Vorlesung Quanteninformationstheorie

Andreas Knauf*

Wintersemester 2001/2002

Zusammenfassung

Vorlesungsbegleitendes Skript. Anregungen und Kritik sind willkommen! Version 13. Februar 2002

Inhaltsverzeichnis

*Mathematisches Institut, Universität Erlangen-Nürnberg, Bismarckstr. $1\frac{1}{2}$, D–91054 Erlangen, Germany. e-mail: knauf@mi.uni-erlangen.de

1 Übersicht

Physikalische Grundlagen der Informationsverarbeitung

Während sich die Computertechnologie stetig änderte, blieb das Konzept der *Turingmaschine* (dem mathematischen Gegenstück realer Computer) von diesen Veränderungen weitgehend unberührt.

Die Turingmaschine modelliert den Vorgang der Berechnung und bildet damit eine Grundlage für die Definition von Berechenbarkeit und Entscheidbarkeit.

Gleichzeitig erlaubt sie die Einführung von Maßen für die Komplexität einer Berechnung.

Für diese Theorie ist es nun unwesentlich, ob ein Bit durch ein Loch in einer Lochkarte oder die Magnetisierung einer Magnetschicht repräsentiert wird.

Erst durch diese — im Begriff der Turingmaschine codierten — Unabhängigkeit der Theorie von den physikalisch-technischen Grundlagen konnte sie sich als mathematische Disziplin etablieren. Scheinbar also reiht sie sich in das allgemeine Schema moderner Mathematik ein, die autonom von ihren Anwendungen ihre eigenen Grundlagen schafft.

In diesem Sinn ist der Titel dieser *mathematischen* Vorlesung irritierend, denn er verbindet eine physikalische Theorie — die Quantenmechanik — mit dem abstrakt mathematischen Begriff der Information.

Dabei soll es *nicht* um die angewandt-mathematische Frage gehen, wie sich mit Hilfe der Quantenmechanik Eigenschaften von elektronischen Bauteilen voraussagen und optimieren lassen. Stattdessen wird im Mittelpunkt der Vorlesung die Frage stehen, inwieweit zwei Theorien — Automaten- und Informationstheorie und Quantenmechanik — miteinander vereinbar sind, und welche Veränderungen am vorhandenen Modell von Berechnung vorgenommen werden müssen, um eine solche Vereinbarkeit sicherzustellen.

Das Mooresche Gesetz

Ein Motiv für diese Fragestellung ist die Tatsache, dass die Miniaturisierung der Computerbauteile heute eine quantenmechanische Beschreibung ihrer Funktion erfordert.

Das empirische *Mooresche Gesetz* stellt fest, dass wichtige Kenngrößen von Computern sich exponentiell in der Zeit verändern.

- So halbiert sich die für eine logische Operation angewandte *Energie* ca. alle 15 Monate, während sich
- die Zahl der zur Darstellung eines Bits benötigten *Atome* ca. alle 13 Monate halbiert.

Eine Trendverlängerung ergibt für das Jahr 2020

- eine Energie, die der thermischen Energie eines Atoms bei Zimmertemperatur entspräche,

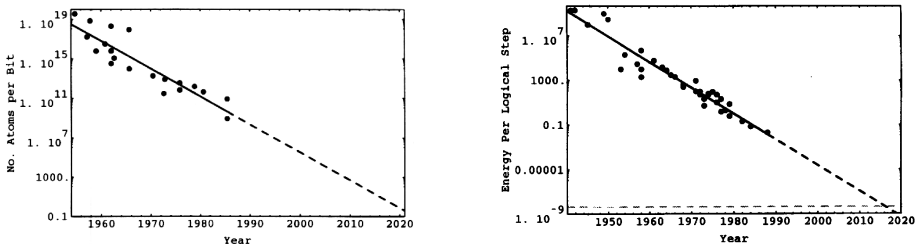


Abbildung 1: Zahl der zur Codierung eines Bits benötigten Atome (links); bei einer logischen Operation verbrauchte Energie (rechts), aus [WC]

- während zur Darstellung eines Bits nur ein Atom benötigt würde [WC].

Die Fragestellung, ob eine solche Miniaturisierung denkbar ist, könnte man also das "year 2020 problem" nennen.

Sie ist nicht rein akademisch, wie man am Beispiel der DNS sehen kann, bei der die Informationen durch Nukleotide gespeichert werden, und der Energieaufwand pro Nukleotid bei dem 20 - 100-fachen des atomar-thermischen Rauschens liegt.

Reversibilität

Es stellt sich hier zunächst die Frage, ob eine physikalische *untere Schranke* für die zur Berechnung benötigte Energie existiert. Betrachten wir dazu eine logische Verknüpfung

$$F : B \times B \to B$$

Dabei bezeichnet das *Bit* $B := \{w, f\}$ die Möglichkeiten 'wahr und falsch'. Die Wahl $F(w, w) := w$ und $F(w, f) := F(f, w) := F(f, f) := f$ codiert die 'und'-Verknüpfung.

Da F eine vierelementige Frage in eine zweielementige Menge abbildet, kann F nicht injektiv sein. Daher geht bei Berechnung von $F(a, b)$ und Löschen der beiden Argumente irreversibel Information verloren, und zwar im Mittel über die möglichen Eingaben mindestens ein Bit.

Nach physikalischen Gesetzen entspricht dies einer Erzeugung von Energie in Form von Wärme. John von Neumann berechnete diese minimal benötigte Energie 1949 auf $kT \ln 2 \approx 3 \cdot 10^{-21}$ Joule bei Zimmertemperatur (k ist die Boltzmannkonstante, T die Absolute Temperatur).

Abbildung 2: Links: Container mit vielen Gasatomen (irrreversibel). Rechts: Container mit wenigen Gasatomen (reversibel).

Beide fundamentalen Theorien zur Beschreibung der Materie — klassische Mechanik und Quantenmechanik — besitzen eine mikroskopisch *reversible Dynamik*, d.h. der Anfangszustand lässt sich aus einem späteren Zustand im Prinzip mit beliebiger Genauigkeit zurückberechnen. Makroskopisch dagegen kann die Dynamik irreversibel sein (siehe Abb. 2). Es stellt sich also die Frage nach der Möglichkeit einer *reversiblen Informationsverarbeitung*, die der mikroskopischen Reversibilität der Dynamik angepasst ist.

Zu diesem Zweck identifizieren wir die zweielementige Menge B mit dem Restklassenkörper $\mathbb{Z}_2 := \mathbb{Z}/2\mathbb{Z}$ mit den beiden Repräsentanten $\mathbb{Z}_2 = \{0, 1\}$ (und der Addition und Multiplikation modulo 2). Die 1 entspreche dabei dem Wert $w \in B$, 0 dem Wert f.

Wir betten die logische Verknüpfung $F : B^2 \to B$ nun in die Funktion

$$\hat{F} : B^3 \to B^3 \quad , \quad \hat{F}(a, b, c) := (a, b, F(a, b) + c)$$

ein. Diese ist zu sich selbst invers, d.h. $\hat{F} \circ \hat{F} = \mathrm{Id}$, und damit insbsesondere invertierbar, also reversibel[1].

Es wurde nun einerseits gezeigt, dass ein solcher reversibler Computer durch eine geeignete Anordnung von Billiardkugeln und Banden *prinzipiell* als System der klassischen Mechanik *realisierbar* ist.

Quantenmechanische Berechnung

Ebenso ist er, wie sich zeigen läßt, prinzipiell durch ein Quantensystem realisierbar, eben einen Quantencomputer.

Allerdings kann diese hypothetische Maschine im Prinzip Berechnungen auf einem anderen als dem durch die klassische Logik vorgegebenen Weg durchführen.

Dies liegt daran, dass die Quantensysteme mehr Zustände als ihnen analoge klassische Systeme annehmen können. Dies zeigt sich schon im Vergleich

[1] Für $F(a, b) := a \cdot b$, d.h. 'oder', ist $\hat{F}$ unter dem Namen *Toffoli-Gatter* bekannt

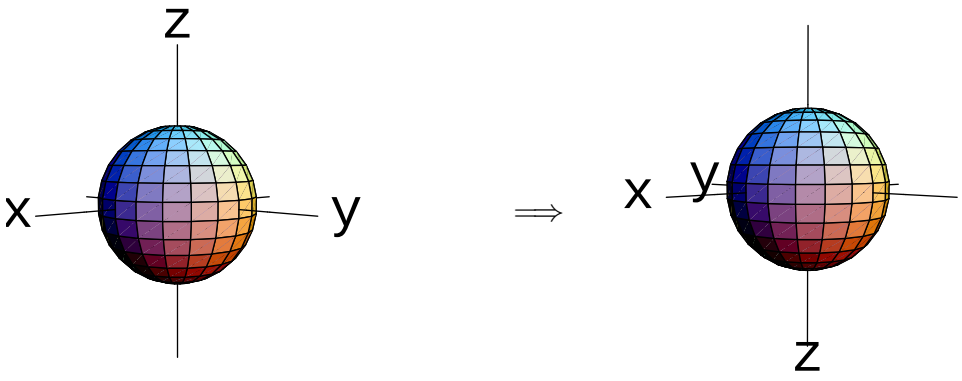


Abbildung 3: Bewegung eines quantenmechanischen Spins. Bsp. 1: Rotation um x–Achse mit Winkel $180°$. Diese vertauscht Nord- und Südpol und entspricht auf B einer Negation.

zwischen der klassischen Informationseinheit — dem Bit — und seinem quantenmechanischen Analogon, dem sog. *Qubit* $\mathcal{B}$.

Allgemein entsprechen den reinen Zuständen eines quantenmechanischen Systems eindimensionale Unterräume eines Hilbertraums $\mathcal{H}$ über den komplexen Zahlen. Im einfachsten Fall, eben dem des Qubits, ist $\mathcal{H} := \mathbb{C}^2$, der zweidimensionale komplexe Hilbertraum. Physikalisch entspricht dies den Einstellmöglichkeiten eines Teilchens mit Spin $1/2$.

Ein eindimensionaler Unterraum $V \subseteq \mathcal{H}$ entspricht damit einem Punkt des projektiven Raumes $\mathbb{C}P_1$: Spannt ein Vektor $v = \left(\begin{smallmatrix} v_1 \\ v_2 \end{smallmatrix}\right) \in \mathcal{H} \setminus \{0\}$ diesen Unterraum auf, dann besitzen genau diejenigen Vektoren $w = \left(\begin{smallmatrix} w_1 \\ w_2 \end{smallmatrix}\right) \in \mathcal{H} \setminus \{0\}$, die in U liegen das gleiche Verhältnis $w_1/w_2 = v_1/v_2 \in \mathbb{C} \cup \{\infty\}$ seiner Komponenten wie v. Daher besitzt der projektive Raum $\mathbb{C}P_1$ die Form der Einpunktkompaktifizierung der komplexen Ebene, oder (unter Benutzung der stereographischen Projektion) der Sphäre $S^2 = \{x \in \mathbb{R}^3 \mid \|x\| = 1\}$.

Ein Qubit ist also die Menge $\mathcal{B} \cong S^2$, und das Bit B ist als Nord- und Südpol in diese Kugeloberfläche eingebettet: $B \cong \left\{ \left(\begin{smallmatrix} 0 \\ 0 \\ 1 \end{smallmatrix}\right), \left(\begin{smallmatrix} 0 \\ 0 \\ -1 \end{smallmatrix}\right) \right\} \subseteq S^2$. Während die Menge B nur die identische Abbildung und die Negation als Bijektionen zuläßt, gibt es viele Isometrien von $\mathcal{B} \cong S^2$, und diese lassen sich prinzipiell zu Berechnungen nutzen, siehe Abbildungen 3 und 4.

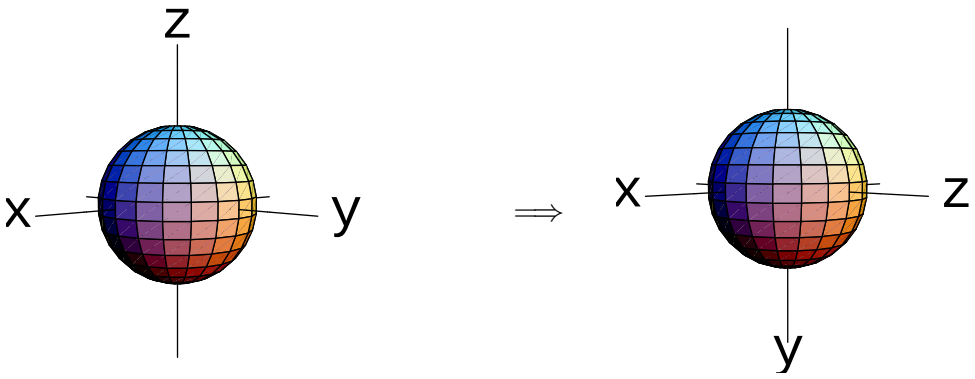


Abbildung 4: Bsp. 2: Rotation um x–Achse mit Winkel $90°$. Dies entspricht keiner klassischen Berechnung, sondern $\sqrt{\neg}$, weil zweifache Anwendung dieser Rotation eine Rotation um die$x$–Achse mit Winkel $180°$ ergibt.

Das entscheidende Motiv, sich mit Quantencomputern zu beschäftigen, ist die Existenz quantenmechanischer Algorithmen, deren Laufzeit wesentlich geringer als die der bekannten klassischen Algorithmen ist. So zeigte Shor 1994, dass sich eine natürliche Zahl in einer Zeit in ihre Primfaktoren zerlegen lässt, die nur polynomial mit der in Bits gemessenen Größe dieser Zahl anwächst, siehe [Sh2]. Ein klassischer Algorithmus mit der entsprechenden Eigenschaft ist nicht bekannt.

Fehlerkorrektur

Kehren wir zum erwähnten reversiblen Modell der klassischen Berechnung zurück. Dort würden sich beim Stoß der Billiardkugeln Fehler in ihrer Ausrichtung verstärken, sodass ohne eine *Korrektur* dieser Fehler an eine praktische Realisierung nicht zu denken ist.

In der (klassischen) Informationstheorie ist ein reichhaltiges Instrumentarium zur Korrektur von Fehlern entwickelt worden. Einfachstes Beispiel ist der

Beispiel 1.1 Wiederholungscode. Codiert wird durch Verdreifachung der Nachricht:

$$C : B \to B^3 \quad , \quad C(a) := (a, a, a)$$

mit Decodierung

$$D : B^3 \to B \quad , \quad D(a_1, a_2, a_3) := \begin{cases} 0 & \text{, höchstens ein } a_k = 1 \\ 1 & \text{, höchstens ein } a_k = 0 \end{cases} .$$

Die Decodierung besteht also in einem Mehrheitsentscheid. Tritt zwischen Codierung und Decodierung höchstens ein Fehler auf, so kann dieser korrigiert werden.

Offensichtlich enthält eine solche Fehlerkorrektur ein irreversibles Element, denn die Decodierung D ist keine injektive Abbildung. D.h.: Physikalisch erzeugt Decodierung Wärme.

Ein weiteres Problem stellt sich im Kontext der Quantenmechanik. Dem Bit B können wir als zweielementige Menge die kanonische Basis $e_0 := \binom{1}{0}$, $e_1 := \binom{0}{1}$ von $\mathcal{H}$ zuordnen. N Qubits entsprechen den eindimensionalen Unterräumen im N–fachen Tensorprodukt $\mathcal{H}^N := \mathcal{H} \otimes \ldots \otimes \mathcal{H}$, einem komplexen Vektorraum der Dimension 2^N.

In der Quantenmechanik entsprechen den *Observablen* selbstadjungierte Operatoren, und deren Eigenwerte (allgemeiner: Spektralwerte) tauchen als Messwerte auf. Nach der Messung ist der quantenmechanische Zustand Eigenvektor zum entsprechenden Messwert (genauer: der von diesem aufgespannte Unterraum).

Insbesondere bedeutet dies, dass der quantenmechanische Zustand durch die Messung verändert wird.

Beispiel 1.2 Wiederholungs-Code. Hier wird der Hilbertraum $\mathcal{H}^3$ von drei Qubits verwandt, und die Decodierung erfolgt durch den Operator $D \in \mathcal{B}(\mathcal{H}^3)$, der bez. der Orthonormalbasis $e_{a_1 a_2 a_3} := e_{a_1} \otimes e_{a_2} \otimes e_{a_3}$ des Tensorproduktes in lexikalischer Ordnung $(e_{000}, e_{001}, e_{010}, e_{011}, e_{100}, e_{101}, e_{110}, e_{111})$ die Diagonalgestalt

$$D = \mathrm{diag}(0, 0, 0, 1, 0, 1, 1, 1)$$

besitzt.

Nun ist in diesem Beispiel der Eigenraum zum Eigenwert 0 genauso wie der zum Eigenwert 1 vierdimensional. Nach der Messung von 0 wissen wir damit, dass der quantenmechanische Zustandsvektor sich in einem der beiden Unterräume des 8-dimensionalen Hilbertraumes $\mathcal{H}^3$ befindet, denn die Messung hat den Zustand entsprechend verändert.

Quantencomputer sind gegenüber ihren klassischen Pendants gerade deshalb potentiell im Vorteil sind, weil es viel mehr N–Qubit-Zustände als N–Bit-Zustände gibt.

Die für eine Fehlerkorrektur notwendige *Messung* reduziert aber den Zustand wie im obigen Beispiel auf einen Unterraum hoher Codimension, und droht damit, den potentiellen Vorteil von Quantencomputern zunichtezumachen.

Gleichzeitig ist wegen der quantenmechanisch größeren Zahl der Zustände auch die Zahl der möglichen Fehler größer als in der klassischen Informationstheorie.

Daher ist die Entdeckung eines *quantenfehlerkorrigierenden Codes* 1995 durch Peter Shor und andere bedeutend, denn dieser ist in der Lage, Fehler zu korrigieren, ohne gleichzeitig die codierte Information zu zerstören, siehe [Sh1].

Vorlesungsthemen

Wir sehen, dass eine Vorlesung über Quantencomputer als mathematische Theorie die Quantenmechanik, Automatentheorie und Informationstheorie benutzt.

Diese Vorlesung richtet sich nicht nur an Mathematiker, sondern auch an Physiker und Informatiker. Entsprechend werden die genannten Theorien nicht vorausgesetzt, sondern, soweit für diese Vorlesung notwendig, bereitgestellt.

Im zweiten Kapitel werde ich den Begriff der *Turingmaschine* definieren, und in Kapitel 3 etwas zu *Komplexitätsmaßen* sagen.

In Kapitel 4 werden Klassische und Quantenmechanik axiomatisch eingeführt, und einige Unterschiede beleuchtet.

Die in Kapitel 5 diskutierten *Zustände* eines physikalischen Systems beschreiben unsere Kenntnis dieses Systems und ordnen observablen Größen Erwartungswerte zu, die den Mittelwerten vieler Messungen dieser Observablen entsprechen. Die *Entropie* des Zustandes misst unsere Unkenntnis, und ist ein Zentralbegriff der physikalisch basierten Informationstheorie.

Wie in Kapitel 6 gezeigt wird, entspricht quantenmechanisch vollständige Kenntnis eines Gesamtsystems nicht notwendig vollständiger Kenntnis seiner Teile. Diese Tatsache hat in der Klassischen Physik kein Pendant, und ist Voraussetzung für die besonderen Eigenschaften von Quantencomputern.

Der *Suchalgorithmus* von Grover wird in Kap. 7 vorgestellt.

In einem späteren Kapitel soll auch den *Shor-Algorithmus* zur Faktorisierung natürlicher Zahlen vorgeführt werden.

Was die Funktionsweise von Quantenrechnern angeht, wird die *Shannonsche Informationstheorie* und ihre quantenmechanische Verallgemeinerung Thema von Kapitel 9 bzw. 10. Klassische bzw. quantenmechanische fehlerkorrigierende Codes werden in Kapitel 11 und 12 behandelt.

Weitere Themen sollen der Aufbau eines Quantencomputers aus elementaren unitären Schaltelementen und die formale Definition einer *Quanten-Turingmaschine* sein. Es wird sich zeigen, dass Quantenberechnungen zumindest mit einem sog. *Orakel* qualitativ schneller als entsprechende klassische Berechnungen sein können.

In einer zweistündigen Vorlesung kann ein so umfangreiches Programm nur oberflächlich abgearbeitet werden. Ziel ist es, eine Einführung in die Literatur zu geben; leider werden die Definitionen gegenüber Sätzen überwiegen, und in vielen Fällen werde ich statt eines Beweises nur eine Literaturstelle angeben.

2 Turingmaschinen

Zu Beginn des zwanzigsten Jahrhunderts führten die Paradoxa der naiven Mengenlehre zu einer verstärkten Beschäftigung mit den Grundlagen der Mathematik. Ein Resultat dieser Bemühungen war die Formalisierung intuitiver Begriffe wie dem des Algorithmus.

Intuitiv verstehen wir unter einem *Algorithmus*, beispielsweise zur Berechnung einer zahlentheoretischen Funktion $f : \mathbb{N}_0^r \to \mathbb{N}_0$, eine endlich beschreibbare Methode, für beliebige $(a_1, \ldots, a_r) \in \mathbb{N}_0^r$ in endlich vielen Schritten determiniert das Ergebnis $f(a_1, \ldots, a_r)$ zu berechnen.

Bemerkungen 2.1

1. Ein Beispiel für eine solche zahlentheoretische Funktion ist die Summe: $f : \mathbb{N}_0^2 \to \mathbb{N}_0$, $(a, b) \mapsto a + b$. Die übliche Implementation der Addition in Computern ist im obigen Sinn kein Algorithmus, denn Definitions- und Wertebereich sind ja beschränkt. Daher wird die Turingmaschine als mathematische Idealisierung des Computers als Ein-und Ausgabemedium wie auch als Speicher ein unendlich langes Band besitzen.

2. Ich habe mich hier auf die Berechnung zahlentheoretischer Funktionen bezogen. Allgemeiner können wir an Algorithmen denken, die auf Zeichenreihen operieren. Z.B. gibt es einen Algorithmus zur alphabetischen Ordnung einer beliebig langen Namensliste.

Da man alles mit Zahlen codieren kann, könnten wir ohne Verlust an Allgemeinheit bei den zahlentheoretischen Algorithmen bleiben.

Wir müssen bei der Codierung nur sicherstellen, dass diese und die Decodierung effektiv durchführbar sind, und dass wir die Codewörter als solche erkennen können.

Der mathematische Begriff der *Turingmaschine* ermöglicht nun eine Formalisierung der intuitiven Vorstellung von Algorithmen. Obwohl Turing seine Arbeit 1936, also kurz vor dem Bau der ersten Computer, veröffentlichte, kommt die Turingmaschine unserer Vorstellung eines Computerprogrammes recht nahe.

Da kein Grund besteht, sich auf Turingmaschinen zur Berechnung zahlentheoretischer Funktionen zu beschränken, definieren wir allgemein:

Definition 2.2 • Ein *Alphabet* ist eine endliche, nicht leere Menge $A = \{a_1, \ldots, a_k\}$, deren Elemente a_l *Buchstaben* heißen.
• Ein *Wort der Länge* $n \in \mathbb{N}$ ist ein Element $w = (w_1, \ldots, w_n) \in A^n$. Es gibt ein ausgezeichnetes *leeres Wort* ε der *Länge* 0, und die *Hülle* $A^* := \cup_{n=0}^{\infty} A^k$ (mit $A^0 := \{\varepsilon\}$) bezeichnet die Menge aller Wörter.
• Die Teilmengen von A^* werden auch *Sprachen* (über dem Alphabet A) genannt.

Notation 2.3 Der Einfachheit halber schreiben wir statt $(w_1, \ldots, w_n) \in A^n$ die nicht durch Kommas abgetrennte Buchstabenfolge $w_1 w_2 \ldots w_n$, und wir bezeichnen mit $|w| := n$ die Länge des Wortes.

Bezüglich der Konkatentation $\circ$ mit $a \circ b$ mit $v \circ w := v_1 v_2 \ldots v_m w_1 w_2 \ldots w_n$ von $v \in A^m$, $w \in A^n$ ist A^* ein *Monoid*, denn die Verknüpfung ist assoziativ und besitzt das neutrale Element ε. Natürlich ist $\circ$ für ein mehrbuchstabiges Alphabet nicht kommutativ.

Notation 2.4 Für $a \in A$ setzen wir $a^0 := \varepsilon$ und $a^n := a \circ a^{n-1}$, $n \in \mathbb{N}$.

Die heuristische Vorstellung von einer Turingmaschine ist die eines Automaten mit endlich vielen Zuständen, der als Speicher ein unendlich langes Band besitzt. Er liest in jedem Arbeitsschritt ein Zeichen. Abhängig von seinem Zustand und diesem Zeichen schreibt er mit dem Schreib-Lese-Kopf ein Zeichen, geht zu einem neuen Zustand über, und verschiebt den Schreib-Lese-Kopf um eine Einheit nach links (-1) bzw. rechts ($+1$), siehe Abb. 5.

Da das ganze Band beschrieben ist, Wörter aber endliche Länge besitzen, ist die Bandinschrift kein Wort. Wir vereinbaren aber, dass nur endlich viele

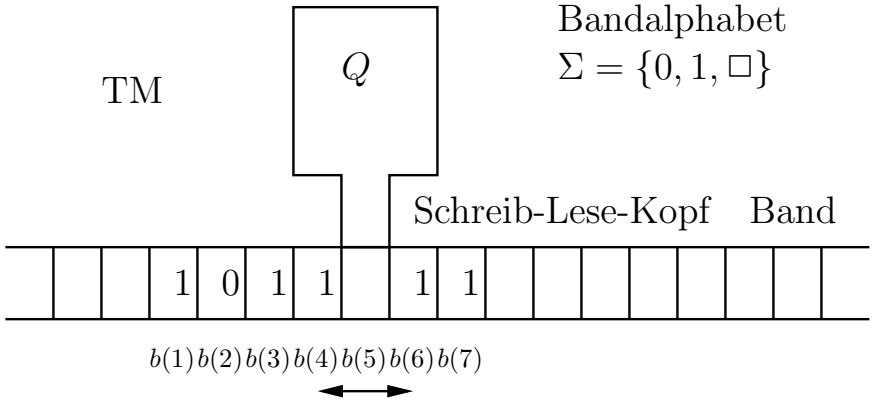


Abbildung 5: Turingmaschine TM mit Bandinschrift b

Felder des Bandes vom Leersymbol $\square$ (analog der Leertaste der Schreibmaschine) verschieden sind.

Definition 2.5 Eine (*deterministische*) *Turingmaschine* $\mathrm{TM} := (\Sigma, Q, \delta)$ besteht aus:

- dem Alphabet Σ, genannt *Bandalphabet*, mit $\square \in \Sigma$.
- der endlichen *Zustandsmenge* Q. Diese enthält einen ausgezeichneten *Anfangszustand* q_i und *Endzustand* q_f, mit $q_f \neq q_i$.
- der *Übergangsfunktion*

  $$\delta : Q \times \Sigma \to Q \times \Sigma \times \{-1, 1\},$$

  und wir schreiben $\delta = (\delta_1, \delta_2, \delta_3)$.

Eine *Konfiguration einer Turingmaschine* von TM ist ein Element von $K := Q \times B \times \mathbb{Z}$, wobei

$$B := \{b : \mathbb{Z} \to \Sigma \mid b(i) \neq \square \text{ nur für endlich viele } i \in \mathbb{Z}\}$$

die Menge der *Bandinschriften* ist.

- $(q, b, j) \in K$ heißt *Anfangskonfiguration*, wenn $q = q_i$, $j = 0$ und $b(i) = \square$ fur $i < 0$ ist (also der Anfangszustand angenommen wird, der Bandkopf sich in Position 0 befindet, und links von ihm nur Leerzeichen stehen).

- $(q, b, j) \in K$ heißt *Endkonfiguration*, wenn $q = q_f$ (also der Endzustand angenommen wird).
- Die *Nachfolgekonfiguration* $N(k)$ der Konfiguration $k = (q, b, j) \in K$ ist

  $$N(k) := (\delta_1(q, b(j)), b', j + \delta_3(q, b(j)))$$

  mit $b'(j) := \delta_2(q, b(j))$ und $b'(i) := b(i)$ für $i \neq j$.

Anschaulich entspricht die Übergangsfunktion einem Computerprogramm, während eine Konfiguration eine Momentaufnahme der gesamten Turingmaschine darstellt.

Beispiel 2.6 Wir wollen eine Turingmaschine konstruieren, die die Funktion $f : \mathbb{N}_0 \to \mathbb{N}_0$, $n \mapsto n + 1$ berechnet, also eine beliebige natürliche Zahl um Eins vergrößert.

- Codieren wir n binär als $n = \sum_{i=0}^{k} n_i 2^i$ mit $n_i \in \{0, 1\}$, dann kommen wir mit dem Alphabet $\Sigma := \{0, 1, \square\}$ aus.
- Die Anfangskonfiguration ist also

  $$b = \ldots \square\square \overset{q_i}{n_k}\, n_{k-1} \ldots n_0 \square\square \ldots$$

  wobei ich über die Bandinschrift an der Stelle $j = 0$ den Anfangszustand q_i der Turingmaschine gesetzt habe.
- Zunächst soll sich der Bandkopf der Turingmaschine bis zur Stelle $k + 1$ nach rechts vortasten, auf dem das erste rechte Leerzeichen steht. Danach sollte der übliche Additionsalgorithmus beginnen. Entsprechend wählen wir die Zustandsmenge $Q := \{q_i, q_a, q_f\}$ und die Übergangsfunktion

δ	0	1	$\square$
q_i	$(q_i, 0, 1)$	$(q_i, 1, 1)$	$(q_a, \square, -1)$
q_a	$(q_f, 1, -1)$	$(q_a, 0, -1)$	$(q_f, 1, -1)$
q_f			

Für $n = 5$, also Anfangskonfiguration $b = \ldots \square\square \overset{q_i}{1}\, 01\square\square \ldots$, ergibt sich die Folge

$$\square \overset{q_i}{1}\, 01\square \mapsto \square 1 \overset{q_i}{0}\, 1\square \mapsto \square 10 \overset{q_i}{1}\, \square \mapsto \square 101 \overset{q_i}{\square} \mapsto$$

$$\mapsto \square 10 \overset{q_a}{1} \square \mapsto \square 1 \overset{q_a}{0} 0\square \mapsto \square \overset{q_f}{1} 10\square$$

von Konfigurationen. Hierbei habe ich den Zähler j der Bandposition einfach durch den Ort des Zustandes markiert.

In der Endkonfiguration codiert die Bandinschrift $f(5) = 6$.

Bemerkungen 2.7 1. Ich habe die Werte der Übergangsfunktion δ für den Endzustand q_f nicht definiert, denn die Turingmaschine soll stoppen, wenn q_f erreicht wird. Bequemlichkeitshalber kann man δ auch für Argumente, die nicht erreicht werden können, undefiniert lassen.

Daher verlangt man oft nur, dass δ eine partielle Funktion ist. (Eine *partielle Funktion* $f : A \to B$ ordnet nur den Elementen x eines *Definitionsbereichs* $D \subseteq A$ Werte $f(x) \in B$ zu. f heißt *total*, wenn $D = A$).

2. Wie in der Programmierung werden in der Automatentheorie Techniken entwickelt, die das Notieren von Algorithmen erleichtern.

So ist es eine nützliche Übung, eine Turingmaschine zu entwerfen, die zwei (binär codierte) Zahlen $a = \sum_{i=0}^{k} a_i 2^i$ und $b = \sum_{i=0}^{l} b_i 2^i$ addiert, also ausgehend von der Bandinschrift

$$\ldots \square\square a_k a_{k-1} \ldots a_0 \square b_l b_{l-1} \ldots b_0 \square\square \ldots$$

die Summe $c := a + b = \sum_{i=0}^{m} c_i 2^i$ berechnet, die c_i aufs Band schreibt und dann stoppt. Hier ist es sinnvoll, die Bit-Stellen zu markieren, die gerade bearbeitet werden. Technisch kann dies durch die Vergrößerung des Bandalphabetes geschehen: Statt $\Sigma := \{0, 1, \square\}$ kann man etwa $\Sigma \times \{\square, *\}$ benutzen, und die Stellen mit dem Stern kennzeichnen.

Ebenso bieten sich Unterprogrammtechniken an, bei denen etwa die obige Addition der Eins wiederverwendet wird.

Turingmaschinen sollen nicht reale Computerprogramme ersetzen (dazu wären sie zu primitiv), sondern beweistechnische Hilfsmittel sein.

Definition 2.8 • Die Turingmaschine TM *akzeptiert die Eingabe* $b_i \in B$, falls für ein $t \in \mathbb{N}$ die t–te Konfiguration $N^t(q_i, b, 0)$ von der Form $(q_f, b_f, j_f) \in K$ (mit beliebiger *Ausgabe* $b_f = b_f(b_i) \in B$ und Bandposition $j_f \in \mathbb{Z}$) ist.

• Das kleinste solche t heißt dann die *Laufzeit von* TM *für die Eingabe* b_i.
• Setze $\hat{\Sigma} := \Sigma \setminus \{\square\}$ und

$$I : \hat{\Sigma}^* \to B \ , \ I(w_0 \dots w_k)(l) := \begin{cases} w_l & 0 \le l \le k \\ \square & \text{sonst} \end{cases} .$$

Dann heißt die Menge

$$L(\mathrm{TM}) := \{w \in \hat{\Sigma}^* \mid \mathrm{TM} \text{ akzeptiert } I(w)\}$$

von Worten aus $\hat{\Sigma}^*$ die *von* TM *akzeptierte Sprache* .
• TM *berechnet* die partielle Funktion $f : \hat{\Sigma}^* \to \hat{\Sigma}^*$, wenn sie die Worte $w \in D$ aus dem Definitionsbereich $D \subseteq \hat{\Sigma}^*$ akzeptiert, und die Ausgabe $b_f = I(f(w))$ schreibt, andere Eingaben $I(w)$, $w \notin D$ aber nicht akzeptiert.
• f heißt *partiell rekursiv*, wenn es eine Turingmaschine TM gibt, die f berechnet.
• Ist f außerdem total, so heißt f *rekursiv*.
• Eine Sprache $L \subseteq \Sigma^*$ heißt *entscheidbar*, wenn ihre charakteristische Funktion $\mathbb{1}_L : \Sigma^* \to \{0, 1\}$ (mit $\mathbb{1}_L(w) = 1$ genau dann wenn $w \in L$) rekursiv ist.

Natürlich setzen wir voraus, dass das Alphabet $\hat{\Sigma}$ nicht leer ist.

Satz 2.9 *Die Menge der Funktionen $f : \hat{\Sigma}^* \to \hat{\Sigma}^*$ ist überabzählbar[2], während die Menge der rekursiven Funktionen abzählbar ist.*

Bew.: Im einfachsten Fall $\hat{\Sigma} = \{a\}$ ist die Hülle von der Form $\hat{\Sigma}^* = \{a^k \mid k \in \mathbb{N}_0\}$, man kann die Worte a^k also durch ihre Länge k abzählen. Damit betrachten wir die Menge der arithmetischen Funktionen $f : \mathbb{N}_0 \to \mathbb{N}_0$. Aber schon die Menge der Funktionen $f : \mathbb{N}_0 \to \{0, 1\}$ ist überabzählbar, wie das bekannte Diagonalargument zeigt. Wäre nämlich $n \mapsto f_n$ eine Abzählung, dann ist $g : \mathbb{N}_0 \to \{0, 1\}$ mit $g(n) := 1 - f_n(n)$ von allen f_n verschieden.

Andererseits ist die Menge der rekursiven, d.h. Turing-berechenbaren Funktionen abzählbar, da wir die Turingmaschinen selbst (bis auf Isomorphie) abzählen können. Denn die Zahl von Turingmaschinen mit einer gegebenen Menge $Q = \{q_1, \dots, q_n\}$ von Zuständen übersteigt nicht die (endliche) Anzahl von Übergangsfunktionen $\delta : Q \times \Sigma \to Q \times \Sigma \times \{-1, 1\}$. □

Die partiell-rekursiven Funktionen stellt man sich nun als diejenigen partiellen Funktionen vor, für die ein Algorithmus existiert.

[2]**Def.:** Eine Menge M heißt *abzählbar*, falls eine injektive Abbildung $M \to \mathbb{N}$ existiert, sonst *überabzählbar*.

Bemerkung 2.10 Man könnte annehmen, dass die nicht partiell-rekursiven Funktionen auch praktisch nicht von Interesse sind, sodass Satz 2.9 keine Beschränkung der Berechenkeit relevanter Funktionen darstellt. Dies ist jedoch nicht der Fall.

Ein bekanntes Beispiel bildet das *Halteproblem*. Dies besteht darin, zu entscheiden, ob eine vorgegebene Turingmaschine TM eine vorgegebene Wort w als Eingabe akzeptiert. Wir können die Turingmaschine TM durch ein Wort v effektiv codieren. Den beiden hintereinander geschriebenen (durch das Zeichen $*$ getrennten) Wörtern entspricht die Bandinschrift $I(v * w)$. Die Klasse aller solcher Wörter $v * w$ ist eine Sprache.

Es läßt sich nun zeigen, dass diese Sprache *nicht entscheidbar* ist. Insbesondere kann es kein Computerprogramm geben, das überprüft, ob Arbeitsprogramme bei gegebener Eingabe von selbst anhalten. Ein solches Programm wäre von praktischem Nutzen, weil man mit ihm verhindern könnte, dass ein fehlprogrammiertes Arbeitsprogramm in einer Endlosschleife sinnlos vor sich hin rechnet.

Es gibt sicher gute Gründe, anzuzweifeln, dass ausgerechnet die Def. 2.5 den intuitiven Algorithmus-Begriff formalisiert. Tatsächlich findet man in der Literatur fast so viele verschiedene Definitionen des Begriffes 'Turingmaschine' wie Lehrbücher über Automatentheorie. Außerdem sind Dutzende von Varianten erfunden worden: Turingmaschinen mit mehreren Bandköpfen oder Bändern, mit mehrdimensionalen Bändern, mit *random access memory*, Zellenautomaten, nicht deterministische Turingmaschinen etc.

Immer ließ sich aber zeigen, dass diese die gleiche Klasse von Funktionen berechnen wie unser Modell. Dies hat den Glauben in die folgende heuristische These verstärkt:

Church-Turing-These: Jede 'im intuitiven Sinn berechenbare' Funktion ist Turing-berechenbar.

Diese These ist nun sicher keine Vermutung im mathematischen Sinn, da sie den undefinierten Begriff der 'im intuitiven Sinn berechenbaren' Funktion enthält.

Andererseits wurde sie von D. Deutsch (1985) so umformuliert, das sie beinahe zu einer mathematischen Vermutung über eine Eigenschaft der physikalischen Grundgesetze wurde:

Modifizierte Church-Turing-These: Jede durch physikalische Prozesse berechenbare Funktion ist Turing-berechenbar.

Wir können nun fragen, welche Berechnungen im Rahmen physikalischer Theorien

wie der klassischen Mechanik oder der Quantenmechanik möglich sind. Natürlich bleibt dabei immer noch offen, was wir noch als Berechnung bezeichnen wollen (z.B. ob die Evolution der Vögel die aerodynamisch günstige Form der Flügel 'berechnet').

Es ist aber möglich, *Modelle* physikalischer Berechnung zu definieren (z.B. ein System von Billiardkugeln und Banden mit Interpretation von An- bzw. Abwesenheit von Kugeln als Wahrheitswerten), und dann die Klasse der durch ein solches Modell berechenbaren Funktionen zu untersuchen.

Ein solches Modell *quantenmechanischer* Berechnung ist die Quanten-Turingmaschine, und für sie lässt sich die obige modifizierte Church-Turing-These zeigen. Sie kann also genau die (partiell) rekursiven Funktionen berechnen. In diesem Sinn ist die Church-Turing-These noch einmal untermauert worden.

3 Komplexitätstheorie

Die Komplexität einer Berechnung kann durch verschiedene Resourcen quantifiziert werden: die Laufzeit des Rechners, den benötigten Speicherplatz etc. Diese Größen sind offensichtlich abhängig vom benutzten Computer und von der Güte des Algorithmus.

Außerdem lässt sich schnell einsehen, dass die eigentliche Fragestellung die nach der Komplexität einer *Funktion* ist, und nicht die nach der Komplexität der Berechnung eines oder mehrerer Funktions*werte*.

Beispiel 3.1 Die Zugehörigkeit zur Menge $\mathbb{P} \subseteq \mathbb{N}$ der Primzahlen wird durch die charakteristische Funktion $\mathbb{1}_{\mathbb{P}} : \mathbb{N} \to \{0,1\}$ von $\mathbb{P}$ (mit $\mathbb{1}_{\mathbb{P}}(n) = 1$ genau dann wenn $n \in \mathbb{P}$) codiert.

Eine einfache (aber nicht sehr schnelle) Turingmaschine zur Berechnung der *Funktion* $\mathbb{1}_{\mathbb{P}}$ liest die Zahl n. Falls $n > 2$ ist, testet sie auf Division ohne Rest durch die Zahlen $l = 2, \ldots, \lfloor\sqrt{n}\rfloor$, indem sie $n \mod l$ berechnet.

Eine Turingmaschine zur Berechnung eines einzelnen Funktions*wertes*, etwa von $\mathbb{1}_{\mathbb{P}}(255811)$, schreibt dagegen einfach die 0 und stoppt dann (denn $255811 = 491 \times 521$).

Es ist klar, dass jede Turingmaschine, die eine Funktion f berechnet, eine obere Schranke für die zur Berechnung von f nötigen Resourcen liefert. Dagegen sind untere Schranken i.A. schwerer erhältlich.

Weiter ist zu erwarten, dass (im Gegensatz zum letzten Kapitel) Schranken an die Laufzeit von der genauen Definition des Begriffes 'Turingmaschine' abhängen.

Beispiel 3.2 Da bei der in Bem. 2.7.2 angesprochenen Addition zweier natürlicher Zahlen $a = \sum_{i=0}^{k} a_i 2^i$ und $b = \sum_{i=0}^{k} b_i 2^i$ die Turingmaschine sich nicht alle Bits merken kann, muss der Bandkopf hin- und herlaufen, sodass die Laufzeit stärker als k wächst.

Lässt man jedoch Turingmaschinen mit mehr als einem Band zu, dann lässt sich eine solche finden, deren Laufzeit $\mathcal{O}(k)$ ist.

Wir beschränken uns im Folgenden auf die Komplexität von charakteristischen Funktionen $\mathbb{1}_L : \hat{\Sigma}^* \to \{0,1\}$ einer Sprache $L \subseteq \hat{\Sigma}^*$ über einem Alphabet $\hat{\Sigma}$. Für die Zwecke dieser Vorlesung verzichten wir auch darauf, Turingmaschinen mit mehreren Bändern zuzulassen[3], und definieren:

Definition 3.3 Eine Turingmaschine TM akzeptiere die Sprache L, und es sei $\mathsf{Lauf}(\mathrm{TM}, w)$ die Laufzeit von TM bei Eingabe $I(w)$.
• Für $T : \mathbb{N}_0 \to \mathbb{R}$ heißt TM *T–zeitbeschränkt*, wenn

$$\max_{w \in L, |w|=n} \mathsf{Lauf}(\mathrm{TM}, w) \leq T(n) \qquad (n \in \mathbb{N}_0).$$

• Die Sprache L besitzt dann *Zeitkomplexität* T, und die Familie der Sprachen mit Zeitkomplexität T wird mit $\mathsf{DZeit}(T)$ bezeichnet.
• Die Sprachfamilie der in *deterministisch polynomialer Zeit* akzeptierten Sprachen wird mit

$$\mathcal{P} := \bigcup_{T \text{ Polynom}} \mathsf{DZeit}(T)$$

bezeichnet.

Eine *nichtdeterministische Turingmaschine* NDTM wird ähnlich wie in Def. 2.5 definiert, wobei die Übergangsfunktion eine (totale) Funktion

$$\delta : Q \times \Sigma \to \mathcal{P}(Q \times \Sigma \times \{-1, 1\}),$$

ist, also $(q, w) \in Q \times \Sigma$ eine *Teilmenge* von $Q \times \Sigma \times \{-1, 1\}$ zuordnet, die Menge aller *möglichen* Übergänge.

[3] Diese können von solchen mit nur einem Band unter quadratischem Zeitverlust simuliert werden, siehe [HU].

In Abwandlung von Def. 2.8 *akzeptiert* eine NDTM genau dann eine Eingabe, wenn es eine Folge von Konfigurationen *gibt*, die zu einem Endzustand führt.

In Analogie mit Def. 3.3 wird die Familie der Sprachen mit nichtdeterministischer Zeitkomplexität T mit NZeit(T) bezeichnet, und die Sprachfamilie der in *nichtdeterministisch polynomialer Zeit* akzeptierten Sprachen besitzt das Symbol

$$\mathcal{NP} := \bigcup_{T \text{ Polynom}} \mathsf{NZeit}(T).$$

Es gilt $\mathcal{P} \subseteq \mathcal{NP}$, denn einer TM mit Übergangsfunktion $(q, a) \mapsto \delta(q, a)$ entspricht die NDTM mit Übergangsfunktion $(q, a) \mapsto \{\delta(q, a)\}$, und die Laufzeiten sind identisch.

Der Beweis der (1971 aufgestellten) Vermutung $\mathcal{P} \neq \mathcal{NP}$ ist dagegen eine der wichtigsten ungelösten Fragen der Mathematik[4].

Bemerkungen 3.4

1. Eine Möglichkeit, sich eine nichtdeterministische Turingmaschine NDTM vorzustellen, besteht darin, die bei gegebener Bandinschrift realisierbaren Folgen von Konfigurationen in Form eines sich in der Zeit verzweigenden Baumes anzuordnen (siehe Anhang A). Dabei sind die Knoten des Baumes Elemente von $Q \times \Sigma$. Die Wurzel dieses Baumes ist $(q_i, b(0))$ (siehe Def. 2.5). An jedem Knoten (q, a) des Baumes zweigen dann höchstens

   $$A := |Q \times \Sigma \times \{-1, 1\}|$$

   Kanten ab, entsprechend dem Wert $\delta(q, a)$ der Übergangsfunktion.

   Man kann nun zeigen, dass sich die NDTM mit höchstens exponentiellem Zeitverlust durch eine deterministische Turingmaschine TM simulieren lässt. Genauer existiert für ein $\mathcal{NP}$–Problem eine TM , die die Sprache in der Zeit 2^p akzeptiert, wobei p ein Polynom ist. Diese TM simuliert dabei alle Wege von der Wurzel zu den Blättern im Wurzelbaum der NDTM . In diesem Sinn sind die $\mathcal{NP}$–Probleme solche mit höchstens exponentieller Laufzeit.

2. Nichtdeterministische Turingmaschine sind nicht in gleicher Weise realisierbar wie die deterministischen. Falls ein realer Computer bei gleicher Eingabe zu verschiedenen Ergebnissen kommt, muss er repariert werden.

[4] siehe `http://www.claymath.org/prizeproblems/pvsnp.htm`

Trotzdem ist die durch die NDTM definierte Sprachfamilie $\mathcal{NP}$ von großer praktischer Bedeutung: Oft handelt es sich hier um Probleme, deren Lösbarkeit in polynomialer Zeit *bestätigt* werden kann. Hier entsprechen die verschiedenen Wege von der Wurzel zu einem Blatt den Kandidaten für eine Lösung des Problems.

Beispiel 3.5 1. Für den Primzahltest aus Bsp. 3.1 ist kein deterministisch polynomialer Algorithmus bekannt, es ist also nicht bekannt, ob die Sprache $\mathbb{P} \subseteq \{0,1\}^*$ der (binär codierten) Primzahlen zur Sprachklasse $\mathcal{P}$ gehört. Da die binäre Darstellung von $n \in \mathbb{N}$ genau $\lceil \log_2(n+1) \rceil$ Bits benötigt, wächst die maximale Rechenzeit des 'Probieralgorithmus' 3.1 exponentiell, also mehr als polynomial in dieser Größe.[5]

Dagegen ist das *Primzahlproblem* in $\mathcal{NP}$, denn wir können in bez. der Bitzahl von $n \in \mathbb{N}$ polynomialer Zeit alle als Faktor von n in Frage kommenden Zahlen $l \in \{2, \ldots, \lfloor \sqrt{n} \rfloor\}$ nichtdeterministisch erzeugen, und jeweils durch Berechnung von $n \mod l$ auf die Teilereigenschaft testen.

2. Manche Graphen besitzen einen geschlossenen Kantenzug, in dem jeder Knoten des Graphen genau einmal auftritt (sog. *Hamilton-Zyklus*), siehe Abb. 6.

   Es ist kein in der Zahl der Knoten des Graphen polynomialer Algorithmus bekannt, der entscheidet, ob ein solcher Hamilton-Zyklus existiert. Wir wissen also nicht, ob das Problem in $\mathcal{P}$ liegt.

   Wohl aber können wir in einer in $|V|$ polynomialer Zeit deterministisch überprüfen, ob ein vorgegebener Kantenzug die gewünschte Eigenschaft besitzt, und damit gegebenenfalls den Beweis der Hamilton-Eigenschaft führen. Andererseits ist es auch möglich, in polynomialer Zeit nichtdeterministisch alle in Frage kommenden $2^{|V| \cdot (|V|-1)/2}$ Kantenmengen E zu generieren.

   Damit ist das Problem in $\mathcal{NP}$.

Da die Familie der Polynome unter Addition, Multiplikation und Komposition abgeschlossen ist, haben die Komplexitätsklassen $\mathcal{P}$ und $\mathcal{NP}$ entsprechende Abgeschlossenheitseigenschaften bez. Hintereinanderausführung, Unterprogrammaufruf etc.

[5] Der beste bekannte deterministsche Primzahltest hat eine Laufzeit der Ordnung $(\log n)^{\log\log\log n}$, ist also nur knapp superpolynomial [CLR].

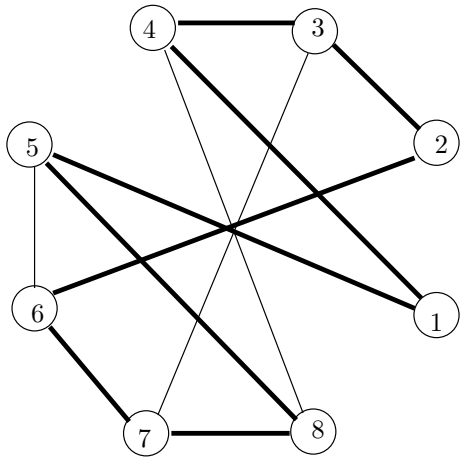


Abbildung 6: Hamilton-Zyklus: $1 \to 4 \to 3 \to 2 \to 6 \to 7 \to 8 \to 5$

Die Komplexitätsklasse $\mathcal{P}$ wird als die Klasse der praktisch handhabbaren Probleme angesehen. Man mag das für Algorithmen, deren Laufzeit wie die hundertste Potenz der Eingabelänge divergiert, bezweifeln, aber viele in der Praxis auftretende polynomiale Algorithmen haben Laufzeiten mit viel niedrigeren Potenzen.

Es ist für viele Zwecke sinnvoll, sich die Klasse $\mathcal{NP}$ als die derjenigen Probleme vorzustellen, für die man eine angebotene Lösung in polynomialer Zeit *verifizieren* kann.

Wenn auch das $\mathcal{P} \neq \mathcal{NP}$–Problem ungelöst ist, ist doch abstrakt bekannt, dass es *beliebig komplexe* rekursive Funktionen bzw. Sprachen gibt:

Satz 3.6 *Es sei $T : \mathbb{N}_0 \to \mathbb{N}_0$ eine rekursive Funktion. Dann gibt es eine rekursive Sprache L, die nicht in der Sprachklasse DZeit(T) liegt.*

Bew.: Da T rekursiv ist, existiert eine immer haltende Turingmaschine TM_T, die T berechnet.

Es sei w_i, $i \in \mathbb{N}_0$ das i–te Wort in der *kanonischen Anordnung*

$$\varepsilon, 0, 1, 00, 01, 10, 11, 000, 001, \ldots$$

nach Wortlänge und lexikographischer Ordnung der Hülle $\hat{\Sigma}^*$ von $\hat{\Sigma} := \{0, 1\}$.

Man beachte, dass sich die Nummer eines Wortes $w = a_1 \dots a_n \in \hat{\Sigma}^*$ der Länge n durch $w = w_i$ mit $i = 2^n - 1 + \sum_{l=1}^{n} a_l 2^{n-l}$ einfach berechnen lässt.

Weiter sei TM_i die i–te Turingmaschine (bezüglich einer effektiven Abzählung nach Zahl der Zustände und Übergangsfunktion).

Wir setzen

$$L := \left\{ w_i \in \hat{\Sigma}^* \mid i \in \mathbb{N}_0, \mathrm{TM}_i \text{ akzeptiert nicht } w_i \text{ in der Laufzeit } T(|w_i|) \right\}.$$

Wäre $L \in \mathrm{DZeit}(T)$, dann müsste eine T–zeitbeschränkte Turingmaschine TM_j existieren, die L akzeptiert. Dafür aber dürfte das Wort w_j nicht in der Sprache L liegen. Dann aber würde es von TM_j nicht akzeptiert, denn kein Wort in $\hat{\Sigma}^* \setminus L$ wird von der L erkennenden Turingmaschine L akzeptiert. Widerspruch!

Um zu zeigen, dass L rekursiv ist, muss man eine Turingmaschine TM_L bauen, die L akzeptiert:

- TM_L liest ein Wort w und berechnet seine Nummer i.
- Durch Aufruf der Turingmaschine TM_T als Unterprogramm berechnet sie die Laufzeit $T(|w|)$.
- Sie berechnet TM_i, und simuliert TM_i mit Eingabe w_i. Akzeptiert TM_i das Wort w in der Zeit $T(|w|)$, so akzeptiert TM_L es nicht, und umgekehrt. □

Entsprechendes lässt sich für die Sprachfamilie $\mathrm{NZeit}(T)$ zeigen.

Es ist also insbesondere sicher, dass $\mathcal{P}$ nicht alle rekursiven Sprachen umfasst. Analog zum letzten Kapitel wird die Bedeutung dieser Sprachklasse durch die folgende heuristische These unterstrichen:

Quantitative Church-Turing-These: Jede durch physikalische Prozesse in polynomialer Zeit berechenbare Funktion ist auch polynomial Turing-berechenbar.

Ob allerdings diese These wahr ist, kann bezweifelt werden, denn das Beispiel des Shor-Algorithmus zeigt, dass es Funktionen gibt, die quantenmechanisch in polynomialer Zeit berechnet werden können, während klassisch wie erwähnt kein polynomialer deterministischer Algorithmus bekannt ist.

4 Klassische Mechanik und Quantenmechanik

In diesem Kapitel werde ich kurz die Strukturen dieser beiden physikalischen Theorien beschreiben. Während diese oberflächlich gesehen so unterschiedlich sind, dass beide Theorien als unvergleichbar erscheinen, wird der Begriff der C^*–Algebra ermöglichen, Gemeinsamkeiten und Unterschiede herauszuarbeiten.

4.1 Klassische Mechanik

Ein System der Klassischen Mechanik wird üblicherweise durch folgende Daten gegeben:

- Einer d–dimensionalen Mannigfaltigkeit P, dem *Phasenraum*. Im einfachsten Fall ist $P = \mathbb{R}^d$.

- Einem *Vektorfeld* $v : P \to TP$ (das jedem Punkt $x \in P$ einen Tangentialvektor $v(x)$ bei x an P zuordnet). Diesem ist das *dynamische System* $\dot{x} = v(x)$ zugeordnet, d.h. in lokalen Koordinaten $x_1, \ldots, x_d$ das Differentialgleichungssystem

  $$\dot{x}_1 = v_1(x) \quad , \quad \ldots \quad , \quad \dot{x}_d = v_d(x).$$

  Spielen Reibungseffekte keine Rolle, dann wird v durch eine 'Hamiltonfunktion' $H : P \to \mathbb{R}$ fixiert. Im einfachsten Fall ist P eine offene Teilmenge des $\mathbb{R}^d$ mit $d = 2n$. In kanonischen Koordinaten $(p_1, \ldots, p_n, q_1, \ldots, q_n)$ des $\mathbb{R}^d = \mathbb{R}^n \times \mathbb{R}^n$ gelten dann die *Hamiltonschen Gleichungen*

  $$\dot{p}_i = -\frac{\partial H}{\partial q_i} \quad , \quad \dot{q}_i = \frac{\partial H}{\partial p_i} \qquad (i = 1, \ldots, n). \tag{4.1}$$

- Einer Anfangskonfiguration $x_0 \in P$.

  Allgemeiner kann man bei unvollständiger Kenntnis der Anfangswerte (die wegen endlicher Messgenauigkeit die Regel ist), den *Anfangszustand* als Wahrscheinlichkeitsmaß auf P modellieren. Der Spezialfall genauer Kenntnis des Anfangswerts $x_0 \in P$ entspricht dem Diracmaß δ_{x_0} bei x_0.

- Dem Raum $C(P, \mathbb{R})$ der stetigen reellen Funktionen auf dem Phasenraum, genannt *Observablen*.

Unter günstigen Umständen existiert dann eine Lösung des *dynamischen Systems* im Sinne eines Flusses

$$\Psi^t : P \to P \qquad (t \in \mathbb{R}),$$

wobei $\Psi^t(x_0)$ die Lösung des Anfangswertproblems $\dot{x} = v(x)$, $x(0) = x_0$ zur Zeit t ist.

In diesem Fall wird auf dem Raum $C(P, \mathbb{R})$ der Observablen O durch die Familie

$$\widehat{\Psi}^t : C(P, \mathbb{R}) \to C(P, \mathbb{R}) \quad , \quad O \mapsto O \circ \Psi^t \qquad (t \in \mathbb{R})$$

linearer Abbildungen eine Zeitevolution definiert.

Beispiel 4.1 (Himmelsmechanisches 1-Zentren-Problem)
Es wird die Bewegung der Erde im Kraftfeld der Sonne beschrieben. Beide Himmelskörper werden als punktförmig angesehen, wobei sich die (schwerere) Sonne im Koordinatenursprung des $\mathbb{R}^3$ befindet.

Der sog. *Konfigurationsraum* der Erde ist damit $\mathbb{R}^3 \setminus \{0\}$, der Phasenraum der Erde durch $P := \mathbb{R}^3 \times (\mathbb{R}^3 \setminus \{0\})$ gegeben, und ein Punkt $(\vec{p}, \vec{q}) \in P$ fixiert den Ort $\vec{q}$ und die Geschwindigkeit[6] $\vec{p}$ der Erde.

Da die Geschwindigkeit $\vec{p}$ die zeitliche Änderung $\dot{\vec{q}}$ des Ortes, und die Geschwindigkeitsänderung $\dot{\vec{p}}$ proportional der auf die Erde ausgeübten Schwerkraft $- - \vec{q}/|\vec{q}|^3$ ist, gilt (in geeigneten Einheiten, in denen die Masse der Erde 1 ist)

$$\dot{x} = (\dot{\vec{p}}, \dot{\vec{q}}) = v(x) := \left(-\vec{q}/|\vec{q}|^3, \vec{p}\right).$$

Dies sind die Hamiltonschen Gleichungen für die Hamiltonfunktion

$$H(\vec{p}, \vec{q}) := \tfrac{1}{2}\vec{p}^2 - \frac{1}{|\vec{q}|}.$$

Die Komponenten des *Drehimpulses* $\vec{L} : P \to \mathbb{R}^3$, $\vec{L}(\vec{p}, \vec{q}) := \vec{q} \times \vec{p}$ sind neben den Komponenten p_i, q_i von Impuls und Ort weitere Beispiele für physikalisch relevante Observablen.

$\vec{L}$ ist zeitlich konstant, da $dL_i(v) = 0$, $i = 1, 2, 3$. Die durch Anfangsort und Anfangsimpuls aufgespannte Ebene senkrecht zu $\vec{L}(x_0)$ ist damit Bahnebene. Ist $\vec{L}(x_0) = \vec{0}$, dann findet in endlicher Zeit eine Kollision statt, sonst nicht.

Die Bahn $t \mapsto \vec{q}(t)$ ist ein Kegelschnitt mit der Sonne im einen Brennpunkt.

[6] Genauer: ihren Impuls

Wir werden keine tieferen Aussagen und Konzepte der klassischen Mechanik benötigen. Eine gute Einführung findet man in Arnol'd [Ar], siehe auch[7].

4.2 Quantenmechanik

Ein System der Quantenmechanik wird durch folgende Daten gegeben:

- dem $\mathbb{C}$–*Hilbertraum*[8] $(\mathcal{H}, \langle\cdot,\cdot\rangle)$. Die Elemente $\psi \in \mathcal{H}$, $\|\psi\| = 1$ der Einheitssphäre heißen (reine) *Zustände*. Diesen ist die Orthogonalprojektion $\hat{P}_\psi : \mathcal{H} \to \mathcal{H}$ auf den Unterraum $\mathrm{span}(\psi)$ zugeordnet.

  Verallgemeinert betrachtet man bei unvollständiger Kenntnis des Zustandes *Dichtematrizen* $\hat{\rho} \in \mathcal{B}(\mathcal{H})$, d.h. positive Spurklasseoperatoren [9] mit Spur 1.

- einem linearen Operator $\hat{H}$ auf $\mathcal{H}$, dem sog. 'Hamiltonoperator';

- einem *Anfangszustand* $\psi_0 \in \mathcal{H}$.

- dem Banachraum $\mathcal{B}(\mathcal{H})$ der beschränkten Operatoren. $\hat{O} \in \mathcal{B}(\mathcal{H})$ mit Operatornorm. $\hat{O}$ heißt *Observable*, wenn der adjungierte Operator $\hat{O}^* = \hat{O}$. Bezüglich eines Zustandes ψ hat $\hat{O}$ den *Erwartungswert* $\left\langle \psi, \hat{O}\,\psi \right\rangle = \mathrm{tr}(P_\psi \hat{O})$.

---

[7] Vorlesungsskript *Mathematische Physik 1 (Klassische Mechanik)*, erhältlich unter `http://www.mi.uni-erlangen.de/~knauf/Skripte/skripte.html`

[8] **Def.:** • Ein $\mathbb{C}$–*unitärer Vektorraum* $\mathcal{H}$ ist ein Vektorraum über den komplexen Zahlen mit einem sogenannten *Skalarprodukt*, d.h. einer Abbildung $\langle\cdot,\cdot\rangle : \mathcal{H} \times \mathcal{H} \to \mathbb{C}$, für die gilt:

- $\langle \varphi, \psi_1 + c\psi_2 \rangle = \langle \varphi, \psi_1 \rangle + c\,\langle \varphi, \psi_2 \rangle \qquad (c \in \mathbb{C},\ \varphi, \psi_i \in \mathcal{H})$
- $\langle \varphi, \psi \rangle = \overline{\langle \psi, \varphi \rangle} \qquad (\varphi, \psi \in \mathcal{H})$
- $\langle \psi, \psi \rangle \geq 0$, und $\langle \psi, \psi \rangle \geq 0 \iff \psi = 0$.

• Ein $\mathbb{C}$–*unitärer Vektorraum* $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ heißt *Hilbertraum*, wenn er bezüglich der Norm $\|\psi\| := \sqrt{\langle \psi, \psi \rangle}$ vollständig ist, d.h. alle Cauchyfolgen konvergieren.

[9] **Def:** • Ein Operator $\hat{O} \in \mathcal{B}(\mathcal{H})$ heißt *positiv*, wenn $\hat{O}^* = \hat{O}$ und sein Spektrum $\sigma(\hat{O}) \subset [0, \infty)$. • Ein positiver Operator $\hat{O} \in \mathcal{B}(\mathcal{H})$ in einem Hilbertraum mit Orthonormalbasis $(\varphi_i)_{i\in\mathbb{N}}$ ist ein *Spurklasseoperator*, wenn seine *Spur* $\mathrm{tr}(\hat{O}) := \sum_{i=1}^{\infty} \left\langle \varphi_i, \hat{O}\,\varphi_i \right\rangle$ endlich ist; für nichtpositive Operatoren betrachtet man die Spur von $|\hat{O}|$.

Nach dem Satz von Stone ist $\hat{U}(t) := \exp(-i\hat{H}\,t)$, $t \in \mathbb{R}$, genau dann unitär (d.h. $\hat{U}^{-1} = \hat{U}^{*}$), wenn $\hat{H} : \mathcal{H} \to \mathcal{H}$ selbstadjungiert ist (d.h. $\hat{H}^{*} = \hat{H}$), siehe z.B. Thirring [Th] Bd. 3.

Ist der Hamiltonoperator $\hat{H}$ also selbstadjungiert, dann erzeugt er auf $\mathcal{H}$ die Zeitevolution $\psi_0 \mapsto \psi_t := \hat{U}(t)\psi_0$, die Zustände in Zustände überführt, und die *Schrödingergleichung*

$$i\frac{d}{dt}\psi = \hat{H}\,\psi \tag{4.2}$$

löst. Auf $\mathcal{B}(\mathcal{H})$ erzeugt $\hat{H}$ die Zeitevolution

$$\hat{O} \mapsto \hat{O}_t := \hat{U}(-t)\hat{O}\,\hat{U}(t) \qquad (t \in \mathbb{R}).$$

Beispiel 4.2 Wasserstoffatom Für das Elektrons im Kraftfeld eines im Ursprung des $\mathbb{R}^3$ befindlichen Atomkerns (also dem quantenmechanischen Analogon des Keplerproblems) ist der Hilbertraum $\mathcal{H} := L^2(\mathbb{R}^3, \mathbb{C})$ der Raum der quadratintegrablen komplexen Funktionen (mit innerem Produkt $\langle f, g\rangle = \int_{\mathbb{R}^3} \bar{f}(x)g(x)dx$).

Bezeichnen wir mit $\hat{p} := (\hat{p}_1, \hat{p}_2, \hat{p}_3)$ den Impulsvektor, dessen Komponenten die Differenzialoperatoren $\hat{p}_i := -i\partial/\partial q_i$ sind, dann ist dieser nicht beschränkt, und gleiches gilt für den Ortsvektor $\hat{q} := (\hat{q}_1, \hat{q}_2, \hat{q}_3)$, dessen Komponenten die Multiplikationsoperatoren $\hat{q}_i$ mit der Funktion q_i sind.

Entsprechend ist auch der Hamiltonoperator $\hat{H} := \frac{1}{2}\hat{p}^2 - 1/|\hat{q}|$ unbeschränkt, aber doch (auf einem dichten Unterraum von $\mathcal{H}$) selbstadjungiert. Damit existiert die quantenmechanische Zeitevolution für alle Anfangszustände (im Gegensatz zur klassischen Mechanik!).

Neben den $\hat{p}_i$ und $\hat{q}_i$ sind die Komponenten des Drehimpulsvektors $\hat{L} := \hat{q} \times \hat{p}$ physikalisch wichtige Observablen. Wegen des verschwindenden *Kommutators* $[\hat{H}, \hat{L}_i] = \hat{H}\,\hat{L}_i - \hat{L}_i\hat{H} = 0$ von Hamiltonoperator und Drehimpulskomponenten ist

$$\frac{d}{dt}\hat{L}_t = U(-t)i[\hat{H}, \hat{L}]U(t) = 0,$$

der Drehimpulsoperator also zeitinvariant.

4.3 Vergleich zwischen klassischer und Quantenmechanik

Ein Hauptproblem der Mathematischen Physik ist die Berechnung der *Dynamik*, d.h. des Flusses Ψ^t bzw. der unitären Zeitevolution $U(t)$ zur Zeit $t \in \mathbb{R}$. Mathematisch gesehen entspricht dies normalerweise der Lösung der Hamiltonschen

Gleichungen (4.1), einer gewöhnlichen Differenzialgleichung bzw. der Schrödingergleichung (4.2), einer partiellen Differenzialgleichung.

Auch wenn unser Fernziel darin besteht, einen Quantencomputer mit kontrollierter Funktionsweise, d.h. Dynamik zu modellieren, müssen wir zunächst die konzeptionelle Frage beantworten, was eine solche Maschine denn von einem klassischen Computer unterscheidet.

Insbesondere werden sich die Frage nach der Bedeutung von Korrelationen zwischen Teilsystemen und die nach der Rolle von Messungen als entscheidend herausstellen.

Die bisherigen Darstellungen der beiden physikalischen Theorien sind zu unterschiedlich, um sie direkt zu vergleichen. Eine gemeinsame Formulierung ist aber auf Basis der sog. C^*–Algebren möglich. Kapitel 2.1 und 2.2 des Buches [BR] von Bratteli und Robinson geben eine gute Einführung in die Theorie der C^*–Algebren.

Wir wollen Observablen addieren und multiplizieren können. Da wir an der Quantenmechanik interessiert sind, wollen wir Algebren über den komplexen Zahlen $\mathbb{C}$ betrachten, aber auch über die Komplexkonjugation verfügen, die uns zum reellen (und damit zu den Messwerten) zurückkommen lässt. Wir wollen auch Abstände zwischen Observablen messen:

Definition 4.3 • Eine *C^*–Algebra* $\mathcal{A} = (\mathcal{A}, +, \cdot, *, \|\cdot\|)$ ist ein Banachraum $(\mathcal{A}, +, \|\cdot\|)$ über $\mathbb{C}$ mit Multiplikation $\cdot : \mathcal{A}\times\mathcal{A} \to \mathcal{A}$ und Involution $* : \mathcal{A} \to \mathcal{A}$, sodass (mit $a, b, b_i \in \mathcal{A}$, $\alpha, \beta \in \mathbb{C}$):

- $a \cdot (b_1 + b_2) = a \cdot b_1 + a \cdot b_2 \quad , \quad a \cdot (b \cdot c) = (a \cdot b) \cdot c$ und $(\alpha a) \cdot (\beta b) = \alpha\beta(a \cdot b)$.
- $a^{**} = a \quad , \quad (\alpha a)^* = \bar{\alpha} a^* \quad , \quad (a + b)^* = a^* + b^* \quad , \quad (a \cdot b)^* = b^* \cdot a^*$
- $\|a \cdot b\| \leq \|a\|\ \|b\| \quad , \quad \|a^* \cdot a\| = \|a\|^2$.

• $\mathcal{A}$ heißt *abelsch*, wenn zusätzlich $a \cdot b = b \cdot a$ gilt.
• $\mathcal{A}$ heißt *unital*, wenn ein Element $\mathbb{1} \in \mathcal{A}$ mit $\mathbb{1} \cdot a = a \cdot \mathbb{1} = a$ existiert.

Wegen $\|a\|^2 = \|a^* \cdot a\| \leq \|a^*\| \cdot \|a\|$ gilt $\|a\| \leq \|a^*\|$, also besitzt die *Adjungierte* a^* die gleiche Norm wie a.

Beispiel 4.4 1. Es sei P ein kompakter topologischer Raum (z.B. eine kompakte Teilmenge $P \subset \mathbb{R}^n$). Dann ist der Banachraum

$$\mathcal{A}_{\mathrm{cl}} \equiv \mathcal{A}_{\mathrm{cl}}(P) := (C(P, \mathbb{C}), +, \cdot, *, \|\cdot\|)$$

der stetigen komplexwertigen Funktionen mit Supremumsnorm $\|a\| := \sup_{x \in P} |a(x)|$, punktweiser Addition $(a+b)(x) := a(x)+b(x)$, punktweiser Multiplikation $(a \cdot b)(x) := a(x)b(x)$ und Konjugation $a^*(x) := \overline{a(x)}$ ist eine abelsche unitale C^*–Algebra.

Besonders wird uns im folgenden die Observablenalgebra über dem 2^n–elementige Phasenraum $P = B^n$ von n Bits $B = \{0, 1\}$ interessieren. Wichtige Observablen sind dann die Funktionen

$$a_i : P \to \{0, 1\},\ a_i(b_1, \ldots, b_n) := b_i,$$

die den Wert des i-ten Bits auslesen.

2. Es sei $(\mathcal{H}, \langle \cdot, \cdot \rangle)$ ein ein $\mathbb{C}$–Hilbertraum. Der Banachraum

$$\mathcal{A}_{\mathrm{qm}} \equiv \mathcal{A}_{\mathrm{qm}}(\mathcal{H}) := (\mathcal{B}(\mathcal{H}), +, \cdot, *, \|\cdot\|)$$

der beschränkten Operatoren $a : \mathcal{H} \to \mathcal{H}$ mit *Operatornorm*

$$\|a\| := \sup_{v \in \mathcal{H} \setminus \{0\}} \frac{\|av\|}{\|v\|} = \sup_{v \in \mathcal{H},\, \|v\|=1} \|av\|, \tag{4.3}$$

Komposition $\cdot$ von Operatoren und durch

$$\langle a^* v, w \rangle = \langle v, aw \rangle \quad , \quad (v, w \in \mathcal{H})$$

fixierter Konjugation ist eine unitale (für $\dim \mathcal{H} > 1$ nicht abelsche) C^*–Algebra.

Baustein der quantenmechanischen Informationstheorie ist die Observablenalgebra $\mathcal{B}(\mathbb{C}^2)$ eines Qubits. Wichtige Observablen sind die sog. *Pauli-Matrizen*

$$\hat{\sigma}_1 := \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right) \quad , \quad \hat{\sigma}_2 := \left(\begin{smallmatrix} 0 & i \\ -i & 0 \end{smallmatrix}\right) \quad , \quad \hat{\sigma}_3 := \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right).$$

Diese sind selbstadjungiert und haben die Eigenwerte ± 1.

Im ersten Fall wird man ein Modell der klassischen Observablen über einem Phasenraum P erkennen, im zweiten ein Modell der quantenmechanischen Observablen.

Bemerkungen 4.5 1. In beiden Fällen erscheint die *Beschränktheit der Observablen* (die in Bsp. 4.4.1. durch die Kompaktheit von P und die Stetigkeit der Funktionen erzwungen wird) unphysikalisch. Diese lässt sich aber umgehen (um z.B. die schon in den Beispielen 4.1 und 4.2 auftauchenden unbeschränkten Observablen zu behandeln). In unserem informationstheoretischen Kontext ist sie unproblematisch.

2. Besteht ein Phasenraum P aus mehreren Zusammenhangskomponenten (wie dies im Fall $P = B^n$ der Fall ist), dann bleibt ein Anfangspunkt $x \in P$ unter der Wirkung eines stetigen Flusses in der gleichen Zusammenhangskomponente. Bezogen auf das Beispiel B^n würde dies bedeuten, dass die Dynamik trivial wäre.

   Wir stellen uns hier aber vor, dass der Phasenraum P in einen zusammenhängenden Phasenraum $\tilde{P}$ eingebettet ist, und auf $\tilde{P}$ ein Fluss $\tilde{\Psi}^t$ ($t \in \mathbb{R}$) mit der Eigenschaft existiert, dass seine Einschränkung $\Psi^t := \tilde{\Psi}^t{\restriction}_P$ ($t \in \mathbb{Z}$) auf P und ganzzahlige Zeiten eine Abbildung $\Psi^t : P \to P$ ist. Dies ist z.B. beim einleitend erwähnten reversiblen Billiardball–Computer der Fall (allerdings ist in diesem Fall $\tilde{P}$ nicht mehr kompakt).

Das *Spektrum* eines Elements $a \in \mathcal{A}$ einer C^*–Algebra besteht im Fall $\mathcal{A} = \mathcal{B}(\mathcal{H})$ der linearen Endomorphismen eines endlich-dimensionalen $\mathbb{C}$–Hilbertraumes $\mathcal{H}$ aus den Eigenwerten $\lambda \in \mathbb{C}$ von a. Genau für diese Werte ist $a - \lambda \mathbb{1}$ nicht bijektiv, also nicht invertierbar. Dies verallgemeinert sich folgendermaßen:

Definition 4.6 Ist die C^*–Algebra $\mathcal{A}$ unital, dann

- heißt $b \in \mathcal{A}$ *zu* $a \in \mathcal{A}$ *invers*, wenn $ba = ab = \mathbb{1}$. Existiert b, dann ist es eindeutig, und wird a^{-1} geschrieben.
- Das *Spektrum von* a ist die Menge $\sigma(a) := \{\lambda \in \mathbb{C} \mid \nexists (a - \lambda \mathbb{1})^{-1}\}$.
- Der *Spektralradius* von a ist $\mathrm{spr}(a) := \sup\{|\lambda| \mid \lambda \in \sigma(a)\}$.

Das Komplement $\mathbb{C} - \sigma(a)$ des Spektrums heißt *Resolventenmenge* von a, und für λ aus der Resolventenmenge wird $(a - \lambda \mathbb{1})^{-1}$ die *Resolvente von* a *bei* λ genannt.

Satz 4.7 *In einer unitalen C^*–Algebra $\mathcal{A}$ ist für alle $a \in \mathcal{A}$ das Spektrum $\sigma(a)$ eine nicht leere abgeschlossene Menge, und es gilt*

$$\mathrm{spr}(a) = \lim_{n\to\infty} \|a^n\|^{1/n} \leq \|a\|.$$

Bew.: • Es sei $R := \liminf_{n\to\infty} \|a^n\|^{1/n}$ und $N \in \mathbb{N}$ mit $\|a^N\|^{1/N} \leq R + \varepsilon/2$. Weiter setzen wir $C := \max(1, \|a\|)$. Dann gilt es für $n \in \mathbb{N}$ eine eindeutige Zerlegung $n = kN + l$ mit $0 \leq l < N$, und

$$\begin{aligned} \|a^n\|^{1/n} &\leq \|a^{kN}\|^{1/n} \cdot \|a^l\|^{1/n} \leq \|a^N\|^{k/n} \cdot C^{l/n} \\ &\leq (R + \varepsilon/2)^{\frac{Nk}{n}} C^{N/n} < R + \varepsilon \end{aligned}$$

für n groß. Also existiert $\lim_{n\to\infty} \|a^n\|^{1/n}$ und ist gleich R.
• Es sei $|\lambda| > \|a^n\|^{1/n}$ für ein $n \in \mathbb{N}$. Dann konvergiert die Cauchyfolge

$$\lambda^{-1} \sum_{k=0}^{\infty} \left(\frac{a}{\lambda}\right)^k$$

wegen der Vollständigkeit des Banachraumes $\mathcal{A}$, und zwar gegen das Element $(\lambda\,\mathbb{1} - a)^{-1}$. Damit ist $\lambda \notin \sigma(a)$, also $\mathrm{spr}(a) \leq R$.
• Die umgekehrte Ungleichung $\mathrm{spr(a)} \geq R$ wird z.B. in [BR], Seite 26 bewiesen.
• Ist $\lambda_0 \notin \sigma(a)$ und $\lambda \in \mathbb{C}$ ein Punkt mit $|\lambda - \lambda_0| < \|(a - \lambda_0\,\mathbb{1})^{-1}\|$, dann konvergiert die sogenannte *Neumannreihe*

$$\sum_{k=0}^{\infty} (\lambda_0 - \lambda)^k (a - \lambda_0 \mathbb{1})^{-k-1},$$

und zwar gegen $(a - \lambda_0 \mathbb{1})^{-1}$. Also ist $\mathbb{C} - \sigma(a)$ offen bzw. $\sigma(a) \subset \mathbb{C}$ abgeschlossen.
• Wäre $\sigma(a) = \emptyset$, dann wäre die analytische Funktion $\lambda \mapsto (a - \lambda\,\mathbb{1})^{-1}$ auf ganz $\mathbb{C}$ definiert und nach dem Satz von Liouville[10] gleich 0 (da $\lim_{\lambda\to\infty} \|(a - \lambda\,\mathbb{1})^{-1}\| = 0$). Widerspruch! □

Es gilt wegen $\mathbb{1} = \big((a - \lambda\,\mathbb{1})(a - \lambda\,\mathbb{1})^{-1}\big)^* = ((a - \lambda\,\mathbb{1})^{-1})^*(a^* - \lambda\,\mathbb{1})$

$$\sigma(a^*) = \overline{\sigma(a)} \qquad (a \in \mathcal{A}).$$

Wir sind aus der linearen Algebra gewohnt, Operatoren mit den folgenden Eigenschaften auszusondern:

[10] **Satz (Liouville)**: Eine ganze beschränkte Funktion ist konstant.

Definition 4.8 Ein Element $a \in \mathcal{A}$ der C^*–Algebra $\mathcal{A}$ heißt

- *normal*, wenn $aa^* = a^*a$,
- *selbstadjungiert* oder *hermitesch*, wenn $a^* = a$,
- *(orthogonale) Projektion*, wenn $a^2 = a = a^*$,
- *positiv* (Notation: $a \geq 0$), wenn $b \in \mathcal{A}$ mit $a = bb^*$ existiert.
- Ist $\mathcal{A}$ unital, dann heißt a *unitär*, wenn $a^*a = aa^* = \mathbb{1}$,

Bemerkungen 4.9

1. Die hermiteschen und unitären Operatoren a sind offensichtlich normal; ist $a \geq 0$, dann ist a hermitesch, und die Projektionen sind positiv.
2. In einer abelschen C^*–Algebra sind offensichtlich alle Elemente normal, während z.B. die nilpotente Matrix $a := \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right) \in \mathcal{B}(\mathbb{C}^2)$ diese Eigenschaft nicht besitzt. Hier gilt auch $0 = \|a^2\| < \|a\|^2 = 1$.
3. Für die abelsche C^*–Algebra $\mathcal{A}_{\mathrm{cl}} = C(P, \mathbb{C})$ ist $\sigma(a) = a(P) \subset \mathbb{C}$.
4. Für $\mathcal{A}_{\mathrm{qm}} = \mathcal{B}(\mathcal{H})$ sind die Eigenwerte von a Bestandteil des Spektrums $\sigma(a)$. Ist aber $\dim(\mathcal{H}) = \infty$, so kann es noch weitere Bestandteile geben.

Satz 4.10 *Für ein Element a aus einer unitalen C^*–Algebra $\mathcal{A}$ gilt:*

1. a *normal* $\implies \mathrm{spr}(a) = \|a\|$
2. a *hermitesch* $\implies \sigma(a) \subseteq [\,-\|a\|, \|a\|\,]$
3. a *invertierbar* $\implies \sigma(a^{-1}) = (\sigma(a))^{-1}$
4. a *unitär* $\implies \sigma(a) \subseteq S^1 := \{\lambda \in \mathbb{C} \mid |\lambda| = 1\}$
5. $a \in \mathcal{A}$ *und* $p \in \mathbb{C}[x]$ *Polynom* $\implies \sigma(p(a)) = p(\sigma(a))$
6. a *Projektion* $\implies \sigma(a) \subseteq \{0, 1\}$
7. a *positiv* $\implies \sigma(a) \subseteq [\,0, \|a\|\,]$.

Ist $\dim(\mathcal{A}) < \infty$*, dann ist* $\mathcal{A}$ *isomorph zu einer direkten Summe*

$$\mathcal{A} \cong \bigoplus_{k=1}^{k_{\max}} \mathcal{B}(\mathbb{C}^{m_k}), \tag{4.4}$$

und normale Operatoren $a \in \mathcal{A}$ *besitzen die Spektraldarstellung*

$$a = \sum_{\lambda \in \sigma(a)} \lambda P_\lambda \quad \textit{mit Projektionen} \quad P_\lambda \in \mathcal{A}, \tag{4.5}$$

für die $\sum_{\lambda\in\sigma(a)} P_\lambda = \mathbb{1}$*, und* $P_\lambda P_\mu = \delta_{\lambda,\mu} P_\lambda$ *gilt.*

Bemerkungen 4.11 1. Diese Aussagen verallgemeinern bekannte Aussagen über die Eigenwerte komplexer Matrizen.

2. Wegen 1. ist die Norm $\|\cdot\| : \mathcal{A} \to \mathbb{R}$ schon durch die algebraische Stuktur von $\mathcal{A}$ festgelegt:

$$\|a\| = \|a^*a\|^{1/2} = \mathrm{spr}(a^*a)^{1/2} \qquad (a \in \mathcal{A}).$$

Bew.:

1. Nach Satz 4.7 ist $\mathrm{spr}(a) = \lim_{n\to\infty} \|a^n\|^{1/n} = \lim_{m\to\infty} \|a^{2^m}\|^{2^{-m}}$. Nun ist

$$\begin{aligned} \|a^{2^m}\|^2 &= \|(a^{2^m})^* a^{2^m}\| = \|(a^*)^{2^m} a^{2^m}\| \\ &= \|(a^*a)^{2^m}\| = \|(a^*a)^{2^{m-1}}\|^2 = \ldots = \|a\|^{2^{m+1}}, \end{aligned}$$

   wobei wir in der dritten Gleichung die Normalität von a benutzt haben. Also ist $\|a^{2^m}\|^{2^{-m}} = \|a\|$.

2. Da a normal ist, reicht es wegen 1. aus zu zeigen, dass $\sigma(a) \subseteq \mathbb{R}$. Wir nehmen nun an, dass $\lambda \in \mathbb{C}\backslash\mathbb{R}$ zum Spektrum von a gehört.

   Setzen wir nun $\alpha := \mathrm{Re}(\lambda)$ und $\beta := \mathrm{Im}(\lambda)$, dann gilt $i\beta \in \sigma(a - \alpha\mathbb{1})$ und auch für alle $n \in \mathbb{N}$

$$i(n+1)\beta \in \sigma(a - (\alpha - in\beta)\mathbb{1})$$

oder

$$\begin{aligned}(n+1)^2\beta^2 &\leq (\mathrm{spr}(a-(\alpha-in\beta)\mathbb{1}))^2\\ &\leq \|(a-(\alpha-in\beta)\mathbb{1})(a-(\alpha+in\beta)\mathbb{1})\|\\ &= \|(a-\alpha\mathbb{1})^2+n^2\beta^2\mathbb{1}\| \leq n^2\beta^2+\|a-\alpha\mathbb{1}\|^2,\end{aligned}$$

was für große n unmöglich ist, falls $\beta \neq 0$.

3. Für alle $\lambda \in \mathbb{C}\backslash\{0\}$ gilt

$$\lambda\,\mathbb{1}-a = \lambda a(a^{-1}-\lambda^{-1}\mathbb{1}) \quad \text{und} \quad \lambda^{-1}\mathbb{1}-a^{-1} = \lambda^{-1}a^{-1}(a-\lambda\,\mathbb{1}).$$

Ist nun $\lambda\,\mathbb{1}-a$ nicht invertierbar, dann ist wegen der ersten Gleichung auch $\lambda^{-1}\mathbb{1}-a^{-1}$ nicht invertierbar. Analog folgt wegen der zweiten Gleichung aus der Invertierbarkeit von $\lambda\,\mathbb{1}-a$ die Invertierbarkeit von $\lambda^{-1}\mathbb{1}-a^{-1}$.

Wegen der Invertierbarkeit von a und a^{-1} ist 0 weder in $\sigma(a)$ noch in $\sigma(a^{-1})$.

4. Da $a \in \mathcal{A}$ unitär ist, gilt $\|a\|^2 = \|a^*a\| = \|\mathbb{1}\| = 1$, also wegen Satz 4.7 $\sigma(a) \subseteq \{\lambda \in \mathbb{C} \mid |\lambda| \leq 1\}$.

Andererseits ist a invertierbar mit inverser a^*, und es gilt auch $\|a^*\| = 1$, also $\sigma(a^{-1}) \subseteq \{\lambda \in \mathbb{C} \mid |\lambda| \leq 1\}$. Da nach 3. $\sigma(a^{-1}) = (\sigma(a))^{-1}$ ist, folgt $\sigma(a) \subseteq S^1$.

5. Allgemein gilt für das Produkt $b := a_1 \cdot \ldots \cdot a_n$ kommutierender Elemente $a_i \in \mathcal{A}$: b ist genau dann invertierbar, wenn $a_1, \ldots, a_n$ invertierbar.

Denn im Fall der Invertierbarkeit von b gilt $a_i^{-1} = a_1 \cdot \ldots \cdot a_{i-1} b^{-1} a_{i+1} \cdot \ldots \cdot a_n$, während umgekehrt $b^{-1} = a_1^{-1} \cdot \ldots \cdot a_n^{-1}$ ist.

Ist nun für das Polynom p und $\lambda \in \mathbb{C}$

$$p(x) - \lambda = \alpha_0 \prod_{i=1}^{n} (x - \alpha_i),$$

dann setzen wir $b := p(a) - \lambda\,\mathbb{1} = \alpha_0 \prod_{i=1}^{n}(a-\alpha_i\mathbb{1})$ und $a_i := a - \alpha_i\mathbb{1}$ (für konstante Polynome p ist der Satz offensichtlich erfüllt, sodass wir $\alpha_0 \neq 0$ und $n \geq 1$ voraussetzen).

- Ist $\lambda \in \sigma(p(a))$, also b nicht invertierbar, dann ist auch für ein i der Faktor a_i nicht invertierbar, also $\alpha_i \in \sigma(a)$ und $p(\alpha_i) = \lambda \in p(\sigma(a))$.
- Ist umgekehrt $\lambda \in p(\sigma(a))$, also ex. ein $\gamma \in \sigma(a)$ mit $\lambda = p(\gamma)$, dann muss $\gamma \in \{\alpha_1, \ldots, \alpha_n\}$ gelten. Dann ist ein a_i nicht invertierbar und damit $\lambda \in \sigma(p(a))$.

6. Ist a eine Projektion, dann gilt $a^2 = a$, also $p(a) = 0$ für $p(x) := x^2 - x$. Damit ist nach 5. $p(\sigma(a)) = \sigma(p(a)) = \sigma(0) = \{0\}$, also $\sigma(a) \subseteq \{0, 1\}$.

7. Ich beweise den Satz nur im Fall $\dim(\mathcal{A}) < \infty$. Der allgemeine Fall wird z.B. in [BR] behandelt.

   Da $a = b^*b$ selbstadjungiert, also insbesondere normal ist, ist nach (4.4) $a = \sum_{\lambda \in \sigma(a)} \lambda P_\lambda$, und $\sigma(a) \subseteq [\, -\|a\|, \|a\| \,]$.

   Gäbe es ein $\lambda < 0$ aus dem Spektrum von a, dann auch einen Vektor $\psi \in \mathrm{Im}(P_\lambda) \backslash \{0\}$ mit der Eigenschaft $\langle \psi, a\psi \rangle = \langle \psi, \lambda P_\lambda \psi \rangle = \lambda \langle \psi, \psi \rangle < 0$. Andererseits würde gelten $\langle \psi, a\psi \rangle = \langle b\psi, b\psi \rangle \geq 0$. Widerspruch!

Einen Beweis der Isomorphie (4.4) findet man z.B. in Davidson [Da], Thm. III.1.1.

Im Fall endlich-dimensionaler Algebren folgt wegen der Isomorphie (4.4) die Spektraldarstellung (4.5) aus der analogen Aussage[11] in $\mathcal{B}(\mathbb{C}^n)$ mit $n := \sum_{k=1}^{k_{\max}} n_k$, denn dieser enthält $\oplus_{k=1}^{k_{\max}} \mathcal{B}(\mathbb{C}^{n_k})$ als Unteralgebra, und die Spektralprojektoren P_λ respektieren die Summenaufspaltung. □

5 Isomorphismen von C^*-Algebren

Wir schauen uns als Nächstes die strukturerhaltenden Abbildungen von C^*–Algebren an:

Definition 5.1 • Eine lineare Abbildung $\alpha : \mathcal{A} \to \mathcal{B}$ zwischen C^*–Algebren heißt $*$–*Morphismus*, wenn für alle $a, b \in \mathcal{A}$ gilt:

$$\alpha(ab) = \alpha(a)\,\alpha(b) \quad , \quad \alpha(a^*) = \alpha(a)^*.$$

• Ein $*$–Morphismus $\alpha : \mathcal{A} \to \mathcal{B}$ heißt $*$–*Isomorphismus*, wenn α bijektiv ist,
• $*$–*Automorphismus*, wenn zusätzlich $\mathcal{B} = \mathcal{A}$.

[11] Jede Einführung in die Lineare Algebra, z.B. mein Skript *Lineare Algebra 2*, Satz 4.6 und 8.14, erhältlich unter `http://www.mi.uni-erlangen.de/~knauf/Skripte/skripte.html`

Im informationstheoretischen Kontext kann $\mathcal{A}$ die Observablenalgebra des Senders, $\mathcal{B}$ die Observablenalgebra des Empfängers einer Nachricht sein, und α modelliert den Kanal, über den die Nachricht geschickt wird. Ebenso kann α eine Berechnung modellieren.

Wir werden später die entsprechenden Unterschiede zwischen dem klassischen (d.h. $\mathcal{A}$ und $\mathcal{B}$ abelsch) und dem quantenmechanischen Fall untersuchen.

Satz 5.2 • *Ein $*$–Morphismus $\alpha : \mathcal{A} \to \mathcal{B}$ ist* **positiv**, *d.h. aus $a \geq 0$ folgt $\alpha(a) \geq 0$.*
• *$\alpha(\mathcal{A}) \subseteq \mathcal{B}$ ist eine C^*–Unteralgebra.*
• *Sind $\mathcal{A}$ und $\mathcal{B}$ unital und ist $\alpha(\mathbb{1}_{\mathcal{A}}) = \mathbb{1}_{\mathcal{B}}$, dann ist $\|\alpha(a)\| \leq \|a\|$, α also insbesondere stetig.*
• *Für $*$–Isomorphismen α gilt $\|\alpha(a)\| = \|a\| \qquad (a \in \mathcal{A})$.*

Bew.:

- $a \geq 0$ bedeutet ja, dass ein $c \in \mathcal{A}$ mit $a = c^*c$ existiert. Dann ist

$$\alpha(a) = \alpha(c^*c) = \alpha(c^*)\alpha(c) = \alpha(c)^*\alpha(c) \geq 0.$$

- Um zu zeigen, dass der Untervektorraum $\alpha(\mathcal{A})$ eine C^*–Unteralgebra ist, müssen wir zeigen, dass für $b, b' \in \alpha(\mathcal{A})$ auch $bb' \in \alpha(\mathcal{A})$ und $b^* \in \alpha(\mathcal{A})$ gilt. Das folgt aber mit $b = \alpha(a)$, $b' = \alpha(a')$ aus $bb' = \alpha(aa')$ und $b^* = \alpha(a^*)$.

  Außerdem müssen wir beweisen, dass $\alpha(\mathcal{A}) \subseteq \mathcal{B}$ (mit der Norm von $\mathcal{B}$) ein Banachraum ist. Dies ist für $\dim \mathcal{A} < \infty$ trivial, denn endlich-dimensionale Unterräume sind immer abgeschlossen. Für den allgemeinen Fall siehe [BR], Prop. 2.3.1.

- Es gilt $\|a\|^2 = \|a^*a\|$ und $\|\alpha(a)\|^2 = \|\alpha(a)^*\alpha(a)\| = \|\alpha(a^*)\alpha(a)\| = \|\alpha(a^*a)\|$, wir können uns also auf den Fall selbstadjungierter $b \in \mathcal{A}$ beschränken. Für diese gilt wegen Satz 4.10.1.

$$\|b\| = \mathrm{spr}(b) \quad \text{und} \quad \|\alpha(b)\| = \mathrm{spr}(\alpha(b)),$$

  denn auch $\alpha(b)$ ist selbstadjungiert. Die Ungleichung $\|\alpha(b)\| \leq \|b\|$ folgt damit aus[12]

$$\sigma_{\mathcal{B}}(\alpha(b)) \subseteq \sigma_{\mathcal{A}}(b) \qquad (b \in \mathcal{A}). \tag{5.6}$$

[12] Es gilt sogar $\sigma_{\mathcal{B}}(\alpha(b)) = \sigma_{\mathcal{A}}(b)$ für alle $b \in \mathcal{A}$, siehe [BR], Prop. 2.2.7.

- Wir nehmen nun an, dass $\lambda \notin \sigma(b)$, also dass $(b - \lambda \mathbb{1}_{\mathcal{A}})^{-1} \in \mathcal{A}$ existiert. Dann folgt

$$\begin{aligned} \mathbb{1}_{\mathcal{B}} &= \alpha(\mathbb{1}_{\mathcal{A}}) = \alpha\left((b - \lambda \mathbb{1}_{\mathcal{A}})^{-1}(b - \lambda \mathbb{1}_{\mathcal{A}})\right) \\ &= \alpha\left((b - \lambda \mathbb{1}_{\mathcal{A}})^{-1}\right) \alpha\,(b - \lambda \mathbb{1}_{\mathcal{A}}) = \alpha\left((b - \lambda \mathbb{1}_{\mathcal{A}})^{-1}\right)(\alpha(b) - \lambda \mathbb{1}_{\mathcal{B}}), \end{aligned}$$

  also $\lambda \notin \sigma(\alpha(b))$.

- Ist $\alpha : \mathcal{A} \to \mathcal{B}$ ein $*$–Isomorphismus, dann lässt sich α auf $\alpha(\mathcal{A})$ invertieren, und $\|a\| = \|\alpha^{-1}(\alpha(a))\| \leq \|\alpha(a)\| \leq \|a\|$. □

Die Automorphismen α einer C^*–Algebra $\mathcal{A}$ bilden eine Gruppe, die mit $\mathrm{Aut}(\mathcal{A})$ bezeichnet wird.

Wir haben schon Beispiele von $*$–Automorphismen klassischer und quantenmechanischer C^*–Algebren kennengelernt:

Beispiel 5.3 1. Ist P ein kompakter Phasenraum und $\mathcal{A}_{\mathrm{cl}} := C(P, \mathbb{C})$, dann ist für einen Homöomorphismus $\Phi : P \to P$

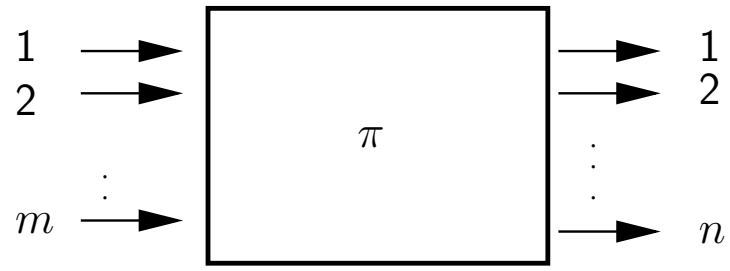
$$\alpha : \mathcal{A}_{\mathrm{cl}} \to \mathcal{A}_{\mathrm{cl}} \quad , \quad \alpha(a) := a \circ \Phi$$

ein $*$–Automorphismus.

Ist insbesondere P eine endliche Menge[13], dann ist für jede Permutation $\pi : P \to P$

$$a \mapsto a \circ \pi \qquad (a \in \mathcal{A}_{\mathrm{cl}})$$

ein $*$–Automorphismus. Insbesondere können wir reversible Berechnungen durch Permutationen von $P = B^n$ modellieren (entsprechend einem reversiblen Schaltglied mit m Eingängen und $n = m$ Ausgängen).



[13] mit der diskreten Topologie

2. Ist $\mathcal{H}$ ein Hilbertraum und $u \in \mathcal{A}_{\mathrm{qm}} := B(\mathcal{H})$ unitär, dann ist

$$a \mapsto u^* a u \qquad (a \in \mathcal{A}_{\mathrm{qm}})$$

ein $*$–Automorphismus von $\mathcal{A}_{\mathrm{qm}}$.

Automorphismen α einer C^*–Algebra $\mathcal{A}$, für die ein unitäres $u \in \mathcal{A}$ existiert, sodass $\alpha(a) = u^* a u$ gilt, heißen *innere Automorphismen*. Offensichtlich ist für abelsche $\mathcal{A}$ die Identische Abbildung der einzige innere Automorphismus.

3. Die klassische bzw. quantenmechanische Dynamik auf dem Phasenraum bzw. dem Hilbertraum induziert *einparametrige Automorphismengruppen*

$$(\alpha^t)_{t\in\mathbb{R}} \quad \text{mit} \quad \alpha^0 = \mathbb{1}_{\mathcal{A}} \quad \text{und} \quad \alpha^s \circ \alpha^t = \alpha^{s+t}$$

der entsprechenden Observablenalgebren $\mathcal{A}$:

- $\alpha^t(a) := a \circ \Psi^t \qquad (a \in \mathcal{A}_{\mathrm{cl}}, t \in \mathbb{R})$
- $\alpha^t(a) := U(-t) a U(t) \qquad (a \in \mathcal{A}_{\mathrm{qm}}, t \in \mathbb{R}).$

Die Theorie unendlich-dimensionaler C^*–Algebren ist sehr reichhaltig und entsprechend kompliziert. Wir sind jedoch an klassischen bzw. quantenmechanischen Computern mit nur endlich vielen Bauteilen interessiert, weswegen wir uns auf die Betrachtung endlich-dimensionaler C^*–Algebren beschränken können.

Für eine endlich-dimensionale C^*–Algebra $\mathcal{A}$ aber lässt sich jeder $*$–Automorphismus aus den ersten beiden Beispielen kombinieren.

Wir schreiben dazu zunächst gemäß (4.4)

$$\mathcal{A} = \bigoplus_{k=1}^{k_{\max}} \mathcal{M}_{m_k} \quad \text{mit} \quad \mathcal{M}_m := B(\mathbb{C}^m)$$

Dabei können die m_k noch untereinander gleich sein. Diese gleichen Faktoren mit Multiplizitäten s_l sammeln wir auf:

$$\mathcal{A} = \bigoplus_{l=1}^{l_{\max}} \bigoplus_{t=1}^{s_l} \mathcal{M}_{r_l} \quad \text{mit} \quad r_1 < \ldots < r_{l_{\max}}. \tag{5.7}$$

Satz 5.4 *Für jeden $*$–Automorphismus $\alpha : \mathcal{A} \to \mathcal{A}$ gibt es eine unitäre Abbildung $U \in \mathcal{A}$ und Permutationen π_l von s_l Elementen $(1 \leq l \leq l_{\max})$, sodass*

$$\alpha(a) = U^* \left(\bigoplus_{l=1}^{l_{\max}} \bigoplus_{t=1}^{s_l} a_{l,\pi_l(t)} \right) U \qquad \left(a = \bigoplus_{l=1}^{l_{\max}} \bigoplus_{t=1}^{s_l} a_{l,t} \in \mathcal{A} \right). \tag{5.8}$$

Umgekehrt ist jede Abbildung α der Form (5.8) ein $*$–Automorphismus. Dabei sind die π_l eindeutig; ist $U' \in \mathcal{A}$ eine zweite derartige unitäre Abbildung, dann ist

$$U'U^* = \bigoplus_{l=1}^{l_{\max}} \bigoplus_{t=1}^{s_l} \lambda_{t,l} \mathbb{1}_{r_l} \quad \textit{mit} \quad |\lambda_{t,l}| = 1$$

Bew.:

- Es sei $P_{l,t} \in \mathcal{A} \;\; (l = 1, \ldots, l_{\max}, t = 1, \ldots, s_l)$ die Projektion auf den t–ten Faktor der Form $\mathcal{M}_{r_l}$ in (5.7). Diese bilden eine Zerlegung der Eins:

  $$P_{l,t} P_{l',t'} = 0 \quad \text{für} \quad (l,t) \neq (l',t') \quad \text{und} \quad \sum_{l=1}^{l_{\max}} \sum_{t=1}^{s_l} P_{l,t} = \mathbb{1}_{\mathcal{A}}. \tag{5.9}$$

  Gleiches gilt auch für die Projektionen $\alpha(P_{l,t})$.

  Alle $a \in \mathcal{A}$ kommutieren mit diesen Projektionen: $P_{l,t} a = a P_{l,t}$. Damit sind auch die Elemente $\alpha(P_{l',t'}) P_{l,t} \in \mathcal{A}$ Projektionen.

- Ist nun $\alpha(P_{l_1,t_1}) P_{l,t} \neq 0$ und $\alpha(P_{l_2,t_2}) P_{l,t} \neq 0$, dann ist $(l_1, t_1) = (l_2, t_2)$. Denn für beliebige Unterräume $\mathcal{H}_1, \mathcal{H}_2 \subseteq \mathrm{Im}(P_{l,t})$ mit $\dim(\mathcal{H}_1) \leq \dim(\mathcal{H}_2)$ gibt es ein unitäres $V \in \mathcal{A}$ mit $V(\mathcal{H}_1) \subseteq \mathcal{H}_2$. Daher gilt ohne Beschränkung der Allgemeinheit

  $$\{0\} \subsetneq \mathrm{Im}(V \alpha(P_{l_1,t_1}) P_{l,t}) \subseteq \mathrm{Im}(\alpha(P_{l_2,t_2}) P_{l,t})$$

  oder

  $$\{0\} \subsetneq \mathrm{Im}(\alpha^{-1}(V P_{l,t}) P_{l_1,t_1}) \subseteq \mathrm{Im}(\alpha^{-1}(P_{l,t}) P_{l_2,t_2}),$$

  also

  $$\{0\} \subsetneq \mathrm{Im}(P_{l_1,t_1} \alpha^{-1}(V P_{l,t})) \subseteq \mathrm{Im}(P_{l_2,t_2}),$$

  was wegen (5.9) nur für $(l_1, t_1) = (l_2, t_2)$ möglich ist.

- Also gibt es zu (l', t') genau ein (l, t) mit $\alpha(P_{l',t'})P_{l,t} \neq 0$, und wir erhalten eine Permutation $(l', t') \mapsto (l, t)$.

  Damit muss wegen

  $$\mathrm{rank}(\alpha(P_{l',t'})P_{l,t}) = \mathrm{rank}(\alpha(P_{l',t'}))$$

  $r_l = \mathrm{rank}(P_{l,t}) \geq \mathrm{rank}(P_{l',t'}) = r_{l'}$ gelten, also gemäß (5.7) $l \geq l'$. Dies ist nur für $l = l'$ möglich. Damit erhalten wir für jedes l eine Permutation π_l von s_l Elementen, mit $\pi_l(t) = t'$.

- Es genügt damit nachzuweisen, dass jeder Automorphismus $\beta : \mathcal{M}_m \to \mathcal{M}_m$ einer Matrixalgebra $\mathcal{M}_m$ ein innerer Automorphismus ist, also ein unitäres $W \in \mathcal{M}_m$ mit

  $$\beta(a) = WaW^* \qquad (a \in \mathcal{M}_m)$$

  existiert. Es sei dazu $l \in \mathbb{C}^m$ ein Vektor der Länge 1, und P_l die Projektion auf $\mathrm{span}(l)$. Dann ist $\alpha(P_l) = P_f$ für einen geeigneten Vektor $f \in \mathbb{C}^m$ der Länge 1.

  Setzen wir $W \in \mathcal{M}_m$ gleich

  $$Wv := \beta(vl^*)f \qquad (v \in \mathbb{C}^m),$$

  dann ist W unitär:

  $$\begin{aligned} \|Wv\|^2 &= \langle \alpha(vl^*)f, \alpha(vl^*f)\rangle = \langle \alpha(lv^*vl^*)f, f\rangle \\ &= \|v\|^2 \langle \alpha(ll^*)f, f\rangle = \|v\|^2 \langle P_f f, f\rangle = \|v\|^2. \end{aligned}$$

  Andererseits gilt für alle $v \in \mathbb{C}^m$

  $$\begin{aligned} WaW^*v &= \beta(aW^*vl^*f) = \beta(a)\beta((W^*v)l^*f) \\ &= \beta(a)W(W^*v) = \beta(a)v, \end{aligned}$$

  sodass $\beta(a) = WaW^*$ ist (siehe auch [Da], Lemma 5.6.1).

- Ein α der Form (5.8) ist ein $*$–Automorphismus. Denn α ist bijektiv, da sowohl unitäre Abbildungen als auch Permutationen bijektiv sind. Ebenso ist α linear, und es gilt $\alpha(ab) = \alpha(a)\alpha(b)$ und $\alpha(a)^* = \alpha(a^*)$.

- Gilt für ein unitäres $V \in \mathcal{M}_m$ ebenfalls $\beta(a) = VaV^*$ für alle $a \in \mathcal{M}_m$, dann auch $a = V^*\beta(a)V$, also $a = XaX^*$ oder $Xa = aX$ mit dem unitären $X := V^*W \in \mathcal{M}_m$. Dann muß aber $X = \lambda\mathbb{1}$ sein, also $W = \lambda V$.

  Denn hätte X zwei verschiedene Eigenwerte, dann würde X mit einem unitären Operator a, der den einen Eigenraum in den anderen abbildet, nicht vertauschen. □

Insbesondere sehen wir, dass in den in Beispiel 5.3 angesprochenen Extremfällen klassischer bzw. quantenmechanischer C^*–Algebren *alle* Automorphismen von der dort angegebenen Form sind:

Beispiel 5.5 1. Besteht der Phasenraum P nur aus n Punkten, so ist $\mathcal{A}_{\rm cl} = C(P, \mathbb{C}) \cong \mathbb{C}^n$ n–dimensional (und in der Darstellung (4.4) $n_1 = \ldots = n_n = 1$). Wir haben in Beispiel 5.3.1 gesehen, dass die Wirkungen $a \mapsto a \circ \pi$ der Permutationen $\pi : P \to P$ Automorphismen von $\mathcal{A}_{\rm cl}$ sind.

   Thm. 5.4 besagt nun, dass *jeder* Automorphismus von $\mathcal{A}_{\rm cl}$ von dieser Form ist.

2. Ist der Hilbertraum $\mathcal{H}$ n–dimensional, dann ist $\mathcal{A}_{\rm qm} = \mathcal{M}_n$ die (n^2–dimensionale) C^*–Algebra der komplexen $n \times n$–Matrizen. Wir haben in Beispiel 5.3.2 die inneren Automorphismen von $\mathcal{A}_{\rm qm}$ kennengelernt.

   Thm. 5.4 besagt nun, dass *jeder* Automorphismus dieser Matrixalgebra ein innerer Automorphismus ist, es also ein unitäres $U \in \mathcal{M}_m$ gibt mit $\alpha(a) = U^*aU$ für $a \in \mathcal{M}_m$.

Da wir die (reversiblen) Funktionen α von m Bits als Automorphismen der Algebra $\mathcal{A}_{\rm cl} = C(B^m, \mathbb{C})$ auffassen, stellt sich im Zusammenhang der Komplexität einer solchen Permutation α die Frage, wieviele auf nur wenigen (z.B. höchstens drei) Bits wirkende Permutationen $\alpha_1, \ldots, \alpha_{k_{\rm cl}} : \mathcal{A}_{\rm cl} \to \mathcal{A}_{\rm cl}$ wir benötigen, um jene in der Form

$$\alpha = \alpha_{k_{\rm cl}} \circ \alpha_{k_{\rm cl}-1} \circ \ldots \circ \alpha_2 \circ \alpha_1$$

darstellen zu können. Die α_l entsprechen dabei klassischen Schaltelementen mit drei Ein- und Ausgängen. Wir fragen also nach der Minimalzahl k_{cl} der benötigten Schaltelemente.

$\mathcal{A}_{\rm cl}$ kann als Unteralgebra der Diagonalmatrizen in $\mathcal{A}_{\rm qm} = \mathcal{M}_n$ (mit $n := |B^m| = 2^m$) verstanden werden. Hier stellt sich die Frage, wie viele wiederum

auf nur wenigen Qubits wirkenden Automorphismen $\beta_1, \ldots, \beta_{k_{\mathrm{qm}}} : \mathcal{A}_{\mathrm{qm}} \to \mathcal{A}_{\mathrm{qm}}$ wir benötigen, um α in der Form

$$\alpha = \beta_{k_{\mathrm{qm}}} \circ \beta_{k_{\mathrm{qm}}-1} \circ \ldots \circ \beta_2 \circ \beta_1$$

darstellen zu können.

Es gilt $k_{\mathrm{qm}} \leq k_{\mathrm{cl}}$, denn jede Permutation π von $\mathcal{A}_{\mathrm{cl}}$ kommt von dem (inneren) Automorphismus

$$a \mapsto U^* a U \quad \text{mit} \quad (U)_{ik} := \begin{cases} 1 & , k = \pi(i) \\ 0 & , \text{sonst} \end{cases}$$

von $\mathcal{A}_{\mathrm{qm}} \supset \mathcal{A}_{\mathrm{cl}}$.

Wie wir in Beispielen sehen werden, gibt es Fälle, in denen k_{qm} wesentlich kleiner als k_{cl} ist.

Bemerkung 5.6 Da physikalische Messwerte reelle Zahlen sind, werden die *selbstadjungierten* Operatoren als die eigentlich observablen Größen angesehen.

Allerdings ist dieses Argument im Fall von $\mathcal{A}_{\mathrm{qm}}$ nicht hinreichend, denn wie man am Beispiel der Matrix $a := \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right) \in \mathcal{B}(\mathbb{C}^2)$ mit $\sigma(a) = \{0\}$ sieht, gibt es dort auch nicht selbstadjungierte Operatoren mit reellem Spektrum.

Für nicht selbstadjungierte $a \in \mathcal{A}$ mit reellem Spektrum wäre allerdings der im nächsten Kapitel eingeführte Wert $\Phi(a) \in \mathbb{C}$ des Zustandes Φ im Allgemeinen nicht mehr reell, und damit insbesondere nicht mehr Erwartungswert der möglichen Ausgänge der Messung.

6 Zustände

Neben den Observablen sind die *Zustände*, in denen sich das betrachtete System befinden kann, Teil der physikalischen Modellierung.

Definition 6.1 • Ein *Zustand* Φ einer unitalen C^*–Algebra $\mathcal{A}$ ist ein lineares Funktional

$$\Phi : \mathcal{A} \to \mathbb{C},$$

das *positiv* ($\Phi(a) \geq 0$ für $a \geq 0$) und *identitätserhaltend* ($\Phi(\mathbb{1}) = 1$) ist.
• Die Menge aller Zustände von $\mathcal{A}$ wird mit $\mathcal{S}(\mathcal{A})$ bezeichnet.
• Ein Zustand Φ heißt *rein*, wenn er sich nicht als Konvexkombination $\Phi = \lambda\Phi_1 + (1-\lambda)\Phi_0$ mit $\Phi_i \in \mathcal{S}(\mathcal{A})$, $\Phi_0 \neq \Phi_1$ und $0 < \lambda < 1$ schreiben lässt.

$\mathcal{S}(\mathcal{A})$ ist eine Teilmenge des Dualraums $\mathcal{A}^*$ des Banachraums $\mathcal{A}$, und zwar eine *konvexe* Teilmenge, d.h. mit $\Phi_0, \Phi_1 \in \mathcal{S}(\mathcal{A})$ und $0 < \lambda < 1$ ist auch $\lambda\Phi_1 + (1-\lambda)\Phi_0 \in \mathcal{S}(\mathcal{A})$.

Trotzdem ist die Geometrie von $\mathcal{S}(\mathcal{A})$ im klassischen und im quantenmechanischen Fall sehr unterschiedlich.

Beispiel 6.2 1. Für $\mathcal{A}_{\rm cl} = C(P, \mathbb{C})$ induzieren Wahrscheinlichkeitsmaße μ auf dem (kompakten) Phasenraum P Zustände Φ_μ mit

$$\Phi_\mu(a) := \int_P a(x)\, d\mu(x).$$

Auch umgekehrt findet man nach dem Satz von Riesz-Markov[14] zu jedem Zustand $\Phi \in \mathcal{S}(\mathcal{A}_{\rm cl})$ ein eindeutiges Wahrscheinlichkeitsmaß μ mit $\Phi = \Phi_\mu$. Die reinen Zustände sind von der Form

$$\Phi(a) = a(x) \qquad (a \in \mathcal{A}_{\rm cl})$$

für ein beliebiges $x \in P$, entsprechen also Delta-Maßen δ_x auf P.

2. Für $\mathcal{A}_{\rm qm} = \mathcal{B}(\mathcal{H})$ induzieren *Dichtematrizen*, d.h. Spurklasseoperatoren $\rho \geq 0$ mit $\mathrm{tr}(\rho) = 1$, Zustände Φ_ρ mit

$$\Phi_\rho(a) := \mathrm{tr}(\rho a).$$

Ist $\dim(\mathcal{H}) < \infty$, dann kann man auch jeden Zustand Φ in der Form Φ_ρ darstellen. Für allgemeine Hilberträume $\mathcal{H}$ gilt das aber nicht mehr.

In einer endlich-dimensionalen Algebra $\mathcal{A}$ ist der Dualraum zu $\mathcal{A}$ isomorph, sodass folgender Satz gilt.

Satz 6.3 *Hat die C^*–Algebra $\mathcal{A}$ endliche Dimension, dann kann jeder Zustand $\Phi : \mathcal{A} \to \mathbb{C}$ in der Form*

$$\Phi(a) = \mathrm{tr}(\rho a) \qquad (a \in \mathcal{A})$$

mit $\rho \in \mathcal{A}$ geschrieben werden. Die Dichtematrix ρ ist eindeutig und positiv. Die kompakte konvexe Menge $\mathcal{S}(\mathcal{A}) \subseteq \mathcal{A}^*$ hat die reelle Dimension

$$\dim_{\mathbb{R}}(\mathcal{S}(\mathcal{A})) = \dim_{\mathbb{C}}(\mathcal{A}) - 1.$$

[14] **Satz von Riesz-Markov:** Es sei P ein kompakter Hausdorffraum. Dann gibt es für jedes positive lineare Funktional $\ell : C(P) \to \mathbb{C}$ ein eindeutiges Bairemaß μ auf P mit $\ell(f) = \int_P f\, d\mu$.

Bemerkung 6.4 $\mathbb{C}$–Vektorräume V können auch als $\mathbb{R}$-Vektorräume aufgefasst (*reellifiziert*) werden, wenn die Skalarmultiplikation auf reelle Zahlen eingeschränkt ist. Es gilt dann im endlich-dimensionalen Fall $\dim_{\mathbb{R}}(V) = 2\dim_{\mathbb{C}}(V)$.

Die *Dimension* $\dim(S)$ einer Teilmenge $S \subseteq \mathcal{A}$ eines Vektorraums $\mathcal{A}$ wird als die Dimension des kleinsten S enthaltenden affinen Unterraums $U \subseteq \mathcal{A}$ definiert.

Bew.: • Im Fall einer endlich-dimensionalen C^*–Algebra $\mathcal{A}$ können wir wegen der Isomorphie (4.4) zu einer direkten Summe von Matrixalgebren $\mathcal{A}$ als Unteralgebra einer Matrixalgebra $\mathcal{B}(\mathbb{C}^n)$ auffassen. Wegen der Hilbertraumstruktur von $\mathcal{B}(\mathbb{C}^n)$ mit innerem Produkt $\langle a, b\rangle := \mathrm{tr}(a^*b)$ können wir beliebige lineare Funktionale $\Phi : \mathcal{A} \to \mathbb{C}$ in der Form $\Phi(a) = \mathrm{tr}(\rho a)$ für ein geeignetes $\rho \in \mathcal{B}(\mathbb{C}^n)$ schreiben, wobei tr die Spur der $\mathcal{A}$ enthaltenden Matrixalgebra $\mathcal{B}(\mathbb{C}^n)$ ist. Wenn wir ρ in den in $\mathcal{A}$ gelegenen und den dazu bezüglich des Skalarproduktes senkrechten Anteil zerlegen, sehen wir, dass der Zustand vom letzteren Anteil unabhängig ist. Wir können also $\rho \in \mathcal{A}$ annehmen, wodurch die Dichtematrix ρ durch den Zustand fixiert ist.
• Wäre ρ nicht selbstadjungiert, also $c := \rho - \rho^* \neq 0$, dann besäße dieser anti-selbstadjungierte Operator einen imaginären Eigenwert $\lambda \in i\mathbb{R} \setminus \{0\}$. Bezeichnet $P_\lambda \in \mathcal{A}$ die Projektion auf den Eigenraum von λ, dann wäre $\mathrm{tr}(cP_\lambda) = \lambda\,\mathrm{tr}(P_\lambda)$, also

$$\Phi(P_\lambda) = \mathrm{tr}(\rho P_\lambda) = \tfrac{1}{2}\big(\mathrm{tr}((c - c^*)P_\lambda)\big) = \lambda\,\mathrm{tr}(P_\lambda) \notin \mathbb{R}.$$

Dies widerspräche aber der Positivität des Zustands Φ.
• $\rho = \rho^*$ besitzt also die Spektraldarstellung $\rho = \sum_{\lambda\in\sigma(\rho)} \lambda P_\lambda$ mit $\lambda \in \mathbb{R}$. Wäre ein $\lambda < 0$, so würde auch $\Phi(P_\lambda) = \lambda\,\mathrm{tr}(P_\lambda) < 0$ gelten, was wieder der Positivität des Zustands Φ widersprechen würde. Man kann also die positive Wurzel $\rho^{1/2} := \sum_{\lambda\in\sigma(\rho)} \sqrt{\lambda} P_\lambda \in \mathcal{A}$ aus ρ ziehen.
• Wäre die Dichtematrix $\rho \in \mathcal{A}$ nicht eindeutig, dann gäbe es ein $\tilde{\rho} \in \mathcal{A}\setminus\{0\}$ mit $\mathrm{tr}(\tilde{\rho}a) = 0$ für alle $a \in \mathcal{A}$, insbesondere also für $a := \tilde{\rho}^*$. Widerspruch!
• Der Spurzustand mit Dichtematrix $\frac{1}{n}\mathbb{1}$ ist in $\mathcal{S}(\mathcal{A})$ enthalten.

Bezeichnet $\mathrm{Sym}(\mathcal{A}) := \{a \in \mathcal{A} \mid a^* = a\}$ den reellen Unterraum der selbstadjungierten Elemente, dann ist wegen der eindeutigen Zerlegung $b = b_1 + ib_2$ eines beliebigen Elements $b \in \mathcal{A}$ mit $b_1, b_2 \in \mathrm{Sym}(\mathcal{A})$ (nämlich $b_1 = \frac{b+b^*}{z}, b_2 = \frac{b-b^*}{z_i}$) $\dim_{\mathbb{R}}(\mathrm{Sym}(\mathcal{A})) = \dim_{\mathbb{C}}(\mathcal{A}) = n$.

Mit dem reellen Unterraum $\mathrm{Sym}_0(\mathcal{A}) := \{a \in \mathrm{Sym}(\mathcal{A}) \mid \mathrm{tr}(a) = 0\}$ ist $\mathcal{S}(\mathcal{A})$ in dem affinen Unterraum $\frac{1}{n}\mathbb{1} + \mathrm{Sym}_0(\mathcal{A})$ der $\mathcal{S}(\mathcal{A})$ Dimension $n-1$ enthalten.

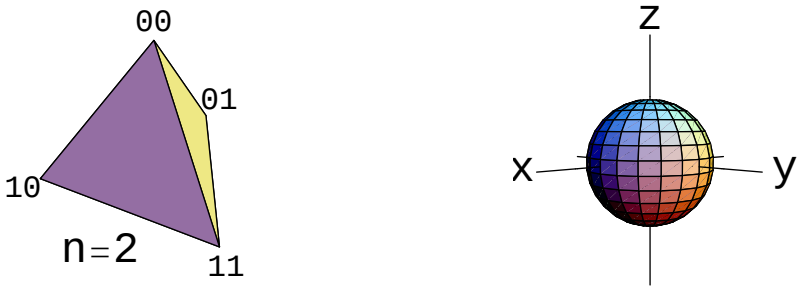


Abbildung 7: Simplex der Zustände für $n = 2$ Bits (links), Raum der Zustände eines Qubit (rechts)

Andererseits ist für alle $a \in \mathrm{Sym}_0(\mathcal{A})$ mit $\|a\| \leq \frac{1}{n}$ das selbstadjungierte Element $\frac{1}{n}\mathbb{1} + a$ positiv, denn mit $a = \sum_{\lambda\in\sigma(a)} \lambda P_\lambda$ ist $\lambda \geq -\frac{1}{n}$, also $\frac{1}{n}\mathbb{1} + a = \sum_{\lambda\in\sigma(a)}(\lambda + \frac{1}{n})P_\lambda \geq 0$. □

Der Einfachheit halber bezeichnen wir die ρ ebenfalls als Zustände.

Beispiel 6.5 1. Ist der Phasenraum $P = \{x_1, \ldots, x_n\}$, dann ist die Menge der Zustände von der Form

$$\mathcal{S}(\mathcal{A}_{\mathrm{cl}}) = \left\{ \sum_{l=1}^{n} \lambda_l \Phi_l \mid \lambda_l \geq 0, \sum_{l=1}^{n} \lambda_l = 1 \right\}, \quad \text{mit} \quad \Phi_l(a) := a(x_l).$$

Es gibt also genau n reine Zustände $\Phi_1, \ldots, \Phi_n$, und geometrisch hat $\mathcal{S}(\mathcal{A}_{\mathrm{cl}})$ die Form des $(n-1)$–dimensionalen *Simplex*

$$\left\{ \lambda \in \mathbb{R}^n \mid \lambda_l \geq 0, \sum_{l=1}^{n} \lambda_l = 1 \right\},$$

siehe Abb. 1. Nur die Eckpunkte dieses Simplex, nicht etwa sein gesamter Rand, sind reine Zustände.

2. Für $\dim(\mathcal{H}) = 2$ sei $\vec{\sigma} = (\hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_3) = \left(\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right), \left(\begin{smallmatrix} 0 & i \\ -i & 0 \end{smallmatrix}\right), \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right) \right)$ der Vektor der Pauli-Matrizen. Diese bilden zusammen mit der Einheitsmatrix $\mathbb{1} = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ eine Basis selbstadjungierter (und unitärer!) Operatoren in $\mathcal{A}_{\mathrm{qm}} := \mathcal{B}(\mathbb{C}^2)$.

Die selbstadjungierten Matrizen mit Spur 1 haben die Form

$$\rho_{\vec{b}} = \tfrac{1}{2}(\mathbb{1} + \vec{b} \cdot \vec{\sigma}) \qquad (\vec{b} \in \mathbb{R}^3), \tag{6.10}$$

denn wegen der Basiseigenschaft von $(\mathbb{1}, \hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_3)$ lässt sich jede Matrix mit Spur 1 eindeutig in dieser Form mit $\vec{b} \in \mathbb{C}^3$ schreiben, und genau für reelle $\vec{b}$ ist $\rho_{\vec{b}}$ selbstadjungiert. Die Forderung $\rho_{\vec{b}} \geq 0$ ist genau für die Vektoren in der Vollkugel $D^{(3)} := \{\vec{b} \in \mathbb{R}^3 \mid \|\vec{b}\| \leq 1\}$ des $\mathbb{R}^3$ erfüllt, denn $\rho_{\vec{0}} = \frac{1}{2}\mathbb{1} \geq 0$, und für $\vec{b} \neq \vec{0}$ gilt mit $b := \|\vec{b}\|$ und $\hat{b} := \vec{b}/b$:

$$\rho_{\vec{b}} = \lambda_+ P_+ + \lambda_- P_- \quad \text{mit} \quad \lambda_\pm := \tfrac{1}{2}(1 \pm b) \quad \text{und} \quad P_\pm := \tfrac{1}{2}(\mathbb{1} \pm \hat{b} \cdot \vec{\sigma}).$$

Dass die $P_\pm$ Projektionen sind, überprüft man anhand der Relation $\hat{\sigma}_m \hat{\sigma}_n + \hat{\sigma}_n \hat{\sigma}_m = 2\delta_{mn}\mathbb{1}$; die Eigenwerte $\lambda_\pm$ sind nur für $b \leq 1$ beide positiv.

Damit entspricht $\mathcal{S}(\mathcal{A}_{\mathrm{qm}})$ hier geometrisch der dreidimensionalen Vollkugel $D^{(3)}$, und die reinen Zustände entsprechen ihres Randes, der 2-Sphäre

$$\partial D^{(3)} = \left\{\vec{b} \in \mathbb{R}^3 \mid \|\vec{b}\| = 1\right\}.$$

Diese Kugel wird oft *Blochkugel* genannt.

Definition 6.6 Ein *Qubit* $\mathcal{B}$ ist die Menge der reinen Zustände von $\mathcal{B}(\mathbb{C}^2)$.

Identifiziert man wie üblich Zustände mit Dichtematrizen, dann ist also mit (6.10)

$$\mathcal{B} = \{\rho_{\vec{b}} \mid \|\vec{b}\| = 1\}.$$

Physikalisch ist $\mathcal{B}(\mathbb{C}^2)$ die Observablenalgebra des Spins eines Spin-1/2-Teilchens (z.B. eines Elektrons, Protons oder Neutrons). Also beschreiben die $\Phi \in \mathcal{B}$ die reinen Zustände eines solchen Teilchens. Der selbstadjungierten Observablen $\frac{1}{2}\vec{b} \cdot \vec{\sigma}$, $\|\vec{b}\| = 1$ entspricht die Messung der Spinkomponente in $\vec{b}$–Richtung.

Lemma 6.7 *Für $*$–Automorphismen $\alpha : \mathcal{A} \to \mathcal{A}$ einer endlich-dimensionalen C^*–Algebra $\mathcal{A}$ gilt*

$$\mathrm{tr}(\alpha(a)) = \mathrm{tr}(a) \qquad (a \in \mathcal{A}).$$

Bew.: Dies folgt aus Satz 5.4. □

Der wichtigste Unterschied zwischen den klassischen und quantenmechanischen Zustands-Räumen liegt in der Frage nach der *eindeutigen Zerlegbarkeit* eines Zustandes in reine Zustände.

Jeder Punkt eines Simplex lässt sich eindeutig als Konvexkombination der Eckpunkte schreiben, sodass klassische Zustände sich eindeutig zerlegen lassen.

Dagegen ist dies quantenmechanisch schon für $\mathcal{B}(\mathbb{C}^2)$ nicht mehr der Fall, denn beispielsweise gilt für den *Spurzustand* $\rho_{\vec{0}} = \frac{1}{2}\mathbb{1}$

$$\rho_{\vec{0}} = \tfrac{1}{2}\rho_{\vec{b}} + \tfrac{1}{2}\rho_{-\vec{b}} \quad \text{für beliebige} \quad \vec{b} \in \partial D^{(3)}.$$

In diesem Beispiel kann man den Zustand sogar aus beliebig vielen voneinander verschiedenen reinen Zuständen konvex kombinieren.

Da unreine Zustände unvollständige Information über den wirklichen (reinen) Zustand codieren, sehen wir hier, dass die Quantenmechanik keine rein wahrscheinlichkeitstheoretische Deutung zulässt.

Die Frage ist keineswegs akademisch. Beispielsweise können wir die Situation betrachten, in der ein (klassisches) Bit, d.h. ein Element x des Phasenraums $B = \{0, 1\}$, durch einen gestörten Kanal gesendet wird, der mit Wahrscheinlichkeit p das Bit x zu $1 - x$ invertiert. Messen wir ...

6.1 Messungen

Im uns interessierenden endlich-dimensionalen Fall können wir einen normalen Operator $a \in \mathcal{A}$ eindeutig in der Form (4.5) schreiben (Spektralzerlegung).

Es werde nun die selbstadjungierte Observable $a \in \mathcal{A}$ gemessen.

Definition 6.8 Der Zustand des (klassischen oder quantenmechanischen) Systems vor der Messung sei $b \mapsto \Phi(b)$ $(b \in \mathcal{A})$.

- *Nach* der Messung von a, aber *vor* der Kenntnisnahme des Messergebnisses ist der Zustand gleich

  $$b \mapsto \sum_{\lambda \in \sigma(a)} \Phi(P_\lambda b P_\lambda) \qquad (b \in \mathcal{A}).$$

  Die möglichen *Messwerte* sind die $\lambda \in \sigma(a)$, die *Wahrscheinlichkeit* ihres Auftretens $\Phi(P_\lambda)$.

- *Nach* der Kenntnisnahme des Messergebnisses $\mu \in \sigma(a)$ ist der Zustand gleich

$$b \mapsto \Phi(P_\mu b P_\mu)/\Phi(P_\mu) \qquad (b \in \mathcal{A}).$$

Wir sehen, dass die quantenmechanische im Gegensatz zur klassischen Messung den Zustand i.a. verändert (außer wenn die Dichtematrix mit den Projektionen kommutiert). Physikalisch gesehen liegt dies an der Rückwirkung des Messapparates auf das System.

Dass dagegen die Kenntnisnahme des Messergebnisses den Zustand verändert, ist ein schon klassisch auftauchendes Phänomen. Da der (gemischte) Zustand unsere Kenntnis vom vorliegenden System beschreibt, und unsere Information durch die Kenntnisnahme des Messergebnisses verändert wird, ist dies nicht sonderlich erstaunlich.

Bemerkung 6.9 Auch klassisch können wir aber nicht erwarten, dass uns *eine* Messung vollständige Information über einen Zustand liefert, er nach Kenntnisnahme der Messung also rein ist. Beispielsweise müssen wir im Allgemeinen alle n Observablen $a_i : B^n \to \{0,1\}, a_i(b_1,\ldots,b_n) = b_i$ messen, um die Bitfolge $b_1,\ldots,b_n$ zu kennen. Allerdings würde hier *eine* Messung der Observable $a := \sum_{i=1}^n a_i 2^{-i}$ genügen. Denn während die Spektralprojektoren $P_1 = a_i$ und $P_0 = \mathbb{1} - P_1$ jeweils Rang 2^{n-1} besitzen, projezieren die Projektoren in der Spektraldarstellung von a auf einen nur eindimensionalen Unterraum.

6.2 Ordnung der Zustände

Betrachten wir nun das Spektrum $\sigma(\rho)$ mit Multiplizität $\mathrm{rank}(P_\lambda)$ der Projektionen, dann bekommen wir n Zahlen, die wir nach absteigender Größe geordnet in der Form

$$1 \geq \rho_1 \geq \rho_2 \geq \ldots \geq \rho_n \geq 0 \quad \text{mit} \quad \sum_{i=1}^n \rho_i = 1$$

notieren, und setzen $\rho(k) := \sum_{i=1}^k \rho_i$, $k = 1,\ldots,n$.

Definition 6.10 (siehe [Th]) Ein Zustand σ von $\mathcal{A}$ heißt *gemischter als* ein Zustand ρ (bzw. ρ *reiner als* σ), symbolisch $\sigma \succeq \rho$, wenn

$$\sigma(k) \leq \rho(k) \qquad (k = 1,\ldots,n). \tag{6.11}$$

Damit werden die Zustände zwar nicht total geordnet, denn für $n > 2$ kann das Vorzeichen von $\rho(k) - \sigma(k)$ mit k variieren. Aber immerhin sind die reinen Zustände ρ mit $\rho_1 = 1$, $\rho_i = 0, i > 1$ mit allen anderen vergleichbar, und natürlich reiner als diese.

Die Dichtematrizen ρ reiner Zustände sind eindimensionale Projektionen und umgekehrt, denn wäre $\mathrm{rank}(\rho) > 1$, dann könnte man ρ als Konvexkombination orthogonaler Projektionen schreiben.

Schreibt man $\sigma \sim \rho$ falls $\sigma \succeq \rho \succeq \sigma$, so ist dies eine Äquivalenzrelation, die den Raum der Zustände in gleich gemischte zerlegt, also solche, deren Dichtematrizen gleiches Spektrum (mit gleicher Multiplizität) besitzen.

Satz 6.11 *1. Automorphismen $\alpha : \mathcal{A} \to \mathcal{A}$ erhalten den Mischungsgrad:*

$$\rho \succeq \alpha(\rho) \succeq \rho.$$

2. *Konvexkombination erhöht den Mischungsgrad:*
   Für Dichtematrizen $\rho_0, \rho_1, \sigma \in \mathcal{A}$ und $\rho_\lambda := (1-\lambda)\rho_0 + \lambda\rho_1$, $\lambda \in (0,1)$ gilt:

$$\rho_0 \succeq \sigma \quad \textit{und} \quad \rho_1 \succeq \sigma \implies \rho_\lambda \succeq \sigma.$$

3. *Messungen ohne Kenntnisnahme des Ergebnisses erhöhen den Mischungsgrad:*

$$\tilde{\rho} := \sum_{\lambda \in \sigma(a)} P_\lambda \, \rho P_\lambda \succeq \rho \quad \textit{für Messung von} \quad a = \sum_{\lambda \in \sigma(a)} \lambda P_\lambda \in \mathcal{A}.$$

Bew.: 1) Die erste Aussage folgt aus der Invarianz des Spektrums unter Automorphismen.
2) Die zweite folgt aus Min-Max-Prinzip für $k = 1, \ldots, n$ (siehe [RS2])

$$\begin{aligned} \rho_\lambda(k) &= \sup_{q_k \in Q_k} \mathrm{tr}(\rho_\lambda q_k) \leq (1-\lambda) \sup_{q_k \in Q_k} \mathrm{tr}(\rho_0 q_k) + \lambda \sup_{q_k \in Q_k} \mathrm{tr}(\rho_1 q_k) \\ &= (1-\lambda)\rho_0(k) + \lambda\rho_1(k), \end{aligned}$$

wobei $Q_k \subset \mathcal{A}$ die Menge der Projektionen vom Rang k bezeichnet.
3) Für die unmittelbar nach der Messung von $a \in \mathcal{A}$ vorliegende Dichtematrix

$$\tilde{\rho} = \sum_{\lambda \in \sigma(a)} P_\lambda \, \rho P_\lambda = \sum_{l=1}^{n} \tilde{\rho}_l \tilde{P}_l$$

mit Spektraldarstellung durch eindimensionale Projektionen $\tilde{P}_l$ gilt nach dem Min-Max-Prinzip für $k = 1, \ldots, n$

$$\tilde{\rho}(k) = \mathrm{tr}(\tilde{\rho}\tilde{q}_k) = \mathrm{tr}(\rho\tilde{q}_k) \leq \sup_{q_k \in Q_k} \mathrm{tr}(\rho q_k) = \rho(k),$$

wobei $\tilde{q}_k := \sum_{l=1}^{k} \tilde{P}_l$. Die zweite Gleichung gilt, da (bei einer Wahl der $\tilde{P}_l$, für die P_λ in eine Summe von eindimensionalen Projektionen $\tilde{P}_l$ zerlegt wird) der Kommutator $[P_\lambda, \tilde{P}_l] = 0$ ist. □

Bemerkungen 6.12 1. Während sich diese Ordnung auf Dichtematrizen in unendlichdimensionalen Algebren übertragen lässt, können wir dort nicht die Existenz eines gemischtesten Zustandes erwarten. Im Fall $\dim(\mathcal{A}) = n$ ist der *Spurzustand* mit Dichtematrix $n^{-1}\mathbb{1}_n$ der (einzige) gemischteste, denn es gibt keine andere positive Matrix ρ mit den Eigenwerten $\rho_i = 1/n$.

2. Die Zeitevolution ändert als Automorphismengruppe den Zustand nicht.

3. Natürlich wird klassisch (d.h. für $\mathcal{A}$ abelsch) der Zustand bei der Messung nicht verändert.

4. Selbst im klassischen Fall verringert die Kenntnisnahme des Messergebnisses μ den Mischungsgrad nicht immer, d.h. es gilt nicht notwendig

   $$P_\mu \, \rho P_\mu / \mathrm{tr}(P_\mu \, \rho P_\mu) \preceq \rho.$$

   Ein Gegenbeispiel ist die drei-dimensionale abelsche Algebra, mit $\rho := \mathrm{diag}(0.8, 0.1, 0.1)$. Kenntnisnahme des Messwertes $\mu = 1$ der Observable $a := \mathrm{diag}(0, 1, 1)$ mit Projektion $P_\mu = a$ führt zur Dichtematrix

   $$P_\mu \, \rho P_\mu / \mathrm{tr}(P_\mu \, \rho P_\mu) = \mathrm{diag}(0, 0.5, 0.5),$$

   die weder reiner noch gemischter als ρ ist.

   Oft ist aber P_μ eindimensional, der Zustand nach der Messung also rein.

6.3 Entropie

Die Entropie $S(\Phi)$ eines Zustandes $\Phi : \mathcal{A} \to \mathbb{C}$ mit Dichtematrix ρ ist eine nichtnegative Zahl, die seine Gemischtheit quantifiziert. Eine solche Kenngröße sollte die beiden folgenden Eigenschaften besitzen:

1. Sie sollte nur von der Äquivalenzklasse abhängen, also nur vom Spektrum der Dichtematrix (mit Multiplizität). Diese Abhängigkeit sollte stetig sein.

2. Ist die C^*–Algebra $\mathcal{A}$ eine direkte Summe $\mathcal{A} = \bigoplus_{l=1}^{n} \mathcal{A}_l$, dann lässt sich eine Dichtematrix ρ auf $\mathcal{A}$ in der Form $\rho = \bigoplus_{l=1}^{n} p_l \rho_l$ mit Dichtematrix ρ_l auf $\mathcal{A}_l$ schreiben. Diese Darstellung ist (bis auf den Fall $p_l = 0$) eindeutig, und die Koeffizienten $p_l \geq 0$ haben die Summe 1.

   Wir bezeichnen mit $p := (p_1, \ldots, p_n)$ den Vektor der Wahrscheinlichkeiten p_l, dass sich das durch ρ modellierte Teilchen in $\mathcal{A}_l$ befindet. Dieser lässt sich als Dichtematrix auf der n–dimensionalen abelschen Algebra auffassen.

   Es soll nun

$$S(\rho) = S(p) + \sum_{l=1}^{n} p_l \cdot S(\rho_l) \tag{6.12}$$

   gelten, die Gesamtentropie also die Summe aus der Entropie von p und dem gewichteten Mittel der Entropien der Zustände ρ_l auf den Algebren $\mathcal{A}_l$ sein.

Man beachte, dass die zweite Forderung Entropien von Zuständen auf verschiedenen Algebren miteinander verknüpft. Dies wird später wichtig, wenn wir das Tensorprodukt und physikalische Teilsysteme betrachten.

Erfüllt S diese Forderungen, dann gilt gleiches für $c \cdot S$, wenn $c \geq 0$. Wir können also noch die Maßeinheit der Entropie festlegen. Weitere Freiheiten existieren aber nicht:

Satz 6.13 *Bis auf eine positive multiplikative Konstante ist die* **von Neumann–Entropie**

$$S(\rho) := -\mathrm{tr}(\rho \ln(\rho))$$

das einzige Funktional auf den Dichtematrizen, das die obigen Bedingungen erfüllt.

Bew.: • Dass die von Neumann-Entropie die beiden Bedingungen erfüllt, sieht man sofort.
• Wegen der in 1) geforderten Abhängigkeit nur vom Spektrum reicht es aus, diagonale Dichtematrizen zu betrachten. Wir betrachten zunächst die Entropie der Spurzustände $\rho_n := \frac{1}{n}\mathbb{1}_n$ auf den n–dim. Hilberträumen. Es ist

$$\rho_{mn} = \bigoplus_{l=1}^{n} p_l \rho_m,$$

mit Wahrscheinlichkeitsvektor $p = (p_1, \ldots, p_n) := (\frac{1}{n}, \ldots, \frac{1}{n})$, sodass wegen Forderung (6.12)

$$S(\rho_{mn}) = S(p) + \sum_{l=1}^{n} \frac{1}{n} S(\rho_m) = S(\rho_m) + S(\rho_n)$$

ist. Die einzigen Lösungen dieser Funktionalgleichung sind (wegen der geforderten Stetigkeit, siehe Thirring [Th] Band 4, Kap. 2.2) durch

$$S(\rho_n) = c \ln(n) \qquad (n \in \mathbb{N}),$$

wobei wegen unserer Forderung $S \geq 0$ der Parameter c positiv gewählt werden muss. Wir wählen $c := 1$, und schreiben $\eta(x) := -x \ln(x)$ für $x > 0$, $\eta(0) := 0$.
• In einem zweiten Schritt betrachtet man Diagonalmatrizen mit rationalen Einträgen $p = (p_1, \ldots, p_n) = (q_1/m, \ldots, q_n/m)$, $q_l \in \mathbb{N}_0$. Wegen $\rho_m = \bigoplus_{l=1}^{n} p_l \rho_{q_l}$ und (6.12) folgt

$$S(p) = S(\rho_m) - \sum_{l=1}^{n} p_l S(\rho_{q_l}) = \ln(m) - \sum_{l=1}^{n} p_l \ln(q_l) = \sum_{l=1}^{n} \eta(p_l).$$

• Wegen der geforderten Stetigkeit der Abbildung $\rho \mapsto S(\rho)$ gilt diese Identität dann auch für irrationale Wahrscheinlichkeiten p_l. □

Es lässt sich nun zeigen [Th], dass die Entropie die Ordnung der Zustände widerspiegelt in dem Sinn

$$\rho \succeq \sigma \Longrightarrow S(\rho) \geq S(\sigma). \tag{6.13}$$

Daraus folgt insbesondere, dass $0 \leq S(\rho) \leq S(\frac{1}{n}\mathbb{1}_n) = \ln(n)$ gilt.

Corollar 6.14 *Ist ρ Dichtematrix in einer (endlich-dimensionalen) C^*–Algebra $\mathcal{A}$, dann gilt für die Dichtematrix ρ' nach einer Messung aber vor Kenntnisnahme des Messergebnisses*

$$S(\rho') \geq S(\rho)$$

(ist $\mathcal{A}$ abelsch, dann gilt $S(\rho') = S(\rho)$).

Bew.: Dies folgt aus Satz 6.11 und (6.13). □

Bemerkung 6.15 Auch klassisch kann paradoxerweise die *Kenntnisnahme* einer Messung unsere Information über den Zustand *vermindern*. Im Beispiel der dreidimensionalen abelschen Algebra aus Bemerkung 6.12.4. ist die Entropie des Zustandes $\rho = \mathrm{diag}(0.8, 0.1, 0.1)$ gleich

$$S(\rho) = \eta(0.8) + 2\eta(0.1) \approx 0.639$$

und damit geringer als die Entropie $S(\mathrm{diag}(0, 0.5, 0.5)) = \ln(2) \approx 0.693$ nach Kenntnisnahme des Messwertes $\mu = 1$ der Observable $\mathrm{diag}(0, 1, 1)$.

Immerhin wird *im Mittel* vieler Messungen klassisch die Information über den Zustand erhöht. Dann ist in einem Anteil $w_\lambda := \Phi(P_\lambda)$ der Fälle der Messwert $\lambda \in \sigma(a)$ der Observablen a aufgetreten, und nach Kenntnisnahme von λ besitzt der Zustand die Dichtematrix

$$\rho_\lambda := \frac{P_\lambda \rho P_\lambda}{w_\lambda}, \tag{6.14}$$

unter der Annahme, dass der Messwert λ mit Wahrscheinlichkeit $w_\lambda > 0$ auftritt. Für den Fall $w_\lambda = 0$ können wir eine beliebige Dichtematrix ρ_λ wählen, z.B. $\rho_\lambda := P_\lambda / \mathrm{tr}(P_\lambda)$.

Im Mittel über die Wahrscheinlichkeiten w_λ ist also damit die Entropie nach Kenntnisnahme des Messergebnisses gleich

$$\sum_{\lambda \in \sigma(a)} w_\lambda S(\rho_\lambda), \tag{6.15}$$

und zwar sowohl klassisch wie quantenmechanisch. Man beachte, dass dieser Ausdruck *nicht* stetig vom Zustand vor der Messung abhängt, dessen Dichtematrix ρ in die Definition (6.14) eingeht, auch nicht im abelschen Fall von Diagonalmatrizen. Sind nämlich mehrere Eigenwerte von ρ gleich, dann werden i.A. nicht alle

ρ_λ reine Zustände sein, sodass (6.15) größer als Null ist. In jeder Umgebung von $\rho \in \mathcal{A}$ liegen dann aber Dichtematrizen $\tilde{\rho}$ mit nicht degenerierten Eigenwerten, für die dann $\sum_{\lambda\in\sigma(a)} \tilde{\mu}_\lambda S(\tilde{\rho}_\lambda) = 0$ gilt.

Satz 6.16 *Ist $\mathcal{A}$ eine endlich–dimensionale C^*–Algebra, dann gilt für die Dichtematrix $\tilde{\rho} = \sum_{\lambda\in\sigma(a)} P_\lambda \rho P_\lambda$ nach der Messung ohne Kennnisnahme*

$$\sum_{\lambda\in\sigma(a)} w_\lambda S(\rho_\lambda) = S(\tilde{\rho}) - \sum_{\lambda\in\sigma(a)} \eta(w_\lambda).$$

Im Mittel wird also durch Kenntnisnahme des Messergebnisses die Information über den Zustand *nach* der Messung erhöht, und zwar klassisch wie quantenmechanisch um die Entropie der Wahrscheinlichkeitsverteilung w der Messergebnisse.

Bew.: Es ist $\tilde{\rho} = \sum_{\lambda\in\sigma(a)} w_\lambda \rho_\lambda$, also

$$\begin{aligned} S(\tilde{\rho}) - \sum_{\lambda\in\sigma(a)} w_\lambda S(\rho_\lambda) &= \sum_{\lambda\in\sigma(a)} w_\lambda \mathrm{tr}(-\rho_\lambda \ln(w_\lambda\rho_\lambda) + \rho_\lambda \ln(\rho_\lambda)) \\ &= \sum_{\lambda\in\sigma(a)} w_\lambda \mathrm{tr}(-\rho_\lambda \ln(w_\lambda P_\lambda)) \\ &= \sum_{\lambda\in\sigma(a)} -w_\lambda \ln(w_\lambda)\mathrm{tr}(\rho_\lambda) = \sum_{\lambda\in\sigma(a)} \eta(w_\lambda) \end{aligned}$$ □

Quantenmechanisch kann unsere Information über den ursprünglichen Zustand nach Kenntnisnahme des Messergebnisses auch im Mittel geringer sein als vor der Messung, während dies im klassischen Fall nicht vorkommen kann.

Beispiel 6.17 Es sei $\rho_{\vec{b}}$ der Zustand eines quantenmechanischen Qubits (siehe Beispiel 6.5.2. Wir messen die Observable $a := \vec{c}\cdot\vec{\sigma}$ mit einem zu $\vec{b}$ senkrechten Vektor $\vec{c} \in \mathbb{R}^3$, $\|\vec{c}\| = 1$. Die Observable besitzt die Form $a = P_+ - P_-$ mit den Projektoren $P_\pm = \frac{1}{2}(\mathbb{1} \pm \vec{c}\cdot\vec{\sigma})$ vom Rang 1.

Nach Messung ist der Zustand zwar rein (mit Dichtematrix P_+ oder P_- entsprechend den Messergebnissen 1 oder -1), aber unabhängig von $\vec{b}$ haben beide Messergebnisse Wahrscheinlichkeit $1/2$.

7 Zusammengesetzte Systeme

Wir können zwei disjunkte Raumbereiche als Aufenthaltsgebiete eines Teilchens vereinigen, und wir können den Zustand zweier Teilchen aus Zuständen der beiden Teilchen zusammensetzen. Die erste Konstruktion führt zur direkten Summe, die zweite zum direkten Produkt = Tensorprodukt.

7.1 Direkte Summe

Ist der (kompakte) Phasenraum P eines *klassischen* Systems disjunkte Vereinigung

$$\mathcal{P} = \bigcup_{l=1}^{k} \mathcal{P}_l$$

von Phasenräumen $\mathcal{P}_l$ ($\mathcal{P}_l \cap \mathcal{P}_m = \emptyset$ für $l \neq m$), dann ist die $\mathcal{P}$ zugeordnete abelsche C^*–Algebra $\mathcal{A}_{\rm cl}(\mathcal{P}) = C(\mathcal{P}, \mathbb{C})$ von der Form

$$\mathcal{A}_{\rm cl}(\mathcal{P}) = \bigoplus_{l=1}^{k} \mathcal{A}_{\rm cl}(\mathcal{P}_l).$$

Ist dagegen der Hilbertraum $\mathcal{H}$ eines *quantenmechanischen* Systems direkte Summe

$$\mathcal{H} = \bigoplus_{l=1}^{k} \mathcal{H}_l$$

von Hilberträumen $\mathcal{H}_l$, dann gilt für die quantenmechanische C^*–Algebra $\mathcal{A}_{\rm qm}(\mathcal{H}) = B(\mathcal{H})$

$$\mathcal{A}_{\rm qm}(\mathcal{H}) \supseteq \bigoplus_{l=1}^{k} \mathcal{A}_{\rm qm}(\mathcal{H}_i),$$

und außer im Trivialfall nulldimensionaler $\mathcal{H}_i$ herrscht keine Gleichheit.

Beispiel 7.1 Betrachten wir ein Teilchen in dem Doppelmuldenpotential

$$V : \mathbb{R} \to \mathbb{R} \quad , \quad V(q) = (q^2 - 1)^2.$$

- *Klassisch* ist der Phasenraum der $\mathbb{R}^2$ und die Hamiltonfunktion $H(p,q) = \frac{1}{2}p^2 + V(q)$. Für Energien $0 < E < 1$ besteht die Energieschale $P := H^{-1}(E)$ aus zwei Komponenten $P = P_- \cup P_+$, die beide kompakt sind und unter der Transformation $(p,q) \mapsto (p,-q)$ ineinander übergehen. Das klassische Teilchen mit der Energie E befindet sich entweder in der linken Mulde (d.h. in P_-) oder der rechten Mulde (d.h. in P_+) des Potentials V.

- *Quantenmechanisch* ist der Hilbertraum $\mathcal{H} = L^2(\mathbb{R})$ direkte Summe $\mathcal{H} = \mathcal{H}_- \oplus \mathcal{H}_+$ der Hilberträume $\mathcal{H}_- := L^2((-\infty, 0])$ und $\mathcal{H}_+ := L^2([0,\infty))$, und der Hamiltonoperator $H := -\frac{\hbar^2}{2}\frac{d^2}{dq^2} + V(q)$ auf $\mathrm{tr}(\rho H)$ ist der Erwartungswert der Energie des duch die Dichtematrix ρ beschriebenen Teilchens. Fasst man das Plancksche Wirkungsquantum $\hbar > 0$ als kleinen Parameter auf, dann ist die Eigenfunktion $\varphi_n \in \mathcal{H}$ mit Eigenwert $E_n \leq 1 - \varepsilon$ im wesentlichen auf den Intervallen $\{q \in \mathbb{R} \mid V(q) \leq E_n\}$, also den klassisch für diese Energie erlaubte Gebiete, lokalisiert:

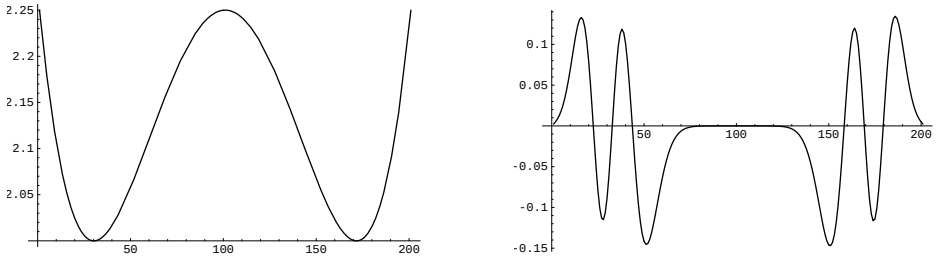


Abbildung 8: Doppelmuldenpotential V (links); Eigenfunktion φ_6 des Schrödingeroperators (rechts)

Der Operator $\Pi : \mathcal{H} \to \mathcal{H}$, $(\Pi\psi)(q) := \psi(-q)$ ist selbstadjungiert und unitär, zerlegt also $\mathcal{H}$ in die Eigenräume zu den Eigenwerten ± 1 (Parität). Da Π mit $\mathcal{H}$ kommutiert, sind die φ_n auch Eigenfunktionen von Π, und $\Pi\varphi_n = (-1)^n\varphi_n$. Es ist

$$\Pi \in B(\mathcal{H}) \text{ , aber } \Pi \notin B(\mathcal{H}_-) \oplus B(\mathcal{H}_+).$$

Metaphorisch könnte man sagen, dass die quantenmechanische Vereinigung mehr als die Summe ihrer Teile ist. Insbesondere gibt es reine Zustände

$$\Phi : \mathcal{A}_{\mathrm{qm}}(\mathcal{H}) \to \mathbb{C} \quad \text{auf} \quad \mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2$$

die weder auf $\mathcal{A}_{\rm qm}(\mathcal{H}_1) \oplus 0$ noch auf $0 \oplus A_{\rm qm}(\mathcal{H}_2)$ verschwinden. Wir müssen ja nur die Projektion auf $v = (v_1, v_2) \in \mathcal{H}$ mit $v_i \neq 0$ als Dichtematrix von Φ benutzen. Klassisch dagegen sind die reinen Zustände in einer Phasenraumkomponente lokalisiert, denn sie entsprechen ja den δ–Funktionen.

7.2 Das Tensorprodukt

Wir führen in diesem Kapitel das Tensorprodukt von Vektorräumen und das Tensorprodukt von (endlich-dimensionalen) C^*-Algebren ein.

Das Tensorprodukt $\mathbb{C}^m \otimes \mathbb{C}^n$ der Vektorräume $\mathbb{C}^m$ und $\mathbb{C}^n$ ist isomorph zum Vektorraum $\mathbb{C}^{mn}$ der Produktdimension.

Schon als Menge ist $\mathbb{C}^m \otimes \mathbb{C}^n$ verschieden vom *kartesischen Produkt* $\mathbb{C}^m \times \mathbb{C}^n$ der geordneten Paare (v_1, v_2) von Vektoren $v_1 \in \mathbb{C}^m$ und $v_2 \in \mathbb{C}^n$. $\mathbb{C}^m \times \mathbb{C}^n$ wird ja durch die Definition

$$(v_1, v_2) + (w_1, w_2) := (v_1 + w_1, v_2 + w_2) \quad , \quad a(v_1, v_2) := (av_1, av_2)$$

zum Vektorraum $\mathbb{C}^m \oplus \mathbb{C}^n$. Dieser Summenraum besitzt die Dimension $m + n$ und ist isomorph zum Vektorraum $\mathbb{C}^{m+n}$.

Tensorprodukt von Vektorräumen

Die jetzt folgende allgemeine Definition des Tensorproduktes basiert trotzdem auf dem kartesischen Produkt.

Sie ist nur schwer eingänglich, und wir werden sie für praktische Rechnungen nicht brauchen, da wir uns auf endlichdimensionale Räume $\mathcal{H}_i$ beschränken. Sie legt aber die Sicht auf den geometrischen Gehalt frei.

Es seien $\mathcal{H}_1, \ldots, \mathcal{H}_k$ $\mathbb{C}$–Vektorräume und

$$\mathcal{M} := \{f : \mathcal{H}_1 \times \ldots \times \mathcal{H}_k \to \mathbb{C} \mid |\mathrm{supp}(f)| < \infty\}$$

der $\mathbb{C}$–Vektorraum der Funktionen mit endlichem *Träger* $\mathrm{supp}(f) := \{x \mid f(x) \neq 0\}$ auf ihrem kartesischen Produkt.

Damit können wir f als endliche Linearkombinationen von Funktionen

$$\delta_x : \times_{l=1}^k \mathcal{H}_l \to \mathbb{C} \quad , \quad \delta_x(y) := \begin{cases} 1 & , \quad y = x \\ 0 & , \quad y \neq x \end{cases}$$

mit Träger $x \in \times_{l=1}^k \mathcal{H}_k$ darstellen.

$\mathcal{M}_0 \subseteq \mathcal{M}$ sei der von den Vektoren der Form

$$\delta_{(x_1,\ldots x'_j+x''_j,\ldots,x_k)} - \delta_{(x_1,\ldots,x'_j,\ldots,x_k)} - \delta_{(x_1,\ldots,x''_j,\ldots,x_k)}$$

und

$$\delta_{(x_1,\ldots,ax_j,\ldots,x_k)} - a\delta_{(x_1,\ldots,x_k)} \quad (a \in \mathbb{C})$$

aufgespannte Unterraum.

Definition 7.2 Das *(algebraische) Tensorprodukt*

$$\mathcal{H} := \bigotimes_{l=1}^{k} \mathcal{H}_l \equiv \mathcal{H}_1 \otimes \ldots \otimes \mathcal{H}_k$$

ist der Quotientenraum $\mathcal{H} := \mathcal{M}/\mathcal{M}_0$.

Es sei

$$t : \times_{l=1}^{k} \mathcal{H}_l \to \mathcal{H} \quad , \quad x_1 \otimes \ldots \otimes x_k := t(x_1, \ldots, x_k) := \delta_{(x_1,\ldots,x_k)} + \mathcal{M}_0 = .$$

t ist multilinear, und da die $\delta_{(x_1,\ldots,x_k)}$ eine Basis von $\mathcal{M}$ bilden, ist

$$\mathrm{span}(t(\times_{l=1}^{k} \mathcal{H}_l)) = \mathcal{H}.$$

Definition 7.3 Vektoren $v \in t(\times_{l=1}^{k} \mathcal{H}_l) \subseteq \mathcal{H}$ heißen *faktorisierbar*, die anderen Vektoren aus $\mathcal{H}$ heißen *verschränkt*.

Beispiel 7.4 Es sei $e_1^{(l)}, \ldots, e_{n_l}^{(l)}$ eine Basis von $\mathcal{H}_l := \mathbb{C}^{n_l}$. Dann bilden für $I := I_1 \times \ldots \times I_k$, $I_l := \{1, \ldots, n_l\}$ die $n := n_1 \cdot n_2 \cdot \ldots \cdot n_k$ Vektoren

$$e_i \equiv e_{i_1}^{(1)} \otimes \ldots \otimes e_{i_n}^{(n)} = t\left(e_{i_1}^{(1)}, \ldots, e_{i_k}^{(k)}\right) \in \mathcal{H} \qquad \left(i := (i_1, \ldots, i_k) \in I\right)$$

eine Basis von $\mathcal{H} := \otimes_{l=1}^{k} \mathcal{H}_l$. Damit ist $\dim(\mathcal{H}) = n$.

Die Vektoren in $\mathcal{H} = \mathbb{C}^2 \otimes \mathbb{C}^2$ können in der Form $v = a_{11}e_{11} + a_{12}e_{12} + a_{21}e_{21} + a_{22}e_{22}$ mit $a_{ij} \in \mathbb{C}$ geschrieben werden und sind genau dann faktorisierbar, wenn

$$a_{11}a_{22} - a_{12}a_{21} = 0,$$

denn $t\left(b_1e_1^{(1)} + b_2e_2^{(1)}, c_1e_1^{(2)} + c_2e_2^{(2)}\right) = b_1c_1e_{11} + b_1c_2e_{12} + b_2c_1e_{21} + b_2c_2e_{22}$.

Beispiel 7.5 Es seien $\mathcal{H}_l := C(P_l, \mathbb{C})$ die Vektorräume der (stetigen) Funktionen $f : P_l \to \mathbb{C}$ auf den endlichen Mengen P_l, $l = 1, \ldots, k$. Dann ist deren Tensorprodukt $\mathcal{H} := \mathcal{H}_1 \otimes \ldots \otimes \mathcal{H}_k$ isomorph zu $C(P, \mathbb{C})$ mit $P := P_1 \times \ldots \times P_k$, mit dem Isomorphismus $I : \mathcal{H} \to C(P, \mathbb{C})$,

$$I(f_1 \otimes \ldots \otimes f_k)(p_1, \ldots, p_k) := f_1(p_1) \cdot \ldots \cdot f_n(p_k) \qquad (p_l \in P_l)$$

für $f_l \in \mathcal{H}_l$, linear fortgesetzt.

Hier sind die faktorisierbaren Funktionen $f \in C(P, \mathbb{C})$ gerade diejenigen, die Produkte von Koordinatenfunktionen sind.

Tensorprodukt von Hilberträumen

Definition 7.6 Sind $\mathcal{H}_1, \ldots, \mathcal{H}_k$ $\mathbb{C}$–Hilberträume, dann wird auf dem Tensorprodukt $\mathcal{H} := \bigotimes_{l=1}^{k} \mathcal{H}_l$ durch

$$\langle \varphi_1 \otimes \ldots \otimes \varphi_k \,|\, \psi_1 \otimes \ldots \otimes \psi_k \rangle := \prod_{l=1}^{k} \langle \varphi_l \,|\, \psi_l \rangle \tag{7.16}$$

und lineare Fortsetzung eine Sesquilinearform $\langle \cdot \,|\, \cdot \rangle : \mathcal{H} \otimes \mathcal{H} \to \mathbb{C}$ definiert.

Ist $\mathcal{H}$ endlich-dimensional, dann ist diese auch schon ein Skalarprodukt auf $\mathcal{H}$.

Versieht man in Beispiel 7.4 die Vektorräume mit $\mathcal{H}_l$ dem kanonischen Skalarprodukt, dann ist das durch (7.16) definierte Skalarprodukt auf $\mathcal{H}$ dasjenige, bezüglich dessen die Basisvektoren e_i, $i \in I$, orthonormal sind.

Tensorprodukt von C^*–Algebren

Definition 7.7 Es seien $\mathcal{H}_1, \ldots, \mathcal{H}_k$ $\mathbb{C}$–Vektorräume und $a_l \in B(\mathcal{H}_l)$, $l = 1, \ldots, k$. Dann ist $a_1 \otimes \ldots \otimes a_k \in B(\mathcal{H})$, $\mathcal{H} := \bigotimes_{l=1}^{k} \mathcal{H}_l$ durch

$$a_1 \otimes \ldots \otimes a_k(v_1 \otimes \ldots \otimes v_k) := a_1(v_1) \otimes a_2(v_2) \otimes \ldots \otimes a_k(v_k)$$

gegeben.

Das Tensorprodukt $\mathcal{A} := \mathcal{A}_1 \otimes \ldots \otimes \mathcal{A}_k$ von C^*–Algebren $\mathcal{A}_l$ wird folgendermaßen definiert. Zunächst sind die Faktoren $\mathcal{A}_l$ $\mathbb{C}$–Vektorräume, sodass $\mathcal{A}$ mit Def. 7.2 ebenfalls ein $\mathbb{C}$–Vektorraum wird. Sind die $\mathcal{A}_l$ endlich-dimensional[15], dann sind sie

[15] Der Fall unendlicher Dimensionen ist subtiler.

nach (4.4) isomorph zu Matrixalgebren auf Hilberträumen $\mathbb{C}^{n_l}$, also $\mathcal{A}$ isomorph zu einer Matrixalgebra auf $\otimes_{l=1}^{k}\mathbb{C}^{n_l} \cong \mathbb{C}^n$. Die Operatornorm des Hilbertraumes $\mathbb{C}^n$ definiert eine Norm auf $\mathcal{A}$, mit der diese Algebra zu einer C^*–Algebra wird.

In der klassischen bzw. Quantenmechanik ist nun $\bigotimes_{l=1}^{k}\mathcal{A}_l$ die Observablenalgebra von k (unterscheidbaren) Teilchen, falls diesen die Observablenalgebren $\mathcal{A}_l$ zugeordnet sind.

Da in endlicher Dimension alle C^*–Algebra unital sind, ist jeder Faktor im Tensorprodukt $\mathcal{A}_1 \otimes \mathcal{A}_2$ als Unteralgebra enthalten.

Die Abbildung

$$\mathcal{A}_1 \to \mathcal{A}_1 \otimes \mathcal{A}_2 \quad , \quad a_1 \mapsto a_1 \otimes \mathbb{1} \qquad\qquad (a_1 \in \mathcal{A}_1)$$

ist nämlich nicht nur linear, sondern erhält auch das Operatorprodukt, die Adjungierte und die Norm.

Oft können wir nur an Teilsystemen Messungen vornehmen. ist

$$\Phi : \mathcal{A}_1 \otimes \mathcal{A}_2 \to \mathbb{C}$$

ein Zustand auf dem Tensorprodukt, dann wird durch

$$\Phi_1 : \mathcal{A}_1 \to \mathbb{C} \qquad\qquad \Phi_1(a) := \Phi(a \otimes \mathbb{1})$$

ein Zustand auf $\mathcal{A}_1$ induziert, analog auf $\mathcal{A}_2$.

Ist nun Φ durch $\Phi(a) = \mathrm{tr}(\rho a) \qquad\qquad (a \in \mathcal{A}_1 \otimes \mathcal{A}_2)$ mit Dichtematrix $\rho \in \mathcal{A}_1 \otimes \mathcal{A}_2$ gegeben, dann ist $\Phi(a_1 \otimes \mathbb{1}) = \mathrm{tr}(\rho \cdot a_1 \otimes \mathbb{1}) = \mathrm{tr}_{\mathcal{H}_1}(\rho_1 a_1)$ mit

$$\rho_1 = \sum_{i_1,i_2=1}^{\dim\mathcal{H}_1} \sum_{j=1}^{\dim\mathcal{H}_2} \langle i_1| \otimes \langle j|\, \rho\, |i_2\rangle \otimes |j\rangle \;\; |i_1\rangle \langle i_2|\,,$$

wie ein Koeffizientenvergleich ergibt[16]. Man nennt diese Abbildung *partielle Spurbildung*.

[16] **Notation.** In der Quantenmechanik wird oft die *Bra-ket*-Notation benutzt, in der

- Vektoren v des Hilbertraumes $\mathcal{H}$ in der Form $|v\rangle$ notiert werden (*ket*),
- Vektoren w des Dualraumes $\mathcal{H}^* \cong \mathcal{H}$ als $\langle w|$ geschrieben werden (*bra*),
- die Bezeichnung der Vektoren verkürzt wird z.B. $|i\rangle := e_i = (\,0 \cdots 0\; 1\; 0 \cdots 0\,)^t$ für den i–ten Vektor der kanonischen Basis des $\mathbb{C}^n$ oder $A\,|\lambda\rangle = \lambda\,|\lambda\rangle$ für einen (normierten) Eigenvektor $|\lambda\rangle$ mit Eigenwert λ.
- $\langle v\,|\,w\rangle := \langle v, w\rangle$ bezeichnet das Skalarprodukt.

Beispiel 7.8 Sind $\mathcal{A}_1$ und $\mathcal{A}_2$ abelsch, so auch $\mathcal{A}_1 \otimes \mathcal{A}_2$, und die Dichtematrix ρ ist diagonal:

$$\rho = \sum_{i=1}^{\dim \mathcal{A}_1} \sum_{j=1}^{\dim \mathcal{A}_2} \mu_{ij} P_i^{(1)} \otimes P_j^{(2)}$$

mit den eindimensionalen Projektionen $P_i^{(1)}$ auf $e_i^{(1)} \in \mathcal{H}_1$ und $P_j^{(2)}$ auf $e_j^{(2)} \in \mathcal{H}_2, \mu_{ij} \geq 0, \sum_{i,j} \mu_{ij} = 1$.

Damit ist $\rho_1 = \sum_{i=1}^{\dim \mathcal{A}_1} \mu_i^{(1)} P_i^{(1)}$ mit $\mu_i^{(1)} = \sum_j \mu_{ij}$ die Randverteilung.

Insbesondere gilt Satz 6.11.1: Sind $\mathcal{A}_1$ und $\mathcal{A}_2$ abelsche, unitale C^*–Algebren, dann gilt für $\Phi : \mathcal{A}_1 \otimes \mathcal{A}_2 \to \mathbb{C}$

$$\Phi \text{ rein } \iff \Phi_i : \mathcal{A}_i \to \mathbb{C} \text{ rein.}$$

Entsprechendes gilt nicht mehr für die Quantenmechanik.

Beispiel 7.9 Es sei $\mathcal{A}_1 = \mathcal{A}_2 = \mathcal{B}(\mathbb{C}^2)$ und die Dichtematrix auf $\mathcal{A}_1 \otimes \mathcal{A}_2$ die Projektion P_φ auf den von $\varphi := \frac{1}{\sqrt{2}}(e_1 \otimes e_1 + e_2 \otimes e_2)$ aufgespannten Unterraum. Dann ist

$$P_\varphi = \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}$$

und $\rho_1 = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, also Dichtematrix des Spurzustandes von $\mathcal{A}_1$.

Schema von Berechnungen

Diese Lage der Dinge hat für klassische und Quantencomputer die folgende Konsequenz. Da wir in beiden Fällen mikroskopisch d.h. reversibel rechnen wollen, ordnen wir einer zu berechnenden Funktion

$$g : B^m \to B^n \quad \text{mit} \quad B = \{0, 1\}$$

zunächst die reversible, d.h. bijektive Funktion

$$\tilde{g} : B^{m+n} \to B^{m+n} \quad , \quad \tilde{g}(x, y) := (x, g(x) + y)$$

zu. Die Addition in der abelschen Gruppe B^k erfolgt dabei bitweise, also

$$(x_1, \ldots, x_k) + (y_1, \ldots, y_k) = (x_1 + y_1 (\mathrm{mod}\, 2), \ldots, x_k + y_k (\mathrm{mod}\, 2)).$$

$\tilde{g}$ ist nicht nur bijektiv, sondern es ist sogar $\tilde{g}^{-1} = \tilde{g}$.

Da $\tilde{g}(x,0) = (x, g(x))$ ist, berechnet $\tilde{g}$ bei der Eingabe $(x,0)$ den Funktionswert $g(x)$. Im Gegensatz zu g ist in unserer Modellierung $\tilde{g}$ physikalisch berechenbar, und zwar als Automorphismus der abelschen (klassischen) C^*–Algebra $\mathcal{A}_{\mathrm{cl}}(B^{m+n}))$

Auf dem Hilbertraum $\mathbb{C}^2$ bezeichne $e_0 := \left(\begin{smallmatrix}1\\0\end{smallmatrix}\right)$, $e_1 := \left(\begin{smallmatrix}0\\1\end{smallmatrix}\right)$ die kanonische Basis, auf dem Tensorprodukt

$$\mathcal{H}^k := \bigotimes_{l=1}^{k} \mathbb{C}^2$$

benutzen wir die (lexikalisch, also entsprechend der Dualzahl $\sum_{i=1}^{k} x_i 2^{k-i}$ geordnete) Orthonormalbasis

$$e_x := e_{x_1} \otimes \ldots \otimes e_{x_k} \in \mathcal{H}^k \qquad (x = (x_1, \ldots, x_k) \in B^k).$$

Quantenmechanisch benutzen wir nun auf $\mathcal{H}^m \otimes \mathcal{H}^n = \mathcal{H}^{m+n}$ die lineare Abbildung

$$U_g \in \mathcal{B}(\mathcal{H}^{m+n}) \quad , \quad U_g(e_z) := e_{\tilde{g}(z)} \qquad (z \in B^{m+n}). \tag{7.17}$$

Da U_g die Basis von $\mathcal{H}^{m+n}$ permutiert, ist U_g auf diesen Hilbertraum ein (sehr spezieller) unitärer Operator.

Nun gibt es natürlich viele Operatoren auf $\mathcal{H}^k$, die unitär, aber keine Permutationsoperatoren sind, z.B. die *Hadamard-Transformation* $H^k := \bigotimes_{l=1}^{k} H$ mit

$$H : \mathbb{C}^2 \to \mathbb{C}^2 \quad , \quad H\begin{pmatrix} u \\ v \end{pmatrix} := \frac{1}{\sqrt{2}} \begin{pmatrix} u+v \\ u-v \end{pmatrix}. \tag{7.18}$$

Die darstellende Matrix dieser *diskreten Fouriertransformation* ist also das k–fache Tensorprodukt von $\frac{1}{\sqrt{2}} \left(\begin{smallmatrix}1 & 1\\ 1 & -1\end{smallmatrix}\right)$.

- In beiden Fällen beginnt die Berechnung mit *Eingabedaten*, die durch einen reinen Zustand auf der Observablenalgebra modelliert werden.
- Die *Berechnung* entspricht einem Automorphismus der C^*–Algebra (klassisch einer Permutation, quantenmechanisch einer Konjugation mit einer unitären Abbildung).

  Nach der Berechnung ist der Zustand gemäß Satz 6.11.1 immer noch rein.

- Nicht alle in der Rechnung vorkommenden Bits bzw. Qubits entsprechen Ausgabedaten, d.h. die C^*–Algebra ist um die während der Berechnung benutzten Variablen vergrößert. Beobachtet man am Ende nur die *Ausgabedaten*, dann ist das Ergebnis klassisch rein, also determiniert, quantenmechanisch aber i.A. probabilistisch. Das hat zur Folge, dass man den Quantencomputer eventuell zur Erhöhung der Wahrscheinlichkeiten mehrmals mit der gleichen Eingabe laufen lassen sollte.

8 Klassische und quantenmechanische Suchprobleme

In diesem Kapitel soll der quantenmechanische Grover-Algorithmus vorgestellt werden, mit dessen Hilfe in einer unstrukturierten Datenbank von n Einträgen ein Eintrag in etwa $\sqrt{n}$ Schritten gefunden wird. Klassisch muss man dagegen im Mittel etwa $n/2$ Einträge überprüfen, im schlechtesten Fall $n-1$. Man denke etwa an das Berliner Telefonbuch mit vielleicht $n = 10^6$ Einträgen. Die Suche nach einer Telefonnummer würde quantenmechanisch nur noch ca. 10^3 Anfragen benötigen.

8.1 Klassische Suchalgorithmen

Um das Problem in einen etwas weiteren Kontext zu stellen, werde ich einige Begriffe der klassischen Suchtheorie vorstellen. Dies führt uns zum Shannonschen Hauptsatz der Suchtheorie und zu Resultaten der klassischen Codierungstheorie.

Beispiel 8.1 Genau eine von $n \geq 3$ Münzen ist gefälscht, hat also ein kleineres Gewicht als das konstante Gewicht der echten Münzen. Wir wollen diese Münze durch Wägungen auf einer Balkenwaage finden. Legen wir $k \leq n/2$ Münzen auf die linke Schale und k auf die rechte, dann können drei Resultate eintreten: die linken Münzen sind leichter, die rechten Münzen sind leichter, oder beide sind gleich schwer. Entsprechend befindet sich die gefälschte Münze in der linken Schale, der rechten Schale, oder dem übrigen Haufen von $n-2k$ Münzen.

Ein Suchalgorithmus besteht z.B. darin, die erste Münze nacheinander mit der zweiten, der dritten usw. zu vergleichen.

Es ergibt sich der in Abbildung 9 dargestellte Suchbaum.

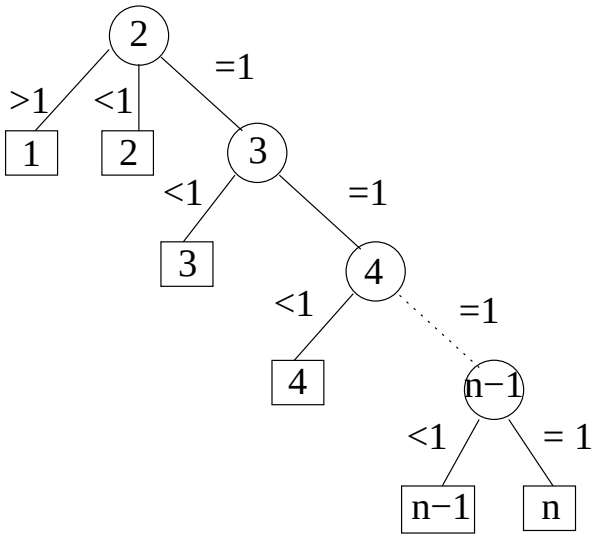


Abbildung 9: Suchbaum für den Vergleich der k–ten Münze mit der ersten Münze, $k = 2, \ldots, n-1$

Wir formalisieren, indem wir den Begriff des Suchalgorithmus als Wurzelbaum formalisieren (siehe Anhang A).

Definition 8.2 • Ein Wurzelbaum T heißt (n,q)–*Baum*, falls T n Blätter hat und alle innere Ecken höchstens q unmittelbare Nachfolger besitzen, *regulärer* (n,q)–Baum, falls sie genau q unmittelbare Nachfolger besitzen.
• $\mathcal{T}(n,q)$ bezeichnet die Menge der (n,q)–Bäume.

Beispiel 8.3 Der Suchbaum aus dem Beispiel 8.1 ist ein $(n,3)$–Baum der Länge $n-2$. In Abbildung 9 sind die Blätter quadratisch und die inneren Ecken rund gezeichnet.

Definition 8.4 • Ein *Suchbereich* $\mathcal{S}$ ist eine endliche, nicht leere Menge. Ein *Suchalgorithmus für* $\mathcal{S}$ ist ein Wurzelbaum (T,v^*) mit Blättern $\mathcal{S}$.
• Er heißt $(|\mathcal{S}|,q)$–Suchalgorithmus, wenn (T,v^*) ein $(|\mathcal{S}|,q)$–Baum ist.

Fragen:

1. Finde einen $(|\mathcal{S}|,q)$–Suchalgorithmus *minimaler Länge*

$$L := \min_{T\in\mathcal{T}(|\mathcal{S}|,q)} L(T) \quad \text{mit} \quad L(T) = \max_{s\in\mathcal{S}} l(s).$$

2. Ist $p : \mathcal{S} \to [0,1]$ eine Wahrscheinlichkeitsverteilung auf dem Suchbereich $\mathcal{S}$, (also $\sum_{s\in\mathcal{S}} p(s) = 1$), dann finde einen $(|\mathcal{S}|,q)$–Suchalgorithmus *minimaler mittlerer Länge*

$$\bar{L}(p) := \min_{T\in\mathcal{T}(|\mathcal{S}|,q)} \bar{L}(T,p)$$

mit der mittleren Blattlänge $\bar{L}(T,p)$ des Baumes T, also

$$\bar{L}(T,p) := \sum_{s\in\mathcal{S}} p(s)l(s).$$

Beispiel 8.5 Wir haben das Gefühl, dass der in Beispiel 8.1 angesprochende $(n,3)$–Suchalgorithmus nicht sehr effizient ist. Wir haben die Balkenwaage nicht gut genutzt, denn die meisten inneren Ecken des Suchbaums aus Abbildung 9 besitzen nur zwei statt drei unmittelbare Nachfolger. Ist z.B. $n = 3^m$, dann können wir die Münzen $1,\ldots,3^{m-1}$ auf die linke und die Münzen $3^{m-1}+1,\ldots,2\cdot 3^{m-1}$ auf die rechte Schale legen. Eine Messung entscheidet dann, in welcher der drei Teilmengen der Größe 3^{m-1} die gefälschte Münze liegt, und es reichen m statt 3^m-2 Wägungen.

Satz 8.6 (Kraftsche Ungleichung)

1. *Es seien $l_1, \ldots, l_n$ die Längen der Blätter des Baumes $T \in \mathcal{T}(n, q)$. Dann gilt $\sum_{i=1}^{n} q^{-l_i} \leq 1$. Gleichheit gilt genau dann, wenn T regulär ist.*

2. *Es seien $l_1, \ldots, l_n \in \mathbb{N}_0$ gegeben, und $\sum_{i=1}^{n} q^{-l_i} \leq 1$. Dann gibt es einen Baum $T \in \mathcal{T}(n, q)$ mit den Längen $l_1, \ldots, l_n$.*

Bew.: 1) Ein nicht regulärer (n, q)–Baum lässt sich durch das Hinzufügen von Blättern zu einem regulären (n', q)–Baum ergänzen. Dabei vergrößert sich die Summe, sodass wir die erste Behauptung nur für reguläre (n, q)–Bäume beweisen müssen. Diese lassen sich durch Anwendung der in Abb. 10 dargestellten Substitution verkürzen, wobei der entstehende Baum immer noch ein regulärer

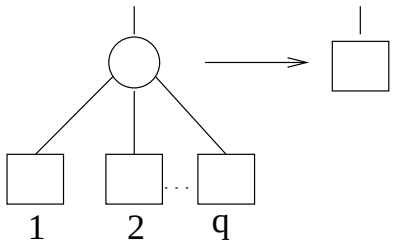


Abbildung 10: Substitution bei regulären (n, q)–Bäumen

$(\tilde{n}, q)$–Baum (mit $\tilde{n} = n - q + 1$) ist, und die Summe $\sum_{i=1}^{n} q^{-l_i}$ sich nicht ändert.
Für das Endresultat, den regulären (n, q)–Baum der Länge 1 ist $n = q$, und

$$\sum_{i=1}^{n} q^{-l_i} = \sum_{i=1}^{q} q^{-1} = 1.$$

2) Es sei $w_k := |\{i \mid l_i = k\}|$ die Zahl der Blätter der Länge k des gesuchten Baumes, mit $k = 0, \ldots, L := \max_i(l_i)$. Es gilt also

$$\sum_{k=0}^{L} w_k q^{-k} \leq 1$$

oder

$$\begin{aligned} w_{k+1} &\leq q(q^k - \sum_{i=0}^{k} w_i q^{k-i} - \sum_{i=k+2}^{L} w_i q^{k-i}) \\ &\leq q(q^k - \sum_{i=0}^{k} w_i q^{k-i}) \in \mathbb{N}_0. \end{aligned}$$

Diesem Bedarf an Blättern der Länge $k+1$ steht ein Angebot von $q^k - \sum_{i=0}^{k} w_i q^{k-i}$ inneren Ecken der Länge k gegenüber. Er kann also gedeckt werden, denn an jede dieser inneren Ecken können q Blätter angehängt werden. □

Corollar 8.7 *Für $q \geq 2$ und $n \in \mathbb{N}$ ist die minimale Länge eines (n,q)–Suchalgorithmus gleich $L = \lceil \log_q(n) \rceil$.*

Bew.:

- Nach Teil (2) der Kraftschen Ungleichung gibt es einen (n,q)–Baum mit den Längen $l_1 = \ldots = l_n = \lceil \log_q(n) \rceil$, denn $\sum_{i=1}^{n} q^{-l_i} = nq^{-\lceil \log_q(n) \rceil} \leq nq^{-\log_q(n)} = \frac{n}{n} = 1$.
- Nach Teil (1) der Kraftschen Ungleichung gibt es aber keinen kürzeren Baum, denn für alle $x \in \mathbb{R}$ gilt $\lceil x \rceil - 1 < x$. □

Beispiel 8.8 Im Fall der gefälschten Münze benötigen wir also höchstens m Wägungen, falls $3^{m-1} < n \leq 3^m$ gilt.

Wenn wir die minimale mittlere Länge $\bar{L}(p)$ bez. einer Wahrscheinlichkeitsverteilung p auf dem Suchbereich $\mathcal{S} = \{1, \ldots, n\}$ abschätzen wollen, werden wir in natürlicher Art und Weise auf entropische Abschätzungen geführt.

8.2 Relative Entropie

Definition 8.9 Es seien ω und φ Zustände auf einer endlichen C^*–Algebra $\mathcal{A}$, und $\rho_\omega, \rho_\varphi \in \mathcal{A}$ seien ihre Dichtematrizen.

Dann heißt

$$S(\omega, \varphi) := \begin{cases} \operatorname{tr}(\rho_\omega(\ln \rho_\omega - \ln \rho_\varphi)) & , \text{Kern } \rho_\omega \supseteq \text{Kern } \rho_\varphi \\ \infty & , \text{sonst} \end{cases}$$

die *relative Entropie von ω und φ*.

$S(\omega, \varphi)$ wird auch *Kullback-Leibler*-Abstand genannt. Zwar ist S nicht symmetrisch in den beiden Argumenten. Trotzdem misst die relative Entropie in folgendem Sinn den Abstand von ω und φ:

Satz 8.10

$$S(\omega, \varphi) \geq \tfrac{1}{2}\mathrm{tr}((\rho_\omega - \rho_\varphi)^2).$$

Bew.: Wir können Kern $\rho_\omega \supseteq$ Kern ρ_φ annehmen und benutzen die Spektraldarstellungen

$$\rho_\varphi = \sum_{\lambda \in \sigma(\rho_\varphi)} \lambda P_\lambda \quad \text{und} \quad \rho_\omega = \sum_{\mu \in \sigma(\rho_\omega)} \mu Q_\mu.$$

Da ρ_φ eine Dichtematrix ist, gilt $\sigma(\rho_\varphi) \subset [0, 1]$, und wir brauchen über einen etwaigen Eigenwert $\lambda = 0$ nicht zu summieren. Analoges gilt für ρ_ω.

Für ein ξ zwischen λ und μ gilt nach dem Mittelwertsatz

$$-\eta(\lambda) + \eta(\mu) + (\lambda - \mu)\eta'(\mu) = -\tfrac{1}{2}(\lambda - \mu)^2\eta''(\xi) \geq \tfrac{1}{2}(\lambda - \mu)^2,$$

wobei die Ungleichung wegen $-\frac{d^2}{d\xi^2}\eta(\xi) = \frac{d^2}{d\xi^2}\xi \ln \xi = \frac{1}{\xi} \geq 1$ gilt (siehe Abbildung 11).

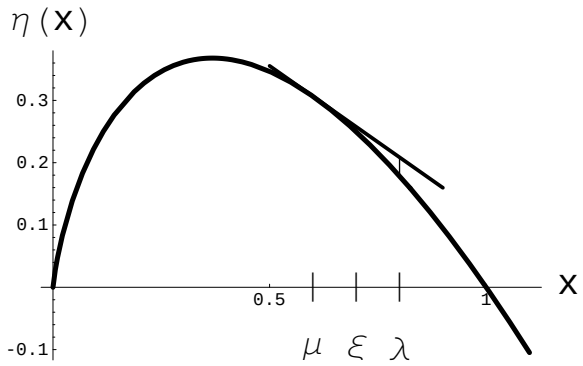


Abbildung 11: Ungleichung für die relative Entropie

Damit ist (unter Benutzung von $\mathrm{tr}[(\rho_\omega - \rho_\varphi)\mathbb{1}] = 1 - 1 = 0$)

$$
\begin{aligned}
& S(\omega,\varphi) - \tfrac{1}{2}\mathrm{tr}((\rho_\omega - \rho_\varphi)^2) \\
&= \mathrm{tr}\left(-\eta(\rho_\omega) + \eta(\rho_\varphi) + (\rho_\omega - \rho_\varphi)(-\mathbb{1} - \ln\rho_\varphi) - \tfrac{1}{2}(\rho_\omega - \rho_\varphi)^2\right) \\
&= \sum_{\lambda\in\sigma(\rho_\omega)} \sum_{\mu\in\sigma(\rho_\varphi)} \left[-\eta(\lambda) + \eta(\mu) + (\lambda-\mu) \underbrace{(-1 - \ln\mu)}_{\eta'(\mu)} - \tfrac{1}{2}(\lambda-\mu)^2 \right] \mathrm{tr}(P_\lambda Q_\mu) \\
&\geq 0,
\end{aligned}
$$

da $\mathrm{tr}(P_\lambda Q_\mu) = \mathrm{tr}((P_\lambda Q_\mu)^*(P_\lambda Q_\mu)) \geq 0$. □

Insbesondere ist die relative Entropie nicht negativ und Null nur, falls die beiden Zustände gleich sind.

Satz 8.11 *Für $q \geq 2$ und $n \in \mathbb{N}$ gilt für die minimale mittlere Länge $\bar{L}(p)$ eines (n,q)–Suchalgorithmus zur Wahrscheinlichkeitsverteilung p auf $\mathcal{S}$ die Abschätzung*

$$\frac{S(p)}{\ln q} \leq \bar{L}(p) \leq \frac{S(p)}{\ln q} + 1. \tag{8.19}$$

Bew.: Blätter $i \in \mathcal{S}$ mit Wahrscheinlichkeit $p_i = 0$ haben wir nicht zugelassen.
1) Es sei T ein (n,q)–Baum mit den Längen $l_1, \ldots, l_n$, und $r_i := q^{-l_i}$.

Ist T regulär, dann ist $\sum_{i=1}^n r_i = 1$, also $r := (r_1, \ldots, r_n)$ ein Wahrscheinlichkeitsmaß, und nach Satz 8.10 gilt $S(p,r) \geq 0$.
Ist T nicht regulär, gilt trotzdem die analoge Ungleichung

$$\sum_{i=1}^n p_i(\ln p_i - \ln r_i) \geq 0, \tag{8.20}$$

denn wegen $\ln(1+x) \leq x$ ist $\sum_{i=1}^n p_i(-\ln(\frac{r_i}{p_i})) \geq \sum_{i=1}^n -p_i \cdot (\frac{r_i}{p_i} - 1) = 1 - \sum_{i=1}^n r_i > 0$.

Mit der Definition von r_i und $S(p) = -\sum_{i=1}^n p_i \ln p_i$ lässt sich (8.20) in der Form $\sum_{i=1}^n p_i\, l_i \ln q \geq S(p)$ schreiben.

Da $\bar{L}(T,p) = \sum_{i=1}^n p_i\, l_i$, folgt daraus die erste Ungleichung in (8.19).
2) Zum Beweis der zweiten Ungleichung in (8.19) setzen wir $l_i := \lceil -\log_q p_i \rceil$, $i = 1, \ldots, n$, sodass $q^{-l_i} \leq p_i$ und $\sum_{i=1}^n q^{-l_i} \leq \sum_{i=1}^n p_i = 1$.

Die Kraftsche Ungleichung ist also erfüllt, und wir können einen (n, q)–Baum T mit den Längen $l_1, \ldots, l_n$ finden. Dessen mittlere Länge ist

$$\begin{aligned} L(T,p) &= \sum_{i=1}^n p_i\ l_i = \sum_i p_i \lceil -\log_q p_i \rceil < \sum_i p_i \cdot (1 - \log_q p_i) \\ &= 1 - \tfrac{1}{\ln q} \sum_{i=1}^n p_i \ln p_i = 1 + \tfrac{S(p)}{\ln q}. \end{aligned}$$ □

Eine kleine Zusatzüberlegung wird uns später dazu führen, dass diese Abschätzungen auch die Kompression festlegen, die wir für eine Übertragung von Zeichen aus $\mathcal{S}$ mit Wahrscheinlichkeiten p_i über einen Kanal finden können.

Oft kann man in der Suchtheorie nicht beliebige Tests zulassen:

Definition 8.12 Eine Familie $\mathcal{F}$ von Funktionen $f : \mathcal{S} \to \mathbb{N}_0$ heißt *Testfamilie*, und das Paar $(\mathcal{S}, \mathcal{F})$ *Suchprozess*.

Beispiel 8.13 Unstrukturierte Datenbank mit N Einträgen, d.h. $\mathcal{S} = \{x_1, \ldots, x_N\}$,

$$\mathcal{F}_0 = \{f_1, \ldots, f_N\} \quad \text{mit} \quad f_i := 1_{x_i} \quad , \quad 1_{x_i}(x) = \begin{cases} 1 & , \ x = x_i \\ 0 & , \ \text{sonst.} \end{cases}$$

Ein Algorithmus für diesen Suchprozess besteht in der in Abbildung 12 dargestellten elementweisen Suche.

Er besitzt die Länge $L(\mathcal{S}, \mathcal{F}_0) = N - 1$. Bei einer Wahrscheinlichkeitsverteilung p auf $\mathcal{S}$ fragen wir am besten in der Reihenfolge absteigender Wahrscheinlichkeit und erhalten nach Umnummerierung

$$\bar{L}(\mathcal{S}, \mathcal{F}_0, p) = \sum_{i=1}^{N-1} (i p_i) + (N-1) p_N,$$

also für $p_1 = \ldots = p_N = \frac{1}{N}$ $\qquad \bar{L}(\mathcal{S}, \mathcal{F}_0, p) = \frac{N+1}{2} - \frac{1}{N}$.

8.3 Quantenmechanische Suche: Der Grover-Algorithmus

Wir wollen nun den *Grover-Algorithmus* zur quantenmechanischen Suche in einer unstrukturierten Datenbank besprechen. Der Einfachheit halber sei ihre Größe $m := 2^N$.

Klassisch würde man die Elemente $x \in B^m$ des Suchbereiches mit dem Suchprozess aus Beispiel 8.13 durchtesten, d.h. $1_y(x)$ berechnen, bis $1_y(y) = 1$ auftritt, also y gefunden ist.

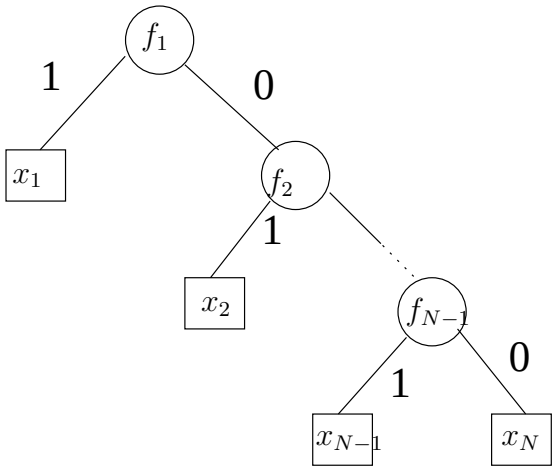


Abbildung 12: Klassische Suche in einer umstrukturierten Datenbank

In unserer reversiblen Formulierung der klassischen Berechnung würde man die Funktion

$$\widetilde{1}_y : B^{m+1} \to B^{m+1} \quad , \quad \widetilde{1}_y(x,i) = (x, i + \delta_{x,y})$$

berechnen.

Quantenmechanisch wenden wir stattdessen auf den *Startvektor*

$$s := (H^m e_0) \otimes (He_1) = \frac{1}{\sqrt{2^{m+1}}} \sum_{x \in B^{m+1}} (-1)^{x_{m+1}} e_x$$

die in (7.17) eingeführte unitäre Transformation U_{1_y} an, wobei H die in (7.18) eingeführte Hadamard-Transformation ist. Unser Ziel ist es, das gesuchte $y \in B^m$ zu finden, d.h. die Funktion 1_y zu berechnen.

Lemma 8.14 *Die Spiegelung an der zu einem Vektor $v \in \mathcal{H}^{m+1} \setminus \{0\}$ senkrechten Ebene werde durch*

$$S_v := \mathbb{1} - 2P_v : \mathcal{H}^{m+1} \to \mathcal{H}^{m+1}$$

bezeichnet. Dann gilt

$$U_{1_y} = S_{e_y \otimes He_1}.$$

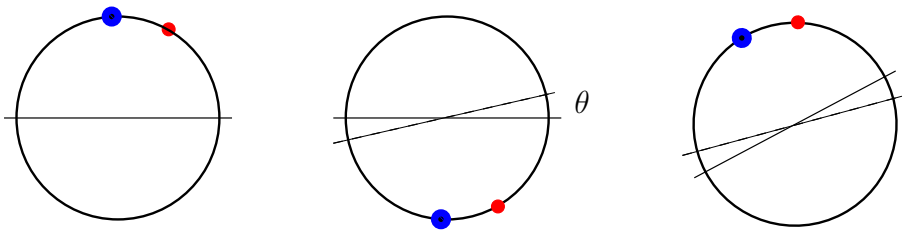


Abbildung 13: Links: Spiegelung an der horizontalen Geraden. Mitte: Spiegelung an einer um θ gedrehten Geraden. Rechts: Drehung um den Winkel 2θ.

Bew.: Allgemein ist für einen Vektor $a = \sum_{x\in B^m} a_x e_x \in \mathcal{H}^m$ und $i = 0, 1$

$$\begin{aligned} & U_{1_y}(a \otimes (He_i)) \\ = & \sum_{x\in B^m} a_x U_{1_y}(e_x \otimes (He_i)) = 2^{-1/2} \sum_{x\in B^m} a_x \left(e_{\widetilde{1_y}(x,0)} + (-1)^i e_{\widetilde{1_y}(x,1)} \right) \\ = & 2^{-1/2} \left(a_y (e_{(y,1)} + (-1)^i e_{(y,0)}) + \sum_{x\in B^m\setminus\{y\}} a_x (e_{(x,0)} + (-1)^i e_{(x,1)}) \right) \\ = & S_{e_y\otimes He_1}(a \otimes (He_i)). \end{aligned}$$ □

Als Nächstes führen wir die unitäre Transformation S_s auf $\mathcal{H}^{m+1}$ durch, spiegeln also am Startvektor.

Lemma 8.15 *Das Produkt $S_v S_w$ zweier Spiegelungen an Einheitsvektoren v, w ist eine Drehung um den Winkel $2\arccos(\langle v, w\rangle)$ in der von v und w aufgespannten Ebene.*

Bew.: Als Produkt zweier (komplexifizierter) Spiegelungen ist $S_v S_w$ eine (komplexifizierte) Drehung. Die Drehung muss in $\mathrm{span}(v, w)$ stattfinden, denn im zu v und w senkrechten Unterraum wirken S_v und S_w trivial. Den Drehwinkel liest man z.B. aus der Identität

$$P_v S_v S_w P_v = P_v(\mathbb{1} - 2P_v)(\mathbb{1} - 2P_w)P_v = -P_v(\mathbb{1} - 2P_w)P_v - P_v(\mathbb{1} - 2P_w)P_v$$

ab, denn $1 - 2(\cos\varphi)^2 = \cos(2\varphi)$, siehe Abbildung 13. □

In unserem Fall ist $v = s = (H^m e_0) \otimes (He_1)$ und $w = e_y \otimes He_1$, also

$$\langle v, w \rangle = \langle H^m e_0, e_y \rangle \cdot \langle He_1, He_1 \rangle = \langle H^m e_0, e_y \rangle = 2^{-m/2},$$

die beiden Vektoren stehen also nahezu senkrecht aufeinander. Entsprechend ist der Drehwinkel von $S_v S_w$ etwa π. Daher wird die *Grover-Transformation* durch

$$G_y := -S_v S_w$$

definiert. In der komplexen Ebene $\mathrm{span}(v, w)$ wirkt diese dann als Drehung um den (kleinen) Winkel

$$\theta := \pi - 2\arccos(\langle v, w \rangle) = 2\arcsin(\langle v, w \rangle) = 2 \cdot 2^{-m/2}(1 + \mathcal{O}(2^{-m})).$$

Wir drehen in dieser Ebene $k := \left\lfloor \frac{\pi/2}{\theta} \right\rfloor = 2^{(m-2)/2}(1 + \mathcal{O}(2^{-m}))$ mal mit G_y, entsprechend einem Gesamtwinkel von $k\,\theta = \pi/2 + \mathcal{O}(2^{-m})$. Eine anschliessende Messung des reinen Zustandes $(G_y)^k s \approx e_y \otimes He_1$ in der kanonischen Basis ergibt mit Wahrscheinlichkeit $1 - \mathcal{O}(2^{-m})$ den richtigen Wert y.

Wichtig ist zu bemerken, dass k unabhängig vom unbekannten Wert y ist, und dass k von der Größenordnung $\sqrt{N}$ mit der Größe $N = 2^m$ des Suchbereiches ist.

Weiter läßt sich zeigen, dass die verwandten Operationen lokal sind, d.h. auf einer m-unabhängigen Zahl von Qubits wirken.

9 Klassische Informationstheorie

Wir haben gesehen, dass die Entropie $S(p)$ der Wahrscheinlichkeitsverteilung p auf dem Suchbereich $\mathcal{S}$ im wesentlichen die minimale mittlere Länge $\bar{L}(p)$ der Suche bestimmt.

Auch in der Nachrichtenübertragung spielt die Entropie eine dominante Rolle. Die durch ein 1948 von C. Shannon initiierte Informationstheorie hat die Nachrichtenübertragung zum Inhalt. Da Kanäle wie z.B. Telefonleitungen nur eine begrenzte Übertragungskapazität haben und die Nachrichten mit einer positiven Fehlerrate übertragen werden, müssen in der Informationstheorie zwei konfligierende Ziele erreicht werden:

- Die zu übertragende Nachricht muss möglichst stark *komprimiert* werden (*Quellencodierung*).

- Sie muss so codiert werden, dass *Übertragungsfehler erkannt und beseitigt* werden (*Kanalcodierung*).

- Eine zusätzliche Frage ist die nach der *Verschlüsselung* der Nachricht zu Geheimhaltungszwecken, also der *Kryptographie*.

Die Informationstheorie beschäftigt sich also nicht mit dem *Inhalt* von Nachrichten, sondern ihrer günstigen *Übertragung* und Verarbeitung.

Alle drei Problemstellungen haben ihre Pendants in der Quanteninformationstheorie. Besonders wichtig für die Frage, ob Quantencomputer technisch realisierbar sind, ist aber die quantenmechanische Verallgemeinerung der *Kanalcodierung*, denn hier geht es um die Kontrolle der allen Bauelementen innewohnenden Fehler.

Das Schema des Nachrichtenflusses ist in Abb. 14 skizziert.

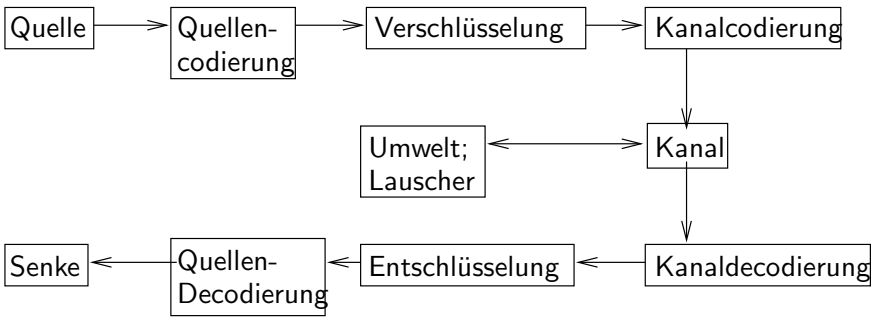


Abbildung 14: Nachrichtenübertragung vom Sender (Quelle) zum Empfänger (Sender)

In der Informationstheorie wird dieses Schema zunächst präzisiert und die statistischen Eigenschaften von Quelle, Kanal und Senke werden modelliert.

Danach wird die bestmögliche Effizienz der verschiedenen (De-) Codierungsschritte untersucht und es werden günstige Algorithmen entwickelt.

9.1 Entropie einer Quelle

Da die Sprache der (klassischen) Informationstheorie die Wahrscheinlichkeitstheorie ist, erinnere ich zunächst an deren Grundbegriffe.

Definition 9.1 Ein System $\mathcal{F}$ von *Teilmengen* einer Menge Ω heißt σ*–Algebra*, wenn

1. $\Omega \in \mathcal{F}$
2. $A \in \mathcal{F} \Longrightarrow \Omega - A \in \mathcal{F}$
3. $A_n \in \mathcal{F}\ (n \in \mathbb{N}) \Longrightarrow \bigcup_{n\in\mathbb{N}} A_n \in \mathcal{F}$.

Beispiel 9.2 1. $\mathcal{F}_1 := 2^\Omega := \{A \subseteq \Omega\}$ (Potenzmenge von Ω).

2. $\mathcal{F}_2 := \{\Omega, \emptyset\}$

Alle σ– von Ω sind Teilmengen von $\mathcal{F}_1$ und enthalten $\mathcal{F}_2$. Für die meisten Anwendungen ist $\mathcal{F}_2$ zu grob und $\mathcal{F}_1$ ist oft zu fein (enthält zu viele Elemente).

Definition 9.3 Ein Tripel $(\Omega, \mathcal{F}, P)$ mit

- Ω Menge von sog. *Elementarereignissen* $\omega \in \Omega$
- $\mathcal{F}$ σ–Algebra von Ω von sog. *Ereignissen* $A \in \mathcal{F}$
- $P : \mathcal{F} \to [0,1]$ mit $P(\Omega) = 1$ und $P(\bigcup_{n\in\mathbb{N}} A_n) = \sum_{n=1}^{\infty} P(A_n)$ für disjunkte $(A_k \cap A_l = \emptyset$ für $k \neq l)$ Ereignisse $A_n \in \mathcal{F}$, genannt *Wahrscheinlichkeitsmaß*,

heißt *Wahrscheinlichkeitsraum*.

Beispiel 9.4 $\Omega := \Omega\{-, a, b, c, \ldots, z\}$ Alphabet ($-$ entspricht Zwischenraum) $\mathcal{F} = 2^\Omega$, $P(\{\omega\}) := p_\omega\ (\omega \in \Omega)$ mit
$p_- = 0.1515$, $p_l = 0.1470$, $p_n = 0.0884$, $p_r = 0.0638, \ldots, p_x = 0.001$

Dies sind die Wahrscheinlichkeiten für das Auftreten von Buchstaben in der deutschen Sprache [BG].

Quellen produzieren mit vorgegebenen Wahrscheinlichkeiten Buchstabenfolgen über einem Alphabet $A = \{a_1, \ldots, a_n\}$.

Eine solche (zweifach unendliche) Buchstabenfolge ist ein Element von

$$\Omega := \{\omega : \mathbb{Z} \to A\} = \prod_{k\in\mathbb{Z}} A_k \quad \text{mit} \quad A_k := A;$$

der k–te Buchstabe der Folge ω ist $\omega(k) \equiv \omega_k$.

Um aus Ω einen Wahrscheinlichkeitsraum zu machen, betrachten wir zunächst für endliche $\{k_1, \ldots, k_n\} = U \subseteq \mathbb{Z}$ und $B_i \subseteq A, i = 1, \ldots, n$ die *Zylindermengen*

$$:= \{\omega \in \Omega \mid \omega_{k_1} \in B_{k_1}, \ldots, \omega_{k_n} \in B_{k_n}\},$$

bei denen die k_i–ten Buchstaben aus B_i stammen.

Es sei $\mathcal{F}_A \subseteq 2^\Omega$ die kleinste σ–Algebra, die alle diese Zylindermengen enthält.

Nach einem Satz von Kolmogorov können wir nun eindeutig ein Wahrscheinlichkeitsmaß

$$P : \mathcal{F}_A \to [0, 1]$$

fixieren, indem wir die Wahrscheinlichkeiten

$$P(Z_n(B_{k_1}, \ldots, B_{k_n}))$$

konsistent als Zahlen in $[0, 1]$ definieren. Dabei bedeutet Konsistenz, dass für je zwei disjunkte Zylindermengen $Z_1, Z_2 \in \mathcal{F}_A$

$$P(Z_1 \cup Z_2) = P(Z_1) + P(Z_2)$$

ist.

Definition 9.5

- Die *Shift*abbildung $T : \Omega \to \Omega$ ist durch

  $$(T\omega)(k) := \omega(k + 1) \qquad (k \in \mathbb{Z})$$

  gegeben.

- Ist P *invariant* unter T, d.h. $P(T(E)) = P(E)$ für $E \in \mathcal{F}_A$, dann heißt

  $$(\Omega, \mathcal{F}_A, P)$$

  (stationäre) Quelle.

- Sie heißt *ergodisch*, wenn für ein invariantes Ereignis

  $$E = T(E) \in \mathcal{F}_A \qquad P(E) = 0 \quad \text{oder} \quad P(E) = 1$$

  gilt.

- Sie heißt Quelle *ohne Gedächtnis*, wenn

$$P(Z_n(B_{k_1}, \ldots, B_{k_n})) = \prod_{l=1}^{n} \tilde{p}(B_{k_i}) \tag{9.21}$$

für ein Wahrscheinlichkeitsmaß $\tilde{p}$ auf A gilt.

Beispiel 9.6 1. die Wahrscheinlichkeitsverteilung $\tilde{p}$ auf dem Alphabet $A := \{-, a, b, c, \ldots, z\}$ aus Beispiel 9.4 definiert via 9.21 eine Quelle ohne Gedächtnis.

Eine typische Folge $\omega \in \Omega$ (im Sinne typischer Entropie) für diesen Wahrscheinlichkeitsraum ist

`...eme-gkheet-ers-titbl...`

2. Ähnlich wie die Häufigkeit der einzelnen Buchstaben kann man auch die Häufigkeit von Gruppen von n Buchstaben durch statistische Auswertung von Texten ermitteln. Beispielsweise wird die Wahrscheinlichkeit von $p_{\omega_1\omega_2}$ Bigrammen $(\omega_1, \omega_2) \in A \times A$ angeführt durch[17]

$$\begin{aligned} p_{en} &= 0.0447, \\ p_{er} &= 0.0340, \\ p_{ch} &= 0.0280. \end{aligned}$$

Dies führt dann über die Definition der Wahrscheinlichkeit von Zylindermengen zu Quellen mit Gedächtnis.

Legt man die Häufigkeit von Vierergruppen in der deutschen Sprache zugrunde, ergibt sich als eine typische Folge $\omega \in \Omega$ z.B.

`...ich-folgemaeszig-bis-stehen-disponin-seele-namen...`

3. Ähnlich kann man Wörter und deren Folgehäufigkeiten zugrunde legen (siehe z.B. Cover, Thomas [CT], Bsp. 7, p. 135).

Definition 9.7 Die *Entropie* der Quelle $(\Omega, \mathcal{F}_A, P)$ ist durch $H := \lim_{n\to\infty} \frac{H(\mathcal{P}n)}{n}$ gegeben, wobei $\mathcal{P} := \{Z_{\{1,\ldots,n\}}(\{a_1\}, \ldots, \{a_n\} \mid a_i \in A\}$ eine Partition von Ω in disjunkte Zylindermengen ist und

$$H(\mathcal{P}_n) := \sum_{a_1,\ldots,a_n \in A} \eta\left(P(Z_{\{1,\ldots,n\}}(\{a_1\}, \ldots, \{a_n\}))\right).$$

[17] Ohne Berücksichtigung des Leerzeichens

Beispiel 9.8 1. Die in Bit (also zur Basis 2, statt zur Basis e) berechnete Entropie der gedächtnislosen Quelle ist $H \approx 4.11$ (Bit pro Zeichen), also deutlich geringer als $\log_2 27 \approx 4.75$ für die Gleichverteilung.

2. Für die deutsche Sprache rechnet man mit einer Entropie $H \approx 1.3$. Schon bei Fixierung von Tripelhäufigkeiten ergibt sich $H \approx 1.6$.

   Dass die Entropie einer Quelle wirklich definiert ist, also dass der Limes in der Definition von H existiert, folgt aus der *Subadditivität der Entropie*, die auch quantenmechanisch gilt:

Satz 9.9 *Es sei ρ eine Dichtematrix auf einem Tensorprodukt $\mathcal{A}_1 \otimes \mathcal{A}_2$ endlichdimensionaler C^*–Algebren. Dann gilt für die partiellen Spuren $\rho_1 := \mathrm{tr}_2(\rho)$, $\rho_2 := \mathrm{tr}_1(\rho)$*

$$S(\rho) \leq S(\rho_1) + S(\rho_2).$$

Gleichheit gilt nur für den Produktzustand $\rho = \rho_1 \otimes \rho_2$.

Bew.: Wegen der in Satz 8.10 gezeigten Positivität der relativen Entropie

$$S(\rho, \tilde{\rho}) = \mathrm{tr}(\rho(\ln \rho - \ln \tilde{\rho})) \geq \tfrac{1}{2}\mathrm{tr}\left((\rho - \tilde{\rho})^2\right)$$

mit Dichtematrix $\tilde{\rho} := \rho_1 \otimes \rho_2$ auf $\mathcal{A}_1 \otimes \mathcal{A}_2$,

$$\ln(\rho_1 \otimes \rho_2) = \ln(\rho_1 \otimes \mathbb{1}) + \ln(\mathbb{1} \otimes \rho_2)$$

und den Identitäten

$$\begin{aligned} S(\rho_1) &= -\mathrm{tr}_1(\rho_1 \ln \rho_1) = -\mathrm{tr}(\rho \ln(\rho_1 \otimes \mathbb{1})), \\ S(\rho_2) &= -\mathrm{tr}_2(\rho_2 \ln \rho_2) = -\mathrm{tr}(\rho \ln(\mathbb{1} \otimes \rho_2)) \end{aligned}$$

folgt die Behauptung. □

Bemerkung 9.10 Wie wir in Satz 9.20 sehen werden, gilt *klassisch* (d.h. im abelschen Fall) $S(\rho) \geq S(\rho_i) \quad i = 1, 2$. Ein quantenmechanisches Gegenbeispiel ist der in Bsp. 7.9 beschriebene reine Zustand (Entropie 0) auf $\mathcal{B}(\mathbb{C}^2) \otimes \mathcal{B}(\mathbb{C}^2)$, dessen partielle Spuren zu dem Spurzustand auf $\mathcal{B}(\mathbb{C}^2)$ (Entropie $\ln 2$) führen.

Satz 9.11 *Die Entropie einer Quelle ist wohldefiniert und*

$$H \leq \frac{H(P_n)}{n}.$$

Bew.: Wir setzen $H_n := H(\Psi_n)$, und folgern aus Subadditivität und Stationarität

$$H_{kn} \leq k\ H_n,$$

also insbesondere

$$\frac{H_k}{k} \leq H_1 < \infty.$$

Sei $h := \liminf_k \frac{H_k}{k}$, $\varepsilon > 0$ und $l \in \mathbb{N}$ so, dass $\frac{H_l}{l} \leq h + \varepsilon$. Dann gilt für $m > l$, $(k-1)l < m \leq kl$

$$\frac{H_m}{m} \leq \frac{H_{kl}}{(k-1)l} \leq \frac{k}{k-1}\frac{H_l}{l} \leq \frac{k}{k-1}(k+\varepsilon),$$

also für große m $\frac{H_m}{m} \leq k + 2\varepsilon$. □

9.2 Quellencodierung

Oft wird die Quelle über ein anderes Alphabet A definiert sein als das Alphabet B, das man über den technischen Kanal schicken kann.

Eine Methode der Quellencodierung besteht darin, die Buchstabenfolge $\omega \in \Omega = A^{\mathbb{Z}}$ zunächst in Teilfolgen

$$(\omega_{kr}, \ldots, \omega_{kr+r-1}) \in A^r \qquad (k \in \mathbb{Z})$$

der konstanten Länge $r \in \mathbb{N}$ zu verhackstücken und diese Elemente von A^r dann zu codieren und hintereinanderzuschreiben.

Definition 9.12

- Ein *Code* C (über dem Alphabet B) ist eine endliche Teilmenge von $B^* = \bigcup_l B^l$. Die Kardinalität $|C|$ von C heißt *Ordnung von* C.
- Die *Länge* eines *Codewortes* $a = (a_1, \ldots, a_n) \equiv a_1 \ldots a_n \in C$ ist $|a| = n$.
- Besitzen alle Codewörter die gleiche Länge n, dann heißt C *Blockcode* (der Länge n).
- Ein Code C heißt *Präfixcode*, wenn kein Paar von Worten $a_1 \ldots a_m \in C$, $a_1 \ldots a_n \in C$ mit $m < n$ existiert.

- Eine Abbildung $\varphi : A^r \to B^*$ heißt *Codierung von A^r in B*.

Vernünftige Codierungen φ sollten injektiv sein, sodass eine *Decodierung*

$$\psi : C := \varphi(A^r) \to A^r \quad \text{mit} \quad \psi \circ \varphi = \mathrm{id}$$

existiert; ich werde die Injektivität im Weiteren voraussetzen. Weiter sei $|B| \geq 2$.

Dies ist aber nur eine notwendige, keine hinreichende Bedingung für die Decodierung der Codierung von $\omega \in \Omega$ mittels φ, bei der man die Codewörter

$$b^{(k)} := \varphi(\omega_{kr} \dots \omega_{kr+r-1}) \in C \qquad (k \in \mathbb{Z}) \tag{9.22}$$

in der Form $\dots b^{(-1)} b^{(0)} b^{(1)} b^{(2)} \dots$ hintereinanderschreibt.

Beispiel 9.13 Die Abbildung

$$\varphi : \{a, b, c\} \to \{0, 1\}^*$$
$$\varphi(a) = 0 \quad , \quad \varphi(b) = 1 \quad , \quad \varphi(c) = 01$$

ist ein injektiver Code, aber $\varphi(c) = \varphi(a)\varphi(b) = 01$.

Wissen wir dagegen bei einer (injektiven) *Präfixcodierung*, wo das erste von mehreren Wörtern beginnt, dann können wir die Wortfolge eindeutig decodieren, denn wir müssen ja nur schauen, welches das eindeutige erste Wort in der zu decodierenden Folge von Buchstaben ist.

Abbildung 15:

Beispiel 9.14 $\varphi : \{a, b, c\} \to \{0, 1\}^*,\ \varphi(a) = 00,\ \varphi(b) = 1,\ \varphi(c) = 01$ ist ein Präfixcode, denn hier stehen die Buchstaben des Alphabets $\{a, b, c\}$ an den Blättern des Baumes.

$$\begin{array}{cccc} 1 & 0\,1 & 0\,0 & 0\,1 \\ \cup & \cup & \cup & \cup \\ b & c & a & c \end{array}$$

codiert nur die Buchstabenfolge $bcac$.

Natürlich können wir auch bei einem *(injektiven) Blockcode* die Wortfolge eindeutig decodieren, wenn wir wissen, wo das erste Wort beginnt.

Der Zweck der Quellencodierung ist es ja, die zu übermittelnde Nachricht möglichst kompakt zu schreiben, ohne ihre eindeutige Decodierung zu gefährden.

Wollen wir *jede* Nachricht eindeutig rekonstruieren, dann müssen wir einen Präfixcode verwenden (und den Anfang der Nachricht markieren). Um welchen Faktor können wir bei der Quellencodierung die Nachricht im Mittel verkürzen?

Um diese Frage zu beantworten, bemerken wir zunächst, dass bei einer stationären Quelle $(\Omega, \mathcal{F}_A, P)$ nach Definition 9.5 die Wahrscheinlichkeit einer mit der Shiftabbildung T verschobenen Zylindermenge gleich der der unverschobenen Menge ist. Wollen wir also eine günstige Präfixcodierung in (9.22) auswählen, dann genügt es nach Wahl der Blocklänge r, bezüglich des Wahrscheinlichkeitsmaßes P_r auf A^r, das durch $P_r(a_1 \dots a_r) := P(Z_{\{1,\dots,r\}}(a_1, \dots, a_r)$ gegeben ist, eine Präfixcodierung

$$\varphi : A^r \to B^*$$

minimaler mittlerer Länge

$$\bar{L}(\varphi, P_r) = \sum_{a \in A^r} P_r(a)|\varphi(a)|$$

finden.

Satz 9.15 *Die Präfixcodierung $\varphi : A^r \to B^*$ minimaler mittlerer Länge erfüllt die Abschätzung*

$$\frac{H(P_r)}{\ln |B|} \leq \bar{L}(\varphi, P_r) \leq \frac{H(P_r)}{\ln |B|} + 1$$

Bew.: Wir fassen im Sinn des letzten Kapitels A^r als Suchbereich auf. Dieser besitzt $|A|^r$ Elemente. Die Kanten eines $(|A|^r, q)$–Suchbaumes T mit $q := |B|$ (also der Zahl der Buchstaben in B), können wir so die Buchstaben von B zuordnen, dass die von einer Ecke von T abgehenden, von der Wurzel wegweisenden Kante alle verschiedene Buchstaben tragen. Damit wird jedem Blatt von T das Wort zugeordnet, das sich ergibt, wenn man die Buchstaben von der Wurzel bis zum Blatt hintereinanderschreibt und wir erhalten so einen Präfixcode, dessen Wortlängen durch die Länge der Blätter von T gegeben ist. Die Aussage folgt

also aus Satz 8.11. □

Insbesondere sehen wir, daß für $|A| = |B|$ die Quellencodierung die Nachricht stets verkürzt, natürlich nur im Mittel. Wegen der Injektivität der Codierung gibt es andererseits (selten auftretende) Nachrichten, die durch die Quellencodierung verlängert werden.

Oft möchte man statt eines Präfixcodes einen Blockcode verwenden, z.B. um nachrichenunabhängig die Übertragung bis zu einem bestimmten Zeitpunkt zu beenden.

Will man mit einem Blockcode die Nachricht komprimieren, dann muss man auf die eindeutige Decodierbarkeit der seltenen Nachrichten verzichten.

9.3 Kanalcodierung

Wir kommen jetzt zum Begriff des klassischen Kanals. Die Nachricht, also eine unendliche Buchstabenfolge $\omega \in \Omega$, wird von der Quelle mit einer vorgegebenen Wahrscheinlichkeitsverteilung produziert, über den Kanal geschickt und empfangen. Wir wollen zulassen, dass der Kanal Nachrichten in einem Alphabet empfängt und in einem anderen wieder ausgibt. Gibt man z.B. die Folge $(\ldots \omega_{-1}, \omega_0, \omega_1 \ldots) = \omega \in A^{\mathbb{Z}}$ über dem Alphabet A ein, dann gibt der Kanal eine Folge $(\ldots \nu_{-1}, \nu_0, \nu_1 \ldots) = \nu \in B^{\mathbb{Z}}$ aus. Die bedingte Wahrscheinlichkeit der Ausgaben ν bei Eingabe von ω definiert den Kanal; nur wenn der Kanal *deterministisch* ist, ist diese bedingte Wahrscheinlichkeit genau für ein $\nu(\omega)$ gleich 1. I.A. sind Kanäle *gestört*, sodass die Ausgabe nicht sicher durch die Eingabe bestimmt ist.

In praktischen Anwendungen hängt die Wahrscheinlichkeit $P(\{\nu_n = b\}|\omega)$ dafür, dass der n–te Ausgabebuchstabe ν_n bei Eingabe von ω ein $b \in B$ ist, meist nur von den Buchstaben $\ldots \omega_{n-2}, \omega_{n-1}, \omega_n$ ab, die schon eingegeben wurden: der Kanal ist *nicht antizipierend*. In dem einfachen hier zu behandelnden Fall wird diese Wahrscheinlichkeit nur von ω_n abhängen.

Definition 9.16 Ein *Kanal ohne Gedächtnis* besteht aus einem

- *Eingabealphabet* A, einem
- *Ausgabealphabet* B und einer

- *Kanalmatrix* M mit Einträgen $M_{ij} \geq 0 \qquad (i \in A,\ j \in B)$

$$\sum_{j \in B} M_{ij} = 1 \qquad (i \in A)$$

Ein solcher Kanal ohne Gedächtnis (A, B, M) bestimmt die bedingte Wahrscheinlichkeit

$$P\left(\{\nu \in B^{\mathbb{Z}} | (\nu_m, \nu_{m+1}, \ldots, \nu_n) = (b_m, b_{m+1}, \ldots, b_n)\} | \omega\right)$$

dass an der m–ten bis n–ten Stelle die Buchstaben $b_m, \ldots, b_n$ ausgegeben werden, zu $\prod_{i=m}^{n} M_{\omega_i, b_i}$.

Beispiel 9.17 Binärer symmetrischer Kanal (A, B, M) mit

$$A = B = \{0, 1\}, \quad \begin{matrix} M_{00} = M_{11} = 1 - p, \\ M_{01} = M_{10} = p \end{matrix} \quad \text{mit} \quad p \in [0, 1].$$

p ist die Fehlerrate, also der mittlere Anteil an Zeichen, die verändert werden.

Ziel ist es nun, die Nachricht mit hoher Wahrscheinlichkeit wieder richtig zu decodieren. Dies gelingt gut, wenn p klein ist, aber auch, wenn p nahe bei 1 ist, denn in diesem Fall können wir als Decodierung einfach die Vertauschung der Zeichen 0 und 1 benutzen.

Nun wird die Fehlerrate bei der Codierung nicht nur vom Kanal, sondern auch von der Quelle abhängen, denn die Quelle beschreibt ja die Häufigkeit, mit der die Buchstaben des Alphabetes A vorkommen.

Beispiel 9.18 Wir schreiben mit einer alten Schreibmaschine eine Geschichte auf. Wir sind die Quelle, die Schreibmaschine der Kanal, das Papier die Senke. Die Maschine funktioniert gut, nur dass der Buchstabe ß bei ihr wie ein B aussieht. Nach der Rechtschreibreform beeinträchtigt dies die Decodierbarkeit deutlich weniger.

Betrachten wir einfachheitshalber eine Quelle ohne Gedächtnis, die die Buchstaben $\omega_i \in A$ mit Wahrscheinlichkeiten $p(\omega_i)$ ausgibt. Dann ist bei Übertragung des i–ten Buchstabens die Wahrscheinlichkeit, dass dieser $a \in A$ ist und der Kanal ihn in $b \in B$ umwandelt, gleich

$$\hat{p}\,(a, b) := p(a) M(a, b)$$

$\hat{p}$ ist damit eine Wahrscheinlichkeitsverteilung auf $A \times B$. Die Wahrscheinlichkeit, dass der Kanal bei Übertragung des i–ten Buchstabens ein $b \in B$ produziert, ist gleich

$$q(b) := \sum_{a \in A} p(a) M(a,b) = \sum_{a \in A} \hat{p}\,(a,b),$$

und q ist eine Wahrscheinlichkeitsverteilung auf dem Ausgabealphabet B.

Damit haben wir neben $\hat{p}$ noch eine zweite Wahrscheinlichkeitsverteilung auf $A \times B$, die *Produktverteilung*

$$p \times q,$$

die (a,b) die Wahrscheinlichkeit $p(a) \cdot q(b)$ zuordnet. Bei der Produktverteilung bestehen überhaupt keine Korrelationen zwischen Ein- und Ausgabe, denn für sie ist die bedingte Wahrscheinlichkeit der Eingabe $a \in A$ beim Auftreten von $b \in B$ gleich $\frac{p \times q(a,b)}{q(b)} = p(a)$ und damit unabhängig vom beobachteten Zeichen b. Wir setzen daher

Definition 9.19 Es heißen

- $H(p)$ *Eingangsentropie*,
- $H(q)$ *Ausgangsentropie*,
- $H(\hat{p}\,)$ *Verbundentropie*,
- $I(\hat{p}\,) := H(p \times q) - H(\hat{p}\,)$ *(mittlere) Transinformation*.

Nach Satz 9.9 ist die Entropie der Produktverteilung die Summe von Eingangs- und Ausgangsentropie:

$$H(p \times q) = H(p) + H(q), \tag{9.23}$$

und die Transinformation von $\hat{p}$ ist der Kullback-Leibler-Abstand zur Produktverteilung:

$$I(\hat{p}\,) = S(\hat{p}\,, p \times q).$$

Die Transinformation ist damit ein *Maß für die Abhängigkeit von Ein- und Ausgabe*, und es gilt:

Satz 9.20 $0 \leq I(\hat{p}\,) \leq \min(H(p), H(q))$, *und die Transinformation* $I(\hat{p}\,)$ *ist genau dann gleich Null, wenn* $\hat{p} = p \times q$ *gilt, Eingangs -und Ausgangsverteilungen also unabhängig sind.*

Bew.: • Wegen der Symmetrie bei Vertauschung von A und B reicht es, die Ungleichung

$$0 \leq I(\hat{p}\,) \leq H(q)$$

zu zeigen. Wegen (9.23) ist

$$I(\hat{p}\,) = H(q) + H(p) - H(\hat{p}\,). \tag{9.24}$$

• $I(\hat{p}\,) \geq 0$ folgt also aus der in Satz 9.9 gezeigten Subadditivität der Entropie.

• $I(\hat{p}\,) \leq H(q)$ ist nach (9.24) eine Folge der Monotonieeigenschaft

$$H(\hat{p}\,) \geq H(p) \tag{9.25}$$

der Entropie (und wir hatten in Bem. 9.10 gesehen, dass etwas entsprechendes quantenmechanisch nicht gilt). Um nun (9.25) zu zeigen, schreiben wir die Differenz in der Form

$$H(\hat{p}\,) - H(p) = \sum_{a \in A} \sum_{b \in B} -p(a) M(a,b) [\ln(p(a)) + \ln(M(a,b))] + \sum_{a \in A} p(a) \ln p(a).$$

Da $\sum_{b \in B} M(a,b) = 1$ ist, heben sich der erste und der dritte Term gegeneinander weg. Der übrigbleibende dritte Term ist aber positiv:

$$H(\hat{p}\,) - H(p) = \sum_{a \in A} p(a) \sum_{b \in B} \eta(M(a,b)) \geq 0. \qquad \square \tag{9.26}$$

Wir schauen uns nun den Fall $I(\hat{p}\,) = H(q)$ etwas genauer an, bei dem die Transinformation gleich der Ausgangsentropie ist. Nach (9.24) bedeutet dies, dass $H(\hat{p}\,) = H(p)$, oder unter der Voraussetzung positiver Wahrscheinlichkeiten $p(a) > 0$ für alle $a \in A$ nach (9.26).

$$\sum_{b \in B} M(a,b) \ln M(a,b) = 0 \qquad (a \in A). \tag{9.27}$$

Nun ist aber $M(a,\cdot)$ eine Wahrscheinlichkeitsverteilung auf B. (9.27) verlangt, dass diese Entropie Null hat, dass a also auch genau ein Zeichen b' abgebildet wird, also $M(a,b')=1$ und $M(a,b)=0$ für $b\in B\backslash\{b'\}$.

Einen solchen Kanal nennt man, wie schon erwähnt, deterministisch.

Analog folgern wir, dass für

$$I(\hat{p}\,)=H(p)$$

der Kanal *ungestört* ist, man also mit Sicherheit auf das Eingangssignal zurückschließen kann.

Wegen der Definition von $\hat{p}$ hängt die Transinformation nicht nur von den durch M fixierten Eigenschaften des Kanals, sondern auch von den durch p fixierten Eigenschaften der gedächtnislosen Quelle ab. Es liegt daher nahe, durch

$$C:=\max_{\substack{p:A\to[0,1]\\ \sum_a p(a)=1}} I(\hat{p}\,(p,M))$$

die bei geeigneter Quelle erreichbare maximale *Kapazität des Kanals* (A,B,M) zu definieren. Dies wird durch den Kanalcodierunssatz präzisiert, der unser nächstes Ziel ist.

Unser Ziel ist es, einen effektiven Blockcode zu finden, der mit hoher Wahrscheinlichkeit richtig decodierbar ist und gleichzeitig eine hohe Übertragungsrate garantiert.

Definition 9.21 Für $\varepsilon\geq 0$ ist ein ε*–Code der Ordnung* N und *Länge* t für den Kanal (A,B,M) ein N–Tupel

$$\big((\alpha^{(1)},E^{(1)}),\dots(\alpha^{(N)},E^{(N)})\big)$$

mit $\alpha^{(1)},\dots,\alpha^{(N)}\in A^t$ und $E^{(1)},\dots,E^{(N)}\subseteq B^t$, wobei die $E^{(i)}$ disjunkt sind und für die Wahrscheinlichkeitsverteilung $\hat{p}^{\,t}=\underbrace{\hat{p}\times\dots\times\hat{p}}_{t\text{ mal}}$ auf $A^t\times B^t$ gilt:

$$\hat{p}^{\,t}\left(\{a^{(i)}\}\times E^{(i)}\right)\geq 1-\varepsilon\quad,\quad(i=1,\dots,N).$$

Mit Wahrscheinlichkeit $\geq 1-\varepsilon$ wandelt der Kanal also das Wort $\alpha^{(i)}$ in eines der Wörter aus $E^{(i)}\subseteq B^t$ um; da die $E^{(i)}$ disjunkt sind, lässt sich also das Eingabewort mit einer Wahrscheinlichkeit $\geq 1-\varepsilon$ rekonstruieren. Ist $\varepsilon<\frac{1}{2}$, dann sind die $\alpha^{(i)}$ voneinander verschieden.

Für beliebig kleine $\varepsilon>0$ existieren ε–Codes der Ordnung N. Man muss nur Codewörter $\alpha^{(i)}$ mit großem Hammingabstand wählen.

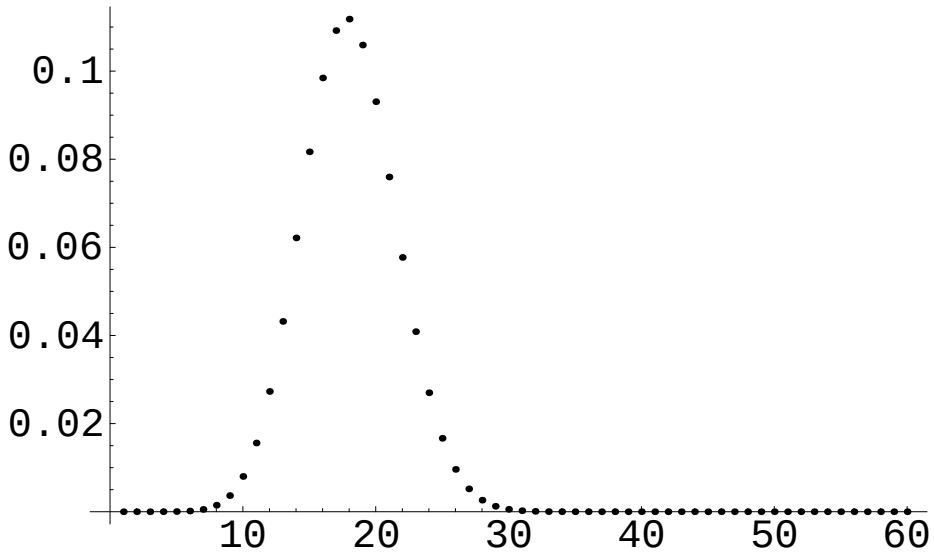


Abbildung 16: Binomial-Verteilung (Fehlerwahrscheinlichkeit $p = 0.3$, $t = 60$).

Definition 9.22 Der *Hammingabstand* $d : A^t \times A^t \to \mathbb{N}_0$ ist durch $d(\alpha, \alpha') := |\{i \in \{1, \ldots, t\} | \alpha_i \neq \alpha'_i\}|$ gegeben.

Es handelt sich also um eine Metrik auf dem Raum A^t der t–buchstabigen Wörter über dem Alphabet A.

Beispiel 9.23 t**–Wiederholungscode** für $A = B = \{1, 2\}$

Es wird $1 \in A$ durch $\alpha^{(1)} = (1, \ldots, 1) \in A^t$ codiert und $2 \in A$ durch $\alpha^{(2)} = (2, \ldots, 2) \in A^t$. Dann ist $d(\alpha^{(1)}, \alpha^{(2)}) = t$, und für ungerade $t = 2t' + 1$ (Mehrheitsentscheid!) setzen wir

$$E^{(i)} := \{\alpha' \in A^t | d(\alpha', \alpha^{(i)}) \leq t'\}.$$

Damit ist $E^{(1)} \cap E^{(2)} = \emptyset$ (Dreiecksungleichung), und $\hat{p}^{\,t}(\{\alpha^{(i)}\} \times E^{(i)})$ ist für einen symmetrischen Kanal mit Fehlerwahrscheinlichkeit ε' und $p(1) = p(2) = \frac{1}{2}$, also $\hat{p} = \frac{1}{2} \begin{pmatrix} 1 - \varepsilon' & \varepsilon' \\ \varepsilon' & 1 - \varepsilon' \end{pmatrix}$, durch den Wert

$$\hat{p}^{\,t}\left(\{\alpha^{(i)}\} \times E^{(i)}\right) = \sum_{n=0}^{t'} (1 - \varepsilon')^{t-n} \varepsilon'^n \binom{t}{n}$$

der Verteilungsfunktion der Binomialverteilung gegeben, der für $\varepsilon' < \frac{1}{2}$ und $t \to \infty$ mit exponentieller Rate gegen 1 strebt:

$$1 - \hat{p}^{\,t}\left(\{\alpha^{(i)}\} \times E^{(i)}\right) = \sum_{n=t'+1}^{t} (1-\varepsilon')^{t-n}\varepsilon'^{n}\binom{t}{n} \leq 2^t[(1-\varepsilon')\varepsilon']^{\frac{t}{2}} = c^t$$

mit $c := 2\sqrt{(1-\varepsilon')\varepsilon'} < 1$.

Die Kunst besteht also nicht so sehr darin, durch naheliegende Modifikationen des vorangegangenen Beispiels ε–Codes der Ordnung N zu finden, sondern solche mit hoher *Übertragungsrate* Basis!

$$R := \frac{\log(N)}{t},$$

also kleiner Länge. Hier gilt die folgende fundamentale Aussage:

Satz 9.24 (Kanalcodierungssatz) *Die maximal erreichbare Übertragungsrate R für einen ε–Code der Länge t ist durch*

$$C - \frac{k}{\sqrt{t}} \leq R \leq C + \frac{k}{\sqrt{t}}$$

gegeben, wobei C die Kapazität des Kanals und k eine (u.a. von ε abhängige) Konstante ist.

Bew.: Eine Beweismethode für die linke Ungleichung basiert auf *zufälligen Codes* der Ordnung N (man wählt also die $\alpha^{(i)} \in A^t$) zufällig). Man zeigt dann, dass für die meisten zufälligen Codes die Wörter $\alpha^{(i)}$ großen Hammingabstand voneinander haben und damit gut decodiert werden können. Diesen auf Shannon zurückgehenden Beweis (zusammen mit einem zweiten) findet man z.B. in [Ja], die (nicht optimale) rechte Ungleichung in [Wo]. □

Ist also die Zahl t der Buchstaben des Blockcodes groß, dann wird die maximal erreichbare Übertragungsrate wirklich durch die Kanalkapazität C, d.h. durch die maximale Transinformation bestimmt.

Es ist aber eine Angelegenheit, die *Existenz* effektiver Codes der Ordnung N zu beweisen und eine andere, solche Codes tatsächlich zu *konstruieren*. Während der Existenzbeweis in den Bereich der Wahrscheinlichkeitstheorie fällt, ist das

zweite Problem eher algebraisch-zahlentheoretisch, besitzt aber auch einen geometrischen Anteil.

Für eine Primzahl p und das Alphabet $A := \mathbb{F}_p \equiv \mathbb{Z}/p\mathbb{Z}$, d.h. den Körper der Restklassen (mod p), ist A^t ein t–dimensionaler Vektorraum über $\mathbb{F}_p$. Die Aufgabe, möglichst viele Codewörter $\alpha^{(i)}, \ldots, \alpha^{(N)} \in A^t$ mit Hammingabstand $> 2h$ zu finden, bedeutet ja geometrisch, N disjunkte Kugeln vom Radius h in A^t zu packen. Dass wir nun A^t nicht einfach als p^t–elementige Menge, sondern als Vektorraum auffassen, hat weitreichende Folgen. Insbesondere liegt es nahe, einen Unterraum als Menge von Codewörtern zu benutzen (sog. *linearer Code*), denn dann sieht die Umgebung jedes Codewortes gleich aus.

Wir kehren zu dieser Frage zurück, wenn wir quantenmechanische Codes behandeln.

10 Quantenmechanische Informationstheorie

Unser aktuelles Ziel ist eine quantenmechanische Verallgemeinerung der Begriffe der klassischen Informationstheorie, insbesondere von Quelle und Kanal. Zu diesem Zweck erinnern wir uns an den in Definition 4.3 eingeführten Begriff der C^*–Algebra, der einen Vergleich von klassischer und Quantenmechanik ermöglichte. In beiden Fällen beschrieb eine C^*–Algebra $\mathcal{A}$ die Menge der Observablen, aber während sie klassisch abelsch war (und damit eine Algebra von Funktionen auf einem Phasenraum), war sie quantenmechanisch eine nicht abelsche Algebra beschränkter Operatoren auf einem Hilbertraum. In beiden Fällen war die Zeitevolution eine Automorphismengruppe von $\mathcal{A}$, und diese Automorphismen wollen wir uns zunächst anschauen.

10.1 Der Quantenmechanische Kanal

Wir betrachten jetzt eine (endlich-dimensionale) C^*–Algebra $\mathcal{G}$, unser *gesamtes* physikalisches System. Dieses soll zwei eventuell verschiedene Aufspaltungen

$$\mathcal{G} = \mathcal{A}_- \otimes \mathcal{E}_- = \mathcal{A}_+ \otimes \mathcal{E}_+ \tag{10.28}$$

besitzen, wobei wir mit dem Index '-' die Zeit vor einer durch den $*$– Automorphismus

$$\alpha : \mathcal{G} \to \mathcal{G}$$

beschriebenen Dynamik auf $\mathcal{G}$, und mit dem Index '+' die Zeit danach gekennzeichnet wird.

$\mathcal{A}_\pm$ soll das von uns zum entsprechenden Zeitpunkt beobachtete System sein, $\mathcal{E}_\pm$ die von uns ignorierte Umwelt. Es sei $\epsilon_- \in \mathcal{S}(\mathcal{E}_-) \subseteq \mathcal{E}_-$ der Zustand der Umwelt vor der Wechselwirkung α.

Dann ist

$$T : \mathcal{A}_- \to \mathcal{A}_+ \quad , \quad \rho_- \mapsto \rho_+ := \mathrm{tr}_{\mathcal{E}_+}(\alpha(\rho_- \otimes \epsilon_-)) \tag{10.29}$$

eine lineare Abbildung.

Definition 10.1 Lineare Abbildungen $T \in B(\mathcal{A}_-, \mathcal{A}_+)$ heißen *Transformationen*. Eine Transformation T heißt *positiv*, wenn für $a_- \in \mathcal{A}_-$ mit $a_- \geq 0$ auch $T(a_-) \geq 0$ folgt.

Eine Transformation T heißt *zulässig*, wenn sie von der Form (10.29) ist.

Satz 10.2 *Zulässige Transformationen T sind positiv und bilden Zustände $\rho_- \in \mathcal{S}(\mathcal{A}_-) \subseteq \mathcal{A}_-$ auf Zustände $\rho_+ = T(\rho_-) \in \mathcal{S}(\mathcal{A}_+) \subseteq \mathcal{A}_+$ ab.*

Bew.: Es gilt nach Lemma 6.7

$$\mathrm{tr}_{\mathcal{A}_+}(\rho_+) = \mathrm{tr}_{\mathcal{G}}(\alpha(\rho_- \otimes \epsilon_-)) = \mathrm{tr}_{\mathcal{G}}(\rho_- \otimes \epsilon_-) = 1,$$

und mit $\rho_- = \hat{\rho}_-^* \hat{\rho}_-$, $\epsilon_- = \hat{\epsilon}_-^* \hat{\epsilon}_-$

$$\begin{aligned} \rho_+ &= \mathrm{tr}_{\mathcal{E}_+}\left(\alpha(\hat{\rho}_-^* \hat{\rho}_- \otimes \hat{\epsilon}_-^* \hat{\epsilon}_-)\right) = \mathrm{tr}_{\mathcal{E}_+}\left(\alpha(\hat{\rho}_-^* \otimes \hat{\epsilon}_-^* \cdot \hat{\rho}_- \otimes \hat{\epsilon}_-)\right) \\ &= \mathrm{tr}_{\mathcal{E}_+}\left(\alpha(\hat{\rho}_- \otimes \hat{\epsilon}_-)^* \alpha(\hat{\rho}_- \otimes \hat{\epsilon}_-)\right) \geq 0. \end{aligned}$$ □

Sind nun die C^*–Algebren $\mathcal{A}_\pm$ abelsch, dann können wir die Paare $(\mathcal{A}_\pm, \rho_\pm)$ als Quellen ohne Gedächtnis im Sinne von Definition 9.5 auffassen, denn

- $\mathcal{A}_\pm \cong C(A_\pm, \mathbb{C})$ mit den Alphabeten $A_\pm := \{1, \ldots, n_\pm\}, n_\pm := \dim(\mathcal{A}_\pm)$
- $p_\pm : A_\pm \to [0,1]$ mit $p_\pm(i) := \rho_\pm(\mathbb{1}_i)$ definiert eine Wahrscheinlichkeitsverteilung auf $A_\pm$.

Andererseits definiert eine lineare Abbildung $T : \mathcal{A}_- \to \mathcal{A}_+$ der Form (10.29) einen Kanal ohne Gedächtnis im Sinn von Definition 9.16:

Satz 10.3 *Sind die C^*–Algebren $\mathcal{A}_\pm$ abelsch, dann hat mit den obigen Bezeichnungen jede zulässige Transformation $T : \mathcal{A}_- \to \mathcal{A}_+$ die Form*

$$p_+(b) = \sum_{a \in A_-} M_{a,b}\, p_-(a) \qquad (b \in A_+) \tag{10.30}$$

mit

$$M_{a,b} \geq 0 \quad , \quad \sum_{b \in A_+} M_{a,b} = 1 \qquad (a \in A_-). \tag{10.31}$$

Umgekehrt existiert für jede $n_- \times n_+$–Matrix M mit den Eigenschaften (10.31) eine zulässige Transformation $T : \mathcal{A}_- \to \mathcal{A}_+$.

Bew.: Der erste Teil des Satzes folgt aus Satz 10.2, denn Transformationen, die klassische Zustände auf klassische Zustände abbilden, müssen durch eine Matrix der Form (10.31) beschrieben werden.

Zum Beweis der Umkehrung benutzen wir die abelschen C^*–Algebren $\mathcal{E}_\pm$ in (10.28), die die Gestalt

$$\mathcal{E}_\pm := C(E_\pm, \mathbb{C})$$

mit Mengen $E_\pm$ der Kardinalität

$$|E_-| := n_- \cdot n_+^2 \quad , \quad |E_+| := n_-^2 \cdot n_+$$

besitzen. Die den Zustand ϵ_- auf $\mathcal{E}_-$ induzierende Wahrscheinlichkeitsverteilung e_- auf dem Alphabet E_- definieren wir folgendermaßen:

- Wir ordnen die $n := n_- \cdot n_+$ Zahlen

  $$\sum_{c=1}^{b} M_{a,c} \qquad (a \in A_-, b \in A_+) \tag{10.32}$$

  in aufsteigender Größe und nennen diese $z_1, \ldots, z_n$.

  Es gilt $0 \leq z_1 \leq z_2 \leq \ldots \leq z_n = 1$.

- Wir setzen (mit $z_0 := 0$)

  $$e_-(i) := z_i - z_{i-1} \qquad (i = 1, \ldots, n)$$

  und füllen mit Nullen auf, d.h. $e_-(i) := 0 \qquad (i = n+1, \ldots, n_- \cdot n_+^2)$.

Damit ist e_- tatsächlich eine Wahrscheinlichkeitsverteilung auf E_-.

Weiter gibt es für jeden Eingabebuchstaben $a \in A_-$ eine Partition $E_- = \bigcup_{b\in A_+} D_{a,b}$ von E_- mit Kardinalitäten $|D_{a,b}| = n$ und der Eigenschaft

$$M_{a,b} = \sum_{i\in D_{a,b}} e_-(i), \tag{10.33}$$

denn ist in (10.32) $z_i = \sum_{c=1}^{b-1} M_{a,c}$ und $z_j = \sum_{c=1}^{b} M_{a,c}$, dann ist

$$M_{a,b} = z_j - z_i = \sum_{r=i+1}^{j} e_-(r);$$

außerdem ist $j-i \leq n$, und wir können $D_{a,b}$ so wählen, dass es neben den Indizes $i+1,\ldots,j$ noch entsprechend viele Indizes $k > n$ enthält, für die $e_-(k) = 0$ ist.

Nach Satz 5.4 wird der $*$–Automorphismus α in (10.29) durch eine Permutation $\pi : G \to G$ auf $G \cong A_- \times E_- \cong A_+ \times E_+$ beschrieben, und wir können beliebige π wählen.

Wenn π die Produktverteilung $p_- \times e_-$ (entsprechend dem anfänglichen Produktzustand $\rho_- \otimes \epsilon_-$ in (10.29)) so umordnet, dass zu $p_+(b)$ genau die Terme

$$p_-(a)e_-(i) \quad \text{mit} \quad i \in E_{a,b} \qquad (a \in A_-)$$

beitragen, ergibt sich (10.30) aus (10.33). □

Wir haben nun einerseits mit diesem Satz eine physikalische Begründung des klassischen Kanals ohne Gedächtnis gefunden: Jeder solche Kanal kann durch Wechselwirkung mit einer geeigneten Umwelt realisiert werden. Einerseits bewirkt jede Umwelt, ob klassisch oder quantenmechanisch, das aus der Informationstheorie bekannte Verhalten eines Kanals.

Andererseits können wir den Satz zum Ausgangspunkt einer Verallgemeinerung des klassischen Kanalbegriffes machen, indem wir einfach beliebige zulässige Transformationen $T : \mathcal{A}_- \to \mathcal{A}_+$ von C^*–Algebren als *Kanal* bezeichnen.

10.2 Vollständige Positivität

Hier zeigt sich nun sofort eine Besonderheit der Quantenmechanik. Klassisch sind nach dem letzten Satz die zulässigen Transformationen genau diejenigen,

die Zustände auf Zustände abbilden, also positiv und spurerhaltend $(\mathrm{tr}(T(a)) = \mathrm{tr}(a))$ sind.

Quantenmechanisch aber können nicht alle solche Transformationen physikalisch realisiert werden. Sie sind also nicht notwendig zulässig. Dazu fehlt ihnen die vollständige Positivität:

Definition 10.4 Für eine Transformation $T : \mathcal{A}_- \to \mathcal{A}_+$ sei (mit $\mathcal{M}_n = B(\mathbb{C}^n)$)

$$T \otimes \mathbb{1}_{\mathcal{M}_n} : \mathcal{A}_- \otimes \mathcal{M}_n \to \mathcal{A}_+ \otimes \mathcal{M}_n$$

die durch $a \otimes b \mapsto T(a) \otimes b$ definierte Transformation. T heißt *vollständig positiv*, wenn für alle $n \in \mathbb{N}$ $\quad T \otimes \mathbb{1}_{\mathcal{M}_n}$ positiv ist.

- Vollständig positive Transformationen sind natürlich positiv (man nehme $n = 1$, also $\mathcal{M}_n = \mathbb{C}$).
- In unser physikalischen Deutung bedeutet vollständige Positivität einer spurerhaltenden Transformation, dass auch Zustände des Gesamtsystems $\mathcal{A}_- \otimes \mathcal{E}$ in Zustände des Gesamtsystems $\mathcal{A}_+ \otimes \mathcal{E}$ übergeführt werden, wenn der Zustand der Umwelt $\mathcal{E}$ unverändert bleibt.

Satz 10.5 (Stinespring [St]) *Ist mindestens eine der C^*–Algebren $\mathcal{A}_\pm$ abelsch, dann ist jede positive Transformation $T : \mathcal{A}_- \to \mathcal{A}_+$ vollständig positiv.*

Bew.: Wir nehmen $\dim(\mathcal{A}_\pm) < \infty$ an; für den Beweis im allgemeinen Fall siehe auch Takesaki [Ta], Par. IV.3.

- Es sei $\mathcal{A}_-$ abelsch, T positiv und $n \in \mathbb{N}$. Dann ist $A \in \mathcal{A}_- \otimes \mathcal{M}_n$ von der Form

  $$A = [A_{ik}]_{i,k=1}^n \quad \text{mit} \quad A_{ik} \in \mathcal{A}_-.$$

  Wir können annehmen, dass $\mathcal{A}_- = C(P_-, \mathbb{C})$ mit $|P_-| = \dim(\mathcal{A}_-)$ ist. Damit ist das Spektrum

  $$\sigma(A) = \bigcup_{p \in P_-} \sigma(A(p)) \quad \text{mit} \quad A(p) = [A_{ik}(p)]_{i,k=1}^n.$$

  Also gilt $A \geq 0$ genau dann, wenn $A(p) \geq 0$ für alle $p \in P_-$.

Nun ist $B := (T \otimes \mathbb{1})(A) \in \mathcal{A}_+ \otimes \mathcal{M}_n$ von der Form $B = [B_{ik}]_{i,k=1}^n$ mit $B_{ik} = T(A_{ik}) \in \mathcal{A}_+$.

Wir zerlegen $A \geq 0$ in die Summe

$$A = \sum_{q \in P_-} A^{(q)} \quad \text{mit} \quad A^{(q)}(p) := \delta_q(p) A(q)$$

positiver Operatoren $A^{(q)} \in \mathcal{A}_- \otimes \mathcal{M}_n$ (mit $\delta_q(p) = 1$ für $p = q$, sonst $\delta_q(p) = 0$.

Damit ist

$$B = \sum_{q \in P_-} T(\delta_q) \otimes A(q),$$

und nach Voraussetzung ist $T(\delta_q) \geq 0$.

Damit ist $B \geq 0$.

- Es sei $\mathcal{A}_+$ abelsch, also $\mathcal{A} \cong \mathbb{C}^r$.

  Damit ist $A \in \mathcal{A}_+ \otimes \mathcal{M}_n$ von der Form $A = (A^{(l)}, \ldots, A^{(r)})$ mit $A^{(l)} \in \mathcal{M}_n$), und $\sigma(A) = \bigcup_{l=1}^r \sigma(A^{(l)})$. Es reicht also aus, den Fall $r = 1$, also $T : \mathcal{A}_- \to \mathbb{C}$ zu untersuchen.

  Analog zu Satz 6.3 ist die positive Transformation T von der Form

  $$T(a) = \mathrm{tr}(\rho a) \qquad (a \in \mathcal{A}_-)$$

  mit positivem $\rho \in \mathcal{A}_-$ (aber im Allgemeinen $\mathrm{tr}(\rho) \neq 1$).

  Damit ist für $A \in \mathcal{A}_- \otimes \mathcal{M}_n$, $A \geq 0$

  $$(T \otimes \mathbb{1})(A) = \mathrm{tr}((\rho \otimes \mathbb{1})A) = \mathrm{tr}(C^* A C) \geq 0$$

  mit $C := \sqrt{\rho} \otimes \mathbb{1}$. □

Allerdings ist schon im einfachsten nichtabelschen Fall $\mathcal{A}_\pm = \mathcal{B}(\mathbb{C}^2)$ nicht jede positive Transformation vollständig positiv:

Beispiel 10.6 Ein Beispiel für eine quantenmechanische positive spurerhaltende aber nicht vollständig positive Transformation ist die Transposition

$$T : \mathcal{M}_n \to \mathcal{M}_n \quad , \quad T(\rho) = \rho^T.$$

Mit $\rho = \hat{\rho}^* \hat{\rho}$ gilt $T(\rho) = \hat{\rho}^T \hat{\rho}^{*T} = \hat{\sigma}^* \hat{\sigma}$ mit $\hat{\sigma} := \bar{\hat{\rho}}$ (komplexe Konjugation der Matrixeinträge), und $\mathrm{tr}(T(\rho)) = \mathrm{tr}(\rho)$.

Andererseits ist (unter Verwendung der Bra-ket-Notation, siehe 16) schon für $n = 2$ und für die verschränkten Vektoren

$$|\varphi\rangle := \frac{1}{\sqrt{2}}(|11\rangle + |22\rangle) \equiv \frac{1}{\sqrt{2}}(e_1 \otimes e_1 + e_2 \otimes e_2),$$

$$|\psi\rangle := \frac{1}{\sqrt{2}}(|12\rangle - |21\rangle)$$

aus $\mathbb{C}^2 \otimes \mathbb{C}^2$ der Länge 1 und den reinen Zustand $\rho := |\varphi\rangle \langle\varphi|$

$$\begin{aligned} \langle \psi, T \otimes \mathbb{1}(\rho)\psi \rangle &= \left\langle \psi, \tfrac{1}{2} \left(|11\rangle \langle 11| + |22\rangle \langle 22| + |12\rangle \langle 21| + |21\rangle \langle 12| \right) \psi \right\rangle \\ &= -\tfrac{1}{2}(\langle 12 \,|\, 12\rangle \langle 21 \,|\, 21\rangle + \langle 21 \,|\, 21\rangle \langle 12 \,|\, 12\rangle) = -1. \end{aligned}$$

$\frac{1}{2}$?

Satz 10.7 *Genau die vollständig positiven spurerhaltenden Transformationen sind zulässig.*

Bew.: Siehe Kitaev [Ki], Thm. 3.4 und [Da], Thm. IX.4.3. □

An der Struktur der vollständig positiven Transformationen zeigt sich der Unterschied zwischen klassischer und Quantenmechanik sehr deutlich, und auch zwischen klassischer und quantenmechanischer Berechnung.

Beispiel 10.8 Eine der einfachsten klassischen Rechenoperationen ist die *Kopie von Bits*.

Technisch kann diese z.B. realisiert werden, indem man eine Abzweigung an einen Stromleiter legt. Mathematisch gesehen bedeutet die Kopie eines durch

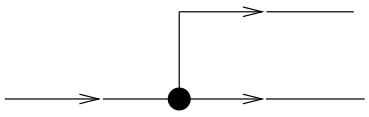


Abbildung 17: Kopie eines Bit

$\rho = \sum_{\lambda\in\sigma(\rho)} \lambda P_\lambda$ (mit minimalen Projektionen P_λ) dargestellten Zustandes $\rho \in \mathcal{S}(\mathcal{A})$ auf einer C^*–Algebra $\mathcal{A}$ die Abbildung

$$\rho \mapsto T(\rho) := \sum_{\lambda\in\sigma(\rho)} \lambda P_\lambda \otimes P_\lambda.$$

$T(\rho)$ ist dann ein Zustand auf $\mathcal{A}\otimes\mathcal{A}$, dessen partielle Spuren ρ ergeben.

Nun ist T für eine abelsche C^*–Algebra zulässig, quantenmechanisch aber noch nicht einmal linear (und bei höheren Multiplizitäten von Eigenwerten λ nicht wohldefiniert!).

Quantenmechanisch können wir also Zustände nicht einfach kopieren.

Wir können quantenmechanische Zustände auch nicht *messen*, ohne sie zu verändern.

Wir können Messungen zunächst als zulässige Transformationen auffassen: Es werde die selbstadjungierte Observable $a = a^* \in \mathcal{A}$ mit Spektraldarstellung $a = \sum_{\lambda\in\sigma(a)} \lambda P_\lambda$ gemessen. Nach Kapitel 6.1 wird ein Zustand mit Dichtematrix $\rho \in \mathcal{S}(\mathcal{A})$ durch die Messung in den $\sum_{\lambda\in\sigma(a)} P_\lambda \rho P_\lambda$ entsprechenden Zustand übergeführt.

Die lineare Transformation

$$\begin{aligned} T_a : \mathcal{A} &\to \mathbb{C}^{\sigma(a)} \otimes \mathcal{A} \cong \bigoplus_{\lambda\in\sigma(a)} \mathcal{A} \\ b &\mapsto \sum_\lambda |\lambda\rangle\langle\lambda| \otimes P_\lambda b P_\lambda \cong \bigoplus_{\lambda\in\sigma(a)} P_\lambda b P_\lambda \end{aligned} \tag{10.34}$$

(mit den kanonischen Basisvektoren $|\lambda\rangle$ des $\mathbb{C}^{\sigma(a)}$) lässt sich in der Form (10.29) schreiben.

Um dies zu sehen, betrachten wir den (endlichdimensionalen) Hilbertraum $\mathcal{H}$ auf dem die Matrixalgebra $\mathcal{A}$ wirkt.

Die Abbildung

$$\mathcal{H} \to \mathbb{C}^{\sigma(a)} \otimes \mathcal{H} \quad , \quad |\psi\rangle \mapsto \sum_{\lambda\in\sigma(a)} |\lambda\rangle \otimes P_\lambda |\psi\rangle$$

ist ja isometrisch und lässt sich als solche zu einer unitären Abbildung $U : \mathcal{H}_- \to \mathcal{H}_+ = \mathbb{C}^{\sigma(a)} \otimes \mathcal{H}$ ausdehnen mit $U|\psi\rangle \otimes |\epsilon\rangle = \sum_{\lambda\in\sigma(a)} |\lambda\rangle \otimes P_\lambda |\psi\rangle$.

Die durch T_a beschriebene Messung besitzt zwei Aspekte:

- T_a, gefolgt von einer partiellen Spurbildung bez. $\mathcal{A}$, liefert eine zulässige Transformation $\mathcal{A} \to \mathbb{C}^{\sigma(a)}$, die einem Zustand auf $\mathcal{A}$ ein Wahrscheinlichkeitsmaß auf dem Spektrum $\sigma(a)$ der gemessenen Observablen a zuordnet.

- T_a, gefolgt von einer partiellen Spurbildung bez. $\mathbb{C}^{\sigma(a)}$ liefert die zulässige Transformation $\mathcal{A} \to \mathcal{A} \quad , \quad b \mapsto \sum_\lambda P_\lambda b P_\lambda$ entsprechend der Veränderung des Zustandes durch Wechselwirkung mit der Messapparatur.

Klassisch ist letztere Abbildung die Identität, d.h. der Zustand wird durch die Messung nicht verändert. Dies ist in der Quantenmechanik nicht möglich.

Es sei $\mathcal{A}_{\rm qm} = \mathcal{M}_m$ und $\mathcal{O}$ die den Zustand der Messapparatur beschreibende C^*–Algebra (also z.B. $\mathcal{O} = \mathbb{C}^{\sigma(a)}$).

Satz 10.9 *Ist $T : \mathcal{A}_{\rm qm} \to \mathcal{O} \otimes \mathcal{A}_{\rm qm}$ eine zulässige Transformation mit*

$$tr_{\mathcal{O}}(T(b)) = b \qquad (b \in \mathcal{A}_{\rm qm}),$$

dann ist für ein $\tau \in \mathcal{S}(\mathcal{O})$ die Transformation von der Form $T(a) = \tau \otimes a$.

Bemerkung 10.10 Das bedeutet, dass zwischen Messapparat und quantenmechanischem Zustand keine Wechselwirkung stattfand und durch exklusive Betrachtung von $\mathrm{tr}_{\mathcal{A}_{\rm qm}}(T(\rho))$ keine Information über den Zustand ρ gewonnen werden kann.

Bew.: Es sei $\mathcal{A}_- := \mathcal{A}_{\rm qm} \quad , \quad \mathcal{A}_+ := \mathcal{O} \otimes \mathcal{A}_{\rm qm}$ und $T : \mathcal{A}_- \to \mathcal{A}_+$ von der Form (10.29). Nach dem Beweis von Satz 10.7 können wir O.B.d.A. annehmen, dass ϵ_- ein reiner Zustand auf einer Matrixalgebra $\mathcal{E}_- = \mathcal{M}_{d_-}$ ist, also $\epsilon_- = |e_-\rangle \langle e_-|$ für einen Vektor $|e_-\rangle \in \mathbb{C}^{d_-}$ der Länge 1. Ist nun auch der Zustand ρ auf $\mathcal{A}_-$ rein, $\rho = |\psi\rangle \langle\psi|$, dann gilt

$$U^* |\psi\rangle \otimes |e_-\rangle = |\tau(\psi)\rangle \otimes |\psi\rangle ,$$

und wegen Linearität von U^* ist $|\tau(\psi)\rangle$ von φ unabhängig. Damit ist
$U^* \rho \otimes \epsilon_- U = \tau \otimes \rho$. □

Mit dem Begriff der zulässigen Transformation, oder gleichbedeutend, des Kanals, können wir alle Vorgänge in einen klassischen oder quantenmechanischen Computer beschreiben:

- Den ungestörten Berechnungen entsprechen die $*$–Automorphismen.

- Störungen aller Art werden durch Kanäle mit gleicher Eingangs- und Ausgangsalgebra modelliert,
- während Codierung und Decodierung i.A. verschiedene Algebren in Beziehung setzen.
- Schließlich entspricht der Messung des Rechenergebnisses der in (10.34) beschriebene Kanal.

Allerdings scheint Satz 10.9 gegen die Möglichkeit zu sprechen, effektiv Fehler zu korrigieren, ohne die Rechnung zu verfälschen. Die Entwicklung solcher quantenmechanischen fehlerkorrigierenden Codes stellte einen entscheidenden theoretischen Durchbruch dar.

11 Fehlerkorrigierende Quantencodes

Wir betrachten zunächst klassische fehlerkorrigierende Codes, und zwar der Einfachheit halber für einen Kanal mit Eingangs- und Ausgangsalphabet $B = \{0, 1\}$.

Die Codewörter werden aus B^t gewählt, sind also Folgen von t Bits und der Code (die Menge der Codewörter) ist eine Teilmenge $C \subseteq B^t$.

Wir gehen davon aus, dass die Fehler in den einzelnen Bits unabhängig und gleichverteilt auftreten und dass die Fehlerrate klein ist. Finden wir daher am Ausgang des Kanals ein Wort $w \in B^t$ vor, dann werden wir es als dasjenige Codewort $c \in C$ interpretieren, das den kleinsten Hammingabstand $d(w, c)$ besitzt.

Besitzen die Codewörter untereinander einen Minimalabstand $\geq 2k+1$, dann ist für $d(w, c) \leq k$ diese Zuordnung eindeutig und wir sprechen davon, dass der Code C k *Fehler korrigiert*.

Nun ist B^t ein t–dimensionaler Vektorraum über dem Körper $B = \mathbb{F}_2$, und in der klassischen Codierungstheorie wird dieser Umstand genutzt, um möglichst viele Vollkugeln vom Radius k in B^t zu packen.

Definition 11.1 $C \subseteq B^t$ heißt *linearer Code*, wenn C ein Unterraum von B^t ist. Ist $k = \dim C$, dann nennt man C einen (t, k)*–Code*.

Ist $c_1, \ldots, c_k \in C$ eine Basis, dann heißt die $k \times t$–Matrix $G = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}$ eine *Generatormatrix* von C.

Der Vorteil der Linearität besteht nun u.a. darin, dass wir, wenn wir ein Wort $w \in B^t$ korrigieren wollen, nicht die Hammingabstände zu allen $c \in C$ berechnen müssen.

Zunächst ist, mit der Bezeichnung

$$Gew(w) := d(w, 0) \qquad (w \in B^t)$$

für das *Gewicht* des Vektors w, der Minimalabstand eines linearen Codes $\mathcal{C}$ durch

$$\begin{aligned} Dist(\mathcal{C}) &= \min_{c \in \mathcal{C} \setminus \{0\}} Gew(c) \end{aligned}$$

gegeben, wie sich aus der Unterraum–Eigenschaft von $\mathcal{C} \subseteq B^t$ und $d(w_1, w_2) = Gew(w_1 - w_2)$ ergibt. Der Minimalabstand eines linearen Codes lässt sich damit leichter errechnen als der eines allgemeinen Codes $C \subseteq B^t$.

Der k–dimensionale Unterraum $C \subseteq B^t$ lässt sich als Kern einer surjektiven linearen Abbildung

$$\Delta : B^t \to B^{t-k}$$

darstellen. Die darstellende Matrix von Δ heißt *Kontrollmatrix*. Diese $(t-k) \times t$–Matrix bezeichnen wir auch mit Δ. Ein Wort $w \in B^t$ ist genau dann ein Codewort, wenn $\Delta t = 0$.

Beispiel 11.2 (Paritätsbit) Der lineare $(t-1, t)$–Code $C := \{(b_1, \ldots, b_t) \in B^t \mid \sum_{i=1}^{t} b_i = 0\}$ besitzt die Generatormatrix FEHLT!!
und die Kontrollmatrix $\Delta = (1, \ldots, 1)$.
Er erlaubt es, einzelne Fehler zu erkennen, nicht aber, sie zu korrigieren.
Der Vektorraum B^t ist bez. der Bilinearform $(\cdot, \cdot) : B^t \times B^t \to B$,

$$(w, w') := \sum_{i=1}^{t} w_i w'_i \qquad (w, w' \in B^t)$$

zu sich selbst dual.

Definition 11.3 Es sei $C \subseteq B^t$ ein linearer Code. Dann heißt

$$C^\perp := \{w \in B^t \mid \forall c \in C : (c, w) = 0\}$$

der zu C *duale Code*. Gilt $C^\perp = C$, dann heißt C *selbstdual*.

Da die Bilinearform auf B^t nicht degeneriert ist, gilt $C^{\perp\perp} = C$.

Damit gilt:

Lemma 11.4 *M Generatormatrix von $C \iff M$ Kontrollmatrix von $C^\perp$.*

Um nun effektiv zu decodieren, zerlegen wir den Vektorraum B^t in Nebenklassen $w_1^{(i)} + C$ nach dem Unterraum C. Dabei wählen wir als Repräsentanten $w^{(i)} \in B^t$ der Nebenklasse einen Vektor minimalen Gewichts, den sog. *Nebenklassenführer*. Nun besitzen alle Elemente einer Nebenklasse unter der Surjektion Δ das gleiche Bild. Empfangen wir also am Ende eines Kanals das Wort $w \in B^t$, dann können wir seine Nebenklasse $w + C$ bestimmen, indem wir sein sog. *Syndrom*

$$w\Delta^T \in B^{t-k}$$

berechnen. Wir vergleichen dieses Syndrom mit einer Liste der Syndrome $w^{(i)}\Delta^T$ der Repräsentanten $w^{(i)}$ und decodieren, indem wir w den Code

$$c := w - w^{(j)} \quad \text{für} \quad w\Delta^T = w^{(j)}\Delta^T$$

zuordnen. Da $w^{(j)}$ in seiner Nebenklasse minimales Gewicht besitzt, hat c minimalen Abstand von w. Es kann aber noch weitere Codewörter $c' \in C$ mit $d(w, c') = d(w, c)$ geben.

Beispiel 11.5 *Simplex– und Hammingcodes*

Zur Konstruktion des binären *Simplex–Codes* $Sim(k) \subseteq B^t$ betrachten wir die $k \times t$–Matrix G mit $t := 2^k - 1$, deren t *Spalten* die (beliebig angeordneten) von Null verschiedenen Vektoren aus B^k sind. Für $k = 3$ ergibt sich (in lexikalischer Ordnung)

$$G = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$$

Jede Zeile von G besitzt 2^{k-1} Einsen und $2^{k-1} - 1$ Nullen, denn es fehlt ja nur der Null(spalten)vektor aus B^k.

Das gleiche gilt für jede von Null verschiedene Linearkombination dieser Zeilen.

G ist nun Generatormatrix von $Sim(k)$; wir erkennen, dass die Codewörter *äquidistant* sind mit Abstand 2^{k-1}.

Fassen wir nun G als Kontrollmatrix auf, dann nennen wir den entsprechenden zum Simplexcode dualen Code $Sim(k)^{\perp}$ den *Hamming–Code* $Ham(k) \subseteq B^t$. Der Minimalabstand zweier Wörter in diesem Code ist

$$Dist(Ham(k)) = 3.$$

Dies sehen wir durch Angabe eines Wortes $c \in Ham(k) \setminus \{0\}$ minimalen Gewichts. Die den Einsen in diesem Wort entsprechenden Spalten von G sind linear abhängig, denn $Gc^T = 0$ und ihre Zahl ist die Gesamtzahl solcher abhängiger Spalten, also gleich 3.

Der Hamming–Code gestaltet damit die Korrektur eines Fehlers, entsprechend den Hammingkugeln vom Radius 1 um die Codewörter. Diese Kugeln umfassen jeweils neben dem Codewort noch $t = 2^k - 1$ weitere Wörter, also insgesamt 2^k Wörter. Es gibt insgesamt $|Ham(r)| = 2^{t-k}$ Codewörter. Damit ist der Hammingcode *perfekt*, d.h. die Hammingkugeln füllen B^t aus.

Um für den Hammingcode die Decodierung durchzuführen, erstellen wir zunächst die Liste der Nebenklassenführer und ihrer Syndrome. Wegen der Perfektheit von $Ham(r)$ können und müssen wir die Kugel vom Radius 1 um die Null als Menge der Nebenklassenführer wählen, also

$$w^{(0)} := (0, \ldots, 0) \in B^t \quad ,$$

$$w^{(i)} := (0, \ldots, 0, \underbrace{1}_{i-\text{te Stelle}}, 0, \ldots, 0) \quad , \quad (i = 1, \ldots, t).$$

Damit ist das Syndrom $\Delta w^{(0)} = (0, \ldots, 0) \in B^k$ und $\Delta w^{(i)}$ ist für $i = 1, \ldots, t$ der i–te Spaltenvektor von G.

Ist z.B. $k = 3$ und empfangen wir das Wort $w = 0010101$, so ist sein Syndrom $w\Delta^T = (1, 0, 0)$, entsprechend dem des Nebenklassenführers $w^{(1)} = 1000000$. Wir korrigieren also, indem wir w als das Codewort

$$c = w - w^{(1)} = 1010101$$

interpretieren.

Da wir die Spalten der Kontrollmatrix von $Ham(k)$ lexikalisch geordnet haben, können wir sogar aus Δw direkt ablesen, an welcher Stelle von w wir gegebenenfalls ein Bit korrigieren müssen, hier also wegen $\Delta w = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ an der Stelle $0 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2^0 = 1$.

Bevor wir die Frage nach den quantenmechanischen Codes untersuchen, wollen wir noch einmal die benutzten Strukturen rekapitulieren.

Unser Ausgangspunkt war ein *Fehlermodell*, d.h. eine Beschreibung derjenigen Fehler, die wir korrigieren wollen. Die Form und Wahrscheinlichkeit auftretender Fehler wird durch den (klassischen) Kanal beschrieben.

Im einfachsten und wichtigsten Fall wird angenommen, dass die Fehler in den einzelnen Bits unabhängig und gleichverteilt auftreten und dass die Fehlerrate für die Verfälschung der Null gleich groß ist wie die für die Verfälschung der Eins.

Ich erinnere an die Grundbegriffe der klassischen Codierungstheorie und formuliere sie so im Schema von C^*–Algebren, dass wir die quantenmechanischen Codes als Verallgemeinerung einführen können.

- Ein *klassischer Code* C war zunächst eine Teilmenge der Menge B^t von Wörtern der Länge t in den Symbolen $\{0, 1\}$.

- Er sollte zur Korrektur von *Fehlern* dienen, und diese bildeten eine Teilmenge $E \subseteq B^t$. Typisch ist der Raum $E(t, k) := \{w \in B^t \mid d(w, 0) \leq k\}$ von höchstens k Fehlern.

- Dies war möglich, wenn $(c_1 + E) \cap (c_2 + E) = \emptyset$ für $c_1, c_2 \in C$, $c_1 \neq c_2$. Für $E = E(t, k)$ bedeutet dies, dass die Hammingkugeln vom Radius k um die Codewörter $c \in C$ nicht überlappen.

- Die Korrektur erfolgte dann durch eine Abbildung

  $$\tilde{P} : B^t \to C,$$

  bei der $c + E$ auf $c \in C$ abgebildet wurde, und der Rest der Wörter aus B^t irgendwie abgebildet wurde.

Wir algebraisieren nun:

- Statt B^t betrachten wir die abelsche C^*–Algebra $\mathcal{A}^{(t)} := C(B^t, \mathbb{C})$ der komplexen Funktionen auf B^t. Damit entspricht C der Algebra

  $$\mathbb{1}_C \mathcal{A}^{(t)} \subseteq \mathcal{A}^{(t)}$$

  der Funktionen mit Träger in C.

- Jedem Fehler $e = e_1 \dots e_t \in E$ entspricht die Permutation $\pi_e : B^t \to B^t$, die genau die Bits an den Stellen $1 \leq i \leq t$ mit $e_i = 1$ umkehrt.

  Wir erinnern uns daran, dass die Permutationen Automorphismen von $\mathcal{A}^{(t)}$ erzeugten und wir wollen als (potentiell) korrigierbare Fehler beliebige zulässige Transformationen $T : \mathcal{A}^{(t)} \to \mathcal{A}^{(t)}$ zulassen, die sich als Konvexkombinationen dieser Automorphismen schreiben lassen.

- Korrigierbarkeit der Fehler ist nun gleichbedeutend mit $\mathrm{tr}(T_1(\rho_1)^* T_2(\rho_2)) = 0$ für $\rho_i \in \mathbb{1}_C \mathcal{A}^{(t)}$ mit $\mathrm{tr}(\rho_1^* \rho_2) = 0$ und beliebigen zulässigen Transformationen T_i.

- $P : \mathcal{A}^{(t)} \to \mathcal{A}^{(t)}$ mit $(P\rho)_w := \sum_{v \in \tilde{P}^{-1}(w)} \rho(v)$ bildet in $\mathbb{1}_C \mathcal{A}^{(t)}$ ab und ist eine zulässige Transformation, also prinzipiell physikalisch realisierbar.

  Tatsächlich korrigiert P die E entsprechenden Fehler, denn für $\rho \in \mathbb{1}_C \mathcal{A}^{(t)}$ und Fehler-Transformation T ist $PT\rho = \rho$.

Wir versuchen das Schema zu quantisieren:

Definition 11.6 Ein *Quantencode* ist ein Unterraum $\mathcal{M} \subseteq \mathcal{N}$ eines endlichdimensionalen Hilbertraumes $\mathcal{N}$.

Es sei $\mathcal{A}_\mathcal{N} := B(\mathcal{N})$ und $R \subseteq B(\mathcal{A}_\mathcal{N})$ eine Menge von Transformationen.

$\mathcal{M}$ *korrigiert Fehler aus* R, wenn eine zulässige Transformation $P \in B(\mathcal{A}_n)$ und ein Funktional $c : R \to \mathbb{C}$ mit

$$PT(\rho) = c(T)\rho \quad \text{für} \quad T \in R \quad , \quad \rho \in \mathcal{A}_\mu := B(\mathcal{M})$$

existiert.

Bemerkung 11.7 Eigentlich sind wir hauptsächlich an *zulässigen* Transformationen $T \in R$ interessiert. Für diese muss $c(T) = 1$ sein, denn die zulässige Transformation PT bildet $\mathbb{1}_\mu \in \mathcal{A}_\mu$ auf $\mathbb{1}_\mu$ ab.

Es sei $R \subseteq B(\mathcal{A}_\mathcal{N})$ eine Menge vollständig positiver Transformationen. Jedes $T \in R$ können wir in der Form $T(\rho) = \mathrm{tr}_{\mathcal{E}^+}(V(\rho \otimes \varepsilon^-)V^*)$ schreiben. Nach Definition der partiellen Spur $\mathrm{tr}_\mathcal{E}^+$ auf der Algebra $G = \mathcal{A}_+ \otimes \mathcal{E}_+$ ist

$$T(\rho) = \sum_{i_1, i_2 = 1}^{n_A} \sum_{j=1}^{n_\mathcal{E}} \langle i_1| \otimes \langle j| \, V(\rho \otimes \varepsilon^-)V^* \, |i_2\rangle \otimes |j\rangle \cdot |i_1\rangle \langle i_2| \, .$$

Wählen wir als Basisvektoren $|j\rangle$ für $\mathcal{E}_+$ solche, bezüglich derer die Dichtematrix der Umwelt diagonal ist, d.h.

$$\varepsilon^- = \sum_{j=1}^{n} e_j^- \, |j\rangle \langle j| \,,$$

dann folgt

$$T(\rho) = \sum_{i_1,i_2=1}^{n_A} \sum_{j_1,j_2=1}^{n_\mathcal{E}} e_j^- \, \langle i_1| \otimes \langle j_1| \, V(\rho \otimes |j_2\rangle \langle j_2|)V^* \, |i_2\rangle \otimes |j_1\rangle \cdot |i_1\rangle \langle i_2|$$

$$T(\rho) = \sum_{j_1,j_2=1}^{n_\mathcal{E}} M_{j_1 j_2} \rho M_{j_1 j_2}^* \tag{11.35}$$

mit

$$M_{j_1 j_2} \in \mathcal{A}_+ \quad , \quad M_{j_1 j_2} = \sum_{i_1,i_3=1}^{n_A} \sqrt{e_{j_2}^-} \, \langle i_1| \otimes \langle j_1| \, V \, |i_3\rangle \otimes |j_2\rangle \cdot |i_1\rangle \langle i_3|$$

Insbesondere gibt es damit einen *minimalen Unterraum* $U_R \subseteq \mathcal{A}_N$ von Operatoren, sodass wir $T(\rho)$ für alle $T \in R$ in der Form (11.35) darstellen können.

Satz 11.8 *Es sei $R \subseteq B(\mathcal{A}_N)$ eine Menge vollständig positiver Transformationen. Dann sind die folgenden Bedingungen äquivalent:*

1. *Der Code $\mathcal{M} \subseteq \mathcal{N}$ korrigiert Fehler aus U_R.*
2. *$\mathcal{M}$ korrigiert Fehler aus R.*
3. $|\xi\rangle, |\eta\rangle \in \mathcal{M} \quad , \quad \langle \xi \mid \eta \rangle = 0 \quad , \quad X, Y \in U_R \Longrightarrow \langle X\xi \mid Y\eta \rangle = 0.$
4. $\mathbb{1}_\mathcal{M} Y^* X \mathbb{1}_\mathcal{M} \in \mathbb{C}\mathbb{1}_\mathcal{M} \qquad (X, Y \in U_R)$

Bemerkung 11.9 Der praktische Wert des Satzes besteht in der Feststellung, dass wir, um Fehler R zu korrigieren, nur darauf achten müssen, dass die Bedingung 3. erfüllt ist, dass also für orthogonale $|\xi\rangle, |\eta\rangle \in \mathcal{M}$ die Unterräume

$$\{|X\xi\rangle \mid X \in U_R\} \subseteq \mathcal{M}$$

und

$$\{|Y\eta\rangle \mid Y \in U_R\} \subseteq \mathcal{M}$$

immer noch orthogonal sind.

Bew.: Siehe Kitaev [Ki].

12 Symplektische Quantencodes

Es sei

- $\mathcal{H} := \mathbb{C}^2$ (der Hilbertraum eines Spins),
- $\mathcal{H}^n := \bigotimes_{i=1}^n \mathcal{H}$ (der Hilbertraum von n Spins),
- $\mathcal{A}^n := B(\mathcal{H}^n)$ die Observablenalgebra.

Weiter sei

$$P \subseteq \{1, \ldots, n\}$$

und

$$\mathcal{A}^n(P) := \mathrm{span}\left\{ \bigotimes_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{A_i} \subseteq \mathcal{A}^{\mathrm{n}} \mid \mathrm{A_i} \in \mathcal{A} \quad ; \quad \mathrm{i} \notin \mathrm{P} \Longrightarrow \mathrm{A_i} = \mathbb{1} \right\}$$

die Unteralgebra, bei der nur die Observablen für die durch P bezeichneten Spinpositionen von der Identität verschieden sind. Diese Observablenalgebra entspricht also Beobachtungen bei P.

Wir setzen

$$\varepsilon(n,k) := \sum_{\substack{P \subseteq \{1,\ldots,n\} \\ |P| \leq k}} \mathcal{A}^n(P) \subseteq \mathcal{A}^n.$$

Dies entspricht einem Fehlermodell von höchstens k Fehlern und wir müssen nach dem Satz über die Quantencodes nachprüfen, ob für beliebige $X, Y \in \varepsilon(n,k)$

$$\mathbb{1}_{\mathcal{M}} Y^* X \mathbb{1}_{\mathcal{M}} \in \mathbb{C}\mathbb{1}_{\mathcal{M}}$$

gilt.

Um einen guten Quantencode, also einen nicht zu kleinen Unterraum $\mathcal{M} \subseteq \mathcal{N}$, zu finden, der diese Bedingung erfüllt, wollen wir $\mathcal{M}$ als gemeinsamen Unterraum zum Eigenwert 1 zu einer Familie kommutierender Observablen $X_j \in \mathcal{A}^n$ beschreiben.

Hier ist es zunächst sinnvoll, eine Basis von $\mathcal{A}^n$ zu wählen, und zwar solche, die bez. des Skalarpoduktes auf $\mathcal{A}^n$,

$$\langle A \,|\, B \rangle := 2^{-n} \mathrm{tr}_{H^n}(A^* B) \qquad (A, B \in \mathcal{A}^n)$$

orthonormal ist. Für $n = 1$ kennen wir schon eine solche Basis. Sie wird gebildet von den Matrizen

$$\mathbb{1} \quad , \quad \sigma_x = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad , \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \text{und} \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Um die zwischen den vier Matrizen bestehenden algebraischen Relationen besser notieren zu können, setzen wir

$$\sigma_{00} := \mathbb{1} \quad , \quad \sigma_{10} := \sigma_x \quad , \quad \sigma_{01} := \sigma_z \quad \text{und} \quad \sigma_{11} := \sigma_y.$$

Mit $G^n := \{0,1\}^{2n} \equiv (\mathbb{F}_2)^n \times (\mathbb{F}_2)^n$ können wir dann beliebige Produkte von n solcher Matrizen in der Form

$$\sigma_a := \bigotimes_{i=1}^{n} \sigma_{a_i^{(1)} a_i^{(2)}} \quad , \quad a = (a_1^{(1)}, \ldots, a_1^{(n)}, a_2^{(2)}, \ldots, a_n^{(2)}) \in G^n$$

schreiben.

Es gilt für alle $a, b \in G^n$

1. $\mathrm{tr}(\sigma_a) = 2^n \delta_{a,0}$
2. $\sigma_a^* = \sigma_a$
3. $\sigma_a^{-1} = \sigma_a$
4. $\sigma_a \sigma_b = (-i)^k \sigma_{a+b}$, mit

$$k = \sum_{i=1}^{n} f\left(a_i^{(1)}, b_i^{(2)}\right) - f\left(a_i^{(2)}, b_i^{(1)}\right),$$

wobei

$$f(i,j) := \begin{cases} 1 & i = j = 1 \\ 0 & \text{sonst.} \end{cases}$$

Aus 1. , 2. und 4. folgern wir die Orthonormalität 5.

$$\langle \sigma_a \,|\, \sigma_b \rangle = \delta_{a,b},$$

denn

$$\langle \sigma_a \,|\, \sigma_b \rangle = 2^{-n} \mathrm{tr}(\sigma_a^* \sigma_b) = 2^{-n} (-i)^k \mathrm{tr}(\sigma_{a+b}).$$

Aus 4. folgern wir die (Anti-)Kommutativität

$$\sigma_a \sigma_b = (-1)^{\omega(a,b)} \sigma_b \sigma_a$$

der Basiselemente, mit

$$\omega(a,b) := a^{(1)} \cdot b^{(2)} - a^{(2)} \cdot b^{(1)} \in \mathbb{F}_2.$$

Damit ist ω eine symplektische Form auf dem $2n$–dimensionalen Vektorraum $G^n = \mathbb{F}_2^{2n}$ über dem 2–elementigen Körper $\mathbb{F}_2$, d.h. eine nichtdegenerierte antisymmetrische Bilinearform.

Weil in $\mathbb{F}_2 \;\; -a = a$ gilt, ist ω gleichzeitig symmetrisch.

In der algebraischen Theorie symplektischer Vektorräume werden folgende Bezeichnungen benutzt:

Definition 12.1 Es sei $F \subseteq E$ Unterraum des symplektischen Vektorraums (E, ω). Dann heißt

- $F^\perp := \{e \in E \mid \omega(e,f) = 0 \quad \text{für alle} \quad f \in F\}$ *ω–orthogonales Komplement* von F.
- F heißt *isotrop*, wenn $F \subseteq F^\perp$.
- F heißt *Lagrangesch*, wenn $F = F^\perp$.

Isotrope Unterräume haben höchstens die halbe Dimension von E und Langrange-Unterräume besitzen genau diese Dimension, denn es gilt

$$\dim(F^\perp) = \dim(E) - \dim(F).$$

Beispielsweise ist in unserem Fall $E = G^n$ der Raum $F' := \{(0, a^{(2)}) \mid a^{(2)} \in \mathbb{F}_2^n\}$ ein Lagrange-Unterraum.

- Wählen wir nun irgendeinen isotropen Unterraum $F \subseteq G^n$ aus, dann kommutieren alle Operatoren σ_a, $a \in F$ miteinander.
- Andererseits erzeugen *beliebige* Unterräume $F \subseteq G^n$ Unteralgebren $\hat{\sigma}(F) := \mathrm{span}\{\sigma_a \mid a \in F\} \subseteq \mathcal{A}^n$. Dies gilt wegen Rechenregel 4.
- Wegen 2. und 3. lässt sich für $a \in G^n$ der Operator σ_a in der Spektraldarstellung

  $$\sigma_a = P_1(a) - P_{-1}(a)$$

  schreiben, wobei die Projektionen $P_{\pm 1}(a)$ auf die Unterräume zum Eigenwert ± 1 abbilden.

  Wir können $P_{\pm 1}(a) = \frac{1}{2}(\mathbb{1} \pm \sigma_a)$ schreiben.

Daher kommutieren die $P_{\pm 1}(a)$ und $P_{\pm 1}(b)$ genau dann, wenn $\omega(a,b) = 0$. Ist $a(1), \ldots, a(k) \in G^n$ eine Basis eines isotropen Unterraums

$$F = \mathrm{span}(a(1), \ldots, a(k))$$

der Dimension k (also $k \leq n$), dann können wir zunächst den Bits $\varepsilon_1, \ldots, \varepsilon_k \in \{\pm 1\}$ denjenigen Unterraum zuordnen, auf den $\prod_{i=1}^k P_{\varepsilon_i}(a(i))$ projeziert.

Es gilt

$$\begin{aligned}
&\mathrm{tr}\left(\prod_{i=1}^k P_{\varepsilon_i}(a(i))\right) \\
&= 2^{-k}\mathrm{tr}\left(\prod_{i=1}^k (\mathbb{1} + \varepsilon_i \sigma(a(i)))\right) \\
&= 2^{-k}\mathrm{tr}\left(\sum_{U \subseteq \{1,\ldots,k\}} \left(\prod_{i \in U} \varepsilon_i\right) \sigma_{\left(\sum_{i \in U} a(i)\right)}\right) \\
&= 2^{-k}\mathrm{tr}(\mathbb{1}) = 2^{n-k},
\end{aligned}$$

denn nur $U = \emptyset$ liefert wegen der Basiseigenschaft eine von Null verschiedene Summe.

Wegen der (Anti-) Kommutationsrelation

$$\sigma_a \sigma_b = (-1)^{\omega(a,b)} \sigma_b \sigma_a$$

gilt

$$\sigma_a \mathcal{M} = \mathcal{L}_{\mu(a)} \qquad (a \in G^n) \tag{12.36}$$

mit

$$\mu(a) = \omega(a, \cdot) \in F^*.$$

Insbesondere vertauschen die Fehler $a \in E \subseteq G^n$ die Unterräume.

Weiter mit Notation $\sigma(g) \equiv \sigma_g$

Satz 12.2 *Der symplektische Code $\mathcal{M} \subseteq \mathcal{H}^n$ korrigiert genau dann die Fehler aus $\hat{\sigma}(E)$ $(E \subseteq G^n)$, wenn für die Menge $E - E := \{g' - g'' \mid g', g'' \in E\}$*

$$(E - E) \cap F^\perp \subseteq F$$

gilt.

Bew.: Wir wissen schon, dass eine notwendige und hinreichende Bedingung für die Fehlerkorrektur ist, dass für $X, Y \in \hat{\sigma}(E)$

$$\mathbb{1}_{\mathcal{M}} X^* Y \mathbb{1}_{\mathcal{M}} \in \mathbb{C}\mathbb{1}_{\mathcal{M}} \tag{12.37}$$

ist.

Dies lässt sich für X, Y in der Basis $\{\sigma(g) \mid g \in E\}$ überprüfen. Es ist aber

$$\sigma(g'')^* \sigma(g') = i^k \sigma(g' - g'').$$

Ist $g := g' - g'' \notin F^\perp$, dann ist nach (12.36)

$$\mathbb{1}_{\mathcal{M}} \sigma(g'')^* \sigma(g') \mathbb{1}_{\mathcal{M}} = 0,$$

und die Bedingung erfüllt.

Wir haben also gesehen, dass die Auswahl eines isotropen Unterraumes $F \subseteq G^n$ und des Strings $(\varepsilon_1, \ldots, \varepsilon_n) \in \{-1, 1\}^k$ zur Basis von F einen Informationsraum

$$\mathcal{M} := \prod_{i=1}^{k} P_{\varepsilon_i}(a(i)) \mathcal{H}^n$$

der Dimension 2^{n-k} fixiert. Die anderen Eigenwertkombinationen liefern orthogonale Teilräume gleicher Dimensionen, also insgesamt 2^k Stück.

Für $a \in F$ mit der Darstellung $a = \sum_{i=1}^{k} \lambda_i a(i)$ $(\lambda_i \in \mathbb{F}_2)$ ist

$$\sigma_a \mathbb{1}_{\mathcal{M}} = \prod_{i=1}^{k} \varepsilon_i^{\lambda_i} \mathbb{1}_{\mathcal{M}}, \tag{12.38}$$

sodass mit $X(a) := \left(\pi_{i=1}^{k} \varepsilon_i^{\lambda_i}\right) \sigma_a$ gilt:

$$X(a)\mathbb{1}_{\mathcal{M}} = \mathbb{1}_{\mathcal{M}}.$$

Setzen wir für $h \in F^* \equiv L(F, \mathbb{F}_2)$

$$\mathcal{L}_h := \{\xi \in \mathcal{H}^n \mid X(a)\xi = (-1)^{h(a)}\xi \qquad (a \in F)\},$$

dann ist

$$\mathcal{L}_0 = \mathcal{M}$$

und

$$\mathcal{H}^n = \bigoplus_{h \in F^*} \mathcal{L}_h.$$

Ist andererseits $g \in F^{\perp}$, dann ist $\sigma(g)\xi \in \mathcal{M}$, falls $\xi \in \mathcal{M}$.

Damit nun (12.37) erfüllt ist, ist notwendig und hinreichend, dass $g \in F$. Hinreichend wegen (12.38), notwendig, weil für $g \in F^{\perp} \backslash F$ die Projektionen $P \pm (g)$ von $\sigma(g)\mathcal{M}$ in zwei Unterräume halber Dimension zerlegen würden. □

Beispiel 12.3 1. **Shor's 9 qubit-Code:** $F = \mathrm{span}(\mathrm{a}(1), \ldots, \mathrm{a}(8))$ mit

$$\begin{aligned}
a(i)^{(2)} &= 0 \qquad (i = 1, \ldots, 6), \\
a(i)^{(1)} &= 0 \qquad (i = 7, 8) \quad \text{und} \\
a(1)^{(1)} &:= (110\ 000\ 000) \\
a(2)^{(1)} &:= (011\ 000\ 000) \\
a(3)^{(1)} &:= (000\ 110\ 000) \\
a(4)^{(1)} &:= (000\ 011\ 000) \\
a(5)^{(1)} &:= (000\ 000\ 110) \\
a(6)^{(1)} &:= (000\ 000\ 011) \\
a(7)^{(2)} &= (111\ 111\ 000) \\
a(8)^{(2)} &= (000\ 111\ 111).
\end{aligned}$$

F ist isotrop und $\dim(F) = 8$, also $\dim(F^\perp) = \dim(G^9) - \dim(F) = 10$.

$F^\perp$ wird von F und den Vektoren

$$\begin{aligned} a(9) &:= (111\ 111\ 111\ ,\ 000\ 000\ 000) \\ a(10) &:= (000\ 000\ 000\ ,\ 111\ 111\ 111) \end{aligned}$$

aufgespannt.

Ist $a \in F^\perp \backslash F$, dann ist $|\mathrm{supp}(a)| \geq 3$, sodass der Shor-Code einen beliebigen Fehler korrigiert. Wegen $2^{9-\dim F} = 2$ wird pro Codewort ein qubit übertragen.

2. $LMPZ$**– Code (Laflamme, Miguel, Paz, Zurek)**

   Hier enthalten die Codewörter fünf qubits und der Informationsraum ist durch

   $$F = \mathrm{span}(a(1), a(2), a(3), a(4))$$

   mit

   $$\begin{aligned} a(1) &:= (11000\ ,\ 00110) \\ a(2) &:= (10100\ ,\ 01001) \\ a(3) &:= (00010\ ,\ 10101) \\ a(4) &:= (00001\ ,\ 11010) \end{aligned}$$

   fixiert. F ist isotrop und $\dim(F) = 4$, also $\dim(F^\perp) = \dim(G^5) - \dim(F) = 6$.

   $F^\perp$ wird von F und den Vektoren

   $$\begin{aligned} a(5) &:= (11111\ ,\ 11111) \\ a(6) &:= (10000\ ,\ 00011) \end{aligned}$$

   aufgespannt.

   Es gilt $(E - E) \cap F^\perp = \{0\}$ für $E = E(5, 1)$.

   Da der Informationsraum 2–dimensional ist, wird pro Codewort ein qubit übermittelt und es wird ein beliebiger Fehler korrigiert.

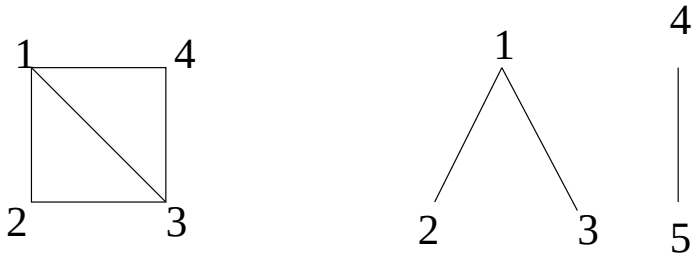


Abbildung 18: Zwei Graphen, die keine Bäume sind

A Elementare Begriffe der Graphentheorie

Definition A.1 • Ein *Graph* (V, E) besteht aus einer (endlichen) Menge V von *Ecken* (oder *Knoten*) und einer Menge E von Paaren $\{u, v\} \subseteq V, u \neq v$, genannt *Kanten*.
• $w \in V$ mit $\{w, v\} \in E$ heißt *Nachbar* von $v \in V$.
• Ein *Weg* (der Länge $n-1$) ist eine Folge $P = (v_1, \ldots, v_n)$ voneinander verschiedener Ecken $v_i \in V$ mit $\{v_1, v_{i+1}\} \in E$.
• Ein *Kreis* ist eine Folge $C = (v_1, \ldots, v_n)$ voneinander verschiedener Ecken $v_i \in V$ mit $\{v_i, v_{i+1}\} \in E$ und $\{v_n, v_1\} \in E$.
• (V, E) heißt *zusammenhängend*, wenn es für $v, w \in V$ einen Weg mit $v_1 = v$, $v_n = w$ gibt.
• Der Graph (V, E) heißt *Baum*, wenn er zusammenhängend ist, aber keinen Kreis besitzt.

Beispiel A.2

1. $V := \{1, \ldots, 4\}$, $E := \{\{1,2\}, \{2,3\}, \{3,4\}, \{4,1\}, \{1,3\}\}$.
   (V, E) ist zusammenhängend, $(1, 3, 4)$ ist ein Kreis.

2. $V := \{1, 2, 3, 4, 5\}$, $E := \{\{1,2\}, \{1,3\}, \{4,5\}\}$.
   (V, E) ist nicht zusammenhängend und enthält keinen Kreis.

Bemerkung A.3 In einem Baum $T = (V, E)$ gibt es zu $v, w \in V$ genau einen Weg mit $v_1 = v$, $v_n = w$.

Definition A.4 • Ein *Wurzelbaum* (T, v^*) ist ein Baum $T = (V, E)$ mit einem ausgezeichneten Element $v^* \in V$, der *Wurzel*.

• Die *Länge* $l(v)$ der Ecke $v \in V$ ist die Länge des Wegs von v^* nach v.
• Besitzt $v \in V$ keine *unmittelbaren Nachfolger*, d.h. Nachbarn w mit $l(w) > l(v)$, dann heißt V *Blatt* von T. Die anderen Ecken heißen *innere Ecken*.
• $L(T) := \max_{v \in V} l(v)$ heißt *Länge von* T.

Literatur

[Ai] Martin Aigner: Combinatorial Search. Wiley, Teubner 1988

[Ar] Vladimir I. Arnol'd: Mathematical Methods of Classical Mechanics. Springer GTM 60

[BG] Friedrich L. Bauer, Gerhard Goos: Informatik 1. Teil. Berlin: Springer 1971

[Be] Gennady Berman, Fary Doolen, Ronnie Mainieri, Vladimir Tsifrinovich: Introduction to Quantum Computers. World Scientific, 1998

[BEZ] Dirk Bouwmeester, Artur Ekert, Anton Zeilinger (Eds): The Physics of Quantum Information. Quantum Cryptography, Quantum Teleportation, Quantum Computation. Springer 2001

[BR] Ola Bratteli, Derek Robinson: Operator Algebras and Quantum Statistical Mechanics 1. Springer 1987

[CLR] Thomas Cormen, Charles Leiserson, Ronald Rivest: Introduction to Algorithms. M.I.T. Press 1997

[CT] Thomas Cover, Joy Thomas: Elements of Information Theory. New York: Wiley 1991

[Da] Kenneth Davidson: C^*–Algebras by Example. American Mathematical Society 1996

[Gr] Jozef Gruska: Quantum Computing. London: McGraw Hill 1999

[FT] Edward Fredkin, Tommaso Toffoli: Conservative logic. Int. J. Theor. Physics **21**, 467–488 (1982)

[Hi] Mika Hirvensalo: Quantum Computing. Springer 2001

[HU] John E. Hopcroft, Jeffrey D. Ullman: Einführung in die Automatentheorie, formale Sprachen und Komplexitätstheorie. Bonn: Addison Wesley 1988

[IKO] Roman S. Ingarden, Andrzej Kossakowski, Masanori Ohya: Information Dynamics and Open Systems. Classical and Quantum Approach. Dordrecht: Kluwer 1997

[Ja] Konrad Jacobs: Discrete Stochastics. Birkhäuser 1992

[Ki] Alexey Yu. Kitaev: Quantum Computations: Algorithms and Error Correction. Russian Math. Surveys **52**, 1191-1249 (1997)

[Ko] Dexter A. Kozen: Automata and Computability. Springer 1997

[NC] Michael A. Nielsen, Isaac L. Chuang: Quantum Computation and Quantum Information. Cambridge University Press 2001

[Pr] John Preskill: Quantum Information and Computation. Vorlesung, *Erhältlich unter* `http://theory.caltech.edu/~preskill/ph229/`

[RS] M. Reed, B. Simon: Methods of Modern Mathematical Physics, Vol. I: Functional Analysis. Academic Press 1980

[RS2] M. Reed, B. Simon: Methods of Modern Mathematical Physics, Vol. II: Fourier Analysis, Self–Adjointness. Academic Press 1975

[Sh] Peter W. Shor: Quantum Computing. In: Proc. of the International Congress of Mathematicians, Berlin 1998, Vol. I. Documenta Mathematica 1998. *Erhältlich unter* http://www.mathematik.uni-bielefeld.de/documenta/xvol-icm/00/00.html

[Sh1] Peter W. Shor: Fault-tolerant quantum computations. In: Proc. 37nd Annual Symposium of Foundations of Computer Science, IEEE Computer Society Press, 55-65 (1996)

[Sh2] Peter W. Shor: Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. SIAM Rev. **41**, 303–332 (1999)

[St] W. Forrest Stinespring: Positive Functions on C^*–Algebras. Proc. Amer. Math. Soc. **6**, 211–216 (1955)

[Ta] M. Takesaki: Theory of Operator Algebras, Bd. 1. Springer 1979

[Th] Walter Thirring: Lehrbuch der Mathematischen Physik, Band 1-4. Wien: Springer 1980

[WC] Colin P. Williams, Scott H. Clearwater: Explorations in quantum computing. TELOS, 1998

[Wo] Jacob Wolfowitz: Coding Theorems of Information Theory. Springer 1961

Index