}$ die inverse Hesse–Matrix an der Stelle $f^{(i)}$:

$$\mathcal{O}^{(i)} = \left(\nabla_f \nabla_{f'} F[f^{(i)}] \right)^{-1}, \quad (1.37)$$

vorausgesetzt sie ist invertierbar und positiv–definit. Dieses ist aber in der Nähe des Minimums f^* zu erwarten. In jedem Iterationsschritt wird F durch eine quadratische Approximation genähert und bezüglich dieser ein Iterationsschritt vorgenommen. Dadurch wird eine gegenüber dem Gradientenabstiegsverfahren bessere Konvergenz erreicht. Die Nachteile liegen in dem verhältnismäßig hohen Rechenaufwand bei der Inversion aus (1.37). Ist die jeweilige Bestimmung von (1.37) zu umständlich, so kann man versuchen mit einer genäherten Matrix \mathcal{O} zu iterieren, beispielsweise indem man nur $\mathcal{O}^{(0)}$ anwendet oder indem man sich bei der Bestimmung der Hesse–Matrix auf die Diagonalelemente $\nabla_f^2 F$ beschränkt.

Für eine detaillierte Darstellung der numerischen Behandlung nichtlinearer Gleichungen ist das Buch von Bertsekas [9] zu empfehlen.

Kapitel 2

Lernen von zeitabhängigen Wahrscheinlichkeitsdichten

2.1 Formulierung in Wahrscheinlichkeitsdichten

Dieses Kapitel handelt von Regressions- und Dichteapproximationsproblemen aus der Sicht der Bayesschen Theorie. Eine grundlegende Behandlung für die zeitunabhängige Dichteapproximation wird in [10] erbracht. Die Ergebnisse dieser Veröffentlichung werden in diesem Kapitel durch Betrachtung von sich zeitlich ändernden Dichteverteilungen und dem Einbau einer Regressionsfunktion generalisiert.

Ein Problem könnte beispielsweise sein, die Trajektorie $X(t)$ eines Teilchens zu rekonstruieren. Liegen für das Teilchen Ortsmessungen $D = \{(x_i, t_i) | i = 1, \dots, N\}$ vor, so geht es bei der Regressionstheorie darum, eine Funktion $X(t)$ oder eine Verteilung $p[X|D]$ zu berechnen, welche den "Informationsgehalt" der Daten unter vorgegebenen Gesichtspunkten optimal wiedergibt. Handelt es sich um ein quantenmechanisches Teilchen oder möchte man der endlichen Meßgenauigkeit Rechnung tragen, so kann man diese Unsicherheit in den Formalismus einbauen, indem man eine Wahrscheinlichkeitsdichte $Q(x, t)$ für das Teilchen einführt und das Problem dadurch künstlich erweitert. Durch Einführung eines Hilfsintegrals über die Dichten $Q(x, t)$ erhält man für die Aposteriori-Wahrscheinlichkeit

$$p[X|D] = \int [dQ(x, t)] p[X, Q|D] = \int [dQ(x, t)] p[X|Q, D] p[Q|D]. \quad (2.1)$$

Unter der obigen Interpretation für $Q(x, t)$ kann für die Abhängigkeiten in den bedingten Wahrscheinlichkeiten

$$p[X|Q, D] = p[X|Q] \quad (2.2)$$

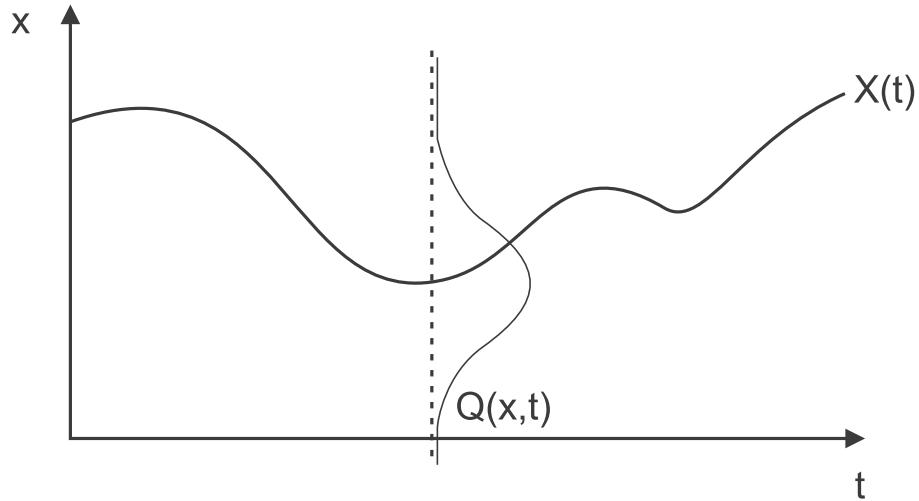
ausgenutzt werden. Wendet man in einem zweiten Schritt das Bayes-Theorem für

$$p[Q|D] = \frac{p[D|Q] p[Q]}{p[D]} \quad (2.3)$$

an und setzt die beiden Ergebnisse in (2.1) ein, so erhält man:

$$p[X|D] = \frac{1}{p[D]} \int [dQ(x, t)] p[Q] p[D|Q] p[X|Q]. \quad (2.4)$$

Das Funktionalintegral über alle möglichen Dichten $Q(x, t)$ aus (2.4) wird in der Regel nicht analytisch ausführbar sein. Die Form von (2.4) ist aber anwendungsfreundlich gegenüber Näherungen.

Abbildung 2.1: Erläuterung der Größen $X(t)$ und $Q(x, t)$.

Im weiteren Verlauf werden Annahmen für die einzelnen Wahrscheinlichkeiten aus (2.4) gemacht und das Funktionalintegral mittels Stationärer Phasenapproximation bearbeitet.

Die Wahrscheinlichkeit für die Daten $p[D|Q]$ ist durch die Definition von Q durch

$$p[D|Q] = \prod_{i=1}^N Q(x_i, t_i) = e^{\sum_{i=1}^N \ln Q(x_i, t_i)} \quad (2.5)$$

vorgegeben. Freiheiten hat man bei der Wahl der bedingten Wahrscheinlichkeit $p[X|Q]$. Im wesentlichen erwartet man, daß der Ortserwartungswert unter der Verteilung $Q(x, t)$ mit der Regression $X(t)$ zusammenfällt. Diese kann über eine strikte Forderung¹

$$p[X|Q] = \prod_t \delta\left(X(t) - \langle x(t) \rangle_Q\right) = \prod_t \delta\left(X(t) - \int dx x Q(x, t)\right) \quad (2.6)$$

implementiert werden. Eine schwächere Variante ist eine Gauß-Verteilung

$$p[X|Q] = \frac{1}{Z} e^{-\frac{1}{2\sigma^2} \int dt \left(X(t) - \langle x(t) \rangle_Q\right)^2} \quad (2.7)$$

mit der Normkonstanten

$$Z = \int [dX(t)] e^{-\frac{1}{2\sigma^2} \int dt \left(X(t) - \langle x(t) \rangle_Q\right)^2} \quad (2.8)$$

und vorgegebener Standardabweichung σ . Man beachte, daß die Normierung Z aus (2.8) unabhängig von $Q(x, t)$ und $\langle x \rangle_Q$ ist.

¹In diesem Abschnitt werden die Variablen x und t aus Gründen der Notation kontinuierlich ausgeschrieben, obwohl sie im Hinblick auf numerische Anwendungen diskret gedacht sind. So steht das Symbol \prod_t für ein endliches Produkt über eine diskretisierte Zeitachse, das Integral $\int dt$ für die diskrete Summe \sum_t und das Pfadintegral $\int [dX(t)]$ für die Summation über alle möglichen Pfade $X(t)$ des Gitters. Ableitungsoperationen $\frac{\partial}{\partial t}$ können durch die in Kapitel 5.4 genannten Matrizen mit der Gitterkonstanten $\Delta t = 1$ implementiert werden.

Als weiteres Beispiel sei hier die Verteilung

$$p[X|Q] = \prod_t Q(X(t), t) \quad (2.9)$$

angegeben. (2.9) entspricht im Vergleich zu (2.5) der Wahrscheinlichkeit, die einzelnen Funktionswerte der Regressionsfunktion unabhängig voneinander zu messen. Alle drei Wahrscheinlichkeitsverteilungen (2.6), (2.7) und (2.9) sind korrekt normiert, erfüllen also die erforderliche Bedingung $1 = \int [dX(t)] p[X|Q]$.

Das Vorwissen für die Wahrscheinlichkeitsdichte Q wird in $p[Q]$ wiedergegeben. Dabei kann zwischen den notwendigen Bedingungen, die an Q gestellt werden – wie die Forderung $Q(x, t) \geq 0$ für alle (x, t) und die Normierungsbedingung $\int dx Q(x, t) = 1$ für alle t – und problemspezifischem Wissen, wie zum Beispiel die Erwartung, daß Q sowohl in der Orts- als auch in der Zeitkoordinate glatt ist, unterschieden werden. So wird in den Anwendungen mit einer Verteilung der Form

$$p[Q] = \frac{1}{Z_0} e^{-\Gamma_0[Q]} \left[\prod_t \delta \left(1 - \int dx Q(x, t) \right) \right] \left[\prod_{x t} \theta(Q(x, t)) \right] \quad (2.10)$$

gearbeitet. Mittels des Energiefunktionalen Γ_0 kann die Erwartung großer Glattheit durch

$$\Gamma_0[Q] = \frac{1}{2\sigma_1^2} \int dt dx (\nabla_x Q(x, t))^2 + \frac{1}{2\sigma_2^2} \int dt dx \left(\frac{\partial Q(x, t)}{\partial t} \right)^2 \quad (2.11)$$

ausgedrückt werden. Eine andere Möglichkeit besteht in der Verwendung von

$$\Gamma_0[Q] = k \int dt dx Q(x, t) \ln Q(x, t) \quad (2.12)$$

und entspricht der Forderung nach einer großen Entropie von Q .

Die δ - und θ -Funktionen aus (2.10) können über die Integraldarstellungen

$$\prod_t \delta \left(1 - \int dx Q(x, t) \right) = \int \frac{[d\lambda(t)]}{\mathcal{N}} e^{i \int dt \lambda(t) (1 - \int dx Q(x, t))} \quad (2.13)$$

und

$$\prod_{x t} \theta(Q(x, t)) = \int \frac{[ds(x, t)]}{\mathcal{N}'} e^{\int dx dt (i s(x, t) Q(x, t) - \ln(s(x, t) - i\kappa))} \quad (2.14)$$

implementiert werden. Die Einführung der Hilfsintegrale macht die unangenehmen Bedingungen analytisch handhabbar. Die Integrale werden im weiteren Verlauf durch Stationäre Phasenapproximation behandelt. Im Falle von (2.13) ist diese äquivalent zu der Methode der Lagrange-Multiplikatoren.

Für den Fall, daß eine sehr große Zahl von Daten vorliegt oder daß das Vorwissen die möglichen Dichten Q stark einschränkt, ist die Verteilung $p[Q|D]$ scharf an einer Stelle \tilde{Q} lokalisiert. Diese Kenntnis kann dafür genutzt werden, das Integral aus (2.1) zu nähern, indem $p[X|Q]$ an der Stelle \tilde{Q} herausgezogen wird:

$$p[X|D] \approx p[X|\tilde{Q}] \int [dQ] p[Q|D] = p[X|\tilde{Q}]. \quad (2.15)$$

Beim Lernen der Regressionsfunktion besteht das Problem also darin, durch eine Maximum-Aposteriori-Approximation von (2.3) die Dichte \tilde{Q} zu finden, welche $p[Q|D]$ maximiert. Somit kann mit (2.15) das Regressionsproblem im wesentlichen als ein Dichteapproximationsproblem angesehen werden. Im Falle von (2.6) folgt ein deterministischer Zusammenhang von Regressionsfunktion $X(t)$ und Dichte $Q(x, t)$:

$$p[X|D] \approx p[X|\tilde{Q}] = \prod_t \delta \left(X(t) - \langle x(t) \rangle_{\tilde{Q}} \right). \quad (2.16)$$

2.2 Formulierung in logarithmischen Wahrscheinlichkeiten

Die Hauptproblematik beim Arbeiten mit Wahrscheinlichkeitsdichten $Q(x, t)$ ist die analytische Implementierung der Positivität $Q(x, t) \geq 0$. Eine günstige Art, dies schon von vornherein zu implementieren, ist die Einführung von logarithmischen Wahrscheinlichkeiten nach

$$\phi(x, t) = -\ln Q(x, t) \quad \text{oder} \quad Q(x, t) = e^{-\phi(x, t)}. \quad (2.17)$$

Die Vorzeichen für ϕ wurde dabei so gewählt, daß ϕ gleichzeitig die Interpretation einer Energie erhält: Für kleine ϕ ist die dazugehörige Wahrscheinlichkeitsdichte Q groß. Durch den positiven Wertebereich der Exponentialfunktion wird das positive Vorzeichen von Q erzwungen, und ϕ muß als notwendige Bedingung nur die Normen

$$\int dx e^{-\phi(x, t)} = 1 \quad \forall t \quad (2.18)$$

erfüllen. Schreibt man die Wahrscheinlichkeiten des vorherigen Abschnitts auf die logarithmischen Wahrscheinlichkeiten um, so erhält man für die Apriori-Wahrscheinlichkeit von ϕ analog zu (2.10) eine Verteilung der Form

$$p[\phi] = \frac{1}{Z_0} e^{-\Gamma_0[\phi]} \left[\prod_t \delta \left(1 - \int dx e^{-\phi(x, t)} \right) \right]. \quad (2.19)$$

Wählt man ein Energiefunktional nach

$$\Gamma_0[\phi] = \frac{1}{2\sigma_1^2} \int dt dx (\nabla_x \phi(x, t))^2 + \frac{1}{2\sigma_2^2} \int dt dx \left(\frac{\partial \phi(x, t)}{\partial t} \right)^2 \quad (2.20)$$

und bestraft somit die Abweichung von der Glattheit für ϕ , so ist dies ein prinzipieller Unterschied gegenüber (2.11). Denn wegen

$$\int dx dt \left(\frac{\partial}{\partial t} \ln Q(x, t) \right)^2 = \int dx dt \frac{1}{Q^2(x, t)} \left(\frac{\partial Q(x, t)}{\partial t} \right)^2 \quad (2.21)$$

wird im Vergleich zu (2.11) an den Stellen (x, t) stärker bestraft, an denen $Q(x, t)$ besonders klein ist. Dies ist ein durchaus willkommener Effekt, denn die Stellen kleiner Wahrscheinlichkeitsdichte sind in größerem Abstand von den Datenpunkten lokalisiert. Dort sind Abweichungen von der Glattheit am wenigsten zu erwarten.

Die Likelihood-Wahrscheinlichkeit für die Daten $p[D|\phi]$ bekommt mit (2.17) die Form

$$\prod_{i=1}^N p[(x_i, t_i)|\phi] = e^{-\sum_{i=1}^N \phi(x_i, t_i)}, \quad (2.22)$$

und der Wahrscheinlichkeit für die Regressionsfunktion bei gegebener Dichte $p[X|Q]$ aus (2.9) entspricht der Ausdruck

$$p[X|\phi] = \prod_t e^{-\phi(X(t), t)} = e^{-\int dt \phi(X(t), t)}. \quad (2.23)$$

Man beachte, daß beim Arbeiten mit der Dichte Q in den Exponenten der Normierungsbedingung nach (2.13) Q linear und in der Likelihood-Dichte (2.5) nichtlinear auftritt. Die Rollen sind in der Darstellung mittels ϕ nach (2.18) und (2.22) genau vertauscht. Dies hat im weiteren Verlauf

Konsequenzen bei der Durchführung einer Maximum–Aposteriori–Approximation. Der nichtlineare Charakter der Exponenten führt auf nichtlineare Differentialgleichungen.

Zuletzt sei noch die Möglichkeit erwähnt, die Positivität von Q zu erzwingen, indem mit reellen $\varphi(x, t)$ als Wurzel von Q gearbeitet wird:

$$\pm \sqrt{Q(x, t)} = \varphi(x, t) \quad \text{oder} \quad \varphi^2(x, t) = Q(x, t), \quad (2.24)$$

mit der Normbedingung

$$\int dx \varphi^2(x, t) = 1 \quad \forall t. \quad (2.25)$$

Es zeigt sich, daß die Wahl von (2.24) sich besonders für den Einbau von Mustern und das Arbeiten mit Überlappen eignet.

2.3 Behandlung von Mustern und Korrelationen

Oft ist zu erwarten, daß die Regressionsfunktion $X(t)$ zusätzlich einem oder mehreren gegebenen Mustern ähnlich ist oder daß aus einer Menge von Mustern eines zu $X(t)$ paßt. Als Abkürzung für die Musterfunktionen wird hier die Menge $\Lambda = \{\tau_j(t) | j = 1, \dots, s\}$ eingeführt. Genauso können Muster für die Wahrscheinlichkeitsdichte $Q(x, t)$ vorhanden sein. Zur Unterscheidung werden diese mit

$$\Lambda = \{\chi_j(x, t) | \chi_j(x, t) \geq 0 \forall x, t ; \int dx \chi_j(x, t) = 1 \forall t ; j = 1, \dots, s\} \quad (2.26)$$

gekennzeichnet. Der Index j kann dabei sowohl eine diskrete als auch eine kontinuierliche Menge parametrisieren. Die für das Lernen interessante Verteilung ist dann die Wahrscheinlichkeit für die Regressionsfunktion bei gegebenen Daten und Mustern $p[X|D, \Lambda]$.

Geht man analog zu (2.1) vor und führt die Dichten $Q(x, t)$ mit der gleichen Interpretation ein, so erhält man

$$\begin{aligned} p[X|D, \Lambda] &= \int [dQ(x, t)] p[X, Q|D, \Lambda] \\ &= \int [dQ(x, t)] p[X|Q, D, \Lambda] p[Q|D, \Lambda], \end{aligned} \quad (2.27)$$

wobei hier erneut

$$p[X|Q, D, \Lambda] = p[X|Q] \quad (2.28)$$

ausgenutzt werden kann. Anwendung des Bayes–Theorems zur Vertauschung der Abhängigkeiten von Q und D liefert

$$p[Q|D, \Lambda] = \frac{p[D|Q, \Lambda] p[Q|\Lambda]}{p[D|\Lambda]} = \frac{p[D|Q] p[Q|\Lambda]}{p[D|\Lambda]}, \quad (2.29)$$

und damit bekommt man für (2.27) eine Verallgemeinerung von (2.4):

$$p[X|D, \Lambda] = \frac{1}{p[D|\Lambda]} \int [dQ(x, t)] p[X|Q] p[D|Q] p[Q|\Lambda]. \quad (2.30)$$

In der Regel wird die Menge Λ nur aus einem Muster $\chi(x, t)$ bestehen, und es wird erwartet, daß die Verteilung $Q(x, t)$ diesem "ähnlich" ist. Die bedingte Wahrscheinlichkeit $p[Q|\Lambda]$ kann dann analog zu (2.10) von der Form

$$p[Q|\Lambda] \sim e^{-(\Gamma_0[Q] + \Gamma[Q, \chi])} \left[\prod_t \delta \left(1 - \int dx Q(x, t) \right) \right] \left[\prod_{x t} \theta(Q(x, t)) \right] \quad (2.31)$$

gewählt werden. In (2.31) wird zwischen einem χ -unabhängigen Anteil Γ_0 , zum Beispiel (2.11), und einem χ -abhängigen Anteil $\Gamma[Q, \chi]$ im Energiefunktional unterschieden. Als Funktional $\Gamma[Q, \chi]$ bietet sich beispielsweise der Abstand der beiden Ortserwartungswerte unter den jeweiligen Verteilungen an:

$$\Gamma[Q, \chi] = \frac{\gamma}{2} \int dt \left(\langle x(t) \rangle_Q - \langle x(t) \rangle_\chi \right)^2 = \frac{\gamma}{2} \int dt \left(\int dx x (Q(x, t) - \chi(x, t)) \right)^2 \quad (2.32)$$

und im Fall von einem scharfen Muster $\chi(x, t) = \delta(\tau(t) - x)$

$$\Gamma[Q, \tau] = \frac{\gamma}{2} \int dt \left(\langle x(t) \rangle_Q - \tau(t) \right)^2. \quad (2.33)$$

Die Energiefunktionale (2.32) und (2.33) alleine sind für die Beschreibung des Apriori-Wissens wenig geeignet, da sie kaum Information über die Form der Verteilung Q beinhalten. So gibt es eine große Zahl sehr unterschiedlicher Verteilungen Q , die denselben Erwartungswert $\langle x(t) \rangle_Q$ besitzen. Diese Freiheit kann eingeschränkt werden, indem zusätzlich eine kleine Varianz für die Verteilung mittels

$$\int dt \left(\langle x^2(t) \rangle_Q - \langle x(t) \rangle_Q^2 \right) = \int dt dx Q(x, t) x^2 - \int dt \left(\int dx Q(x, t) x \right)^2 \quad (2.34)$$

gefordert wird. Eine zweite Möglichkeit ist, die Funktionswerte der Dichten für alle Orte und Zeiten zu vergleichen:

$$\Gamma[Q, \chi] = \frac{\gamma}{2} \int dt dx \left(Q(x, t) - \chi(x, t) \right)^2. \quad (2.35)$$

Motiviert durch (2.9) und (2.23) bietet sich ein Verwendung von

$$\Gamma[Q, \tau] = \gamma \int dt \ln Q(\tau(t), t) \quad (2.36)$$

an, und bei der Darstellung mittels logarithmischer Wahrscheinlichkeiten bekommt man schließlich

$$\Gamma[\phi, \tau] = \gamma \int dt \phi(\tau(t), t). \quad (2.37)$$

Sind mehrere Muster vorhanden und soll Q sämtlichen Mustern χ_j gleichzeitig ähnlich sehen, so entspricht dies einer logischen AND-Verknüpfung der Muster. Die Aposteriori-Wahrscheinlichkeit $p[Q|\Lambda]$ ist für solche Dichten Q groß, bei denen die Merkmale der einzelnen Muster im Mittel auftreten. Dies leistet eine Verteilung der Form

$$p[Q|\Lambda] \sim e^{-\sum_{j=1}^s \Gamma_j[Q, \chi_j]} \left[\prod_t \delta \left(1 - \int dx Q(x, t) \right) \right] \left[\prod_x \theta(Q(x, t)) \right] \quad (2.38)$$

mit den einzelnen Energiefunktionalen Γ_j .

Arbeitet man wie in (2.24) mit reellen $\xi_j(x, t) = \pm \sqrt{\chi_j(x, t)}$ und $\varphi(x, t) = \pm \sqrt{Q(x, t)}$, so können die zeitabhängigen Überlape von Muster und Verteilung Q

$$\alpha_j(t) = \int dx \xi_j(x, t) \varphi(x, t) \quad (2.39)$$

definiert werden. Die Verwendung von $\alpha(t)$ hat den Vorteil, daß durch die Normbedingungen für Q und χ_j der Wertebereich von $\alpha_j(t)$ auf das Intervall $[-1, 1]$ beschränkt wird. Umgekehrt folgt aus

dem Fall $\alpha_j(t_0) = \pm 1$ für eine feste Zeit t_0 sofort die Gleichheit von Dichteverteilung und Muster $Q(x, t_0) = \chi_j(x, t_0)$ für alle x . Somit kann entsprechend (2.31) mit (2.35) für ein Muster ξ in dieser Darstellung

$$p[\varphi|\xi] \sim e^{-\frac{1}{2\sigma^2} \int dt (1-\alpha^2(t))^2} \prod_t \delta\left(1 - \int dx \varphi^2(x, t)\right) \quad (2.40)$$

gewählt werden.

Oft erwartet man, daß nur genau eines der gegebenen Muster zu der Verteilung Q paßt. Dies entspricht einer logischen OR-Verknüpfung der Muster. Definiert man die zeitlichen Mittelwerte der Überlappe von φ und einzelnen Mustern ξ_i und der Muster untereinander:

$$\bar{\alpha}_i = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T dt \alpha_i(t) \in [-1, 1] \quad \bar{\beta}_{ij} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T dt dx \xi_i(x, t) \xi_j(x, t) \in [-1, 1], \quad (2.41)$$

so können Energiefunktionale in diesen Überlappen für den OR-Fall konstruiert werden. Man beachte, daß aus $\bar{\alpha}_i = \pm 1$ sofort $Q(x, t) = \chi_i(x, t)$ für alle (x, t) folgt. Als direkte Implementierung der OR-Verknüpfung bietet sich, motiviert durch (2.40), eine Bestrafung der Form

$$\Gamma(\bar{\alpha}_1, \dots, \bar{\alpha}_s) = -\ln \left(\sum_{j=1}^s e^{-(1-\bar{\alpha}_j^2)^2} \right) \quad (2.42)$$

an. Ein weiteres Beispiel ist eine Produktform nach

$$\Gamma(\bar{\alpha}_1, \dots, \bar{\alpha}_s) = \prod_{j=1}^s (1 - \bar{\alpha}_j^2)^2. \quad (2.43)$$

Der Nachteil von (2.42) und (2.43) liegt in der schlechten analytischen und numerischen Handhabbarkeit und im möglicherweise hochdimensionalen, gemischten Polynom in $\bar{\alpha}_j$. Die Energiefunktionale bestrafen, wenn der Überlap $\xi_1-\varphi$ oder $\xi_2-\varphi$ oder \dots oder $\xi_s-\varphi$ nicht gleich Eins ist. Eine ähnliche Eigenschaft kann konstruiert werden, indem man bestraft, wenn der Überlap $\xi_1-\varphi$ weder groß noch klein ist und \dots und der Überlap $\xi_s-\varphi$ weder groß noch klein ist. Ein solches Funktional ist

$$\Gamma(\bar{\alpha}_1, \dots, \bar{\alpha}_s) = \sum_{j=1}^s \bar{\alpha}_j^2 (1 - \bar{\alpha}_j^2)^2 = \begin{cases} 0 & \text{falls ein } \bar{\alpha}_j = \pm 1 \text{ und } \bar{\beta}_{ij} = \delta_{ij} \\ 0 & \text{falls alle } \bar{\alpha}_j = 0 \\ \frac{4s}{27} = \max & \text{falls alle } \bar{\alpha}_j = \pm \frac{1}{\sqrt{3}} \end{cases}. \quad (2.44)$$

Im Unterschied zu (2.42) und (2.43) wird nicht mehr bestraft, wenn alle Überlappe mit den Mustern verschwinden. Der Vorteil von (2.44) liegt in der logischen AND-Verknüpfung. Es treten keine gemischten Terme in $\bar{\alpha}_j$ auf. Die Wahl von (2.44) ist aber ungeschickt für den Fall, daß sich die Muster untereinander überlappen ($\bar{\beta}_{ij} \neq \delta_{ij}$), denn dann würde unter Umständen die Bestrafung für den Fall, daß für ein $\bar{\alpha}_i = \pm 1$ gilt, nicht verschwinden. Dies kann korrigiert werden, indem

$$\Gamma(\bar{\alpha}_1, \dots, \bar{\alpha}_s, \bar{\beta}_{ij}) = \sum_{i=1}^s \bar{\alpha}_i^2 \prod_{j=1}^s (\bar{\beta}_{ij}^2 - \bar{\alpha}_i^2)^2 = \begin{cases} 0 & \text{falls mindestens ein } \bar{\alpha}_i = \pm 1 \\ 0 & \text{falls alle } \bar{\alpha}_i = 0 \\ > 0 & \text{sonst} \end{cases} \quad (2.45)$$

gewählt wird, denn für den Fall $\bar{\alpha}_k = \pm 1$ für einen der k -Werte folgt mit $\varphi = \pm \xi_k$ und der Definition von (2.41): $\bar{\alpha}_i^2 = \bar{\beta}_{ik}^2$ für alle i . Das heißt, jeder Summand aus (2.45) verschwindet, da in dem Produkt der $(k=j)$ -Term identisch Null ist.

Als Spezialfall von (2.45) für orthogonale Muster $\bar{\beta}_{ij} = \delta_{ij}$ erhält man mit

$$\Gamma(\bar{\alpha}_1, \dots, \bar{\alpha}_s) = \sum_{i=1}^s \bar{\alpha}_i^2 \prod_{j=1}^s (\delta_{ij} - \bar{\alpha}_i^2)^2 = \sum_{i=1}^s \bar{\alpha}_i^2 (1 - \bar{\alpha}_i^2)^2 \prod_{j \neq i} (\delta_{ij} - \bar{\alpha}_i^2)^2 = \sum_{i=1}^s \bar{\alpha}_i^{2s} (1 - \bar{\alpha}_i^2)^2 \quad (2.46)$$

ein gegenüber (2.44) leicht abgeändertes Ergebnis.

2.4 Anwendungen für die Dichteapproximation

In diesem Abschnitt wird der Bayessche Formalismus für die zeitabhängige Dichteapproximation an unterschiedlichen Beispielen angewandt. In einem ersten Beispiel wird die Formulierung in logarithmischen Wahrscheinlichkeiten $\phi(x, t)$ gewählt. Es wird davon ausgegangen, daß zusätzlich zu den Meßdaten D ein scharfes Muster $\tau(t)$ vorliegt. Für die Aposteriori-Wahrscheinlichkeit gilt dann:

$$p[\phi|D, \tau] = \frac{p[D|\phi] p[\phi|\tau]}{p[D|\tau]}, \quad (2.47)$$

und die Verteilung $p[\phi|\tau]$ soll eine zu (2.31) analoge Form

$$p[\phi|\tau] \sim e^{-(\Gamma_0[\phi] + \Gamma[\phi, \tau])} \left[\prod_t \delta \left(1 - \int dx e^{-\phi(x, t)} \right) \right] \quad (2.48)$$

mit Γ_0 aus (2.20) besitzen. Es werden zwei Möglichkeiten durchgespielt, das Muster in die Rechnung einzubauen. Die erste Variante besteht darin, daß Muster mittels (2.37) einzuführen. Eine Alternative dazu ist die Verwendung von (2.33) und (2.34). Unter diesen beiden Annahmen bekommt man für die Aposteriori-Wahrscheinlichkeit

$$p[\phi|D, \tau] \sim \int [d\lambda(t)] e^{-F[\phi, \lambda, D, \tau]} \quad (2.49)$$

mit dem Energiefunktional

$$\begin{aligned} F[\phi, \lambda, D, \tau] = & -i \int dt \lambda(t) \left(1 - \int dx e^{-\phi(x, t)} \right) + \sum_{i=1}^N \phi(x_i, t_i) \\ & + \frac{1}{2\sigma_1^2} \int dt dx \left((\nabla_x \phi(x, t))^2 + \zeta (\Delta_x \phi(x, t))^2 \right) \\ & + \frac{1}{2\sigma_2^2} \int dt dx \left(\left(\frac{\partial \phi(x, t)}{\partial t} \right)^2 + \zeta \left(\frac{\partial^2 \phi(x, t)}{\partial t^2} \right)^2 \right) \\ & + \begin{cases} \gamma \int dt \phi(\tau(t), t) \\ \frac{\gamma}{2} \int dt \left(\langle x(t) \rangle_\phi - \tau(t) \right)^2 + \frac{\vartheta}{2} \int dt \left(\langle x^2(t) \rangle_\phi - \langle x(t) \rangle_\phi^2 \right) \end{cases} . \end{aligned} \quad (2.50)$$

Dabei wurde die Normierungsbedingung in der integralen Darstellung analog zu (2.13) verwendet. Man beachte, daß zusätzlich zur Glattheit von ϕ eine mit ζ gewichtete Glattheit der Ableitungen von ϕ gefordert wird. Die Motivation hierfür ist die Vermeidung von Unstetigkeiten der Ableitungen von ϕ in den Datenpunkten und damit eine abgerundetere Wahrscheinlichkeitsverteilung als Lösung.

Die Glattheitsterme aus (2.50) können mittels partieller Integration bearbeitet werden. Unter Voraussetzung von zyklischen Randbedingungen für ϕ in Orts- und Zeitkoordinaten verschwinden die Randterme, und es folgt:

$$\int dt \left(\frac{\partial \phi(x, t)}{\partial t} \right)^2 = - \int dt \phi(x, t) \frac{\partial^2 \phi(x, t)}{\partial t^2} , \quad \int dt \left(\frac{\partial^2 \phi(x, t)}{\partial t^2} \right)^2 = \int dt \phi(x, t) \frac{\partial^4 \phi(x, t)}{\partial t^4} . \quad (2.51)$$

Führt man nun eine Maximum–Aposteriori–Approximation für die Verteilungen (2.49) unter gleichzeitiger Sattelpunktsnäherung für das λ –Integral durch, so wird die Normbedingung

$$1 = \int dx e^{-\phi(x,t)} \quad \forall t \quad (2.52)$$

und die Stationaritätsgleichung

$$\begin{aligned} 0 = \frac{\delta F[\phi, \lambda, D, \tau]}{\delta \phi(x, t)} &= \frac{1}{\sigma_1^2} \left(-\Delta_x \phi(x, t) + \zeta \Delta_x^2 \phi(x, t) \right) + \frac{1}{\sigma_2^2} \left(-\frac{\partial^2 \phi(x, t)}{\partial t^2} + \zeta \frac{\partial^4 \phi(x, t)}{\partial t^4} \right) \\ &\quad - i\lambda(t) e^{-\phi(x, t)} + \sum_{i=1}^N \delta(x - x_i) \delta(t - t_i) \\ &\quad + \begin{cases} \gamma \delta(x - \tau(t)) \\ \gamma e^{-\phi(x, t)} x \left(\tau(t) - \langle x(t) \rangle_\phi \right) + \frac{\vartheta}{2} e^{-\phi(x, t)} x \left(2 \langle x(t) \rangle_\phi - x \right) \end{cases} \end{aligned} \quad (2.53)$$

erhalten. Die Lagrange–Parameter $\lambda(t)$ können eliminiert werden, indem in Gleichung (2.53) die Ortsvariable ausintegriert und die Normbedingung ausgenutzt wird. Dies führt mit den zyklischen Randbedingungen auf

$$\begin{aligned} -i\lambda(t) &= -\frac{1}{\sigma_2^2} \int dx \left(-\frac{\partial^2 \phi(x, t)}{\partial t^2} + \zeta \frac{\partial^4 \phi(x, t)}{\partial t^4} \right) - \sum_{i=1}^N \delta(t - t_i) \\ &\quad - \begin{cases} \gamma \\ \gamma \langle x(t) \rangle_\phi \left(\tau(t) - \langle x(t) \rangle_\phi \right) + \frac{\vartheta}{2} \left(2 \langle x(t) \rangle_\phi^2 - \langle x^2(t) \rangle_\phi \right) \end{cases} \end{aligned} \quad (2.54)$$

Man beachte, daß bei Wegfall von sämtlichen Vorwissens – das heißt, $\tau(t)$ tritt nicht auf, und es gilt $\sigma_1 = \sigma_2 = \infty$ – die Stationaritätsgleichung (2.53) zusammen mit (2.54) auf

$$0 = \sum_{i=1}^N \delta(t - t_i) \left(\delta(x - x_i) - e^{-\phi(x, t)} \right) \quad (2.55)$$

zusammenfällt. Erwartungsgemäß ist die Lösung an den Stellen t_i durch die δ –Distributionen

$$Q(x, t_i) = e^{-\phi(x, t_i)} = \delta(x - x_i) \quad (2.56)$$

gegeben, falls zur Zeit t_i nur ein Datum x_i vorliegt. Zu anderen Zeiten $t \neq t_i$ ist jede normierte Verteilung eine Lösung.

In einer numerischen Anwendung² wurde für beide Varianten die Stationaritätsgleichung (2.53) für den zweidimensionalen Fall $\phi(x, t)$ gemäß Kapitel 1.6 durch Iteration mittels

$$\phi^{(j+1)}(x, t) = \phi^{(j)}(x, t) - \eta^{(j)} \int dt' dx' \mathcal{O}(x, t; x', t') \frac{\delta F[\phi^{(j)}]}{\delta \phi(x', t')} \quad (2.57)$$

gelöst. Dafür wurde die Funktion $\phi(x, t)$ auf einem zweidimensionalen Gitter diskretisiert ($\Delta x = 1$, $\Delta t = 1$) und entsprechend sämtliche Operatoren als Matrizen dargestellt (siehe Kapitel 5.4). Ein Gitter der Größe 30×40 war dabei numerisch gut handhabbar. Es ist zu beachten, daß in den Ableitungsoperatoren zyklische Randbedingungen verwendet werden mußten. Als Iterationsmatrix bot sich

$$\mathcal{O}^{-1}(x, t; x', t') = \delta(x - x') \delta(t - t') \left(\frac{1}{\sigma_1^2} \left(-\frac{\partial^2}{\partial x^2} + \zeta \frac{\partial^4}{\partial x^4} \right) + \frac{1}{\sigma_2^2} \left(-\frac{\partial^2}{\partial t^2} + \zeta \frac{\partial^4}{\partial t^4} \right) + \nu \right) \quad (2.58)$$

²In den numerischen Anwendungen der gesamten Arbeit wird mit dimensionslosen Größen gerechnet.

als genäherte Hesse-Matrix von (2.53) an. Die "Masse" ν wurde dabei aus Gründen der Invertierbarkeit eingeführt. Gestartet wurde die Iteration mit der Gleichverteilung. Es hat sich eine Schrittweitenoptimierung von $\eta^{(j)}$ als günstig herausgestellt, da im Laufe der Iteration aufgrund der zunehmenden Bedeutung des Datenanteils für das Energiefunktional die Anwendung von (2.58) als Näherung immer schlechter wurde. Das Ergebnis einer Suche nach Lösungen der oberen Variante von (2.53) mittels (2.57) ist in Abbildung 2.2 dargestellt. Als Musterfunktion wurde $\tau(t) = 20 + 10 \cos(\frac{2\pi}{30}t)$ auf dem verwendeten Gitter angenommen. Diese Wahl für das Muster war konform mit den zyklischen Randbedingungen bezüglich der Zeitachse. Die Daten wurden erzeugt, indem aus den Intervallen $t \in [8, 15]$ und $t \in [22, 25]$ zufällige Zeiten t_i gleichverteilt gezogen wurden und die Ortskoordinaten mittels $x_i(t_i) = 20 + 15 \cos(\frac{2\pi}{30}t_i) + \Upsilon$ mit einer ganzzahligen Zufallszahl $\Upsilon \in [0, 3]$ berechnet wurden.

Zu den Abbildungen lohnen sich folgende Bemerkungen: Wie zu erwarten, versucht die Lösung für die Regressionsfunktion den Verlauf der Daten unter den Gesichtspunkten des Vorwissens wiederzugeben. Durch die Forderung einer glatten Funktion ϕ ist auch eine glatte Regressionsfunktion $X(t)$ zu erwarten. Die Kopplung von Muster und Lösung Q ist aufgrund eines kleinen Hyperparameters γ eher schwach. Das Abstellen des Musters $\tau(t)$ hat aber ein deutlich schlechteres Ergebnis zur Folge. Man beachte, daß sich an Stellen hoher Datendichte erwartungsgemäß die Wahrscheinlichkeitsdichte an den Daten konzentriert und damit die Unsicherheit für die Lösung $X(t)$ abnimmt, während an den übrigen Stellen die Tendenz zur Gleichverteilung dominiert.

Das Ergebnis zu der unteren Variante von (2.53) ist in Abbildung 2.3 dargestellt. Es sind die gleichen Effekte für die Lösung zu erkennen wie in Abbildung 2.2. Es zeigt sich allerdings eine härtere Implementierung des Musters $\tau(t)$. Nur an den Stellen, an denen die Daten für eine Abweichung von der Musterfunktion sprechen, verläßt die Regressionsfunktion das Muster.

Im Vergleich der beiden Varianten läßt sich feststellen, daß in der ersten Variante das Muster effektiv als zusätzliche Daten bei den Funktionswerten von τ angesehen werden kann, während in der zweiten Variante das Muster in die Rechnung eingeht, indem es durch Fixierung des Erwartungswerts von Q Einfluß auf die Gestalt der Lösung nimmt. Die Dichteverteilungen Q der Lösungen für die beiden Varianten unterscheiden sich wenig. Abweichungen treten im wesentlichen beim Vergleich der Regressionsfunktionen an Orten geringer Datendichte auf.

In einem zweiten Beispiel wird die direkte Darstellung in $Q(x, t)$ benutzt. Die unangenehme Positivitätsforderung für Q wird auf passive Art und Weise behandelt. Darunter ist zu verstehen, daß erst im nachhinein für die Lösung überprüft wird, ob die Bedingung $Q(x, t) \geq 0$ erfüllt ist. Dies setzt einen geeigneten Startpunkt für die Iteration, eine günstige Wahl für die Iterationsmatrix und hinreichend kleine Iterationsschritte η voraus, damit die Lösung $Q(x, t)$ nicht lokal in den negativen Wertebereich abgleitet.

Falls kein Muster vorhanden ist, kann für die Aposteriori-Wahrscheinlichkeit der Dichte

$$p[Q|D] = \frac{p[D|Q] p[Q]}{p[D]} \sim \int [d\lambda(t)] e^{-F[Q, \lambda, D]} \prod_{x, t} \theta(Q(x, t)) \quad (2.59)$$

geschrieben werden, wobei das Funktional F von der analytischen Form

$$\begin{aligned} F[Q, \lambda, D] = & -i \int dt \lambda(t) \left(1 - \int dx Q(x, t) \right) - \sum_{i=1}^N \ln Q(x_i, t_i) \\ & + \frac{1}{2\sigma_1^2} \int dt dx \left((\nabla_x Q(x, t))^2 + \zeta (\Delta_x Q(x, t))^2 \right) \\ & + \frac{1}{2\sigma_2^2} \int dt dx \left(\left(\frac{\partial Q(x, t)}{\partial t} \right)^2 + \zeta \left(\frac{\partial^2 Q(x, t)}{\partial t^2} \right)^2 \right) \end{aligned} \quad (2.60)$$

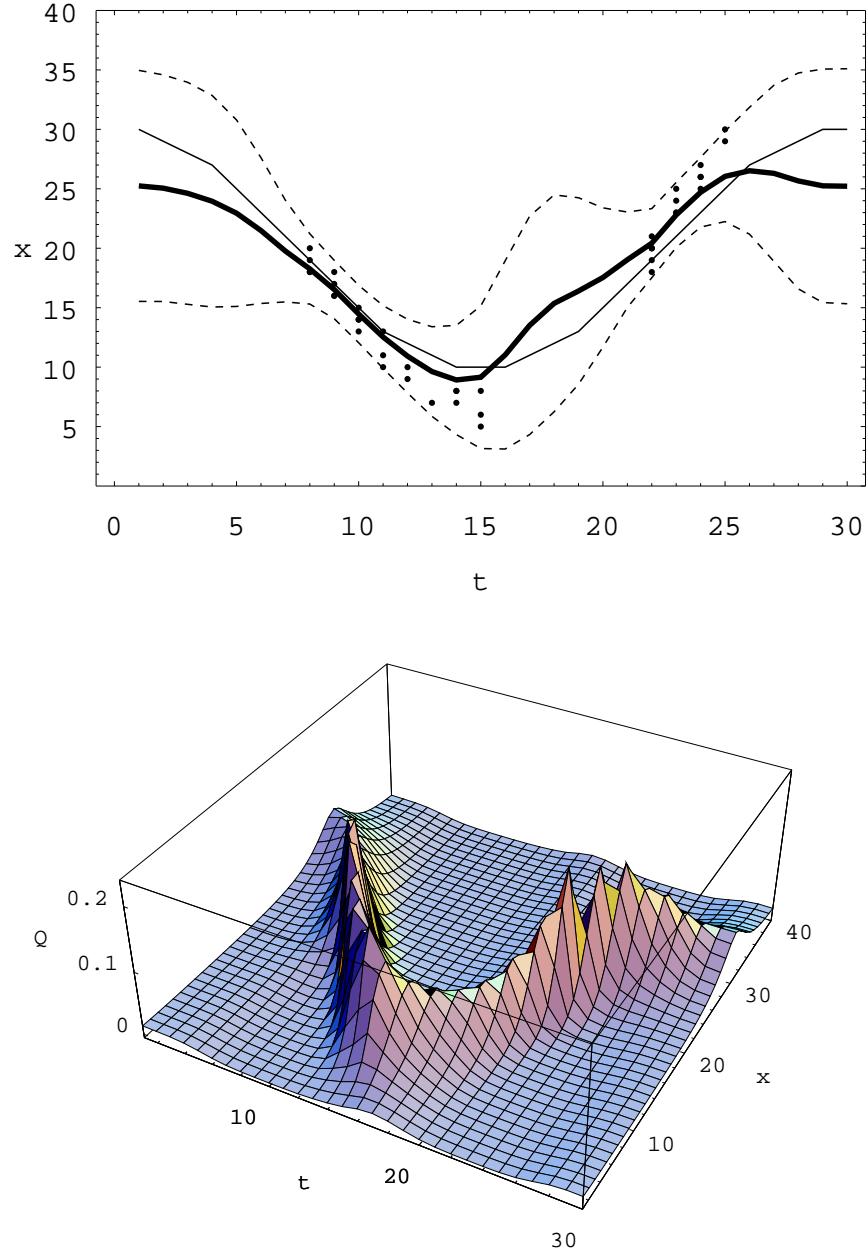


Abbildung 2.2: Lernen von Wahrscheinlichkeitsdichten mittels oberer Variante von (2.53). Die Zeitachse wurde in 30, die Ortsachse in 40 Gitterpunkten diskretisiert. Oben: Die Punkte markieren die Daten (x_i, t_i) , und die dünne durchgezogene Linie kennzeichnet das Muster $\tau(t)$. Die gelernte Regressionsfunktion $X(t)$ als Ortserwartungswert von $Q(x, t) = e^{-\phi(x, t)}$ ist durch die dicke durchgezogene Linie dargestellt. Die gestrichelten Linien zeigen die Standardabweichung für die Lösung $X(t)$ unter der Verteilung Q an. Unten ist die dazugehörige Wahrscheinlichkeitsverteilung $Q(x, t)$ dargestellt. Die Parameter sind: $\sigma_1 = \sigma_2 = 4.08$, $\gamma = 1$, $\zeta = 5$, $N = 40$, $\nu = 0.1$.

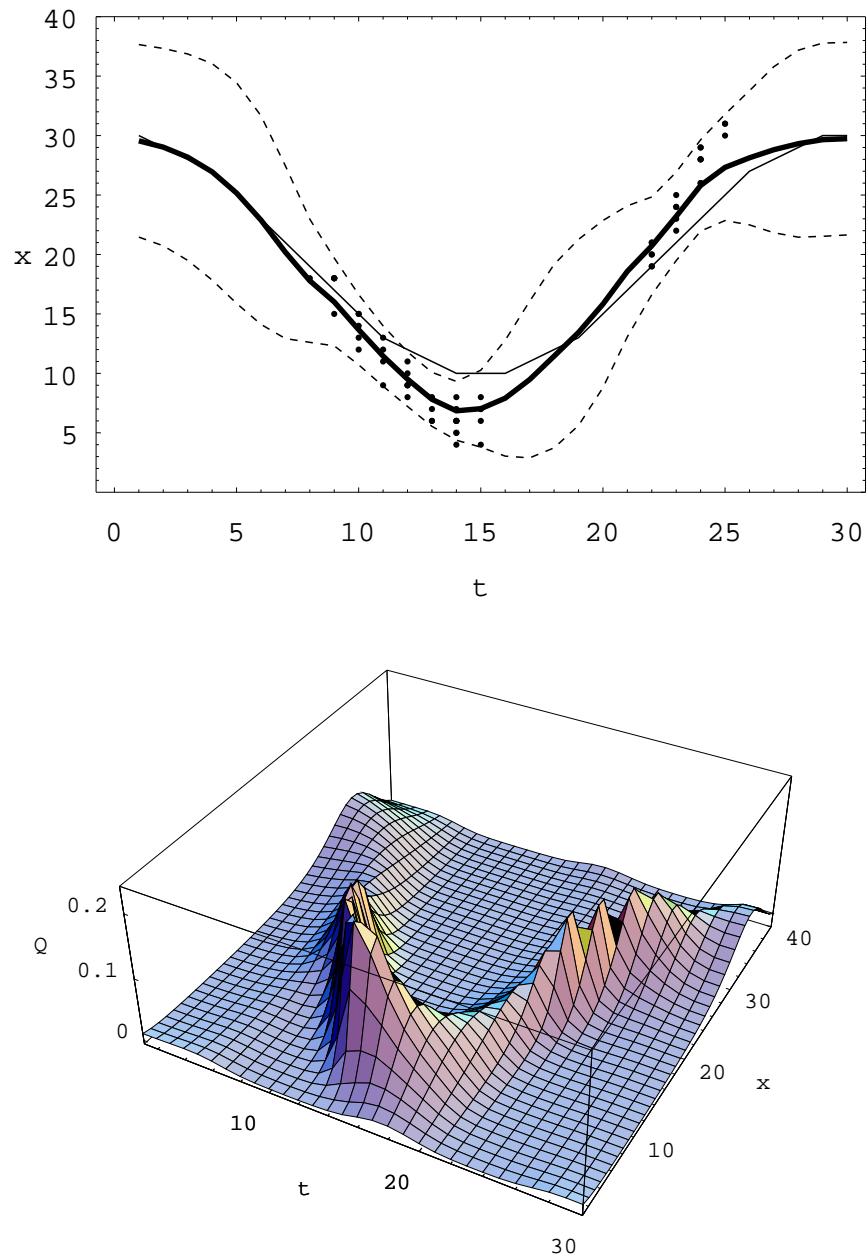


Abbildung 2.3: Lernen von Wahrscheinlichkeitsdichten mittels unterer Variante von (2.53). Die Bedeutung der Graphen ist in Abbildung 2.2 erläuert. Die Parameter sind $\gamma = 7$, $\vartheta = 0.5$, $\sigma_1 = \sigma_2 = 4.08$, $\zeta = 8$, $N = 60$, $\nu = 0.1$.

ist. Eine Durchführung einer Maximum–Aposteriori–Approximation mittels gleichzeitiger Sattelpunktsnäherung für die Wirkung bezüglich Q und λ führt auf die Normbedingung $1 = \int dx Q(x, t)$ zurück. Als zweite Gleichung erhält man die nichtlineare Differentialgleichung

$$0 = \frac{\delta F[Q, \lambda, D]}{\delta Q(x, t)} = \frac{1}{\sigma_1^2} \left(-\Delta_x Q(x, t) + \zeta \Delta_x^2 Q(x, t) \right) + \frac{1}{\sigma_2^2} \left(-\frac{\partial^2 Q(x, t)}{\partial t^2} + \zeta \frac{\partial^4 Q(x, t)}{\partial t^4} \right) + i\lambda(t) - \sum_{i=1}^N \frac{\delta(x - x_i)\delta(t - t_i)}{Q(x, t)}. \quad (2.61)$$

Multiplikation der Gleichung (2.61) mit $Q(x, t)$ von links mit anschließender Integration über die Ortskoordinaten ermöglicht die Berechnung des Hilfsfeldes. So erhält man:

$$i\lambda(t) = -\frac{1}{\sigma_2^2} \int dx Q(x, t) \left(-\frac{\partial^2 Q(x, t)}{\partial t^2} + \zeta \frac{\partial^4 Q(x, t)}{\partial t^4} \right) - \frac{1}{\sigma_1^2} \int dx Q(x, t) \left(-\Delta_x Q(x, t) + \zeta \Delta_x^2 Q(x, t) \right) + \sum_{i=1}^N \delta(t - t_i). \quad (2.62)$$

In Abbildung 2.4 ist das Ergebnis einer Suche nach Lösungen von (2.61) vorgestellt. Die Regressionsfunktion gibt den Verlauf der Daten wieder. An Stellen geringer Datendichte dominiert die Tendenz zur Gleichverteilung. Die Varianzen zeigen die zu erwartenden Unsicherheiten für die Regressionsfunktion. Das Beispiel zeigt, daß die "Qualität" der Lösung auch ohne Anwesenheit eines Musters nicht nachhaltig nachläßt.

Insgesamt kann man sagen, daß die Verwendung der logarithmischen Wahrscheinlichkeitsdichten aus numerischen Gründen zu bevorzugen ist. Durch das Ignorieren der θ –Bedingung beim Arbeiten mit Q für die Lösung der Stationaritätsgleichung tendiert die Dichte während der Iteration dazu, zu Zeiten t hoher Datendichte an den Rändern negative Werte anzunehmen, damit die Dichte in den Datenpunkten möglichst groß werden kann, ohne die Normbedingung zu verletzen. Um diesen Effekt zu vermeiden, muß in der Nähe der Lösung äußerst "vorsichtig" iteriert werden. Bei der Verwendung von (2.14) zur Erzwingung der Positivität erhält das Energiefunktional (2.60) die zusätzlichen Summanden

$$F[Q, \lambda, D, s] = \dots - i \int dx dt s(x, t) Q(x, t) + \int dx dt \ln s(x, t). \quad (2.63)$$

Der Sattelpunkt für das Hilfsintegral (2.14) liefert die Bedingung $s(x, t) = -\frac{i}{Q(x, t)}$. Dadurch bekommt die Stationaritätsgleichung (2.61) einen extra Summanden

$$0 = \frac{\delta F[Q, \lambda, D, s]}{\delta Q(x, t)} = \dots - is(x, t) = \dots - \frac{1}{Q(x, t)} = \dots - \frac{\delta \int dx' dt' \ln Q(x', t')}{\delta Q(x, t)}. \quad (2.64)$$

Das heißt, die Positivitätsbedingung führt effektiv auf einen Bestrafungsterm der Form

$$\Gamma[Q] = - \int dx' dt' \ln Q(x', t'). \quad (2.65)$$

Dieser verhindert durch zu kleine Funktionswerte von $Q(x, t)$. Der Lagrange–Multiplikator λ aus (2.62) muß dann mittels $i\lambda(t) = \dots + \int dx$ verändert werden.

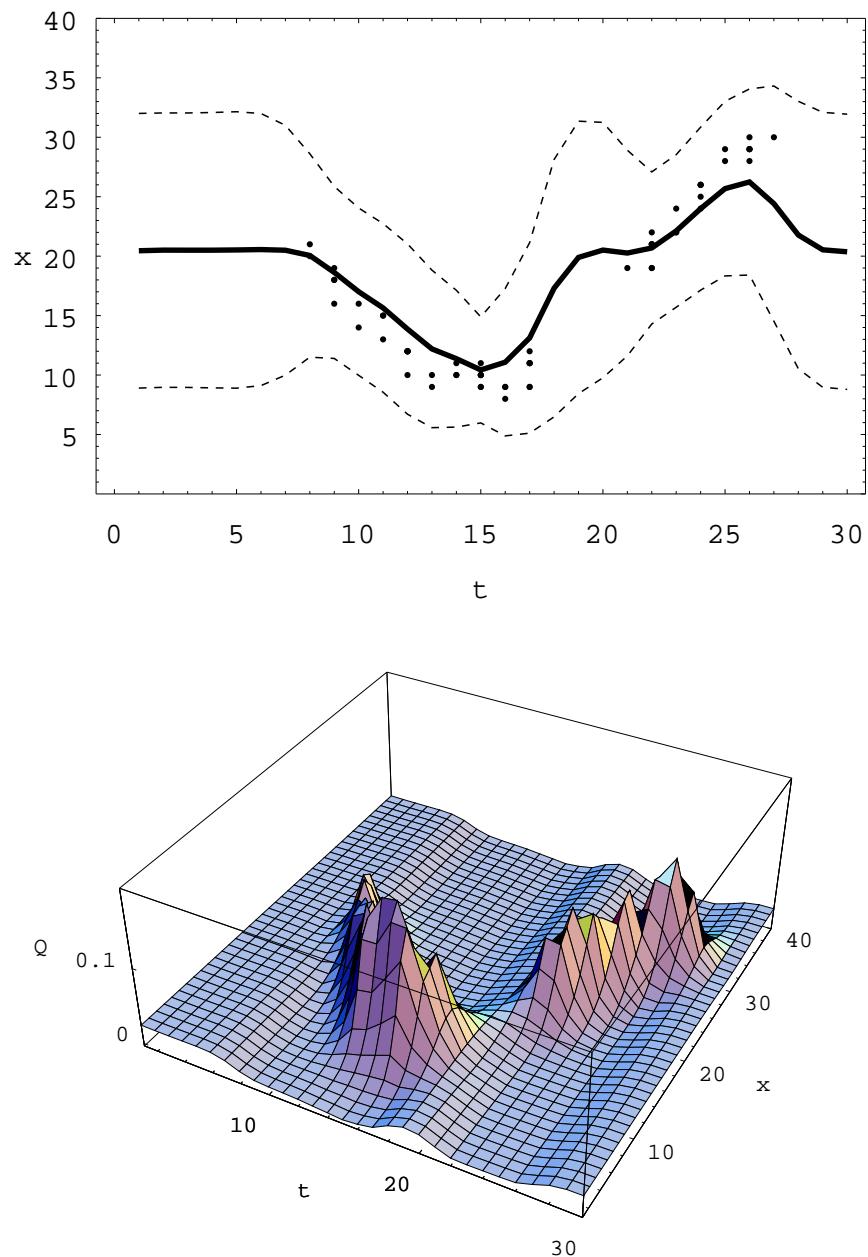


Abbildung 2.4: Lernen von Wahrscheinlichkeitsdichten mittels (2.61). Die Parameter sind: $N = 60$, $\nu = 1$, $\zeta = 1$, $\sigma_1 = 2.24$, $\sigma_2 = 0.07$

Kapitel 3

Lernen von statistischen Operatoren

3.1 Statistische Operatoren

Der Zustand eines quantenmechanischen Systems wird durch den statistischen Operator ρ festgelegt. Hier sind noch einmal in Kürze die wichtigsten Eigenschaften von ρ zusammengefaßt. ρ ist ein hermitischer ($\rho^\dagger = \rho$), positiv-semidefiniter Operator, d.h. diagonalisierbar mit Eigenwerten größer gleich Null. Zusätzlich ist seine Spur auf Eins normiert ($\text{Tr}[\rho] = 1$). Für die Darstellung von ρ in seiner (ohne Beschränkung der Allgemeinheit diskreten) Eigenbasis folgt dann:

$$\rho = \sum_n p_n |n\rangle\langle n|, \quad \sum_n p_n = 1 \quad \text{mit } p_n \geq 0 \quad \forall n. \quad (3.1)$$

(3.1) legt die Interpretation von ρ als Zustandsgemisch nahe, wobei p_n die Wahrscheinlichkeit für den reinen Zustand $|n\rangle$ angibt.

Die Erwartungswerte für Messungen einer Observablen \mathcal{A} an einem durch ρ beschriebenen Zustand können mittels

$$\langle \mathcal{A} \rangle_\rho = \text{Tr}[\rho \mathcal{A}] = \sum_n p_n \langle n | \mathcal{A} | n \rangle \quad (3.2)$$

berechnet werden. Wenn $|\epsilon\rangle\langle\epsilon|$ den Projektor auf den Eigenzustand des Meßoperators \mathcal{A} zum Eigenwert ϵ darstellt, so gibt

$$\text{Tr}[\rho |\epsilon\rangle\langle\epsilon|] = \sum_n p_n |\langle n | \epsilon \rangle|^2 = p[\epsilon | \mathcal{A}, \rho] \quad (3.3)$$

die Wahrscheinlichkeit für die Messung des Eigenzustands zum Eigenwert ϵ . Hier wurde die Notation von (1.7) mit $f \Rightarrow \rho$, $x \Rightarrow \mathcal{A}$ und $y \Rightarrow \epsilon$ wieder aufgenommen. Die Wahrscheinlichkeitsinterpretation aus (3.3) bleibt im Falle von Entartung erhalten, wenn bei e -facher Entartung des Eigenraums von \mathcal{A} zum Eigenwert ϵ ($\mathcal{A} | \epsilon_i \rangle = \epsilon | \epsilon_i \rangle$, $\forall i = 1, \dots, e$) der Erwartungswert von $|\epsilon\rangle\langle\epsilon| = \sum_{i=1}^e |\epsilon_i\rangle\langle\epsilon_i|$ betrachtet wird:

$$\text{Tr}[\rho |\epsilon\rangle\langle\epsilon|] = \sum_{i=1}^e \sum_n p_n |\langle n | \epsilon_i \rangle|^2 = p[\epsilon | \mathcal{A}, \rho]. \quad (3.4)$$

Wichtig für die Angabe von Meßwahrscheinlichkeiten ist also, daß zu jedem Meßoperator Eigenzustände, Eigenwerte und Entartungsgrade bekannt sein müssen.

Ist der statistische Operator gleichzeitig ein Projektionsoperator ($\rho^2 = \rho$), dann (und nur dann) beschreibt ρ einen reinen Zustand, denn wegen der Bedingungen $p_n \in \{0, 1\}$ und $\sum_n p_n = 1$ folgt sofort $p_n = \delta_{nj}$ für ein bestimmtes j und damit $\rho = |j\rangle\langle j|$.

3.2 Lernen unter Maximierung der Entropie

Dieses Kapitel handelt davon, Aussagen über den statistischen Operator nach Messung von Observablen zu machen¹. Es wird davon ausgegangen, daß ein Ensemble von identisch präparierten Zustandsgemischen vorliegt. Alle Ensemblemitglieder werden durch den gleichen statistischen Operator ρ beschrieben. An einem Gemisch wird höchstens eine Messung vorgenommen, um die Unabhängigkeit der Meßdaten zu gewährleisten.

Bei dem Meßprozeß kann es sich sowohl um die Anwendung von unterschiedlichen Operatoren als auch um wiederholtes Messen desselben Operators handeln. Die Daten werden in diesem Kapitel durch die Menge $D = \{|\chi_i\rangle\langle\chi_i| \mid i = 1, \dots, N\}$ abgekürzt, wobei $|\chi_i\rangle\langle\chi_i|$ der bei der Messung des i -ten Operators auftretende Projektor zum Meßwert χ_i ist. Bei Entartung von χ_i ist der gesamte Eigenraum gemeint.

Zusätzlich kann es sein, daß ein Referenzoperator ρ_0 für ρ bekannt ist – also zu erwarten ist, daß ρ und ρ_0 ähnliche statistische Gemische beschreiben. Die Festlegung von ρ_0 könnte auf den Erwartungen nach vorangegangenen Messungen beruhen.

Der Bayessche Ansatz der Behandlung der Aposteriori-Wahrscheinlichkeit $p[\rho|D, \rho_0]$ lautet dann:

$$p[\rho|D, \rho_0] = \frac{p[D|\rho, \rho_0] p[\rho|\rho_0]}{p[D|\rho_0]} = \frac{p[D|\rho] p[\rho|\rho_0]}{p[D|\rho_0]}, \quad (3.5)$$

denn unter der Voraussetzung, daß ρ der wahre, datenerzeugende statistische Operator ist, gilt $p[D|\rho, \rho_0] = p[D|\rho]$. Unter Annahme von unabhängigen Messungen folgt aus der Quantenmechanik

$$p[D|\rho] = \prod_{i=1}^N \text{Tr} \left[|\chi_i\rangle\langle\chi_i| \rho \right]. \quad (3.6)$$

Freiheiten sind bei der Wahl von $p[\rho|\rho_0]$ gegeben. Folgender Fall wird jetzt studiert: Es wird eine Verteilung konstruiert, welche das Prinzip der Entropimaximierung erfüllt und gleichzeitig eine Tendenz zu den wahrscheinlichsten Zuständen von ρ_0 beinhaltet:

$$p[\rho|\rho_0] = \frac{1}{Z} e^{-k \text{Tr}[\rho \ln \rho] + \mu \text{Tr}[\rho_0 \rho]} \delta(\text{Tr}[\rho] - 1). \quad (3.7)$$

Denn mit $\rho_0 = \sum_m p_m^0 |\zeta_m\rangle\langle\zeta_m|$ in der Darstellung seiner Eigenbasis und ρ wie in (3.1) entspricht

$$\langle \rho_0 \rangle_\rho = \langle \rho \rangle_{\rho_0} = \text{Tr}[\rho \rho_0] = \sum_m p_m^0 \left[\sum_n p_n |\langle n | \zeta_m \rangle|^2 \right] \quad (3.8)$$

der mittleren Wahrscheinlichkeit die Basiszustände von ρ zu messen, wenn ρ_0 der wahre statistische Operator wäre (oder umgekehrt!). Mit (3.7) folgt für (3.5) unter Einführung der δ -Funktion in ihrer Integraldarstellung:

$$p[\rho|D, \rho_0] \sim \int d\lambda e^{-F[\rho, \rho_0, D, \lambda]} \quad (3.9)$$

¹Zu der Rekonstruktion von statistischen Operatoren für Spin- $\frac{1}{2}$ -Systeme gibt es einige Veröffentlichungen. Diese beruhen ebenfalls auf dem Prinzip der Entropimaximierung [11] oder nehmen den Bayesschen Formalismus auf [12]. Im Unterschied zu diesem Kapitel wird dort allerdings die exakte Kenntnis von Erwartungswerten oder die Durchführung einer unendlichen Anzahl von Messungen vorausgesetzt.

mit der Wirkung

$$F[\rho, \rho_0, D, \lambda] = k \text{Tr}[\rho \ln \rho] - \mu \text{Tr}[\rho \rho_0] - i\lambda(\text{Tr}[\rho] - 1) - \sum_{i=1}^N \ln \text{Tr} \left[|\chi_i\rangle \langle \chi_i| \rho \right]. \quad (3.10)$$

Eine Maximum–Aposteriori–Approximation von (3.9) entspricht der Suche nach einem Lagrange–Parameter, der die Nebenbedingung erfüllt

$$0 = \frac{\partial F}{\partial \lambda}, \quad \text{also} \quad \text{Tr}[\rho] = 1, \quad (3.11)$$

und nach Stationarität bezüglich des Operators ρ . Im Anhang B ist vorgeführt, wie mittels Ableitung nach den Matrixelementen von ρ in beliebiger Basis die Operatoridentität

$$0 = \frac{\delta F}{\delta \rho} \quad \Rightarrow \quad 0 = k \ln \rho - (i\lambda - k) \mathcal{I} - \mu \rho_0 - \sum_{i=1}^N \frac{|\chi_i\rangle \langle \chi_i|}{\text{Tr}[\rho |\chi_i\rangle \langle \chi_i| \rho]} \quad (3.12)$$

hergeleitet werden kann. Multiplikation mit ρ von links, anschließende Spurbildung und Ausnutzung von (3.11) ergibt die Bedingung für λ :

$$i\lambda - k = k \text{Tr}[\rho \ln \rho] - \mu \text{Tr}[\rho \rho_0] - N, \quad (3.13)$$

und damit erhält man aus (3.12) die Gleichung

$$\rho = \frac{e^{A(\rho, \rho_0, D)}}{\mathcal{Z}(\rho)} \quad \text{mit} \quad A(\rho, \rho_0, D) = \frac{\mu}{k} \rho_0 + \frac{1}{k} \sum_{i=1}^N \frac{|\chi_i\rangle \langle \chi_i|}{\text{Tr}[\rho |\chi_i\rangle \langle \chi_i| \rho]} \quad (3.14)$$

mit der Normierungskonstante

$$\mathcal{Z}(\rho) = e^{-\text{Tr}[\rho \ln \rho] + \frac{\mu}{k} \text{Tr}[\rho \rho_0] + \frac{N}{k}} = \text{Tr}[e^{A(\rho)}], \quad (3.15)$$

welche durch Iteration selbstkonsistent für ρ gelöst werden muß.

Diskutiert man die beiden Grenzfälle $k \rightarrow \infty$ oder $k \rightarrow 0$, so entspricht dies einer Übergewichtung beziehungsweise einer Vernachlässigung des Entropieterms aus (3.10). Wie zu erwarten ist, ist im Fall von $k \rightarrow \infty$ die Gleichverteilung die Lösung des Problems (3.14):

$$\rho = e^{\text{Tr}[\rho \ln \rho]} \mathcal{I} \quad \Rightarrow \quad \rho = \frac{1}{\dim \mathcal{H}} \mathcal{I}. \quad (3.16)$$

Interessanter ist der Fall $k \rightarrow 0$. Die Gleichung (3.14) ist von der Form

$$\rho = e^{\text{Tr}[\rho \ln \rho] + \frac{1}{k} B(\rho)}, \quad (3.17)$$

wobei gleichzeitig $\text{Tr}[\rho] = 1$ erfüllt ist. Da für einen normierten Operator die Spur $\text{Tr}[\rho \ln \rho]$ durch $\rho \sim \mathcal{I}$ nach unten beschränkt ist, ist im Grenzwert $k \rightarrow 0$ die Bedingung (3.14) mit (3.11) nur für den Fall

$$0 = B = \left(-\mu \text{Tr}[\rho \rho_0] - N \right) \mathcal{I} + \mu \rho_0 + \sum_{i=1}^N \frac{|\chi_i\rangle \langle \chi_i|}{\text{Tr}[\rho |\chi_i\rangle \langle \chi_i| \rho]} \quad (3.18)$$

erfüllbar.

Als weiteres Vorgehen bietet sich die Diagonalisierung des hermitischen Operators A aus (3.14) nach

$$A(\rho) |\mathbf{n}(\rho)\rangle = a_n(\rho) |\mathbf{n}(\rho)\rangle \quad \text{und} \quad A(\rho) = \sum_n a_n(\rho) |\mathbf{n}(\rho)\rangle \langle \mathbf{n}(\rho)| \quad (3.19)$$

an. Es sei hier nochmals angemerkt, daß die Zustände $|\chi_i\rangle$ im allgemeinen nicht Eigenzustände ein und desselben Meßoperators sind. Mit

$$e^{\sum_n a_n |n\rangle\langle n|} = \sum_n e^{a_n} |n\rangle\langle n| \quad (3.20)$$

folgt für (3.14) dann die Darstellung von ρ in der Eigenbasis

$$\rho = \frac{e^{A(\rho)}}{\text{Tr}[e^{A(\rho)}]} = \sum_n p_n(\rho) |n(\rho)\rangle\langle n(\rho)| \quad \text{mit } p_n(\rho) = \frac{e^{a_n(\rho)}}{\sum_m e^{a_m(\rho)}}. \quad (3.21)$$

Beim iterativen Vorgehen zum Lösen der Gleichung (3.21) startet man mit einer Vermutung für ρ (zum Beispiel $\rho = \rho_0$ im Falle großer μ), löst das Eigenwertproblem (3.19) und berechnet nach (3.21) einen neuen statistischen Operator. Man wiederholt dieses bis zur Selbstkonsistenz.

In einem numerischen Beispiel werden Orts- und Impulsmessungen an einem unbekannten Ensemble betrachtet. Als wahrer statistischer Operator wurde

$$\rho_D = \frac{e^{-\beta H}}{\text{Tr}[e^{-\beta H}]} \quad \text{mit} \quad H = \frac{\hat{p}^2}{2m} + \frac{1}{2}m\omega^2(\hat{x} - x_0)^2 \quad (3.22)$$

angenommen. Die Daten wurden aus den Likelihood-Verteilungen $p[\tilde{p}|\hat{p}, \rho_D]$ und $p[\hat{x}|\hat{x}, \rho_D]$ gewonnen (siehe Anhang A). Es wurde davon ausgegangen, daß kein Referenzoperator ρ_0 vorliegt. Alle Operatoren wurden numerisch jeweils auf einem Gitter der Größe $n = 20$ mit der Gitterkonstanten $\Delta x = 1$ diskretisiert in ihrer Ortsdarstellung behandelt. Dies wird für den Hamilton-Operator – und damit auch für den statistischen Operator – im Abschnitt 5.4 näher erläutert. Der Impulsoperator in Ortsdarstellung $\hat{p} = -i\hbar\nabla_x$ wurde numerisch durch die Matrix

$$\hat{p} \Rightarrow -\frac{i\hbar}{2(\Delta x)} \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ -1 & \ddots & \ddots & & \vdots \\ 0 & \ddots & & & 0 \\ \vdots & & \ddots & \ddots & 1 \\ 0 & \cdots & 0 & -1 & 0 \end{pmatrix} \quad (3.23)$$

implementiert. Durch diese Wahl für den Ableitungsoperator im Ortsraum wurden die Hermitizität und die damit verbundenen reellen Eigenwerte gewährleistet. Durch Diagonalisierung wurden die numerischen Eigenvektoren und Eigenwerte \tilde{p} als mögliche Meßergebnisse gewonnen. In Abbildung 3.1 sind die Wahrscheinlichkeitsdichten für die einzelnen statistischen Operatoren im Orts- und Impulsraum dargestellt. Es zeigt sich, daß die Lösung von (3.21) in beiden Räumen die Datenverteilungen wiedergibt. Das oben beschriebene iterative Vorgehen versagt, falls die Gewichtung des Entropieterms durch k zu klein ist. Dieses liegt an dem "empfindlichen" Nenner des Datenterms aus (3.14). In der Anwendung wurde vorsichtiger mittels

$$\rho^{(j+1)} = \rho^{(j)} + \eta \left(\frac{e^{A(\rho^{(j)})}}{\text{Tr}[e^{A(\rho^{(j)})}]} - \rho^{(j)} \right) \quad (3.24)$$

iteriert.

3.3 Spezialfall gleicher Messungen

In diesem Abschnitt wird ein häufig auftretender Spezialfall behandelt: So wird jetzt davon ausgegangen, daß nur eine feste Observable \mathcal{A} wiederholt gemessen wird. Die dazugehörigen Eigenzustände $|\chi_n\rangle$ erfüllen dann Vollständigkeits- und Orthogonalitätsrelation. Die bei den Messungen erhaltenen Daten können jetzt auch durch die absolute Häufigkeitsverteilung h_n der gemessenen

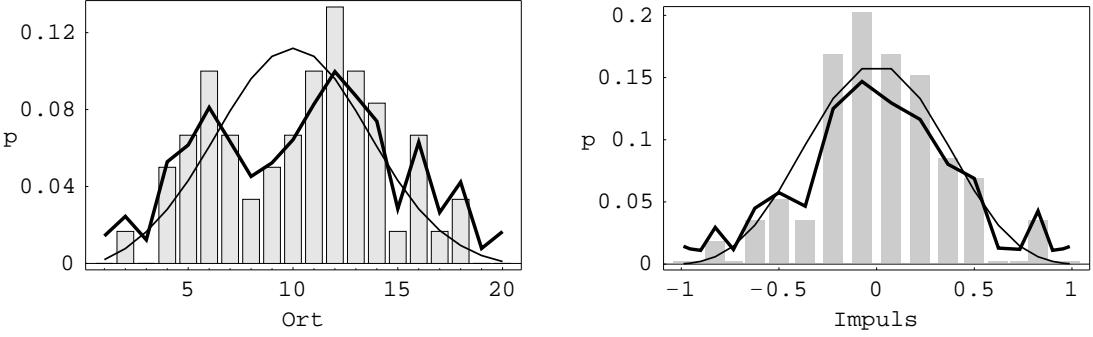


Abbildung 3.1: Lernen statistischer Operatoren für Orts- und Impulsmessungen. Links sind die Wahrscheinlichkeitsverteilungen der statistischen Operatoren im Ortsraum, rechts im Impulsraum dargestellt. Die dicken durchgezogenen Linien sind die Verteilungen der Lösung ρ nach (3.14) mittels Iteration nach (3.24). Die dünnen durchgezogenen Linien kennzeichnen die Verteilung des Operators ρ_D , für den die Daten (Balkendiagramm) erzeugt wurden. Die Masse des betrachteten Teilchens betrug $m = 0.1$, bei $\hbar = 1$. Die Anzahl der Daten der Orts- und Impulsmessungen ist $N_x = N_{\tilde{p}} = 60$. Die Iteration wurde mit der Gleichverteilung gestartet. Die übrigen Parameter sind: $k = 30$, $\beta = 0.8$, $\omega = 1$, $x_0 = 10$.

Eigenwerte von \mathcal{A} ausgedrückt werden – also $\sum_n h_n = N$. Außerdem wird angenommen, daß ρ_0 diagonal in dieser Basis ist: $\rho_0 = \sum_n p_n^0 |\chi_n\rangle\langle\chi_n|$.

Der Operator A aus (3.14) bekommt dann die Form

$$A = \sum_n \left[\frac{h_n}{k \text{Tr}[\rho |\chi_n\rangle\langle\chi_n|]} + \frac{\mu}{k} p_n^0 \right] |\chi_n\rangle\langle\chi_n| \quad (3.25)$$

und ist bereits in der Basis von \mathcal{A} diagonal. Damit wird auch ρ nach (3.20) diagonal und durch die Angabe von $p_n(\rho)$ nach

$$\rho = \sum_n p_n(\rho) |\chi_n\rangle\langle\chi_n| \quad (3.26)$$

einheitig festgelegt. Im Gegensatz zu (3.21) hängen die Basiszustände nicht mehr von ρ ab. Mit

$$\text{Tr}[\rho |\chi_n\rangle\langle\chi_n|] = p_n(\rho) \quad , \quad \text{Tr}[\rho \ln \rho] = \sum_n p_n(\rho) \ln p_n(\rho) \quad \text{und} \quad \text{Tr}[\rho \rho_0] = \sum_n p_n(\rho) p_n^0 \quad (3.27)$$

bekommt man schließlich für (3.14) und (3.15) gekoppelte Gleichungen für die Wahrscheinlichkeiten p_n :

$$p_n = \frac{1}{Z} e^{\frac{\mu_n}{k p_n} + \frac{\mu}{k} p_n^0} \quad \text{mit} \quad Z = \sum_m e^{\frac{\mu_m}{k p_m} + \frac{\mu}{k} p_m^0} = e^{-\sum_m p_m \ln p_m + \frac{\mu}{k} \sum_m p_m p_m^0 + \frac{N}{k}}. \quad (3.28)$$

Die Darstellung (3.28) lässt sich unter Verwendung des Produktlogarithmus auf eine andere Form bringen. Die Funktion $w(x)$ wird der Produktlogarithmus von x genannt, wenn sie die Gleichung $x = w e^w$ löst. Eine Gleichung der Form $x = e^{\frac{c}{x} + t}$ kann somit mittels $x = \frac{c}{w(c e^{-t})}$ nach x aufgelöst werden. Also folgt:

$$p_n = \frac{h_n}{k} \frac{1}{w\left(\frac{h_n Z}{k} e^{-\frac{\mu}{k} p_n^0}\right)}, \quad (3.29)$$

wobei Z nach (3.28) p_n -abhängig ist. Für $h_n = 0$ liegt in (3.29) eine hebare Singularität vor, und es gilt in Übereinstimmung mit (3.28) $p_n = \frac{1}{Z} e^{\frac{\mu}{k} p_n^0}$. Für das Lösen der Gleichungen (3.28) beziehungsweise (3.29) wird eine günstige Vorgehensweise sein, erst die Abhängigkeit der Zustandssumme Z von p_n zu ignorieren, indem durch die Wahl von $Z = 1$ in (3.29) unnormierte Wahrscheinlichkeiten p'_n berechnet und diese danach durch $Z = \sum_n p'_n$ normiert werden. Diese Prozedur kann bis zur Konvergenz fortgesetzt werden.

Entscheidend für die Lösung von (3.28) ist die Gewichtung der Parameter μ und N . Dominiert einer der beiden Summanden, so liegt die Lösung in der Nähe der beiden Extremfälle:

$$p_n = \frac{e^{\frac{\mu}{k} p_n^0}}{\sum_m e^{\frac{\mu}{k} p_m^0}} \quad \text{für } h_n = 0 \quad \forall n, \quad (3.30)$$

$$p_n = \frac{e^{\frac{h_n}{k p_n}}}{\sum_m e^{\frac{h_m}{k p_m}}} = \frac{h_n}{k w\left(\frac{h_n Z}{k}\right)} \quad \text{für } \mu = 0. \quad (3.31)$$

Man beachte, daß für den Fall, daß keine Daten vorhanden sind ($N = 0$), für großes μ die Wahrscheinlichkeiten p_n nicht gegen p_n^0 , sondern δ -förmig gegen die Stelle n_0 des Maximalwerts von p_n^0 strebt. Ein weiterer wichtiger Grenzfall ist das Abschalten des Entropieterms durch $k \rightarrow 0$. Für die Operatorgleichung (3.18) bekommt man unter obigen Voraussetzungen:

$$\mathcal{I} = \sum_n \left(\frac{\mu p_n^0 + \frac{h_n}{p_n}}{\mu \sum_m p_m p_m^0 + N} \right) | \chi_n > < \chi_n | . \quad (3.32)$$

Die Operatoridentität kann aufgrund der Vollständigkeit nur erfüllt werden, wenn alle Koeffizienten identisch Eins sind:

$$1 = \frac{\mu p_n^0 + \frac{h_n}{p_n}}{\mu \sum_m p_m p_m^0 + N} \quad \forall n. \quad (3.33)$$

Vereinfachung von (3.33) bringt schließlich

$$p_n = \frac{h_n}{\mu \left(\sum_m p_m p_m^0 - p_n^0 \right) + N}, \quad (3.34)$$

woran $p_n \rightarrow \frac{h_n}{N}$ für den Fall von kleiner Gewichtung des Musters ρ_0 durch μ beziehungsweise für große N zu erkennen ist; d.h. die Wahrscheinlichkeiten entsprechen den relativen Häufigkeiten. In einer numerischen Anwendung für diesen Fall wird ein in der Ortsdarstellung diagonaler statistischer Operator $\rho_0 = \int dx p^0(x) | x > < x |$ mit vorgegebenen Wahrscheinlichkeiten $p^0(x)$ als Referenz verwendet. In dem Beispiel aus Abbildung 3.2 wurde eine um $x_0 = 44$ lokalisierte Gauß-Verteilung angenommen. Die Daten wurden aus einer dazu verschobenen Gauß-Verteilung ($x_0 = 55$) mit gleicher Varianz erzeugt. Für verschiedene Kombinationen von μ und k wurde nach Lösungen von (3.29) gesucht. Beim iterativen Lösen von (3.29) muß numerisch mit der Singularität bei $h(x) = 0$ vorsichtig umgegangen werden, indem diese Stellen separat nach (3.28) berechnet werden. Bei diesem Vorgehen zeigte sich eine stabile Konvergenz nach 10 bis 20 Iterationsschritten. In den Ergebnissen können die wesentlichen Vorhersagen für $p(x)$ – wie Tendenz zum Maximalwert von $p^0(x)$ für große μ , Tendenz zur Gleichverteilung für große k und Tendenz zur empirischen Häufigkeit für kleine k und μ – beobachtet werden.

3.4 Lernen unter Minimierung der Entropie

Alternativ kann in (3.7) eine Apriori-Verteilung angesetzt werden, welche einer Minimierung der Entropie entspricht. Dieses ist genau dann der Fall, wenn ρ ein Projektor auf einen reinen Zustand

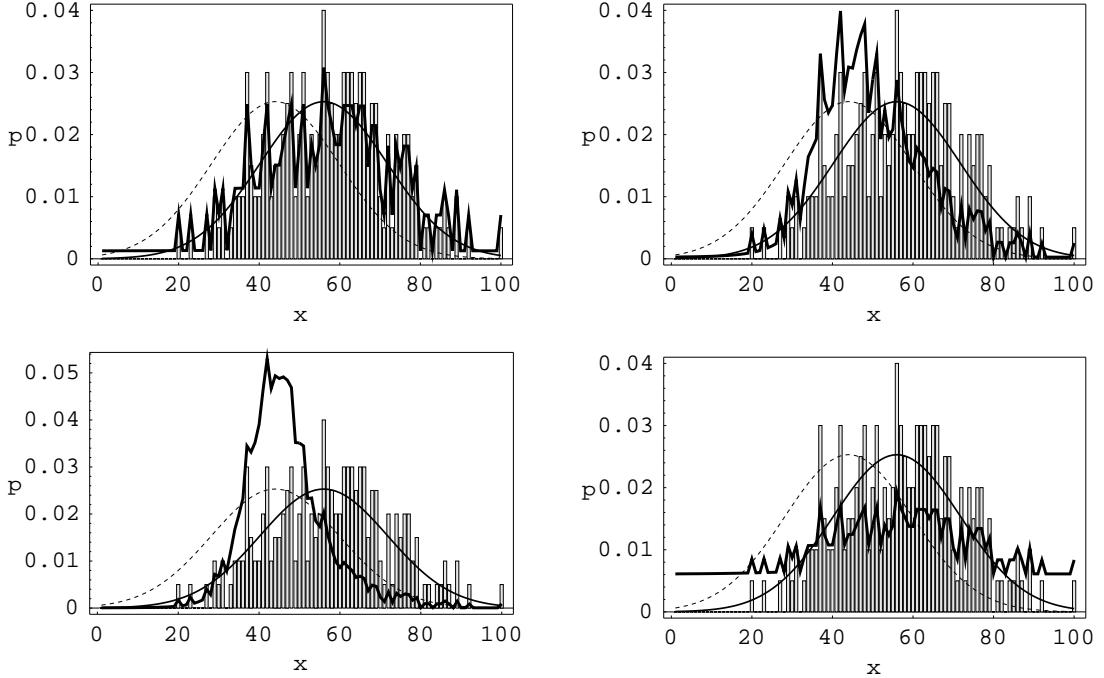


Abbildung 3.2: Lernen statistischer Operatoren. Die dicke durchgezogene Linie kennzeichnet die Wahrscheinlichkeitsverteilung der Lösung $p(x)$. Die dünne durchgezogene Linie stellt die Verteilung dar, aus der die Daten (Balkendiagramm) generiert wurden. Die gestrichelte Linie kennzeichnet die Referenzwahrscheinlichkeiten $p^0(x)$. Die Anzahl der Daten beträgt $N = 200$. Die jeweiligen Parameter sind: oben links $k = 3, \mu = 10$; oben rechts $k = 3, \mu = 500$; unten links $k = 3, \mu = 800$; unten rechts $k = 30, \mu = 100$.

ψ ist, also $\rho^2 = \rho$ oder $\rho = \frac{|\psi\rangle\langle\psi|}{\langle\psi|\psi\rangle}$:

$$p[\rho|\rho_0] = \frac{1}{Z} e^{\mu \text{Tr}[\rho\rho_0]} \delta(\text{Tr}[\rho] - 1) \delta(\rho^2 - \rho). \quad (3.35)$$

Die Bedingung $\rho^2 = \rho$ ist genau dann erfüllt, wenn in einer beliebigen Basis alle Matrixelemente übereinstimmen. Nach Einführung der δ -Funktion in ihrer Spektraldarstellung:

$$\begin{aligned} \delta(\rho^2 - \rho) &= \prod_{\alpha\beta} \delta((\rho^2 - \rho)_{\alpha\beta}) = \int \left(\prod_{\alpha\beta} \frac{d\Delta_{\alpha\beta}}{2\pi} \right) e^{i \sum_{\alpha\beta} \Delta_{\beta\alpha} (\rho^2 - \rho)_{\alpha\beta}} \\ &= \int \left(\prod_{\alpha\beta} \frac{d\Delta_{\alpha\beta}}{2\pi} \right) e^{i \text{Tr}[\Delta(\rho^2 - \rho)]} \end{aligned} \quad (3.36)$$

und Einsetzen von (3.35) mit (3.36) in die Verteilung $p[\rho|\rho_0, D]$ erhält man als Energiefunktional anstatt (3.10):

$$F[\rho, \rho_0, \Delta, \lambda, D] = -\mu \text{Tr}[\rho\rho_0] - \sum_{i=1}^N \ln \left(\text{Tr} \left[|\chi_i\rangle\langle\chi_i| \rho \right] \right) - i\lambda(\text{Tr}[\rho] - 1) - i\text{Tr}[\Delta(\rho^2 - \rho)]. \quad (3.37)$$

Die Stationaritätsgleichungen bezüglich ρ , Δ und λ sind einfach. Zusätzlich zu

$$\text{Tr}[\rho] = 1 \quad \text{und} \quad \rho^2 = \rho \quad (3.38)$$

bekommt man:

$$0 = \frac{\partial F}{\partial \rho_{st}} = -\mu \rho_{ts}^0 - \sum_{i=1}^N \frac{\langle t | \chi_i \rangle \langle \chi_i | s \rangle}{\text{Tr}[\rho | \chi_i \rangle \langle \chi_i |]} - i\lambda \delta_{ts} - i \sum_{\alpha} \Delta_{\alpha s} \rho_{t\alpha} - i \sum_{\beta} \Delta_{t\beta} \rho_{\beta s} + i \Delta_{ts}. \quad (3.39)$$

Damit gilt für die Operatoren

$$0 = \mu \rho_0 + \sum_{i=1}^N \frac{\langle \chi_i | \chi_i \rangle}{\text{Tr}[\rho | \chi_i \rangle \langle \chi_i |]} + i\lambda \mathcal{I} + i\rho \Delta + i\Delta \rho - i\Delta. \quad (3.40)$$

Multipliziert man die Gleichung (3.40) einmal von links und einmal von rechts mit ρ , nutzt dabei $\rho^2 = \rho$ aus und subtrahiert die beiden Gleichungen voneinander, so werden gleichzeitig λ und Δ eliminiert. Das Ergebnis ist der Kommutator

$$0 = \left[\mu \rho_0 + \sum_{i=1}^N \frac{\langle \chi_i | \chi_i \rangle}{\text{Tr}[\rho | \chi_i \rangle \langle \chi_i |]}, \rho \right] = [kA, \rho] \quad (3.41)$$

mit dem Operator A aus (3.14). Zum Lösen von (3.41) wird eine Basis gesucht, in der die beiden Operatoren des Kommutators gleichzeitig diagonal werden. Dabei müssen die Bedingungen (3.38) erfüllt sein. Dies kann erreicht werden, indem man einen statistischen Operator ρ vorgibt, dann den Operator $A(\rho)$ diagonalisiert und einen neuen statistischen Operator wählt, der den Eigenvektor zum größten Eigenwert von A als reinen Zustand beschreibt. Man wiederholt dies bis zur Selbstkonsistenz.

Möchte man alternativ zu (3.41) eine Zustandsgleichung für $|\psi\rangle$ gewinnen, so kann man $\rho = \frac{|\psi\rangle\langle\psi|}{\langle\psi|\psi\rangle}$ in (3.41) einsetzen und anschließend die Operatorgleichung mit $|\psi\rangle$ oder $\langle\psi|$ multiplizieren. Dies führt auf

$$-\sum_{i=1}^N \frac{\langle\psi|\psi\rangle}{\langle\psi|\chi_i\rangle} |\chi_i\rangle = \left(\mu \rho_0 - \mu \frac{\langle\psi|\rho_0|\psi\rangle}{\langle\psi|\psi\rangle} - N \right) |\psi\rangle \quad (3.42)$$

beziehungsweise die adjungierte Gleichung. Eine Möglichkeit zum Lösen von (3.42) wäre das Invertieren des Operators der rechten Seite. Somit könnte eine teilweise nach $|\psi\rangle$ aufgelöste Form erhalten werden. Dieses Vorgehen hat sich in der Praxis als ungünstig herausgestellt, da die Invertierbarkeit des betreffenden Operators während des sukzessiven Einsetzens von $|\psi\rangle$ nicht gewährleistet war. Das Arbeiten mit dem Gradientenabstiegsverfahren nach

$$|\psi^{(j+1)}\rangle = |\psi^{(j)}\rangle - \eta \left(-\sum_{i=1}^N \frac{\langle\psi^{(j)}|\psi^{(j)}\rangle}{\langle\psi^{(j)}|\chi_i\rangle} |\chi_i\rangle - \left(\mu \rho_0 - \mu \frac{\langle\psi^{(j)}|\rho_0|\psi^{(j)}\rangle}{\langle\psi^{(j)}|\psi^{(j)}\rangle} - N \right) |\psi^{(j)}\rangle \right) \quad (3.43)$$

vermeidet die Inversion und führt zum Erfolg. In einer numerischen Anwendung aus Abbildung 3.3 wurde als statistischer Referenzoperator

$$\rho_0 = \frac{e^{-\beta(t+v)}}{\text{Tr}[e^{-\beta(t+v)}]} \quad \text{mit} \quad v(x) = \frac{1}{2}m\omega^2(x-20)^2 \quad (3.44)$$

verwendet. Für den wahren Zustand des Systems wurde ein Impulseigenzustand angenommen.

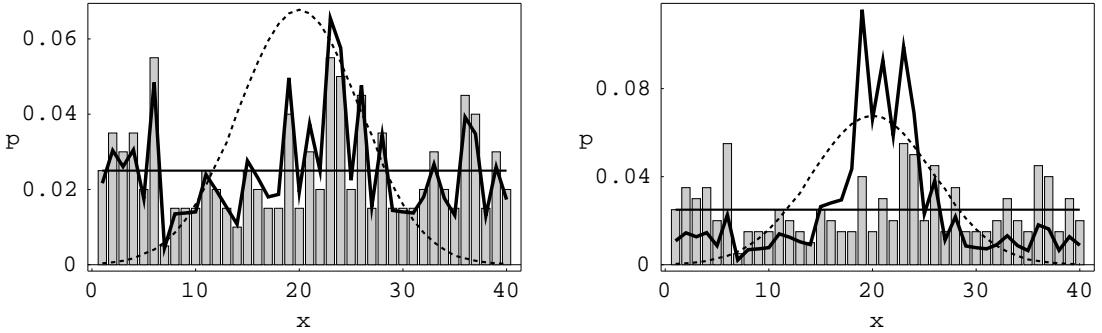


Abbildung 3.3: Lernen statistischer Operatoren unter Verwendung von (3.43). Es sind die Ergebnisse im Ortsraum dargestellt. Die durchgezogene Linie zeigt die Verteilung der Lösung von (3.42), die gestrichelte Linie die Verteilung für den Referenzoperator und die dünne Linie die Verteilung für die Datenerzeugung (inklusive empirischer Datenverteilung). Dabei sind: $N = 200$, $m = 1$, $\eta = 0.002$. Für den Referenzoperator wurden $\beta = 0.08$, $\omega = 0.6$ gewählt. Es wurden unterschiedliche Gewichtungen der Referenz mittels μ getestet. Links: $\mu = 1000$, Rechts: $\mu = 6000$.

Daher wurden die Daten aus einer Gleichverteilung im Ortsraum gewonnen. Die Diagramme, in denen die Verteilungen im Ortsraum dargestellt sind, zeigen vergleichbare Ergebnisse wie in Abbildung 3.2. Bei der Interpretation für diese Verteilungen muß unterschieden werden: Während die Lösung für den statistischen Operator aus Kapitel 3.3 in Ortsdarstellung diagonal ist, also für einen Hilbert–Raum der Dimension n

$$\rho = \begin{pmatrix} p_1 & 0 & \cdots & 0 \\ 0 & p_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & p_n \end{pmatrix} \quad (3.45)$$

gilt, ist dies in diesem Abschnitt nicht der Fall. Hier besteht der statistische Operator aus dem dyadischen Produkt

$$\rho = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_n \end{pmatrix} \cdot (\psi_1^*, \psi_2^*, \dots, \psi_n^*) \quad (3.46)$$

mit den Komponenten $\psi_j = \psi(x_j)$.

Kapitel 4

Lernen von parametrisierten Potentialen

Dieses Kapitel handelt von der Bestimmung von Einteilchenpotentialen, bei denen eine parametrisierte Form vorliegt. Dabei wird davon ausgegangen, daß wiederholt Ortsmessungen an einem kanonischen Ensemble unabhängig voneinander vorgenommen werden. Als Anwendungen werden anharmonische Potentiale diskutiert. Durch die Kenntnis der Parametrisierungen hat man die statistischen Operatoren für die Hoch- und Tieftemperaturgrenzfälle analytisch unter Kontrolle. Die Maximum-Likelihood-Approximation führt auf Stationaritätsgleichungen, welche für die Parameter selbstkonsistent gelöst werden müssen.

4.1 Doppelmuldenpotentiale

Als erstes Beispiel werden hier die sogenannten Doppelmuldenpotentiale betrachtet, welche folgende analytische Struktur besitzen:

$$v(x) = \frac{s}{4}(x - \mu)^4 - \frac{t}{2}(x - \mu)^2 + \frac{t^2}{4s} \quad t, s \geq 0 \quad (4.1)$$

mit den unbekannten Parametern μ , s und t , welche die Verschiebung des Potentials in Ortsrichtung und die jeweiligen Koeffizienten vor quartischem und quadratischem Term festlegen. Die Parametrisierung ist so gewählt, daß die beiden Minima bei den Stellen $x = \mu \pm a$ – mit der Abkürzung $a = \sqrt{t/s}$ – mit dem Funktionswert $v(\mu \pm a) = 0$ auftreten. Dabei gilt für die zweite Ableitung an diesen Stellen $v''(\mu \pm a) = 2t > 0$. Der Funktionswert des lokalen Maximums bei $x = \mu$ ist $v(\mu) = \frac{t^2}{4s}$.

Es wird in diesem Abschnitt davon ausgegangen, daß unabhängige Ortsmessungen an einem statistischen Ensemble im Tieftemperaturfall ($T = 0$) vorgenommen werden. Für zwei "gut getrennte" Minima ist die Wellenfunktion des Grundzustands von $v(x)$ in Näherung eine symmetrische Überlagerung der beiden bei den Minima lokalisierten Grundzustände [13]. Legt man approximativ Oszillatorenpotentiale in die beiden Mulden, so wird die Form durch $v''(\mu \pm a)$ bestimmt. Die Wellenfunktion ψ_{osz} und Grundzustandsenergie E_0 eines bei $x = 0$ lokalisierten harmonischen Oszillatorenpotentials $v_{osz}(x) = \frac{1}{2}m\omega^2x^2$ mit Masse m und Kreisfrequenz ω ist bekanntermaßen durch

$$\psi_{osz}(x) = \left(\frac{1}{\sqrt{\pi}x_0} \right)^{\frac{1}{2}} e^{-\frac{x^2}{2x_0^2}} \quad \text{und} \quad E_0 = \frac{\hbar\omega}{2} \quad (4.2)$$

mit der charakteristischen Länge $x_0 = \sqrt{\frac{\hbar}{m\omega}}$ gegeben. Die Adaption der Größen des Oszillators auf das obige Potential entspricht dann

$$t = \frac{1}{2}m\omega^2 \quad \text{und damit} \quad x_0 = \left(\frac{\hbar^2}{2mt} \right)^{\frac{1}{4}}. \quad (4.3)$$

Berechnet man die Wellenfunktion einer symmetrischen Überlagerung der bei $\mu \pm a$ lokalisierten Oszillatorenwellenfunktionen, so erhält man

$$\Psi(x) = \frac{1}{N_0} [\psi_{osz}(x-\mu-a) + \psi_{osz}(x-\mu+a)] = \frac{1}{N_0} \left(\frac{1}{\sqrt{\pi}x_0} \right)^{\frac{1}{2}} e^{-\frac{(x-\mu)^2+a^2}{2x_0^2}} 2 \cosh \left(\frac{(x-\mu)a}{x_0^2} \right) \quad (4.4)$$

mit der Normierung

$$N_0^2 = 2 \left[1 + e^{-\frac{a^2}{x_0^2}} \right]. \quad (4.5)$$

Die Näherung (4.4) für die Gesamtwellenfunktion des Grundzustands ist nur dann anwendbar, wenn die Energie E_0 wesentlich kleiner als das Potential im Maximum ist. Daraus folgt mit

$$E_0 \ll \frac{t^2}{4s} \quad \Rightarrow \quad \frac{8\hbar^2}{m} \ll \frac{t^3}{s^2} \quad \Rightarrow \quad 2 \ll \frac{a}{x_0} \quad (4.6)$$

eine Bedingung für den Mindestabstand der Minima – skaliert in der charakteristischen Länge x_0 . Dann gilt für die Wahrscheinlichkeit von N unabhängigen Ortsmessungen mit den Ausgängen x_i bei gegebenem Potential:

$$p[x_1, \dots, x_N | \hat{x}, a, x_0, \mu] = \prod_{i=1}^N |\Psi(\mu, a, x_0, x_i)|^2 \sim e^{-F[\mu, a, x_0, x_i]} \quad (4.7)$$

mit dem Exponenten

$$F[\mu, a, x_0, x_i] = N \frac{a^2}{x_0^2} + \frac{1}{x_0^2} \sum_{i=1}^N (x_i - \mu)^2 - 2 \sum_{i=1}^N \ln \left(\cosh \frac{a(x_i - \mu)}{x_0^2} \right) + N \ln \left(x_0 \left(1 + e^{-\frac{a^2}{x_0^2}} \right) \right). \quad (4.8)$$

Für den Fall, daß kein Vorwissen für die Parameter vorhanden ist ($p[a, x_0, \mu] = \text{const}$), entspricht die Maximum-Likelihood-Approximation der Maximum-Aposteriori-Approximation. Im Gegensatz zu den nächsten Kapiteln, wo nichtparametrisierte Potentiale betrachtet werden, bei denen jeder Funktionswert $v(x)$ als eigener Freiheitsgrad aufgefaßt wird, reicht hier die strikte Einschränkung auf Potentiale der Form (4.1) aus, um pathologische Lösungen zu vermeiden. Auch bei Abwesenheit von Vorwissen kann das Potential die Daten nicht durch δ -förmige Lösungen beliebig gut adaptieren. In gewissem Sinn kann die Kenntnis, daß v eine Struktur von (4.1) besitzt, auch als Vorwissen betrachtet werden. Führt man eine Maximum-Likelihood-Approximation durch, so müssen die Ableitungen von (4.8) nach den Parametern μ , a und x_0 berechnet werden.

Die Stationarität von F bezüglich μ liefert direkt mit

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i - \frac{a}{N} \sum_{i=1}^N \tanh \left(\frac{a}{x_0^2} (x_i - \mu) \right) \quad (4.9)$$

eine implizite Beziehung, die die Verschiebung des Potentials festlegt; d.h. μ weicht von dem empirischen Mittelwert der Messungen ab. Allerdings wird für eine große Anzahl von Messungen N der

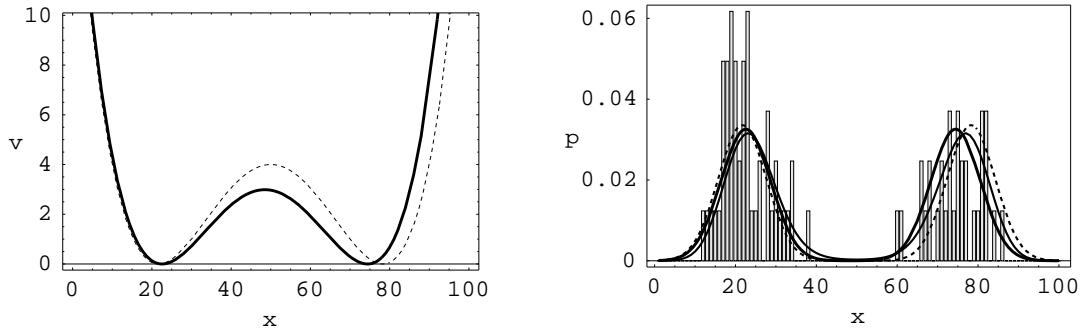


Abbildung 4.1: Lernen von Doppelmuldenpotentialen. Links sind das wahre Potential (gestrichelte Linie) mit den Parametern $t_0 = 0.02$, $s_0 = 0.000025$, $\mu_0 = 50$ und das rekonstruierte Potential (durchgezogene Linie) mit $t = 0.0178$, $s = 0.0000264$, $\mu = 48.50$ angegeben. Rechts sind die dazugehörigen Likelihood–Wahrscheinlichkeiten für den wahren Grundzustand (dünne durchgezogene Linie), den nach (4.4) genäherten Grundzustand des wahren Potentials (gestrichelte Linie) und den Grundzustand des rekonstruierten Potentials dargestellt. Das Balkendiagramm zeigt die empirische Häufigkeit für die Meßdaten.

zweite Summand verschwinden. Mit der Näherung, daß für große N der empirische Erwartungswert einer Funktion $f(x)$ gegen den Erwartungswert unter der datenerzeugenden Wahrscheinlichkeitsverteilung $p(x)$ konvergiert ($\frac{1}{N} \sum_{i=1}^N f(x_i) \rightarrow \int dx f(x)p(x)$), und bei Berücksichtigung, daß der \tanh –Term ungerade und die Wahrscheinlichkeitsverteilung $|\Psi|^2$ gerade bezüglich des Punktes $x = \mu$ ist, folgt:

$$\mu \longrightarrow \frac{1}{N} \sum_{i=1}^N x_i \quad \text{für } N \rightarrow \infty. \quad (4.10)$$

Man beachte weiter, daß im Falle eines großen Abstands der beiden Minima des Potentials wegen

$$\tanh\left(\frac{a}{x_0^2}(x_i - \mu)\right) \longrightarrow \begin{cases} 1 & , x_i > \mu \\ -1 & , x_i < \mu \end{cases} \quad \text{für } \frac{a}{x_0} \rightarrow \infty \quad (4.11)$$

der zweite Summand effektiv zählt, wieviele Daten rechts und links von μ liegen. Stationarität von F bezüglich a liefert für den Abstand

$$a = \frac{1}{N} \left(1 + e^{-\frac{a^2}{x_0^2}} \right) \sum_{i=1}^N (x_i - \mu) \tanh\left(\frac{a}{x_0^2}(x_i - \mu)\right), \quad (4.12)$$

wobei wegen (4.11) der Limes

$$a \longrightarrow \frac{1}{N} \sum_{i=1}^N |x_i - \mu| \quad \text{für } \frac{a}{x_0} \rightarrow \infty \quad (4.13)$$

angegeben werden kann. Nach der Ableitung von F nach x_0 bekommt man unter Ausnutzung von (4.12) die kompakte Form für die charakteristische Länge:

$$x_0^2 = \frac{2}{N} \sum_{i=1}^N (x_i - \mu)^2 - \frac{2a^2}{1 + e^{-\frac{a^2}{x_0^2}}} \longrightarrow 2 \left(\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2 - a^2 \right) \quad \text{für } \frac{a}{x_0} \rightarrow \infty. \quad (4.14)$$

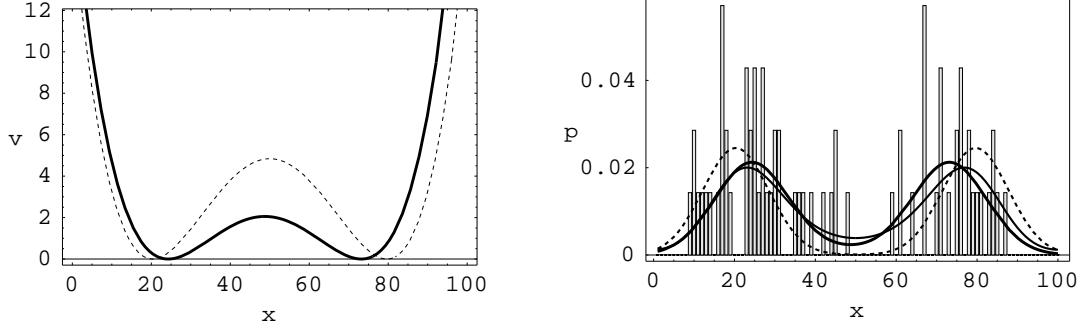


Abbildung 4.2: Verletzung der Bedingung $\frac{a}{x_0} \gg 2$. Die Grundzustandsenergien liegen bei $E_0 = 2.41$ für v_D und $E_0 = 2.30$ für v . Die Parameter sind $t_0 = 0.022$, $s_0 = 0.000025$, $\mu_0 = 50$, $t = 0.014$, $s = 0.000023$, $\mu = 48.7$. Es wurde $N = 70$ Daten betrachtet.

Die Gleichungen (4.9), (4.12) und (4.14) müssen gleichzeitig selbstkonsistent gelöst werden. Bei dem Ergebnis muß darauf geachtet werden, daß (4.6) erfüllt ist. Aus den stationären Werten für a und x_0 können s und t berechnet werden.

Numerisch bietet sich ein Iterationsverfahren an, bei dem mit den obigen Abschätzungen für μ , a und x_0 für den Fall $a \gg x_0$ gestartet wird. Diese Werte können auf den rechten Seiten der Bestimmungsgleichungen (4.9), (4.12) und (4.14) eingesetzt werden, und mit den jeweiligen Ergebnissen kann analog bis zur Selbstkonsistenz fortgefahrene werden. Dieses Vorgehen zeigt eine gute Konvergenz nach sehr wenigen Iterationsschritten. In der Anwendung wurden die Werte μ_0 , t_0 und s_0 des wahren Potentials v_D vorgegeben und bezüglich dieses Potentials Daten erzeugt, indem v_D auf einem Gitter mit $n = 100$ Stützstellen diskretisiert wurde (Gitterkonstante: $\Delta x = 1$). Für den diskreten Hamilton-Operator mit gegebener Teilchenmasse $m = 0.005$ und $\hbar = 1$ wurde der Grundzustand bestimmt und aus der Grundzustandswahrscheinlichkeitsdichte 81 Daten gezogen. In Abbildung 4.1 sind die Ergebnisse des Computerexperiments zusammengefaßt. Es zeigt sich eine passable Rekonstruktion des wahren Potentials, wobei in der Lösung datenspezifische Merkmale enthalten sind. So wird die Dichte der linken Potentialmulde durch die Daten besser wiedergegeben als die der rechten. Die Grundzustandsenergien liegen mit $E_0 = 1.34$ für v_D und $E_0 = 1.33$ für v eindeutig unterhalb der jeweiligen Werte der lokalen Maxima. Das Verhältnis $\frac{a}{x_0}$ erfüllt mit 3.36 für v_D und 2.99 für v die notwendige Bedingung aus (4.6). Es zeigt sich aber auch eine Abweichung von exakter und genauerer Verteilung für v_D . Dies zeigt die Grenzen des Modells auf. Um diese noch zu verdeutlichen, wird nun die Bedingung (4.6) bewußt verletzt, beispielsweise indem die Masse des betrachteten Teilchens auf $m = 0.0013$ reduziert und dadurch die Grundzustandsenergie vergrößert wird. Abbildung 4.2 zeigt das Resultat unter diesen Voraussetzungen. Die hier auftretenden Verhältnisse $\frac{a}{x_0}$ betragen 2.6 für v_D und 1.9 für v . Bei erhöhter Energie wird durch Tunneleffekte die Antreffwahrscheinlichkeit des Teilchens zwischen den beiden Mulden größer. Die Tunnelprozesse werden bei der einfachen Überlagerung der Wellenfunktionen nach (4.4) unzureichend berücksichtigt, und somit gibt auch die Lösung v die in diesem Fall falschen Voraussetzungen wieder. Dieses zeigt sich auch beim Betrachten der beiden Verteilungen in Abbildung 4.2.

4.2 Anharmonischer Oszillator

Als zweites Beispiel wird ein Potential betrachtet, daß aus einem gegebenen Potential eines harmonischen Oszillators besteht und durch einen quartischen Anteil mit unbekanntem Gewicht gestört wird. Dies entspricht einer ähnlich parametrisierten Form wie (4.1), aber mit unterschiedlichen

Vorzeichen:

$$v(x) = \frac{1}{2}\alpha x^2 + \frac{1}{4}gx^4 \quad g \geq 0 \quad \alpha > 0, \quad (4.15)$$

wobei die unbekannte Größe g durch Ortsmessungen an einem kanonischen Ensemble bestimmt werden soll. Wählt man die Darstellung mittels euklidischer Pfadintegrale [14], so können die Matrixelemente des statistischen Operators und damit die Wahrscheinlichkeiten für die einzelnen Ortsmessungen bei beliebiger Temperatur über

$$p[x_i | \hat{x}, g] = \langle x_i | \frac{e^{-\beta H}}{Z} | x_i \rangle = \frac{1}{Z} \int_{q(0)=x_i}^{q(\beta\hbar)=x_i} \frac{[dq(t)]}{\mathcal{N}} e^{-\frac{1}{\hbar} \int_0^{\beta\hbar} dt \left[\frac{1}{2}m\dot{q}^2(t) + \frac{1}{2}\alpha q^2(t) + \frac{1}{4}gq^4(t) \right]} \quad (4.16)$$

ausgedrückt werden. Die Benutzung der Pfadintegraldarstellung hat den Vorteil, daß eine analytische Verarbeitung der Likelihood-Dichte vorgenommen werden kann. Dies ist bei der weiter unten in Kapitel 5 nach (5.14) verwendeten Energiedarstellung nicht der Fall, da Eigenfunktionen und Eigenwerte des Hamilton-Operators für beliebige g nicht exakt vorliegen.

Das weitere Vorgehen ist dadurch gekennzeichnet, eine Näherung für das Pfadintegral aus (4.16) zu finden, indem der quartische Term durch Einführung eines Hilfsintegrals in einen effektiven quadratischen Term umgewandelt wird. Danach kann eine auf komplexe Kreisfrequenzen verallgemeinerte Rechnung des harmonischen Oszillators vorgenommen werden.

Wie gewöhnlich diskretisiert man die Zeitvariable t , indem man das Zeitintervall $[0, \beta\hbar]$ in M Abschnitte der Länge $\epsilon = \frac{\hbar\beta}{M}$ aufteilt. Mit der Abkürzung $q(\epsilon k) = q_k$ für $k \in \{1, \dots, M\}$ bekommt man

$$e^{-\frac{g}{4\hbar} \int_0^{\beta\hbar} dt q^4(t)} \longrightarrow e^{-\frac{g}{4\hbar} \sum_{k,k'}^M \epsilon q_k^2 q_{k'}^2 \delta_{kk'}} = e^{\frac{1}{2} \sum_{k,k'} \left(\frac{i\epsilon g}{\hbar} q_k^2 \right) A(k, k') \left(\frac{i\epsilon g}{\hbar} q_{k'}^2 \right)} \quad (4.17)$$

mit der positiv-definiten Matrix $A(k, k') = \frac{\hbar}{2g\epsilon} \delta_{kk'}$. Mittels Einführung eines Gauß-Integrals über eine Hilfsvariable λ_k gelingt die Umwandlung des quartischen Terms:

$$e^{-\frac{g}{4\hbar} \sum_{k,k'}^M \epsilon q_k^2 q_{k'}^2 \delta_{kk'}} = \frac{1}{\sqrt{\det A}} \int \left(\prod_{k=1}^M \frac{d\lambda_k}{\sqrt{2\pi}} \right) e^{-\frac{1}{2} \sum_{k,k'} \lambda_k A^{-1}(k, k') \lambda_{k'} - i \frac{\epsilon g}{\hbar} \sum_k \lambda_k q_k^2}. \quad (4.18)$$

Bildet man den Grenzübergang $\epsilon \rightarrow 0$, $M \rightarrow \infty$, wobei $\epsilon M = \beta\hbar$ endlich bleibt, und setzt das Ergebnis in (4.16) ein, so erhält man für die Matrixelemente

$$\langle x_i | e^{-\beta H} | x_i \rangle = \int \frac{[d\lambda(t)]}{\mathcal{N}'(g)} \int \frac{[dq(t)]}{\mathcal{N}} e^{-\frac{1}{\hbar} \int_0^{\beta\hbar} dt \left[\frac{1}{2}m\dot{q}^2(t) + \frac{1}{2}(\alpha + i2g\lambda(t))q^2(t) + g\lambda^2(t) \right]}. \quad (4.19)$$

Bei der Gleichung (4.19) handelt es sich um eine Identität. Der anharmonische Anteil des Potentials wurde mit Hilfe des λ -Integrals in ein Oszillatormodell mit zeitabhängiger Kreisfrequenz umgewandelt.

Die Vorteile der Darstellung (4.19) liegen in der Möglichkeit, Näherungen vorzunehmen. Schränkt man beispielsweise die Integration der Variablen $\lambda(t)$ auf statische Felder λ ein, so erhält man für (4.19) bis auf eine x_i -unabhängige Normierungskonstante:

$$\langle x_i | e^{-\beta H} | x_i \rangle \sim \int d\lambda e^{-g\beta\lambda^2} \int_{q(0)=x_i}^{q(\beta\hbar)=x_i} \frac{[dq(t)]}{\mathcal{N}} e^{-\frac{1}{\hbar} \int_0^{\beta\hbar} dt \left[\frac{1}{2}m\dot{q}^2(t) + \frac{1}{2}m\omega^2(\lambda)q^2(t) \right]} \quad (4.20)$$

mit

$$\omega(\lambda) = \pm \sqrt{\frac{\alpha}{m} + i \frac{2g\lambda}{m}}. \quad (4.21)$$

Es wird bei der Wahl des Vorzeichens von ω angenommen, daß es einen positiven Realteil gewährleistet. (4.20) entspricht einer Summation über harmonische Oszillatoren mit abgeänderten, komplexen Kreisfrequenzen. Nimmt man eine Variablentransformation $r(t) = q(t) - q_c(t)$ vor, wobei q_c die Lösung der klassischen Bewegungsgleichung

$$\ddot{q}_c(t) = \omega^2 q_c(t) \quad \text{mit} \quad q_c(0) = q_c(\beta\hbar) = x_i \quad (4.22)$$

ist, so bekommt man

$$\begin{aligned} \langle x_i | e^{-\beta H} | x_i \rangle &\sim \int d\lambda e^{-g\beta\lambda^2 - \frac{1}{\hbar} \int_0^{\beta\hbar} dt \left[\frac{1}{2}m\dot{q}_c^2(t) + \frac{1}{2}m\omega^2(\lambda)q_c^2(t) \right]} \times \\ &\quad \int_{r(0)=0}^{r(\beta\hbar)=0} \frac{[dr(t)]}{\mathcal{N}} e^{-\frac{1}{\hbar} \int_0^{\beta\hbar} dt \left[\frac{1}{2}mr^2(t) + \frac{1}{2}m\omega^2(\lambda)r^2(t) \right]}. \end{aligned} \quad (4.23)$$

Dabei wurde mit (4.22) ausgenutzt, daß die Wirkung additiv für die Argumente ist:

$$S[q] = S[q_c] + S[r] \quad \text{mit} \quad S[q] = \int_0^{\beta\hbar} dt \left[\frac{1}{2}m\dot{q}^2(t) + \frac{1}{2}m\omega^2 q^2(t) \right]. \quad (4.24)$$

Die Bewegungsgleichung (4.22) kann durch einen Ansatz $q_c(t) = Ae^{\omega t} + Be^{-\omega t}$ gelöst werden. Dies führt auf das Ergebnis:

$$q_c(t) = \frac{x_i}{\sinh(\beta\hbar\omega)} (\sinh(\omega t) + \sinh(\omega(\beta\hbar - t))). \quad (4.25)$$

Damit erhält man für den stationären Wert $S[q_c]$ durch einmalige partielle Integration und Ausnutzen von (4.22)

$$\begin{aligned} S[q_c] &= \int_0^{\beta\hbar} dt \left(\frac{1}{2}m\dot{q}_c^2(t) + \frac{1}{2}m\omega^2 q_c^2(t) \right) = \left[\frac{1}{2}m q_c(t) \dot{q}_c(t) \right]_0^{\beta\hbar} \\ &= \frac{m\omega x_i^2}{\sinh(\beta\hbar\omega)} (\cosh(\beta\hbar\omega) - 1) = m\omega x_i^2 \tanh \frac{\beta\hbar\omega}{2}. \end{aligned} \quad (4.26)$$

Das Pfadintegral aus (4.23) über Fluktuationen $r(t)$ ist ebenfalls exakt ausführbar. Integriert man das Integral der kinetischen Energie einmal partiell und nutzt die Randbedingung $r(0) = r(\beta\hbar) = 0$ aus, so bekommt man:

$$I = \int \frac{[dr(t)]}{\mathcal{N}} e^{-\frac{m}{2\hbar} \int_0^{\beta\hbar} dt dt' r(t) \delta(t-t') \left(-\frac{\partial^2}{\partial t'^2} + \omega^2 \right) r(t')} \quad (4.27)$$

mit der Integralnormierung [14]

$$\mathcal{N} \sim \int [dy(t)] e^{-\frac{m}{2\hbar} \int_0^{\beta\hbar} dt dt' y(t) \delta(t-t') \left(-\frac{\partial^2}{\partial t'^2} \right) y(t')}. \quad (4.28)$$

Die Pfadintegrale aus (4.27) und (4.28) sind hochdimensionale Gauß–Integrale und werden durch die Determinanten von Operatoren der Form

$$A_\omega(t, t') = \frac{m}{\hbar} \delta(t - t') \left(-\frac{\partial^2}{\partial t'^2} + \omega^2 \right) \quad (4.29)$$

bestimmt. Berechnet man die Determinanten über das Produkt der Eigenwerte von A_ω , so müssen die Eigenwertgleichungen

$$\left(-\frac{\partial^2}{\partial t'^2} + \omega^2 \right) r_n(t) = \mu_n r_n(t) \quad \text{mit} \quad r_n(0) = r_n(\beta\hbar) = 0 \quad (4.30)$$

gelöst werden. Die normierten Eigenfunktionen r_n mit den dazugehörigen Eigenwerten können mit

$$r_n(t) = \sqrt{\frac{2}{\beta\hbar}} \sin \frac{\pi nt}{\beta\hbar} \quad \text{mit} \quad \mu_n = \frac{\pi^2 n^2}{\beta^2 \hbar^2} + \omega^2 \quad n > 0 \quad (4.31)$$

analytisch angegeben werden. Das Gauß-Integral (4.27) ist konvergent. Denn setzt man ω aus (4.21) in die Eigenwerte von A_ω ein, so folgt mit

$$\operatorname{Re} \mu_n = \frac{\pi^2 n^2}{\beta^2 \hbar^2} + \frac{\alpha}{m} > 0 \quad (4.32)$$

ein positiver Realteil. Damit erhält man für (4.27) unter Ausnutzung der Produktdarstellung des Sinushyperbolicus

$$I \sim \sqrt{\frac{\det A_{\omega=0}}{\det A_\omega}} = \sqrt{\prod_{n>0} \frac{\frac{\pi^2 n^2}{\beta^2 \hbar^2}}{\frac{\pi^2 n^2}{\beta^2 \hbar^2} + \omega^2}} = \frac{1}{\sqrt{\prod_{n>0} \left(1 + \frac{\beta^2 \hbar^2 \omega^2}{\pi^2 n^2}\right)}} = \sqrt{\frac{\beta \hbar \omega}{\sinh(\beta \hbar \omega)}}. \quad (4.33)$$

Man beachte, daß ω weiterhin komplex ist. Zusammenfassend sind die Matrixelemente dann bis auf einen x_i -unabhängigen Proportionalitätsfaktor, der sich später durch die Zustandssumme wegbürtzt, in der Näherung statischer λ -Felder durch

$$\langle x_i | e^{-\beta H} | x_i \rangle \sim \int d\lambda e^{-g\beta\lambda^2 - \frac{m\omega(\lambda)x_i^2}{\hbar}} \tanh\left(\frac{\omega(\lambda)\hbar\beta}{2}\right) \sqrt{\frac{\beta\hbar\omega(\lambda)}{\sinh(\beta\hbar\omega(\lambda))}} \quad (4.34)$$

gegeben.

Des Weiteren werden wegen der schlechten Handhabbarkeit der λ -Abhängigkeiten in (4.34) Hoch- und Tieftemperaturgrenzwerte betrachtet und eine Stationäre Phasenapproximation des λ -Integrals vorgenommen. Im Hochtemperaturgrenzwert $\beta \rightarrow 0$ können die hyperbolischen Funktionen durch ihr Argument genähert werden ($\tanh z \rightarrow z$ und $\sinh z \rightarrow z$), so daß für (4.34) folgt

$$\langle x_i | e^{-\beta H} | x_i \rangle \sim \int d\lambda e^{-g\beta\lambda^2 - \frac{m\beta\omega^2}{2}x_i^2}. \quad (4.35)$$

Die Stationaritätsgleichung lautet dann:

$$0 = -2g\beta\lambda - \frac{1}{2}m\beta x_i^2 \frac{\partial(\omega^2)}{\partial\lambda} = -2g\beta\lambda - i\beta g x_i^2, \quad (4.36)$$

wobei $\frac{\partial\omega}{\partial\lambda} = i\frac{g}{m\omega}$ nach (4.21) ausgenutzt wurde. Damit folgt für die Sattelpunkte

$$\lambda = -i\frac{x_i^2}{2} \quad \text{und} \quad \omega^2 = \frac{\alpha}{m} + \frac{g}{m}x_i^2. \quad (4.37)$$

Setzt man die stationären Werte für λ wieder in den Integranden ein und nutzt aus, daß die quadratische Korrektur zur Sattelpunktsnäherung unabhängig von x_i ist, da nach (4.36) die zweite Ableitung des Exponenten nach λ unabhängig von x_i ist, so ergibt sich:

$$\langle x_i | e^{-\beta H} | x_i \rangle \sim e^{-\frac{\alpha\beta}{2}x_i^2 - \frac{\beta g}{4}x_i^4} = e^{-\beta v(x_i)}. \quad (4.38)$$

Für diesen Grenzfall ist dann die Wahrscheinlichkeit für eine Ortsmessung für gegebenes Potential näherungsweise:

$$p[x_i | \hat{x}, g] = \frac{1}{Z(g)} e^{-\frac{\alpha\beta}{2}x_i^2 - \frac{\beta g}{4}x_i^4} \quad \text{für } \beta \rightarrow 0 \quad (4.39)$$

mit der Zustandssumme

$$Z(g) = \int dx e^{-\frac{\alpha\beta}{2}x^2 - \frac{\beta g}{4}x^4}. \quad (4.40)$$

Somit werden die Likelihood–Wahrscheinlichkeiten im Hochtemperaturgrenzfall unabhängig von m und \hbar . Dieses Resultat ist nicht überraschend. So entspricht (4.39) dem Ergebnis der klassischen statistischen Mechanik für das kanonische Ensemble. Betrachtet man den Phasenraum mit den dazugehörigen Orts– und Impulskoordinaten \tilde{q}, \tilde{p} , so ist die Hamilton–Funktion von der Form $H(\tilde{q}, \tilde{p}) = \frac{1}{2m}\tilde{p}^2 + v(\tilde{q})$. Die verbundene Wahrscheinlichkeit $p[\tilde{q}, \tilde{p}]$ im kanonischen Ensemble, das Teilchen gleichzeitig mit dem Koordinatenpaar \tilde{q}, \tilde{p} zu messen, ist dann:

$$p[\tilde{q}, \tilde{p}] = \frac{1}{Z} e^{-\beta H(\tilde{q}, \tilde{p})} \quad \text{mit} \quad Z = \int d\tilde{q} d\tilde{p} e^{-\beta H(\tilde{q}, \tilde{p})}. \quad (4.41)$$

Bei der Berechnung der Wahrscheinlichkeit $p[\tilde{q}]$ müssen die Impulse ausintegriert werden. Aufgrund der rein additiven Form der Hamilton–Funktion kürzen sich die Beiträge der kinetischen Energie in der Boltzmann–Verteilung heraus, und man erhält (4.39):

$$p[\tilde{q}] = \int d\tilde{p} p[\tilde{q}, \tilde{p}] = \frac{e^{-\beta v(\tilde{q})}}{\int d\tilde{q} e^{-\beta v(\tilde{q})}}. \quad (4.42)$$

Im Tieftemperaturfall kann analog vorgegangen werden: Durch die Vorzeichenwahl für ω in (4.21) kann $\text{Re}\omega > 0$ erzwungen werden. Daraus folgen die Grenzwerte

$$\tanh \frac{\beta \hbar \omega}{2} \rightarrow 1 \quad \text{und} \quad \sinh(\beta \hbar \omega) \rightarrow \frac{1}{2} e^{\beta \hbar \omega} \quad \text{für } \beta \rightarrow \infty. \quad (4.43)$$

Mit dieser Abschätzung gilt für die Matrixelemente (4.34)

$$\langle x_i | e^{-\beta H} | x_i \rangle \sim \int d\lambda e^{-g\beta\lambda^2 - \frac{m\omega(\lambda)}{\hbar}x_i^2 + \frac{1}{2}\ln(\beta\hbar\omega(\lambda)) - \frac{1}{2}\beta\hbar\omega(\lambda)}, \quad (4.44)$$

und die Stationaritätsgleichung lautet unter Berücksichtigung der führenden Terme in β für $\beta \rightarrow \infty$:

$$0 = -2g\beta\lambda - \frac{1}{2}\beta\hbar\frac{\partial\omega}{\partial\lambda} = -2g\beta\lambda - i\frac{\beta\hbar g}{2m\omega}. \quad (4.45)$$

Damit ergibt sich für Zusammenhang von λ und ω :

$$\lambda = -i\frac{\hbar}{4m\omega} \quad \text{oder} \quad \omega^2 = \frac{\alpha}{m} + \frac{g\hbar}{2m^2\omega}. \quad (4.46)$$

Als stationären Wert erhält man eine Gauß–Verteilung in x_i , allerdings mit einer gegenüber dem ungestörten harmonischen Oszillatoren abgeänderten Kreisfrequenz, die durch die kubische Gleichung (4.46) gegeben ist. Die Zustandssumme lässt sich als Gauß–Integral leicht ausführen, und man erhält schließlich:

$$\langle x_i | \frac{e^{-\beta H}}{Z} | x_i \rangle = \left(\frac{m\omega}{\pi\hbar} \right)^{\frac{1}{2}} e^{-\frac{m\omega}{\hbar}x_i^2} \quad \text{mit} \quad \omega^3 - \frac{\alpha}{m}\omega - \frac{g\hbar}{2m^2} = 0. \quad (4.47)$$

Die kubische Gleichung für ω kann mit Hilfe von Mathematica gelöst werden. Es zeigt sich, daß für $\alpha, g > 0$ nur eine der drei Nullstellen den erforderlichen positiven Realteil besitzt. Die effektive Kreisfrequenz ist erwartungsgemäß größer als die des ungestörten harmonischen Oszillators: $\omega \geq \omega_0 = \sqrt{\frac{\alpha}{m}}$.

Die Qualität der für den Hoch– und Tieftemperaturfall gemachten Näherungen ist in Abbildung 4.3 dargestellt. Es zeigt sich eine sehr gute Approximation der Likelihood–Verteilung durch die

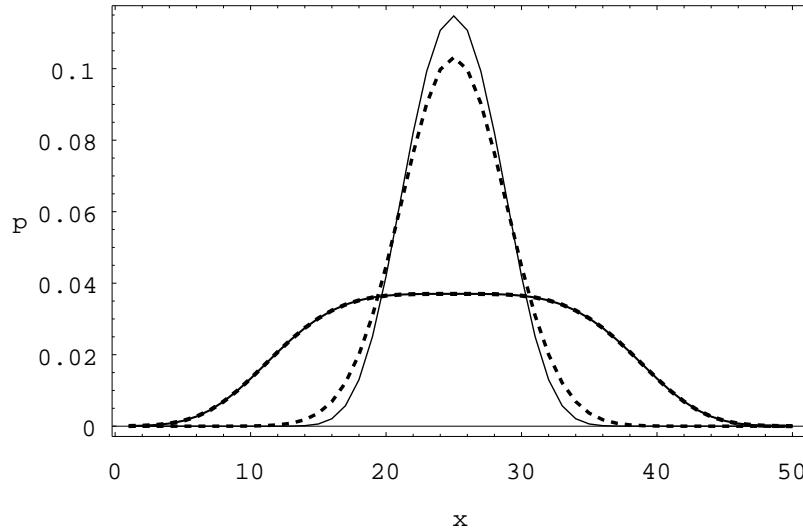


Abbildung 4.3: Exakte und genäherte Likelihood-Wahrscheinlichkeiten des anharmonischen Oszillators im Hoch- und Tieftemperaturfall. Die durchgezogenen Linien zeigen die auf einem Gitter von 50 Stützstellen exakten Dichten für eine hohe Temperatur $\beta_1 = 0.008$ und eine sehr kleine Temperatur $\beta_2 = 10000$ für einen bei $x = 25$ lokalisierten anharmonischen Oszillatoren mit den Parametern $g = 0.01$, $\alpha = 0.01$, $m = 0.007$. Die gestrichelten Linien zeigen die in (4.39) und (4.47) berechneten Näherungen. Es wurde $\hbar = 1$ und $\Delta x = 1$ angenommen.

klassischen Wahrscheinlichkeiten für größere Temperaturen, während bei der Betrachtung des exakten und genäherten Grundzustands im Tieftemperaturlimes Abweichungen auftreten. Die nach (4.47) genäherte Verteilung weist eine größere Varianz als die auf dem Gitter exakte Verteilung auf. Im Grenzwert $g \rightarrow 0$ oder $m \rightarrow \infty$ verschwindet diese Diskrepanz, und die Ergebnisse des harmonischen Oszillators werden reproduziert.

Im folgenden wird eine Maximum-Likelihood-Methode für die Grenzfälle durchgeführt. Dadurch, daß die Likelihood-Dichten analytisch vorliegen, ist der numerische Aufwand zum Lösen der entsprechenden Sattelpunktsgleichungen sehr einfach.

Für die Likelihood-Wahrscheinlichkeit für unabhängige Messung von N Daten bekommt man bei $\beta \rightarrow 0$:

$$p[x_1, \dots, x_N | \hat{x}, g] = \frac{1}{Z^N(g)} e^{-\frac{\alpha\beta}{2} \sum_{i=1}^N x_i^2 - \frac{\beta g}{4} \sum_{i=1}^N x_i^4}, \quad (4.48)$$

und die Maximum-Likelihood-Approximation für (4.48) liefert die Bedingung

$$\frac{1}{N} \sum_{i=1}^N x_i^4 = \frac{1}{Z(g)} \int dx e^{-\frac{\alpha\beta}{2} x^2 - \frac{\beta g}{4} x^4} x^4 = \langle x^4 \rangle \quad (4.49)$$

für die Bestimmung von g . Es muß also nach Potentialen gesucht werden, bei denen der empirische Erwartungswert von x^4 mit dem statistischen Wert zusammenfällt. In einer numerischen Anwendung auf einem Gitter der Größe $n = 50$ wurde eine Kopplungskonstante g_0 für einen bei $x = 25$ lokalisierten Oszillatoren für die Datenerzeugung vorgegeben. Verwendet man das in Kapitel 1.6 vorgestellte Newton-Verfahren zum Lernen von g nach Gleichung (4.49), so zeigt sich eine sehr

gute Konvergenz schon nach wenigen Iterationsschritten. In Abbildung 4.4 ist die Anwendbarkeit der Theorie demonstriert. Für den Fall von $N = 70$ Datenpunkten liegt die gelernte Kopplungskonstante g nahe bei g_0 .

Im Tieftemperaturfall ist die Likelihood-Dichte durch

$$p[x_1, \dots, x_N | \hat{x}, g] = \left(\frac{m\omega}{\pi\hbar} \right)^{\frac{N}{2}} e^{-\frac{m\omega}{\hbar} \sum_{i=1}^N x_i^2} \quad (4.50)$$

mit $\omega = \omega(g)$ gegeben. Die Maximierung von (4.50) bezüglich g liefert sofort

$$\omega = \frac{N\hbar}{2m} \frac{1}{\sum_{i=1}^N x_i^2}, \quad (4.51)$$

und mit (4.47) folgt als Ergebnis für g :

$$g = \frac{2m^2}{\hbar} \left(\omega^3 - \frac{\alpha}{m} \omega \right) = \frac{N^3 \hbar^2}{4m \left(\sum_{i=1}^N x_i^2 \right)^3} - \frac{\alpha N}{\sum_{i=1}^N x_i^2}. \quad (4.52)$$

Als Spezialfall $\alpha = 0$ gilt dann direkt:

$$g = \frac{N^3 \hbar^2}{4m \left(\sum_{i=1}^N x_i^2 \right)^3}. \quad (4.53)$$

Abbildung 4.5 zeigt ein numerisches Beispiel für das Lernen nach (4.52).

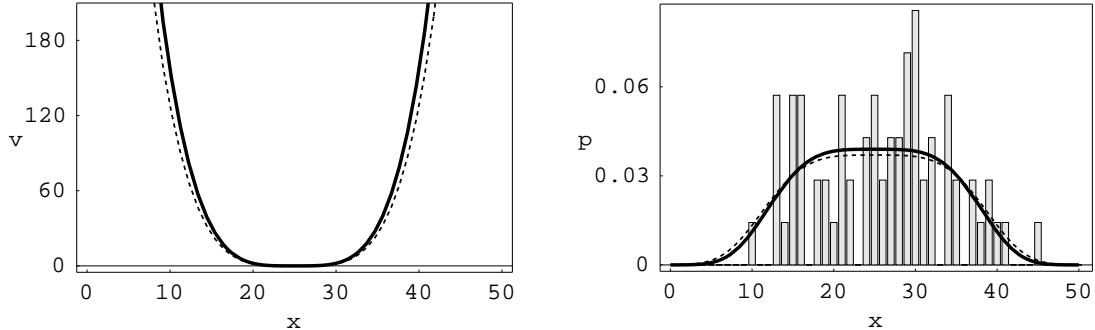


Abbildung 4.4: Lernen am anharmonischen Oszillatoren für große Temperaturen. Für ein Teilchen der Masse $m = 0.1$ wurden $N = 70$ Daten für ein Potential mit $\alpha = 0.01$ und $g_0 = 0.01$ bei einer inversen Temperatur von $\beta = 0.008$ gewonnen ($\hbar = 1$, $\Delta x = 1$). Das wahre Potential und die Wahrscheinlichkeitsverteilung sind in den Diagrammen gestrichelt dargestellt. Die gelernte Kopplungskonstante beträgt $g = 0.0123$. Das gelernte Potential und die dazugehörige Dichte sind die durchgezogenen Linien. Der Erwartungswert für eine Energiemessung liegt bei $\langle H \rangle = 41.0$.

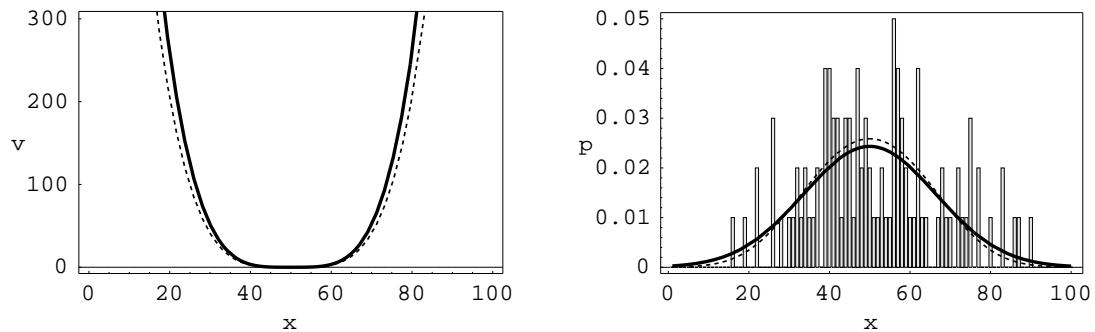


Abbildung 4.5: Lernen am anharmonischen Oszillatoren für tiefe Temperaturen. Die Bedeutung der Kurven ist analog zu Abbildung 4.4 gewählt. Die wahre und gelernte Kopplungskonstante betragen $g_0 = 0.001$ und $g = 0.00127$. Die übrigen Parameter sind $N = 100$, $m = 0.00001$, $\alpha = 0.005$. Die Grundzustandsenergie liegt bei $E_0 = 90.4$.

Kapitel 5

Lernen von Potentialen (Energiedarstellung)

Ziel dieses Kapitels ist, durch Ortsmessung an einem kanonischen Ensemble bei bekannter, fixierter inverser Temperatur $\beta = 1/(kT)$ Aussagen über den zugrundeliegenden Einteilchen–Hamilton–Operator $H = t + v$ bei unbekanntem Potential v zu machen. Um die Unabhängigkeit der Ortsmessungen zu gewährleisten, wird davon ausgegangen, daß im Falle wiederholter Messungen an einem einzelnen Ensemblemitglied nach jeder Messung dem System genügend Zeit gelassen wird, um erneut in das thermische Gleichgewicht zu relaxieren. Die Datenmenge bei der wiederholten Messung von \hat{x} wird dabei mit $D = \{x_i | i = 1, \dots, N\}$ beschrieben, wobei x_i hier die Meßausgänge und nicht wie in (1.7) die Meßobservablen kennzeichnet. Je nach Dimension des betrachteten Problems handelt es sich bei x_i um einen Vektor oder eine skalare Zahl.

Die Beschränkung auf Ortsmessungen ist für die Problemstellung nicht essentiell, bietet sich aber aufgrund der Annahme an, daß das Potential diagonal in den Ortskoordinaten ist. So ist in der Ortsdarstellung (Basis $|x\rangle$) der Hamilton–Operator für ein Teilchen der Masse m gegeben durch:

$$H(x, x') = \langle x | H | x' \rangle = \delta(x - x') \left(-\frac{\hbar^2}{2m} \Delta_{x'} + v(x') \right) \quad (5.1)$$

mit dem skalaren, reellen Feld v . Es wird momentan stillschweigend angenommen, daß der Hamilton–Operator nach unten beschränkt ist. Hierfür ist hinreichend, daß $v(x')$ nach unten beschränkt ist. Weiter wird davon ausgegangen, daß die Eigenwerte von H nicht entartet sind.

Die orthogonalen Eigenfunktionen mit den dazugehörigen reellen Eigenwerten werden über die Eigenwertgleichung bestimmt:

$$H | \varphi_n \rangle = E_n | \varphi_n \rangle \quad \text{mit} \quad -\infty < E_0 < E_n, \forall n. \quad (5.2)$$

Der quantenmechanische Zustand des Ensembles im thermischen Gleichgewicht wird bekanntermaßen durch den statistischen Operator

$$\rho = \frac{e^{-\beta H}}{\text{Tr}[e^{-\beta H}]} = \sum_n \frac{e^{-\beta E_n}}{Z} \frac{|\varphi_n\rangle \langle \varphi_n|}{\langle \varphi_n | \varphi_n \rangle} \quad , \quad Z = \sum_n e^{-\beta E_n} \quad (5.3)$$

festgelegt; d.h. ρ ist eine eindeutige Parametrisierung von β und H . Für die Wahrscheinlichkeit einer Ortsmessung mit dem Ergebnis x_i folgt dann:

$$p[x_i | \hat{x}, \rho] = \text{Tr} \left[|x_i\rangle \langle x_i| \rho \right] = \langle x_i | \frac{e^{-\beta H}}{Z} | x_i \rangle, \quad (5.4)$$

wobei $\hat{x} | x_i > = x_i | x_i >$ mit $< x_i | x_i > = 1$ gilt. In diesem Fall werden also die Likelihood-Wahrscheinlichkeiten direkt von der Quantenmechanik vorgegeben.

Der Bayessche Ansatz für die Aposteriori-Wahrscheinlichkeit $p[v|D]$ analog zu (1.3) und (1.7) lässt sich dann mittels

$$p[v|D] = \frac{p[D|v] p[v]}{p[D]} \sim e^{-\gamma\Gamma[v]} \prod_{i=1}^N < x_i | \frac{e^{-\beta H}}{Z} | x_i > = e^{-F[v]} \quad (5.5)$$

mit dem Energiefunktional

$$F[v] = - \sum_{i=1}^N \ln < x_i | e^{-\beta H} | x_i > + N \ln Z + \gamma\Gamma[v] \quad (5.6)$$

ausdrücken. Dabei wurde die Apriori-Verteilung $p[v]$ analog zu (1.9) über eine Energiefunktion $\Gamma[v]$ mit Hyperparameter γ eingeführt. Eine Spezifizierung von Γ erfolgt in späteren Abschnitten.

5.1 Klassische Näherung

Im Falle großer Massen ($m \rightarrow \infty$) wird erwartet, daß sich die quantenmechanische Rechnung den klassischen Ergebnissen der statistischen Mechanik annähert. So folgt für die Matrixelemente aus (5.4) für große Massen mit

$$< x | e^{-\beta \left(\frac{\hat{x}^2}{2m} + v(\hat{x}) \right)} | x > \rightarrow < x | e^{-\beta v(\hat{x})} | x > = e^{-\beta v(x)} \quad \text{und} \quad Z = \int dx e^{-\beta v(x)} \quad (5.7)$$

das gleiche Ergebnis wie in (4.42). Übernimmt man die nach (5.7) berechnete Näherung, so gilt für die Likelihood-Verteilung

$$p[D|v] = \prod_{i=1}^N p[x_i | \hat{x}, \rho] = \frac{1}{Z^N} e^{-\beta \sum_{i=1}^N v(x_i)}. \quad (5.8)$$

Damit erhält man für die Aposteriori-Wahrscheinlichkeit

$$p[v|D] \sim e^{-\gamma\Gamma[v] - \beta \sum_{i=1}^N v(x_i) - N \ln Z(v)}. \quad (5.9)$$

Eine Maximum-Aposteriori-Approximation entspricht der Suche nach dem maximalen Funktionalwert von $p[v|D]$. Dies führt auf die Stationaritätsgleichung

$$\frac{\delta\Gamma}{\delta v(x)} = -\frac{\beta}{\gamma} \sum_{i=1}^N \delta(x - x_i) + \frac{\beta N}{\gamma Z} e^{-\beta v(x)}. \quad (5.10)$$

Wählt man beispielsweise eine Bestrafungsfunktion [3]

$$\Gamma[v] = \frac{1}{2} \int dx \left(e^{\frac{\alpha^2}{2} \nabla_x} v(x) \right)^2 = \frac{1}{2} \int dx v(x) e^{-\frac{\alpha^2}{2} \Delta_x} v(x), \quad (5.11)$$

so bekommt (5.10) wegen $\frac{\delta\Gamma}{\delta v} = e^{-\frac{\alpha^2}{2} \Delta_x} v$ den Charakter einer Differentialgleichung. Γ ist eine typische Regularisierungsfunktion, die die Glattheit der Funktion v und ihrer Ableitungen fordert. Die

Verwendung von (5.11) ist vor allem deswegen bequem, weil die Greenschen–Funktionen $G_\alpha(x, x')$ von $e^{-\frac{\alpha^2}{2}\Delta_x}$ unter den Randbedingungen $G_\alpha(x - x') \rightarrow 0$ für $|x - x'| \rightarrow \infty$ Gauß–Funktionen sind:

$$e^{-\frac{\alpha^2}{2}\Delta_x} G_\alpha(x, x') = \delta(x - x') \quad \text{mit} \quad G_\alpha(x, x') = \frac{1}{\sqrt{2\pi\alpha}} e^{-\frac{1}{2\alpha^2}(x-x')^2}. \quad (5.12)$$

Somit kann die Gleichung (5.10) formal nach v aufgelöst werden. Erinnert man sich daran, daß die in (5.8) gemachte klassische Näherung nur im Falle hinreichend kleiner β gültig ist, so kann der zweite Summand aus (5.10) vernachlässigt werden. Denn berücksichtigt man nur die in β linearen Terme, so erhält man $\frac{\beta}{Z} e^{-\beta v(x)} \rightarrow \frac{\beta}{\int dx}$. Damit bekommt man als Endergebnis mit

$$v(x) = -\frac{\beta}{\sqrt{2\pi\alpha\gamma}} \sum_{i=1}^N e^{-\frac{1}{2\alpha^2}(x-x_i)^2} \quad (5.13)$$

eine Faltung von negativen Gauß–Funktionen und Datenverteilung. Das Potential wird umso tiefer desto kleiner die Temperatur beziehungsweise desto kleiner die Glattheitsforderung mittels γ ist. Die Varianz wird alleine durch α festgelegt.

5.2 Spektraler Zugang

Es gibt mehrere Methoden, die Matrixelemente $\langle x_i | e^{-\beta H} | x_i \rangle$ analytisch zu bearbeiten. Die eine besteht in der Möglichkeit, die Übergangsmatrixelemente über ein Pfadintegral auszudrücken. Dieser Weg wird im Kapitel 6 gegangen. Eine andere Möglichkeit ist das Arbeiten mit den Energieeigenfunktionen. Drückt man die Matrixelemente des statistischen Operators ρ in seiner Spektraldarstellung (5.3) aus

$$\langle x_i | e^{-\beta H} | x_i \rangle = \sum_n e^{-\beta E_n} \frac{\langle x_i | \varphi_n \rangle \langle \varphi_n | x_i \rangle}{\langle \varphi_n | \varphi_n \rangle}, \quad Z = \sum_n e^{-\beta E_n}, \quad (5.14)$$

so ist man bei der weiteren Rechnung auf den Umgang mit E_n und φ_n angewiesen. Die Maximum–Aposteriori–Approximation für den Ausdruck (5.5) führt dann auf:

$$\begin{aligned} 0 &= \frac{\delta F[v]}{\delta v(x'')} = \gamma \frac{\delta \Gamma[v]}{\delta v(x'')} + \frac{N}{Z} \frac{\delta Z}{\delta v(x'')} \\ &\quad - \sum_{i=1}^N \frac{1}{\langle x_i | e^{-\beta H} | x_i \rangle} \sum_n e^{-\beta E_n} \frac{\langle x_i | \varphi_n \rangle \langle \varphi_n | x_i \rangle}{\langle \varphi_n | \varphi_n \rangle} \times \\ &\quad \left[-\beta \frac{\delta E_n}{\delta v(x'')} + \frac{\langle x_i | \frac{\delta \varphi_n}{\delta v(x'')} \rangle}{\langle x_i | \varphi_n \rangle} + \frac{\langle \frac{\delta \varphi_n}{\delta v(x'')} | x_i \rangle}{\langle \varphi_n | x_i \rangle} - \frac{\langle \varphi_n | \frac{\delta \varphi_n}{\delta v(x'')} \rangle}{\langle \varphi_n | \varphi_n \rangle} - \frac{\langle \frac{\delta \varphi_n}{\delta v(x'')} | \varphi_n \rangle}{\langle \varphi_n | \varphi_n \rangle} \right]. \end{aligned} \quad (5.15)$$

Es verbleibt also die Bestimmung der Variationsableitungen $\frac{\delta E_n}{\delta v(x'')}$, $\langle \frac{\delta \varphi_n}{\delta v(x'')} | \varphi_n \rangle$ und $\frac{\delta Z}{\delta v(x'')}$. Die Ableitungen können über die Eigenwertgleichung (5.2) gewonnen werden. So folgt mit

$$\frac{\delta H}{\delta v(x'')} | \varphi_n \rangle + H | \frac{\delta \varphi_n}{\delta v(x'')} \rangle = \frac{\delta E_n}{\delta v(x'')} | \varphi_n \rangle + E_n | \frac{\delta \varphi_n}{\delta v(x'')} \rangle \quad (5.16)$$

und Multiplikation mit $\langle \varphi_n |$ von links die Variation der Eigenwerte:

$$\frac{\delta E_n}{\delta v(x'')} = \frac{\langle \varphi_n | x'' \rangle \langle x'' | \varphi_n \rangle}{\langle \varphi_n | \varphi_n \rangle}. \quad (5.17)$$

Mit (5.1) kann für die Variation des Hamilton–Operators eine Operatoridentität errechnet werden: Man erhält mit

$$\langle x | \frac{\delta H}{\delta v(x'')} | x' \rangle = \frac{\delta H(x, x')}{\delta v(x'')} = \delta(x - x')\delta(x' - x'') \quad , \forall x, x' \quad , \quad \frac{\delta H}{\delta v(x'')} = | x'' \rangle \langle x'' | \quad (5.18)$$

den Projektor auf die Ortseigenfunktionen. Mit (5.17) wird auch die Ableitung der Zustandssumme ausführbar. Es ergibt sich mit (5.4):

$$\frac{1}{Z} \frac{\delta Z}{\delta v(x'')} = -\beta \sum_n \frac{e^{-\beta E_n}}{Z} \frac{\delta E_n}{\delta v(x'')} = -\beta \langle x'' | \frac{e^{-\beta H}}{Z} | x'' \rangle = -\beta p[x'' | \hat{x}, \rho]. \quad (5.19)$$

Zur Bestimmung von $\frac{\delta \varphi_n}{\delta v}$ kann die Gleichung (5.16) umgestellt werden:

$$(H - E_n) | \frac{\delta \varphi_n}{\delta v(x'')} \rangle = \left(\frac{\langle \varphi_n | x'' \rangle \langle x'' | \varphi_n \rangle}{\langle \varphi_n | \varphi_n \rangle} - | x'' \rangle \langle x'' | \right) | \varphi_n \rangle . \quad (5.20)$$

Der Operator $H - E_n$ ist offensichtlich nicht invertierbar. Wie durch Arbeiten mit dem pseudoinversen Operator die Auflösung nach $\frac{\delta \varphi_n}{\delta v}$ trotzdem gelingt, wird im Anhang C vorgeführt. Als Ergebnis erhält man:

$$| \frac{\delta \varphi_n}{\delta v(x'')} \rangle = \frac{\langle \varphi_n | \lambda_n \rangle}{\langle \varphi_n | \varphi_n \rangle} | \varphi_n \rangle - \sum_{j \neq n} \frac{1}{E_j - E_n} | \varphi_j \rangle \frac{\langle \varphi_j | x'' \rangle \langle x'' | \varphi_n \rangle}{\langle \varphi_j | \varphi_j \rangle} \quad (5.21)$$

und die entsprechend adjungierte Gleichung für $| \frac{\delta \varphi_n}{\delta v(x'')} \rangle$. Dabei ist $| \lambda_n \rangle$ ein beliebiger Zustand, der die φ_n –Komponente festlegt. Man beachte die Analogie von (5.21) zu den Ergebnissen der zeitunabhängigen Störungstheorie. Es ist zu erwarten, daß hauptsächlich φ_j –Komponenten zu $\frac{\delta \varphi_n}{\delta v}$ betragen, die energetisch nahe bei E_n liegen.

Setzt man die Ergebnisse (5.17), (5.19) und (5.21) in (5.15) ein und nutzt die Orthogonalität der Zustände $\langle \varphi_j | \varphi_n \rangle = \delta_{nj} \langle \varphi_n | \varphi_n \rangle$ aus, so fällt die Abhängigkeit von $| \lambda_n \rangle$ heraus, und man erhält:

$$\begin{aligned} 0 &= \frac{\delta F[v]}{\delta v(x'')} = \gamma \frac{\Gamma[v]}{\delta v(x'')} - N\beta \langle x'' | \frac{e^{-\beta H}}{Z} | x'' \rangle \\ &\quad - \sum_{i=1}^N \frac{1}{\langle x_i | e^{-\beta H} | x_i \rangle} \sum_n e^{-\beta E_n} \left\{ -\beta \frac{\langle x_i | \varphi_n \rangle \langle \varphi_n | x_i \rangle \langle x'' | \varphi_n \rangle \langle \varphi_n | x'' \rangle}{\langle \varphi_n | \varphi_n \rangle^2} \right. \\ &\quad \left. - \frac{2}{\langle \varphi_n | \varphi_n \rangle} \sum_{j \neq n} \frac{1}{E_j - E_n} \frac{\text{Re} \left[\langle x_i | \varphi_j \rangle \langle \varphi_j | x'' \rangle \langle x'' | \varphi_n \rangle \langle \varphi_n | x_i \rangle \right]}{\langle \varphi_j | \varphi_j \rangle} \right\}. \end{aligned} \quad (5.22)$$

Die Stationaritätsgleichung (5.22) ist invariant unter Umnormierung von φ_n , was auf die Invarianz von (5.2) und (5.3) bezüglich Umnormierung zurückzuführen ist.

Es sei noch angemerkt, daß die Gültigkeit von (5.22) im Falle von anderen Meßobservablen erhalten bleibt. Betrachtet man beispielsweise Impulsmessungen \tilde{p}_i , so gilt analog für die Meßwahrscheinlichkeiten

$$p[\tilde{p}_i | \hat{p}, \rho] = \langle \tilde{p}_i | \rho | \tilde{p}_i \rangle = \sum_n \frac{e^{-\beta E_n}}{Z} \frac{\langle \tilde{p}_i | \varphi_n \rangle \langle \varphi_n | \tilde{p}_i \rangle}{\langle \varphi_n | \varphi_n \rangle}. \quad (5.23)$$

Das heißt, der Formalismus für die Variationsableitungen nach v bleibt unverändert. Effektiv müssen in der Gleichung (5.22) nur die Ortseigenzustände $| x_i \rangle$ durch die jeweiligen gemessenen Eigenzustände $| \tilde{p}_i \rangle$ ersetzt werden. Der weitere Verlauf dieses Kapitels beschäftigt sich mit der Diskussion der Gleichung (5.22). Es werden numerische Anwendungen und Näherungen verfolgt.

5.3 Hoch– und Tieftemperaturgrenzfälle

Im Tieftemperaturgrenzfall $\beta \rightarrow \infty$ konvergiert der statistische Operator gegen den Projektator auf den Grundzustand φ_0 von H . Dieser beschreibt also einen reinen Zustand:

$$\rho = \frac{e^{-\beta H}}{Z} \rightarrow \frac{|\varphi_0\rangle\langle\varphi_0|}{\langle\varphi_0|\varphi_0\rangle} \quad \text{für } \beta \rightarrow \infty. \quad (5.24)$$

In der Stationaritätsgleichung (5.22) fällt der zweite Summand mit dem ersten Term des dritten Summanden weg, und man erhält

$$0 = \lim_{\beta \rightarrow \infty} \frac{\delta F[v]}{\delta v(x'')} = \gamma \frac{\delta \Gamma[v]}{\delta v(x'')} + 2 \sum_{i=1}^N \sum_{j \neq 0} \frac{1}{E_j - E_0} \frac{1}{\langle \varphi_j | \varphi_j \rangle} \text{Re} \left[\frac{\langle x_i | \varphi_j \rangle \langle \varphi_j | x'' \rangle \langle x'' | \varphi_0 \rangle}{\langle x_i | \varphi_0 \rangle} \right]. \quad (5.25)$$

Man beachte, daß trotz Tieftemperaturnäherung sämtliche Eigenfunktionen zur Stationarität beitragen – und zwar mit einem Summanden proportional $\frac{1}{E_j - E_0}$. Das heißt, die Eigenzustände in unmittelbarer Nähe zum Grundzustand haben den größten Einfluß. Der numerische Aufwand zum Lösen von (5.25) ist im Vergleich zu (5.22) eindeutig reduziert worden, da eine der Summen über den vollen Hilbert–Raum verschwindet.

Im Falle von sehr hohen Temperaturen konvergiert der statistische Operator gegen

$$\rho \rightarrow \frac{1}{\dim \mathcal{H}} \mathcal{I} \quad \text{für } \beta \rightarrow 0, \quad (5.26)$$

wobei $\dim \mathcal{H}$ für die Dimension des Hilbert–Raums steht. Das Ensemble besteht aus einem vollständigen, gleichverteilten Gemisch. Für die Meßwahrscheinlichkeiten folgt dann sofort: $p[x_i | \hat{x}, \rho] = \text{const.}$

Der statistische Operator wird also von H unabhängig. Diese Tatsache muß sich auch in den obigen Ergebnissen widerspiegeln. Führt man einen Hochtemperaturlimes $\beta \rightarrow 0$ in nullter Entwicklungsordnung in β bei Gleichung (5.22) aus, so bekommt man

$$0 = \lim_{\beta \rightarrow 0} \frac{\delta F[v]}{\delta v(x'')} = \gamma \frac{\delta \Gamma[v]}{\delta v(x'')} + 2 \sum_{i=1}^N \sum_{\substack{n,j \\ n \neq j}} \frac{1}{E_j - E_n} \frac{\text{Re} \left[\langle x_i | \varphi_j \rangle \langle \varphi_j | x'' \rangle \langle x'' | \varphi_n \rangle \langle \varphi_n | x_i \rangle \right]}{\langle \varphi_j | \varphi_j \rangle \langle \varphi_n | \varphi_n \rangle}. \quad (5.27)$$

Mit der Definition

$$a_{nj} \equiv \frac{1}{E_j - E_n} \frac{\text{Re} \left[\langle x_i | \varphi_j \rangle \langle \varphi_j | x'' \rangle \langle x'' | \varphi_n \rangle \langle \varphi_n | x_i \rangle \right]}{\langle \varphi_j | \varphi_j \rangle \langle \varphi_n | \varphi_n \rangle} \quad (5.28)$$

folgt $a_{nj} = -a_{jn}$. Also kann für den letzten Summanden aus (5.27) gezeigt werden, daß er verschwindet:

$$\sum_{\substack{n,j \\ n \neq j}} a_{nj} = \frac{1}{2} \sum_{\substack{n,j \\ n \neq j}} [a_{nj} + a_{jn}] = \frac{1}{2} \sum_{\substack{n,j \\ n \neq j}} [a_{nj} - a_{nj}] = 0. \quad (5.29)$$

Somit trägt erwartungsgemäß nur das Vorwissen $p[v]$ durch Γ zur Bestimmung der Maximum–Aposteriori–Lösung bei.

5.4 Numerische Implementierung

Bei der numerischen Suche nach Lösungen der Stationaritätsgleichung (5.22) ist man auf die Diskretisierung des Hilbert–Raums angewiesen. Es bietet sich dabei das Arbeiten in der Ortsdarstellung an. Erzeugt man einen n –dimensionalen Hilbert–Raum, indem man die Ortsachse durch n äquidistante Stützstellen x_j mit der Gitterkonstanten Δx codiert, so führt dies auf das Arbeiten mit Operatoren als $n \times n$ –Matrizen und Zuständen als n –dimensionale Vektoren. Die diskrete Version des Hamilton–Operators aus (5.1) hat dann in einer Dimension die Form

$$H(x, x') \Rightarrow -\frac{\hbar^2}{2m(\Delta x)^2} \begin{pmatrix} -2 & 1 & 0 & \cdots & 0 & 1 \\ 1 & -2 & 1 & \ddots & & 0 \\ 0 & 1 & -2 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & & 0 \\ 0 & & \ddots & \ddots & 1 & \\ 1 & 0 & \cdots & 0 & 1 & -2 \end{pmatrix} + \begin{pmatrix} v_1 & 0 & \cdots & \cdots & 0 \\ 0 & v_2 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & & \vdots \\ & & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & v_n \end{pmatrix}, \quad (5.30)$$

wobei das Potential an den Stützstellen mit $v(x_j) = v_j$ abgekürzt wurde und der Operator der zweiten Ableitung mit zyklischen Randbedingungen verwendet wurde. Im Gegensatz zu der Wahl von

$$\Delta_x \Rightarrow \frac{1}{(\Delta x)^2} \begin{pmatrix} -2 & 1 & 0 & \cdots & 0 \\ 1 & -2 & 1 & \ddots & \vdots \\ 0 & 1 & -2 & \ddots & \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & -2 \end{pmatrix} \quad (5.31)$$

werden in (5.30) die Wellenfunktionen auf den Rändern nicht notwendigerweise auf Null fixiert. Eine solche Fixierung würde einem unendlich hohen Potential an den Rändern entsprechen. Eine Verwendung von (5.31) hat bei der Wirkung im Potentialraum – beispielsweise durch die Glattheitsforderung $\Gamma[v] = -\frac{1}{2} \sum_{i,j=1}^n v_i \Delta_{ij} v_j$ – den Vorteil, daß das Potential auf den Rändern automatisch auf Null gesetzt wird. Dadurch wird die Invarianz der Lösung von (5.22) gegenüber Verschiebung des Potentials längs der Energieachse aufgehoben. Wird als Vorwissen der Erwartungswert für eine Energiemessung $\langle H \rangle$ für das Potential vorgegeben, so hat dies bei (5.31) einen direkten Einfluß auf die Potentialform – bei der Anwendung von Δ_x mit zyklischen Randbedingungen resultiert nur eine Verschiebung in Energierichtung.

Das numerische Lösen der Gleichung (5.22) analog zu den in Kapitel 1.6 vorgestellten Algorithmen führt auf Iterationsgleichungen der vektoriellen Form

$$v^{(j+1)} = v^{(j)} - \eta^{(j)} \mathcal{O}^{(j)} \frac{\delta F[v^{(j)}]}{\delta v} \quad \text{mit} \quad F[v^{(j)}] = \gamma \Gamma[v^{(j)}] - \sum_{i=1}^N \ln \langle x_i | \frac{e^{-\beta H(v^{(j)})}}{Z(v^{(j)})} | x_i \rangle, \quad (5.32)$$

wobei v ein n –dimensionaler Vektor ist und j den Iterationsschritt angibt. In jedem Iterationsschritt besteht also die Hauptarbeit darin, den jeweiligen Gradienten des Funktionals zu bestimmen, indem die Eigenfunktionen und Eigenwerte des Hamilton–Operators $H(v^{(j)})$ berechnet werden. Es ist eine Schrittweitenoptimierung bei der Wahl von $\eta^{(j)}$ zu empfehlen, da die Iteration sehr empfindlich bei zu großer Schrittweite reagiert. Falls kein Referenzpotential vorhanden ist, bietet sich als Startpotential für die Iteration eine geeignet skalierte Funktion an, die die empirische Datenverteilung

mit negativem Vorzeichen wiedergibt. Als Iterationsmatrix $\mathcal{O}^{(j)}$ hat sich die Einheitsmatrix für das Gradientenabstiegsverfahren als äußerst ungünstig erwiesen. Bei nicht zu kleiner Gewichtung des Vorwissens durch γ zeigt die Verwendung von $\mathcal{O} = (\gamma \frac{\delta^2 \Gamma}{\delta v \delta v'})^{-1}$ als genäherte inverse Hesse-Matrix eine passable Konvergenz. Im Anhang D ist die Berechnung der Hesse-Matrix des Datenanteils im Tieftemperaturfall vorgeführt. Deren zeitlich aufwendige Bestimmung lohnt sich allerdings nur bei sehr keiner Dimension des betrachteten Hilbert-Raums, verbessert aber die Konvergenz der Iteration wesentlich. In den numerischen Anwendungen dieser Arbeit werden Gitterkonstanten $\Delta x = 1$ und Gittergrößen bis $n = 36$ verwendet. Im Tieftemperaturfall sind Gittergrößen von $n = 100$ handhabbar.

5.5 Beispiele für Apriori-Wissen

Vorwissen ist für die Anwendung essentiell. Bei der Abwesenheit von Vorwissen ($\Gamma \equiv 0$) sind pathologische Lösungen zu erwarten. Das Potential als Lösung von (5.22) tendiert dann dazu, an den Datenpunkten δ -förmige Peaks zu bekommen. Als Beispiel sei hier der Tieftemperaturfall aus (5.25) für einen einzelnen Datenpunkt $D = \{x_0\}$ diskutiert.

Für einen Hamilton-Operator mit einem δ -förmigen Potential $v(x) = -a\delta(x - x_0)$ errechnen sich die Eigenfunktionen über:

$$\left[-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} - a\delta(x - x_0) \right] \varphi_0(x) = E_0 \varphi_0(x) \quad a > 0. \quad (5.33)$$

Als normierbare Lösung mit negativem Eigenwert ergibt sich dann

$$\varphi_0(x) = \sqrt{\mu} \left[e^{\mu(x-x_0)} \theta(x_0 - x) + e^{-\mu(x-x_0)} \theta(x - x_0) \right] \quad \text{mit} \quad E_0 = -\frac{\hbar^2 \mu^2}{2m}. \quad (5.34)$$

Durch einmalige Integration $\int_{x_0-\epsilon}^{x_0+\epsilon} dx$ in (5.33) mit infinitesimalem $\epsilon > 0$ kann unter Ausnutzung der Stetigkeitsbedingung für φ_0 eine Sprungbedingung

$$\frac{\partial \varphi_0(x_0 + \epsilon)}{\partial x} - \frac{\partial \varphi_0(x_0 - \epsilon)}{\partial x} = -\frac{2ma\varphi_0(x_0)}{\hbar^2} \quad \rightarrow \mu = \frac{am}{\hbar^2} \quad (5.35)$$

ausgerechnet werden, welche die Quantisierung der Eigenwerte festlegt. Somit existiert nur ein Zustand mit negativer Energie. Dieser Grundzustand hat die Energie $E_0 = -\frac{a^2 m}{2\hbar^2}$ und ist nicht entartet. Im Grenzfall $a \rightarrow \infty$ fällt $|\varphi_0|$ aufgrund von (5.34) und (5.35) mit dem Ortseigenzustand $|x_0\rangle$ zusammen:

$$\varphi_0(x) = \langle x | \varphi_0 \rangle \Rightarrow \delta(x - x_0) = \langle x | x_0 \rangle \quad \text{und} \quad E_0 \Rightarrow -\infty. \quad (5.36)$$

Alle anderen Eigenzustände haben eine positive Energie $E_j > 0$ für $j \neq 0$ (Streuzustände) und die Orthogonalitätsbedingung fordert

$$0 = \langle \varphi_0 | \varphi_j \rangle = \langle x_0 | \varphi_j \rangle = \int dx \varphi_0^*(x) \varphi_j(x) = \varphi_j(x_0) \quad \forall j \neq 0. \quad (5.37)$$

Das konstruierte Potential $v(x) = \lim_{a \rightarrow \infty} (-a\delta(x - x_0))$ erfüllt im Fall von einem Datenpunkt bei x_0 die Stationaritätsgleichung (5.25), da die Überlappe $\langle x_0 | \varphi_j \rangle$ gegen Null und gleichzeitig $E_j - E_0$ gegen Unendlich tendieren. Somit ist bei $\Gamma \equiv 0$ das Potential v eine Lösung der Maximum-Aposteriori-Approximation. In der Abbildung 5.1 werden diese Überlegungen durch ein numerisches Beispiel bestätigt. Es wurde auf einem Gitter der Größe $n = 100$ ein einzelner Datenpunkt bei $x_0 = 50$ angenommen und für den Tieftemperaturfall nach Lösungen von (5.25)

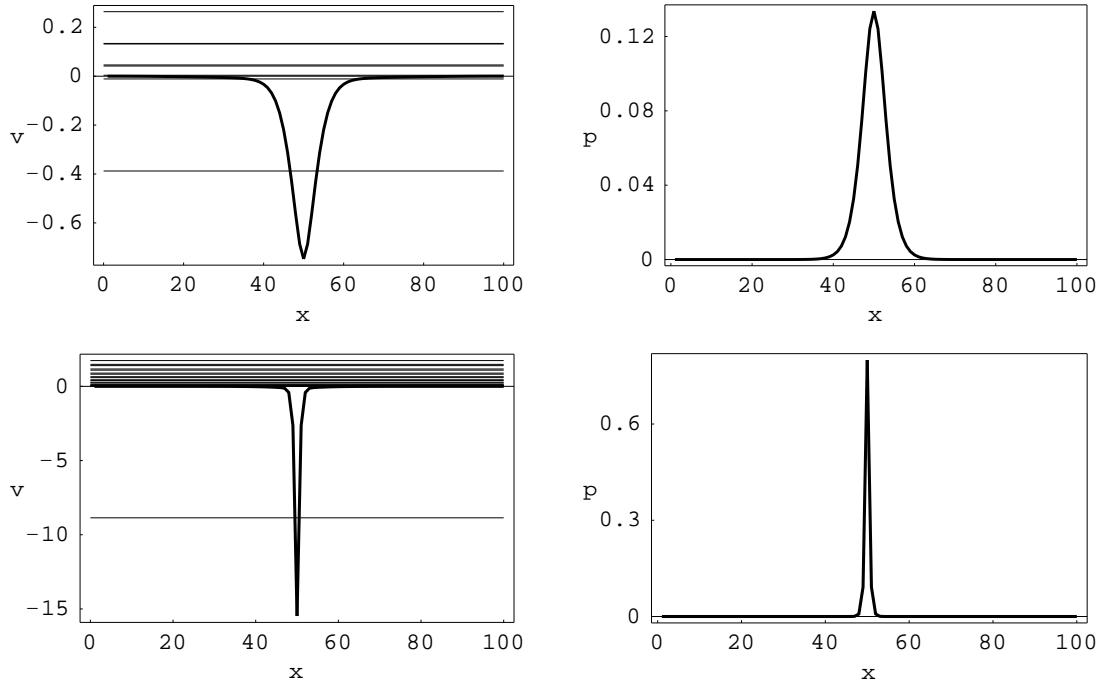


Abbildung 5.1: Abstellen des Vorwissens. Oben sind Potential und Likelihood für den Fall $\gamma = 3$ dargestellt. Bei den unteren Diagrammen wurde γ auf 0.001 reduziert. Die Masse des betrachteten Teilchens beträgt $m = 0.1$, ($\hbar = 1$). Zusätzlich sind die tiefsten Eigenenergien der Potentiale eingezeichnet. Man beachte die unterschiedliche Skalierung der Achsen.

gesucht. Ein abnehmender Einfluß von Vorwissen der Form (5.40) wurde durch Verkleinerung von γ simuliert (siehe unten). Es zeigt sich, daß durch Abstellen des Vorwissens die Lösung für das Potential im Datenpunkt und die Grundzustandsenergie gegen minus Unendlich konvergieren. Die dazugehörige Wahrscheinlichkeitsdichte für den Grundzustand nimmt die Form einer δ -Funktion an. Der Grundzustand ist – im Gegensatz zu dem Fall einer größeren Gewichtung von Γ – der einzige gebundene Zustand.

Das Beispiel zeigt, daß ohne Vorwissen die Flexibilität des unendlichdimensionalen Feldes v zu groß ist. Die endlichdimensionale Datenmenge kann mittels δ -förmiger Lösungen adaptiert werden. Die Möglichkeit zu generalisieren ist also nur bei der Anwesenheit von Vorwissen vorhanden (vgl. Argumentation um (1.5) und (1.6)). Die Aufgabe des Vorwissens besteht darin, durch seinen Einfluß in dem Energiefunktional nichtlokale Information (im Gegensatz zu den Daten) einzubringen und somit pathologische Lösungen auszuschließen.

Falls es möglich ist, wird man ein in v quadratisches Energiefunktional wählen, was einer unendlichdimensionalen Gauß-Verteilung als Apriori-Verteilung $p[v]$ entspricht:

$$p[v] = \left(\det \left(\frac{\gamma}{2\pi} \mathcal{O} \right) \right)^{\frac{1}{2}} e^{-\frac{\gamma}{2} \int dx dx' (v(x) - v_0(x)) \mathcal{O}(x, x') (v(x') - v_0(x'))} \sim e^{-\gamma \Gamma[v]} \quad (5.38)$$

mit symmetrischer, positiv-definiter Matrix $\mathcal{O}(x, x')$, die eine Metrik auf dem Funktionenraum von v definiert, und einem Referenzfeld v_0 , welches als hochdimensionaler Hyperparameter aufgefaßt werden kann – also $p[v] = p[v|v_0]$. Für die Variationsableitung des Energiefunktional Γ folgt eine

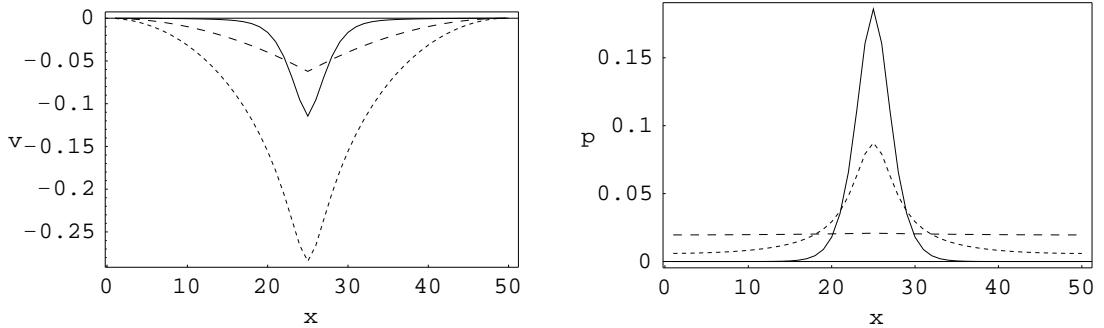


Abbildung 5.2: Potential für einen Datenpunkt bei unterschiedlichen Temperaturen. Links sind die Potentiale, rechts die dazugehörigen Likelihood-Dichten für die Fälle $\beta = 1000$ (durchgezogene Linie), $\beta = 10$ (kleingestrichelte Linie) und $\beta = 1$ (großgestrichelte Linie) gezeichnet. Die Erwartungswerte für eine Energiemessung liegen für kleinere Temperaturen mit $\langle H \rangle_{\beta=1000} = -0.058$ und $\langle H \rangle_{\beta=10} = -0.11$ im Bereich gebundener Zustände, für große Temperaturen im Streubereich: $\langle H \rangle_{\beta=1} = 0.48$. Die übrigen Parameter sind: $\gamma = 100$, $m = 1.25$. Es wurde $\hbar = 1$ angenommen.

lineare Abhängigkeit in v :

$$\frac{\delta \Gamma[v]}{\delta v(x'')} = \int dx' \mathcal{O}(x'', x') (v(x') - v_0(x')) , \quad (5.39)$$

und die Verteilung $p[v]$ wird für $v = v_0$ maximiert.

Oft besteht das Vorwissen aus der Erwartung, daß das Potential glatt ist. Dieses entspricht in (5.38) einer Referenzfunktion $v_0 \equiv 0$ und $\mathcal{O}(x, x') = -\delta(x - x') \frac{d^2}{dx^2}$. Dann gilt mittels partieller Integration unter der Annahme zyklischer Randbedingungen:

$$\Gamma[v] = \frac{1}{2} \int dx \left(\frac{dv(x)}{dx} \right)^2 . \quad (5.40)$$

Für diese Art von Vorwissen wird eine Anwendung in Abbildung 5.2 vorgestellt. Es wurde auf einem Gitter der Größe $n = 50$ für einen einzelnen Datenpunkt bei $x_0 = 25$ der Einfluß der Temperatur des kanonischen Ensembles auf die Lösung von (5.22) untersucht. Hierbei wurde die Bestrafung mittels Γ konstant gehalten. Es zeigt sich, daß sich erwartungsgemäß für steigende Temperaturen die Varianzen der um x_0 konzentrierten Likelihood-Dichten vergrößern, bis schließlich die Gleichverteilung erreicht wird. Komplizierter ist die Situation bei den Lösungen für die Potentiale. Von der Lösung für den Tieftemperaturfall ausgehend vergrößert sich zuerst die Tiefe der Potentialmulde mit steigender Temperatur. Von einer gewissen Grenztemperatur ab kehrt sich dieses Phänomen um: Die Potentialtiefe nimmt wieder ab, und die Potentiale streben gegen die Nullfunktion. Die anfängliche Vertiefung kann durch den wachsenden Einfluß der ersten angeregten Zustände zum Datenanteil des Funktionalen (5.6) erklärt werden. Durch die Vertiefung sind die energetisch untersten Zustände stärker bei x_0 lokalisiert. Dadurch wird der Likelihood-Anteil in F erhöht, und die Verschlechterung des Prior-Anteils zum Energiefunktional wird durch diesen Effekt mehr als kompensiert. Für sehr große Temperaturen wird der statistische Operator unabhängig vom Potential und beschreibt eine Gleichverteilung. Das heißt, nur der Prior-Term trägt effektiv zur Lösung der Stationaritätsgleichung bei. Wegen $\Gamma[v \equiv 0] = 0$ tendiert die Lösung für das Potential gegen die Nullfunktion.

Ein anderer Spezialfall von (5.38) ist:

$$\mathcal{O}(x, x') = \delta(x - x') \quad \text{und} \quad \Gamma[v] = \frac{1}{2} \int dx \left(v(x) - v_0(x) \right)^2. \quad (5.41)$$

Hier mißt Γ den quadratischen Abstand der beiden Potentiale.

Eine Verschiebung des Potentials um eine Konstante v_c hat nach (5.2) eine Verschiebung der Energieeigenwerte nach $E_n + v_c$ zur Folge, wobei die Eigenfunktionen $|\varphi_n\rangle$ invariant gelassen werden. Dies hat keine Auswirkungen auf die Lösung von (5.22), die das gesuchte Potential nur bis auf eine additive Konstante bestimmt. Möchte man trotzdem die energetische Lage des rekonstruierten Potentials eindeutig fixieren, so bietet sich eine Messung der mittleren Energie U für das kanonische Ensemble an. Danach könnte v_c so gewählt werden, daß U mit dem theoretischen Erwartungswert

$$\langle H \rangle = \text{Tr}[\rho H] = \sum_n \frac{e^{-\beta E_n}}{Z} (E_n + v_c) \quad (5.42)$$

zusammenfällt.

Eine andere Art von Vorwissen, ist die Kenntnis von Randbedingungen für v . Oft ist beispielsweise bekannt, daß das Potential im Unendlichen verschwindet: $\lim_{x \rightarrow \pm\infty} v(x) = 0$. Unter dieser Annahme ist das Erzwingen von $U \approx \langle H \rangle$ durch einen Term der Form

$$p[v] \sim e^{-\frac{\gamma}{2}(U - \langle H \rangle)^2} \quad \text{oder} \quad p[v] \sim \delta(U - \langle H \rangle) \quad (5.43)$$

eine echte, zusätzliche Bedingung, die im wesentlichen die Tiefe des Potentials festlegt.

Eine logische AND-Verknüpfung einzelner Prior-Terme führt nach (1.12) zu einer Summe der Energiefunktionale.

5.6 Rekonstruktion von Potentialen

Dieser Abschnitt befaßt sich mit der Rekonstruktion von Potentialen, indem Ortsmessungen für das unbekannte Potential vorgenommen werden. Es zeigt sich, daß der Erfolg bei der Rekonstruktion entscheidend von der Qualität des Vorwissens abhängt. Dies liegt vor allem an der Tatsache, daß es sehr unterschiedliche Potentiale mit sehr ähnlichen Likelihood-Dichten gibt beziehungsweise daß möglicherweise kleine Variationen des Potentials große Änderungen in der Likelihood-Dichte zur Folge haben.

Im Tieftemperaturfall kommt es zu keiner Mischung der Eigenzustände: Die Likelihood-Dichte entspricht der Wahrscheinlichkeitsdichte des Grundzustands. Falls die Grundzustandswellenfunktion $\varphi_0(x)$ und die zugehörige Grundzustandsenergie E_0 bekannt sind, kann das Potential mittels Auflösung der Eigenwertgleichung (5.2) in Ortsdarstellung nach

$$v(x) = E_0 + \frac{\hbar^2}{2m} \frac{\varphi_0^*(x) \Delta_x \varphi_0(x)}{|\varphi_0(x)|^2} \quad (5.44)$$

direkt analytisch bestimmt werden. Die Energieeigenfunktionen können ohne Beschränkung der Allgemeinheit als reell angenommen werden. Denn dadurch, daß der Hamilton-Operator reell ist, erlaubt die komplexkonjugierte Eigenwertgleichung

$$H\varphi_0(x) = E_0\varphi_0(x) \quad \implies \quad H\varphi_0^*(x) = E_0\varphi_0^*(x) \quad (5.45)$$

die Schlußfolgerung $\varphi_0(x) \equiv \varphi_0^*(x)$, falls E_0 nicht entartet ist. Im Falle von Entartung lassen sich reelle Eigenfunktionen durch geeignete Linearkombinationen erzeugen.

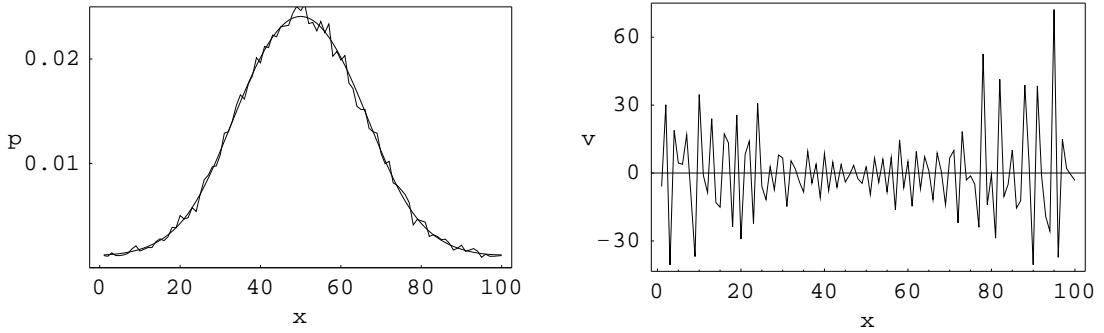


Abbildung 5.3: Rekonstruktion von Potentialen nach (5.44). Das linke Diagramm zeigt die originale Likelihood-Dichte $|\varphi_0(x)|^2$ (glatte Funktion) und die empirische Wahrscheinlichkeitsdichte $h(x)$ (leicht schwankende Funktion) bei einer Datenmenge von $N = 30000$. Rechts ist das mittels $h(x)$ aus (5.44) rekonstruierte Potential dargestellt. Die Masse des Teilchens ist $m = 2.78$, $\hbar = 1$.

Dies legt bei der Rekonstruktion des Potentials ein Vorgehen nahe, bei dem einmal die Grundzustandsenergie gemessen und die Wahrscheinlichkeitsdichte bei sehr häufigen Messungen durch die empirische Datenverteilung $\varphi_0(x) \approx \sqrt{h(x)}$ approximiert wird. Vorzeichenwechsel von $\varphi_0(x)$, welche $h(x)$ nicht registrieren würde, sind nicht zu erwarten, da der Grundzustand betrachtet wird. In Abbildung 5.3 ist dargestellt, daß diese Strategie selbst bei einer sehr großen Datenmenge ($N = 30000$) nicht zum Erfolg führt. Auf einem Gitter der Größe $n = 100$ wurden für ein Woods-Saxon-Potential als angenommenes Originalpotential

$$v_D(x) = -\frac{1}{1 + e^{(|x-50|-25)/3}} \quad (5.46)$$

30000 Daten aus der Verteilung des Grundzustands gewonnen. Obwohl die originale Likelihood-Dichte durch die Häufigkeitsverteilung $h(x)$ gut wiedergegeben wird, versagt die Rekonstruktion von $v(x)$ nach (5.44) mit genähertem φ_0 . Der Grund hierfür liegt darin, daß die Information über das Potential nach (5.44) in der zweiten Ableitung der Wellenfunktion steckt.

Im folgenden wird dargestellt, wie eine Rekonstruktion mittels des in diesem Kapitel vorgestellten Formalismus aussehen kann. In einem ersten Beispiel wird erneut der Fall $T = 0$ betrachtet. Das Vorwissen besteht dabei aus der Erwartung eines glatten Potentials und der ungefähren Kenntnis E_0 – der Grundzustandsenergie des wahren Potentials. Zusammen mit der Bedingung, daß das Potential auf den Rändern verschwindet, müßten somit bei hinreichend großer Datenmenge genügend Informationen vorliegen, um die wesentlichen Merkmale des wahren Potentials zu erkennen. Es ist dabei zu erwarten, daß an Orten, in denen das konstruierte Potential unterhalb von E_0 liegt, die Potentialform am besten erkannt wird, weil dort die meisten Daten auftreten werden. An Orten mit geringer Datendichte wird die Lösung vor allem versuchen, die Kriterien des Vorwissens zu erfüllen.

Die Implementierung des obigen Vorwissens führt für das Energiefunktional aus (5.6) mit der Normierung $\langle \varphi_0 | \varphi_0 \rangle = 1$ auf die Form:

$$F[v] = -\frac{\gamma}{2} \int dx v(x) \Delta_x v(x) - \sum_{i=1}^N \ln \langle x_i | \varphi_0(v) \rangle \langle \varphi_0(v) | x_i \rangle + \frac{\chi}{2} \left(\langle H \rangle - E_0 \right)^2 \quad (5.47)$$

mit Δ_x aus (5.31) und vorgegebenen Hyperparametern γ und χ . Es bestünde auch die Möglichkeit, diese Parameter durch das in Kapitel 1 beschriebene Vorgehen zu lernen. In Abbildung 5.4 ist das

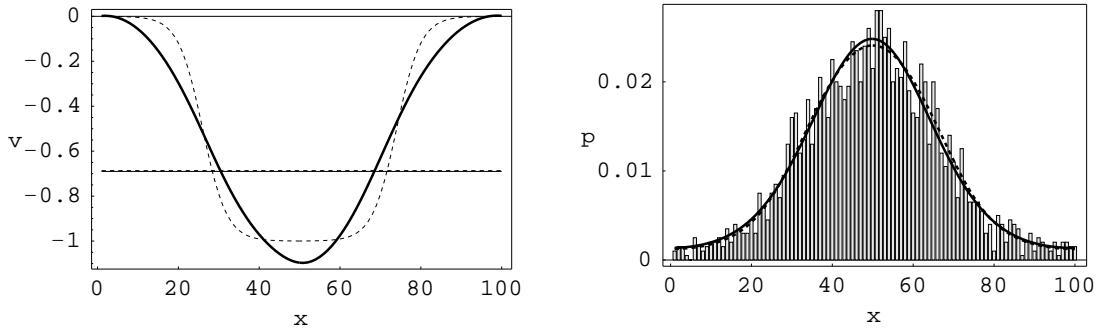


Abbildung 5.4: Rekonstruktion von Potentialen nach (5.47). Links sind Originalpotential (gestrichelt) und rekonstruiertes Potential (durchgezogen) mit den Grundzustandsenergien gezeichnet. Rechts sind die Likelihood-Dichten und die empirische Verteilung der Daten dargestellt. Die übrigen Parameter sind: $\chi = 3000$, $\gamma = 10$, $N = 2000$, $m = 2.78$.

Ergebnis einer Iteration mittels (5.32) mit (5.47) dargestellt: Während die Likelihood-Dichte gut wiedergegeben wird, wird für das Potential nur die grobe Form reproduziert. Dies verdeutlicht die Problematik, Aussagen über das direkt nicht messbare Potential zu machen: Die observable Wahrscheinlichkeitsdichte hängt durch die quantenmechanischen Gesetze über keinen injektiven Zusammenhang von dem gesuchten Potential ab. Bei dem Inversionsproblem, von den Likelihood-Dichten auf die Potentiale zu schließen, ist vor allem dann ein gutes Lernen möglich, wenn das Vorwissen unter der Vielzahl der Potentiale mit ähnlicher Dichteverteilung das richtige herausfiltert. Diese Schlußfolgerungen deuten darauf hin, daß präziseres Vorwissen erstrebenswert ist. Dies könnte beispielsweise die Anwesenheit eines Referenzpotentials sein. Eine weitere Veränderungsmöglichkeit wäre Messungen am kanonischen Ensemble für abgeänderte Temperaturen vorzunehmen. Eine Erhöhung der Temperatur hat eine Erhöhung des Energieerwartungswertes zur Folge. Dadurch besteht die Möglichkeit, an Daten bei energetisch höher liegenden Potentialabschnitten zu gelangen, welche lokale Informationen an diesen Orten mit sich bringen.

Ein zweites Beispiel beschäftigt sich mit der Rekonstruktion eines näherungsweise periodischen Potentials. Dies könnte beispielsweise das Potential eines Festkörpers in einem eindimensionalen Modell sein, bei dem eine Fehlstelle vorliegt. Ziel des Lernens ist, die Stelle der Abweichung eindeutig zu lokalisieren und, wenn möglich, Aussagen über die Art der Störung zu machen. Das unten verwendete Originalpotential hat auf einem Gitter der Größe $n = 30$ die Form

$$v_D(x) = \begin{cases} \sin\left(\frac{2\pi}{6}x\right) & \text{für } x \leq 12, x \geq 25 \\ \sin\left(\frac{2\pi}{12}x\right) & \text{für } 13 \leq x \leq 24 \end{cases} \quad (5.48)$$

Im Unterschied zum obigen Beispiel liegt ein Referenzpotential $v_0(x) = \sin\left(\frac{2\pi x}{6}\right)$ für das ungestörte Potential vor. Es wird der allgemeinere Fall $T \neq 0$ betrachtet. Für verschiedene Fehlerfunktionale, die unterschiedliches Vorwissen beschreiben, können die numerischen Ergebnisse miteinander verglichen werden. Die ersten drei Funktionale versuchen ohne eine Referenzfunktion für die Störstelle auszukommen. Dafür wird eine ungefähre Kenntnis des Energieerwartungswertes $\langle H \rangle$ angenommen. Eine Wahl für das zu minimierende Funktional könnte

$$F[v] = -\frac{\lambda}{2} \int dx (v(x) - v_0) \Delta_x (v(x) - v_0) + \frac{\chi}{2} (\langle H \rangle - U)^2 - \sum_{i=1}^N \ln \langle x_i | \frac{e^{-\beta H(v)}}{Z(v)} | x_i \rangle \quad (5.49)$$

sein. Bei kleinem Hyperparameter λ ist zu erwarten, daß durch die nichtperiodische Likelihood-Dichte die Position der Fehlstelle erkannt wird. (5.49) hat den offensichtlichen Nachteil, daß die Referenz v_0 über das gesamte Gitter Einfluß auf die Potentialgestalt hat. Dieses zeigt auch das Ergebnis der numerischen Rechnung mit der Referenzfunktion als Startpotential ($v^{(0)} = v_0$) in Abbildung 5.5. Während die Potentialform an Orten, in denen das Referenzpotential durch die empirischen Daten unterstützt wird, und bei den Stellen des Störbereichs, in denen das Potential unterhalb von $\langle H \rangle$ liegt, gut rekonstruiert wird, ist sie im Bereich $x \in [14, 20]$, wo wenige Daten auftreten, fehlerhaft. Trotzdem wird die Fehlstelle zufriedenstellend lokalisiert. Man beachte, daß die Likelihood-Dichte es in den Bereichen außerhalb der Fehlstelle schafft, die Individualität der Daten wiederzugeben, ohne daß das Potential an diesen Stellen merklich von der Referenz abweicht. Eine Verbesserung von (5.49) kann durch ein Funktional erreicht werden, bei dem die Wirkung des Referenzpotentials in den Bereichen abgestellt wird, in denen die Daten für eine Abweichung von der Likelihood-Dichte von v_0 sprechen. Ein solches Funktional ist

$$\begin{aligned} F[v] = & \frac{\gamma}{2} \int dx \left(\frac{\partial v(x)}{\partial x} \right)^2 + \frac{\lambda}{2} \int dx L(x) \theta[b - L(x)] + \frac{\chi}{2} \left(\langle H \rangle - U \right)^2 \\ & - \sum_{i=1}^N \ln \langle x_i | \frac{e^{-\beta H(v)}}{Z(v)} | x_i \rangle \end{aligned} \quad (5.50)$$

mit dem euklidischen Abstand

$$L(x) = \left(v(x) - v_0(x) \right)^2. \quad (5.51)$$

Eine Variante von (5.50) ist die Wahl

$$\begin{aligned} F[v] = & \frac{\gamma}{2} \int dx \left(\frac{\partial v(x)}{\partial x} \right)^2 \theta[l(x) - b] + \frac{\lambda}{2} \int dx l(x) \theta[b - l(x)] + \frac{\chi}{2} \left(\langle H \rangle - U \right)^2 \\ & - \sum_{i=1}^N \ln \langle x_i | \frac{e^{-\beta H(v)}}{Z(v)} | x_i \rangle \end{aligned} \quad (5.52)$$

mit dem lokalen, differentiellen Abstand von Potential und Referenz

$$l(x) = \left(\frac{\partial}{\partial x} (v(x) - v_0(x)) \right)^2. \quad (5.53)$$

Während in (5.52) zwischen den beiden Alternativen $v = v_0$ (Muster) und $v = 0$ (Glattheit) umgeschaltet wird, wird in (5.50) nur das Muster v_0 lokal an- oder abgeschaltet. Der Hyperparameter b gibt dabei die Schranke für das Umschalten vor. Entscheidend für die Iteration ist die Wahl des Anfangspotentials. Es besteht die Gefahr, daß bei einem ungünstigen Startpotential die Iteration in einem unerwünschten Nebenminimum des Energiefunktions "hängenbleibt". In den beiden numerischen Anwendungen zu (5.50) und (5.52) (Abbildung 5.5) wurde mit der Referenzfunktion $v_0(x) = \sin(\frac{2\pi x}{6})$ gestartet, wobei im Intervall $x \in [14, 24]$, in dem die Daten für eine Abweichung sprechen, das Startpotential auf Null gesetzt wurde. Das Vorgehen mit (5.50) hat einen entscheidenden Nachteil: Kommt es am Ort der Fehlstelle zu zufälligen Überschneidungen von Potential und Referenzpotential, so wird unerwünschterweise in diesen Bereichen die Referenz v_0 angestellt. Der gleiche Effekt tritt auf, wenn im Falle von (5.52) die Ableitungen übereinstimmen. Ähnlich wie bei der Verwendung von (5.49) läßt sich feststellen, daß die Position der Störstelle identifiziert wird, aber die Nullfunktion als Referenz für das Störstellenpotential wenig geeignet ist. In einem letzten Beispiel wird eine zweite Referenzfunktion für die Störstelle zur Verfügung gestellt.

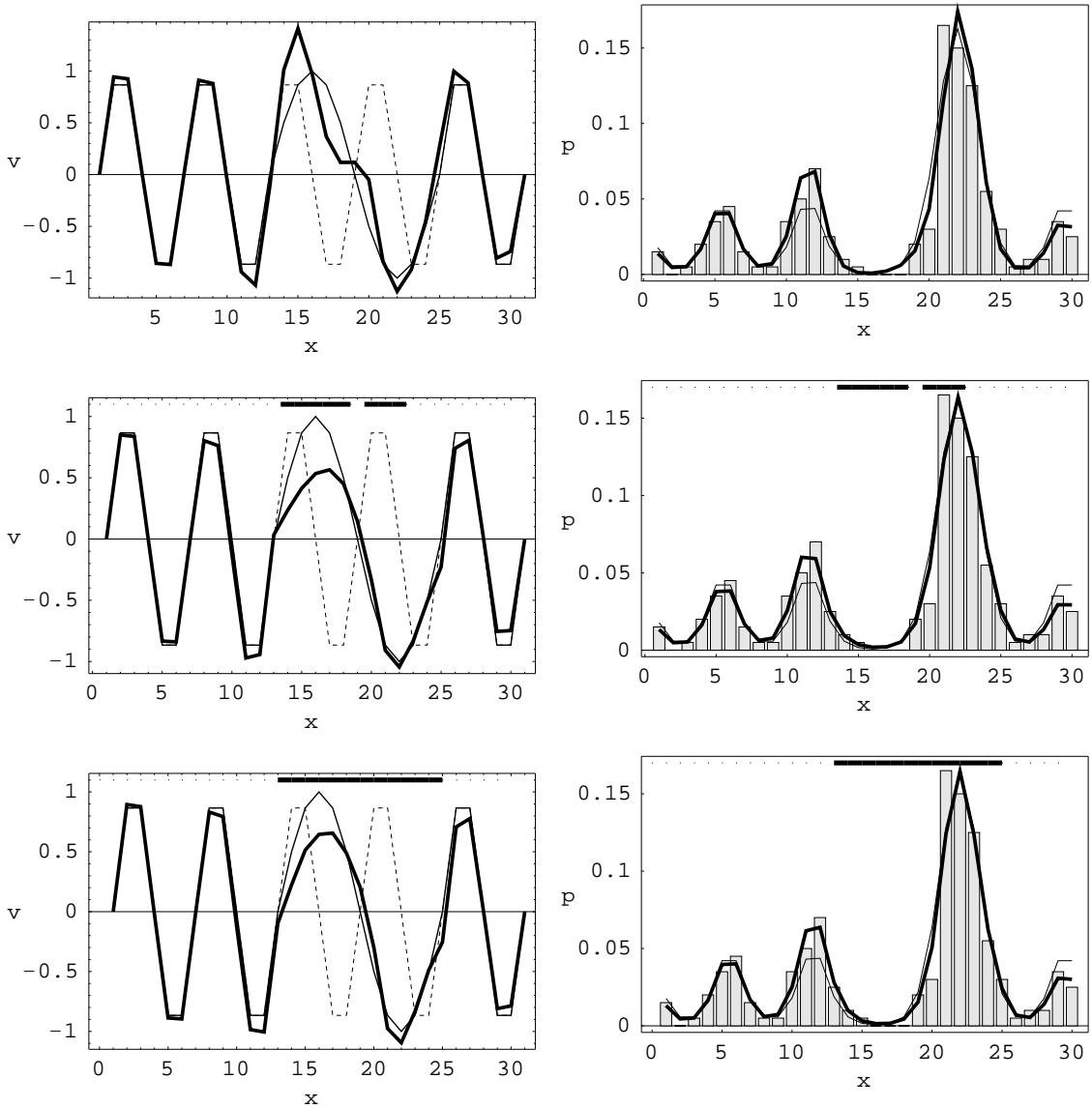


Abbildung 5.5: Rekonstruktion von näherungsweise periodischen Potentialen. Links sind Originalpotential (dünne Linie), rekonstruiertes Potential (dicke Linie) und Referenzpotential v_0 (gestrichelte Linie) eingezeichnet, rechts die dazugehörigen Dichteveerteilungen mit der empirischen Häufigkeit der Daten. Die Temperatur des kanonischen Ensembles beträgt $\beta = 4$. Es wurden $N = 200$ Daten gewonnen, und die Masse des betrachteten Teilchens ist $m = 0.25$. Oben ist das Ergebnis zu (5.49) dargestellt. Die Energieerwartungswerte betragen $\langle H \rangle_v = -0.353$ und $U = \langle H \rangle_{v_D} = -0.388$, die Hyperparameter sind $\chi = 1000$ und $\lambda = 0.05$. In der Mitte wurde das Funktional aus (5.50) verwendet. Die Balken im Diagramm markieren die Stellen, an denen der Einfluß von v_0 abgestellt ist. Die Energien liegen bei $\langle H \rangle_v = -0.382$ und $U = \langle H \rangle_{v_D} = -0.388$. Die vorgegebenen Parameter sind: $\chi = 10$, $\gamma = 1$, $\lambda = 5$ und $b = 0.15$. Unten: Ergebnis bei Gebrauch von (5.52). Die Energien sind $\langle H \rangle_v = -0.354$ und $U = \langle H \rangle_{v_D} = -0.388$. Als Hyperparameter wurden $\chi = 100$, $\gamma = 0.8$, $\lambda = 4$ und $b = 0.025$ gewählt.

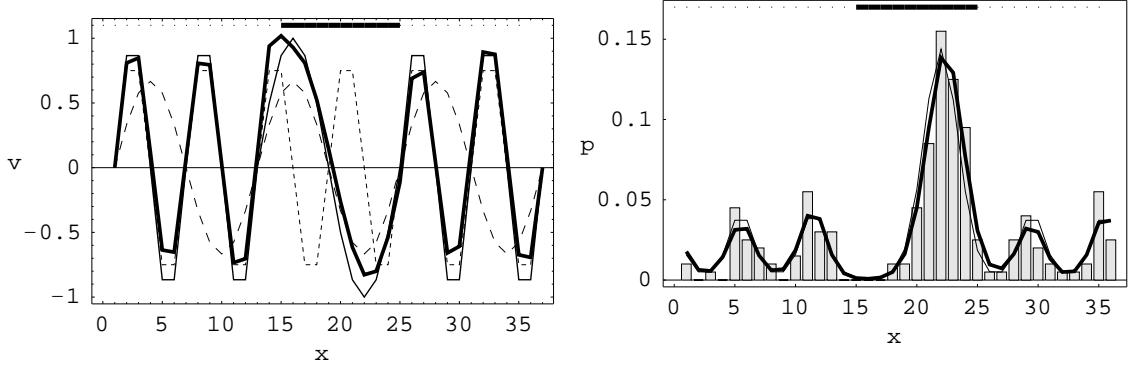


Abbildung 5.6: Rekonstruktion von näherungsweise periodischen Potentialen nach (5.57). Bedeutung der Linien wie in Abbildung 5.5 mit zusätzlichem zweitem Referenzpotential. Die Hyperparameter sind $\lambda_1 = \lambda_2 = 10$, $k = 10$, $N = 200$, $\beta = 4$. Als Startpotential für die Iteration wurde $v^{(0)}(x) = v_2(x)$ für $x \in [16, 25]$ und $v^{(0)}(x) = v_1(x)$ für den Komplementbereich gewählt.

Bei der Verwendung der beiden Muster

$$v_1(x) = \sin^2\left(\frac{2\pi}{6}x\right) \operatorname{sign}\left[\sin\left(\frac{2\pi}{6}x\right)\right] \quad \text{und} \quad v_2(x) = \frac{2}{3} \sin\left(\frac{2\pi}{12}x\right) \quad (5.54)$$

liegen jeweils leicht abgewandelte Komponenten des Originalpotentials (5.48) vor. Die Funktionen

$$l_i(x) = \left(\frac{\partial}{\partial x} (v(x) - v_i(x)) \right)^2 \quad i \in \{1, 2\} \quad (5.55)$$

geben den lokalen Abstand von v zu den jeweiligen Mustern an. Als effektives Muster wird ein zusammengesetztes Muster konstruiert, das lokal aus dem jeweils im Sinne von (5.55) näheren Muster $v_i(x)$ besteht:

$$v_0(x) = \begin{cases} v_1(x) & \text{für } x \text{ mit } \theta[l_1(x) - l_2(x)] = 0 \\ v_2(x) & \text{für } x \text{ mit } \theta[l_1(x) - l_2(x)] = 1 \end{cases} \quad (5.56)$$

Somit wird jedem Potential v ein effektives Muster zugeordnet ($v_0 = v_0(v)$). Ein Problem könnte sein, daß aufgrund zufälliger Ähnlichkeiten vor allem im Bereich der Störstelle unerwünscht häufige Umschaltungen zwischen den Mustern auftreten. Dies ist die Motivation, einen Bestrafungsterm einzuführen, der die Anzahl der Umschaltungen zwischen den Mustern zählt: $\mathcal{Z}(v)$ = "Anzahl der Umschaltungen". Als insgesamt zu minimierendes Funktional wählt man dann

$$\begin{aligned} F[v] = & \frac{\lambda_1}{2} \int dx l_1(x) \theta[l_2(x) - l_1(x)] + \frac{\lambda_2}{2} \int dx l_2(x) \theta[l_1(x) - l_2(x)] \\ & - \sum_{i=1}^N \ln \langle x_i | \frac{e^{-\beta H(v)}}{Z(v)} | x_i \rangle + k \mathcal{Z}(v). \end{aligned} \quad (5.57)$$

Das Resultat einer Iteration ist sehr empfindlich gegenüber dem Startpotential. Dies liegt daran, daß das Potential selber darüber entscheidet, welches effektive Muster es "sieht". In den Anwendungen wurde ein Startpotential gewählt, bei dem Potential und effektives Muster zusammenfallen:

$v^{(0)} \equiv v_0(v^{(0)})$. Damit verschwinden im Energiefunktional (5.57) für $v^{(0)}$ die ersten beiden Summanden, und es gilt:

$$F[v^{(0)}] = - \sum_{i=1}^N \ln \langle x_i | \frac{e^{-\beta H(v^{(0)})}}{Z(v^{(0)})} | x_i \rangle + k \mathcal{Z}(v^{(0)}). \quad (5.58)$$

Vor dem eigentlichen Iterationsprozeß wird nach einem $v^{(0)}$ gesucht, welches das Funktional (5.58) für sich minimiert, und dieses dann als Startpotential für die Iteration nach (5.57) verwendet. In Abbildung 5.6 ist das Ergebnis für dieses Vorgehen dargestellt. Die Suche nach $v^{(0)}$ führt bei $k = 10$ auf ein Startpotential, das im Intervall $x \in [16, 25]$ aus $v_2(x)$ und sonst aus $v_1(x)$ besteht. Das dazugehörige Ergebnis zeigt auch in den Stellen geringer Datenhäufigkeit eine trotz fehlerhafter Muster gute Rekonstruktion.

Kapitel 6

Lernen von Potentialen (Pfadintegraldarstellung)

In diesem Kapitel wird mittels Pfadintegralen eine alternative Methode zur Darstellung der Matrixelemente des statistischen Operators gewählt. Die Verwendung von Pfadintegralen und die resultierende Stationaritätsgleichung für die Maximum–Aposteriori–Approximation ist äquivalent zur entsprechenden Gleichung aus Kapitel 5. Im Unterschied zur Spektraldarstellung ergibt sich mit der Möglichkeit für eine Stationäre Phasenapproximation der Pfadintegrale eine semiklassische Näherungsstrategie für die Berechnung der Matrixelemente. Während man im vorangegangenen Kapitel auf den Umgang mit Eigenwertgleichungen angewiesen war, erfordert die Stationäre Phasenapproximation das Lösen von klassischen Bewegungsgleichungen.

6.1 Erste Variante

Die Pfadintegraldarstellung der Matrixelemente des statistischen Operators als Alternative zu (5.14) besitzt nach [14, 15, 16] die funktionale Form

$$\langle x_i | e^{-\beta H} | x_i \rangle = \int_{q(0)=x_i}^{q(\beta\hbar)=x_i} \frac{[dq(t)]}{\mathcal{N}} e^{-\frac{1}{\hbar} \int_0^{\beta\hbar} dt \left[\frac{1}{2} m \dot{q}^2(t) + v(q(t)) \right]}. \quad (6.1)$$

Die Darstellung (6.1) ist eine Abkürzung für eine unendlichdimensionale Ortsintegration, die sich ergibt, wenn das Zeitintervall $\beta\hbar$ in äquidistante Teilstücke der Länge $\epsilon = \frac{\beta\hbar}{M}$ geteilt wird, der Pfad $q(t)$ an den diskreten Zeiten durch $q_k = q(t_k)$ mit $t_k = \epsilon k$ codiert und anschließend der Grenzwert $M \rightarrow \infty$ ausgeführt wird:

$$\langle x_i | e^{-\beta H} | x_i \rangle = \lim_{M \rightarrow \infty} \left(\frac{m}{2\pi\hbar\epsilon} \right)^{\frac{M}{2}} \int \left(\prod_{k=1}^{M-1} dq_k \right) e^{-\frac{\epsilon}{\hbar} \sum_{k=1}^M \left(\frac{m}{2} \left(\frac{q_k - q_{k-1}}{\epsilon} \right)^2 + v(q_k) \right)}. \quad (6.2)$$

Hierbei sind die Randwerte $q_0 = q_M = x_i$ vorgegeben. Es wurde ohne Beschränkung der Allgemeinheit von einer eindimensionalen Ortskoordinate x_i ausgegangen. \mathcal{N} aus (6.1) ist eine v – und x_i –unabhängige Normierungskonstante:

$$\mathcal{N}^{-1} = \lim_{M \rightarrow \infty} \left(\frac{m}{2\pi\hbar\epsilon} \right)^{\frac{M}{2}}. \quad (6.3)$$

Das Integral (6.1) entspricht also einer Summe über alle möglichen Pfade $q(t)$ des Definitionsbereichs $t \in [0, \beta\hbar]$, wobei die zeitlichen Randbedingungen $q(0) = q(\beta\hbar) = x_i$ erfüllt sein müssen.

Nur in einzelnen Spezialfällen lässt sich das Integral analytisch lösen. Während bei der spektralen Behandlung die Matrixelemente und deren Ableitung exakt bearbeitet wurden, bietet sich hier die Methode der Stationären Phasen als Näherung an. Dabei wird nach Pfaden gesucht, die die euklidische Wirkung minimieren und somit den Hauptbeitrag zum Pfadintegral liefern. Dieses führt auf das Lösen von klassischen Bewegungsgleichungen, wobei durch die Datenpunkte x_i die zeitlichen Randbedingungen vorgegeben werden. Da die Güte dieser Näherung im Limes $\hbar \rightarrow 0$ zunimmt, spricht man auch von einer semiklassischen Näherung.

Die Zustandssumme lässt sich als Spur im Ortsraum durch ein Pfadintegral über alle periodischen Funktionen mit vorgegebener Periode $\beta\hbar$ ausdrücken:

$$\begin{aligned} Z &= \text{Tr}[e^{-\beta H}] = \int dx \langle x | e^{-\beta H} | x \rangle \\ &= \int dx \int_{q(0)=x}^{q(\beta\hbar)=x} \frac{[dq(t)]}{\mathcal{N}} e^{-\frac{1}{\hbar}S[q]} = \int_{q(0)=q(\beta\hbar)} \frac{[dq(t)]}{\mathcal{N}} e^{-\frac{1}{\hbar}S[q]}. \end{aligned} \quad (6.4)$$

Dabei ist die euklidische Wirkung mit

$$S[q] = \int_0^{\beta\hbar} dt \left[\frac{1}{2} m \dot{q}^2(t) + v(q(t)) \right] \quad (6.5)$$

abgekürzt.

Setzt man (6.1) in (5.5) ein und berücksichtigt, daß jedes Datum ein eigenes Pfadintegral erhält, so bekommt man als Ausdruck für die Aposteriori-Wahrscheinlichkeit des Potentials:

$$p[v|D] \sim \int \left(\prod_{i=1}^N [dq_i(t)] \right) e^{-\frac{1}{\hbar}F[q_i, v]}, \quad (6.6)$$

wobei die Gesamtwirkung F durch

$$F[q_i, v] = \sum_{i=1}^N \int_0^{\beta\hbar} dt \left[\frac{1}{2} m \dot{q}_i^2(t) + v(q_i(t)) \right] + \hbar N \ln Z(v) + \hbar \gamma \Gamma[v] \quad (6.7)$$

gegeben ist. In einer ersten Variante für die Behandlung der Wahrscheinlichkeit $p[v|D]$ wird eine Stationäre Phasenapproximation der Pfadintegrale $q_i(t)$ mit einer gleichzeitigen Maximum-Aposteriori-Approximation für v vorgenommen. Dies entspricht dann dem Lösen der Bewegungsgleichungen:

$$0 = \frac{\delta F}{\delta q_i(t)} \implies m\ddot{q}_i(t) = v'(q_i(t)) \quad , \quad q_i(0) = q_i(\beta\hbar) = x_i \quad \forall i = 1, \dots, N, \quad (6.8)$$

welche die Bewegung eines klassischen Teilchens in dem Potential $-v(x)$ unter periodischen Randbedingungen festlegen. Auf jedem Pfad bleibt die Energie

$$E_i = \frac{1}{2} m \dot{q}_i^2(t) - v(q_i(t)) \quad (6.9)$$

zeitlich erhalten. Die zweite, simultan mit (6.8) zu behandelnde Bedingung lautet:

$$0 = \frac{\delta F}{\delta v(x')} = \sum_{i=1}^N \int_0^{\beta\hbar} dt \delta(q_i(t) - x') - \beta\hbar N \langle x' | \frac{e^{-\beta H}}{Z} | x' \rangle + \gamma\hbar \frac{\delta \Gamma}{\delta v(x')}. \quad (6.10)$$

Hierbei wurde für die Ableitung der Zustandssumme das Ergebnis aus (5.19) ausgenutzt. Die Integrale über die δ -Funktionen können im Falle von einfachen Nullstellen im Argument unter Ausnutzung von (6.9) mit

$$\int_0^{\beta\hbar} dt \delta(q_i(t) - x') = \sum_{j_i=1}^{n_i(x')} \frac{1}{|\dot{q}_i(t_{j_i})|} = \sum_{j_i=1}^{n_i(x')} \frac{1}{\sqrt{\frac{2}{m}(E_i + v(q_i(t_{j_i})))}} = \frac{n_i(x')}{\sqrt{\frac{2}{m}(E_i + v(x'))}} \quad (6.11)$$

angegeben werden. Dabei sind mit $n_i(x')$ die Anzahl der Zeiten t_{j_i} mit $q_i(t_{j_i}) = x'$ gekennzeichnet. Die Bedingung aus (6.10) lässt sich kompakt ausdrücken, indem man mit einer beliebigen Observablen $f(x')/(N\beta\hbar)$ multipliziert und danach x' ausintegriert:

$$\begin{aligned} 0 &= \frac{1}{N\hbar\beta} \sum_{i=1}^N \int_0^{\beta\hbar} dt f(q_i(t)) - \int dx' f(x') \langle x' | \frac{e^{-\beta H}}{Z} | x' \rangle + \frac{\gamma}{N\beta} \int dx' f(x') \frac{\delta\Gamma}{\delta v(x')} \\ &= \frac{1}{N} \sum_{i=1}^N \bar{f}_i - \langle f \rangle + \frac{\gamma}{N\beta} \int dx' f(x') \frac{\delta\Gamma}{\delta v(x')} \quad \forall f, \end{aligned} \quad (6.12)$$

wobei \bar{f}_i den zeitlichen Mittelwert von f auf der klassischen Trajektorie $q_i(t)$ angibt und $\langle f \rangle$ für den Erwartungswert der Observablen f steht. Die beiden Gleichungen (6.10) und (6.12) sind äquivalent, da (6.12) für alle f erfüllt sein muß. Funktionalableitung von (6.12) bezüglich f führt zurück auf (6.10). Das Ergebnis (6.12) erinnert sehr an die Aussage der Ergodenhypothese [17] der statistischen Mechanik. Dort ist die Gleichheit von "Zeitmittel" und "Ensemblemittel" eine der wesentlichen Annahmen für alle physikalischen Observablen. Im Gegensatz zu (6.12) erfordert die Ergodenhypothese allerdings eine Zeitmittelung über einen unendlichen Zeitraum, und die Trajektorien $q(t)$ werden durch Anfangsbedingungen für q_i , \dot{q}_i und nicht über ein Randwertproblem (6.8) vorgegeben.

Diskutiert man den Hochtemperaturlimes $\beta \rightarrow 0$ von (6.12), so können im ersten Summanden die Integrale mittels $\int_0^{\beta\hbar} dt f(q_i(t)) \approx \beta\hbar f(x_i)$ genähert werden. Hierbei wurde die zu erwartende Stetigkeit von $q_i(t)$ ausgenutzt. Der zweite Summand wird wegen $e^{-\beta H} \approx \mathcal{I}$ und damit $\langle f \rangle \approx (\int dx' f(x')) / \int dx'$ ebenfalls β -unabhängig. Wie zu erwarten dominiert der Prior-Term in diesem Grenzfall die Stationaritätsgleichung (6.12). Die Gleichung (6.12) darf im Tieftemperaturfall nicht benutzt werden, da für deren Herleitung explizit $\beta \neq \infty$ benutzt wurde.

Falls kein Vorwissen vorhanden ist ($\Gamma = 0$), kann analog zu Abschnitt 5.5 für den Tieftemperaturfall gezeigt werden, daß

$$v(x) = - \lim_{a \rightarrow \infty} a \sum_i \delta(x_i - x) \quad (6.13)$$

die Lösung von (6.10) ist. Denn die Bewegungsgleichungen (6.8) haben dann die instabilen Lösungen $q_i(t) \equiv x_i$, und damit folgt für den ersten Summanden von (6.10):

$$\sum_i \int dt \delta(q_i(t) - x') = \beta\hbar \sum_i \delta(x_i - x'). \quad (6.14)$$

Mit (5.36) gilt für dieses Potential

$$\lim_{\beta \rightarrow \infty} \langle x' | \frac{e^{-\beta H}}{Z} | x' \rangle = \frac{1}{N} \sum_i^N \delta(x_i - x'). \quad (6.15)$$

Damit heben sich der erste und zweite Summand aus (6.10) weg.

Im Gegensatz zu Kapitel 5 ist die Beschränkung auf Ortsmessungen für die Gültigkeit von (6.1)

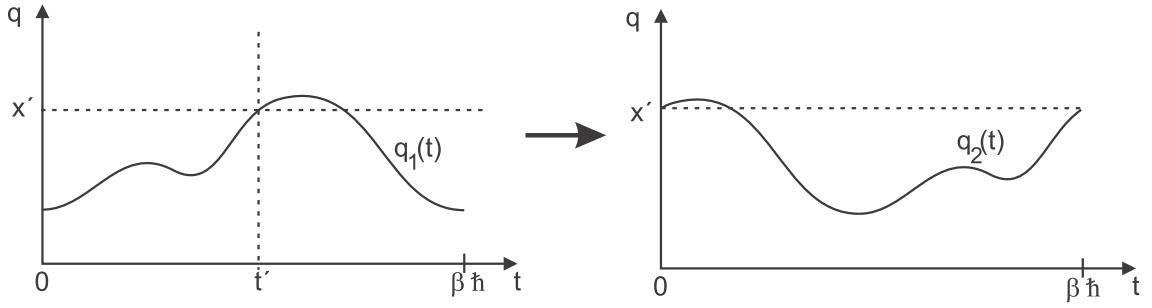


Abbildung 6.1: Graphische Erläuterung der Argumentation für den Schritt von (6.18) nach (6.19).

essentiell. Möchte man mit anderen Meßobservablen arbeiten, aber gleichzeitig den Pfadintegralformalismus beibehalten, so können die Matrixelemente über die Vollständigkeit dargestellt werden. Im Falle von Impulsmessungen \tilde{p} bekommt man mittels Fourier-Transformation

$$\begin{aligned}
 \langle \tilde{p} | e^{-\beta H} | \tilde{p} \rangle &= \int dx dx' \langle \tilde{p} | x' \rangle \langle x' | e^{-\beta H} | x \rangle \langle x | \tilde{p} \rangle \\
 &\sim \int dx dx' \int_{q(0)=x}^{q(\beta \hbar)=x'} \frac{[dq(t)]}{\mathcal{N}} e^{-\frac{1}{\hbar}(S[q] + i\tilde{p}(x' - x))} \\
 &= \int \frac{[dq(t)]}{\mathcal{N}} e^{-\frac{1}{\hbar}(S[q] + i\tilde{p}(q(\beta \hbar) - q(0)))}
 \end{aligned} \tag{6.16}$$

ein Integral über sämtliche Pfade $q(t)$ des Definitionsbereichs $t \in [0, \beta \hbar]$. Für den weiteren Verlauf der Rechnung bietet sich beispielsweise eine Sattelpunktnäherung als Näherungsmethode für die Ortsintegrale an.

6.2 Zweite Variante

In (6.6) wurde das Produkt der Pfadintegrale effektiv über ein einzelnes Pfadintegral ausgedrückt und gleichzeitig eine Maximierung bezüglich v und eine Stationäre Phasenapproximation für das gesamte Integral vorgenommen. In diesem Abschnitt wird mittels Pfadintegralmethoden eine zu (5.22) alternative Darstellung für $\frac{\delta}{\delta v(x')} \langle x_i | e^{-\beta H} | x_i \rangle$ und $\frac{\delta Z}{\delta v(x')}$ hergeleitet und in die Stationaritätsbedingung zu (5.6):

$$0 = \gamma \frac{\delta \Gamma}{\delta v(x')} + \frac{N}{Z} \frac{\delta Z}{\delta v(x')} - \sum_{i=1}^N \frac{\frac{\delta}{\delta v(x')} \langle x_i | e^{-\beta H} | x_i \rangle}{\langle x_i | e^{-\beta H} | x_i \rangle} \tag{6.17}$$

eingesetzt.

Eine Funktionalableitung der Zustandssumme aus (6.4) nach $v(x')$ mit anschließender Vertauschung der Reihenfolge der beiden Integrationen führt auf:

$$\begin{aligned}
 \frac{\delta Z}{\delta v(x')} &= -\frac{1}{\hbar} \int_{q(0)=q(\beta \hbar)} \frac{[dq(t)]}{\mathcal{N}} e^{-\frac{1}{\hbar} \int_0^{\beta \hbar} dt [\frac{1}{2} m \dot{q}^2(t) + v(q(t))]} \int_0^{\beta \hbar} dt' \delta(q(t') - x') \\
 &= -\frac{1}{\hbar} \int_0^{\beta \hbar} dt' \int_{q(0)=q(\beta \hbar)} \frac{[dq(t)]}{\mathcal{N}} e^{-\frac{1}{\hbar} \int_0^{\beta \hbar} dt [\frac{1}{2} m \dot{q}^2(t) + v(q(t))]} \delta(q(t') - x') .
 \end{aligned} \tag{6.18}$$

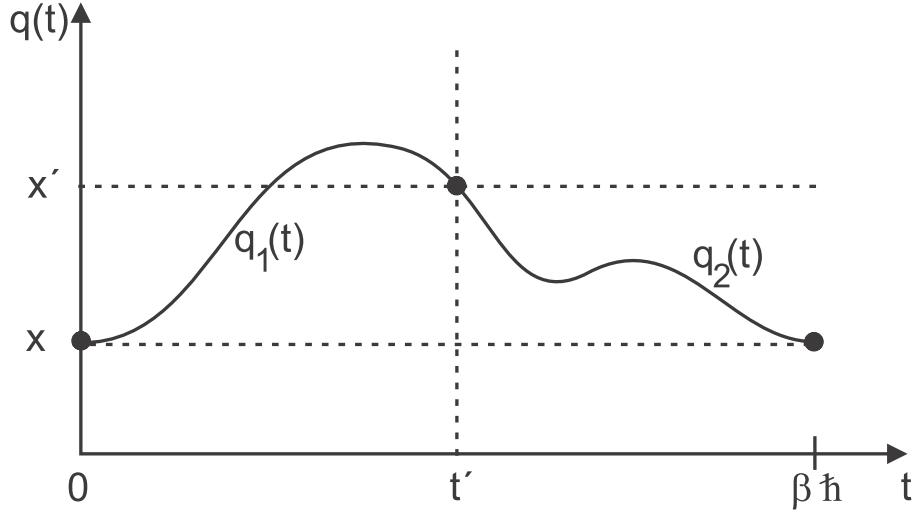


Abbildung 6.2: Aufspaltung des Pfads q in die beiden Anteile q_1 und q_2 .

Der Wert des Pfadintegrals kann direkt angegeben werden, wenn man sich veranschaulicht, daß Wirkung $S[q]$ und Normierung \mathcal{N} invariant unter einer zyklischen Verschiebung jedes Pfades $q(t)$ um den Wert t' auf der Zeitachse sind. Diese Verschiebung ist in der Abbildung 6.1 angedeutet. Beide Pfade q_1 und q_2 nehmen insgesamt die gleichen Funktionswerte an und besitzen deshalb den gleichen Wert für $\int dt v(q(t))$. Die Pfade nehmen ebenfalls auf dem Zeitintervall die gleichen Ableitungswerte $\dot{q}(t)$ an. Somit ist für beide Pfade das Integral $\int dt \dot{q}^2(t)$ identisch. Der Argumentation folgend kann das Pfadintegral aus (6.18), bei dem über alle periodischen Pfade, die zugleich durch den Punkt x' zur Zeit t' gehen, effektiv als ein Pfadintegral über alle Pfade, die bei x' starten und enden, ausgedrückt werden:

$$\int_{q(0)=q(\beta\hbar)} \frac{[dq(t)]}{\mathcal{N}} e^{-\frac{1}{\hbar}S[q]} \delta(q(t') - x') = \int_{q(0)=x'}^{q(\beta\hbar)=x'} \frac{[dq(t)]}{\mathcal{N}} e^{-\frac{1}{\hbar}S[q]} = \langle x' | e^{-\beta H} | x' \rangle. \quad (6.19)$$

Hierbei wurde die Definition (6.1) ausgenutzt. Man beachte, daß die Gleichung (6.19) unabhängig von t' gültig ist. Die Ausführung des t' -Integrals aus (6.18) wird dadurch trivial, und das Ergebnis aus (5.19) wird reproduziert:

$$\frac{\delta Z}{\delta v(x')} = -\beta \langle x' | e^{-\beta H} | x' \rangle. \quad (6.20)$$

Mittels ähnlicher Methoden können die Ableitungen der Matrixelemente $\langle x_i | e^{-\beta H} | x_i \rangle$ in (6.17) berechnet werden. Nach Ableitung

$$\begin{aligned} \frac{\delta}{\delta v(x')} \langle x_i | e^{-\beta H} | x_i \rangle &= -\frac{1}{\hbar} \int_0^{\beta\hbar} dt' \int_{q(0)=x_i}^{q(\beta\hbar)=x_i} \frac{[dq(t)]}{\mathcal{N}} e^{-\frac{1}{\hbar}S[q]} \delta(q(t') - x') \\ &= -\frac{1}{\hbar} \int_0^{\beta\hbar} dt' \int_{\substack{q(t')=x' \\ q(0)=x_i}}^{q(\beta\hbar)=x_i} \frac{[dq(t)]}{\mathcal{N}} e^{-\frac{1}{\hbar}S[q]} \end{aligned} \quad (6.21)$$

kann das Pfadintegral in zwei separate Integrale aufgespalten werden (siehe Abbildung 6.2). Nach Berücksichtigung der zeitlichen Randbedingungen für die Pfade q_1 und q_2 entsteht ein Produkt

von Übergangsmatrixelementen von statistischen Operatoren mit abgeänderten Temperaturen:

$$\begin{aligned} \frac{\delta \langle x_i | e^{-\beta H} | x_i \rangle}{\delta v(x')} &= -\frac{1}{\hbar} \int_0^{\beta \hbar} dt' \int_{q_2(t')=x'}^{q_2(\beta \hbar)=x_i} \frac{[dq_2]}{\mathcal{N}_2} e^{-\frac{1}{\hbar} S_2[q_2]} \int_{q_1(0)=x_i}^{q_1(t')=x'} \frac{[dq_1]}{\mathcal{N}_1} e^{-\frac{1}{\hbar} S_1[q_1]} \\ &= -\frac{1}{\hbar} \int_0^{\beta \hbar} dt' \langle x_i | e^{-(\beta - \frac{t'}{\hbar})H} | x' \rangle \langle x' | e^{-\frac{t'}{\hbar}H} | x_i \rangle \\ &= -\int_0^{\beta} d\beta' \langle x_i | e^{-(\beta - \beta')H} | x' \rangle \langle x' | e^{-\beta' H} | x_i \rangle, \end{aligned} \quad (6.22)$$

wobei $\mathcal{N} = \mathcal{N}_1 \mathcal{N}_2$ gilt, und die Wirkung mittels

$$S_1[q_1] = \int_0^{t'} dt \left[\frac{1}{2} m \dot{q}_1^2(t) + v(q_1(t)) \right] \quad \text{und} \quad S_2[q_2] = \int_{t'}^{\beta \hbar} dt \left[\frac{1}{2} m \dot{q}_2^2(t) + v(q_2(t)) \right] \quad (6.23)$$

aufgeteilt wurde.

Man beachte, daß die Ableitung (6.22) – ebenso wie (5.22) – exakt ist. Deshalb muß (6.22) mit den Ergebnissen aus Kapitel 5 analytisch übereinstimmen. Um die Äquivalenz zu beweisen, führt man in beiden Matrixelementen aus (6.22) eine Identität aus Energieniveaufunktionen (5.2) $\mathcal{I} = \sum_n \frac{|\varphi_n\rangle\langle\varphi_n|}{\langle\varphi_n|\varphi_n\rangle}$ ein und nutzt die Eigenwertgleichung aus. So erhält man:

$$\begin{aligned} \frac{\delta \langle x_i | e^{-\beta H} | x_i \rangle}{\delta v(x')} &= -\sum_{mn} \frac{\langle x_i | \varphi_n \rangle \langle \varphi_n | x' \rangle \langle x' | \varphi_m \rangle \langle \varphi_m | x_i \rangle}{\langle \varphi_n | \varphi_n \rangle \langle \varphi_m | \varphi_m \rangle} \\ &\quad \times e^{-\beta E_n} \int_0^{\beta} d\beta' e^{-\beta'(E_m - E_n)}. \end{aligned} \quad (6.24)$$

Unter Verwendung von

$$\int_0^{\beta} d\beta' e^{-\beta'(E_m - E_n)} = \begin{cases} \frac{\beta}{E_m - E_n} & m = n \\ \frac{1}{E_m - E_n} - \frac{e^{-\beta(E_m - E_n)}}{E_m - E_n} & m \neq n \end{cases} \quad (6.25)$$

bekommt man für (6.24)

$$\begin{aligned} \frac{\delta \langle x_i | e^{-\beta H} | x_i \rangle}{\delta v(x')} &= -\beta \sum_m e^{-\beta E_m} \frac{\langle x_i | \varphi_m \rangle \langle \varphi_m | x' \rangle \langle x' | \varphi_m \rangle \langle \varphi_m | x_i \rangle}{\langle \varphi_m | \varphi_m \rangle^2} \\ &\quad - 2 \sum_{\substack{m \neq n \\ m \neq n}} \frac{e^{-\beta E_n}}{E_m - E_n} \frac{\text{Re} \left[\langle x_i | \varphi_m \rangle \langle \varphi_m | x' \rangle \langle x' | \varphi_n \rangle \langle \varphi_n | x_i \rangle \right]}{\langle \varphi_n | \varphi_n \rangle \langle \varphi_m | \varphi_m \rangle}. \end{aligned} \quad (6.26)$$

Setzt man (6.26) zusammen mit (6.20) in die Gleichung (6.17) ein, so ist die Äquivalenz der Stationaritätsgleichung

$$0 = \gamma \frac{\delta \Gamma}{\delta v(x')} - N \beta \langle x' | \frac{e^{-\beta H}}{Z} | x' \rangle + \sum_{i=1}^N \int_0^{\beta} d\beta' \frac{\langle x_i | e^{-(\beta - \beta')H} | x' \rangle \langle x' | e^{-\beta' H} | x_i \rangle}{\langle x_i | e^{-\beta H} | x_i \rangle} \quad (6.27)$$

mit dem Resultat (5.22) aus Kapitel 5 bewiesen.

Im weiteren Verlauf dieses Kapitels werden analytische Näherungsstrategien für die in Gleichung (6.27) auftretenden Matrixelemente diskutiert. Es kann gezeigt werden, unter welcher Approximation das exakte Ergebnis der zweiten Pfadintegralversion (6.27) in das durch Sattelpunktsnäherung

erhaltene Resultat der ersten Pfadintegralversion (6.10) übergeht. Zusätzlich wird aufgezeigt, wie ein im Vergleich zu (6.10) verbessertes Ergebnis aussehen kann.

Im Anhang E wird eine Verallgemeinerung von (6.20) auf eine N -fache Variationsableitung der Zustandssumme vorgenommen. Durch die Bestimmung höherer Ableitungen werden die analytischen Voraussetzungen zur Festlegung der Hesse-Matrix geliefert.

6.3 Näherungen für die Matrixelemente $\langle x | e^{-\beta H} | x \rangle$

Der Umgang mit den Matrixelementen $\langle x | e^{-\beta H} | x \rangle$ mittels Pfadintegralmethoden ist Standard. So konzentriert man sich bei der näherungsweisen Berechnung von

$$\langle x | e^{-\beta H} | x \rangle = \int_{q(0)=x}^{q(\beta\hbar)=x} \frac{[dq(t)]}{\mathcal{N}} e^{-\frac{1}{\hbar} S[q]} \quad (6.28)$$

auf die Suche nach den Pfaden $q_x(t)$, die die klassische Bewegungsgleichung

$$0 = \frac{\delta S}{\delta q_x(t)} = -m\ddot{q}_x(t) + v'(q_x(t)) \quad \text{mit} \quad q_x(0) = q_x(\beta\hbar) = x \quad (6.29)$$

erfüllen. Bei der Variablentransformation zu neuen Koordinaten $r(t)$ nach $q(t) = q_x(t) + r(t)$ liefert der Pfad q_x den Hauptbeitrag zum Pfadintegral, während $r(t)$ die Fluktuationen um diesen Pfad beschreibt. Es ist zu erwarten, daß gerade kleine Abweichungen von q_x großen Beitrag zum verbleibenden Pfadintegral über $r(t)$ liefern. Somit macht eine Entwicklung des Potentials um q_x nach

$$v(q) \approx v(q_x) + v'(q_x)(q - q_x) + \frac{1}{2}v''(q_x)(q - q_x)^2 \quad (6.30)$$

Sinn. Die Wirkung $S[q]$ verändert sich dann insgesamt zu:

$$\begin{aligned} S[q] &= \int_0^{\beta\hbar} dt \left[\frac{1}{2}m\dot{q}_x^2(t) + \frac{1}{2}m\dot{r}^2(t) + m\dot{q}_x(t)\dot{r}(t) + v(q_x(t) + r(t)) \right] \\ &\approx \int_0^{\beta\hbar} dt \left[\frac{1}{2}m\dot{q}_x^2(t) + v(q_x(t)) \right] + \int_0^{\beta\hbar} dt \left[\frac{1}{2}m\dot{r}^2(t) + \frac{1}{2}v''(q_x(t))r^2(t) \right] \\ &\quad + \int_0^{\beta\hbar} dt \left[m\dot{q}_x(t)\dot{r}(t) + v'(q_x(t))r(t) \right], \end{aligned} \quad (6.31)$$

wobei der letzte Summand aus (6.31) unter Ausnutzung von (6.29) nach

$$\int_0^{\beta\hbar} dt \left[m\dot{q}_x(t)\dot{r}(t) + v'(q_x(t))r(t) \right] = \left[m\dot{q}_x(t)r(t) \right]_0^{\beta\hbar} = 0 \quad (6.32)$$

verschwindet. Die stationäre Wirkung $S[q_x]$ kann aus dem Integral herausgezogen werden, und man erhält schließlich

$$\begin{aligned} \langle x | e^{-\beta H} | x \rangle &\approx e^{-\frac{1}{\hbar} S[q_x]} \int_{r(0)=0}^{r(\beta\hbar)=0} \frac{[dr(t)]}{\mathcal{N}} e^{-\frac{1}{\hbar} \int_0^{\beta\hbar} dt \left[\frac{1}{2}m\dot{r}^2(t) + \frac{1}{2}v''(q_x(t))r^2(t) \right]} \\ &= e^{-\frac{1}{\hbar} S[q_x]} \int_{r(0)=0}^{r(\beta\hbar)=0} \frac{[dr(t)]}{\mathcal{N}} e^{-\frac{1}{2\hbar} \int_0^{\beta\hbar} dt dt' r(t') \left(\frac{\delta^2 S[q]}{\delta q(t') \delta q(t)} \Big|_{q=q_x} \right) r(t)} \end{aligned} \quad (6.33)$$

mit

$$\frac{\delta^2 S[q]}{\delta q(t') \delta q(t)} \Big|_{q=q_x} \approx \delta(t - t') \left(-m \frac{\partial^2}{\partial t^2} + v''(q_x(t)) \right) \quad (6.34)$$

in der Näherung (6.30). Notwendige Bedingung für die Konvergenz des Pfadintegrals aus (6.33) ist eine positiv-definite Hesse-Matrix (6.34). Unter der Entwicklung von (6.30) ist $v''(x) > 0$ für alle x dafür hinreichend.

Das verbleibende Pfadintegral aus (6.33) kann mittels der "Shifting-Methode" [18, 14] exakt gelöst werden. Das Ergebnis wird hier nur angegeben. Es ist unter der Van Vleck-Formel bekannt und beträgt

$$\langle x | e^{-\beta H} | x \rangle \approx \left(\frac{2\pi\hbar\kappa_x(\beta\hbar)\kappa_x(0)}{m} \int_0^{\beta\hbar} \frac{dt}{\kappa_x^2(t)} \right)^{-\frac{1}{2}} e^{-\frac{1}{\hbar}S[q_x]} = \left(-\frac{1}{2\pi\hbar} \frac{\partial^2 S[q_x]}{\partial x^2} \right)^{\frac{1}{2}} e^{-\frac{1}{\hbar}S[q_x]}, \quad (6.35)$$

wobei $\kappa_x(t)$ eine Lösung der Gleichung

$$\frac{1}{m} v''(q_x(t)) = \frac{\ddot{\kappa}_x(t)}{\kappa_x(t)} \quad (6.36)$$

ist. Falls auf dem Pfad q_x die Geschwindigkeit $\dot{q}_x(t)$ nicht verschwindet, kann aufgrund von (6.29) $\kappa_x(t) = \dot{q}_x(t)$ gewählt werden.

Um die Zustandssumme in dieser Darstellung

$$Z = \int dx \langle x | e^{-\beta H} | x \rangle = \int dx \int_{q(0)=x}^{q(\beta\hbar)=x} \frac{[dq(t)]}{\mathcal{N}} e^{-\frac{1}{\hbar}S[q]} \quad (6.37)$$

auszurechnen, müßte eigentlich die Gleichung (6.29) für alle Randbedingungen $q_x(0) = q_x(\beta\hbar) = x$ gelöst werden. Eine zusätzlich ausgeführte Sattelpunktnäherung für das x -Integral vermeidet diese Aufgabe und führt zu der Bedingung, daß die Geschwindigkeiten auf den Rändern ebenfalls übereinstimmen [16, 19]. Es muß also nach Orten x_0 gesucht werden, für die neben (6.29) noch $\dot{q}_{x_0}(0) = \dot{q}_{x_0}(\beta\hbar)$ erfüllt ist. Damit erhält man für die Zustandssumme zu (6.33) für ein einzelnes x_0 :

$$Z \approx e^{-\frac{1}{\hbar}S[q_{x_0}]} \int_{r(0)=0}^{r(\beta\hbar)=0} \frac{[dr(t)]}{\mathcal{N}} e^{-\frac{1}{\hbar} \int_0^{\beta\hbar} dt \left[\frac{1}{2}m\dot{r}^2(t) + \frac{1}{2}v''(q_{x_0}(t))r^2(t) \right]} \quad (6.38)$$

und insgesamt für die Matrixelemente des statistischen Operators

$$\langle x | \frac{e^{-\beta H}}{Z} | x \rangle \approx \frac{A_x}{A_{x_0}} e^{-\frac{1}{\hbar}(S[q_x] - S[q_{x_0}])} \quad (6.39)$$

mit jeweiligen Pfadintegralen A_x und A_{x_0} aus (6.33) und (6.38).

Im Falle großer Massen m müssen die Ergebnisse der klassischen statistischen Mechanik (4.42) beziehungsweise (5.7) reproduziert werden. Dies kann bewiesen werden, indem man sich veranschaulicht, daß aufgrund von

$$\ddot{q}_x = \frac{1}{m} v'(q_x) \rightarrow 0 \quad \text{für } m \rightarrow \infty \quad (6.40)$$

die Lösungen der Bewegungsgleichungen (6.29) die statischen Pfade $q_x(t) \equiv x$ sind. Setzt man dieses Ergebnis in die euklidischen Wirkungen ein, so erhält man für (6.33):

$$\langle x | e^{-\beta H} | x \rangle \approx e^{-\beta v(x)} \int_{r(0)=0}^{r(\beta\hbar)=0} \frac{[dr(t)]}{\mathcal{N}} e^{-\frac{1}{\hbar} \int_0^{\beta\hbar} dt \left[\frac{1}{2}m\dot{r}^2(t) + \frac{1}{2}v''(x)r^2(t) \right]}. \quad (6.41)$$

Das verbleibende Pfadintegral eines harmonischen Oszillators wird im Falle großer Massen im Grenzwert $\frac{v''(x)}{m} \rightarrow 0$ mit

$$\int_{r(0)=0}^{r(\beta\hbar)=0} \frac{[dr(t)]}{\mathcal{N}} e^{-\frac{1}{\hbar} \int_0^{\beta\hbar} dt \left[\frac{1}{2}m\dot{r}^2(t) + \frac{1}{2}v''(x)r^2(t) \right]} \approx \sqrt{\frac{m}{2\pi\hbar^2\beta}} \quad (6.42)$$

unabhängig von $v''(x)$ [14], und damit folgt das klassische Ergebnis

$$\langle x | \frac{e^{-\beta H}}{Z} | x \rangle \approx \frac{e^{-\beta v(x)}}{\int dx' e^{-\beta v(x')}}. \quad (6.43)$$

Eine etwas genauere Rechnung bei großen Massen ergibt sich, indem in (6.33) genähert wird. In der Wirkung des Pfadintegrals dominiert für große m -Werte der Term der kinetischen Energie:

$$-\frac{1}{2\hbar} \int_0^{\beta\hbar} dt \left[m\dot{r}^2(t) + v''(q_x(t))r^2(t) \right] \approx -\frac{m}{2\hbar} \int_0^{\beta\hbar} dt \dot{r}^2(t). \quad (6.44)$$

Dadurch erhält man nach (6.44) eine x -unabhängige quadratische Korrektur und bekommt

$$\langle x | e^{-\beta H} | x \rangle \sim e^{-\frac{1}{\hbar} S[q_x]} \quad (6.45)$$

als Resultat.

6.4 Näherungen für $\int_0^\beta d\beta' \langle x | e^{-(\beta-\beta')H} | x' \rangle \langle x' | e^{-\beta' H} | x \rangle$

Bei Näherungen für das Ergebnis von (6.21) beziehungsweise (6.22) gibt es mehrere Vorgehensweisen, die im wesentlichen darauf abzielen, daß im dritten Summanden von (6.27) das Matrixelement $\langle x_i | e^{-\beta H} | x_i \rangle$ herausgekürzt werden kann. Es werden hier mehrere Möglichkeiten vorgestellt, wie dies geschehen kann.

Eine erste Möglichkeit besteht darin, ein Integral der Form $\int dx g(x)f(x)$ unter der Voraussetzung, daß $f(x)$ um x_0 lokalisiert ist und $g(x)$ dort schwach oszilliert, durch $g(x_0) \int dx f(x)$ zu approximieren. Dies entspricht in (6.21) dem Rausziehen von $\delta(q(t') - x')$ für den Pfad $q(t') = q_{x_i}(t')$, wobei q_{x_i} die klassische Bewegungsgleichung (6.29) mit den Randbedingungen $q_{x_i}(0) = q_{x_i}(\beta\hbar) = x_i$ erfüllt (und damit mit dem Pfad q_i aus (6.8) übereinstimmt):

$$\begin{aligned} \frac{\delta}{\delta v(x')} \langle x_i | e^{-\beta H} | x_i \rangle &= -\frac{1}{\hbar} \int_0^{\beta\hbar} dt' \int_{q(0)=x_i}^{q(\beta\hbar)=x_i} \frac{[dq(t)]}{\mathcal{N}} e^{-\frac{1}{\hbar} S[q]} \delta(q(t') - x') \\ &\approx -\frac{1}{\hbar} \int_0^{\beta\hbar} dt' \delta(q_{x_i}(t') - x') \int_{q(0)=x_i}^{q(\beta\hbar)=x_i} \frac{[dq(t)]}{\mathcal{N}} e^{-\frac{1}{\hbar} S[q]}. \end{aligned} \quad (6.46)$$

Beim Einsetzen in (6.27) kürzt sich gerade das t' -unabhängige Matrixelement heraus

$$\frac{\delta}{\delta v(x')} \frac{\langle x_i | e^{-\beta H} | x_i \rangle}{\langle x_i | e^{-\beta H} | x_i \rangle} = -\frac{1}{\hbar} \int_0^{\beta\hbar} dt' \delta(q_{x_i}(t') - x'), \quad (6.47)$$

und damit geht die Stationaritätsgleichung (6.27) in das Ergebnis von (6.10) über. Möchte man genauer rechnen, so kann analog (6.33) in einem ersten Schritt die klassische Wirkung durch Variablentransformation $q(t) = q_{x_i}(t) + r(t)$ herausgezogen werden. Eine anschließende Näherung nach (6.46) bringt:

$$\begin{aligned} \frac{\delta \langle x_i | e^{-\beta H} | x_i \rangle}{\delta v(x')} &\approx -\frac{e^{-\frac{1}{\hbar} S[q_{x_i}]}}{\hbar} \int_0^{\beta\hbar} dt' \int_{r(0)=0}^{r(\beta\hbar)=0} \frac{[dr(t)]}{\mathcal{N}} e^{-\frac{1}{\hbar} \int_0^{\beta\hbar} dt \left[\frac{1}{2} m\dot{r}^2(t) + \frac{1}{2} v''(q_{x_i}(t))r^2(t) \right]} \delta(q_{x_i}(t') + r(t') - x') \\ &\approx -\frac{e^{-\frac{1}{\hbar} S[q_{x_i}]}}{\hbar} \int_0^{\beta\hbar} dt' \delta(q_{x_i}(t') + r_{x_i}(t') - x') \\ &\quad \times \int_{r(0)=0}^{r(\beta\hbar)=0} \frac{[dr(t)]}{\mathcal{N}} e^{-\frac{1}{\hbar} \int_0^{\beta\hbar} dt \left[\frac{1}{2} m\dot{r}^2(t) + \frac{1}{2} v''(q_{x_i}(t))r^2(t) \right]}, \end{aligned} \quad (6.48)$$

wobei r_{x_i} die Lösung von

$$m\ddot{r}_{x_i}(t) = v''(q_{x_i}(t))r_{x_i}(t) \quad \text{mit} \quad r_{x_i}(0) = r_{x_i}(\beta\hbar) = 0 \quad (6.49)$$

ist. Unter Ausnutzung von (6.33) bekommt man im Vergleich zu (6.47) bei

$$\frac{\delta \langle x_i | e^{-\beta H} | x_i \rangle}{\delta v(x')} = -\frac{1}{\hbar} \int_0^{\beta\hbar} dt' \delta(q_{x_i}(t') + r_{x_i}(t') - x') \quad (6.50)$$

einen Korrekturterm r_{x_i} , der die Fluktuationen um die Mean-Field-Lösung q_{x_i} beschreibt.

Eine zweite Möglichkeit für die Bearbeitung von (6.22) besteht darin, die δ -Funktion aus (6.48) in ihrer Spektraldarstellung einzusetzen. Damit erhält man mit

$$\begin{aligned} \frac{\delta \langle x_i | e^{-\beta H} | x_i \rangle}{\delta v(x')} &\approx -\frac{e^{-\frac{1}{\hbar} S[q_{x_i}]}}{\hbar} \int_0^{\beta\hbar} dt' \int \frac{d\lambda}{2\pi} e^{i\lambda(q_{x_i}(t') - x')} \times \\ &\quad \int_{r(0)=0}^{r(\beta\hbar)=0} \frac{[dr(t)]}{\mathcal{N}} e^{-\frac{1}{\hbar} \int_0^{\beta\hbar} dt [\frac{1}{2}m\dot{r}^2(t) + \frac{1}{2}v''(q_{x_i}(t))r^2(t) - i\lambda\hbar\delta(t-t')r(t)]} \end{aligned} \quad (6.51)$$

ein Pfadintegral, welches in der Hamilton-Funktion neben dem Anteil des harmonischen Potentials zusätzlich ein rein imaginäres, δ -förmiges Potential enthält. Die Suche nach Sattelpunkten für das Pfadintegral führt auf die Bewegungsgleichung

$$-m\frac{\partial^2 r_{x_i t' \lambda}(t)}{\partial t^2} + v''(q_{x_i}(t))r_{x_i t' \lambda}(t) = i\lambda\hbar\delta(t-t') \quad \text{mit} \quad r_{x_i t' \lambda}(0) = r_{x_i t' \lambda}(\beta\hbar) = 0. \quad (6.52)$$

Dabei ist zu beachten, daß die Lösung r_{x_i} sowohl λ - als auch t' -abhängig ist. Der Pfad r_{x_i} ist im wesentlichen die Greensche-Funktion des Operators $-m\frac{\partial^2}{\partial t^2} + v''(q_{x_i}(t))$ mit vorgegebenen Randbedingungen. Denn falls man die Greensche-Funktion R_{x_i} nach

$$-m\frac{\partial^2 R_{x_i}(t, t')}{\partial t^2} + v''(q_{x_i}(t))R_{x_i}(t, t') = \delta(t-t') \quad \text{mit} \quad R_{x_i}(0, t') = R_{x_i}(\beta\hbar, t') = 0, \quad \forall t' \quad (6.53)$$

gefunden hat, so folgt für die Lösung von (6.52) direkt: $r_{x_i t' \lambda}(t) = i\hbar\lambda R_{x_i}(t, t')$.

Weiter ergibt sich durch Multiplikation von (6.52) mit $r_{x_i t' \lambda}$, anschließender t -Integration und einmaliger partieller Integration nach $\int dt r(t)\ddot{r}(t) = -\int dt \dot{r}^2(t)$ für den quadratischen Teil der stationären Wirkung von (6.51):

$$-\frac{1}{\hbar} \int_0^{\beta\hbar} dt \left[\frac{1}{2}m\dot{r}_{x_i t' \lambda}^2(t) + \frac{1}{2}v''(q_{x_i}(t))r_{x_i t' \lambda}^2(t) \right] = -\frac{i\lambda}{2}r_{x_i t' \lambda}(t') = \frac{\lambda^2\hbar}{2}R_{x_i}(t', t') \quad (6.54)$$

für alle λ und t' . Nach erneuter Variablentransformation $r(t) = r_{x_i t' \lambda}(t) + l(t)$ fallen in der Wirkung des Pfadintegrals aus (6.51) mit (6.52) erneut Terme mittels

$$\int_0^{\beta\hbar} dt \left[m\dot{r}_{x_i t' \lambda}(t)\dot{l}(t) + v''(q_{x_i}(t))l(t)r_{x_i t' \lambda}(t) - i\hbar l(t)\delta(t-t') \right] = \left[m\dot{r}_{x_i t' \lambda}(t)l(t) \right]_0^{\beta\hbar} = 0 \quad (6.55)$$

weg, und man erhält

$$\begin{aligned} \frac{\delta \langle x_i | e^{-\beta H} | x_i \rangle}{\delta v(x')} &\approx -\frac{e^{-\frac{1}{\hbar} S[q_{x_i}]}}{\hbar} \int_0^{\beta\hbar} dt' \int \frac{d\lambda}{2\pi} e^{i\lambda(q_{x_i}(t') - x')} e^{i\lambda r_{x_i t' \lambda}(t')} \times \\ &\quad e^{-\frac{1}{\hbar} \int_0^{\beta\hbar} dt [\frac{1}{2}m\dot{r}_{x_i t' \lambda}^2(t) + \frac{1}{2}v''(q_{x_i}(t))r_{x_i t' \lambda}^2(t)]} \int_{l(0)=0}^{l(\beta\hbar)=0} \frac{[dl(t)]}{\mathcal{N}} e^{-\frac{1}{\hbar} \int_0^{\beta\hbar} dt [\frac{1}{2}m\dot{l}^2(t) + \frac{1}{2}v''(q_{x_i}(t))l^2(t)]}. \end{aligned} \quad (6.56)$$

Nach Ausnutzen von (6.54) kann das Gauß–Integral über λ nach

$$\int \frac{d\lambda}{2\pi} e^{-\frac{1}{2}\hbar R_{x_i}(t', t')\lambda^2 + i\lambda(q_{x_i}(t') - x')} = \frac{1}{\sqrt{2\pi\hbar R_{x_i}(t', t')}} e^{-\frac{1}{2\hbar R_{x_i}(t', t')}(q_{x_i}(t') - x')^2} \quad (6.57)$$

ausgeführt werden. Als Endergebnis bekommt man mit (6.33):

$$\frac{\delta_{\delta v(x')}}{\delta v(x')} \frac{< x_i | e^{-\beta H} | x_i >}{< x_i | e^{-\beta H} | x_i >} = -\frac{1}{\hbar} \int_0^{\beta\hbar} dt' \frac{1}{\sqrt{2\pi\hbar R_{x_i}(t', t')}} e^{-\frac{1}{2\hbar R_{x_i}(t', t')}(q_{x_i}(t') - x')^2}. \quad (6.58)$$

Im Vergleich zu (6.47) und (6.50) sind die δ –Funktionen der obigen Ergebnisse hier durch eine über x' normierte Gauß–Verteilung ersetzt worden. Die Varianz dieser Verteilung wird durch die Greensche–Funktion festgelegt. Man beachte, daß durch Multiplikation von (6.53) mit $R_{x_i}(t, t')$ und Zeitintegration aufgrund von $v''(q_{x_i}) > 0$:

$$R_{x_i}(t', t') = \int dt \left[\frac{1}{2} m \left(\frac{\partial R_{x_i}(t, t')}{\partial t} \right)^2 + \frac{1}{2} v''(q_{x_i}(t)) R_{x_i}^2(t, t') \right] > 0 \quad (6.59)$$

folgt. Für besonders gut lokalisierte Minima ($v'' \rightarrow \infty$) geht das Resultat (6.58) in (6.47) über. Abschließend sei eine letzte Möglichkeit für die Behandlung von (6.22) vorgestellt. Die einzelnen Matrixelemente können separat durch ein Pfadintegral ausgedrückt werden. Nach einer Stationären Phasenapproximation für beide Pfadintegrale kann eine Sattelpunktsnäherung für das β' –Integral angeschlossen werden. Die Stationarität von

$$\begin{aligned} - \int_0^{\beta} d\beta' & < x_i | e^{-(\beta-\beta')H} | x' > < x' | e^{-\beta' H} | x_i > \\ & = - \int_0^{\beta} d\beta' \int_{q_2(\beta'\hbar)=x'}^{q_2(\beta\hbar)=x_i} \frac{[dq_2(t)]}{\mathcal{N}_2} \int_{q_1(0)=x_i}^{q_1(\beta'\hbar)=x'} \frac{[dq_1(t)]}{\mathcal{N}_1} e^{-\frac{1}{\hbar}(S_1[q_1] + S_2[q_2])} \end{aligned} \quad (6.60)$$

mit den einzelnen Wirkungen

$$S_1[q_1] = \int_0^{\beta'\hbar} dt \left[\frac{1}{2} m \dot{q}_1^2(t) + v(q_1(t)) \right] \quad \text{und} \quad S_2[q_2] = \int_{\beta'\hbar}^{\beta\hbar} dt \left[\frac{1}{2} m \dot{q}_2^2(t) + v(q_2(t)) \right] \quad (6.61)$$

führt auf die gewohnten Bewegungsgleichungen mit den Randbedingungen $q_1(0) = x_i$, $q_1(\beta'\hbar) = x'$, $q_2(\beta'\hbar) = x'$ und $q_2(\beta\hbar) = x_i$. Die Ableitung nach β' liefert die Bedingung:

$$0 = \frac{\partial}{\partial \beta'} (S_1[q_1] + S_2[q_2]) = \frac{1}{2} m \dot{q}_1^2(\beta'\hbar) + v(q_1(\beta'\hbar)) - \frac{1}{2} m \dot{q}_2^2(\beta'\hbar) - v(q_2(\beta'\hbar)). \quad (6.62)$$

Mit der Randbedingung $v(q_1(\beta'\hbar)) = v(q_2(\beta'\hbar))$ folgt somit, daß sich die Beträge der Geschwindigkeiten entsprechen: $\dot{q}_1(\beta'\hbar) = \pm \dot{q}_2(\beta'\hbar)$. Damit fallen auch die beiden Energien $E_j = \frac{1}{2} m \dot{q}_j^2(t) - v(q_j(t))$ für $j \in \{1, 2\}$ der Pfade zusammen.

Durch die Quotientenbildung kommt man zu dem Ergebnis:

$$\frac{\delta_{\delta v(x')}}{\delta v(x')} \frac{< x_i | e^{-\beta H} | x_i >}{< x_i | e^{-\beta H} | x_i >} = - \frac{A_1 A_2}{A_{x_i}} e^{-\frac{1}{\hbar}(S_1[q_1] + S_2[q_2] - S[q_{x_i}])}, \quad (6.63)$$

wobei q_{x_i} für den Pfad aus (6.8) mit $q_{x_i}(0) = q_{x_i}(\beta\hbar) = x_i$ und A_j für die jeweiligen Fluktuationen um die Mean–Field–Beiträge entsprechend (6.33) stehen soll.

6.5 Numerische Implementierung

Abschließend wird eine mögliche numerische Implementierung der Theorie dieses Kapitels vorgestellt. Das Beispiel beschränkt sich auf die erste Variante aus Abschnitt 6.1 und hat einzig und allein den Beweis der numerischen Anwendbarkeit zum Ziel.

Wesentlicher Bestandteil ist die Lösung der Bewegungsgleichungen (6.8) beziehungsweise (6.29). Für ein beliebiges, gegebenes Potential $v(x)$ ist man auf eine Diskretisierung des Zeitintervalls $[0, \beta\hbar]$ angewiesen. Bei einer Aufteilung in n_t äquidistante Stützstellen und $\hbar = 1$ ergibt sich ein Abstand von $\epsilon = \frac{\beta}{n_t}$. Die Pfade werden durch einen Vektor mit den Komponenten $q_k = q(t_k)$ bei $t_k = \epsilon k$ mit $k \in \{0, \dots, n_t\}$ codiert. Die Bewegungsgleichungen (6.29) bekommen dann in der diskreten Version die Form

$$0 = -\frac{m}{\epsilon^2} (q_{k+1} - 2q_k + q_{k-1}) + v'(q_k) \quad k = 1, \dots, n_t - 1 \quad . \quad (6.64)$$

Zusätzlich müssen die Randbedingungen $q_0 = x$ und $q_{n_t} = x$ erfüllt werden. Zusammen lässt sich dies durch die Matrixgleichung

$$0 = -\frac{m}{\epsilon^2} \begin{pmatrix} \frac{\epsilon^2}{m} & 0 & \cdots & 0 & \cdots & 0 \\ 1 & -2 & 1 & 0 & & 0 \\ 0 & 1 & -2 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \\ & & & \ddots & \ddots & 0 \\ 0 & & \cdots & 0 & 1 & -2 \\ 0 & \cdots & & \cdots & 0 & \frac{\epsilon^2}{m} \end{pmatrix} \begin{pmatrix} q_0 \\ q_1 \\ \vdots \\ q_{n_t-1} \\ q_{n_t} \end{pmatrix} + \begin{pmatrix} x \\ v'(q_1) \\ \vdots \\ v'(q_{n_t-1}) \\ x \end{pmatrix} \equiv \mathcal{A}\vec{q} + \vec{t}(\vec{q}) \quad (6.65)$$

ausdrücken. Wie gewohnt muß die Gleichung (6.65) iterativ selbstkonsistent gelöst werden. Dabei wurde im unteren Beispiel der Algorithmus

$$\vec{q}^{(j+1)} = \vec{q}^{(j)} - \eta_q \left(\vec{q}^{(j)} + \mathcal{A}^{-1} \vec{t}(\vec{q}^{(j)}) \right) \quad (6.66)$$

gewählt, was dem Verfahren (1.36) mit $\mathcal{O} = \mathcal{A}^{-1}$ entspricht.

Für das eigentliche Problem des Lernens – der Suche nach Lösungen $v(x)$ der Stationaritätsgleichung (6.10) – ist ebenfalls eine Diskretisierung der Ortskoordinate erforderlich. Der Wertebereich der Pfade kann mit dem ganzzahligen Definitionsbereich des Potentials in Einklang gebracht werden, indem die Funktionswerte der Pfade gerundet werden oder das Potential zwischen den Stützstellen linear interpoliert wird.

In einem Beispiel wurde ein Potential der Form

$$v(x) = -\frac{1}{1 + e^{\frac{|x-15|-4}{2}}} \quad (6.67)$$

auf einem Gitter der Größe $n_x = 30$ vorgegeben. Für dieses Potential wurden mittels (6.66) die Pfade $q_x(t)$ für alle 30 möglichen Randbedingungen $q_x(0) = q_x(\beta\hbar) = x$ des Ortsgitters bestimmt ($\beta = 6$, $\hbar = 1$). Für die Gittergröße der Zeitkoordinate wurde $n_t = 30$ gewählt. Im Diagramm 6.3 sind die klassischen (nach (6.43) berechneten) Ortsmeßwahrscheinlichkeiten, die quantenmechanischen (nach (5.4) berechneten) Wahrscheinlichkeiten und die semiklassischen (mittels (6.45) bestimmten) Wahrscheinlichkeiten aufgetragen worden. Es zeigt sich, daß beim Übergang von klassischer zu quantenmechanischer Berechnung die Verteilungen aufgrund der Unschärferelation erwartungsgemäß eine größere Varianz aufweisen. Es lohnt sich die Bemerkung, daß bei einer

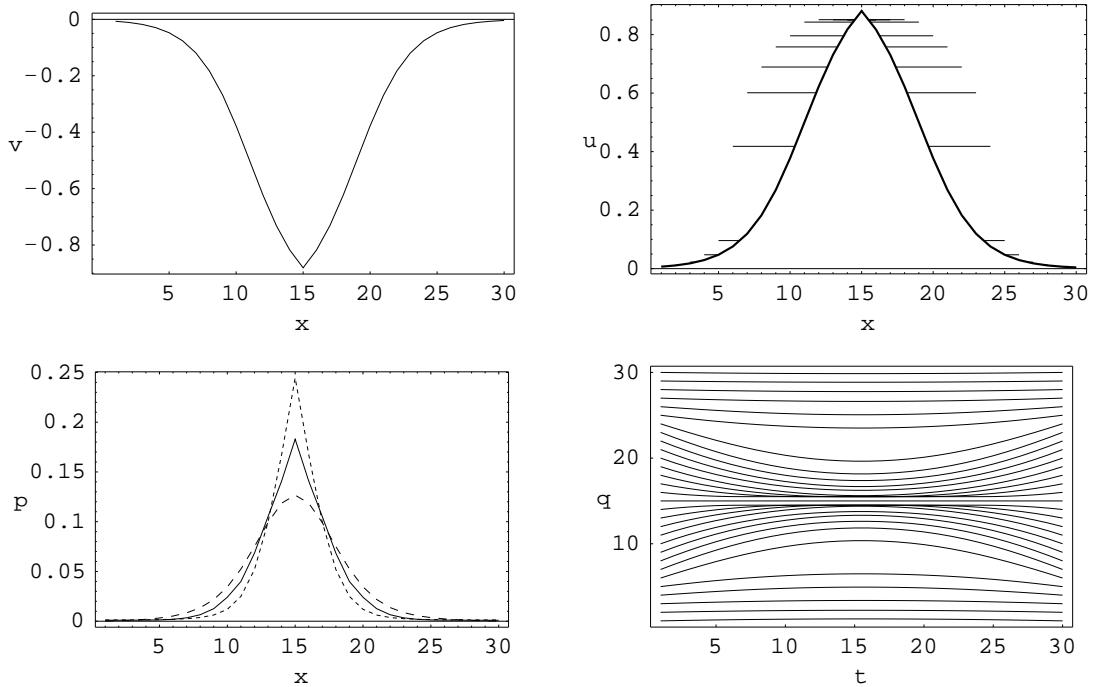


Abbildung 6.3: Vergleich von klassischer, semiklassischer und quantenmechanischer Meßwahrscheinlichkeit. Oben links: vorgegebenes Potential $v(x)$. Oben rechts: Potential $u(x) = -v(x)$, welches für die Lösung der Bewegungsgleichungen (6.29) entscheidend ist. Die dünnen, horizontalen Linien kennzeichnen den Ortsbereich und die Energien der einzelnen Pfade q_x . Unten links: klassische (feingestrichelte Linie), quantenmechanische (grobgestrichelte Linie) und semiklassische (durchgezogene Linie) Ortsmeßwahrscheinlichkeiten. Unten rechts: Darstellung der Pfade q_x . Die Parameter sind: $m = 0.1$, $\beta = 6$, $n_t = 30$, $n_x = 30$.

Erhöhung der Temperatur oder bei einer Erhöhung der Masse die Meßwahrscheinlichkeiten zusammenfallen und die drei Kurven übereinanderliegen.

Nachdem diese Voraussetzungen für das Lösen der Bewegungsgleichungen geschaffen wurden, kann nun nach numerischen Lösungen der Stationaritätsgleichung (6.10) gesucht werden. Die Auswertung des ersten Summanden mittels (6.11) kann bei der Betrachtung einer eindimensionalen Variablen x nicht verwendet werden, da auf allen eindimensionalen, periodischen Pfaden zu mindestens einer Zeit die Geschwindigkeit Null wird. Für diese Zeit liegt keine einfache Nullstelle im Argument der δ -Funktion vor. Bei der numerischen Behandlung wurden im folgenden Beispiel für alle diskreten Zeiten die Funktionswerte der Pfade $q_i(t_j) = q_{ij}$ gerundet und die Bedingung mittels $\int dt \delta(q_i(t) - x') \rightarrow \epsilon \sum_{j=1}^{n_t} \delta_{q_{ij}, x'}$ mit ganzzahligen Orten $x' \in [1, n_x]$ ausgewertet. Die Matrixelemente $\langle x' | e^{-\beta H} | x' \rangle$ des zweiten Summanden aus (6.10) wurden semiklassisch durch (6.45) berechnet. Als Bestrafungsfunktion wurde erneut eine Glattheitsforderung für das Potential $v(x_j) = v_j$ durch $\Gamma[v] = -\frac{1}{2} \sum_{ij=1}^{n_x} v_i \Delta_{ij} v_j$ mit Δ aus (5.31) gewählt. In der Anwendung wurden $N = 15$ Daten aus der diskreten Version des Potentials ($n_x = 30$)

$$v_D(x) = \begin{cases} \frac{1}{4} [\cos(\frac{2\pi}{10}(x - 15)) - 1] & \text{für } x \in [5, 25] \\ 0 & \text{für sonst} \end{cases} \quad (6.68)$$

quantenmechanisch zu vorgegebener Temperatur $\beta = 10$ gewonnen. Die Ergebnisse einer Iteration für (6.10) mit dem Gradientenabstiegsverfahren (1.34) sind in Abbildung 6.4 dargestellt. Dabei wurde der Hyperparameter γ so gewählt, daß die Tiefen der Potentiale näherungsweise zusammenfallen. Es zeigt sich, daß im Falle hinreichend großer Massen die grobe Form des Originalpotentials erkannt werden kann. Die semiklassische Dichte und die exakte quantenmechanische Dichte fallen näherungsweise zusammen.

In Abbildung 6.5 ist vorgeführt, was bei einer Verkleinerung der Masse passiert. Damit weiter getrennt aufgelöste Maxima der quantenmechanischen, datenerzeugenden Verteilung des Originalpotentials vorliegen, wurde die Potentialtiefe in (6.68) vergrößert. Die Doppelmuldenform des Originalpotentials wurde somit erneut erkannt, wobei aber zu beachten ist, daß die auf drei verschiedene Arten berechneten Dichteverteilungen bemerkenswert voneinander abweichen. Eine Verkleinerung des Hyperparameters γ hat eine Vertiefung der Lösung v zur Folge, verzerrt allerdings die Form der Doppelmulde eindeutig.

Insgesamt kann man sagen, daß sich für hinreichend große Massen die Verwendung von semiklassischer beziehungsweise klassischer Rechnung anbietet, falls man nicht zu große Zeitgitter verwendet. Denn der hauptsächliche numerische Aufwand liegt in der Bestimmung der Lösung der Bewegungsgleichungen und ist damit durch die Größe des Zeitgitters vorgegeben. Eine Vergrößerung des Ortsgitters hat im Gegensatz zu Kapitel 5 – wo das Ortsgitter die Größe des Hilbert–Raums vorgibt – geringere Konsequenzen für die Rechnungsdauer.

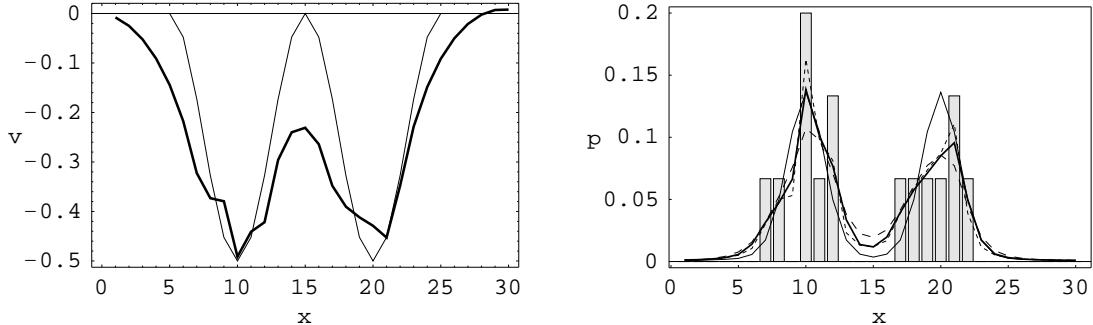


Abbildung 6.4: Lernen von Potentialen mittels Pfadintegralmethoden. Links sind Originalpotential (6.68) (dünne Linie) und die Lösungen von (6.10) (dicke Linie) dargestellt. Rechts sind die empirische Häufigkeit der Daten (Balkendiagramm), die Likelihood-Dichte des Originalpotentials (dünne Linie) und die Likelihood-Dichte des rekonstruierten Potentials v aufgetragen. Hierbei ist die semiklassisch berechnete Dichte, bezüglich der iteriert wurde, durch die dicke Linie dargestellt. Die klassisch berechnete Dichte für v ist als dünngestrichelte Linie, die quantenmechanisch berechnete Dichte als grobgestrichelte Linie gezeichnet. Die Parameter sind: $\beta = 10$, $m = 1$, $\gamma = 5$, $N = 15$.

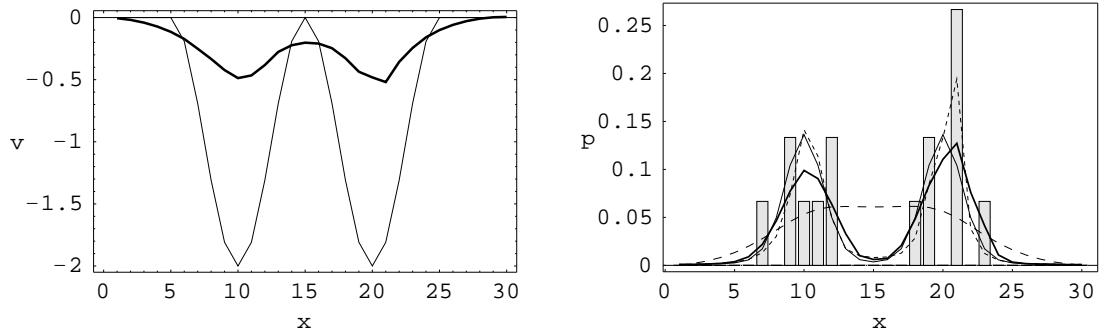


Abbildung 6.5: Lernen von Potentialen mittels Pfadintegralmethoden im Falle sehr kleiner Massen m . Die Bedeutung der Graphen ist analog zu Abbildung 6.4 gewählt. Die Parameter sind: $\beta = 10$, $m = 0.05$, $\gamma = 10$, $N = 15$.

Kapitel 7

Lernen von Zweiteilchenwechselwirkungen

Ziel dieses Kapitels ist, mittels empirischer Meßdaten Aussagen über das zugrundeliegende Potential einer Zweiteilchenwechselwirkung zu machen. Die Wechselwirkungen werden im Sinne effektiver Wechselwirkungen aufgefaßt, da nur diese sich in den Daten widerspiegeln. Das Lernen von Zweiteilchenwechselwirkungen bedeutet im Vergleich zu Kapitel 5 einen eindeutigen Generalisierungsschritt. Die Anzahl der zu betrachtenden Freiheitsgrade ist bei diesem inversen Vielteilchenproblem hoch, da der Hamilton–Operator und der statistische Operator von sämtlichen Koordinaten der auftretenden Teilchen abhängen. Hinzu kommt, daß eine allgemeine, nichtlokale Zweiteilchenwechselwirkung zusätzliche Freiheitsgrade mit sich bringt. Im Falle von ununterscheidbaren, fermionischen Teilchen muß die Antisymmetrie der Gesamtwellenfunktion in der Rechnung berücksichtigt werden. Eine exakte Behandlung des Hamilton–Operators ist für ein allgemeines Vielteilchenpotential und damit für die erforderliche Berechnung der Likelihood–Dichte nicht möglich. Somit wird man im Laufe der Rechnung auf Näherungen angewiesen sein. Diesbezüglich wird hier die Hartree–Fock–Methode als fundamentale Vielteilchennäherung angewandt. Um die Unabhängigkeit der Messungen zu gewährleisten, werden erneut Messungen an einem kanonischen Ensemble vorgenommen. Im vorliegenden Fall müssen dabei Aussagen über die Art der Messung gemacht werden: So besteht die Möglichkeit, daß bei jedem Meßvorgang alle anwesenden Teilchen lokalisiert oder nur jeweils die Koordinaten einzelner Teilchen bestimmt werden.

In Abschnitt 7.1 wird der Einstieg anhand der Betrachtung eines kanonischen Ensembles für den Fall $T = 0$ geschaffen, indem statt des vollen Vielteilchen–Hamilton–Operators der Hartree–Fock–Operator als effektiver Einteilchenoperator verwendet wird. Das weitere Vorgehen ist dem von Kapitel 5 analog. In Abschnitt 7.2 erfolgt eine Generalisierung auf den Fall $T \neq 0$ für ein großkanonisches Ensemble. Abschnitt 7.3 stellt eine Alternative zu 7.2 dar. Es wird mittels Hilfsintegralmethoden der statistische Operator des Vielteilchensystems auf eine Summe von zeitabhängigen Einteilchenoperatoren abgebildet. Anschließend wird eine gleichzeitige Sattelpunktnäherung für das Hilfsintegral und eine Maximum–Aposteriori–Approximation vorgenommen.

7.1 Kanonisches Ensemble bei $T = 0$

In diesem Abschnitt wird das Bayessche Lernen für eine Zweiteilchenwechselwirkung durch Messung an einem kanonischen Ensemble bei Temperatur $T = 0$ angewandt. Es wird ein fermionisches s –Teilchensystem betrachtet. Die Teilchenzahl s ist dabei bekannt und fixiert. Es wird davon ausgegangen, daß bei jeder Messung die Orte aller s Teilchen bestimmt werden können. Aufgrund

der Ununterscheidbarkeit der identischen Teilchen kann das Ergebnis jeder Messung mit einer s -Teilchen Slaterdeterminante $|x_1, \dots, x_s\rangle$ angegeben werden. Die Gesamtheit der Daten aus N Messungen besteht dann aus der Menge $D = \{|x_1^i, \dots, x_s^i\rangle | i = 1, \dots, N\}$.

Aufgrund der Notwendigkeit einer antisymmetrisierten Rechnung bietet sich die Darstellung der zweiten Quantisierung an. Ein Hamilton-Operator mit Anteilen aus kinetischer und potentieller Energie einer Zweiteilchenwechselwirkung bekommt dann in willkürlicher Basis die Form

$$H = \sum_{\alpha\gamma} t_{\alpha\gamma} a_{\alpha}^{\dagger} a_{\gamma} + \frac{1}{2} \sum_{\alpha\beta\gamma\delta} v_{\alpha\beta\gamma\delta} a_{\alpha}^{\dagger} a_{\beta}^{\dagger} a_{\delta} a_{\gamma}. \quad (7.1)$$

Dabei sind die Matrixelemente $v_{\alpha\beta\gamma\delta} = (\alpha\beta | v | \gamma\delta)$ nichtantisymmetrisch, und die Erzeugungs- und Vernichtungsoperatoren erfüllen die fermionischen Vertauschungsrelationen. Eine grundlegende Voraussetzung für das Arbeiten in der zweiten Quantisierung ist, daß die Wechselwirkung symmetrisch unter Vertauschung der Teilchen ist. Dies entspricht der erforderlichen Symmetrie der Matrixelemente

$$v_{\alpha\beta\gamma\delta} = v_{\beta\alpha\delta\gamma} \quad (7.2)$$

und muß explizit durch die Apriori-Wahrscheinlichkeit $p[v]$ gewährleistet werden. Oft ist zusätzlich bekannt, daß es sich bei v um ein lokales Potential handelt. Das heißt, v ist in der Ortsdarstellung diagonal $(x_1 x_2 | v | x_3 x_4) = \delta(x_1 - x_3) \delta(x_2 - x_4) v(x_1, x_2)$, und für die Matrixelemente gilt:

$$v_{\alpha\beta\gamma\delta} = (\alpha\beta | v | \gamma\delta) = \int dx dx' \varphi_{\alpha}^*(x) \varphi_{\beta}^*(x') v(x, x') \varphi_{\gamma}(x) \varphi_{\delta}(x'). \quad (7.3)$$

Man beachte, daß durch dieses Wissen die Zahl der Freiheitsgrade für das Potential halbiert wird. Im weiteren Verlauf der Rechnung wird diese Einschränkung nicht gemacht. Es werden auch nicht-lokale Potentiale zugelassen.

Für die Likelihood-Wahrscheinlichkeit, den Zustand $|x_1^i, \dots, x_s^i\rangle$ in einem kanonischen Ensemble im Tieftemperaturfall $\beta \rightarrow \infty$ zu messen, gilt als Verallgemeinerung von (5.4) und (5.24)

$$p[x_1^i, \dots, x_s^i | \hat{x}, \rho] = \langle x_1^i, \dots, x_s^i | \frac{e^{-\beta H}}{Z} | x_1^i, \dots, x_s^i \rangle \longrightarrow \langle x_1^i, \dots, x_s^i | \phi_0 \rangle \langle \phi_0 | x_1^i, \dots, x_s^i \rangle. \quad (7.4)$$

Hierbei ist die Zustandssumme Z eine Spur im s -Teilchenraum, und $|\phi_0\rangle$ steht für die normierte s -Teilchen Slaterdeterminante des Grundzustands von H . Die Bestimmung von $|\phi_0\rangle$ entspricht dem Lösen einer Vielteilchen-Schrödinger-Gleichung und ist nur in den seltensten Fällen direkt durchführbar. Hierin liegt die Motivation, den Hamilton-Operator H durch den Hartree-Fock-Operator des Grundzustands

$$H \longrightarrow h = \sum_{\alpha\gamma} a_{\alpha}^{\dagger} a_{\gamma} \left(t_{\alpha\gamma} + \sum_{n=1}^s \bar{v}_{\alpha\varphi_n\gamma\varphi_n} \right) = \sum_{\alpha\gamma} a_{\alpha}^{\dagger} a_{\gamma} \left(t_{\alpha\gamma} + \sum_{n=1}^s \sum_{\beta\delta} \bar{v}_{\alpha\beta\gamma\delta} \langle \varphi_n | \beta \rangle \langle \delta | \varphi_n \rangle \right) \quad (7.5)$$

zu ersetzen. Dabei sind mit

$$\bar{v}_{\alpha\varphi_n\gamma\varphi_n} = \langle \alpha\varphi_n | v | \gamma\varphi_n \rangle = (\alpha\varphi_n | v | \gamma\varphi_n) - (\alpha\varphi_n | v | \varphi_n\gamma) \quad (7.6)$$

antisymmetrische Matrixelemente abgekürzt, und die in (7.5) auftretenden Hartree-Fock-Orbitale werden durch die Einteilchengleichungen

$$h | \varphi_{\nu} \rangle = \varepsilon_{\nu} | \varphi_{\nu} \rangle \quad \nu = 1, \dots, s \quad (7.7)$$

mit den s niedrigsten Eigenenergien ε_{ν} festgelegt. Die Hartree-Fock-Gleichungen (7.5) und (7.7) müssen gleichzeitig selbstkonsistent gelöst werden. Der genäherte Grundzustand ist dann durch

$|\phi_0\rangle \rightarrow |\varphi_1, \dots, \varphi_s\rangle$ gegeben.

Bei der Apriori-Wahrscheinlichkeit für das Potential muß die Symmetrieforderung (7.2) berücksichtigt werden. Aus diesem Grund wird von einer Verteilung der Form

$$p[v] \sim e^{-\gamma\Gamma[v]} \prod_{\alpha\beta\gamma\delta} \delta(v_{\alpha\beta\gamma\delta} - v_{\beta\alpha\delta\gamma}) \quad (7.8)$$

ausgegangen. Nach der gewohnten Einführung der Integraldarstellung der δ -Distribution

$$\prod_{\alpha\beta\gamma\delta} \delta(v_{\alpha\beta\gamma\delta} - v_{\beta\alpha\delta\gamma}) = \int \left(\prod_{\alpha\beta\gamma\delta} \frac{d\lambda_{\alpha\beta\gamma\delta}}{2\pi} \right) e^{i \sum_{\alpha\beta\gamma\delta} \lambda_{\alpha\beta\gamma\delta} (v_{\alpha\beta\gamma\delta} - v_{\beta\alpha\delta\gamma})} \quad (7.9)$$

kann die Aposteriori-Wahrscheinlichkeit des Potentials mittels

$$p[v|D] \sim \int \left(\prod_{\alpha\beta\gamma\delta} d\lambda_{\alpha\beta\gamma\delta} \right) e^{-F[\lambda, v]} \quad (7.10)$$

ausgedrückt werden. Dabei ist das Energiefunktional durch

$$\begin{aligned} F[\lambda, v] &= \gamma\Gamma[v] - i \sum_{\alpha\beta\gamma\delta} \lambda_{\alpha\beta\gamma\delta} (v_{\alpha\beta\gamma\delta} - v_{\beta\alpha\delta\gamma}) \\ &\quad - \sum_{i=1}^N \left[\ln \langle x_1^i, \dots, x_s^i | \varphi_1, \dots, \varphi_s \rangle + \ln \langle \varphi_1, \dots, \varphi_s | x_1^i, \dots, x_s^i \rangle \right] \end{aligned} \quad (7.11)$$

gegeben. Bei der anschließenden Maximum-Aposteriori-Approximation für (7.10) ist man auf die Ableitung der Slaterdeterminante nach den Matrixelementen von v angewiesen. Nach Verwendung des Antisymmetrisierers $\mathcal{A} = \frac{1}{\sqrt{s!}} \sum_{\mathcal{P}} (-1)^{\mathcal{P}}$, Ausnutzen des Kommutators $[\frac{\delta}{\delta v}, \mathcal{P}] = 0$ und Ableiten nach der Produktregel erhält man:

$$\begin{aligned} \frac{\delta}{\delta v} |\varphi_1, \dots, \varphi_s\rangle &= \frac{\delta}{\delta v} \mathcal{A} \prod_{\nu=1}^s |\varphi_\nu\rangle = \frac{\delta}{\delta v} \frac{1}{\sqrt{s!}} \sum_{\mathcal{P}} (-1)^{\mathcal{P}} \mathcal{P} [\varphi_1, \dots, \varphi_s] \\ &= \frac{1}{\sqrt{s!}} \sum_{\mathcal{P}} (-1)^{\mathcal{P}} \frac{\delta}{\delta v} \mathcal{P} [\varphi_1, \dots, \varphi_s] = \mathcal{A} \frac{\delta}{\delta v} \prod_{\nu=1}^s |\varphi_\nu\rangle = \mathcal{A} \sum_{\nu=1}^s \left(\prod_{n \neq \nu}^s |\varphi_n\rangle \right) \left| \frac{\delta \varphi_\nu}{\delta v} \right\rangle \\ &= \sum_{\nu=1}^s \mathcal{A} \left(\prod_{n \neq \nu}^s |\varphi_n\rangle \right) \left| \frac{\delta \varphi_\nu}{\delta v} \right\rangle = \sum_{\nu=1}^s |\varphi_1, \dots, \frac{\delta \varphi_\nu}{\delta v}, \dots, \varphi_s\rangle . \end{aligned} \quad (7.12)$$

Nachdem diese Voraussetzungen geschaffen wurden, kann durch Ableitung nach den Matrixelementen von v die Stationaritätsbedingung gewonnen werden:

$$0 = \gamma \frac{\delta\Gamma[v]}{\delta v_{1234}} - 2 \operatorname{Re} \left[\sum_{i=1}^N \sum_{\nu=1}^s \frac{\langle x_1^i, \dots, x_s^i | \varphi_1, \dots, \frac{\delta \varphi_\nu}{\delta v_{1234}}, \dots, \varphi_s \rangle}{\langle x_1^i, \dots, x_s^i | \varphi_1, \dots, \varphi_s \rangle} \right] - i\lambda_{1234} + i\lambda_{2143} . \quad (7.13)$$

Die Abhängigkeit von den Lagrange-Parametern kann aufgehoben werden, indem man Gleichung (7.13) mit der Gleichung addiert, die durch gleichzeitigen Austausch von $1 \leftrightarrow 2$ und $3 \leftrightarrow 4$ entsteht. Unter Ausnutzung der Determinanteneigenschaften erhält man mit

$$0 = \gamma \left(\frac{\delta\Gamma[v]}{\delta v_{1234}} + \frac{\delta\Gamma[v]}{\delta v_{2143}} \right) - 2 \operatorname{Re} \left[\sum_{i=1}^N \sum_{\nu=1}^s \frac{\langle x_1^i, \dots, x_s^i | \varphi_1, \dots, \left(\frac{\delta \varphi_\nu}{\delta v_{1234}} + \frac{\delta \varphi_\nu}{\delta v_{2143}} \right), \dots, \varphi_s \rangle}{\langle x_1^i, \dots, x_s^i | \varphi_1, \dots, \varphi_s \rangle} \right] \quad (7.14)$$

eine symmetrierte Stationaritätsgleichung.

Das Ergebnis aus (7.14) läßt sich kompakter ausdrücken, indem man die Überlappmatrizen M_i von Datenvektor und besetzten Orbitalen

$$M_{\nu n}^i = \langle x_\nu^i | \varphi_n \rangle \quad i = \{1, \dots, N\} \quad \nu, n = \{1, \dots, s\} \quad (7.15)$$

und deren Adjunkten \tilde{M}_i

$$\tilde{M}_i = \left(\det M_i \right) M_i^{-1} \quad , \quad \tilde{M}_{\nu n}^i = (-1)^{\nu+n} |M_{\nu n}^i| \quad (7.16)$$

einführt. Hierbei wird mit $|M_{\nu n}^i|$ die Determinante der Matrix abgekürzt, die durch Streichen der n -ten Zeile und ν -ten Spalte von M_i entsteht. Der aus zwei Slaterdeterminanten bestehende Quotient aus (7.14) läßt sich dann mittels

$$\frac{\langle x_1^i, \dots, x_s^i | \varphi_1, \dots, \left(\frac{\delta \varphi_\nu}{\delta v_{1234}} + \frac{\delta \varphi_\nu}{\delta v_{2143}} \right), \dots, \varphi_s \rangle}{\langle x_1^i, \dots, x_s^i | \varphi_1, \dots, \varphi_s \rangle} = \sum_{n=1}^s \frac{(-1)^{\nu+n}}{\det M_i} \langle x_n^i | \frac{\delta \varphi_\nu}{\delta v_{1234}} + \frac{\delta \varphi_\nu}{\delta v_{2143}} \rangle |M_{\nu n}^i| = \sum_{n=1}^s \left(M_i^{-1} \right)_{\nu n} \langle x_n^i | \frac{\delta \varphi_\nu}{\delta v_{1234}} + \frac{\delta \varphi_\nu}{\delta v_{2143}} \rangle \quad (7.17)$$

umformen, und man erhält als vorläufiges Endergebnis die Stationaritätsgleichung

$$0 = \gamma \left(\frac{\delta \Gamma[v]}{\delta v_{1234}} + \frac{\delta \Gamma[v]}{\delta v_{2143}} \right) - 2 \operatorname{Re} \left[\sum_{i=1}^N \sum_{\nu=1}^s \sum_{n=1}^s \left(M_i^{-1} \right)_{\nu n} \langle x_n^i | \frac{\delta \varphi_\nu}{\delta v_{1234}} + \frac{\delta \varphi_\nu}{\delta v_{2143}} \rangle \right]. \quad (7.18)$$

Es fehlt für die Auswertung von (7.18) letztlich die Bestimmung der Orbitalableitungen $| \frac{\delta \varphi_\nu}{\delta v_{1234}} |$. Diese erfolgt analog zum Vorgehen aus Kapitel 5 – mit dem Unterschied, daß die Abhängigkeit des Hartree–Fock–Operators (7.5) von den Hartree–Fock–Orbitalen berücksichtigt werden muß. Die Ableitung der Eigenwertgleichung (7.7) liefert:

$$\frac{\delta h}{\delta v} | \varphi_\nu \rangle + h | \frac{\delta \varphi_\nu}{\delta v} \rangle = \frac{\delta \varepsilon_\nu}{\delta v} | \varphi_\nu \rangle + \varepsilon_\nu | \frac{\delta \varphi_\nu}{\delta v} \rangle. \quad (7.19)$$

Setzt man zusätzlich die Normierung der Hartree–Fock–Orbitale ($\langle \varphi_\nu | \varphi_\nu \rangle = 1$) voraus, so erhält man mit

$$0 = \langle \frac{\delta \varphi_\nu}{\delta v} | \varphi_\nu \rangle + \langle \varphi_\nu | \frac{\delta \varphi_\nu}{\delta v} \rangle \quad (7.20)$$

eine Beziehung für die Überlappe der Orbitale und deren Ableitung. Nach Auflösung von (7.19)

$$\left(h - \varepsilon_\nu \right) | \frac{\delta \varphi_\nu}{\delta v_{1234}} \rangle = \left(\frac{\delta \varepsilon_\nu}{\delta v_{1234}} - \frac{\delta h}{\delta v_{1234}} \right) | \varphi_\nu \rangle \quad (7.21)$$

und der Pseudoinversion von $h - \varepsilon_\nu$ durch das Vorgehen analog zu Anhang C bekommt man

$$| \frac{\delta \varphi_\nu}{\delta v_{1234}} \rangle = | \varphi_\nu \rangle \langle \varphi_\nu | \lambda_\nu \rangle - \sum_{k \neq \nu} \frac{1}{\varepsilon_k - \varepsilon_\nu} | \varphi_k \rangle \langle \varphi_k | \frac{\delta h}{\delta v_{1234}} | \varphi_\nu \rangle \quad (7.22)$$

und die entsprechend adjungierte Gleichung. Außerdem muß die Bedingung (7.20) erfüllt werden. Deshalb kann ohne Beschränkung der Allgemeinheit $| \lambda_\nu \rangle = 0$ gewählt werden. Durch Einsetzen von (7.22) in den zweiten Summanden von (7.18) kann gezeigt werden, daß von der k –Summe effektiv nur die nichtbesetzten Orbitale zur Stationaritätsgleichung beitragen. Denn nach Aufspaltung

der k -Summe in Loch- und Teilchenorbitale und unter Ausnutzung der Definition (7.15) erhält man

$$\begin{aligned}
 & \sum_{\nu=1}^s \sum_{n=1}^s \left(M_i^{-1} \right)_{\nu n} \langle x_n^i | \frac{\delta \varphi_\nu}{\delta v_{1234}} + \frac{\delta \varphi_\nu}{\delta v_{2143}} \rangle = \sum_{\nu=1}^s \sum_{n=1}^s \sum_{k \neq \nu} \left(M_i^{-1} \right)_{\nu n} \langle x_n^i | \varphi_k \rangle A_{k\nu} \\
 & = \sum_{\nu=1}^s \sum_{n=1}^s \sum_{k \neq \nu} \left(M_i^{-1} \right)_{\nu n} M_{nk}^i A_{k\nu} + \sum_{\nu=1}^s \sum_{n=1}^s \sum_{k > s} \left(M_i^{-1} \right)_{\nu n} \langle x_n^i | \varphi_k \rangle A_{k\nu} \\
 & = \sum_{\nu=1}^s \sum_{n=1}^s \sum_{k > s} \left(M_i^{-1} \right)_{\nu n} \langle x_n^i | \varphi_k \rangle A_{k\nu}. \tag{7.23}
 \end{aligned}$$

Dabei wurde die Abkürzung

$$A_{k\nu} = \frac{1}{\varepsilon_k - \varepsilon_\nu} \langle \varphi_k | \frac{\delta h}{\delta v_{1234}} + \frac{\delta h}{\delta v_{2143}} | \varphi_\nu \rangle, \quad k \neq \nu \tag{7.24}$$

verwendet. Die Ableitung des Hartree–Fock–Operators kann aus (7.5) gewonnen werden. Dies führt auf das Ergebnis

$$\begin{aligned}
 \frac{\delta h}{\delta v_{1234}} &= \sum_{n=1}^s \langle \varphi_n | 2 \rangle \langle 4 | \varphi_n \rangle a_1^\dagger a_3 - \sum_{n=1}^s \langle \varphi_n | 2 \rangle \langle 3 | \varphi_n \rangle a_1^\dagger a_4 \\
 &+ \sum_{\alpha\gamma} \sum_{n=1}^s a_\alpha^\dagger a_\gamma \left[\langle \alpha \frac{\delta \varphi_n}{\delta v_{1234}} | v | \gamma \varphi_n \rangle + \langle \alpha \varphi_n | v | \gamma \frac{\delta \varphi_n}{\delta v_{1234}} \rangle \right]. \tag{7.25}
 \end{aligned}$$

Mit Hilfe von (7.25) können nun die Ableitungen der Hartree–Fock–Energien berechnet werden. Durch Multiplikation von (7.19) mit $\langle \varphi_\nu |$ gewinnt man für die Variationen

$$\begin{aligned}
 \frac{\delta \varepsilon_\nu}{\delta v_{1234}} &= \langle \varphi_\nu | \frac{\delta h}{\delta v_{1234}} | \varphi_\nu \rangle \\
 &= \sum_{n=1}^s \langle \varphi_\nu | 1 \rangle \langle \varphi_n | 2 \rangle \left[\langle 3 | \varphi_\nu \rangle \langle 4 | \varphi_n \rangle - \langle 4 | \varphi_\nu \rangle \langle 3 | \varphi_n \rangle \right] \\
 &+ \sum_{n=1}^s \left[\langle \varphi_\nu \frac{\delta \varphi_n}{\delta v_{1234}} | v | \varphi_n \rangle + \langle \varphi_\nu \varphi_n | v | \varphi_\nu \frac{\delta \varphi_n}{\delta v_{1234}} \rangle \right]. \tag{7.26}
 \end{aligned}$$

Und unter Einsetzen von (7.25) in (7.22) erhält man schließlich für die Orbitalableitungen

$$\begin{aligned}
 \left| \frac{\delta \varphi_\nu}{\delta v_{1234}} \right\rangle &= - \sum_{k \neq \nu} \frac{1}{\varepsilon_k - \varepsilon_\nu} \langle \varphi_k \rangle \times \\
 &\left[\sum_{n=1}^s \langle \varphi_n | 2 \rangle \langle \varphi_k | 1 \rangle \left[\langle 4 | \varphi_n \rangle \langle 3 | \varphi_\nu \rangle - \langle 3 | \varphi_n \rangle \langle 4 | \varphi_\nu \rangle \right] \right. \\
 &\left. + \sum_{n=1}^s \left[\langle \varphi_k \frac{\delta \varphi_n}{\delta v_{1234}} | v | \varphi_n \rangle + \langle \varphi_k \varphi_n | v | \varphi_\nu \frac{\delta \varphi_n}{\delta v_{1234}} \rangle \right] \right]. \tag{7.27}
 \end{aligned}$$

Man beachte, daß die Gleichung (7.27) für alle besetzten Orbitale bestimmt werden muß. Die Indizes ν und n stehen für die s Lochraumorbitale, während der Index k über die volle Hartree–Fock–Basis läuft.

Die Aufgabe bei der Durchführung der Maximum–Aposteriori–Approximation besteht darin, eine Lösung v für die Stationaritätsgleichung (7.18) zu finden, wobei die Orbitale und deren Ableitungen durch die Gleichungen (7.7) und (7.27) gegeben sind. Diese hängen wiederum selbst von v ab. Man muß iterativ vorgehen, indem man mit einer Vermutung für das Potential v startet und hierfür die Hartee–Fock–Gleichung löst. Sind die Hartree–Fock–Orbitale bekannt, kann die Gleichung für die Orbitalableitungen (7.27) für sich selbstkonsistent gelöst werden. Danach kann mit Hilfe eines geeigneten Algorithmus mittels (7.18) ein neues Potential berechnet werden. Dieser Zyklus muß bis zur Konvergenz fortgesetzt werden.

Der Inhalt dieses Kapitels ist in [2] veröffentlicht worden. Der Artikel enthält zusätzlich eine von Lemm programmierte numerische Anwendung. Als einfachstes Beispiel wurde ein eindimensionales Zweiteilchenproblem gewählt und ein lokales Potential angenommen. Auf Anregung eines Gutachters wurde ebenfalls der Fall untersucht, daß bei jeder Messung jeweils nur eines der s Teilchen lokalisiert wurde, also eine Datenmenge $D = \{x^i | i = 1, \dots, N\}$ vorliegt. Der Informationsgehalt aus diesen Messungen ist weitaus geringer, da die Information über die Teilchenwechselwirkungen in den Korrelationen der Teilchenkoordinaten steckt. Analytisch müssen hierfür in der Likelihood–Dichte (7.4) die nicht gemessenen Koordinaten ausintegriert werden. Durch das Einführen der Slaterdeterminanten als Überlappmatrizen, das jeweilige Entwickeln nach der Zeile beziehungsweise der Spalte, in denen x^i auftritt, und Ausnutzen der Vollständigkeit erhält man:

$$\begin{aligned}
p[x^i | \hat{x}, \rho] &= \int dx_2 \cdots dx_s p[x^i, x_2, \dots, x_s | \hat{x}, \rho] \\
&\approx \int dx_2 \cdots dx_s \langle x^i, x_2, \dots, x_s | \varphi_1, \dots, \varphi_s \rangle \langle \varphi_1, \dots, \varphi_s | x^i, x_2, \dots, x_s \rangle \\
&= \sum_{n\nu=1}^s (-1)^{n+\nu} \langle x^i | \varphi_n \rangle \langle \varphi_\nu | x^i \rangle \times \\
&\quad \int dx_2 \cdots dx_s \underbrace{\langle \varphi_1, \dots, \varphi_s | x_2, \dots, x_s \rangle}_{\text{ohne } \varphi_\nu} \langle x_2, \dots, x_s | \underbrace{\varphi_1, \dots, \varphi_s}_{\text{ohne } \varphi_n} \rangle \\
&= \sum_{n\nu=1}^s (-1)^{n+\nu} \langle x^i | \varphi_n \rangle \langle \varphi_\nu | x^i \rangle \underbrace{\langle \varphi_1, \dots, \varphi_s |}_{\text{ohne } \varphi_\nu} \underbrace{\varphi_1, \dots, \varphi_s}_{\text{ohne } \varphi_n} \rangle^{=\delta_{n\nu}} \\
&= \sum_{\nu=1}^s \langle x^i | \varphi_\nu \rangle \langle \varphi_\nu | x^i \rangle. \tag{7.28}
\end{aligned}$$

Somit verändert sich die Wirkung (7.11) zu

$$F[\lambda, v] = \gamma \Gamma[v] - i \sum_{\alpha\beta\gamma\delta} \lambda_{\alpha\beta\gamma\delta} (v_{\alpha\beta\gamma\delta} - v_{\beta\alpha\delta\gamma}) - \sum_{i=1}^N \ln \left(\sum_{\nu=1}^s \langle x^i | \varphi_\nu \rangle \langle \varphi_\nu | x^i \rangle \right). \tag{7.29}$$

Eine Funktionalableitung im Rahmen einer Maximum–Aposteriori–Approximation liefert statt (7.18) das abgeänderte Ergebnis

$$0 = \gamma \left(\frac{\delta \Gamma[v]}{\delta v_{1234}} + \frac{\delta \Gamma[v]}{\delta v_{2143}} \right) - 2 \operatorname{Re} \left[\sum_{i=1}^N \frac{\sum_{\nu=1}^s \langle \varphi_\nu | x^i \rangle \langle x^i | \varphi_\nu \rangle}{\sum_{\nu=1}^s \langle x^i | \varphi_\nu \rangle \langle \varphi_\nu | x^i \rangle} \right], \tag{7.30}$$

wobei die Orbitalableitungen weiterhin durch (7.27) gegeben sind.

7.2 Großkanonisches Ensemble (Energiedarstellung)

Nun erfolgt eine Generalisierung der Situation von 7.1. Es wird der Fall eines großkanonischen Ensembles für eine beliebige Temperatur betrachtet. Dadurch ist die Teilchenzahl nicht mehr fixiert, sondern nur deren Erwartungswert. Es wird davon ausgegangen, daß bei jeder der N Messungen alle Teilchen des Ensembles lokalisiert werden können. Die Daten können über eine Menge von Slaterdeterminanten $D = \{|\chi^i\rangle = |x_1^i, \dots, x_{s_i}^i\rangle |i = 1, \dots, N\}$ mit variabler Teilchenzahl s_i angegeben werden. Das Prinzip der Rechnung bleibt aber im Vergleich zu Abschnitt 7.1 insgesamt erhalten.

Der statistische Operator für das großkanonische Ensemble ist durch

$$\mathcal{D} = \frac{1}{Z} e^{-\beta(H - \mu \hat{N})} \quad Z = \text{Tr} \left[e^{-\beta(H - \mu \hat{N})} \right] \quad (7.31)$$

gegeben. Dabei ist $\hat{N} = \sum_{\alpha} a_{\alpha}^{\dagger} a_{\alpha}$ der Teilchenzahloperator und μ das chemische Potential des Systems. Es ist zu beachten, daß die Spurbildung jetzt über den gesamten Fock–Raum auszuführen ist. In volliger Analogie zu dem vorherigen Abschnitt wird der exakte statistische Operator als Näherung durch die Hartree–Fock–Variante ersetzt:

$$\mathcal{D} \longrightarrow \frac{1}{Z_{HF}} e^{-\beta(h - \mu \hat{N})}. \quad (7.32)$$

Hierbei ist h jetzt der Hartree–Fock–Hamilton–Operator des großkanonischen Ensembles und selbstkonsistent über die Matrixelemente

$$\langle \alpha | h | \gamma \rangle = \langle \alpha | t | \gamma \rangle + \sum_{\beta \delta} \langle \alpha \beta | v | \gamma \delta \rangle \varrho_{\delta \beta} = t_{\alpha \gamma} + \sum_m \langle \alpha \varphi_m | v | \gamma \varphi_m \rangle n_m \quad (7.33)$$

definiert. Wie üblich wurde in (7.33) der Einteilchendichteoperator

$$\varrho = \frac{1}{1 + e^{\beta(h - \mu)}} \quad (7.34)$$

eingeführt. Dieser ist nicht mit dem statistischen Operator ρ der vorangegangenen Kapitel zu verwechseln. Die Orbitale φ_m sind mit

$$h | \varphi_m \rangle = \varepsilon_m | \varphi_m \rangle \quad (7.35)$$

die Eigenzustände von h und damit gleichzeitig Eigenzustände von ϱ :

$$\varrho | \varphi_m \rangle = n_m | \varphi_m \rangle \quad (7.36)$$

mit den Besetzungszahlen der Fermi–Verteilung

$$n_m = \frac{1}{1 + e^{\beta(\varepsilon_m - \mu)}}. \quad (7.37)$$

Der Operator (7.33) stellt eine Verallgemeinerung von (7.5) auf den Fall $T > 0$ dar. Die Summen aus (7.33) laufen über die gesamte Einteilchenbasis. Die Zustandssumme aus (7.32) nimmt im Hartree–Fock–Fall den Wert

$$Z_{HF} = \text{Tr} \left[e^{-\beta(h - \mu \hat{N})} \right] = \prod_m \left(1 + e^{-\beta(\varepsilon_m - \mu)} \right) \quad (7.38)$$

an. Der Vollständigkeit halber seien hier noch der Erwartungswert für eine Teilchenzahlmessung

$$\langle \hat{N} \rangle = \text{Tr} [\mathcal{D} \hat{N}] = \sum_m n_m = \sum_m \frac{1}{1 + e^{\beta(\varepsilon_m - \mu)}} \quad (7.39)$$

und für die Messung der Gesamtenergie

$$\langle H \rangle = \text{Tr} [\mathcal{D} H] = \sum_m \langle \varphi_m | t | \varphi_m \rangle n_m + \frac{1}{2} \sum_{mn} \langle \varphi_m \varphi_n | v | \varphi_m \varphi_n \rangle n_m n_n \quad (7.40)$$

angegeben.

Die Likelihood–Wahrscheinlichkeit für ein Datum χ^i berechnet sich durch das Matrixelement des statistischen Operators beziehungsweise als Erwartungswert für den Projektor auf den gemessenen Zustand:

$$\begin{aligned} \langle x_1^i, \dots, x_{s_i}^i | \frac{e^{-\beta(h - \mu \hat{N})}}{Z_{HF}} | x_1^i, \dots, x_{s_i}^i \rangle &= \text{Tr} \left[\frac{e^{-\beta(h - \mu \hat{N})}}{Z_{HF}} a_{x_1^i}^\dagger \cdots a_{x_{s_i}^i}^\dagger a_{x_{s_i}^i} \cdots a_{x_1^i} \right] \quad (7.41) \\ &= \det \mathcal{W}_i = e^{\text{tr}_i [\ln \mathcal{W}_i]}. \end{aligned}$$

Dabei lässt sich die Spurbildung für die Erwartungswertberechnung in (7.41) analytisch exakt ausführen [20]. Die in (7.41) eingeführte $s_i \times s_i$ –Matrix \mathcal{W}_i beinhaltet die Matrixelemente des Einteilchendichteoperators (7.34) für die gemessenen Eigenzustände:

$$\mathcal{W}_{pq}^i = \langle x_p^i | \varrho | x_q^i \rangle = \sum_n \langle x_p^i | \varphi_n \rangle \langle \varphi_n | x_q^i \rangle n_n \quad , \quad p, q \in \{1, \dots, s_i\}. \quad (7.42)$$

” tr_i “ kürzt die Spurbildung im $s_i \times s_i$ –Raum ab und darf nicht mit der Spur über den Fock–Raum verwechselt werden. Für die Likelihood–Verteilung aller Daten ergibt sich somit

$$p[D|v] = \prod_{i=1}^N \langle x_1^i, \dots, x_{s_i}^i | \frac{e^{-\beta(h(v) - \mu \hat{N})}}{Z_{HF}(v)} | x_1^i, \dots, x_{s_i}^i \rangle = e^{\sum_{i=1}^N \text{tr}_i [\ln \mathcal{W}_i(v)]}. \quad (7.43)$$

Verwendet man die Apriori–Verteilung $p[v]$ von der Form (7.8), so folgt als Maximum–Aposteriori–Approximation von $p[v|D]$ anstatt von (7.14) die Bedingung

$$\begin{aligned} 0 &= \gamma \left(\frac{\delta \Gamma[v]}{\delta v_{1234}} + \frac{\delta \Gamma[v]}{\delta v_{2143}} \right) - \frac{\delta}{\delta v_{1234}} \sum_{i=1}^N \text{tr}_i [\ln \mathcal{W}_i] - \frac{\delta}{\delta v_{2143}} \sum_{i=1}^N \text{tr}_i [\ln \mathcal{W}_i] \quad (7.44) \\ &= \gamma \left(\frac{\delta \Gamma[v]}{\delta v_{1234}} + \frac{\delta \Gamma[v]}{\delta v_{2143}} \right) - \sum_{i=1}^N \sum_{p_i q_i=1}^{s_i} \left(\frac{\delta \mathcal{W}_i}{\delta v_{1234}} + \frac{\delta \mathcal{W}_i}{\delta v_{2143}} \right)_{p_i q_i} (\mathcal{W}_i^{-1})_{q_i p_i}. \end{aligned}$$

Dabei wurde für die Umformung der letzten beiden Summanden

$$\frac{\delta}{\delta v_{1234}} \sum_{i=1}^N \text{tr}_i [\ln \mathcal{W}_i] = \sum_{i=1}^N \text{tr}_i \left[\frac{\delta \mathcal{W}_i}{\delta v_{1234}} \mathcal{W}_i^{-1} \right] = \sum_{i=1}^N \sum_{p_i q_i=1}^{s_i} \left(\frac{\delta \mathcal{W}_i}{\delta v_{1234}} \right)_{p_i q_i} (\mathcal{W}_i^{-1})_{q_i p_i} \quad (7.45)$$

verwendet. Dieser Schritt kann durch Vorgehen analog zum Anhang B bewiesen werden. Im weiteren Verlauf der Rechnung ist man also auf die Bestimmung der Ableitung $\frac{\delta \mathcal{W}_i}{\delta v}$ angewiesen. Aus (7.42) erhält man dabei sofort

$$\left(\frac{\delta \mathcal{W}_i}{\delta v_{1234}} \right)_{pq} = \langle x_p^i | \frac{\delta \varrho}{\delta v_{1234}} | x_q^i \rangle. \quad (7.46)$$

Und die Variation des Dichteoperators (7.34) in seiner Darstellung in Eigenfunktionen liefert

$$\frac{\delta \varrho}{\delta v} = \sum_n \left| \frac{\delta \varphi_n}{\delta v} \right\rangle \langle \varphi_n | n_n + \sum_n \left| \varphi_n \right\rangle \langle \frac{\delta \varphi_n}{\delta v} | n_n + \sum_n \left| \varphi_n \right\rangle \langle \varphi_n | \frac{\delta n_n}{\delta v}. \quad (7.47)$$

Dabei lässt sich die Variation der Besetzungszahlen aus (7.37)

$$\frac{\delta n_n}{\delta v} = -\frac{\beta \delta \varepsilon_n}{2 \delta v} \frac{1}{1 + \cosh(\beta(\varepsilon_n - \mu))} \quad (7.48)$$

gewinnen, und es verbleibt die Berechnung der Ableitungen der Hartree–Fock–Orbitale und der Eigenenergien. Da die Gleichung (7.22) ihre Richtigkeit für den großkanonischen Fall behält:

$$\left| \frac{\delta \varphi_n}{\delta v_{1234}} \right\rangle = - \sum_{k \neq n} \frac{1}{\varepsilon_k - \varepsilon_n} \left| \varphi_k \right\rangle \langle \varphi_k | \frac{\delta h}{\delta v_{1234}} | \varphi_n \rangle, \quad (7.49)$$

kann im folgenden analog zu Kapitel 7.1 vorgegangen werden. Durch Ableitung der Hartree–Fock–Gleichung (7.33) folgt

$$\begin{aligned} \frac{\delta h}{\delta v_{1234}} &= \sum_m n_m \left| \varphi_m \right| 2 \rangle \langle 4 | \varphi_m \rangle a_1^\dagger a_3 - \sum_m n_m \left| \varphi_m \right| 2 \rangle \langle 3 | \varphi_m \rangle a_1^\dagger a_4 \\ &+ \sum_{\alpha\gamma} \sum_m a_\alpha^\dagger a_\gamma n_m \left[\left\langle \alpha \frac{\delta \varphi_m}{\delta v_{1234}} | v | \gamma \varphi_m \right\rangle + \left\langle \alpha \varphi_m | v | \gamma \frac{\delta \varphi_m}{\delta v_{1234}} \right\rangle \right] \\ &+ \sum_{\alpha\gamma} \sum_m a_\alpha^\dagger a_\gamma \left\langle \alpha \varphi_m | v | \gamma \varphi_m \right\rangle \frac{\delta n_m}{\delta v_{1234}} \end{aligned} \quad (7.50)$$

und damit direkt

$$\begin{aligned} \frac{\delta \varepsilon_n}{\delta v_{1234}} &= \sum_m n_m \left| \varphi_n \right| 1 \rangle \langle \varphi_m | 2 \rangle \left[\left\langle 3 | \varphi_n \right\rangle \langle 4 | \varphi_m \rangle - \left\langle 4 | \varphi_n \right\rangle \langle 3 | \varphi_m \rangle \right] \\ &+ \sum_m n_m \left[\left\langle \varphi_n \frac{\delta \varphi_m}{\delta v_{1234}} | v | \varphi_n \varphi_m \right\rangle + \left\langle \varphi_n \varphi_m | v | \varphi_n \frac{\delta \varphi_m}{\delta v_{1234}} \right\rangle \right] \\ &+ \sum_m \left\langle \varphi_n \varphi_m | v | \varphi_n \varphi_m \right\rangle \frac{\delta n_m}{\delta v_{1234}} \end{aligned} \quad (7.51)$$

und

$$\begin{aligned} \left| \frac{\delta \varphi_n}{\delta v_{1234}} \right\rangle &= - \sum_{k \neq n} \sum_m \frac{1}{\varepsilon_k - \varepsilon_n} \left| \varphi_k \right\rangle \times \\ &\left\{ n_m \left| \varphi_m \right| 2 \rangle \langle \varphi_k | 1 \rangle \left[\left\langle 4 | \varphi_m \right\rangle \langle 3 | \varphi_n \rangle - \left\langle 3 | \varphi_m \right\rangle \langle 4 | \varphi_n \rangle \right] \right. \\ &+ n_m \left\langle \varphi_k \frac{\delta \varphi_m}{\delta v_{1234}} | v | \varphi_n \varphi_m \right\rangle + n_m \left\langle \varphi_k \varphi_m | v | \varphi_n \frac{\delta \varphi_m}{\delta v_{1234}} \right\rangle \\ &\left. + \left\langle \varphi_k \varphi_m | v | \varphi_n \varphi_m \right\rangle \frac{\delta n_m}{\delta v_{1234}} \right\} \end{aligned} \quad (7.52)$$

als Verallgemeinerung von (7.26) und (7.27).

Das Lösen der Stationaritätsgleichung (7.44) ist durch mehrere Schritte gekennzeichnet. In einem

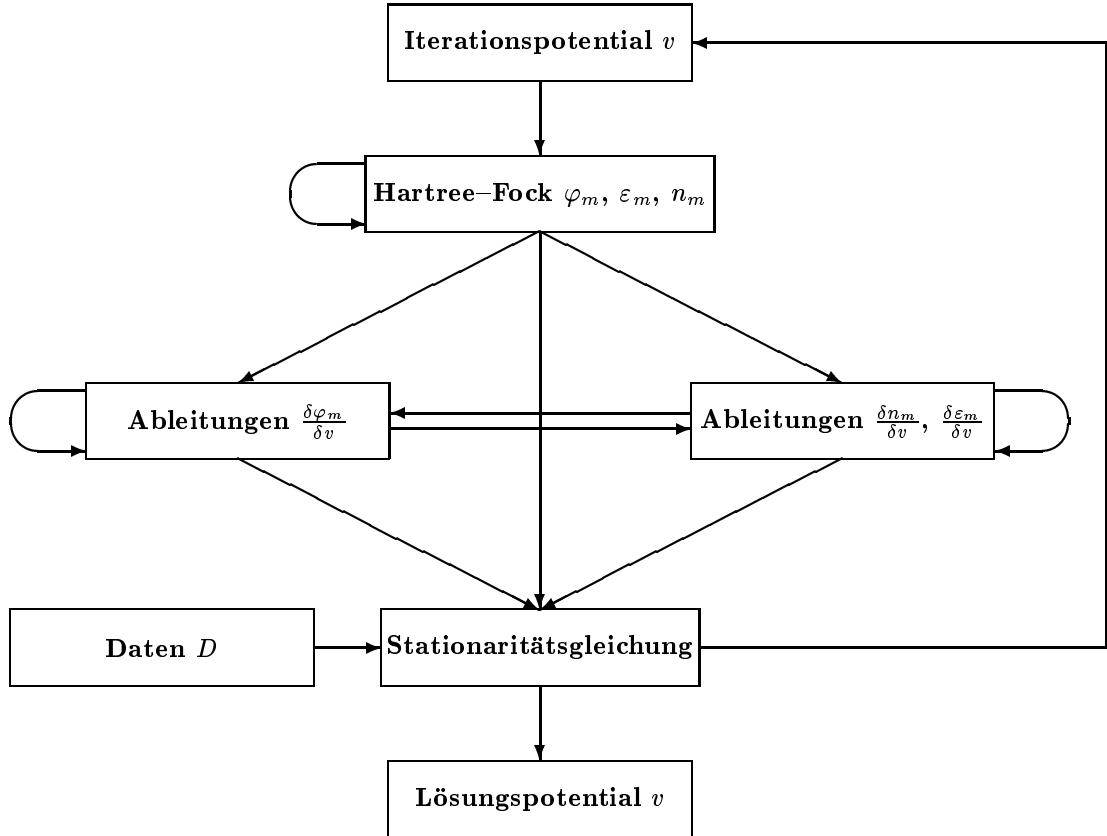


Abbildung 7.1: Erläuterung der Lösungsschritte für die Maximum–Aposteriori–Approximation nach (7.44). Die Pfeile symbolisieren die logischen Abhängigkeiten der einzelnen Größen. Die runden Pfeile bedeuten eine zusätzliche, selbstkonsistente Abhängigkeit.

ersten Schritt werden für ein gegebenes Potential v die Hartree–Fock–Gleichungen (7.33) und (7.35) gelöst. Hiermit können in einem zweiten Schritt die Ableitungen der Orbitale (7.52) und Energien (7.51) gleichzeitig selbstkonsistent bestimmt werden. Mittels eines geeigneten Algorithmus kann im letzten Schritt durch die Stationaritätsgleichung ein neues Potential gewonnen werden, mit dem die Iteration bis zur Konvergenz fortgeführt wird. Die Reihenfolge der Lösungsschritte während der Iteration und die logischen Abhängigkeiten der einzelnen Ausdrücke sind in Abbildung 7.1 dargestellt.

Im Tieftemperaturfall $\beta \rightarrow \infty$ werden Ergebnisse des kanonischen Ensembles aus 7.1 reproduziert. Wählt man das chemische Potential μ so, daß $\langle \hat{N} \rangle$ und s zusammenfallen, so ergibt sich für die Besetzungszahlen: $n_m = 1$ für $m \leq s$ und $n_m = 0$ für $m > s$. Der Dichteoperator fällt damit auf den Projektor auf die s Orbitale mit den niedrigsten Energien zusammen:

$$\varrho = \sum_{m=1}^s |\varphi_m\rangle \langle \varphi_m| \quad \text{und} \quad \varrho^2 = \varrho. \quad (7.53)$$

Damit geht auch der Hartree–Fock–Operator des großkanonischen Ensembles (7.33) in den des kanonischen Ensembles bei $T = 0$ (7.5) über, und im Tieftemperaturfall verschwinden die Variationen der Besetzungszahlen $\frac{\delta n_m}{\delta v} \rightarrow 0$. Die Likelihood–Wahrscheinlichkeiten (7.41) nehmen unter

Ausnutzung von $\det(\mathcal{A}\mathcal{B}) = (\det \mathcal{A})(\det \mathcal{B})$ für die Matrizenmultiplikation mit

$$\begin{aligned} \det \langle x_p^i | \varrho | x_q^i \rangle &= \det \sum_{m=1}^s \langle x_p^i | \varphi_m \rangle \langle \varphi_m | x_q^i \rangle \\ &= \langle x_1^i, \dots, x_s^i | \varphi_1, \dots, \varphi_s \rangle \langle \varphi_1, \dots, \varphi_s | x_1^i, \dots, x_s^i \rangle \end{aligned} \quad (7.54)$$

die Matrixelemente für die Meßwahrscheinlichkeiten aus (7.4) an. Gleichzeitig geht die Stationaritätsgleichung (7.44) in die Gleichung (7.18) über: Unter Benutzung der Überlappmatrizen M_i aus (7.15) und der adjungierten Matrizen M_i^+ können die Matrixelemente des Dichteoperators mittels

$$\mathcal{W}_{pq}^i = \sum_m \langle x_p^i | \varphi_m \rangle \langle \varphi_m | x_q^i \rangle = (M_i M_i^+)^{pq} \quad (7.55)$$

ausgedrückt werden. Unter Ausnutzung von $(\mathcal{A}\mathcal{B})^{-1} = \mathcal{B}^{-1}\mathcal{A}^{-1}$ erhält man mit

$$\begin{aligned} &\sum_{i=1}^N \sum_{pq=1}^s \left(\frac{\delta \mathcal{W}_i}{\delta v} \right)_{pq} (\mathcal{W}_i^{-1})_{qp} \\ &= \sum_{i=1}^N \sum_{pq=1}^s \sum_{nm=1}^s \left(\langle x_p^i | \frac{\delta \varphi_m}{\delta v} \rangle (M_i^+)^{mq} + M_{pm}^i \langle x_q^i | \frac{\delta \varphi_m}{\delta v} \rangle \right) ((M_i^+)^{-1})_{qn} (M_i^{-1})_{np} \\ &= \sum_{i=1}^N \sum_{pm=1}^s \left(\langle x_p^i | \frac{\delta \varphi_m}{\delta v} \rangle (M_i^{-1})_{mp} + \langle x_q^i | \frac{\delta \varphi_m}{\delta v} \rangle ((M_i^+)^{-1})_{pm} \right) \\ &= 2 \operatorname{Re} \left[\sum_{i=1}^N \sum_{pm=1}^s \langle x_p^i | \frac{\delta \varphi_m}{\delta v} \rangle (M_i^{-1})_{mp} \right] \end{aligned} \quad (7.56)$$

das gleiche Ergebnis wie in (7.18).

7.3 Großkanonisches Ensemble (Pfadintegraldarstellung)

Während im Abschnitt 7.2 schon apriori mit dem Hartree–Fock–Hamilton–Operator gearbeitet wurde und die weitere Rechnung auf der Benutzung der Hartree–Fock–Orbitale und deren Funktionalableitung aufbaute, wird hier die Zweiteilchenwechselwirkung mittels eines Pfadintegrals auf ein Integral über Einteilchenoperatoren abgebildet. Die Maximum–Aposteriori–Approximation wird gleichzeitig mit der Sattelpunktsnäherung für das Pfadintegral durchgeführt. Hierin unterscheidet sich das Vorgehen im Vergleich zu Abschnitt 7.2.

Die Linearisierung der Zweiteilchenwechselwirkung eines statistischen Operators ist Standard [16, 21]. Bei der Variante mittels Hubbard–Stratonovich–Transformation geschieht diese Linearisierung auf Kosten einer künstlichen Einführung eines Gauß–Integrals. Nach Einführung der Operatoren $\xi_{\alpha\gamma} = a_\alpha^\dagger a_\gamma$ kann der Vielteilchen–Hamilton–Operator (7.1) mittels

$$H = \sum_{\alpha\gamma} K_{\alpha\gamma} \xi_{\alpha\gamma} + \frac{1}{2} \sum_{\alpha\beta\gamma\delta} v_{\alpha\beta\gamma\delta} \xi_{\alpha\gamma} \xi_{\beta\delta} \quad (7.58)$$

in eine quadratische Form umgeordnet werden. Hierbei sind mit K die Einteilchenmatrixelemente $K_{\alpha\gamma} = t_{\alpha\gamma} - \frac{1}{2} \sum_\beta v_{\alpha\beta\beta\gamma}$ abgekürzt. Das weitere Vorgehen ist durch infinitesimale Zerlegung der inversen Temperatur β des statistischen Operators gekennzeichnet. Durch Erweiterung von β mit \hbar erhält der statistische Operator die Form eines Zeitentwicklungsoperators mit rein imaginärer Zeit.

Bei Aufteilung des Zeitintervalls $\beta\hbar$ im M äquidistante Teilstücke der Länge $\epsilon = \frac{\beta\hbar}{M}$ kommutieren im Limes $\epsilon \rightarrow 0$ die Operatoren des Exponenten, und es kann

$$\begin{aligned} e^{-\beta(H-\mu\hat{N})} &\approx \prod_{k=1}^M e^{-\frac{\epsilon}{\hbar}(\sum_{\alpha\gamma}(K_{\alpha\gamma}-\mu\delta_{\alpha\gamma})\xi_{\alpha\gamma}+\frac{1}{2}\sum_{\alpha\beta\gamma\delta}v_{\alpha\beta\gamma\delta}\xi_{\alpha\gamma}\xi_{\beta\delta})} \\ &\stackrel{\epsilon \rightarrow 0}{\approx} \prod_{k=1}^M e^{-\frac{\epsilon}{\hbar}\sum_{\alpha\gamma}(K_{\alpha\gamma}-\mu\delta_{\alpha\gamma})\xi_{\alpha\gamma}} e^{-\frac{\epsilon}{2\hbar}\sum_{\alpha\beta\gamma\delta}v_{\alpha\beta\gamma\delta}\xi_{\alpha\gamma}\xi_{\beta\delta}} \end{aligned} \quad (7.59)$$

geschrieben werden. Nun kann die Linearisierung des quadratischen Terms vorgenommen werden. Dafür wird für jeden einzelnen Faktor aus (7.59) der Exponent umgeordnet und anschließend ein Gauß-Integral über die Hilfsfelder σ eingeführt:

$$\begin{aligned} e^{-\frac{\epsilon}{2\hbar}\sum_{\alpha\beta\gamma\delta}\xi_{\alpha\gamma}v_{\alpha\beta\gamma\delta}\xi_{\beta\delta}} &= e^{\frac{1}{2}\sum_{\alpha\beta\gamma\delta}(\frac{\epsilon}{\hbar}\xi_{\alpha\gamma})(-\frac{\hbar}{\epsilon}v_{\alpha\beta\gamma\delta})(\frac{\epsilon}{\hbar}\xi_{\beta\delta})} \\ &= \sqrt{\det\left(-\frac{\epsilon}{\hbar}v^{-1}\right)} \int \left(\prod_{\alpha\gamma} \frac{d\bar{\sigma}_{\alpha\gamma}}{\sqrt{2\pi}}\right) e^{\frac{\epsilon}{2\hbar}\sum_{\alpha\beta\gamma\delta}\bar{\sigma}_{\alpha\gamma}(v^{-1})_{\alpha\beta\gamma\delta}\bar{\sigma}_{\beta\delta}-\frac{\epsilon}{\hbar}\sum_{\alpha\gamma}\bar{\sigma}_{\alpha\gamma}\xi_{\alpha\gamma}} \\ &= \sqrt{\det\left(-\frac{\epsilon}{\hbar}v\right)} \int \left(\prod_{\alpha\gamma} \frac{d\sigma_{\alpha\gamma}}{\sqrt{2\pi}}\right) e^{\frac{\epsilon}{2\hbar}\sum_{\alpha\beta\gamma\delta}\sigma_{\alpha\gamma}v_{\alpha\beta\gamma\delta}\sigma_{\beta\delta}-\frac{\epsilon}{\hbar}\sum_{\alpha\beta\gamma\delta}\sigma_{\alpha\gamma}v_{\alpha\beta\gamma\delta}\xi_{\beta\delta}}. \end{aligned} \quad (7.60)$$

Dabei wurde eine Variablentransformation

$$\sigma_{\alpha\gamma} = \sum_{\beta\delta} (v^{-1})_{\alpha\beta\gamma\delta} \bar{\sigma}_{\beta\delta} \quad (7.61)$$

vorgenommen. Eine notwendige Bedingung für die Konvergenz des Hilfsintegrals ist dabei, daß das Potential $-v$ positiv-definit ist. Diese Bedingung muß durch das Vorwissen explizit gewährleistet werden. Setzt man das Ergebnis von (7.60) in den Ausdruck (7.59) ein, so folgt

$$\begin{aligned} e^{-\beta(H-\mu\hat{N})} &\approx \frac{1}{\mathcal{N}_0(v)} \int \left(\prod_{k=1}^M \prod_{\alpha\gamma} d\sigma_{\alpha\gamma}(k) \right) e^{\frac{\epsilon}{2\hbar}\sum_{k=1}^M \sum_{\alpha\beta\gamma\delta} \sigma_{\alpha\gamma}(k)v_{\alpha\beta\gamma\delta}\sigma_{\beta\delta}(k)} \\ &\quad \times \prod_{k=1}^M e^{-\frac{\epsilon}{\hbar}\sum_{\alpha\gamma}(K_{\alpha\gamma}-\mu\delta_{\alpha\gamma}+\sum_{\beta\delta}\sigma_{\beta\delta}(k)v_{\beta\alpha\delta\gamma})\xi_{\alpha\gamma}}, \end{aligned} \quad (7.62)$$

wobei die Normierung mittels

$$\frac{1}{\mathcal{N}_0(v)} = \prod_{k=1}^M \sqrt{\frac{\det\left(-\frac{\epsilon}{\hbar}v\right)}{\prod_{\alpha\gamma} 2\pi}} \quad (7.63)$$

abgekürzt wurde. Beim Übergang von einer endlichen Länge ϵ zu einem infinitesimalen dt und unter Einführung des Zeitordnungsoperators \mathcal{T} erhält man schließlich das Pfadintegral

$$e^{-\beta(H-\mu\hat{N})} = \frac{1}{\mathcal{N}_0(v)} \int \left(\prod_{\alpha\gamma} [d\sigma_{\alpha\gamma}(t)] \right) e^{\frac{1}{2\hbar}\int_0^{\beta\hbar} dt \sum_{\alpha\beta\gamma\delta} \sigma_{\alpha\gamma}(t)v_{\alpha\beta\gamma\delta}\sigma_{\beta\delta}(t)} \mathcal{T} e^{-\frac{1}{\hbar}\int_0^{\beta\hbar} dt g_{\sigma}(t)} \quad (7.64)$$

mit der Normierung

$$\mathcal{N}_0(v) = \int \left(\prod_{\alpha\gamma} [d\sigma_{\alpha\gamma}(t)] \right) e^{\frac{1}{2\hbar}\int_0^{\beta\hbar} dt \sum_{\alpha\beta\gamma\delta} \sigma_{\alpha\gamma}(t)v_{\alpha\beta\gamma\delta}\sigma(t)_{\beta\delta}} \quad (7.65)$$

und dem zeitabhängigen Einteilchenoperator

$$g_\sigma(t) = \sum_{\alpha\gamma} \left(K_{\alpha\gamma} - \mu\delta_{\alpha\gamma} + \sum_{\beta\delta} \sigma_{\beta\delta}(t) v_{\beta\alpha\delta\gamma} \right) \xi_{\alpha\gamma}. \quad (7.66)$$

Die Matrixelemente des statistischen Operators lassen sich dann mit (7.64) mittels

$$\begin{aligned} & \langle x_1^i, \dots, x_{s_i}^i | e^{-\beta(H-\mu\hat{N})} | x_1^i, \dots, x_{s_i}^i \rangle = \\ & \frac{1}{\mathcal{N}_0(v)} \int \left(\prod_{\alpha\gamma} [d\sigma_{\alpha\gamma}^i(t)] \right) e^{\frac{1}{2\hbar} \int_0^{\beta\hbar} dt \sum_{\alpha\beta\gamma\delta} \sigma_{\alpha\gamma}^i(t) v_{\alpha\beta\gamma\delta} \sigma_{\beta\delta}^i(t)} \\ & \times \langle x_1^i, \dots, x_{s_i}^i | \mathcal{T} e^{-\frac{1}{\hbar} \int_0^{\beta\hbar} dt \sum_{\alpha\gamma} g_{\alpha\gamma}^{\sigma^i}(t) a_\alpha^\dagger a_\gamma} | x_1^i, \dots, x_{s_i}^i \rangle \end{aligned} \quad (7.67)$$

ausdrücken.

Der Operator g enthält die Matrixelemente des Potentials v in der Art, daß er mit den Dichten σ^i in einer "Hartree-Form" koppelt. Die Matrixelemente der kinetischen Energie K besitzen weiterhin eine unphysikalische Selbstenergie, welche durch Umsortieren der Erzeugungs- und Vernichtungsoperatoren in die Rechnung gelangte. Nach [16, 21, 22] besteht bei der Einführung des Pfadintegrals eine Willkür darin, die Matrixelemente im Potentialterm von g von "außen" vorzugeben, ohne den Wert des vollen Pfadintegrals als solches zu verändern. Bei einer Sattelpunktsnäherung für das Pfadintegral hat dieser Eingriff allerdings Konsequenzen, indem der Potentialterm das Mean-Field festlegt. Aus Gründen der Anschaulichkeit wurde hier der Weg über die Einführung der Gauß-Integrale mit diesen unerwünschten Konsequenzen vorgeführt. Im weiteren Verlauf der Rechnung wird davon ausgegangen, daß bereits in (7.66) die Matrixelemente ohne Selbstenergie ($K_{\alpha\gamma} \rightarrow t_{\alpha\gamma}$) in der Hartree-Fock-Form ($v_{\alpha\beta\gamma\delta} \rightarrow \bar{v}_{\alpha\beta\gamma\delta}$) vorliegen.

Die Darstellung (7.67) für die Likelihood-Wahrscheinlichkeiten kann nun für die Aposteriori-Wahrscheinlichkeiten des Potentials verwendet werden. Benutzt man eine Apriori-Verteilung nach (7.8), so folgt:

$$\begin{aligned} p[v|D] & \sim p[D|v] p[v] = p[v] \prod_{i=1}^N \langle x_1^i, \dots, x_{s_i}^i | \frac{e^{-\beta(H-\mu\hat{N})}}{Z} | x_1^i, \dots, x_{s_i}^i \rangle \\ & = \int \left(\prod_{\alpha\beta\gamma\delta} d\lambda_{\alpha\beta\gamma\delta} \right) \int \left(\prod_{i=1}^N \prod_{\alpha\gamma} [d\sigma_{\alpha\gamma}^i(t)] \right) e^{-F[\sigma^i, v, \lambda, D]}. \end{aligned} \quad (7.68)$$

Die bei einer Maximum-Aposteriori-Approximation zu minimierende Wirkung F ist dabei durch

$$\begin{aligned} F[\sigma^i, v, \lambda, D] & = -\frac{1}{2\hbar} \sum_{i=1}^N \int_0^{\beta\hbar} dt \sum_{\alpha\beta\gamma\delta} \sigma_{\alpha\gamma}^i(t) v_{\alpha\beta\gamma\delta} \sigma_{\beta\delta}^i(t) + N \ln Z(v) + \gamma \Gamma[v] + N \ln \mathcal{N}_0(v) \\ & - \sum_{i=1}^N \ln \langle \chi^i | \mathcal{T} e^{-\frac{1}{\hbar} \int_0^{\beta\hbar} dt g_{\sigma^i}(t)} | \chi^i \rangle - i \sum_{\alpha\beta\gamma\delta} \lambda_{\alpha\beta\gamma\delta} (v_{\alpha\beta\gamma\delta} - v_{\beta\alpha\delta\gamma}) \end{aligned} \quad (7.69)$$

mit der Abkürzung $|\chi^i\rangle = |x_1^i, \dots, x_{s_i}^i\rangle$ gegeben. Hierbei ist in der Bestrafungsfunktion Γ weiterhin zu gewährleisten, daß $-v$ positiv-definit ist. Dies könnte indirekt geschehen, indem analog (5.38) eine Funktion v_0 als positiv-definites Muster mit hinreichend großer Gewichtung durch γ gewählt wird oder direkt durch eine explizite Parametrisierung von v nach $v = -w^2$.

Es wird nun eine gleichzeitige Sattelpunktsnäherung für die Felder σ^i und eine Maximum–Aposteriori–Approximation für das zu lernende Potential v durchgeführt. Die Stationaritätsbedingung für die Dichten lautet:

$$0 = \frac{\delta F}{\delta \sigma_{12}^i(t)} = -\frac{1}{\hbar} \sum_{\beta\delta} v_{1\beta 2\delta} \sigma_{\beta\delta}^i(t) + \frac{1}{\hbar} \sum_{\alpha\gamma} v_{1\alpha 2\gamma} \frac{\langle \chi^i | [\mathcal{T} e^{-\frac{1}{\hbar} \int_t^{\beta\hbar} dt' g_{\sigma^i}(t')}] a_{\alpha}^{\dagger} a_{\gamma} [\mathcal{T} e^{-\frac{1}{\hbar} \int_0^t dt' g_{\sigma^i}(t')}] | \chi^i \rangle}{\langle \chi^i | \mathcal{T} e^{-\frac{1}{\hbar} \int_0^{\beta\hbar} dt' g_{\sigma^i}(t')} | \chi^i \rangle} \quad \forall i, \quad (7.70)$$

und durch Auflösung erhält man die selbstkonsistente Gleichung

$$\sigma_{\beta\delta}^i(t) = \frac{\langle \chi^i | [\mathcal{T} e^{-\frac{1}{\hbar} \int_t^{\beta\hbar} dt' g_{\sigma^i}(t')}] a_{\beta}^{\dagger} a_{\delta} [\mathcal{T} e^{-\frac{1}{\hbar} \int_0^t dt' g_{\sigma^i}(t')}] | \chi^i \rangle}{\langle \chi^i | \mathcal{T} e^{-\frac{1}{\hbar} \int_0^{\beta\hbar} dt' g_{\sigma^i}(t')} | \chi^i \rangle}, \quad i \in \{1, \dots, N\}. \quad (7.71)$$

Diese muß für alle Datenvektoren $|\chi^i\rangle$ separat gelöst werden. Dadurch wird auch der Hamilton–Operator g_{σ^i} aus (7.66) datenabhängig. Hierin unterscheidet sich der Ansatz dieses Abschnitts im Vergleich zu Kapitel 7.2, wo der Hartree–Fock–Operator (7.33) datenunabhängig vorliegt.

Es bietet sich die Definition von Einteilchendichteoperatoren ϱ_i an, welche über die Matrixelemente

$$\langle \delta | \varrho_i(t) | \beta \rangle = \sigma_{\beta\delta}^i(t) \quad (7.72)$$

festgelegt werden. Durch Zeitableitung von (7.71) läßt sich dann unter Ausnutzung des Kommutators $[g(t), a_{\beta}^{\dagger} a_{\delta}] = \sum_{\alpha} (g_{\alpha\beta}(t) a_{\alpha}^{\dagger} a_{\delta} - g_{\delta\alpha}(t) a_{\beta}^{\dagger} a_{\alpha})$ zeigen, daß für die Operatoren die Bewegungsgleichung der zeitabhängigen Hartree–Fock–Methode

$$\frac{\partial}{\partial t} \varrho_i(t) = -\frac{1}{\hbar} [g_{\varrho_i}(t), \varrho_i(t)] \quad (7.73)$$

erfüllt ist.

Die Maximum–Aposteriori–Approximation für (7.69) liefert die Gleichung

$$0 = \frac{\delta F}{\delta v_{1234}} = -\frac{1}{2\hbar} \sum_{i=1}^N \int_0^{\beta\hbar} dt \sigma_{13}^i(t) \sigma_{24}^i(t) + N \frac{1}{Z} \frac{\delta Z}{\delta v_{1234}} + \frac{N}{\mathcal{N}_0} \frac{\delta \mathcal{N}_0}{\delta v_{1234}} + \gamma \frac{\delta \Gamma[v]}{\delta v_{1234}} - i\lambda_{1234} + i\lambda_{2143} + \frac{1}{\hbar} \int_0^{\beta\hbar} dt \sum_{i=1}^N \sigma_{13}^i(t) \frac{\langle \chi^i | [\mathcal{T} e^{-\frac{1}{\hbar} \int_t^{\beta\hbar} dt' g_{\sigma^i}(t')}] a_2^{\dagger} a_4 [\mathcal{T} e^{-\frac{1}{\hbar} \int_0^t dt' g_{\sigma^i}(t')}] | \chi^i \rangle}{\langle \chi^i | \mathcal{T} e^{-\frac{1}{\hbar} \int_0^{\beta\hbar} dt' g_{\sigma^i}(t')} | \chi^i \rangle}. \quad (7.74)$$

Die Bedingung (7.71) kann in (7.74) eingesetzt werden. Dabei erhält man

$$0 = \frac{1}{2\hbar} \sum_{i=1}^N \int_0^{\beta\hbar} dt \sigma_{13}^i(t) \sigma_{24}^i(t) + N \frac{1}{Z} \frac{\delta Z}{\delta v_{1234}} + \frac{N}{\mathcal{N}_0} \frac{\delta \mathcal{N}_0}{\delta v_{1234}} + \gamma \frac{\delta \Gamma[v]}{\delta v_{1234}} - i\lambda_{1234} + i\lambda_{2143}. \quad (7.75)$$

Für die weitere Verarbeitung von (7.75) ist die Berechnung der Ableitung der Zustandssumme Z unerlässlich. Die Zustandssumme läßt sich mit Hilfe (7.64) auf die Form

$$\begin{aligned} Z &= \text{Tr} [e^{-\beta(H - \mu \hat{N})}] = \\ &= \frac{1}{\mathcal{N}_0} \int \left(\prod_{\alpha\gamma} [d\tilde{\sigma}_{\alpha\gamma}(t)] \right) e^{\frac{1}{\hbar} \int_0^{\beta\hbar} dt \sum_{\alpha\beta\gamma\delta} \tilde{\sigma}_{\alpha\gamma}(t) v_{\alpha\beta\gamma\delta} \tilde{\sigma}_{\beta\delta}(t)} \text{Tr} [\mathcal{T} e^{-\frac{1}{\hbar} \int_0^{\beta\hbar} dt g_{\tilde{\sigma}}(t)}] \end{aligned} \quad (7.76)$$

bringen. Stationäre Phasenapproximation des $\tilde{\sigma}$ -Integrals führt auf die Mean-Field-Dichten als Hauptbeitrag zum Integral:

$$\tilde{\sigma}_{\beta\delta}(t) = \frac{\text{Tr} \left[[\mathcal{T} e^{-\frac{1}{\hbar} \int_t^{\beta\hbar} dt' g_{\tilde{\sigma}}(t')}] a_{\beta}^{\dagger} a_{\delta} [\mathcal{T} e^{-\frac{1}{\hbar} \int_0^t dt' g_{\tilde{\sigma}}(t')}] \right]}{\text{Tr} \left[\mathcal{T} e^{-\frac{1}{\hbar} \int_0^{\beta\hbar} dt' g_{\tilde{\sigma}}(t')} \right]} \quad (7.77)$$

mit $g_{\tilde{\sigma}}$ analog (7.66). Als Ableitung für die Zustandssumme erhält man

$$\begin{aligned} \frac{1}{Z} \frac{\delta Z}{\delta v_{1234}} &= -\frac{1}{\mathcal{N}_0} \frac{\delta \mathcal{N}_0}{\delta v_{1234}} \\ &+ \frac{1}{\mathcal{N}_0 Z} \int \left(\prod_{\alpha\gamma} [d\tilde{\sigma}_{\alpha\gamma}(t)] \right) e^{\frac{1}{2\hbar} \int_0^{\beta\hbar} dt \sum_{\alpha\beta\gamma\delta} \tilde{\sigma}_{\alpha\gamma}(t) v_{\alpha\beta\gamma\delta} \tilde{\sigma}_{\beta\delta}(t) + \ln \text{Tr} \left[\mathcal{T} e^{-\frac{1}{\hbar} \int_0^{\beta\hbar} dt g_{\tilde{\sigma}}(t)} \right]} \\ &\times \frac{1}{\hbar} \int_0^{\beta\hbar} dt \left\{ \frac{1}{2} \tilde{\sigma}_{13}(t) \tilde{\sigma}_{24}(t) - \tilde{\sigma}_{13}(t) \frac{\text{Tr} \left[[\mathcal{T} e^{-\frac{1}{\hbar} \int_t^{\beta\hbar} dt' g_{\tilde{\sigma}}(t')}] a_2^{\dagger} a_4 [\mathcal{T} e^{-\frac{1}{\hbar} \int_0^t dt' g_{\tilde{\sigma}}(t')}] \right]}{\text{Tr} \left[\mathcal{T} e^{-\frac{1}{\hbar} \int_0^{\beta\hbar} dt' g_{\tilde{\sigma}}(t')} \right]} \right\}. \end{aligned} \quad (7.78)$$

Diese Gleichung läßt sich nicht exakt handhaben. Man ist im weiteren auf Näherungen für das Pfadintegral angewiesen. Wie schon in Kapitel 6.4 liegt im zweiten Summand von (7.78) ein Integral der Form $I = \frac{1}{\mathcal{N}_0 Z} \int [d\tilde{\sigma}] f(\tilde{\sigma}) z(\tilde{\sigma})$ vor, wobei die Zustandssumme selber durch $Z = \frac{1}{\mathcal{N}_0} \int [d\tilde{\sigma}] f(\tilde{\sigma})$ gegeben ist. Unter der Annahme, daß die Funktion f um ihre Mean-Field-Dichte $\tilde{\sigma}_0$ (7.77) hinreichend lokalisiert ist, während die Funktion z dort schwach oszilliert, kann z an der Stelle $\tilde{\sigma}_0$ unter dem Integral herausgezogen werden. Damit erhält man $I \approx z(\tilde{\sigma}_0)$, und unter dieser Näherung ergibt sich:

$$\frac{1}{Z} \frac{\delta Z}{\delta v_{1234}} \approx -\frac{1}{\mathcal{N}_0} \frac{\delta \mathcal{N}_0}{\delta v_{1234}} - \frac{1}{2\hbar} \int_0^{\beta\hbar} dt \tilde{\sigma}_{13}(t) \tilde{\sigma}_{24}(t). \quad (7.79)$$

Nach Eliminierung der Lagrange-Parameter analog zu Kapitel 7.1 kann für die Stationaritätsgleichung (7.75)

$$0 = \frac{1}{\beta\hbar} \int_0^{\beta\hbar} dt \left[\frac{1}{N} \sum_{i=1}^N \sigma_{13}^i(t) \sigma_{24}^i(t) - \tilde{\sigma}_{13}(t) \tilde{\sigma}_{24}(t) \right] + \frac{\gamma}{N\beta} \left(\frac{\delta \Gamma}{\delta v_{1234}} + \frac{\delta \Gamma}{\delta v_{2143}} \right) \quad (7.80)$$

geschrieben werden. (7.80) muß zusammen mit den Gleichungen (7.71) und (7.77) für die Bestimmung der Dichten gelöst werden.

Beschränkt man sich auf die Suche nach statischen Dichten, so können die Matrixelemente aus (7.71) und die Spurbildungen aus (7.77) ausgeführt werden. Für zeitunabhängige Hamilton-Operatoren g erhält man für (7.77) die Bestimmungsgleichung für die Hartree-Fock-Dichten:

$$\tilde{\sigma}_{\beta\delta} = \frac{\text{Tr} [e^{-\beta g_{\tilde{\sigma}}} a_{\beta}^{\dagger} a_{\delta}]}{\text{Tr} [e^{-\beta g_{\tilde{\sigma}}}] } = < \delta | \varrho | \beta > = \sum_m n_m < \delta | \varphi_m > < \varphi_m | \beta >. \quad (7.81)$$

Dabei entspricht ϱ dem Dichteoperator

$$\varrho = \frac{1}{1 + e^{\beta(h_{\tilde{\sigma}} - \mu \tilde{N})}} \quad (7.82)$$

und n_m der Fermi-Verteilung (7.37). φ_m kennzeichnen erneut die Hartree-Fock-Orbitale, welche selbstkonsistent als Eigenzustände von

$$h_{\alpha\gamma}^{\tilde{\sigma}} = t_{\alpha\gamma} + \sum_{\beta\delta} \tilde{\sigma}_{\beta\delta} v_{\beta\alpha\delta\gamma} = t_{\alpha\gamma} + \sum_m < \alpha\varphi_m | v | \gamma\varphi_m > n_m \quad (7.83)$$

definiert sind.

Die statischen, datenabhängigen Dichten σ^i als Lösungen der Gleichung (7.71) müssen – falls sie existieren – aufgrund von (7.73) mit dem Einteilchenoperator (7.66) kommutieren:

$$0 = [g_{\varrho_i}, \varrho_i] \quad \forall i. \quad (7.84)$$

Für die Bestimmungsgleichung der Dichten erhält man dann bei der Wahl von $t = 0$ die Matrixelemente

$$\sigma_{\beta\delta}^i = \frac{\langle \chi^i | e^{-\beta(h_{\sigma^i} - \mu\hat{N})} a_{\beta}^{\dagger} a_{\delta} | \chi^i \rangle}{\langle \chi^i | e^{-\beta(h_{\sigma^i} - \mu\hat{N})} | \chi^i \rangle} \quad (7.85)$$

mit

$$h_{\alpha\gamma}^{\sigma^i} = t_{\alpha\gamma} + \sum_{\beta\delta} \sigma_{\beta\delta}^i v_{\beta\alpha\delta\gamma}. \quad (7.86)$$

Es bietet sich nun eine Einführung von Slaterdeterminanten $\langle \Psi^{\sigma^i} | = \langle \psi_1^{\sigma^i}, \dots, \psi_{s_i}^{\sigma^i} |$ mittels

$$\langle x_1^i, \dots, x_{s_i}^i | e^{-\beta(h_{\sigma^i} - \mu\hat{N})} = \langle \psi_1^{\sigma^i}, \dots, \psi_{s_i}^{\sigma^i} | \quad (7.87)$$

an. Durch den Einteilchencharakter des Hamilton–Operators g_{σ^i} entspricht dies der Entwicklung der Orbitale nach

$$\langle x_n^i | e^{-\beta(h_{\sigma^i} - \mu)} = \langle \psi_n^{\sigma^i} |, \quad n = \{1, \dots, s_i\}, \quad i = \{1, \dots, N\}. \quad (7.88)$$

Nach Einsetzen der Definition können die Matrixelemente mittels Wick–Theorem ausgewertet werden, und man erhält das Ergebnis:

$$\sigma_{\beta\delta}^i = \frac{\langle \Psi^{\sigma^i} | a_{\beta}^{\dagger} a_{\delta} | \chi^i \rangle}{\langle \Psi^{\sigma^i} | \chi^i \rangle} = \langle \delta | \varrho_i | \beta \rangle \quad (7.89)$$

mit dem Dichteoperator

$$\varrho_i = \sum_{mn=1}^{s_i} \langle x_n^i | (B_i^{-1})_{nm} | \psi_m^{\sigma^i} | \quad (7.90)$$

und der Überlappmatrix

$$B_{mn}^i = \langle \psi_m^{\sigma^i} | x_n^i \rangle = \langle x_m^i | e^{-\beta(h_{\sigma^i} - \mu)} | x_n^i \rangle. \quad (7.91)$$

Die Gleichungen (7.84) und (7.89) müssen nun zusammen selbstkonsistent gelöst werden. Dabei ist die Lösung σ^i im allgemeinen nichthermitisch. Dies hat eine mögliche Abweichung von der Hermitizität für v zur Folge. Das Problem kann korrigiert werden, indem eine hermitische Matrix mittels Übergang $\sigma_{\alpha\gamma}^i \rightarrow \frac{1}{2}(\sigma_{\alpha\gamma}^i + (\sigma_{\gamma\alpha}^i)^*)$ konstruiert wird.

Die Stationaritätsgleichung (7.80) geht im statischen Fall in die Bedingung

$$0 = \frac{\gamma}{\beta N} \left(\frac{\delta\Gamma}{\delta v_{1234}} + \frac{\delta\Gamma}{\delta v_{2143}} \right) - \tilde{\sigma}_{13}\tilde{\sigma}_{23} + \frac{1}{N} \sum_{i=1}^N \sigma_{13}^i \sigma_{24}^i \quad (7.92)$$

über. Die beiden Stationaritätsgleichungen (7.92) und (7.44) unterscheiden sich in wesentlichen Punkten. In Abschnitt 7.2 werden die Mean–Field–Approximation und die Maximum–Aposteriori–Approximation nacheinander durchgeführt. Dies hat zur Folge, daß nach Abbildung 7.1 die Ableitungen der Mean–Field–Größen auch selbstkonsistent voneinander abhängen. Die Daten gehen erst

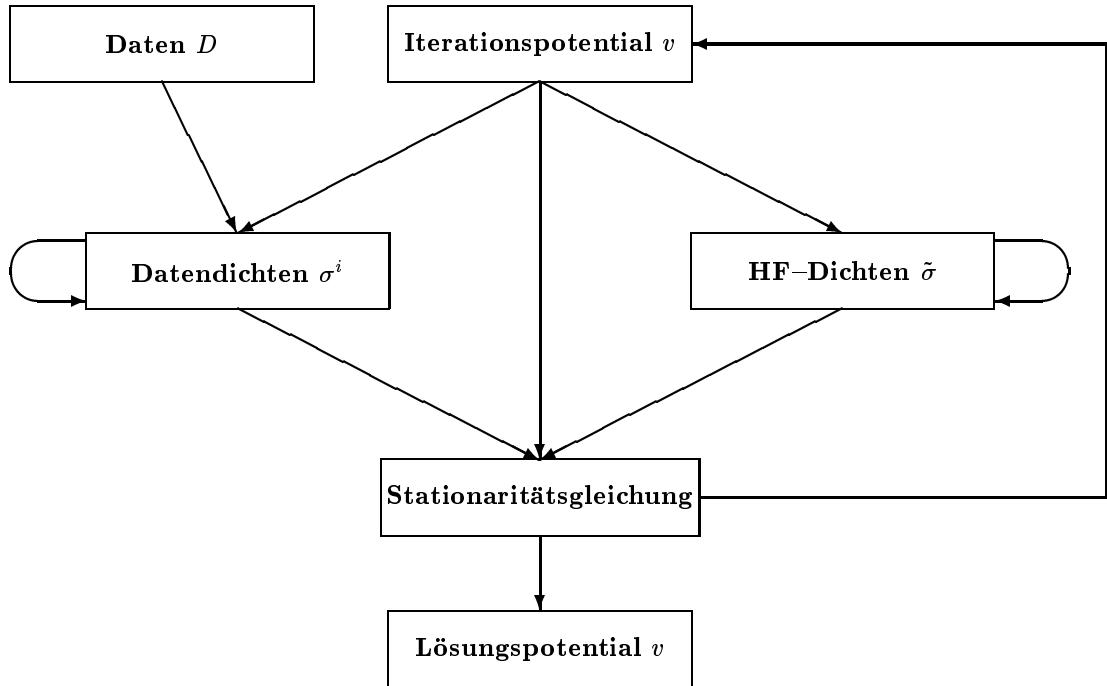


Abbildung 7.2: Erluterung der Lsungsschritte fr die Maximum-Aposteriori-Approximation nach (7.92).

durch die Stationaritatsgleichung in den Formalismus ein. In (7.92) knnen durch die gleichzeitige Anwendung der Approximationen die beiden Dichten $\tilde{\sigma}$ und σ^i als Eingabe fr die Stationaritatsgleichung unabhangig voneinander bestimmt werden. Dies geschieht unter den Kosten, da die Dichten σ^i datenabhangig sind und fr jedes Datum getrennt berechnet werden mussen. Die logischen Abhangigkeiten der fr die Stationaritatsgleichung (7.92) relevanten Groen sind in der Abbildung 7.2 dargestellt. Man beachte, da zwischen $\tilde{\sigma}$ und σ^i keine Pfeile auftreten. Aufgrund der Unabhangigkeit dieser Groen kann man fr die numerische Anwendbarkeit sagen, da bei nicht zu groer Datenmenge D der Formalismus aus diesem Abschnitt dem aus 7.2 vorzuziehen ist.

Zusammenfassung

Diese Arbeit basiert auf der fundamentalen Anwendung des Bayes–Theorems. Hierin liegt der Schlüssel, das inverse Problem des Lernens in ein "direktes" umzuwandeln. Durch die einfache Möglichkeit, die Abhängigkeiten in den bedingten Wahrscheinlichkeiten umzukehren, ergibt sich ein neuer Zugang, an quantenmechanischen Systemen zu lernen.

Der Ansatz des Bayesschen Lernens hat den entscheidenden Vorteil, daß sich sämtliche Erwartungen – physikalischer und nichtphysikalischer Natur – durch das Vorwissen $p[f]$ auf natürliche Weise in den Lernprozeß einbauen lassen. Des Weiteren kann für die Berechnung der Likelihood–Wahrscheinlichkeiten $p[D|f]$ auf die vorhandenen Ergebnisse der Quantenmechanik zurückgegriffen werden. Analytischer Schwerpunkt der Arbeit war die Berechnung der Maximum–Aposteriori–Approximation von $p[f|D] \sim p[D|f] p[D]$ für verschiedenste Anwendungen aus dem Bereich der klassischen Quantenstatistik. Hauptaugenmerk galt der Variation der Likelihood–Dichte $p[D|f]$. Über diese Funktion nimmt die Quantenmechanik Einfluß auf die Rechnung. Damit wurden die wesentlichen Voraussetzungen für das Lernen in der Quantentheorie geschaffen. Die Maximum–Aposteriori–Approximation führt auf nichtlineare Gleichungen. Bei der Suche nach Lösungen ist man auf eine numerische Behandlung und in der Regel auf eine Diskretisierung des Problems angewiesen. In zahlreichen numerischen Beispielen konnte die Anwendbarkeit der Theorie belegt werden. Es zeigt sich, daß die Wahl des Vorwissens dabei von tragender Bedeutung ist. Dies liegt an der Tatsache, daß die Likelihood–Verteilung $p[D|f]$ als Funktion des Zustands f für fixierte Daten keine injektive Funktion ist. Das heißt, es liegt keine eindeutige Abbildung vor, da in der Regel eine große Zahl sehr unterschiedlicher Zustände f die gleiche oder eine sehr ähnliche Likelihood–Verteilung haben. Den observablen Funktionen, wie den Verteilungen $|\psi(x)|^2$ und $\langle x | \rho | x \rangle$, geht Information durch die Betragsquadratbildung von ψ oder durch die Projektion des Operators ρ verloren. Dieser Informationsverlust kann nur durch geeignetes Vorwissen kompensiert werden. Dadurch kann eine Konvexität der Energiefunktionale erzwungen werden. Ein wesentliches Merkmal der Theorie ist die probabilistische Unabhängigkeit der Messungen. Dies setzt eine gleichartige Präparation von identischen Systemen jeweils vor den Messungen voraus und kann als Nachteil der Methode angesehen werden. Die Betrachtung von kanonischen Ensembles ist ein natürlicher Ausweg, der gleichzeitig aber den oft vorliegenden experimentellen Voraussetzungen entgegenkommt. Lemm hat in [24] eine Anwendung untersucht, bei der die Unabhängigkeit der Messungen fallen gelassen wurde.

Insgesamt kann man sagen, daß die Idee, das Bayessche Lernen in der Quantentheorie anzuwenden, um statistische Operatoren und Ein– und Zweiteilchenpotentiale zu lernen, analytisch und numerisch verwirklicht werden konnte. In bezug auf die nichtrelativistische Quantentheorie kann der Formalismus als abgerundet angesehen werden. Es gibt einige Veränderungsmöglichkeiten in den Details: So könnten beim Lernen der Potentiale andere als Ortsmessungen berücksichtigt werden. Hierbei seien bei Vielteilchenexperimenten erhaltene Zweiteilchenkorrelationen im Impulsraum erwähnt, die zusätzlich zu den Einteilchendichten verwendet werden könnten. Des Weiteren könnte die in Kapitel 7 beschriebene Freiheit bei der Wahl des Mean–Fields dafür genutzt werden, einen von der Hartree–Fock–Methode abweichenden Näherungsformalismus anzuwenden. Eine

Weiterführung dieses Gedankens führt auf die Betrachtung von Quasiteilchen und Pairing–Mean–Field im Rahmen der Hartree–Fock–Bogoliubov–Methode. Eine hierzu alternative Möglichkeit, die Ergebnisse der Hartree–Fock–Methode zu verbessern, wäre der Einbau einer Random–Phase–Approximation. Ein aus numerischen Gründen vorteilhafter Ansatz wäre, die Basis des vollen Hilbert–Raums auf eine endlichdimensionale Basis einzuschränken. Die Variation der Wellenfunktionen und die Diagonalisierung des vollen Hamilton–Operators müßten in diesem Teilraum ausgeführt werden. Weitere Generalisierungsmöglichkeiten könnten Anwendungen aus dem Bereich der relativistischen Quantenmechanik sein. Dies führt auf das Arbeiten mit der Klein–Gordon–beziehungsweise Dirac–Gleichung.

Anhang A

Datenerzeugung aus vorgegebener Wahrscheinlichkeitsverteilung

Möchte man Daten aus einer auf $x \in [x_{min}, x_{max}]$ definierten, vorgegebenen Wahrscheinlichkeitsverteilung $p[x]$ gewinnen, wobei nur auf gleichverteilte Zufallszahlen aus dem Intervall $[0, 1]$ zurückgegriffen werden kann, so betrachtet man die Verteilungsfunktion ϕ von $p[x]$:

$$\phi : [x_{min}, x_{max}] \rightarrow [0, 1] \quad \phi(x) = \int_{x_{min}}^x dx' p[x'] . \quad (\text{A.1})$$

Für $p[x] > 0$ ist die Funktion ϕ streng monoton steigend und damit umkehrbar. Für die nach

$$\phi^{-1} : [0, 1] \rightarrow [x_{min}, x_{max}] \quad x = \phi^{-1}(y) \quad (\text{A.2})$$

definierte Umkehrfunktion kann die Ableitung

$$\frac{dx}{dy} = \frac{d\phi^{-1}(y)}{dx} = \frac{1}{\phi'(x)} = \frac{1}{p[x]} \quad (\text{A.3})$$

bestimmt werden.

Die Motivation für die Einführung von ϕ wird deutlich, wenn man den Erwartungswert für eine beliebige Funktion $f(x)$ unter der Verteilung $p[x]$ berechnet. Durch Substitution nach (A.2) lässt sich mit (A.3) die Identität

$$\int_{x_{min}}^{x_{max}} dx p[x] f(x) = \int_0^1 dy f(\phi^{-1}(y)) \quad \forall f \quad (\text{A.4})$$

zeigen. Da (A.4) unabhängig von f erfüllt ist, ist es also bei einer empirischen Auswertung der Integrale äquivalent, die aus der Verteilung $p[x]$ gewonnenen x -Werte oder die gleichverteilten y -Werte zu verwenden. Das heißt, die Daten x_i aus der Verteilung $p[x]$ können erzeugt werden, indem gleichverteilte Zufallszahlen $y_i \in [0, 1]$ gezogen und die zugehörigen $x_i = \phi^{-1}(y_i)$ berechnet werden.

Anhang B

Operatorenableitung

Unter Ableitung der Wirkung (3.10) nach dem Operator ρ versteht man die Ableitung nach den Matrixelementen $\rho_{st} = \langle s | \rho | t \rangle$ in der festen, aber beliebigen Basis, in der auch die Spur ausgewertet wird. In Matrixelementen ausgedrückt lässt sich die Wirkung als

$$\begin{aligned} F[\rho] &= k \sum_{\alpha\beta} \rho_{\alpha\beta} (\ln \rho)_{\beta\alpha} - \mu \sum_{\alpha\beta} \rho_{\alpha\beta} \rho_{\beta\alpha}^0 - i\lambda \left(\sum_{\alpha} \rho_{\alpha\alpha} - 1 \right) \\ &\quad - \sum_{i=1}^N \ln \left(\sum_{\alpha\beta} \langle \alpha | \chi_i \rangle \langle \chi_i | \beta \rangle \rho_{\beta\alpha} \right) \end{aligned} \quad (\text{B.1})$$

schreiben. Die Ableitungen nach Matrixelementen ρ_{st} sind bis auf den ersten Summanden trivial. Im folgenden wird das intuitiv zu erwartende Ergebnis

$$\frac{\partial}{\partial \rho_{st}} \sum_{\alpha\beta} \rho_{\alpha\beta} (\ln \rho)_{\beta\alpha} = (\ln \rho)_{ts} + \delta_{ts} \quad (\text{B.2})$$

hergeleitet.

Unter der Annahme, daß die Eigenwerte von ρ im Intervall $p_n \in [0, 2]$ liegen, kann die Taylor-Reihe des Logarithmus verwendet werden:

$$\ln \rho = \sum_{n=1}^{\infty} c_n (\rho - \mathcal{I})^n \quad \text{mit} \quad c_n = \frac{(-1)^{n+1}}{n}. \quad (\text{B.3})$$

Setzt man die Potenzreihe für den Logarithmus ein, addiert und subtrahiert $\delta_{\alpha\beta}$ und drückt das n -fache Matrixprodukt mittels y_1, \dots, y_{n-1} -Summationen aus, so bekommt man nach

$$\begin{aligned} \sum_{\alpha\beta} \rho_{\alpha\beta} (\ln \rho)_{\beta\alpha} &= \sum_{n=1}^{\infty} c_n \sum_{\alpha\beta} \left[(\rho - \mathcal{I})_{\alpha\beta} + \delta_{\alpha\beta} \right] (\rho - \mathcal{I})_{\beta\alpha}^n \\ &= \sum_{n=1}^{\infty} c_n \sum_{\alpha\beta} \sum_{y_1 \dots y_{n-1}} (\rho - \mathcal{I})_{\alpha\beta} (\rho - \mathcal{I})_{\beta y_1} (\rho - \mathcal{I})_{y_1 y_2} \dots (\rho - \mathcal{I})_{y_{n-1} \alpha} \\ &\quad + \sum_{n=1}^{\infty} c_n \sum_{\alpha\beta} \delta_{\alpha\beta} \sum_{y_1 \dots y_{n-1}} (\rho - \mathcal{I})_{\beta y_1} (\rho - \mathcal{I})_{y_1 y_2} \dots (\rho - \mathcal{I})_{y_{n-1} \alpha} \end{aligned} \quad (\text{B.4})$$

eine in $(\rho - \mathcal{I})$ zyklische Struktur unter den Summen. Dadurch erhält man nach Ableiten mit der Produktregel und Zusammenfassen der beiden Summanden das Ergebnis (B.2):

$$\begin{aligned}
 \frac{\partial}{\partial \rho_{st}} \sum_{\alpha\beta} \rho_{\alpha\beta} (\ln \rho)_{\beta\alpha} &= \sum_{n=1}^{\infty} c_n (n+1) \sum_{y_1 \dots y_{n-1}} (\rho - \mathcal{I})_{ty_1} (\rho - \mathcal{I})_{y_1 y_2} \dots (\rho - \mathcal{I})_{y_{n-1} s} + \\
 &\quad + \sum_{n=1}^{\infty} c_n n \sum_{y_1 \dots y_{n-2}} (\rho - \mathcal{I})_{ty_1} (\rho - \mathcal{I})_{y_1 y_2} \dots (\rho - \mathcal{I})_{y_{n-2} s} \\
 &= \sum_{n=1}^{\infty} c_n (n+1) (\rho - \mathcal{I})_{ts}^n + \sum_{n=1}^{\infty} c_n n (\rho - \mathcal{I})_{ts}^{n-1} \\
 &= \sum_{n=1}^{\infty} [(c_n + c_{n+1}) (n+1)] (\rho - \mathcal{I})_{ts}^n + c_1 \delta_{ts} = (\ln \rho)_{ts} + \delta_{ts}, \tag{B.5}
 \end{aligned}$$

wobei im letzten Schritt für die Koeffizienten

$$(c_n + c_{n+1}) (n+1) = \frac{(-1)^{n+1}}{n} = c_n \quad \text{und} \quad c_1 = 1 \tag{B.6}$$

angewandt wurde.

Somit erhält man für die Ableitung von (B.1)

$$\frac{\partial F[\rho]}{\partial \rho_{st}} = k(\ln \rho)_{ts} + k\delta_{ts} - \mu \rho_{ts}^0 - i\lambda \delta_{ts} - \sum_{i=1}^N \frac{< t | \chi_i > < \chi_i | s >}{\sum_{\alpha\beta} < \alpha | \chi_i > < \chi_i | \beta > \rho_{\beta\alpha}} \quad \forall s, t. \tag{B.7}$$

Da (B.7) für alle Matrixelemente gilt, kann die Bedingung $\frac{\delta F}{\delta \rho} = 0$ mittels

$$0 = k \ln \rho - (i\lambda - k) \mathcal{I} - \mu \rho_0 - \sum_{i=1}^N \frac{| \chi_i > < \chi_i |}{\text{Tr}[| \chi_i > < \chi_i | \rho]} \tag{B.8}$$

angegeben werden.

Anhang C

Pseudoinversion von $H - E$

Ein pseudoinverser Operator A^\sharp zu einem linearen Operator A wird über:

$$A^\sharp A A^\sharp = A^\sharp \quad \text{und} \quad A A^\sharp A = A \quad (\text{C.1})$$

definiert. Also ist A auch der pseudoinverse Operator zu A^\sharp . Falls A invertierbar ist, ist A^{-1} der eindeutige pseudoinverse Operator. Außerdem kann gezeigt werden, daß jeder lineare Operator einen pseudoinversen Operator besitzt.

Die entscheidende Anwendung für pseudoinverse Operatoren liegt in der Auflösbarkeit von Gleichungen der Form $A |\chi\rangle = |b\rangle$ nach $|\chi\rangle$ für den Fall, daß A nicht invertierbar ist.

So sind die beiden folgenden Aussagen äquivalent [23]:

$$A |\chi\rangle = |b\rangle \quad \text{lösbar für } |\chi\rangle \text{ zu vorgegebenem } |b\rangle \iff A A^\sharp |b\rangle = |b\rangle. \quad (\text{C.2})$$

Ist der Fall aus (C.2) erfüllt, so ist $|\chi\rangle = |\lambda\rangle - A^\sharp A |\lambda\rangle + A^\sharp |b\rangle$ mit beliebigem $|\lambda\rangle$ die allgemeinste Lösung.

Im folgenden wird die Äquivalenz von (C.2) auf das Problem (5.20) angewandt. So ist der pseudoinverse Operator zu $A = H - E_n$ gegeben durch:

$$A^\sharp = \frac{1}{H - E_n} \left(\mathcal{I} - \frac{|\varphi_n\rangle \langle \varphi_n|}{\langle \varphi_n | \varphi_n \rangle} \right) = \sum_{j \neq n} \frac{1}{E_j - E_n} \frac{|\varphi_j\rangle \langle \varphi_j|}{\langle \varphi_j | \varphi_j \rangle}, \quad (\text{C.3})$$

denn mit $H = \sum_n \frac{|\varphi_n\rangle \langle \varphi_n|}{\langle \varphi_n | \varphi_n \rangle} E_n$ und dem Kommutator $[(H - E_n), |\varphi_n\rangle \langle \varphi_n|] = 0$ folgt

$$A A^\sharp A = (H - E_n) \frac{1}{H - E_n} \left(\mathcal{I} - \frac{|\varphi_n\rangle \langle \varphi_n|}{\langle \varphi_n | \varphi_n \rangle} \right) (H - E_n) = H - E_n = A \quad (\text{C.4})$$

und unter Ausnutzung der Projektoreigenschaften

$$A^\sharp A A^\sharp = \frac{1}{H - E_n} \left(\mathcal{I} - \frac{|\varphi_n\rangle \langle \varphi_n|}{\langle \varphi_n | \varphi_n \rangle} \right)^2 = \frac{1}{H - E_n} \left(\mathcal{I} - \frac{|\varphi_n\rangle \langle \varphi_n|}{\langle \varphi_n | \varphi_n \rangle} \right) = A^\sharp. \quad (\text{C.5})$$

Also sind die beiden Bedingungen (C.1) erfüllt.

Für das Problem aus (5.20) – Auflösung von

$$(H - E_n) |\frac{\delta \varphi_n}{\delta v(x'')} \rangle = |b\rangle \quad \text{mit} \quad |b\rangle = \left(\frac{\langle \varphi_n | x'' \rangle \langle x'' | \varphi_n \rangle}{\langle \varphi_n | \varphi_n \rangle} - |x'' \rangle \langle x''| \right) |\varphi_n\rangle \quad (\text{C.6})$$

nach $|\frac{\delta\varphi_n}{\delta v}| > -$ ist die rechte Seite von (C.2) wegen:

$$\begin{aligned} AA^\dagger |b\rangle &= \left(\mathcal{I} - \frac{|\varphi_n\rangle\langle\varphi_n|}{\langle\varphi_n|\varphi_n\rangle} \right) \left(\frac{\langle\varphi_n|x''\rangle\langle x''|\varphi_n\rangle}{\langle\varphi_n|\varphi_n\rangle} - |\langle x''|x''|\rangle \right) |\varphi_n\rangle \\ &= \left(\frac{\langle\varphi_n|x''\rangle\langle x''|\varphi_n\rangle}{\langle\varphi_n|\varphi_n\rangle} - |\langle x''|x''|\rangle \right) |\varphi_n\rangle = |b\rangle \end{aligned} \quad (\text{C.7})$$

erfüllt. Damit ergibt sich für die allgemeinste Lösung

$$\begin{aligned} |\frac{\delta\varphi_n}{\delta v(x'')}| &= |\lambda_n\rangle - A^\dagger A |\lambda_n\rangle + A^\dagger |b\rangle \\ &= \frac{\langle\varphi_n|\lambda_n\rangle}{\langle\varphi_n|\varphi_n\rangle} |\varphi_n\rangle + \\ &\quad \sum_{j \neq n} \frac{1}{E_j - E_n} \frac{|\varphi_j\rangle\langle\varphi_j|}{\langle\varphi_j|\varphi_j\rangle} \left(\frac{\langle x''|\varphi_n\rangle\langle x''|\varphi_n\rangle}{\langle\varphi_n|\varphi_n\rangle} - |\langle x''|x''|\rangle \right) |\varphi_n\rangle \\ &= \frac{\langle\varphi_n|\lambda_n\rangle}{\langle\varphi_n|\varphi_n\rangle} |\varphi_n\rangle - \sum_{j \neq n} \frac{1}{E_j - E_n} \frac{\langle\varphi_j|x''\rangle\langle x''|\varphi_n\rangle}{\langle\varphi_j|\varphi_j\rangle} |\varphi_j\rangle \end{aligned} \quad (\text{C.8})$$

mit beliebigen $|\lambda_n\rangle$ zu gewähltem $|\varphi_n\rangle$.

Anhang D

Hesse-Matrix im Tieftemperaturfall

Zur numerischen Lösung der Differentialgleichung (5.25) mittels Newton-Iteration wird die Hesse-Matrix des Funktionalen

$$F = \gamma \Gamma - \sum_{i=1}^N \ln \frac{\langle x_i | \varphi_0 \rangle \langle \varphi_0 | x_i \rangle}{\langle \varphi_0 | \varphi_0 \rangle} \quad (\text{D.1})$$

benötigt. Die Bestimmung ist mühsam, aber einfach.

Nach zweifacher Ableitung erhält man

$$\begin{aligned} \frac{\delta^2 F}{\delta v(x) \delta v(x')} &= \gamma \frac{\delta^2 \Gamma}{\delta v(x) \delta v(x')} - 2 \sum_{i=1}^N \operatorname{Re} \left\{ \frac{\langle x_i | \frac{\delta^2 \varphi_0}{\delta v \delta v'} \rangle}{\langle x_i | \varphi_0 \rangle} - \frac{\langle x_i | \frac{\delta \varphi_0}{\delta v'} \rangle \langle x_i | \frac{\delta \varphi_0}{\delta v} \rangle}{\langle x_i | \varphi_0 \rangle^2} \right. \\ &\quad \left. - \frac{\langle \varphi_0 | \frac{\delta^2 \varphi_0}{\delta v' \delta v} \rangle}{\langle \varphi_0 | \varphi_0 \rangle} - \frac{\langle \frac{\delta \varphi_0}{\delta v'} | \frac{\delta \varphi_0}{\delta v} \rangle}{\langle \varphi_0 | \varphi_0 \rangle} + \frac{\langle \varphi_0 | \frac{\delta \varphi_0}{\delta v} \rangle}{\langle \varphi_0 | \varphi_0 \rangle^2} \left[\langle \frac{\delta \varphi_0}{\delta v'} | \varphi_0 \rangle + \langle \varphi_0 | \frac{\delta \varphi_0}{\delta v'} \rangle \right] \right\}, \end{aligned} \quad (\text{D.2})$$

wobei die Abkürzung $\frac{\delta^2}{\delta v(x) \delta v(x')} = \frac{\delta^2}{\delta v \delta v'}$ verwendet wurde. Die Variationen zweiter Ordnung der Energieeigenfunktionen können durch erneute Ableitung der Eigenwertgleichung (5.16)

$$\frac{\delta^2 H}{\delta v \delta v'} | \varphi_0 \rangle + \frac{H}{\delta v} | \frac{\delta \varphi_0}{\delta v'} \rangle + \frac{\delta H}{\delta v'} | \frac{\delta \varphi_0}{\delta v} \rangle + H | \frac{\delta^2 \varphi_0}{\delta v \delta v'} \rangle = \quad (\text{D.3})$$

$$\frac{\delta^2 E_0}{\delta v \delta v'} | \varphi_0 \rangle + \frac{\delta E_0}{\delta v} | \frac{\delta \varphi_0}{\delta v'} \rangle + \frac{\delta E_0}{\delta v'} | \frac{\delta \varphi_0}{\delta v} \rangle + E_0 | \frac{\delta^2 \varphi_0}{\delta v \delta v'} \rangle \quad (\text{D.4})$$

gewonnen werden. Durch Ableitung von (5.17) folgt mit (C.8)

$$\frac{\delta^2 E_0}{\delta v \delta v'} = 2 \operatorname{Re} \left[\frac{\langle \frac{\delta \varphi_0}{\delta v'} | x \rangle \langle x | \varphi_0 \rangle}{\langle \varphi_0 | \varphi_0 \rangle} - \frac{\langle \varphi_0 | x \rangle \langle x | \varphi_0 \rangle \langle \frac{\delta \varphi_0}{\delta v'} | \varphi_0 \rangle}{\langle \varphi_0 | \varphi_0 \rangle^2} \right] \quad (\text{D.5})$$

$$= -2 \operatorname{Re} \left[\frac{\langle x | \varphi_0 \rangle \langle \varphi_0 | x' \rangle}{\langle \varphi_0 | \varphi_0 \rangle} \sum_{j \neq 0} \frac{\langle \varphi_j | x \rangle \langle x' | \varphi_j \rangle}{(E_j - E_0) \langle \varphi_j | \varphi_j \rangle} \right]. \quad (\text{D.6})$$

Nach Auflösung von (D.3) und mit $\frac{\delta^2 H}{\delta v \delta v'} = 0$ nach

$$(H - E_0) | \frac{\delta^2 \varphi_0}{\delta v \delta v'} \rangle = \frac{\delta^2 E_0}{\delta v \delta v'} | \varphi_0 \rangle + \left(\frac{\delta E_0}{\delta v} - \frac{\delta H}{\delta v} \right) | \frac{\delta \varphi_0}{\delta v'} \rangle + \left(\frac{\delta E_0}{\delta v'} - \frac{\delta H}{\delta v'} \right) | \frac{\delta \varphi_0}{\delta v} \rangle \quad (\text{D.7})$$

gelingt mittels erneuter Pseudoinversion die Auflösung nach den Orbitalen $|\frac{\delta^2 \varphi_0}{\delta v \delta v'}\rangle$. Dafür werden die Ergebnisse für $|\frac{\delta \varphi_0}{\delta v}\rangle$ und $|\frac{\delta \varphi_0}{\delta v'}\rangle$ aus Anhang C auf der rechten Seite von (D.7) eingesetzt und

$$\begin{aligned} \left(\frac{\delta E_0}{\delta v} - \frac{\delta H}{\delta v} \right) |\frac{\delta \varphi_0}{\delta v'}\rangle &= \frac{\langle \varphi_0 | x \rangle \langle x | \varphi_0 \rangle \langle \varphi_0 | \lambda \rangle}{\langle \varphi_0 | \varphi_0 \rangle^2} |\varphi_0\rangle \\ &- \frac{\langle \varphi_0 | x \rangle \langle x | \varphi_0 \rangle}{\langle \varphi_0 | \varphi_0 \rangle} \sum_{j \neq 0} \frac{1}{E_j - E_0} |\varphi_j\rangle \frac{\langle \varphi_j | x' \rangle \langle x' | \varphi_0 \rangle}{\langle \varphi_j | \varphi_j \rangle} \\ &- |x\rangle \frac{\langle \varphi_0 | \lambda \rangle \langle x | \varphi_0 \rangle}{\langle \varphi_0 | \varphi_0 \rangle} + |x\rangle \sum_{j \neq 0} \frac{1}{E_j - E_0} \frac{\langle x | \varphi_j \rangle \langle \varphi_j | x' \rangle \langle x' | \varphi_0 \rangle}{\langle \varphi_j | \varphi_j \rangle} \end{aligned} \quad (\text{D.8})$$

und die Gleichung mit $v \leftrightarrow v'$ und $x \leftrightarrow x'$ Austausch ausgerechnet. Insgesamt führt dies auf:

$$\begin{aligned} |\frac{\delta^2 \varphi_0}{\delta v \delta v'}\rangle &= \frac{\langle \varphi_0 | \tilde{\lambda} \rangle}{\langle \varphi_0 | \varphi_0 \rangle} |\varphi_0\rangle \\ &+ \left\{ - \sum_{k \neq 0} \frac{1}{E_k - E_0} |\varphi_k\rangle \frac{\langle \varphi_k | x \rangle \langle \varphi_0 | \lambda \rangle \langle x | \varphi_0 \rangle}{\langle \varphi_0 | \varphi_0 \rangle \langle \varphi_k | \varphi_k \rangle} \right. \\ &- \frac{\langle \varphi_0 | x \rangle \langle x | \varphi_0 \rangle}{\langle \varphi_0 | \varphi_0 \rangle} \sum_{j \neq 0} \frac{1}{(E_j - E_0)^2} |\varphi_j\rangle \frac{\langle \varphi_j | x' \rangle \langle x' | \varphi_0 \rangle}{\langle \varphi_j | \varphi_j \rangle} \\ &+ \sum_{j,k \neq 0} \frac{1}{E_j - E_0} \frac{1}{E_k - E_0} |\varphi_k\rangle \frac{\langle \varphi_k | x \rangle \langle x | \varphi_j \rangle \langle \varphi_j | x' \rangle \langle x' | \varphi_0 \rangle}{\langle \varphi_j | \varphi_j \rangle \langle \varphi_k | \varphi_k \rangle} \\ &\left. + x, x' - \text{Austausch} \right\} \end{aligned} \quad (\text{D.9})$$

und die entsprechend adjungierte Gleichung. Setzt man die Ableitungen in die Gleichung (D.2) ein, so fallen die Abhängigkeiten von λ , λ' und $\tilde{\lambda}$ heraus, und man erhält das Endergebnis

$$\frac{\delta^2 F}{\delta v \delta v'} = \gamma \frac{\delta^2 \Gamma}{\delta v \delta v'} - 2 \text{Re} \sum_{i=1}^N \left(a(x, x') + b_i(x, x') + c_i(x, x') + d_i(x, x') \right) \quad (\text{D.10})$$

mit den einzelnen Summanden:

$$a(x, x') = - \sum_{j \neq 0} \frac{1}{(E_j - E_0)^2} \frac{\langle \varphi_0 | x \rangle \langle x | \varphi_j \rangle \langle \varphi_j | x' \rangle \langle x' | \varphi_0 \rangle}{\langle \varphi_0 | \varphi_0 \rangle \langle \varphi_j | \varphi_j \rangle}, \quad (\text{D.11})$$

$$\begin{aligned} b_i(x, x') &= - \sum_{j \neq 0} \frac{1}{(E_j - E_0)^2} \frac{\langle x | \varphi_0 \rangle \langle x' | \varphi_0 \rangle \langle x_i | \varphi_j \rangle}{\langle \varphi_j | \varphi_j \rangle \langle \varphi_0 | \varphi_0 \rangle \langle x_i | \varphi_0 \rangle} \times \\ &\left(\langle \varphi_0 | x \rangle \langle \varphi_j | x' \rangle + \langle \varphi_0 | x' \rangle \langle \varphi_j | x \rangle \right), \end{aligned} \quad (\text{D.12})$$

$$\begin{aligned} c_i(x, x') &= \sum_{j,k \neq 0} \frac{1}{E_j - E_0} \frac{1}{E_k - E_0} \frac{\langle x_i | \varphi_k \rangle}{\langle x_i | \varphi_0 \rangle \langle \varphi_k | \varphi_k \rangle \langle \varphi_k | \varphi_j \rangle} \times \\ &\left(\langle \varphi_k | x \rangle \langle x' | \varphi_0 \rangle \langle x | \varphi_j \rangle \langle \varphi_j | x' \rangle \right. \\ &\left. + \langle \varphi_k | x' \rangle \langle x | \varphi_0 \rangle \langle x' | \varphi_j \rangle \langle \varphi_j | x \rangle \right) \end{aligned} \quad (\text{D.13})$$

und

$$d_i(x, x') = - \sum_{j,k \neq 0} \frac{< x_i | \varphi_j > < x_i | \varphi_k > < \varphi_j | x > < x | \varphi_0 > < \varphi_k | x' > < x' | \varphi_0 >}{(E_j - E_0)(E_k - E_0) < \varphi_j | \varphi_j > < \varphi_k | \varphi_k > < x_i | \varphi_0 >^2}. \quad (\text{D.14})$$

Anhang E

N -fache Ableitung der Zustandssumme

Der Vollständigkeit halber sei hier noch die N -fache Variationsableitung der Zustandssumme vorgeführt. So folgt durch Verallgemeinerung von (6.18):

$$\begin{aligned} \frac{\delta^N Z}{\delta v(x_1) \cdots \delta v(x_N)} &= \left(-\frac{1}{\hbar}\right)^N \int_0^{\beta\hbar} dt_1 \cdots dt_N \int_{q(0)=q(\beta\hbar)} \frac{[dq(t)]}{\mathcal{N}} e^{-\frac{1}{\hbar}S[q]} \prod_{i=1}^N \delta(q(t_i) - x_i) \quad (\text{E.1}) \\ &= \left(-\frac{1}{\hbar}\right)^N \int dx \int_0^{\beta\hbar} dt_1 \cdots dt_N \int_{\substack{q(0)=q(\beta\hbar)=x \\ q(t_i)=x_i \forall i}} \frac{[dq(t)]}{\mathcal{N}} e^{-\frac{1}{\hbar}S[q]}. \end{aligned}$$

Das Ergebnis (E.1) lässt sich kompakt weiterverarbeiten, indem man die Pfadintegrale analog zu (6.22) zerlegt. Der Zeitordnungsoperator \mathcal{T} ist erforderlich, um die richtige Reihenfolge der Integrationsvariablen einzuhalten. Diese Notwendigkeit kann durch Verallgemeinerung der Abbildung 6.2 auf den vorliegenden Fall nachvollzogen werden. Mit der Verabredung $t_0 = 0$, $t_{N+1} = \beta\hbar$ und $x_0 = x_{N+1} = x$ folgt:

$$\begin{aligned} \frac{\delta^N Z}{\delta v(x_1) \cdots \delta v(x_N)} &= \left(-\frac{1}{\hbar}\right)^N \int dx \int dt_1 \cdots dt_N \mathcal{T} \prod_{i=0}^N \langle x_{i+1} | e^{-(t_{i+1}-t_i)\frac{H}{\hbar}} | x_i \rangle \\ &= \left(-\frac{1}{\hbar}\right)^N \int dt_1 \cdots dt_N \text{Tr} \left[e^{-\beta H} \mathcal{T} \left(\prod_{i=1}^N \langle x_i(t_i) | x_i(t_i) \rangle \right) \right], \quad (\text{E.2}) \end{aligned}$$

wobei im zweiten Schritt die Zeitentwicklung von $|x_i\rangle$ mittels

$$|x_i(t_i)\rangle \langle x_i(t_i)| = e^{\frac{t_i}{\hbar}H} |x_i\rangle \langle x_i| e^{-\frac{t_i}{\hbar}H} \quad (\text{E.3})$$

abgekürzt wurde. Im Tieftemperaturfall fällt die Spur auf den Beitrag des Grundzustands $|\varphi_0\rangle$ von H zusammen, und man erhält:

$$\frac{\delta^N Z}{\delta v(x_1) \cdots \delta v(x_N)} \longrightarrow \left(-\frac{1}{\hbar}\right)^N \int dt_1 \cdots dt_N \langle \varphi_0 | \mathcal{T} \left(\prod_{i=1}^N \langle x_i(t_i) | x_i(t_i) \rangle \right) | \varphi_0 \rangle e^{-\beta E_0} \quad (\text{E.4})$$

für $\beta \rightarrow \infty$.

Im Fall $N = 1$ folgt aus (E.2) durch Wegfallen von \mathcal{T} und Ausnutzen der Vollständigkeit sofort

(6.20). Das Ergebnis von (6.22) kann man mit

$$\frac{\delta}{\delta v(x')} \langle x_i | e^{-\beta H} | x_i \rangle = -\frac{1}{\beta} \frac{\delta^2 Z}{\delta v(x') \delta v(x_i)} \quad (\text{E.5})$$

mittels (E.2) herleiten, indem man die Zeitordnung explizit durch den Einbau der Stufenfunktion $\theta(t)$ gewährleistet:

$$\begin{aligned} \frac{\delta^2 Z}{\delta v(x_1) \delta v(x_2)} &= \frac{1}{\hbar^2} \int dx \int_0^{\beta\hbar} dt_1 \int_0^{\beta\hbar} dt_2 \times \\ &\left[\theta(t_2 - t_1) \langle x | e^{-(\beta - \frac{t_2}{\hbar})H} | x_2 \rangle \langle x_2 | e^{-(t_2 - t_1)\frac{H}{\hbar}} | x_1 \rangle \langle x_1 | e^{-\frac{t_1}{\hbar}H} | x \rangle \right. \\ &\left. + \theta(t_1 - t_2) \langle x | e^{-(\beta - \frac{t_1}{\hbar})H} | x_1 \rangle \langle x_1 | e^{-(t_1 - t_2)\frac{H}{\hbar}} | x_2 \rangle \langle x_2 | e^{-\frac{t_2}{\hbar}H} | x \rangle \right]. \end{aligned} \quad (\text{E.6})$$

Das x -Integral verschwindet in beiden Summanden direkt mit der Vollständigkeitsrelation. Nach Aufspaltung in die beiden Integrale, Umbenennung $t_1 \leftrightarrow t_2$ im zweiten Integral, Umparametrisieren von t_2 mittels $t' = t_2 - t_1$ im ersten Integral und $t' = \beta\hbar - t_2 + t_1$ im zweiten Integral erhält man

$$\begin{aligned} \frac{\delta^2 Z}{\delta v(x_1) \delta v(x_2)} &= \frac{1}{\hbar^2} \left(\int_0^{\beta\hbar} dt_1 \int_{-t_1}^{\beta\hbar - t_1} dt' \theta(t') \langle x_1 | e^{-(\beta - \frac{t'}{\hbar})H} | x_2 \rangle \langle x_2 | e^{-\frac{t'}{\hbar}H} | x_1 \rangle + \right. \\ &\quad \left. \int_0^{\beta\hbar} dt_1 \int_{t_1}^{\beta\hbar + t_1} dt' \theta(\beta\hbar - t') \langle x_1 | e^{-(\beta - \frac{t'}{\hbar})H} | x_2 \rangle \langle x_2 | e^{-\frac{t'}{\hbar}H} | x_1 \rangle \right). \end{aligned}$$

Nun können die θ -Bedingungen ausgeführt und die Integrale wieder zusammengefaßt werden. Man erhält

$$\begin{aligned} \frac{\delta^2 Z}{\delta v(x_1) \delta v(x_2)} &= \frac{1}{\hbar^2} \int_0^{\beta\hbar} dt_1 \int_0^{\beta\hbar} dt' \langle x_1 | e^{-(\beta - \frac{t'}{\hbar})H} | x_2 \rangle \langle x_2 | e^{-\frac{t'}{\hbar}H} | x_1 \rangle \\ &= \frac{\beta}{\hbar} \int_0^{\beta\hbar} dt' \langle x_1 | e^{-(\beta - \frac{t'}{\hbar})H} | x_2 \rangle \langle x_2 | e^{-\frac{t'}{\hbar}H} | x_1 \rangle \end{aligned} \quad (\text{E.7})$$

und somit mit (E.5) das Ergebnis (6.22).

Literaturverzeichnis

- [1] Lemm,J.C., Uhlig,J., Weiguny,A.: Phys.Rev.Lett. **84**, 2068 (2000)
- [2] Lemm,J.C., Uhlig,J.: Hartree–Fock–Approximation for Inverse Many–Body Problems. nucl-th/9908056 (erscheint demnächst in Physical Review Letters)
- [3] Bishop,C.M.: Neural Networks for Pattern Recognition. Oxford: Clarendon Press 1995
- [4] Berger,J.O.: Statistical Decision Theory and Bayesian Analysis. New York: Springer 1980
- [5] Lemm,J.C.: Bayesian Field Theory. physics/9912005
- [6] Lemm,J.C.: How to Implement A Priori Information: A Statistical Mechanics Approach. cond-mat/9808039
- [7] Ballentine,L.E.: Quantum Mechanics. New Jersey: Prentice Hall 1990
- [8] Balian,R.: From Microphysics to Macrophysics (Volume 1). Berlin: Springer 1982
- [9] Bertsekas,D.P.: Nonlinear Programming. Belmont, MA: Athena Scientific 1995
- [10] Bialek,W., Callan,C.G., Strong,S.P.: Phys.Rev.Lett. **77**, 4693 (1996)
- [11] Bužek,V., Drobny,G., Adam,G., Derka,R., Knight,P.L.: Reconstruction of quantum states of spin systems via the Jaynes principle of maximum entropy. quant-ph/9701038
- [12] Derka,R., Bužek,V., Adam,G., Knight,P.L.: From quantum Bayesian inference to quantum tomography. quant-ph/9701029
- [13] Gildener,E., Patrascioiu,A.: Phys.Rev. **D16**, 423 (1977)
- [14] Zinn–Justin,J.: Quantum Field Theory and Critical Phenomena. Oxford: Clarendon Press 1996
- [15] Schulman,L.S.: Techniques and Applications of Path Integration. New York: John Wiley 1981
- [16] Negele,J.P., Orland,H.: Quantum Many–Particle Systems. Massachusetts: Addison–Wesley 1981
- [17] Römer,H., Filk,T.: Statistische Mechanik. Weinheim: VCH 1994
- [18] Dashen,R.F., Hasslacher,B., Neveu,A.: Phys. Rev. **D10**, 4114 (1974)
- [19] Levit,S., Negele,J.W., Paltiel,Z.: Phys. Rev. **C22**, 1979 (1980)
- [20] Blaizot,J.-P., Ripka,G.: Quantum Theory of Finite Systems. Cambridge,Massachusetts: MIT Press 1986

- [21] Kerman,A.K., Levit,S., Troudet,T: Ann.Phys. **148**, 436 (1982)
- [22] Uhlig,J., Lemm,J.C., Weiguny,A.: Eur.Phys.J. **A2**, 343 (1998)
- [23] Koecher,M.: Lineare Algebra und analytische Geometrie. Berlin: Springer 1985
- [24] Lemm,J.C.: Inverse Time-Dependent Quantum Mechanics. quant-ph/0002010

Lebenslauf

Name	Joerg Uhlig
geboren	am 07.01.1972 in Münster (Westf.)
Familienstand	ledig
Eltern	Dr. Horst Uhlig Gertrud Uhlig, geb. Longerich
Schulbildung	1978–1982 Paul–Gerhardt–Schule, Münster 1982–1991 Immanuel–Kant–Gymnasium, Münster
Hochschulreife	Abitur am 28.05.1991 in Münster
Studium	Immatrikulation an der Westfälischen–Wilhelms–Universität Münster am 21.08.1991 im Fachbereich Physik
Prüfungen	Physik–Diplom am 08.12.1996 an der Westfälischen–Wilhelms–Universität Münster
Beginn der Dissertation	im Januar 1997 am Institut für Theoretische Physik I der Westfälischen–Wilhelms–Universität Münster bei Prof. Dr. A. Weiguny
Tätigkeiten	Studentische Hilfskraft (01.11.95–31.12.96) Wissenschaftliche Hilfskraft (01.01.97–28.02.97) Wissenschaftlicher Mitarbeiter (01.03.97–Sommer 2000) am Institut für Theoretische Physik I der Westfälischen–Wilhelms–Universität Münster

Bei den beiden Betreuern dieser Arbeit, Prof. Dr. A. Weiguny und Dr. J.C. Lemm, möchte ich mich bedanken — vor allem dafür, daß sie mir die Freiheit für die Wahl des Themengebietes gegeben haben.

Ich danke meinen Eltern für die finanzielle, psychologische und chromosomale Unterstützung.

