

Reconstructie van veeldeeltjes spintoestanden op basis van de (Bayesiaanse) meest aannemelijke schatter toegepast op tensornetwerken

Ivan-Jago Coric

Proefschrift ingediend tot het behalen van de graad van
Master in de Fysica

Promotor: Prof. dr. Jacques Tempere



Departement Fysica
Universiteit Antwerpen
België
2021-2022

Dankwoord

Deze thesis was een uitdagend avontuur dat nog eens bemoeilijkt werd door twee jaar corona. Het ging niet altijd even vlot maar zoals Professor Piet Van Espen ooit zei, —in een totaal andere context—, de aanhouder wint. Ik ben dankbaar voor de begeleiding onder Professor Jacques Tempere, hoewel ik achteraf gezien misschien te weinig gebruik gemaakt heb van de besprekingen. Hij stond altijd klaar om te helpen en om de nodige feedback te geven. Voor de rest ben ik nog dankbaar aan mijn neef Davor Josipovic wiens feedback op het hoofdstuk parameterschatten en modelinferentie bijzonder hulpvol was. Tot slot nog mijn dankbaarheid voor mijn familie en vrienden.

Samenvatting

De laatste decennia is de interesse naar veeldeeltjes kwantumtoestanden enorm toegenomen. Drijfkrachten achter deze interesse is het onderzoek naar complexe kwantumsystemen zoals supergeleiders, superfluïda, kwantumgassen, etc. en de nieuwe vooruitgang in de kwantuminformatica. Een centraal probleem blijft echter het efficiënt beschrijven van veeldeeltjes kwantumtoestanden. Uit de wiskundige structuur van kwantummechanica volgt het fenomeen van verstrengeling dat het onmogelijk maakt om samengestelde verstrengelde systemen te beschrijven in termen van hun deelsystemen. Bijgevolg schaalt de Hilbertruimte exponentieel met het aantal deeltjes.

Kwantumtomografie is de wetenschap om op basis van metingen van kwantumtoestanden de oorspronkelijke toestand zo goed mogelijk te reconstrueren. Deze taak is echter allesbehalve evident wegens de aard van de kwantummechanica. Toch zijn er nieuwe formalismen ontworpen, zogenaamde tensornetwerken, die in staat zijn om specifieke delen van de Hilbertruimte efficiënt te beschrijven. Concreet worden in deze thesis de zogenaamde matrixproduct-toestanden en daarvan afgeleide tensornetwerken onderzocht als mogelijke oplossing voor het tomografieprobleem. De theorie van de tensornetwerken wordt aangevuld met de theorie van het parameterschatten op basis van de (Bayesiaanse) meest aannemelijke schatter methode en met de theorie van modelinferentie via het zogenaamde Akaike informatie criterium.

Tot slot wordt de theorie omgezet in praktijk met een numerieke implementatie. Er is code geïmplementeerd om spin metingen te genereren in verschillende richtingen om dan op basis van die metingen de oorspronkelijke toestand te reconstrueren. Resultaten tonen aan dat de methode werkt als een proof of concept maar dat ook er nog veel plaats is voor verbetering.

Summary

In the last decades research interest into many body quantum states has increased significantly. Driving forces behind the interest are advancements in the engineering of large quantum mechanical complex systems for the study of quantum phases of matter such as superconductors, superfluids and quantum gasses. Additionally there have been significant advancements in the field of quantum computing which also relies on many particle states. However, the central issue with the many particle states is that due to the mathematical structure of the quantum mechanics it is difficult to describe them efficiently. For example, quantum entanglement of a composite state forbids describing the state in terms of its constituents. This results in the Hilbert space scaling exponentially with the number of particles.

Quantum tomography is the science of reconstructing the original wave function based of a set of measurements performed on that state. This is a challenging task again due to the nature of quantum mechanics. Yet, several new formalisms have been developed that allow for an efficient description of a sub space of the Hilbert space. An example of such a formalism is that of the matrix product states which allow for an efficient description of low energy one dimensional systems. In this work we will study those matrix product states and derivatives of them. The formalism will be further enhanced by the introduction of the theory of parameter estimation based on (Bayesian) maximum likelihood methods. Additionally the theory of model inference based on the Akaike information criterion will be introduced which will help us further optimize the matrix product states.

Finally all the theory will be put to practice in an attempt to find a parsimonious description of a quantum state. Code is implemented that performs measurements on quantum states in multiple directions which are then used in an quantum tomography exercise applied on matrix product states. Results show that the method works although there is still a lot of room for improvement. Nonetheless it counts as a proof of concept.

Inhoudsopgave

1. Inleiding	7
1.1. Basisconcepten en notatieconventies	8
1.1.1. De kwantumtoestand	8
1.1.2. Kwantumtoestanden manipuleren	9
1.1.3. Gemengde toestanden	10
1.2. De experimentele kwantummechanica	11
1.2.1. Kwantumcomputaties	12
1.3. Kwantumtomografie	16
1.3.1. De kwantummechanische aard	16
1.3.2. Probleembeschrijving en het doel van de thesis	17
2. Tensornetwerken in de kwantummechanica	19
2.1. Tensoren	19
2.2. Matrixproduct-toestanden	22
2.2.1. Schmidt-decompositie en verstrengeling	23
2.2.2. Matrixproduct-operatoren	25
2.3. Gerichtte verstrengeling matrixproduct-toestanden	27
2.3.1. Tensornetwerk contracties	28
3. Parameterschatten en modelinferentie	30
3.1. Parameterschatten	31
3.1.1. Meest aannemelijke schatter	31
3.1.2. Bayesiaanse schatting	33
3.2. Modelinferentie	34
3.2.1. Kullback-Leibler divergentie	35
3.2.2. Het Akaike informatie criterium	35
3.3. Toepassing: klassieke munten	37
3.3.1. Betafunctie voor normalisatiefactoren	39
3.4. Uitbreiding naar de kwantummechanica en tensornetwerken	41
3.4.1. Het kwantummechanisch karakter: de ongelijkheid van Bell	41
3.4.2. Tensornetwerken en modelinferentie	44
4. Resultaten en bespreking	46
4.1. Klassieke munten	46
4.1.1. Implementatie	46
4.1.2. Resultaten	48

Inhoudsopgave

4.2. Tensornetwerken	49
4.2.1. Resultaten	51
5. Conclusie	57
5.1. Verbeteringen en vooruitzichten	57
A. Appendices	58
A.1. Verwachtingswaarde Bell-ongelijkheid	58

1. Inleiding

Ignorance is bliss.

Cypher
The Matrix (1999)

Ter inleiding van deze thesis en om de context van de probleemstelling te schetsen zullen we beginnen vanuit de experimentele kant van kwantummechanica, met name kwantumtomografie. Tomografie is een gekend begrip, voornamelijk in de medische context, waar het verwijst naar de reconstructie van een driedimensionaal beeld van het menselijk lichaam aan de hand van projecties in verschillende richtingen op basis van bv. Röntgenstraling metingen. In de kwantummechanische context verwijst het naar de reconstructie van een onbekende kwantummechanische toestand op basis van een aantal verschillende metingen van die toestand.

De kwantummechanische aard van het systeem introduceert echter nieuwe moeilijkheden. Een eerste moeilijkheid zit in het verrichten van metingen op kwantumsystemen. De meting verandert de kwantumtoestand op een irreversibele manier. Hier kunnen we enigszins aan ontkomen wanneer het mogelijk is om een ensemble aan systemen te maken die allemaal in dezelfde toestand geprepareerd zijn. Er is uiteraard ook nog de moeilijkheid dat de uitkomst van de metingen een onbepaaldheid, een willekeur vertonen die inherent is aan de kwantummechanica, dit bovenop de al aanwezige willekeur van statistische, klassieke ruis. Bovendien is het zo dat toestanden die niet orthogonaal zijn aan elkaar ook niet te onderscheiden zijn met een enkele meting. De situatie wordt nog eens verergerd door de aanwezigheid van verstrengeling tussen meerdere deeltjes, correlaties die niet klassiek van aard zijn, waardoor systemen van meerdere deeltjes non-lokaal gedrag kunnen vertonen. Een gevolg hiervan is dat het aantal parameters van het systeem exponentieel schaal met het aantal deeltjes en bijgevolg ook de hoeveelheid metingen die men moet verrichten. In de literatuur is dit gekend als de vloek van dimensionaliteit [8, 16].

De opkomst van kwantumcomputers en kwantumalgoritmes, alsook het toenemend onderzoek naar kwantummechanische veeldeeltjessystemen drijft echter de vraag naar methoden die onbekende veeldeeltjes-kwantumtoestanden accuraat kunnen reconstrueren maar die bovendien ook zuinig zijn met de tijd alsook de computationele kost. Deze “zuinigheid” (parsimony in het Engels) kan op verschillende manieren bereikt worden. Men kan bijvoorbeeld zich de vraag stellen of alle toestanden wel even relevant zijn voor

1. Inleiding

de fysica die men tracht te bestuderen. Zo kan het zijn dat de fysica in een systeem goed beschreven wordt met toestanden van een beperkte (verstengeling) complexiteit of zijn er aannames mogelijk die extra structuur opleggen op de toestanden.

De volgende secties introduceren de nodige basisbegrippen en geven een korte samenvatting van de hedendaagse capaciteiten in de experimentele kwantummechanica. De laatste sectie zal kort het centrale probleem nog eens schetsen en toelichten wat het doel van deze thesis is.

1.1. Basisconcepten en notatieconventies

Om de discussie vlot te laten verlopen en direct te linken met de literatuur is het goed om eerst de nodige terminologie en notatieconventies te introduceren. Een meer uitgebreide uitleg is te vinden in de inleidende hoofdstukken in [38, 58] of [51].

1.1.1. De kwantumtoestand

Een kwantumtoestand ψ is een éénheidsvector in de complexe Hilbertruimte $\mathcal{H}$, genoteerd in de gewoonlijke Dirac-notatie als $|\psi\rangle$ (ket). De complex toegevoegde of duale vector van $|\psi\rangle$ wordt gegeven door $\langle\psi| \in \mathcal{H}^*$. Kiest men een geordende basis $\{|i_1\rangle, \dots, |i_D\rangle\}$ in $\mathcal{H}$ dan is de geassocieerde duale basis gedefinieerd door de relatie $\langle j|i\rangle = \delta_i^j$. Eens een basis gekozen dan zijn de vectoren te ontbinden als

$$|\psi\rangle = \sum_i^D c_i |i\rangle = \begin{bmatrix} c_1 \\ \vdots \\ c_D \end{bmatrix}, c_i \in \mathbb{C} \quad \langle\psi| = \sum_i^D \bar{c}_i \langle i| = [\bar{c}_1 \quad \dots \quad \bar{c}_D], \bar{c}_i \in \mathbb{C}. \quad (1.1)$$

Merk op dat de relatie tussen $|\psi\rangle$ en $\langle\psi|$ bijectief is en bijgevolg hun Hilbertruimten isomorf zijn, $\mathcal{H}^* \cong \mathcal{H}$. Veel definities van objecten die inwerken op elementen uit $\mathcal{H}$ en $\mathcal{H}^*$ hebben dan een equivalente definitie waarin ze enkel inwerken op $\mathcal{H}$. Een veelgebruikte basis —de zogenaamde computationele basis— voor een twee-niveau systeem zoals een spin- $\frac{1}{2}$ deeltje, bestaat uit de eigenvectoren van de Pauli-Z operator waarbij we de volgende notatie¹ hanteren:

$$|\uparrow\rangle = |0\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad |\downarrow\rangle = |1\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \quad (1.2)$$

¹Dezelfde volgorde wordt ook doorgetrokken naar systemen van meerdere deeltjes

1. Inleiding

De algemene toestand van een spin- $\frac{1}{2}$ deeltje, of qubit², heeft in de computationele basis de volgende vorm:

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle. \quad (1.3)$$

De relatie tussen de duale basis en de gewone basis laat toe om het inproduct tussen kwantumtoestanden te definiëren. Dit is een afbeelding $\langle \cdot | \cdot \rangle : \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{C}$ gegeven door:

$$\forall |x\rangle, |y\rangle \in \mathcal{H} : \langle y|x\rangle = \sum_{ij} \bar{y}_j x_i \delta_i^j = [\bar{y}_1 \quad \dots \quad \bar{y}_D] \begin{bmatrix} x_1 \\ \vdots \\ x_D \end{bmatrix} \quad (1.4)$$

met als eigenschappen:

1. Toegevoegd symmetrisch: $\langle x|y\rangle = \overline{\langle y|x\rangle}$.
2. Lineair in het eerste argument: $\langle \alpha x + \beta y|z\rangle = \alpha \langle x|z\rangle + \beta \langle y|z\rangle$.
3. Positief definit: $\forall x \in \mathcal{H} \setminus \{0\} : \langle x|x\rangle > 0$.

Gegeven twee toestanden ψ en ϕ dan noemen we het getal $\langle \psi|\phi\rangle$ de waarschijnlijkheids-**amplitude** en het getal $\langle \psi|\phi\rangle \langle \phi|\psi\rangle = |\langle \psi|\phi\rangle|^2$ de waarschijnlijkheids**dichtheid**. Deze laatste zegt ons wat de kans is om toestand ψ te observeren in toestand ϕ . De norm van een toestandsvector $|\psi\rangle$ is gedefinieerd als $\| \psi \| = \| |\psi\rangle \| = \sqrt{\langle \psi|\psi\rangle}$.

1.1.2. Kwantumtoestanden manipuleren

Een manipulatie van kwantumtoestanden is een operatie die $|\psi\rangle$ laat evolueren naar $|\phi\rangle$. Een evolutie kan tijdsafhankelijk zijn waarbij de eindtoestand dan afhankelijk is van de gebruikte operatie en van de tijdsduur; deze zullen niet relevant zijn voor de thesis. Een evolutie kan ook tijdsafhankelijk zijn waarbij de eindtoestand enkel afhangt van de operatie zelf. Aangezien kwantumtoestanden vectoren zijn moeten de verschillende operaties lineaire afbeeldingen zijn. Eens een basis gekozen is dan kunnen deze lineaire afbeeldingen expliciet uitgedrukt worden als lineaire operatoren, *i.e.* matrices.

Er zijn twee belangrijke soorten manipulaties. Enerzijds zijn er de reversibele manipulaties beschreven door unitaire operatoren U waarvoor $U^\dagger U = \mathbb{1}$. Deze operatoren worden ook isometriën genoemd omdat ze de norm behouden. $U|\psi\rangle$ blijft ten alle tijden een goed gedefinieerde toestand.

Een belangrijke groep operatoren voor een spin- $\frac{1}{2}$ systeem zijn de Pauli-matrices:

²Quantum-bit

1. Inleiding

$$\mathbb{1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \quad Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \quad (1.5)$$

Anderzijds is er de groep van projectie-operaties. Dit zijn irreversibele operaties en worden beschreven door Hermitische projectie-operatoren Π waarvoor $\Pi^\dagger = \Pi$ en $\Pi^2 = \Pi$. Hermitische operatoren worden ook observabelen genoemd omdat ze overeenkomen met grootheden die gemeten kunnen worden en waarbij hun positieve eigenwaarden dan de mogelijke meetresultaten zijn. Een observabele M kan ontbonden worden in zijn projectoren zodat $M = \sum_i \lambda_i \Pi_i$ waarbij λ_i de eigenwaarden zijn en $\Pi_i = |i\rangle\langle i|$ is een projector naar de deelruimte opgespannen door eigenvector $|i\rangle$ [38].

Een projectieve meting van een observabele M op een toestand $|\psi\rangle$ heeft dan een kans $p(\lambda_i) = \langle\psi|M|\psi\rangle$ om λ_i te registreren. Ten gevolge van de meting is de oorspronkelijke toestand $|\psi\rangle$ geprojecteerd op een nieuwe toestand gegeven door

$$\frac{\Pi_i |\psi\rangle}{\sqrt{\langle\psi|\Pi_i|\psi\rangle}}. \quad (1.6)$$

Bovenstaand resultaat is de zogenaamde ineenstorting van de golffunctie volgens de Kopenhaagse interpretatie van kwantummechanica. Merk op dat een unitaire operator opgebouwd kan zijn uit meerdere projectie-operatoren maar niet tot dezelfde effecten leidt als een projectieve meting. Tot slot, voor een toestand $|\psi\rangle$ zijn de verwachtingswaarden van een observabele A gegeven door:

$$\langle A \rangle_\psi = \langle\psi|A|\psi\rangle \quad (1.7)$$

1.1.3. Gemengde toestanden

Indien we met zekerheid weten dat een kwantumsysteem in een toestand $|\psi\rangle$ geprepareerd is, dan spreken we over een pure toestand. Daarnaast is het mogelijk om een gemengde toestand te hebben waarin het systeem zich in een ensemble van mogelijke pure toestanden kan bevinden, m.a.w. men kan niet zeggen in welke kwantumtoestand het systeem precies geprepareerd werd. Om zulke toestanden te beschrijven is het nodig om het formalisme uit te breiden, met name van toestandsvectoren naar toestandsmatrices, beter bekend als dichtheidsmatrices. Deze worden geformuleerd als $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$ waarbij $|\psi_i\rangle$ de mogelijke pure toestanden van het systeem zijn en p_i de bijhorende waarschijnlijkheid dat het systeem in toestand $|\psi_i\rangle$ werd geprepareerd. Als men de toestanden ontbindt in een gekozen basis dan ziet men ook dat elementen van de dichtheidsmatrix buiten de hoofddiagonaal een gevolg zijn van superpositie toestanden. De operator $|\psi_i\rangle\langle\psi_i|$ is geconstrueerd als een Kroneckerproduct, op zijn beurt gedefinieerd als een afbeelding $|\cdot\rangle\langle\cdot| : \mathcal{H} \times \mathcal{H} \rightarrow \mathcal{H}^* \otimes \mathcal{H}$ gegeven door:

1. Inleiding

$$\begin{aligned}
\forall |x\rangle, |y\rangle \in \mathcal{H} : |x\rangle\langle y| &= \sum_{ij} x_i \bar{y}_j |i\rangle\langle j| \\
&= \begin{bmatrix} x_1 \\ \vdots \\ x_D \end{bmatrix} \begin{bmatrix} \bar{y}_1 & \dots & \bar{y}_D \end{bmatrix} = \begin{bmatrix} x_1 \bar{y}_1 & \dots & x_1 \bar{y}_D \\ \vdots & \ddots & \vdots \\ x_D \bar{y}_1 & \dots & x_D \bar{y}_D \end{bmatrix} \quad (1.8)
\end{aligned}$$

De dichtheidsmatrices laten direct toe om het idee van ideale tomografie[4] te introduceren. Als men namelijk in staat zou zijn om metingen te verrichten op een ensemble van oneindig veel replica's van het systeem dan kan men de exacte kansen berekenen en zo de exacte dichtheidsmatrix, en daarmee dus de exacte toestand, reconstrueren.

In de realiteit is dit natuurlijk niet mogelijk wegens enerzijds de kwantummechanische willekeur en anderzijds de willekeur geïntroduceerd door statistische onzekerheden wegens ruis (defecten, etc.) en imperfecte preparatieprocedures. Het beste dat men dus kan reconstrueren is een dichtheidsmatrix $\sigma = \sum_i f_i |\psi_i\rangle\langle\psi_i|$ waarbij f_i de relatieve frequenties zijn bekomen uit metingen die trekkingen zijn van p_i .

De “kwaliteit” van een reconstructie kan men op een objectieve manier beoordelen door het invoeren van een afstandsmaat of een overlapmaat tussen ρ en σ . Een veelgebruikte keuze is de kwantumfideliteit[30, 38, 58], een afbeelding $F(\cdot, \cdot) : \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{R}$ gegeven door:

$$\forall \rho, \sigma \in \mathcal{H} : F(\rho, \sigma) = \left(\text{Tr} \left[\sqrt{\sqrt{\rho} \sigma \sqrt{\rho}} \right] \right)^2$$

Voor pure toestanden, *i.e.* wanneer $\rho = |\psi\rangle\langle\psi|$ en $\sigma = |\phi\rangle\langle\phi|$ dan reduceert de kwantumfideliteit zich tot de bekende vergelijking

$$F(|\psi\rangle, |\phi\rangle) = |\langle\psi|\phi\rangle|^2. \quad (1.9)$$

De fideliteit van twee toestanden is gelijk aan één indien $\psi = \phi$ en is gelijk aan nul indien ψ en ϕ orthogonaal zijn. Bovenstaande concepten zijn voorlopig genoeg voor verdere besprekingen. Dit volstaat ter inleiding. De volgende sectie zal kort moderne experimentele capaciteiten toelichten.

1.2. De experimentele kwantummechanica

Deze sectie zal kort de huidige landschap van de experimentele kwantummechanica bespreken waarbij de focus zal liggen op de recente ontwikkelingen rond Rydberg-atomen.

1. Inleiding

Zoals eerder al opgemerkt is de afgelopen decennia de belangstelling voor kwantummechanische veeldeeltjestoestanden toegenomen. De drijfkracht hierachter is te vinden in twee grote onderzoeksvelden. Enerzijds is er het onderzoek naar complexe veeldeeltjesystemen en kwantummechanische aggregatietoestanden waarbij de kwantummechanica op een macroscopische schaal significant wordt en leidt tot allerlei interessante fenomenen. Voorbeelden van zulke systemen zijn superfluida, supergeleiders en ultrakoude condensaten. Anderzijds zijn veeldeeltjestoestanden ook noodzakelijk voor veel kwantumcomputaties. Zo hebben kwantumcomputers veel qubits nodig niet enkel voor het uitvoeren van kwantumalgoritmes maar ook voor de implementatie foutencorrecties om te compenseren voor de effecten van ruis en andere externe invloeden. Grote optimalisatieproblemen die men hoopt op te lossen met kwantumcomputaties hebben evengoed veeldeeltjestoestanden nodig om erop afgebeeld te worden. De centrale uitdaging voor bovenstaande problemen bestaat uit twee delen: het creëren van veeldeeltjestoestanden en deze toestanden zinvol manipuleren. De volgende paragrafen zullen dieper ingaan op de verschillende implementaties voor kwantumcomputaties.

1.2.1. Kwantumcomputaties

Het principe van kwantumsimulatoren is eenvoudig: gegeven een moeilijk (kwantummechanisch) fysisch systeem of optimalisatieprobleem dat onderzocht moet worden, beeld het af op een kwantummechanisch systeem waarover men precieze controle heeft, *i.e.* dat programmeerbaar is. Beginnend van een eenvoudige begintoestand kan het systeem door middel van de juiste operaties evolueren naar de eindtoestand die een oplossing is[35]. Dit is gelijkaardig aan hoe vroeger analoge computers gebruikt werden voor ingewikkelde berekeningen.

Kwantumannealing

Een eerste voorbeeld van een mogelijke manier voor kwantumcomputaties is kwantumannealing. (Klassieke) Annealing is een optimalisatieprocedure waarbij het optimalisatiealgoritme een niet triviale afhankelijkheid heeft van een gesimuleerde temperatuur. De temperatuur zorgt ervoor dat de optimalisatieprocedure kan ontsnappen uit lokale minima waardoor de verkenning van de oplossingsruimte robuuster wordt. Door de temperatuur traag genoeg naar nul te laten zakken zal het algoritme convergeren naar de globale oplossing. Het idee is dat bij langdurige simulaties het algoritme zich vooral in het thermodynamisch equilibrium bevindt. Door de temperatuur traag te laten dalen zal het equilibrium en dus het algoritme mee verschuiven totdat het overeenkomt met de grondtoestand bij temperatuur nul.[11, 27].

Kwantumannealing is een gelijkaardige procedure waarbij bijvoorbeeld het probleem wordt voorgesteld door een rooster van spins en in plaats van temperatuur een magneetveld wordt gebruikt. De werking van kwantumannealing berust op de adiabatische

1. Inleiding

stelling die zegt dat een kwantumsysteem in zijn instantane eigentoestand zal blijven als de inwerkende perturbaties traag genoeg zijn en er een kloof is met de huidige eigenwaarde en de eerstvolgende. De clue is dan om het optimalisatieprobleem te verwerken in een Hamiltoniaan van de volgende vorm[35]:

$$H(t) = A(t) \sum_i X_i + B(t) \left[\sum_i (\Delta_i Z_i + \sum_{j \neq i} J_{ij}, Z_i Z_j) \right]$$

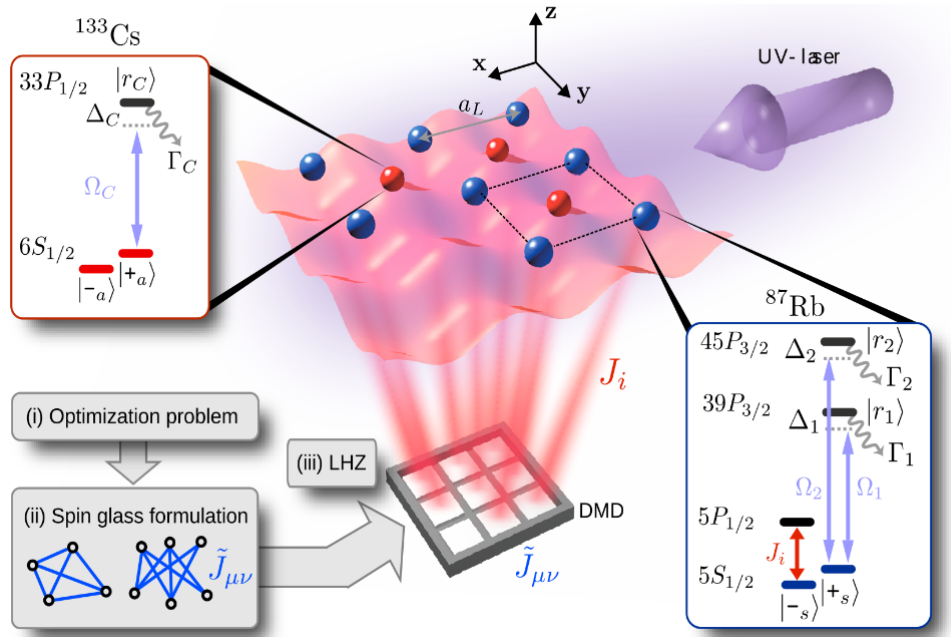
waarbij het optimalisatieprobleem verwerkt zit in de interacties Δ_i en J_{ij} en $B(t)$ bijvoorbeeld een magneetveld is dat als perturbatie dient. Door te beginnen in een “triviale” grondtoestand met $A(0) = 1$ en $B(0) = 0$ en vervolgens voorzichtig het systeem te laten gaan naar $A(T) = 0$ en $B(T) = 1$ zal de grondtoestand van het systeem mee evolueren. De oplossing van het optimalisatieprobleem kan dan achterhaald worden door een meting van de eindtoestand.

Figuur 1.1 illustreert een realisatie van een kwantumannealer op basis van Rydberg-atomen[18]. Dit zijn neutrale atomen waarbij de één of meerdere elektronen geëxciteerd zijn tot toestanden met een groot hoofdkwantumgetal ($n = 60 \sim 80$)[7, 35]. Deze atomen zitten vast in een optisch rooster bestaande uit potentiaalputten gevormd door een interferentiepatroon van intens laserlicht[21]. Door middel van een matrix van microscopische spiegels (digital micro-mirror device - DMD) die individueel instelbaar zijn is het mogelijk om de eigenschappen van het rooster te programmeren voor verschillende toepassingen; zelfs individuele sites zijn hiermee te programmeren. Fysische systemen kunnen vervolgens gesimuleerd worden door de nodige Hamiltoniaan in het rooster te programmeren. Concreet gebruiken ze het onderstaand rooster om spin-glas, tweedimensionale roosters van spins, systemen te simuleren. Deze optische roosters kunnen tegenwoordig bevolkt worden met één Rydberg-atoom per site en ook manipulaties van individuele atomen behoren tot de mogelijkheid[35].

Kwantumcomputers

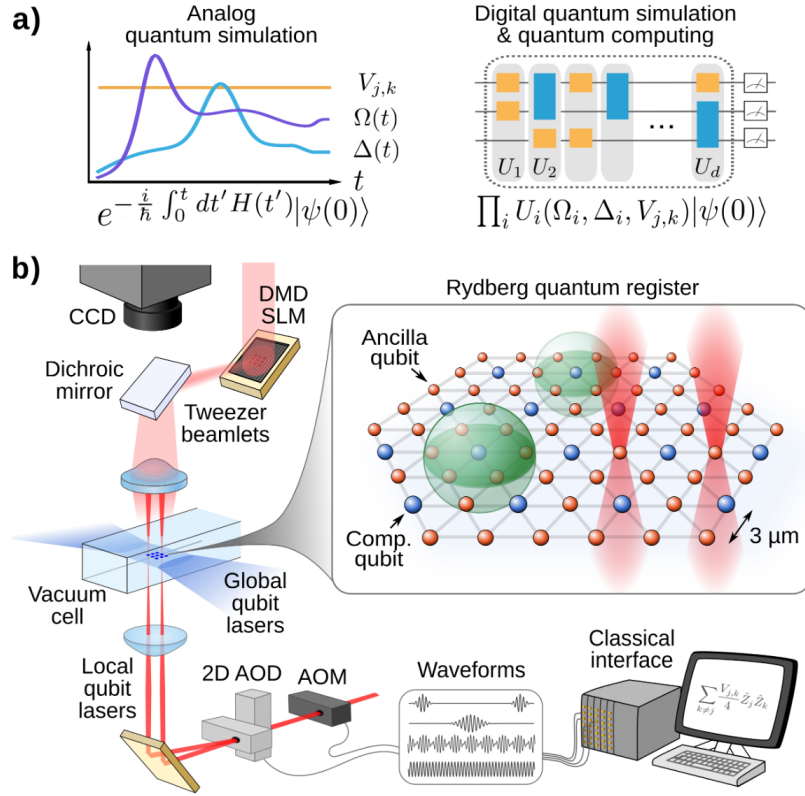
Kwantumcomputers verschillen van kwantumannealers in de zin dat ze algemener zijn. Ze kunnen ontworpen worden voor specifieke kwantumalgoritmes of als men er in slaagt om universele kwantumpoorten te creëren dan kunnen ze ook als universele computers gebruikt worden. Het hoofdbestanddeel van een kwantumcomputer is het qubitregister dat de nodige qubits bevat waarop verschillende operaties uitgevoerd kunnen worden. Figuur 1.2 toont opnieuw een voorbeeld van een kwantumcomputer implementatie op basis van Rydberg-atomen. Deze worden opnieuw vastgezet in een optisch rooster waar vervolgens de nodige operaties uitgevoerd kunnen worden op individuele atomen.

a)



Figuur 1.1.: Voorstelling van een mogelijke opstelling waarbij Rydberg-atomen gevangen zijn door laserlicht in een optisch rooster. De spiegelmatrix, DMD op de figuur, wordt gebruikt om de eigenschappen van het rooster te programmeren. In dit geval dient het om de nodige interacties te implementeren voor spin-glas simulaties van een optimalisatieprobleem. Afbeelding overgenomen uit [18], gepubliceerd onder de CC-BY 4.0 licentie.

1. Inleiding



Figuur 1.2.: Voorbeeld van hoe een kwantumcomputer geïmplementeerd kan worden op basis van Rydberg-atomen in een optisch rooster. De experimentator kan dan verschillende operaties uitvoeren op de atomen via een klassieke interface (computer) verbonden met lasers via akoestisch-optische modulatoren (AOM). Dit systeem kan gebruikt worden voor zowel analoge simulaties als voor digitale digitale (via kwantumpoorten) simulaties. Figuur overgenomen uit [35].

1.3. Kwantumtomografie

Bovenstaande toepassingen zijn veelbelovend maar wel enkel onder voorwaarde dat de geprepareerde kwantumtoestanden correct zijn en dat metingen op die toestanden juist geïnterpreteerd worden. Bij het uitlezen van een qubitregister worden de spins geprojecteerd naar eigenvectoren van de meetoperatoren en gaat er bijgevolg informatie over de kwantummechanische toestand verloren. De verloren kwantumtoestand reconstrueren op basis van de metingen is de taak van kwantumtomografie.

1.3.1. De kwantummechanische aard

Dit is een goed moment om te bespreken wat kwantummechanica juist zo anders maakt dan de klassieke fysica. Een van de eerste breekpunten met de klassieke fysica is dat zelfs perfect gekende kwantumtoestanden willekeurige uitkomsten geven tijdens metingen. In een zekere zin zijn kwantumtoestanden speciale kansverdelingen waar achter de schermen complexe getallen aan het werk zijn[51]. In de volgende secties zullen kort twee belangrijke kwantumfenomenen worden toegelicht. Er zijn er nog vele anderen maar die zijn niet even relevant voor de thesis.

Verstrengeling

Paragraaf 1.1 begint met de opmerking dat kwantumtoestanden éénheidsvectoren zijn. Dit is tweede breekpunt met de klassieke fysica waar toestanden van objecten punten zijn of zodanig behandeld kunnen worden. Het gevolg hiervan is dat de Hilbertruimte $\mathcal{H}$ die twee qubits beschrijft geconstrueerd wordt op basis van het tensorproduct tussen de twee deel-Hilbertruimten, *i.e.* $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$. Bijgevolg is de basis van $\mathcal{H}$ gegeven door $B = \{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$ zodat $\mathcal{H} = \text{span}\{B\}$. Eender welke genormeerde lineaire combinatie van de basisvectoren is een geldige kwantumtoestand. De ruimte opgespannen door B is echter groter dan eender welke ruimte opgespannen door $\text{span}\{|\psi\rangle, |\phi\rangle\}$ met $|\psi\rangle \in \mathcal{H}_1$ en $|\phi\rangle \in \mathcal{H}_2$. Met andere woorden, in $\mathcal{H}$ bestaan er toestanden die niet uitgedrukt kunnen worden als toestanden van individuele deeltjes. Die toestanden worden verstrengeld of niet separabel genoemd.

Aanvankelijk was verstrengeling een reden voor veel kopzorgen maar tegenwoordig zien wetenschappers het als de essentiële “grondstof” voor de verwachte efficiëntie van kwantumcomputers[38] en voor de uitzonderlijke gevoeligheid van kwantumsensoren[43]. Een pijnlijk gevolg van verstrengeling blijft echter dat juist omdat de toestanden niet separabel zijn, het aantal parameters in de theorie exponentieel zal groeien met het aantal deeltjes.

1. Inleiding

Het no-cloning theorema

Het no-cloning theorema zegt simpelweg dat het onmogelijk is om een arbitraire kwantumtoestand $|\psi\rangle$ te kopiëren naar een blanco toestand $|\emptyset\rangle$. Dit volgt eenvoudigweg uit de toepassing van het inproduct[38, 51]. Stel dat zo'n kopieermachine bestaat in de vorm van een unitaire operator C . Deze operator neemt twee toestanden $|\psi\rangle |\emptyset\rangle$ en laat ze evolueren tot $|\psi\rangle |\psi\rangle$. Veronderstel dat er twee toestanden gekopieerd worden, $|\psi\rangle$ en $|\phi\rangle$ zodat

$$C |\psi\emptyset\rangle = |\psi\psi\rangle \quad C |\phi\emptyset\rangle = |\phi\phi\rangle.$$

Neem nu het inproduct van deze twee resultaten:

$$\langle\psi\psi|\phi\phi\rangle = \langle\psi|\phi\rangle^2 \quad \langle\psi\psi|\phi\phi\rangle = \langle\psi\emptyset|C^\dagger C|\phi\emptyset\rangle = \langle\psi|\phi\rangle$$

Hieruit volgt dat $\langle\psi|\phi\rangle^2 = \langle\psi|\phi\rangle$ maar dit is enkel mogelijk indien $\langle\psi|\phi\rangle = 0$ of $\langle\psi|\phi\rangle = 1$, m.a.w. ofwel zijn de twee toestanden orthogonaal ofwel zijn ze identiek. Arbitraire toestanden kunnen dus niet gekopieerd worden. Dit heeft belangrijke gevolgen voor kwantumcommunicatie aangezien het toont dat kwantumtoestanden een inherente beveiliging hebben tegen af luisteraars. Dit theorema is bemoelijk ook de kwantumtomografie. Immers elke meting vernietigt de oorspronkelijke toestand; als er dan veel metingen nodig zijn voor de tomografieprocedure dan moet telkens dezelfde toestand steeds opnieuw geprepareerd worden aangezien het niet mogelijk is om a priori kopieën te maken.

1.3.2. Probleembeschrijving en het doel van de thesis

Het probleem kan als volgt samengevat worden: er is steeds toenemende interesse in veeldeeltjes-kwantumtoestanden maar het beschrijven of modelleren van die toestanden is allesbehalve triviaal. De wiskundige structuur van de kwantummechanica introduceert verstrengeling met als gevolg dat het aantal parameters dat een systeem beschrijft exponentieel schaalst met de grootte van het systeem. Bovendien is het ook zo, en dit zal later toegelicht worden, dat het belangrijk is om de juiste metingen te maken van een toestand. Zo is het noodzakelijk om metingen in verschillende richtingen te verrichten omdat er anders het gevaar bestaat dat het kwantummechanisch gedrag niet waargenomen wordt.

Een oplossing voor het schalingsprobleem bestaat in de vorm van tensornetwerken, alternatieve wiskundige formuleringen voor de complexe toestandsvector. Hoewel deze tensornetwerken juist complexer zijn qua beschrijving kan hun structuur toch uitermate geschikt zijn om specifieke kwantumtoestanden efficiënt te beschrijven. Men moet echter het idee van een exacte beschrijving loslaten. Dit is geen probleem omdat er afhankelijk van het bestudeerd systeem vaak goede argumenten zijn om te focussen op slechts een klein deel van de Hilbertruimte.

1. Inleiding

Het doel van de thesis bestaat dan uit twee delen. Het eerste aspect is de zoektocht naar een goed formalisme in de context van tensornetwerken om toestanden accuraat te beschrijven. Bovendien moet dit een zuinig formalisme zijn, *i.e.* idealiter zo weinig mogelijk parameters. Het tweede aspect bestaat er uit om methoden uit het parameterschatten en modelinferentie toe te passen om de bekomen modellen verder te optimaliseren.

Wegens het gevaar dat de computationele kost gemakkelijk groot kan worden, zeker voor een eerste implementatie, is het belangrijk om op te merken dat het hier vooral gaat om een illustratie van het basisprincipe, *i.e.* aantonen dat het ontwikkeld formalisme inderdaad kwantumtoestanden accuraat kan beschrijven en dat de methodes van modelinferentie hiertoe een acceptabele bijdrage leveren.

De thesis is als volgt gestructureerd: het tweede hoofdstuk introduceert tensornetwerken en meer precies matrixproduct-toestanden en bijhorende operatoren. Het derde hoofdstuk bespreekt de theorie van het parameterschatten en modelinferentie. Er wordt aandacht besteed aan de methode van de meest aannemelijke schatter en de gelijkaardige methode van de Bayesiaanse schattingen. Het vierde hoofdstuk bespreekt de implementatie van tensornetwerken en de relevante algoritmes, gevolgd door de bespreking van enkele numerieke resultaten. Het laatste hoofdstuk concludeert dit werk en bespreekt daarbij wat er al dan niet is gelukt en wat interessant kan zijn voor toekomstig onderzoek.

2. Tensornetwerken in de kwantummechanica

Om het probleem van de exponentiële schaling van de Hilbertruimte aan te pakken zijn wetenschappers sedert bijna twee decennia op zoek naar efficiëntere formalismen voor de kwantummechanica. Een van de veelbelovende formalismen is gebaseerd op tensornetwerken [8, 39, 40, 56] waarvan de zogenaamde matrixproduct-toestanden (MPT) een belangrijke deelgroep van vormen. Deze zijn in staat om een arbitraire veeldeeltjes-toestand exact te beschrijven maar dan zijn ze opnieuw gevoelig aan een exponentiële groei van de parameterruimte. Waar ze echter in uitblinken is in hun benadering van veeldeeltjes-grondtoestanden van Hamiltonianen voor ééndimensionale systemen met locale interacties waarbij slechts een polynomiale schaling van de parameterruimte volstaat om goede resultaten te krijgen.

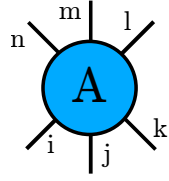
In de volgende secties worden de matrixproduct-toestanden geïntroduceerd alsook enkele belangrijke gerelateerde eigenschappen. Vervolgens wordt het formalisme van matrixproduct-toestanden uitgebreid naar meer complexe netwerken die gebruikt zullen worden voor het tomografieprobleem. Voor een uitgebreid uiteenzetting over tensornetwerken en hun toepassingen kan de lezer [8, 12, 39, 40, 56] raadplegen.

De ontwikkelde tensorcalculus gaat in de literatuur vaak gepaard met een handige diagrammatische voorstelling. Dit is een visualisatie van de verschillende tensor-operaties en is bijzonder handig om ingewikkelde vergelijkingen beknopt te communiceren. Deze thesis zal die traditie voortzetten.

2.1. Tensoren

Vooraleer matrixproduct-toestanden geïntroduceerd worden is het goed om even kort tensoren en enkele belangrijke operaties te bespreken. In de context van deze thesis volstaat het om een tensor te interpreteren als een meerdimensionale datacontainer, waarbij “de positie” van elk datum (een getal) geïndexeerd is. Wiskundig wordt deze indexatie genoteerd als een reeks boven- en onderindices. Bijvoorbeeld:

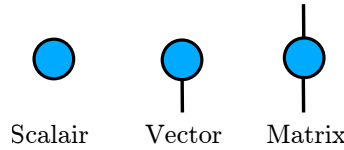
2. Tensornetwerken in de kwantummechanica



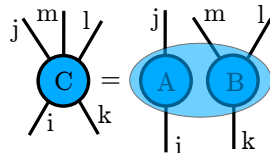
$$A_{ijk}^{lmn} \quad (2.1)$$

Het aantal “pootjes” geeft de rang van de tensor. Zo is bovenstaande tensor een tensor van rang (3,3) of kortweg een (3,3)-tensor. Het eerste getal verwijst naar het aantal onderindices en het tweede getal naar het aantal bovendindices. Elke index heeft ook een dimensie, de “bindingsdimensie”. Dit is het aantal mogelijke waarden dat de index kan aannemen. Bijvoorbeeld, als $k \in \{1, 2, 3, 4\}$, dan is de bindingsdimensie van deze index 4. Het aantal elementen in een tensor wordt dan gegeven door het product van alle dimensies van de indices. In de algemene tensorcalculus krijgen boven- en onderindices nog een betekenis van covariantie en contravariantie maar dat is voor deze thesis minder relevant. De gebruikte conventie is dat gewone toestanden (kets) genoteerd worden met onderindices en dus enkel vrije pootjes onderaan hebben; bovendien worden de indices gelabeld met de letter s . Voor de duale toestanden (bra’s) geldt de omgekeerde oriëntatie en dient de letter k als label. De vrije indices worden ook wel fysische indices genoemd en zij zijn toegankelijk voor de experimentator.

Enkele tensoren hebben een meer bekende naam: een tensor zonder indices is een scalar. Een tensor met maar één index is een vector en indien er twee indices zijn is het een matrix.



Er zijn twee elementaire operaties tussen tensoren. Het tensorproduct waarbij een (m,n)-tensor en een (k,l)-tensor een (n+k, m+l) tensor vormen.



$$C = A \otimes B = A_{i_1 \dots i_m}^{j_1 \dots j_n} B_{p_1 \dots p_k}^{q_1 \dots q_l} = C_{i_1 \dots i_m p_1 \dots p_k}^{j_1 \dots j_n q_1 \dots q_l} \quad (2.2)$$

De contractie volgt de Einsteinsommatieconventie wat wilt zeggen dat er impliciet gesommeerd wordt over alle herhaalde indices. Technisch gezien gaat het dan om indexparen van één bovenindex en één onderindex maar in de context van deze thesis wordt die conventie afgezwakt naar gewoon herhaalde indices.¹

¹Dit geeft geen problemen omdat de tensoren hier in feite ‘array’s zijn. Er is geen opgelegd gedrag onder coördinaattransformaties omdat deze objecten geen coördinaat-afhankelijke tensorvelden voorstellen.

2. Tensornetwerken in de kwantummechanica

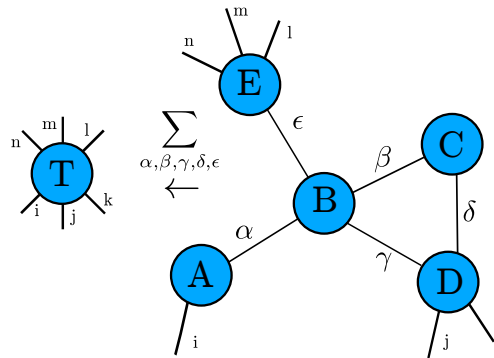
$$\begin{array}{c} \text{k} \\ | \\ \text{C} \\ | \\ \text{i} \quad | \\ \text{l} \end{array} = \begin{array}{c} \text{k} \\ | \\ \text{A} \\ | \\ \text{i} \end{array} \begin{array}{c} \text{j} \\ | \\ \text{B} \\ | \\ \text{l} \end{array} \quad C_{il}^k = \left(\sum_j \right) A_i^{kj} B_l^j \quad (2.3)$$

Nog een operatie die regelmatig wordt toegepast op tensoren is het combineren of splitsen van indices. Bijvoorbeeld een m-dimensionale index combineren met een n-dimensionale index tot een grotere (m×n)-dimensionale index.

$$\begin{array}{c} \text{i} \\ | \\ \text{A} \\ | \\ \text{j} \end{array} \Rightarrow \begin{array}{c} \text{A} \\ \text{---} \text{I} \end{array} \quad A_i^j \Rightarrow A_I \text{ met } I_k \in \{i_1, \dots, i_m, j_1, \dots, j_n\} \quad (2.4)$$

Zo is het Kroneckerproduct tussen twee vectoren uit $\mathbb{C}^2$ een vector in $\mathbb{C}^4$ wat neer komt op een tensorproduct tussen de twee vectoren gevolgd door het combineren van de twee twee-dimensionale indices tot een enkele vier-dimensionale index. Bijvoorbeeld voor een basistoestand van twee deeltjes: $|0\rangle_A \otimes |1\rangle_B \in \mathbb{C}^{2 \times 2} \Rightarrow |01\rangle_{AB} \in \mathbb{C}^4$ [8]. Merk op dat alle andere bewerking zoals matrixvermenigvuldiging, het inproduct, etc. geschreven kunnen worden in termen van een tensorproduct gevolgd door de juiste contracties.

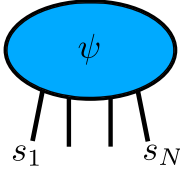
Tot slot nog een woordje uitleg over tensornetwerken. Tensornetwerken zijn simpelweg verschillende tensoren die met elkaar verbonden worden door contracties op een specifieke manier. Belangrijk om te spreken over een netwerk is dat die contracties (nog) niet uitgevoerd worden. Eens de contracties uitgevoerd zijn is de netwerkstructuur verloren en blijft er enkel nog een tensor over. Neem bijvoorbeeld het netwerk weergegeven in paragraaf 2.1. Het object “tensornetwerk” is in feite gewoon het rechterdiagram of in componentnotatie $T_{ijnml} = A_{i\alpha} B_{\alpha,\beta,\gamma,\epsilon} C_{\beta,\delta} D_{jk\gamma} E_{nml\epsilon}$. Tot slot, het aantal parameters in een tensornetwerk is gegeven door $M = \sum_i m_i$ met $m_i = \prod_j d_j^i$ het aantal parameters in de i -de tensor en d_j de dimensie van de j -de index.



Figuur 2.1.: Het verschil tussen een tensornetwerk (rechts) en de tensor (links) tot welke het netwerk zich reduceert na uitvoering van de contracties over de Griekse indices.

2.2. Matrixproduct-toestanden

Nu de nodige basisbewerkingen zijn geïntroduceerd kan de discussie verschuiven naar een eerste toepassing in de kwantummechanica: de matrixproduct-toestanden. We beschouwen een systeem dat bestaat uit N deeltjes. Een ééndeeltjesbasis voor deeltje j is $|s_j\rangle$. Hiermee kunnen we een productbasis opstellen voor de veeldeeltjestoestanden. Een arbitraire N -deeltjestoestand heeft dan de volgende vorm:



$$|\psi\rangle = \sum_{s_1, \dots, s_N} C_{s_1 \dots s_N} |s_1 \dots s_N\rangle \quad (2.5)$$

waarbij de indices s_i lopen over de elementen in de computationele basis van elk deeltje. Bijvoorbeeld, voor spin-1/2 deeltjes is $|s\rangle \in \{|0\rangle, |1\rangle\}$. De veeldeeltjesbasis lijst dan de spins van de verschillende deeltjes op. Een algemene veeldeeltjestoestand kan in deze basis ontbonden worden, met coëfficiënten $C_{s_1 \dots s_N}$ die we in een tensorvorm verzamelen. Deze algemene tensor bevat 2^N getallen en is daarmee zeer duur in termen van opslag en bewerkingen. Zoals besproken in het vorig hoofdstuk hebben veel Hamiltonianen echter interacties van lokale aard. Bijgevolg zullen in de grondtoestand de verstrengelingen tussen de deeltjes ook gelokaliseerd zijn tot naaste burens. In het geval van een ééndimensionaal systeem volstaat het om te werken met matrixproduct-toestanden daar hun structuur voldoende is om de fysica te beschrijven. De MPTs hebben de volgende vorm

$$|\psi_{MPT}\rangle = \sum_{s_1, \dots, s_N} A_{s_1, b_1}^{(1)} A_{s_2, b_1, b_2}^{(2)} \dots A_{s_N, b_{N-1}}^{(N)} |s_1 \dots s_N\rangle \quad (2.6)$$

$$= \sum_{s_1, \dots, s_N} A_{s_1}^{(1)} A_{s_2}^{(2)} \dots A_{s_N}^{(N)} |s_1 \dots s_N\rangle \quad (2.7)$$



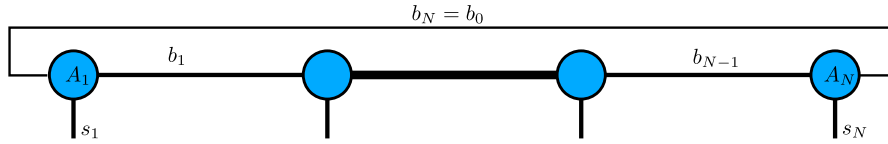
Bovenstaand voorbeeld is een MPT representatie met open randvoorwaarden, *i.e.* de eerste en de laatste tensor zijn niet direct met elkaar verbonden. Bijgevolg zijn de uiterste tensoren van rang twee terwijl de binnenste tensoren van rang drie zijn. Voor een vaste $\{s_1, \dots, s_N\}$ komt het dan neer op een *vector* $\times$ *matrix* $\times$ *vector* procedure wat resulteert in een scalair voor elke basistoestand $|s_1 \dots s_N\rangle$. De indices die de verschillende tensoren verbinden worden bond-indices genoemd en met name hun bindingsdimensies bepalen de hoeveelheid verstrengeling dat een MPT kan voorstellen tussen naburige deeltjes.

2. Tensornetwerken in de kwantummechanica

Wanneer de bindingsdimensie één is, dan hebben we een produkt-toestand gemaakt en zijn de deeltjes onafhankelijk van elkaar. We hebben een separabele toestand, zonder verstrengeling. Met een bindingsdimensie van 2 komt er verstrengeling. Dan heeft elke matrix (behalve de eerste en de laatste) 3 pootjes, en is dus een verzameling van $2^3 = 8$ getallen. Groeit het aantal deeltjes, dan groeit het aantal bij te houden getallen lineair, als $8N$. Dit is veel zuiniger dan de exponentiële groei (2^N) van de Hilbertruimte, en laat tóch toe om een beperkte hoeveelheid verstrengeling mee te nemen in de kwantumtoestand. Vaak volstaat dit om een heel goede (benadering) tot de grondtoestand te bekomen. Door de bindingsdimensie groot genoeg te kiezen (zodat we opnieuw van de orde 2^N parameters bekomen) kan elke N -deeltjes toestand voorgesteld worden.

Er bestaat nog een representatie met periodische randvoorwaarden, deze wordt ter info meegegeven:

$$\begin{aligned} |\psi_{MPT}\rangle &= \sum_{s_1, \dots, s_N} A_{s_1, b_1, b_N}^{(1)} A_{s_2, b_1, b_2}^{(2)} \dots A_{s_N, b_{N-1}, b_N}^{(N)} |s_1 \dots s_N\rangle \\ &= \sum_{s_1, \dots, s_N} \text{Tr} \left[A_{s_1}^{(1)} A_{s_2}^{(2)} \dots A_{s_N}^{(N)} \right] |s_1 \dots s_N\rangle \end{aligned}$$



2.2.1. Schmidt-decompositie en verstrengeling

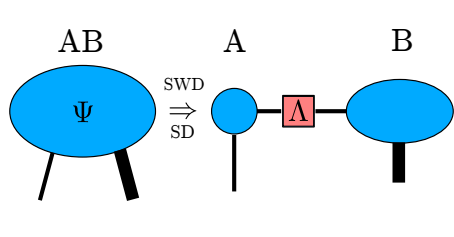
De natuurlijke vraag die zich nu stelt is, gegeven $|\psi\rangle$ hoe vindt men dan $|\psi_{MPT}\rangle$? Het antwoord hierop is: via een Schmidt-decompositie [8, 39, 40]. De volgende uitleg is op basis van [40]. De Schmidt-decompositie (SD) stelling zegt ons dat een bipartiete pure toestand $|\psi\rangle_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B$ altijd geschreven kan worden als een superpositie van orthonormale toestanden [58] via een ontbinding in een Schmidt-basis:

$$|\psi\rangle_{AB} = \sum_{\alpha=1}^{\chi} \lambda_{\alpha} |\alpha\rangle_A |\alpha\rangle_B = \sum_{\alpha=1}^{\chi} \lambda_{\alpha} |\alpha\rangle_A \otimes |\alpha\rangle_B, \quad \lambda_{\alpha} > 0 \text{ en } \sum_{\alpha} \lambda_{\alpha}^2 = 1 \quad (2.8)$$

Hierbij zijn $\{|\alpha\rangle_A\}, \{|\alpha\rangle_B\}$ de zogenaamde Schmidt-basissen voor de twee deelsystemen en is $\chi = \min(D_A, D_B)$ waarbij D de dimensie van de respectievelijke Hilbertruimte is. Het bewijs hiervoor is een toepassing van de singulierewaardenontbinding (SWO of singular value decomposition, SVD). Kies een arbitraire basis voor elk van de twee Hilbertruimten zodat de toestand geschreven kan worden als $|\psi\rangle_{AB} = \Psi_{ij} |i\rangle_A |j\rangle_B$. De matrix Ψ_{ij} heeft dan een singulierewaardenontbinding zodat $\Psi = U\Lambda V^\dagger$ met U en V twee unitaire matrices, *i.e.* $UU^\dagger = \mathbb{1} = VV^\dagger$ en Λ een $(\chi \times \chi)$ -diagonaalmatrix die de

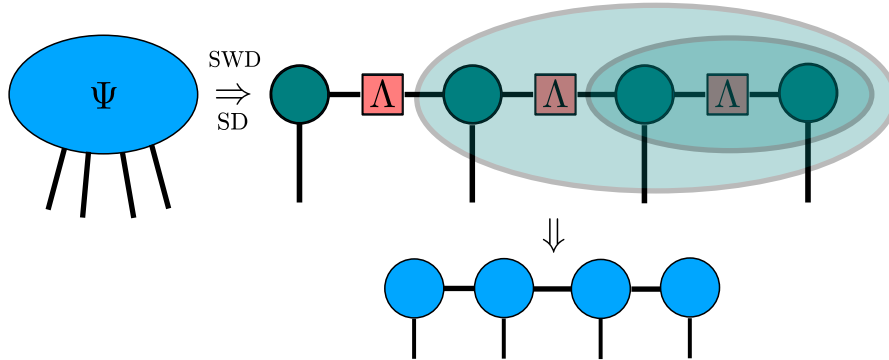
2. Tensornetwerken in de kwantummechanica

singuliere waarden bevat in afnemende volgorde[34]. Het aantal singuliere waarden geeft eveneens de rang van de oorspronkelijke Ψ matrix. Het verband tussen de SWO en de SD volgt dan uit de componentnotatie van de decompositie:



$$\begin{aligned}
 \Psi_{ij} &= \sum_{k,l} U_{ik} \Lambda_{kl} (V^\dagger)_{lj} \\
 &= \sum_k U_{ik} \lambda_k (V^\dagger)_{kj} \quad (\Lambda_{kl} = \lambda_k \delta_k^l) \\
 &= \sum_\kappa \lambda_\kappa \vec{u}_\kappa \otimes \vec{v}_\kappa
 \end{aligned}
 \tag{2.9}$$

De laatste vergelijking gebruikt nieuwe κ -indices om te benadrukken dat het hier niet om de component-indices gaat maar om de κ -de kolomvectoren van U en V alsook de κ -de singuliere waarde. Nu rest enkel nog het maken van de identificatie tussen de kolomvectoren en de Schmidt-basissen van de twee deelsystemen om te zien dat beide equivalent zijn[50]. Met andere woorden een Schmidt-decompositie van een toestandsvector geeft hetzelfde als de singulierewaardeontbinding van zijn coëfficiëntentensor. Een matrixproduct-toestand kan nu gemaakt worden door telkens dezelfde procedure iteratief te herhalen op het rechtersysteem tot dat men evenveel "kleine" tensoren heeft als deeltjes in het systeem. De matrixproduct-toestand volgt dan na de contracties tussen de Λ -matrices en één van hun burens. Matrixproduct-toestanden die uit deze procedure volgen worden ook de canonische MPTs genoemd[8, 42].



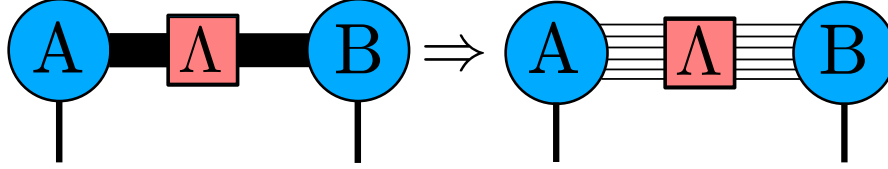
Figuur 2.2.: Diagram van de singulierewaardeontbinding toegepast op een kwantumtoestand om een matrixproduct-toestand te vormen.

De kracht van de stelling zit in de $\chi = \min(D_A, D_B)$ uitdrukking maar dat is misschien nog niet helemaal duidelijk. [58] geeft een mooi voorbeeld om dit te illustreren: $|\psi\rangle_{AB}$ is een toestand van twee kwantumsystemen A en B . Stel dat systeem A een tweedimensionale Hilbertruimte $\mathcal{H}_A$ heeft en systeem B een 2^{100} -dimensionale ruimte $\mathcal{H}_B$. Er is geen beginnend aan om dit ergens op te slaan. Echter, als het geweten is dat het om een pure toestand gaat dan zegt de Schmidt-decompositie ons dat het volstaat om een tweedimensionale deelruimte $\tilde{\mathcal{H}} \subset \mathcal{H}_B$ te vinden samen $\mathcal{H}_A$ en dit genoeg is om

2. Tensornetwerken in de kwantummechanica

$|\psi\rangle_{AB}$ te beschrijven.

Tot slot nog de link met verstrengeling. Intuïtief lijkt het alsof de singuliere waarden in de Λ matrix als een soort van "bindmiddel" dienen tussen de twee deelsystemen[50]. Als het ware dat elke singuliere waarde een verbinding tussen de twee delen vormt en de numerieke waarde als een soort gewicht dient. Een eenvoudige visualisatie volgt uit de splitsing van de bond-indices tussen Λ en de twee deelsystemen, bijvoorbeeld:



Figuur 2.3.: Illustratie van het verband tussen singuliere waarden en verstrengeling.

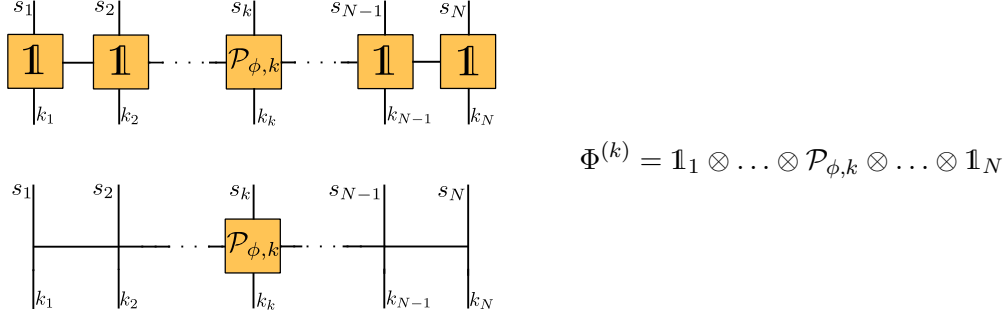
In de zin van bovenstaande figuur stellen de singuliere waarde letterlijk de verstrengeling tussen de twee deelruimten voor. Wanneer een toestand benaderd dient te worden door een toestand met een simpelere verstrengelingsstructuur dan volstaat het om tijdens bovenstaande procedure gewoon een truncatie van de eerste d singuliere waarden uit te voeren door de overige waarden weg te laten. Bovendien moet niet elke ontbinding dezelfde d gebruiken waardoor er lokaal verschillende verstrengelingssterktes kunnen zijn. Voor separabele of klassieke toestanden volstaat het om $d = 1$ te gebruiken maar voor verstrengeling is $d \geq 2$ nodig. Idealiter is de bond-dimensies zo gekozen dat het resulteert in een "zuinige" MPT die enkel de essentiële of dominante verstrengeling meeneemt.

2.2.2. Matrixproduct-operatoren

Enkel een nieuwe representatie voor toestanden volstaat niet. Om fysische grootheden te berekenen en om de toestanden te manipuleren is het nodig om ook operatoren in het nieuw formalisme te krijgen, zogenaamde matrixproduct-operatoren (MPO). De thesis zal zich beperken tot de klasse product-operatoren. Dit zijn operatoren gevormd door een reeks lokale qubit operatoren aaneengeschakeld met tensorproducten en waarvan de bond-dimensies allemaal gelijk aan één zijn. De keuze in deze thesis is verder beperkt tot enkel de vier Pauli-operatoren voor spin-1/2 deeltjes, $\{1, \sigma_x, \sigma_y, \sigma_z\}$, en projectie-operatoren naar de eigenvectoren van de Pauli-operatoren. Deze laatste zullen nodig zijn om metingen op MPTs te simuleren.

Neem bijvoorbeeld een willekeurige projectieoperator die projecteert naar een toestand ϕ zodat $\mathcal{P}_\phi \propto |\phi\rangle\langle\phi|$. Om deze operator uit te breiden naar een MPO die bv. inwerkt op het k -de deeltje dan kan dit door de projectie operator uit te breiden met een reeks identiteit-operatoren die niets doen:

2. Tensornetwerken in de kwantummechanica



De onderste figuur stelt hetzelfde voor maar sommige auteurs gebruiken de conventie om voor de identiteit-operator geen extra symbool te voorzien[39]. Immers wanneer indices verbonden worden voor een contractie impliceert dit al een inwerking van een Kronecker-delta en deze is gelijk aan de identiteit-operator. Enkele belangrijke opmerkingen: uit de structuur van de MPO volgt dat deze zowel inwerkt op $|MPT\rangle$ als op $\langle MPT|$. Bijgevolg als $|MPT\rangle$ een N -tensor is zal de MPO een $2N$ -tensor zijn. Hier geldt de conventie dat de bovenste fysieke indices s -labels krijgen zodat $MPO |MPT\rangle$ al direct de juiste contracties heeft. De overgebleven indices onderaan blijven dan beschikbaar voor een eventuele inwerking op een $\langle MPT|$.

De uitbreiding naar product-operatoren die op meerdere deeltjes inwerken gebeurt geheel analoog. Zo is een algemene MPO (in onze context) gegeven door:

$$\begin{aligned}
 \text{MPO} &= \sum_{\substack{s_1, \dots, s_N \\ k_1, \dots, k_N}} \mathcal{O}_{1,k_1}^{s_1} \dots \mathcal{O}_{N,k_N}^{s_N} |k_1 \dots k_N\rangle \langle s_1 \dots s_N| \\
 &= \sigma_1 \otimes \sigma_2 \otimes \dots \otimes \sigma_N
 \end{aligned}
 \tag{2.10}$$

In de literatuur [8, 39, 56] is het gebruikelijk om de transfermatrix $\mathbb{E}_i$ te definiëren als het i -de stukje van de MPO toegepast op op het i -de stukje van de MPT, zodat de verwachtingswaarde $\langle MPT | MPO | MPT \rangle$ gelijk is aan $\prod_i \mathbb{E}_i$.²

$$\mathbb{E}_i = \sum_{s_i, k_i} \mathcal{O}_{k_i}^{s_i} A_{s_i} \otimes A_{k_i}$$

$$\langle MPO \rangle = \langle MPT | MPO | MPT \rangle = \prod_i \mathbb{E}_i$$

Tot slot, een projectie van een MPT is gedefinieerd als de inwerking van een projectie-MPO op $|MPT\rangle$ gevolgd door een normalisatie van de nieuwe matrixproduct-toestand.

² $\text{Tr} [\prod_i \mathbb{E}_i]$ in geval van periodische randvoorwaarden.

Het normaliseren van een MPT verloopt conceptueel analoog aan de gewone toestanden, *i.e.* $|MPT'\rangle = MPO |MPT\rangle$ delen door de norm $\sqrt{\langle MPT'|MPT'\rangle}$ waarbij:

$$|MPT'\rangle = MPO |MPT\rangle$$

$$\langle MPT'|MPT'\rangle$$

(2.12)

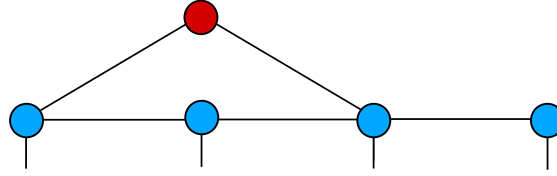
2.3. Gerichte verstrengeling matrixproduct-toestanden

De voorgaande secties hebben matrixproduct-toestanden en matrixproduct-operatoren geïntroduceerd en de Schmidt-decompositie stelling illustreert waarom dit formalisme zo goed is in het efficiënt beschrijven van ééndimensionale toestanden met naaste nabuur interacties. Echter, hiermee is het centraal tomografie probleem niet opgelost. Er is namelijk geen voorkennis van de eigenschappen van ψ , of er al dan niet verstrengeling aanwezig is of welk type verstrengeling. Het enige gegeven zijn een reeks metingen in verschillende basissen verricht op telkens dezelfde toestand. Om het probleem aan te pakken is er dus nood aan een formalisme dat iets breder toepasbaar is dan het MPT formalisme.

Over de jaren heen zijn er enkele uitbreidingen op het MPT formalisme onderzocht. Zo is er een veralgemening van het MPT formalisme naar meerdere dimensies ontwikkeld onder de naam geprojecteerde verstrengelde paar toestanden (Projected Entangled Pair States - PEPS). De PEPS hebben enkele dingen gemeen met de MPTs maar ook enkele belangrijke verschillen. Zo zijn ze bijvoorbeeld in staat om kritische fasen van een systeem te beschrijven terwijl de structuur van MPTs hier te beperkt voor is. De extra complexiteit heeft echter als prijs dat contracties tussen PEPS vaak niet meer exact kunnen gebeuren[39]. Een beknopt overzicht van de verschillende soorten netwerken en hun toepassingen is te vinden in [40].

Het PEPS formalisme is een interessante kandidaat maar de extra computationele complexiteit die nog eens bovenop de complexiteit van kwantumtomografie komt maakt het minder aantrekkelijk. Daarom is er gekozen voor een uitbreiding van de gewone matrixproduct-toestanden door het invoeren van extra hulptensoren. De bedoeling van deze hulptensoren is om een structuur te ontwikkelen die in staat is om interacties buiten naaste naburen efficiënt te beschrijven maar ondertussen compatibel blijft met de bewerkingen etc. geïntroduceerd voor gewone matrixproduct-toestanden. Deze uitbreiding is belangrijk om twee redenen. Enerzijds laat het toe om het aantal parameters in het

tensornetwerk te variëren zodat de hoeveelheid verstrengeling ook variabel is. Dit is op zich niets nieuws aangezien dit al mogelijk was met de gewone MPTs. Anderzijds laat het ook toe om de topologie van het netwerk te variëren wat een significante invloed heeft op de structuur van de verstrengeling. Het gevolg van deze twee eigenschappen is dat er een verstrengeling tussen meerdere arbitraire deeltjes, dus niet per se naaste naburen, opgezet en geschaald kan worden. Vandaar dat deze uitbreiding de naam "gerichte verstrengeling matrixproduct-toestand" (GVMPT) krijgt. Het deel "matrixproduct" is niet meer zo geschikt maar blijft behouden om de compatibiliteit met de gewone matrixproduct-toestanden te benadrukken. Een eenvoudig voorbeeld van een GVMPT is gegeven in fig. 2.4.



Figuur 2.4.: Voorbeeld van een gerichte verstrengeling matrixproduct-toestand. De tensor aangeduid in het rood is de hulptensor en de blauwe tensoren vormen het matrixproduct-toestand deel. De hulptensor verbindt het eerste en het tweede deeltje met elkaar wat zorgt voor een verstrengeling buiten de capaciteiten van een gewone matrixproduct-toestand.

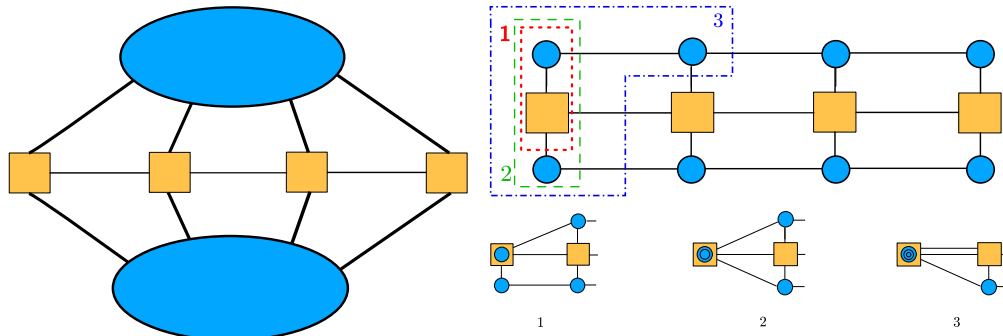
Uit bovenstaande figuur is het duidelijk waarom de GVMPTs compatibel zijn met de gewone matrixproduct-toestanden. De hulptensoren zijn volledig virtueel en de fysische indices blijven ongewijzigd.

2.3.1. Tensornetwerk contracties

Hoewel de tensordigrammen de verschillende bewerkingen goed illustreren, leggen ze geen nadruk op de volgorde van contracties in een netwerk. De volgorde van contracties heeft geen invloed op het eindresultaat maar wel op de benodigde computationele rekenkracht. Neem bijvoorbeeld een gewone separabele matrixproduct-toestand van honderd deeltjes. Aangezien het gaat om een separabele of klassieke toestand is het aantal benodigde parameters gelijk aan 2×100 wat geen probleem is op hedendaagse computers. Bij het berekenen van bv. een verwachtingswaarde mag dan de fout niet gemaakt worden van eerst het hele netwerk te contracteren tot een enkele tensor zoals in fig. 2.5, want in dat geval moeten er 2^{100} getallen opgeslagen worden wat niet haalbaar is. Figuur 2.5 toont de foute contractievolgorde (links) en een betere contractievolgorde (rechts). De optimale contractievolgorde vinden voor een arbitrair netwerk is op zich een heel moeilijk probleem dat afhangt van de aanwezige tensoren en de structuur van het netwerk[8]

Hiermee is de nodige kennis over tensornetwerken geïntroduceerd. De volgende stap is het achterhalen met welke getallen deze tensoren gevuld moeten worden om een kwantumtoestand zo accuraat en efficiënt mogelijk te beschrijven. De procedure van het

2. Tensornetwerken in de kwantummechanica



Figuur 2.5.: Problematische contractievolgorde links en een geoptimalizeerde contractievolgorde rechts.

achterhalen van de parameters staat in de statistiek bekend als parameterschatten en vormt het volgend onderwerp.

3. Parameterschatten en modelinferentie

De tensornetwerken geïntroduceerd in het vorig hoofdstuk kunnen ook geïnterpreteerd worden als verschillende statistische modellen voor meetuitkomsten. Immers, als de kwantumtoestand gekend is dan kan daaruit de exacte kansdichtheid achter de metingen berekend worden. Het kennen van een kwantumtoestand is in deze interpretatie equivalent met het kennen van de parameters van een model. Merk op dat het mogelijk is om eenzelfde kansdichtheid voor meetuitkomsten te bekomen uit verschillende modellen met elk een verschillende parameterstructuur. Bijvoorbeeld, voor een drie-deeltjestoestand kunnen we een aantal modellen voor de toestandskets beschouwen:

- een algemene toestandsvector $|\psi\rangle$ (acht parameters),
- de producttoestand $|\psi_{\text{prod}}\rangle = |\psi_1\rangle \otimes |\psi_2\rangle \otimes |\psi_3\rangle$ (6 parameters),
- de matrixproduct-toestand $|\psi_{\text{MPT}}\rangle$.

In het geval van separabele drie-deeltjestoestanden zijn alle drie de modellen in staat om de juiste kansverdeling voor de meetuitkomsten te genereren. Dan is eigenlijk de algemene toestandsvector een overgeparametriseerd model - er zullen veel verschillende parameterwaarden zijn die compatibel zijn met de meetuitkomsten. Maar in het geval van een verstrengelde driedeeltjestoestand is de producttoestand niet meer toereikend. Moeten we dan naar de algemene toestand, of zal een matrixproduct toestand volstaan als model?

Het is duidelijk dat verschillende modellen voor de kwantumtoestand verschillende capaciteiten hebben. Elk model kan beschouwd worden als een familie¹ van kansdichtheden geïndexeerd door de parameterwaarden[36] en dat er een niet lege doorsnede bestaat tussen de families.

Het probleem bij (kwantum)tomografie is dat de exacte kwantumtoestand onbekend is en bijgevolg dus ook de parameters van het model. De enige uitweg is om een schatting te maken van de parameters op basis van een reeks meetresultaten afkomstig van eenzelfde kansdichtheid, wat het een invers probleem maakt[23]. Daarnaast is er ook nog de vrijheid in het kiezen van een model (en van het aantal modelparameters). Om dat probleem aan te pakken moet er gekeken worden naar de theorie van modelselectie of modelinferentie. De combinatie van deze twee domeinen vormt de ruggengraat van veel

¹Een familie is een functie die een indexverzameling afbeeldt op een andere verzameling $\mathcal{X}$, m.a.w. het indexeert de elementen van $\mathcal{X}$ [60].

onderzoeksdomeinen zoals artificiële intelligentie, psychologie[36], elektronenmicroscopie [46, 47], etc.

Het tomografieprobleem dat hier onderzocht wordt bestaat dus uit twee delen. Enerzijds is er de zoektocht naar de optimale parameterwaarden voor een gegeven model en anderzijds is er de zoektocht naar het optimaal model. Daarom heeft dit hoofdstuk ook twee doelen: het eerste doel is het introduceren van technieken om parameters te schatten. Twee belangrijke technieken zullen besproken worden: schatting volgens de meest aannemelijke schatter (maximum likelihood estimation - MLE) en Bayesiaanse schattingen. Het tweede doel is het introduceren van technieken om aan modelselectie te doen. De theorie wordt toegelicht met een voorbeeld van klassieke munten. Tot slot volgt er bespreking over de uitbreiding naar de kwantummechanica en de uitdagingen die er mee gepaard zijn.

3.1. Parameterschatten

De statistiek kent een drietal grote takken die zich bezighouden met het schatten van parameters: de regressie-analyse (bv. kleinste kwadraten schatter), de meest aannemelijke schatters en de Bayesiaanse schatters[26, 36, 45, 47]. Regressie-analyse zoekt in eerste instantie naar verbanden tussen verschillende variabelen en tracht dan de optimale parameters voor dit verband te schatten via de minimalisatie van een kostfunctie onder specifieke assumpties. Het is een heel rechtstreekse methode maar verder niet interessant voor het huidige probleem.

De methode van de meest aannemelijke schatters en de Bayesiaanse methode maken allebei gebruik van een zogenaamde aannemelijkheidsfunctie (likelihood function) maar bekijken het probleem vanuit verschillende invalshoeken. Zo behoort het principe van de meest aannemelijke schatter tot de frequentie-interpretatie van statistiek die zegt dat de kansen de limieten zijn van de relatieve frequenties. Hoe meer metingen hoe beter de parameters geschat kunnen worden. De Bayesiaanse interpretatie daarentegen gaat verder en construeert een kansdichtheid voor de parameters zelf, *i.e.* de parameters zijn kansvariabelen. Dit is interessant omdat het een uitgebreidere discussie over de parameters mogelijk maakt maar introduceert een zekere subjectiviteit in de interpretatie[26, 45, 52]. Onder bepaalde voorwaarden is het echter wel mogelijk dat beide interpretaties naar hetzelfde resultaat convergeren[59].

3.1.1. Meest aannemelijke schatter

Het principe van de meest aannemelijke schatter werd al gebruikt in de tijd van Laplace en Gauss[24] maar werd pas in het begin van de 20ste eeuw in ruwe vorm geïntroduceerd door F. Y. Edgeworth en vervolgens in 1922 op de kaart gezet door R. A. Fisher. Hij introduceerde daarbij ook veel van de nu nog steeds gebruikte woordenschat [3]. De

3. Parameterschatten en modelinferentie

Bayesiaanse methode was destijds ook al bekend maar was niet populair omdat het wiskundig en computationeel (in de analoge zin) te ingewikkeld was om te gebruiken buiten voor een paar speciale gevallen[26].

De volgende paragrafen volgen de bespreking in [36]. Veronderstel een systeem gemodelleerd door een model $\mathcal{M}(\theta)$ met $\theta = (\theta_1, \theta_2, \dots)$ de parametervector van het model. Het model bepaalt de algemene structuur van de kansdichtheid terwijl de waarden van θ de precieze kansen bepalen. Beschouw een reeks trekkingen uit de onbekende kansverdeling van $\mathcal{M}(\theta)$ voorgesteld door de uitkomstvector $y = (y_1, y_2, \dots)$. Aan het ensemble van de metingen kan nu een experimentele kansdichtheid toegekend worden van de vorm $f(y|\mathcal{M}, \theta) \equiv f(y|\theta)$. Dit is een voorwaardelijke kansverdeling die de kans op een observatie y geeft op voorwaarde dat θ de parameters zijn. Als de individuele metingen onafhankelijk van elkaar zijn dan kan $f(y|\theta)$ gefactoriseerd worden volgens de individuele kansen en bekomt men de multinomiaalverdeling²:

$$f(y|\theta) = \prod_i f_i(y_i|\theta).$$

Parameterschatten is nu het invers probleem[23, 36] om θ te schatten gegeven de observatievector y . Door de rollen van parameters en observaties om te wisselen wordt de likelihood of aannemelijkheidsfunctie geïntroduceerd:

$$L(\theta|y) = f(y|\theta). \quad (3.1)$$

Merk op dat de bovenstaande gelijkheid enkel betrekking heeft tot de wiskundige formule, $L(\theta|y)$ en $f(y|\theta)$ zijn functies met verschillende eigenschappen: f is een genormaliseerde kansdichtheid waarbij θ constant is en enkel y variabel is. L daarentegen is geen genormaliseerde kansverdeling[24]. De gelijkheid geldt enkel wanneer exact dezelfde argumenten gebruikt worden.

Eens de aannemelijkheidsfunctie bekend is dan zegt Fishers principe van de meest aannemelijke schatter dat de beste schatting van de parameters degene is die de aannemelijkheidsfunctie maximaliseert:

$$\theta_{ML} = \arg \max_{\theta} [L(\theta|y)].$$

De interpretatie luidt dan dat θ_{ML} de meest waarschijnlijke of aannemelijke waarde voor θ is gegeven de observaties y of θ_{ML} verklaart de geobserveerde data het best[47]. Merk op dat het omgekeerde niet algemeen waar is, gegeven θ_{ML} dan zal y niet per se het meest waarschijnlijke resultaat zijn.

Een veel gebruikte truuk is om bovenstaande maximalisatie te vervangen door een minimalisatie van

$$-\log [L(\theta|y)] = -\sum_i \log [f_i(y_i|\theta)].$$

²Met de nodige normalisatiefactor.

3. Parameterschatten en modelinferentie

Aangezien de logaritme een monotone functie is zal deze de positie van het extremum niet veranderen. Het voordeel van deze transformatie is dat het de computationele stabiliteit verbetert door kleine waarden af te beelden naar grote waarden en door het product te vervangen door een sommatie. Bovendien volgen veel softwarebibliotheken de conventie om een optimalisatieprobleem te implementeren als een minimalisatieprobleem. In sommige gevallen, zoals de multinomiaalverdeling, is het echter mogelijk om het extremum op een analytische wijze te vinden.

3.1.2. Bayesiaanse schatting

De volgende paragrafen gebruiken dezelfde notatie als hierboven en volgen in grote lijnen de discussie uit [52]. In plaats van direct met de parameters te werken kijkt de Bayesiaanse techniek naar de kansdichtheid van een parameters. De centrale vergelijking achter deze techniek volgt uit de stelling van Bayes:

$$P(\theta|y) = \frac{P(y|\theta)P(\theta)}{P(y)}. \quad (3.2)$$

Het idee achter bovenstaande vergelijking is om de kansdichtheid van θ op te splitsen in twee delen. Enerzijds is er de prior of onvoorwaardelijke kans $P(\theta)$ die dient als een soort nulhypothese steunend op alle a-priori informatie die een experimentator heeft buiten de beschikbare metingen y , ook informatie verkregen na het uitvoeren van de metingen maar voor dat de meetresultaten in rekening worden gebracht. De term “prior” verwijst dus naar het nog niet in rekening brengen van meetdata[24]. Bijvoorbeeld als θ de kans is van een munt om kop te geven, dan is het niet ongewoon dat de experimentator bij gebrek aan zichtbare defecten veronderstelt dat $P(\theta)$ gepiekt is rond $\theta = 1/2$. Dit is een aanname die gemaakt wordt nog voor het opgooien van de munt. Anderzijds is er de posterior of herziene kans $P(\theta|y)$ die fungeert als alternatieve hypothese die gevormd wordt na het in rekening brengen van de meetdata. De voorwaardelijke kans $P(y|\theta)$ heeft dezelfde interpretatie als bij de meest aannemelijke schatter en kan opnieuw geïdentificeerd worden met de aannemelijkheidsfunctie $L(\theta|y)$. Tot slot is er nog de term die “het bewijs” wordt genoemd; de kans $P(y)$ om y te observeren ongeacht enige prior kennis en deze dient om het geheel te normaliseren. Deze kans is gegeven door de volgende integraal over de parameter ruimte Θ :

$$P(y) = \int_{\Theta} P(y|\theta)P(\theta) d\theta \quad (3.3)$$

De vergelijking van Bayes zegt hoe de meetresultaten in rekening gebracht moeten worden met de al aanwezige kennis *i.e.* de prior wordt gebruikt als een gewicht voor de aannemelijkheidsfunctie. Dit verklaart ook waarom de Bayesiaanse methode en de MLE methode uit dezelfde data toch tot verschillende conclusies kunnen komen[24, 45, 47].

3. Parameterschatten en modelinferentie

Bayes' stelling kan ook geïnterpreteerd worden als een protocol om de oude prior $P(\theta) = P_o(\theta)$ te updaten naar een nieuwe prior $P_n(\theta)$ die dan als feedback kan dienen voor een nieuwe Bayesiaanse schatting met nieuwe meetdata y' . Wanneer de prior en de posterior in dezelfde familie van kansverdelingen zitten, dan is de nieuwe Bayesiaanse schatting equivalent aan de oorspronkelijke schatting met de dataset $Y = y + y'$ [52].

Merk op dat de “al aanwezige kennis” kan verschillen van context tot context waardoor de prior een bron van subjectiviteit is. In geval van onwetendheid over de prior of indien een meer neutrale keuze gewenst is dan biedt Jaynes' principe van maximale entropie een uitweg: deze stelt voor om de prior te kiezen die de maximale entropie heeft [52]. Een uitgebreide uitleg is te vinden in [24]. In kort komt het neer op het volgende[25]: gegeven een kansverdeling $p = \{p_i\}$ en een geassocieerde Shannon-entropie $H(p)$ ³

$$H(P) = - \sum_i p_i \log(p_i), \quad (3.4)$$

zoek dan de oplossing p^* waarvoor $H(p)$ maximaal is en p gebonden aan bepaalde voorwaarden. De standaardvoorwaarden zijn die van een goed gedefinieerde kansverdeling *i.e.* $\sum_i p_i = 1$ en $\forall i, p_i \geq 0$. Als er verder geen andere voorwaarden zijn dan is de oplossing een uniforme kansverdeling.

De posterior $P(\theta|y)$ heeft als variabele de parameters θ waarbij y vast is. Deze kan dan ook geïnterpreteerd worden als een nieuwe aannemelijkheidsfunctie $\mathcal{L}(\theta|y)$ die de gewogen Fisher aannemelijkheidsfunctie $L(\theta|y) = p(y|\theta)$ is. Uit de vergelijking van Bayes volgt dat deze nieuwe aannemelijkheidsfunctie wel een genormaliseerde kansverdeling zal zijn. Een schatting van de parameter θ volgt nu uit de toepassing van Fisher's principe van de maximale aannemelijkheid op $\mathcal{L}(\theta|y)$. Hieruit volgt dat indien een uniforme kansdichtheid is gekozen voor de prior de Bayesiaanse schatting en de meest aannemelijke schatting equivalent zijn of deze laatste is een speciaal geval van de Bayesiaanse schatting.

3.2. Modelinferentie

Bovenstaande methodes zijn een vorm van “microscopische” optimalisaties waarbij elk model gewoon zijn optimale parameters krijgt om de meetdata zo goed mogelijk te verklaren. Het biedt echter geen antwoord op de relatieve performantie tussen verschillende modellen. Twee modellen kunnen misschien dezelfde hoeveelheid parameters hebben maar wegens verschillende verbanden tussen parameters, *i.e.* modelstructuur, de data beter of slechter verklaren. Dit speelt bijvoorbeeld een grote rol in het domein van neurale netwerken. Het is de theorie van modelinferentie die deze vragen over de “macroscopische” optimalisatie van modellen bestudeert.

³Voor p_i wordt de l'Hôpital limiet gebruikt: $\lim_{p_i \rightarrow 0} p_i \log(p_i) \stackrel{H}{=} 0$

3.2.1. Kullback-Leibler divergentie

Beschouw opnieuw Bayes' update protocol. De vraag stelt zich hoeveel nieuwe informatie er vrijkomt of verloren gaat door $P_n(\theta)$ te gebruiken in plaats van $P_o(\theta)$. Aangezien de entropie een maat is voor de hoeveelheid (ontbrekende) informatie kan de vraag geherformuleerd worden als: wat is de relatieve entropie tussen de twee priors? Het is exact die grootte die Solomon Kullback en Richard Leibler besloten te gebruiken om de divergentie⁴ tussen twee kansverdelingen te kwantiseren en vandaar de naam Kullback-Leibler divergentie of kortweg KL-divergentie D_{KL} [29, 52]:

$$D_{KL} [P_n(\theta)|P_o(\theta)] = \int_0^1 P_n(\theta) \log \left(\frac{P_n(\theta)}{P_o(\theta)} \right) d\theta. \quad (3.5)$$

De rol van de KL-divergentie moet niet beperkt blijven tot het toekennen van een “score” voor een kansverdeling of model maar het kan ook gebruikt worden als een hulpmiddel in het verbeteren van modellen of experimenten. Experimenten kunnen bijvoorbeeld zodanig opgezet worden om met elke meting een maximale hoeveelheid aan nieuwe informatie te introduceren of, meer interessant voor deze thesis, parameters en modellen updaten zodat er een maximale hoeveelheid aan nieuwe informatie geïntroduceerd wordt.

3.2.2. Het Akaike informatie criterium

Voor het tomografieprobleem blijft een cruciale vraag nog onbeantwoord: hoe ver is een model van de waarheid? Bovenstaande paragrafen tonen hoe modellen tegenover elkaar verbeterd kunnen worden maar dat betekent niet dat de volgende iteraties daarom de fysische realiteit beter beschrijven. Er is dus nood aan een maat voor de “afstand” tussen een model en de fysische realiteit. Anders geformuleerd, hoeveel informatie gaat er verloren door de fysische realiteit te benaderen door een model?

Het goede nieuws is dat de KL-divergentie hiervoor gebruikt kan worden[52]. Gegeven een model $\mathcal{M} = \mathcal{M}(\theta)$ met geassocieerde kansverdeling $p(y|\mathcal{M})$ en de fysische realiteit $\mathcal{F}$ met kansverdeling $f(y)$ dan bedraagt de KL-divergentie:

$$D_{KL} (\mathcal{F}|\mathcal{M}) = \int f(y) \log \left(\frac{f(y)}{p(y|\mathcal{M})} \right) dy. \quad (3.6)$$

Merk op dat bovenstaande vergelijking de vorm heeft van een verwachtingswaarde met $f(y)$ als gewicht. Het logaritme herschrijven geeft

$$D_{KL} (\mathcal{F}|\mathcal{M}) = \mathbb{E}_f [\log(f(y))] - \mathbb{E}_f [\log(p(y|\mathcal{M}))]. \quad (3.7)$$

⁴De divergentie kan geïnterpreteerd worden als een afstandsfunctie. Merk op dat het geen wiskundig formele afstandsfunctie is aangezien het niet de nodige eigenschappen zoals symmetrie heeft.

3. Parameterschatten en modelinferentie

Het slechte nieuws is dat voor het tomografieprobleem de fysische realiteit $\mathcal{F}$ onbekend is. De herformulering als het verschil tussen verwachtingswaarden geeft echter wel hint dat de $\mathcal{F}$ -afhankelijkheid omzeild kan worden. Neem een tweede model $\mathcal{M}'$ en beschouw het verschil tussen de twee divergenties:

$$D_{KL}(\mathcal{F}|\mathcal{M}) - D_{KL}(\mathcal{F}|\mathcal{M}') = \mathbb{E}_f \left[\frac{\log(P(y|\mathcal{M}'))}{\log(P(y|\mathcal{M}))} \right] \quad (3.8)$$

Hieruit kan nog steeds niet de afstand tot de waarheid berekend worden maar het is wel mogelijk om af te leiden welk model de waarheid beter benadert. De linkerterm zal immers positief zijn als $\mathcal{M}'$ dichter bij de waarheid is en negatief al $\mathcal{M}$ dichter bij de waarheid is. Zolang $\mathcal{F}$ onbekend is kan de linkerterm echter niet uitgerekend worden. De rechterterm uitwerken is niet triviaal aangezien deze verwachtingswaarde niet enkel afhangt van data y maar ook van de schattingen $P(\theta|y)$.

Een belangrijk resultaat om bovenstaande situatie te ontrafelen is er in de jaren 70 gekomen dankzij Hirotugu Akaike. Akaike is er in geslaagd om de volgende uitdrukking, het Akaike informatie criterium (AIC), af te leiden in zijn poging om de aannemelijkheidsfunctie te veralgemenen[1, 2, 9, 52]:

$$-\mathbb{E}_f \{ \log[P(y|\mathcal{M})] \} = -\log[L(\theta|y, \mathcal{M})] + K. \quad (3.9)$$

$L(\theta|y, \mathcal{M})$ stelt opnieuw de aannemelijkheidsfunctie die θ schat gegeven y en $\mathcal{M}$. K is het aantal parameters in het model. Akaike interpreteerde het rechterlid als een schatter van de verwachtingswaarde links en toonde aan dat deze bij benadering onvertekend is[52].⁵ Door Akaike's resultaat terug in te vullen in 3.7 is het duidelijk dat deze termen als een straf fungeren die de KL-divergentie vergroten naarmate ze toenemen in waarde. Deze straf is om overfitten te voorkomen.

Gegeven een verzameling van kandidaat modellen is het nu mogelijk om elk model een score toe te kennen volgens het Akaike informatie criterium. Het model met de grootste score (in absolute waarde) heeft dan de kleinste KL-divergentie met de fysische realiteit $\mathcal{F}$ en is daarmee het beste model.

Om het vergelijken van modellen te verduidelijken is het handig om zogenaamde Akaike gewichten te introduceren. Deze zijn gegeven door:

$$\begin{aligned} w_{\mathcal{M}} &= \exp\{\mathbb{E}_f[\log\{P(y|\theta, \mathcal{M})\}]\} = \exp\{\log\{L(\theta|y, \mathcal{M})\} - K_{\mathcal{M}}\} \\ &= L(\theta|y, \mathcal{M})e^{-K_{\mathcal{M}}}. \end{aligned} \quad (3.10)$$

Door deze gewichten vervolgens te normaliseren kan het i -de model een Akaike kans krijgen van de vorm

$$P_i = \frac{w_i}{\sum_j w_j}. \quad (3.11)$$

⁵In het geval dat het aantal observaties N van dezelfde grootteorde is als het aantal parameters K dan wordt er een correctieterm $+\frac{K(K+1)}{N-K-1}$ geïntroduceerd.

3. Parameterschatten en modelinferentie

Deze kansen wijzen niet enkel het beste model aan maar geven ook een maat voor het “vertrouwen” dat dit het beste model is. Dit is analoog aan hoe de Bayesiaanse schatting ook een kans toekent aan de geschatte parameter. Nog een voordeel van de Akaike kansen is dat ze gebruikt kunnen worden in een iteratieve zoektocht naar betere modellen, gelijkaardig aan het idee voorgesteld met de Kullback-Leiblerdivergentie[52]. Definieer de acceptatieverhouding α als

$$\alpha = \frac{P_n}{P_o} = \frac{w_n}{w_o} = \frac{L(\theta_n|y, \mathcal{M}_n)}{L(\theta_o|y, \mathcal{M}_o)} e^{K_o - K_n} \quad (3.12)$$

$$= \exp\{\mathbb{E}_f[\log\{P(y|\mathcal{M}_n)\}] - \mathbb{E}_f[\log\{P(y|\mathcal{M}_o)\}]\}. \quad (3.13)$$

Uit de bovenstaande vorm is het duidelijk dat α omgekeerd evenredig is met de hoeveelheid verloren informatie bij het voorstellen van een nieuw model, immers als het nieuw model beter is dan geldt $P_n > P_o$. Dit laat toe om α te gebruiken bij de implementatie van een Metropolis-Hastingsalgoritme in een Monte Carlo simulatie dat naar het beste model zoekt[33, 52].

Nog een korte opmerking: het Akaike informatie criterium verg. (3.9) hangt niet meer af van de realiteit $\mathcal{F}$ maar zegt enkel iets over de beschikbare meetdata y en het model $\mathcal{M}$. Het is mogelijk dat er een model is die de data beter verklaart dan andere modellen en daarom een beter score krijgt; dat impliceert echter niet dat het model daarom ook dichter bij de realiteit staat[52]. Het is dan goed om enkele “sanity checks” te hebben voor het verkozen model. Bijvoorbeeld afvragen of de complexiteit van het model in lijn is met de (verwachte) complexiteit van de realiteit. Als alternatief kan het misschien interessant zijn om te zien of verschillende informatie criteria met andere eigenschappen, zoals bv. het Schwarz informatie criterium [37], ook hetzelfde model verkiezen.

Alle basistechnieken om aan parameterschatten en modelinferentie te kunnen doen zijn nu geïntroduceerd. De volgende sectie zal de theorie illustreren met een klassiek voorbeeld over klassieke munten. Het laatste deel zal de uitbreiding naar kwantummechanica tensornetwerken bespreken.

3.3. Toepassing: klassieke munten

Om de theorie uit de voorgaande secties te illustreren volgt nu een kort voorbeeld met klassieke munten. Dit deel volgt [52] en in een volgend hoofdstuk wordt er een gelijkaardig probleem numeriek uitgewerkt.

Het probleem is als volgt: gegeven twee munten, elk met mogelijke uitkomsten $\{kop, munt\} = \{k, m\}$, wat zijn dan de gezamenlijke kansverdelingen P_{kk}, P_{km}, P_{mk} en P_{mm} . Dit is in principe niets anders dan het tomografieprobleem geherformuleerd voor klassieke munten. Merk op dat wegens de normalisatie van kansen er slechts drie onafhankelijke of vrije kansvariabelen zijn. Er zijn drie soorten modellen die elk een verschillende behandelingen vereisen. De concrete modelstructuren zijn te vinden in tabel 3.1:

3. Parameterschatten en modelinferentie

1. Onafhankelijk en identiek: onafhankelijk impliceert dat de gezamenlijke kansen van de twee munten factorizeerbaar zijn als het product van de individuele kansen, *i.e.* $P_{AB}(i, j) = P_A(i)P_B(j)$. Identiek betekent dat één set van parameters volstaat om beide munten te beschrijven, *i.e.* $\theta_A = \theta_B = \theta$ bijgevolg $P_A(i) = P_B(i)$. Het globaal model dat het twee-muntensysteem beschrijft kan opgebouwd worden uit individuele muntmodellen. Merk op dat deze situatie equivalent is aan meerdere metingen uitvoeren op één munt.
2. Onafhankelijk: de gezamenlijke kansen zijn nog steeds factorizeerbaar maar de munten zijn niet meer identiek en hebben dus elk hun eigen en verschillende parameters θ_A en θ_B .
3. Gecorreleerd: een meting op één munt onthult informatie over de andere munt, de munten zijn niet meer onafhankelijk. In dit geval zijn de gezamenlijke kansen niet meer factorizeerbaar. Parameters kunnen niet meer toegekend worden aan de individuele modellen en dus moet het globaal model in zijn algemeenheid beschouwd worden. Verstrengeling zal ook resulteren in correlaties maar afhankelijk van de metingen kunnen die correlaties van een heel andere aard zijn; een gedetailleerde bespreking is te vinden in paragraaf 3.4.1

De kansen die niet factorizeerbaar zijn bij correlaties doen denken aan de verstrengelde toestanden die niet separabel zijn. Het zijn echter twee verschillende fenomenen. Verstrengeling is eigen aan de kwantumtoestand en wordt (gedeeltelijk) vernietigd na een meting op een van de subsystemen. Bovendien kan verstrengeling niet waargenomen worden door metingen in één basis zoals het geval bij de munten.

Tabel 3.1.: Overzicht van de verschillende modellen en hun parameterstructuur.

Model ($\mathcal{M}$)	#parameters	Kansen			
		P_{kk}	P_{km}	P_{mk}	P_{mm}
A) Identiek	θ	θ^2	$\theta(1 - \theta)$	$(1 - \theta)\theta$	$(1 - \theta)^2$
B) Onafhankelijk	θ_1, θ_2	$\theta_1\theta_2$	$\theta_1(1 - \theta_2)$	$(1 - \theta_1)\theta_2$	$(1 - \theta_1)(1 - \theta_2)$
C) Gecorreleerd	$\theta_1, \theta_2, \theta_3$	θ_1	θ_2	θ_3	$1 - \theta_1 - \theta_2 - \theta_3$

Één meting bestaat uit het opgooien van alle munten (al dan niet tegelijkertijd) en vervolgens hun uitkomst te registreren. Beschouw de frequentievector van de geregistreerde uitkomsten $n = \{n_{kk}, n_{km}, n_{mk}, n_{mm}\}$ met $N = \sum_i n_i$ het totaal aantal metingen. Alle metingen zijn onafhankelijk van elkaar uitgevoerd en telkens met dezelfde munten, m.a.w. het zijn trekkingen uit eenzelfde kansverdeling. Onder deze twee voorwaarden zal de kansverdeling achter de observaties een multinomiale verdeling zijn:

3. Parameterschatten en modelinferentie

$$P(n|\mathcal{M}_\theta) = \mathcal{N} \times P(kk|\mathcal{M}_\theta)^{n_{kk}} P(km|\mathcal{M}_\theta)^{n_{km}} P(mk|\mathcal{M}_\theta)^{n_{mk}} P(mm|\mathcal{M}_\theta)^{n_{mm}} \quad (3.14)$$

$$= \mathcal{N} \times (\theta_{kk})^{n_{kk}} (\theta_{km})^{n_{km}} (\theta_{mk})^{n_{mk}} (\theta_{mm})^{n_{mm}}, \quad (3.15)$$

met $\mathcal{N}$ de multinomiale normalisatiefactor

$$\mathcal{N} = \frac{N!}{n_{kk}! n_{km}! n_{mk}! n_{mm}!}. \quad (3.16)$$

Concreet voor de drie verschillende modellen geeft dit na enige uitwerking:

$$P(n|A) = \mathcal{N} \times (\theta)^{2n_{kk}+n_{km}+n_{mk}} (1-\theta)^{2n_{mm}+n_{km}+n_{mk}} \quad (3.17)$$

$$P(n|B) = \mathcal{N} \times (\theta_1)^{n_{kk}+n_{km}} (1-\theta_1)^{n_{mm}+n_{mk}} (\theta_2)^{n_{kk}+n_{mk}} (1-\theta_2)^{n_{mm}+n_{km}} \quad (3.18)$$

$$P(n|C) = \mathcal{N} \times (\theta_1)^{n_{kk}} (\theta_2)^{n_{km}} (\theta_3)^{n_{mk}} (1-\theta_1-\theta_2-\theta_3)^{n_{mm}}. \quad (3.19)$$

3.3.1. Betafunctie voor normalisatiefactoren

Merk op dat de bovenstaande vergelijkingen afkomstig zijn van dezelfde multinomiale vorm gegeven in verg. (3.14). De verschillende vormen in verg. (3.17) kunnen echter ook geïnterpreteerd worden als nieuwe multinomiale kansverdelingen. In dit geval zal hun normalisatiefactor ook een andere vorm hebben. Met behulp van de multivariabele bètafunctie kan de normalisatieconstante voor eender welke multinomiale verdeling berekend worden[15]. De bètafunctie is gegeven door[10]:

$$B(b_1, \dots, b_k) = \frac{\Gamma(b_1) \dots \Gamma(b_k)}{\Gamma(b_1 + \dots + b_k)} = \frac{(b_1 - 1)! \dots (b_k - 1)!}{(b_1 + \dots + b_k - 1)!} \quad (3.20)$$

$$= \int_E u_1^{b_1-1} \dots u_{k-1}^{b_{k-1}-1} (1 - u_1 - \dots - u_{k-1})^{b_k-1} du_1 \dots du_{k-1}. \quad (3.21)$$

Het integratiedomein E is de $(k-1)$ -dimensionale simplex gegeven door $E = \{(u_1, \dots, u_{k-1}) : \forall u_i \geq 0, \sum_i u_i \leq 1\}$. Dit is equivalent met het domein van de vrije kansparameters. De normalisatiefactoren voor de kansverdelingen in verg. (3.17) volgen dan uit de volgende identificaties $u_i = \theta_i$ en $b_i = n_i + 1$ zodat

$$\mathcal{N}^{-1} = B(n) = \frac{n_1! \dots n_k!}{(N + k - 1)!}.$$

3. Parameterschatten en modelinferentie

Als alternatief kan de bovenstaande vergelijking ook gefactoriseerd worden voor de individuele subsystemen $S = \{A, B, \dots\}$ zodat

$$\mathcal{N}^{-1} = B(n) = \prod_{i \in S} B(\vec{n}_i)$$

Toepassing van Bayes' stelling verg. (3.2) met een uniforme prior en normalisatiefactor berekend volgens de bètafunctie geeft dan de volgende genormaliseerde aannemelijkheidsfuncties:

$$\begin{aligned} \mathcal{L}(\theta|n, A) &= \frac{(2N+1)!}{(2n_{kk} + n_{km} + n_{mk})!(2n_{mm} + n_{mk} + n_{km})!} \\ &\quad \times (\theta)^{2n_{kk} + n_{km} + n_{mk}} (1-\theta)^{2n_{mm} + n_{km} + n_{mk}} \end{aligned} \quad (3.22)$$

$$\begin{aligned} \mathcal{L}(\theta_1, \theta_2|n, B) &= \left(\frac{(N+1)!}{(n_{kk} + n_{km})!(n_{mm} + n_{mk})!} \right)^2 \\ &\quad \times (\theta_1)^{n_{kk} + n_{km}} (1-\theta_1)^{n_{mm} + n_{mk}} (\theta_2)^{n_{kk} + n_{mk}} (1-\theta_2)^{n_{mm} + n_{km}} \end{aligned} \quad (3.23)$$

$$\begin{aligned} \mathcal{L}(\theta_1, \theta_2, \theta_3|n, C) &= \frac{(N+3)!}{n_{kk}!n_{km}!n_{mk}!n_{mm}!} \\ &\quad \times (\theta_1)^{n_{kk}} (\theta_2)^{n_{km}} (\theta_3)^{n_{mk}} (1-\theta_1-\theta_2-\theta_3)^{n_{mm}}. \end{aligned} \quad (3.24)$$

Voor de aannemelijkheid van model A is de interpretatie genomen dat één munt $2N$ keer is opgegooid i.p.v. twee munten N keer opgegooid. Optimale parameters zijn degene die de aannemelijkheid maximaliseren; deze kunnen numeriek berekend worden of analytisch volgens de voorwaarde $\frac{\partial \mathcal{L}}{\partial \theta} = 0$ of nog beter $\frac{\partial \log(\mathcal{L})}{\partial \theta} = 0$. Hieruit volgt:

$$\text{Model A: } \theta_{ML} = \frac{2n_{kk} + n_{mk} + n_{km}}{2N} \quad (3.25)$$

$$\text{Model B: } \theta_{1,ML} = \frac{n_{kk} + n_{km}}{N}, \quad \theta_{2,ML} = \frac{n_{mm} + n_{mk}}{N} \quad (3.26)$$

$$\text{Model C: } \theta_{1,ML} = \frac{n_{hh}}{N}, \quad \theta_{2,ML} = \frac{n_{km}}{N}, \quad \theta_{3,ML} = \frac{n_{mk}}{N} \quad (3.27)$$

Nu de optimale parameters gekend zijn kunnen de Akaike gewichten en kansen eenvoudigweg uit de definitie berekend worden. Om numerieke resultaten te krijgen volstaat het om enkele uitkomsten te kiezen (of simuleren). In een volgend hoofdstuk ziet men deze theorie echt in actie. De laatste sectie zal kort toelichten hoe de theorie toegepast kan worden in de context van kwantummechanica en tensor netwerken.

3.4. Uitbreiding naar de kwantummechanica en tensornetwerken

Tot nu toe was de voorgaande discussie beperkt tot klassieke systemen en relatief eenvoudige modellen. Het is nu tijd om uit te breiden naar de kwantummechanica. In de vorige sectie werd er al opgemerkt dat er een verschil kan zijn tussen klassieke correlaties en correlaties wegens verstrengeling. Het volgende deel bekijkt dat verschil van dichterbij.

3.4.1. Het kwantummechanisch karakter: de ongelijkheid van Bell

Zoals aangehaald in de inleiding zegt de huidige interpretatie van de kwantummechanica dat voorafgaand aan metingen de kwantumtoestand van een deeltje onbepaald is. Zelfs als een deeltje met absolute zekerheid geprepareerd wordt in de spintoestand $|0\rangle$ dan zal een meting van de spin in een orthogonale richting, X of Y , resulteren in willekeurige realisatie van een meetresultaat.

In 1935, concludeerden Einstein, Podolsky en Rosen in hun befaamde EPR-paradox dat de enige verklaring voor die onbepaaldheid was dat kwantummechanica een onvolledige theorie was; dat er variabelen waren die niet in rekening werden gebracht, zogenaamde lokaal verborgen variabelentheorie[17]. Een dertigtal jaar na de EPR-publicatie publiceerde John Stewart Bell een test om te zien of de kwantumtheorie inderdaad onvolledig was. Het hoofdbestanddeel van deze test is de zogenaamde Bell-ongelijkheid[6] of in verbeterde vorm de Clauser-Horne-Shimony-Holt (CHSH) ongelijkheid[14], afgeleid uit statistische argumenten. Dit deel volgt de discussie in [38] maar er zijn ook alternatieve formuleringen zoals Mermin's gedachte-experiment[31] of het CHSH-spel[58].

Beschouw het volgende experiment: Charlie prepareert deeltjes in een vaste twee-deeltjes-toestand naar zijn keuze en geeft dan telkens een deeltje aan Alice en een deeltje aan Bob. Zij weten niet in welke toestand de deeltjes zich bevinden. Alice en Bob hebben elk meettoestel in hun laboratorium met twee mogelijke instellingen. Voor Alice zijn dat metingen Q en R en voor Bob S en T . Beide zijn meettoestellen met twee meetkanalen zodat voor elke meting de uitkomst $+1$ of -1 kan zijn.

Even ter verduidelijking: in EPR gaan ze er van uit dat de interne vrijheidsgraden van de deeltjes worden vastgelegd zodra de deeltjes in een toestand zijn geprepareerd. Zie het alsof de deeltjes een intern protocol meekrijgen dat zegt “antwoord x als X gevraagd wordt” waarbij X een meetinstelling is en x een mogelijk meetresultaat. Dat protocol, de “verborgen variabele”, reist mee met de deeltjes naar Alice en Bob. Dit is wat in het EPR gedachte-experiment “fysisch realisme” wordt genoemd: iets dat gemeten kan worden moet al bestaan en bepaald zijn vooraleer de meting wordt uitgevoerd. Dit is één van de basisaannames in het EPR gedachte-experiment. Een andere belangrijke aanname is het principe van lokaliteit dat zegt dat causale gebeurtenissen in elkaars

3. Parameterschatten en modelinferentie

lichtkegel moeten zitten, *i.e.* geen superluminale communicatie etc. De combinatie van deze twee aannames is ook bekend als het lokaal realisme.

Terug naar Alice en Bob: zij kiezen hun meetstand⁶ willekeurig en tegelijkertijd zodra ze hun deeltje ontvangen. Dit laatste wilt zeggen dat beide metingen niet in elkaars lichtkegel zitten en er dus “geen tijd is” om informatie van Alice haar meting tot bij Bob te krijgen voor dat hij zijn meting uitvoert en vice versa. Definieer nu de volgende grootheid:

$$QS + RS + RT - QT = (Q + R)S + (R - Q)T. \quad (3.28)$$

Aangezien de meetuitkomsten ± 1 zijn kan Alice voor voor Q en R enkel $(q, r) \in \{(+1, +1), (+1, -1), (-1, +1), (-1, -1)\}$ observeren waaruit volgt dat $(Q + R)S = 0$ of $(R - Q)T = 0$. Bijgevolg moet het linkerlid gelijk zijn aan ± 2 . Afhankelijk van de preparatie zal de kans dat de deeltjes bepaalde waarden hebben gegeven zijn door $P(q, r, s, t)$. Bijvoorbeeld als het deeltje zodanig geprepareerd is dat $(q, r, s, t) = (+1, -1, +1, -1)$ dan is $P(+1, -1, +1, -1) = 1$ en zijn de andere kansen gelijk aan 0. De verwachtingswaarde van bovenstaande grootheid kan nu berekend worden als

$$\begin{aligned} \mathbb{E}[QS + RS + RT - QT] &= \sum_{q,r,s,t \in \{+1,-1\}} P(q, r, s, t)(qs + rs + rt - qt) \\ &\leq \sum_{q,r,s,t} P(q, r, s, t) \times 2 = 2. \end{aligned}$$

Wegens de lineariteit van de verwachtingswaarde kan dit ook uitgedrukt worden als:

$$\mathbb{E}[QS + RS + RT - QT] = \mathbb{E}[QS] + \mathbb{E}[RS] + \mathbb{E}[RT] - \mathbb{E}[QT] \leq 2.$$

Dit is de CHSH ongelijkheid en is helemaal afgeleid op basis van statistische argumenten. Stel nu dat Charlie de deeltjes prepareert in een Bell-toestand $|\psi\rangle = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle)$. Alice en Bob kiezen hun metingen in de volgende richtingen:

$$Q = Z_1 \quad R = X_1 \quad S = \frac{-Z_2 - X_2}{\sqrt{2}} \quad T = \frac{Z_2 - X_2}{\sqrt{2}}$$

De verwachtingswaarden geven nu echter:

$$\mathbb{E}[QS] = \frac{1}{\sqrt{2}} \quad \mathbb{E}[RS] = \frac{1}{\sqrt{2}} \quad \mathbb{E}[RT] = \frac{1}{\sqrt{2}} \quad \mathbb{E}[QT] = \frac{1}{\sqrt{2}}.$$

⁶Bijvoorbeeld de richting waarin ze hun een spin willen meten.

3. Parameterschatten en modelinferentie

Voor een precieze uitwerking zie appendix A.1. Bijgevolg

$$\mathbb{E}[QS + RS + RT - QT] = \mathbb{E}[QS] + \mathbb{E}[RS] + \mathbb{E}[RT] - \mathbb{E}[QT] = 2\sqrt{2}.$$

Dit is een opmerkelijk resultaat, zeker omdat het ook experimenteel bevestigd is[5, 22, 44, 57]. Ten eerste impliceert het dat er iets fout is met de aanname van lokaal realisme. De bovenstaande afleiding vertrekt immers uit die aanname en bestaat uit logische stappen zonder extra aannames. Dat het fysisch realisme niet geldt, m.a.w. dat bepaalde grootheden of eigenschappen niet vooraf bepaald zijn in deeltjes uit zich in de essentiële willekeur van kwantum-metingen. Dat de lokaliteit niet geldt uit zich in verstrengelde toestanden waarbij een meting op een subsysteem de gehele toestand verandert, ook al zitten de deeltjes niet meer in elkaars lichtkegel. Ten tweede toont het aan dat verstrengeling gebruikt kan worden als een “hulpbron” om strategieën of oplossingen te ontwikkelen die beter werken dan klassieke oplossingen. Het CHSH-spel toont dit duidelijk[58]: een klassieke strategie geeft een maximale winkans van 0.75 terwijl het gebruik van een Bell-toestand een winkans van 0.85 mogelijk maakt. Verstrengeling is daarom ook de bron van de rekenkracht van kwantumcomputers en kwantumalgoritmes.

Bovenstaand resultaat is wat er bedoeld wordt met dat verstrengeling correlaties mogelijk maakt die klassieke correlaties overtreffen. Een interessante vraag die zich nu stelt is hoeveel “extra” kwantummechanische correlaties zijn er mogelijk? Dit is voorlopig nog een open vraag voor algemene toestanden maar voor bovenstaande Bell-toestand is het Tsirelson gelukt om een bovengrens af te leiden voor de kwantummechanische CHSH-ongelijkheid: precies $2\sqrt{2}$ [13, 49, 54]. Met andere woorden, voor verstrengelde bipartiete toestanden is de schending van de Bell-ongelijkheid maximaal.

Er bestaat nog een alternatief experiment, het zogenaamde GHZ-experiment van Mermin, bedacht door Greenberger, Horne en Zeilinger voor vier deeltjes en vervolgens door Mermin vereenvoudigd naar drie deeltjes[20, 32]. Het groot verschil met het Bell-experiment is dat het geen statistisch ensemble aan metingen nodig heeft. Één meting volstaat om verborgen variabelen theorie te ontkrachten. Dit is ook al experimenteel bevestigd[41].

Een laatste belangrijke punt: een nodige voorwaarde voor het observeren van verstrengelingsfenomenen is dat er metingen verricht kunnen worden in verschillende richtingen. De exacte oriëntatie van de metingen kan dan verder geoptimaliseerd worden om bv. een maximale schending van de Bell-ongelijkheid te genereren. Als Alice en Bob allebei telkens in dezelfde richting hadden gemeten dan hadden ze nooit een schending van de Bell-test geobserveerd. Goede kwantumtomografie bestaat dus uit zowel een goede reconstructieprocedure als een goede meetprocedure die gevoelig is voor de niet-klassieke aspecten van de kwantumrealiteit.

3.4.2. Tensornetwerken en modelinferentie

Tot slot nog een kort woordje uitleg over het toepassen van modelinferentie op de tensornetwerken. Er is geen technische reden waarom de theorie van modelinferentie niet toegepast zou kunnen worden op tensornetwerken; wel moeten sommige dingen met wat meer voorzichtigheid behandeld worden.

Netwerkstructuur en verkenning van de Hilbertruimte

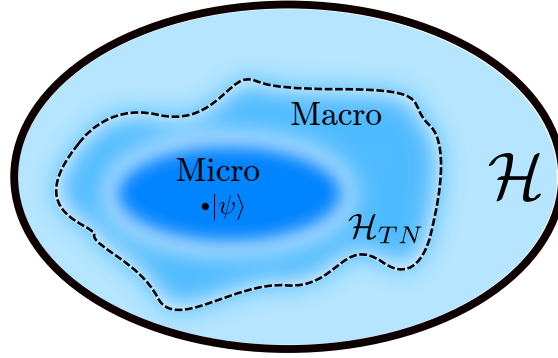
In het tweede hoofdstuk werd al vermeld dat tensornetwerken verschillende modellen op zich zijn met verschillende capaciteiten. Een eerste vraag die daarmee opkomt is: wat zijn de parameters van het model? De tensorcomponenten zijn een vanzelfsprekende groep. Deze zijn te beschouwen als microscopische parameters die een specifieke toestand beschrijven maar beperkt zijn in hun capaciteiten door de structuur van het tensornetwerk. Een tweede groep, de zogenaamde macroscopische of topologische parameters, is verantwoordelijk voor de topologische structuur van het netwerk. De topologische parameters kunnen opgedeeld worden in twee groepen: de bindingsparameters en de tensorparameters. De bindingsparameters bepalen de contractie-indices en hun dimensies terwijl de tensorparameters zeggen welke tensoren er aanwezig zijn. Merk op dat deze opsplitsing eigenlijk overbodig is aangezien de tensorparameters uit de bindingsparameters volgen. Contractie indices moeten immers altijd twee tensoren verbinden⁷ en bijgevolg zijn de aanwezige tensoren hieruit af te leiden. Paragraaf 3.4.2 illustreert hoe de topologische structuur een deelruimte van de Hilbertruimte $\mathcal{H}$ bepaalt dat vervolgens verkend kan worden door de microscopische parameters.

Technische moeilijkheden

Een uitdaging met de verkenning van de Hilbertruimte zoals hierboven is dat de straf in het Akaike informatie criterium mogelijks niet sterk genoeg is om ontsporing van de netwerkstructuur te voorkomen. Ontsporing in deze context betekent ofwel een evolutie naar modellen met steeds meer parameters, *i.e.* overfitten, ofwel convergentie naar naar de 2^N -dimensionale tensor die de exacte veel-deeltjestoestand beschrijft. Een andere strafmaat kiezen is een optie maar dan gaat het verband met de Kullback-Leiblerdivergentie verloren[52]. Om dit probleem te voorkomen zullen de verschillende structuren voorlopig handmatig geïmplementeerd worden.

Een andere uitdaging is het feit dat de microscopische modelparameters –de tensorcomponenten– complexe getallen zijn en bijgevolg aangepaste optimalisatiealgoritmes nodig zijn. De complexe getallen zijn niet meer geordend dus het idee van een parameter

⁷Er bestaan ook hyperindices waarbij een contractie tussen meer dan twee tensoren plaatsvindt maar deze kunnen equivalent geformuleerd worden met gewone indices tussen verschillende (verborgen) Kronecker-deltas[53].



Figuur 3.1.: Macroscopische of topologische parameters bepalen de capaciteiten van een tensor-netwerk om kwantumtoestanden te beschrijven. Dit uit zich als een deelruimte $\mathcal{H}_{TN}$ van de exacte Hilbertruimte $\mathcal{H}$. De microscopische parameters kunnen de toestanden binnen die deelruimte exact beschrijven. Een goede reconstructieprocedure is dan in staat om de macroscopische parameters te vinden zodat $\mathcal{H}_{TN}$ zo klein mogelijk is maar toch nog toelaat om microscopische parameters te vinden die een toestand geven die de metingen goed beschrijft. De kleurgradiënt gaat van meer verstrengeling (licht) naar minder verstrengeling (donker).

verhogen of verlagen bestaat niet meer. Bovendien zijn de meetresultaten invariant onder een globale fase. Hierdoor is het dus mogelijk om verschillende parameters te vinden die even goed zijn en als zulke situaties te vaak voorkomen in bv. een Monte-Carlo simulatie dan kan het de verkenning van de Hilbertruimte verstoren. Op numeriek vlak bestaat ook het gevaar dat functies $f : \mathbb{C} \rightarrow \mathbb{R}$ wegens de eindige precisie toch nog een klein imaginair deel produceren in hun eindresultaat terwijl dat wiskundig niet mogelijk is. Dit vereist dus ook extra aandacht tijdens het implementeren.

Zolang de kostfunctie, de aannemelijkheid in dit geval, reëel is blijft het mogelijk om optimalisatiealgoritmes te implementeren. Voor het complexe optimalisatieprobleem bestaan er enkele oplossingen. Een eerste optie is om het probleem te splitsen in een reëel deel en een imaginair deel waarop dan de gewone optimalisatiealgoritmes toegepast kunnen worden. Dit heeft echter als nadeel dat de link met in de theorie afgeleide formules deels verloren gaat. Als gradiënt-gebaseerde algoritmes gewenst zijn dan is het eventueel mogelijk om deze te herformuleren in termen van de $\mathbb{C}\mathbb{R}$ -calculus[28]. Een laatste optie is het implementeren van Monte-Carlo algoritmes die geen gebruik maken van afgeleiden. In dit geval moeten de stapjes van de toevalsbeweging (random walk) slim gekozen worden. Het volgende hoofdstuk zal de details achter de implementatie verder bespreken.

4. Resultaten en bespreking

Voor tensornetwerken bestaan er verschillende softwarebibliotheken in verschillende talen. Een overzicht is te vinden in [53]. Echter niet alle bibliotheken zijn met hetzelfde doel in gedachte ontworpen. Er zijn bibliotheken waarvan de focus ligt op artificiële intelligentie en gerelateerde toepassingen en er zijn bibliotheken gemaakt voor kwantummechanische toepassingen. Voor deze thesis is er gekozen om te werken met quimb, een Pythonbibliotheek voor tensornetwerken in een kwantummechanische context[19]. De belangrijkste reden om voor deze bibliotheek te kiezen is de beschikbare flexibiliteit in het ontwerpen van netwerken. De flexibiliteit in quimb laat ook toe om eigen algoritmes voor het parameterschatten en modelinferentie te implementeren.

Er zijn twee verschillende implementaties. Om vertrouwd te geraken met de methodes uit het parameterschatten en modelinferentie is er als korte oefening een implementatie voor klassieke munten gemaakt. De tweede implementatie is effectief voor tensornetwerken ontworpen en de twee zijn niet compatibel. Broncode voor beide implementaties is te vinden onder het aanvullend materiaal dat mee is ingediend.

4.1. Klassieke munten

Deze oefening bestond er uit om de parameters van een systeem van vier gecorreleerde munten te achterhalen. De meetdata ter beschikking bestond uit drie datasets met waarnemingen van de vorm¹ $HHHT, TTHH, \dots$ “data 1A” bestaat uit 100 observaties en “data 1B” bestaat uit 500 observaties, elk van dezelfde munten. “data 2” bestond 500 observaties maar van munten met verschillende correlaties.

4.1.1. Implementatie

Zoals gezien in tabel 3.1 is de manier waarop de modelparameters geassocieerd worden met de observatiekansen $P_{HHHH}, P_{HHHT}, \dots$ afhankelijk van het model dat er gebruikt wordt. Echter wanneer er meer dan twee munten zijn dan is het mogelijk dat er een combinatie van verschillende modellen is. Beschouw ter illustratie het volgende niet triviaal voorbeeld:

¹Heads (H), Tails (T)

4. Resultaten en bespreking

Munten 0 en 2 zijn onafhankelijk en identiek. Munten 1, 3 en 4 zijn gecorreleerd. Munt 5 is opnieuw onafhankelijk. Tot slot zijn munten 6 en 7 gecorreleerd en zijn ze identiek aan munten 8 en 9, die dus ook gecorreleerd zijn.

Dit is een globaal model dat bestaat uit zes deelsystemen. Elk deelsysteem wordt beschreven door een verschillend aantal parameters zoals in tabel 3.1. De uitdaging is om, gegeven een uitkomst S_i , bv. $HHHHTTHHH$, de relevante parameters te combineren tot een product. Het globaal model is geïmplementeerd als een woordenboek (dictionary) op de volgende wijze:

$$\text{model} = \left\{ \begin{array}{ll} (0,) & : ((0,), (2,)), \\ (1, 3, 4) & : ((1, 3, 4),), \\ (5,) & : ((5,),) \\ (6, 7) & : ((6, 7), (8, 9)) \end{array} \right\} \quad (4.1)$$

De structuur van een woordenboek is $\{\text{sleutel} : \text{waarde}\}$. De syntaxis is als volgt: elk deelsysteem wordt opgeslagen in een tuple bestaande uit de munt-indices² die behoren tot dat deelsysteem. Verschillende deelsystemen die identiek zijn aan elkaar worden dan opnieuw opgeslagen in een tuple. Van alle verschillende deelsystemen wordt er telkens één aangeduid om zijn groep van identieke deelsystemen te indexeren en wordt daarbij gebruikt als sleutel van de woordenboek.

Op basis van het bovenstaande globaal model moeten enkel de volgende vrije parameters in rekening gebracht worden:

$$\{(\theta_0)_{(0,),(2,)}^2, (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7)_{(1,3,4)}, (\theta_8)_{(5,)}, (\theta_9, \theta_{10}, \theta_{11})_{(6,7),(8,9)}^2\}. \quad (4.2)$$

Het berekenen van de multinomiaalverdeling is dan als volgt geformuleerd. Gegeven een reeks onafhankelijke observaties y van een model $\mathcal{M}$ met parameters Θ , dan is de kans voor het observeren van die reeks gegeven door:

$$P(y|\mathcal{M}, \Theta) = \mathcal{N} \times [\Theta_{S_0}]^{n_{S_0}} \times \dots \times [\Theta_{S_P}]^{n_{S_P}}, \quad (4.3)$$

waarbij $\Theta_{S_i} = (\theta_0 \times \dots \theta_r)_{S_i}$, m.a.w. het product van de parameters gerelateerd aan de observatie S_i . De juiste Θ wordt geconstrueerd op basis van de sleutels in het woordenboek. Het algoritme doorloopt alle individuele deelsystemen waarbij de munt-indices zeggen naar welke letters van S_i er gekeken moet worden. Bijvoorbeeld voor het deelsysteem $(1, 3, 4)$ kijkt het naar de tweede, vierde en vijfde letter van S_i . Op basis van welke letters er zijn geregistreerd kiest het dan de juiste vrije parameter die hoort bij dat deelsysteem.

²Python indices beginnen bij 0.

4. Resultaten en bespreking

De rest van de implementatie bestaat nu gewoon uit het toepassen van de theorie uit hoofdstuk drie. De optimale parameters werden berekend op basis van een numerieke optimalisatiefunctie van Scipy met als voorwaarde voor de vrije parameters $\sum_i \theta_i < 1$ voor elk uniek deelsysteem.

4.1.2. Resultaten

Er werden drie verschillende gecorreleerde modellen met elkaar vergeleken, telkens werden de Akaike gewichten en kansen berekend. De resultaten met betrekking tot modelinferentie staan weergegeven in tabel 4.1. De notatie $\{ij, kl\}$ betekent munten i, j gecorreleerd met elkaar en munten k, l gecorreleerd met elkaar.

Tabel 4.1.: Modelinferentie voor klassieke munten

data 1A	w_j	P_j	data 1B	$w_j(\times 10^3)$	P_j	data 2	$w_j (10^3)$	P_j
$\{12, 34\}$	4237.98	0.28	$\{12, 34\}$	449	0.29	$\{12, 34\}$	642	0.32
$\{13, 24\}$	4524.02	0.30	$\{13, 24\}$	451	0.29	$\{13, 24\}$	761	0.37
$\{14, 23\}$	6491.72	0.43	$\{14, 23\}$	648	0.42	$\{14, 23\}$	633	0.31

De eerste twee tabellen geven duidelijk een voorkeur voor het derde model. De derde tabel verkiest het tweede model voor “data 2” maar is minder overtuigend dan in de andere gevallen.

De geschatte parameters voor de gekozen modellen zijn weergegeven in tabel 4.2.

Tabel 4.2.: De geschatte optimale parameters.

Model	θ_1	θ_2	θ_3	θ'_1	θ'_2	θ'_3	
$\{14, 23\}$	0.49	0.18	0.11	0.13	0.57	0.22	data 1A
	0.49	0.16	0.11	0.16	0.52	0.22	data 1B
$\{13, 24\}$	0.16	0.52	0.22	0.08	0.32	0.12	data 2

De eerste twee rijen komen grotendeels overeen wat in de lijn van verwachtingen zit aangezien het al bekend was dat “data 1A” en “data 1B” afkomstig waren van dezelfde munten. “data 2” is duidelijk gemaakt op basis van verschillende munten.

Tot slot is er nog de gemeenschappelijke informatie (mutual information) tussen gecorreleerde munten berekend. Dit is gedefinieerd als de Kullback-Leiblerdivergentie tussen de gezamenlijke kans van twee kansvariabelen de marginale kansen[52]. Met andere woorden het zegt ons hoeveel informatie er verloren gaat door foutief te veronderstellen dat de munten onafhankelijk zijn:

$$I(\text{munt1}, \text{munt2}) = D_{KL} \left(p_{ij} | p_i^{(1)} p_j^{(2)} \right) = \sum_{i=H,T} \sum_{j=H,T} p_{ij} \log \left(\frac{p_{ij}}{p_i^{(1)} p_j^{(2)}} \right) \quad (4.4)$$

4. Resultaten en bespreking

Het is duidelijk dat als de munten daadwerkelijk onafhankelijk zijn er geen informatie verloren gaat.

De resultaten zijn weergegeven in tabel 4.3 en brengen iets interessants aan het licht. Voor munten die werkelijk gecorreleerd zijn is de gemeenschappelijke informatie significant terwijl voor niet gecorreleerde munten het dichterbij nul zit. Opmerkelijk is dat voor “data 2” de gemeenschappelijke informatie bevestigt dat munten 1 en 3 gecorreleerd maar munten 2 en 4 niet.

Tabel 4.3.: De gemeenschappelijke informatie tussen gecorreleerde munten.

Mutual information of:	(1,4)	(2,3)	(1,3)	(2,4)
Data 1A	0.39	0.68	-0.09	-0.08
Data 1B	0.37	0.36	4.64×10^{-16}	-5.78×10^{-17}
Data 2	2.88×10^{-16}	1.24×10^{-17}	0.36	1.31×10^{-4}

Uiteindelijk werd de geheime sleutel waarmee de data werd gegenereerd vrijgegeven en daaruit bleek dat bovenstaande resultaten juist zijn. De Akaike gewichten hebben de juiste modellen aangeduid en de parameters werden juist geschat. “data 2” bleek uiteindelijk een strikvraag te zijn waarbij munten 2 en 4 inderdaad niet gecorreleerd waren.

4.2. Tensornetwerken

Verschillende functies voor het construeren en verwerken van tensornetwerken zijn geïmplementeerd maar de belangrijkste is de optimalisatiefunctie. Tensornetwerken in quimb zijn niet meer dan een collectie van tensoren waarbij indices die twee keer voorkomen automatisch geïnterpreteerd worden als contracties waarbij deze vaak automatisch geoptimaliseerd worden om grote tensoren tijdens berekeningen te vermijden. Omdat de gerichte verstrengeling matrixproduct-toestanden compatibel zijn met met gewone matrixproduct-toestanden kunnen veel vergelijkingen geïmplementeerd worden zoals ze zijn beschreven in hoofdstuk 2.

We introduceren de conventie om GVMPs voor te stellen door Griekse hoofdletters, *e.g.* $|\Psi\rangle = |\psi_{GVMP}\rangle$, en operatoren als $\mathcal{O}$. Kansen worden ofwel berekend als $|\langle\Phi|\Psi\rangle|^2$ of in geval van een projectie-MPO als $\langle\Psi|\mathcal{O}|\Psi\rangle = |\langle\Psi|\mathcal{O}_\Psi\rangle|^2$ waarbij $|\mathcal{O}_\Psi\rangle$ de genormeerde projectie is. Voor het berekenen van de multinomiaalverdeling gebruiken we projectie-MPOs, bijvoorbeeld voor een observatie “+z+z-x” construeren we eerst de MPO voor een projectie op de |+z+z-x>-toestand en vervolgens berekenen we de kans als $P(+z+z-x|\Psi) = \langle\Psi|\mathcal{O}_{+z+z-x}|\Psi\rangle$. Deze procedure werd dan herhaald voor elke geobserveerde meetuitkomst. Concreet wordt er gewerkt met het logaritme van de aannemelijkheidsfunctie om de numerieke stabiliteit te verbeteren.

4. Resultaten en bespreking

Het algoritme voor het optimaliseren van tensornetwerken is gebaseerd op een Monte-Carlo procedure over de complexe parameterruimte. Een samenvatting van de procedure is weergegeven in algoritme 1. De initiële tensornetwerken zijn willekeurig geïnitieerd. De keuze voor de perturbatie is geïnspireerd op [55] en het stelt de vier richtingen van de complexe ruimte voor. Voor elke tensorcomponent wordt een uniforme trekking gedaan uit die vier mogelijkheden. Nadat alle componenten zijn bijgewerkt wordt $|\Psi\rangle$ opnieuw genormeerd opdat het terug een goede kwantumtoestand is. Er is nog een schalingsfactor geïmplementeerd die bij elke aanvaarding de perturbatie verhoogt met 15% en verlaagt met 15% indien de suggestie verwerpt wordt. Deze keuze resulteert voor de huidige implementatie in een acceptatieverhouding van ongeveer 50% voor de gesuggereerde perturbaties.

De eerste if-test is simpelweg de maximalisatie van de aannemelijkheidsfunctie. De tweede if-test is het zogenaamde Metropolis-Hastings algoritme waarbij de Akaike acceptatieverhouding verg. (3.12) berekend wordt om het algoritme alsnog een kans te geven om de nieuwe schatting te aanvaarden. Als beide testen mislukken wordt de perturbatie verworpen. Merk dat de performantie objectief te kunnen beoordelen wordt na elke iteratie de fideliteit tussen de geschatte kwantumtoestand Ψ en de echte kwantumtoestand ψ berekend. Deze waarde is echter niet betrokken bij de optimalisatieprocedure omdat het in de praktijk onbekend is.

Algorithm 1 Monte Carlo procedure voor het parameterschatten van matrixproducttoestanden

```

1: procedure MAXIMIZE  $L(\Psi|y)$ (tn= $\Psi$ , data= $y$ , iters= $N$ , truth= $\psi$ )
2:    $L \leftarrow L(\Psi|y)$  ▷ Aannemelijkheidfunctie
3:   for  $i \leq N$  do
4:      $\Psi_{new} \leftarrow \text{rng}(\Psi, \{1, i, -1, -i\})$  ▷ Perturbatie
5:      $L_{new} \leftarrow L(\Psi_{new}|y)$ 
6:     if  $L_{new} \geq L$  then
7:        $\Psi \leftarrow \Psi_{new}$ 
8:        $L \leftarrow L_{new}$ 
9:     else if  $\text{rng}(0, 1) \leq \exp(L_{new} - L + K - K_{new})$  then ▷ Akaike
10:       $\Psi \leftarrow \Psi_{new}$ 
11:       $L \leftarrow L_{new}$ 
12:     else
13:        $\emptyset \leftarrow \Psi_{new}$ 
14:     end if
15:     Fidelities  $\leftarrow F(\psi, \Psi)$ 
16:   end for

```

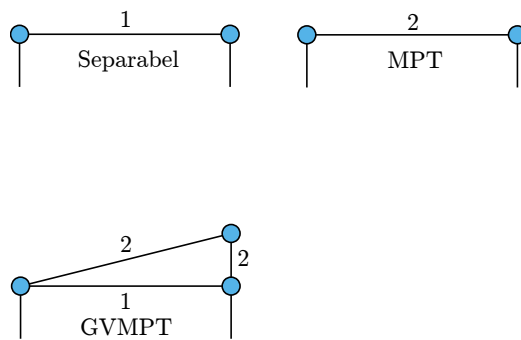
4.2.1. Resultaten

Er zijn maar een beperkt aantal experimenten uitgevoerd op de tensornetwerken, voornamelijk wegens tijdsgebrek. Het algoritme vertraagt significant bij de overgang van twee deeltjes naar drie deeltjes. De hoofdoorzaak ligt vermoedelijk in de implementatie van de functie die de kansen berekent. Er is een bug in die functie waardoor tensoren “inplace” worden aangepast tijdens de berekeningen terwijl dit niet zou mogen gebeuren. Als oplossing wordt er een kopie gemaakt van het netwerk maar dit is in computationele termen een dure operatie. Een tweede oorzaak is het feit dat het bovenstaande algoritme niet ingebouwd is in quimb en bijgevolg niet kan profiteren van ingebouwde optimalisaties. Toch volstaan de uitgevoerde experimenten om aan te tonen dat het algoritme wel degelijk werkt.

Het eerste experiment bestond er uit om de twee deeltjes Bell-toestand $|\psi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ te reconstrueren met de verschillende netwerken weergegeven in Figuur 4.1. Metingen werden evenveel keer uitgevoerd in alle drie richtingen. De fideliteiten van de verschillende netwerken zijn weergegeven in Figuur 4.2. Zoals verwacht is de separabele toestand niet in staat om de Bell-toestand te beschrijven. De GVMPT en MPT gaan allebei boven 90% fideliteit.

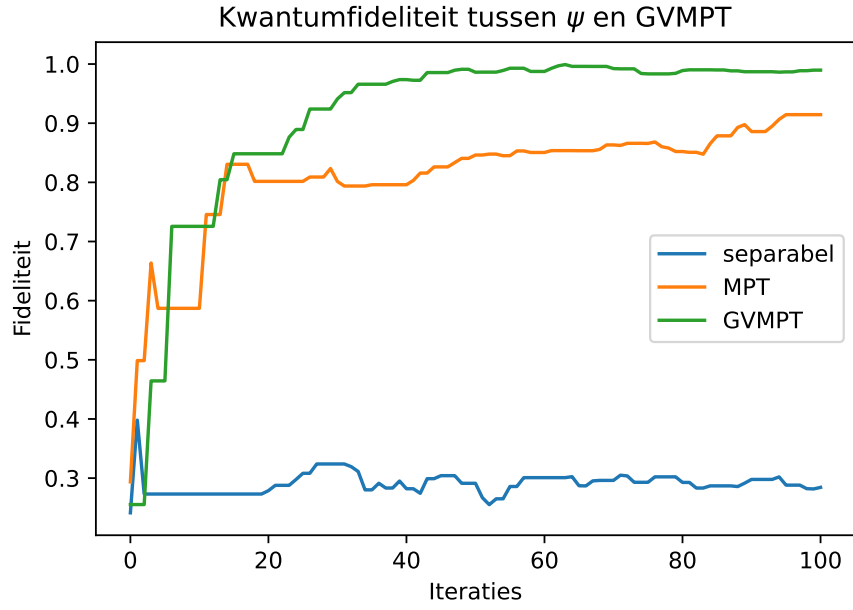
Een tweede experiment keek naar het effect van het aantal metingen op de fideliteit van de MPT en de GVMPT, de resultaten zijn weergegeven in Figuren 4.3 en 4.4. Voor 100 metingen gaan beide netwerken boven de 90%. Voor minder metingen lijkt de GVMPT beter dan de MPT maar er zijn niet genoeg experimenten uitgevoerd om te concluderen GVMPT consistent beter is.

Een laatste experiment keek naar het effect van de diversiteit in metingen op de fideliteit. In eerste instantie zijn er enkel metingen in de z-richting uitgevoerd en in de tweede instantie in zowel de x-richting als de z-richting. De resultaten zijn weergegeven in Figuur 4.5 en Figuur 4.6. De impact is duidelijk, voor enkel metingen in de z-richting wordt 60% fideliteit niet gehaald. Introductie van metingen in x-richting verbetert de fideliteit tot meer dan 80%.

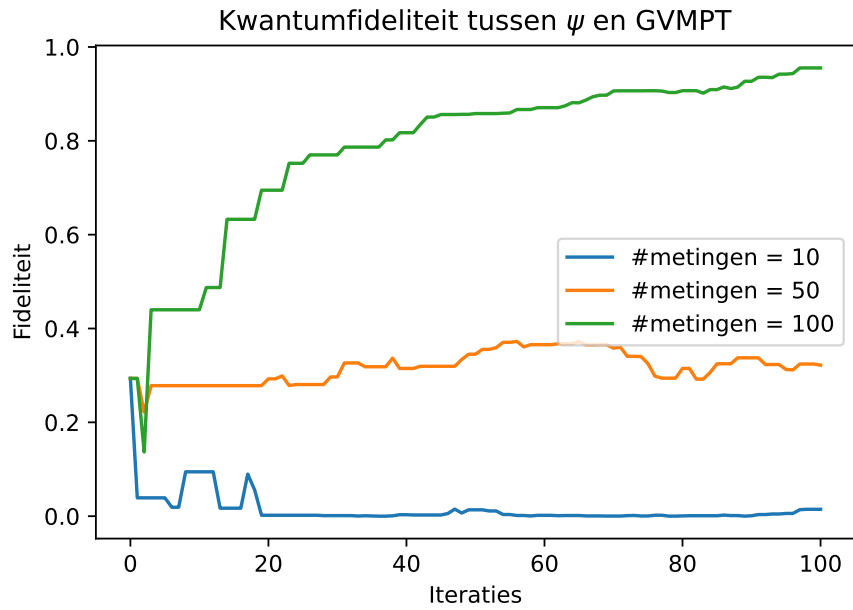


Figuur 4.1.: De verschillende gebruikte netwerken. De getallen geven de bindingsdimensie weer.

4. Resultaten en bespreking

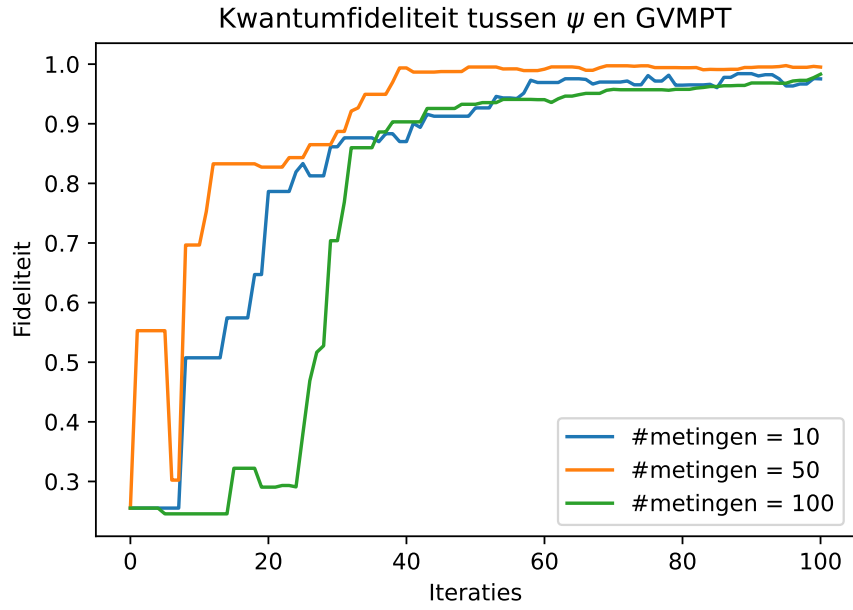


Figuur 4.2.: Vergelijking van de fideliteit tussen een separabele MPT, de exacte MPT en een GVMPT variant daarop. 100 metingen van elke meetconfiguratie.

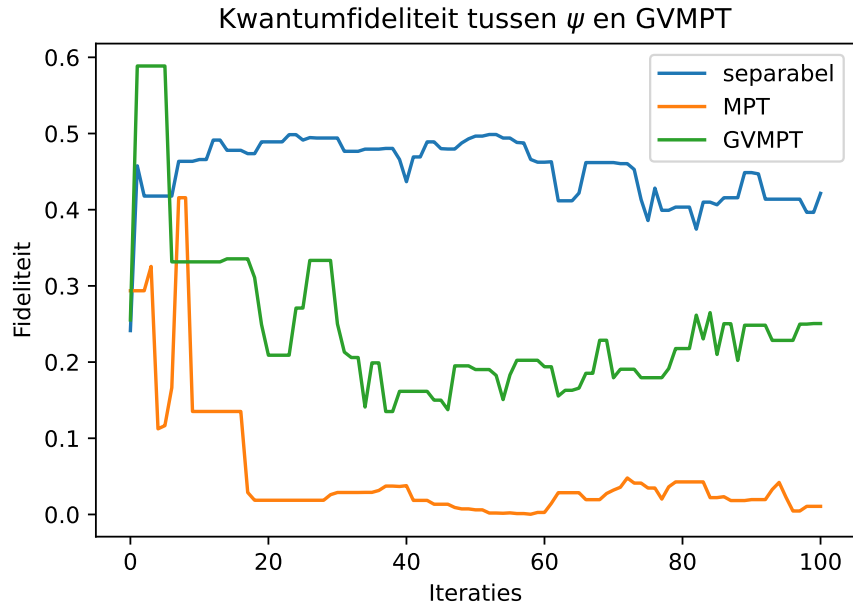


Figuur 4.3.: Vergelijking van de fideliteit voor eenzelfde MPT maar voor verschillende observaties.

4. Resultaten en bespreking

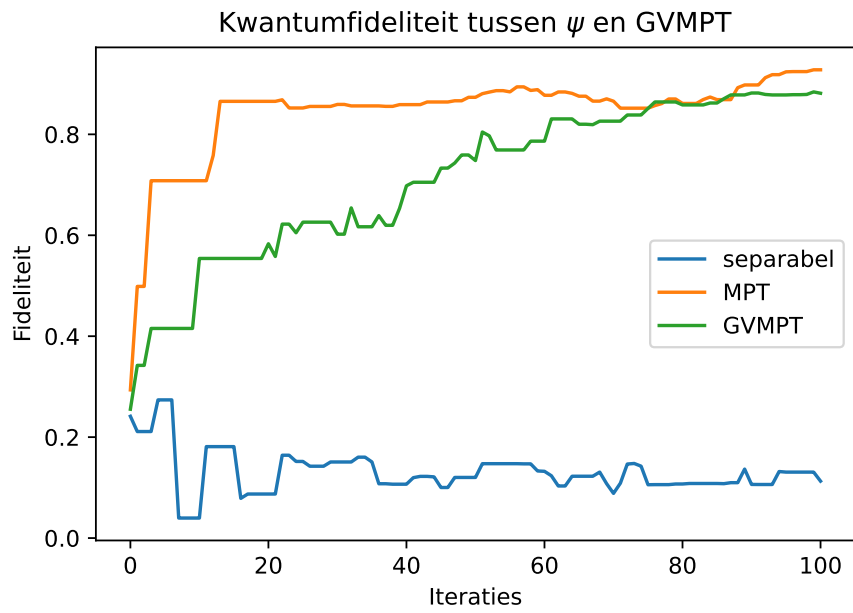


Figuur 4.4.: Vergelijking van de fideliteit voor eenzelfde GVMPT maar voor verschillende observaties.



Figuur 4.5.: Vergelijking van de fideliteit tussen een separabele MPT, de exacte MPT en een GVMPT variant daarop. 100 metingen in de z-richting.

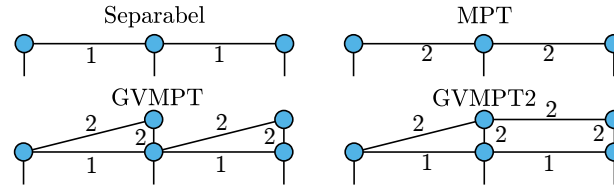
4. Resultaten en bespreking



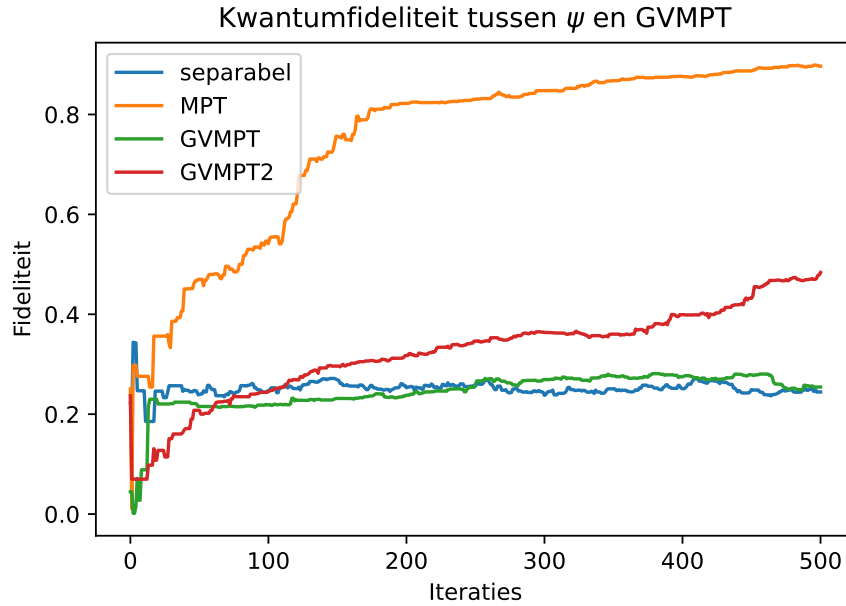
Figuur 4.6.: Vergelijking van de fideliteit tussen een separabele MPT, de exacte MPT en een GVMPT variant daarop. 100 metingen in de x-richting en de z-richting.

4. Resultaten en bespreking

Een deel van de experimenten werd vervolgens herhaald maar dan voor de driedeeltjes Bell-toestand $|\psi\rangle = \frac{1}{\sqrt{2}}(|000\rangle + |111\rangle)$. Een vergelijking van de performantie van de verschillende netwerken uit Figuur 4.7 is te zien in Figuur 4.8. Uit de resultaten blijkt op het eerste zicht dat de gewone MPT de meest performante is. Echter de conclusies zijn te voorbarig wegens de beperkte statistiek. Indien we de experimenten met verschillende hoeveelheid metingen herhalen voor de GVMPT en GVMPT2 netwerken zien we een verbetering in de fideliteit, weergegeven in Figures 4.9 en 4.10. Deze modellen zijn echter niet bijzonder interessant aangezien ze het doel van “zuinigheid” missen. Bovendien zijn er meer experimenten nodig om duidelijke conclusies te kunnen trekken.

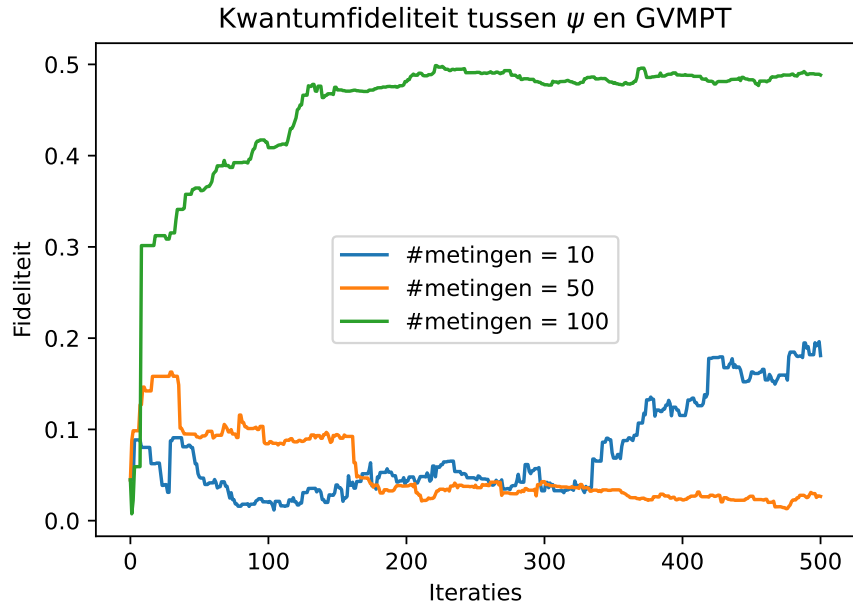


Figuur 4.7.: De verschillende gebruikte netwerken. De getallen geven de bindingsdimensie weer.

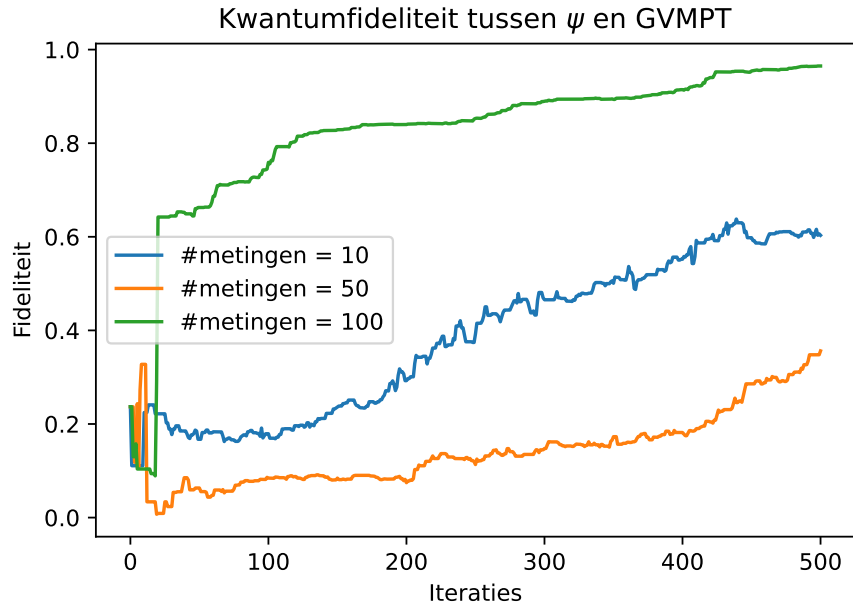


Figuur 4.8.: Vergelijking van de fideliteit tussen een separabele MPT, de exacte MPT en twee GVMPT varianten daarop. 100 metingen van elke meetconfiguratie.

4. Resultaten en bespreking



Figuur 4.9.: Vergelijking van de fideliteit voor dezelfde GVMPT maar met verschillende verschillende observaties.



Figuur 4.10.: Vergelijking van de fideliteit voor dezelfde GVMPT2 maar met verschillende verschillende observaties.

5. Conclusie

De verschillende resultaten tonen dat het algoritme wel degelijk werkt aangezien er regelmatig fideliteiten boven de 90% behaald worden. Echter zijn de uitgevoerde experimenten te beperkt, zowel in het aantal experimenten als in hun complexiteit. Tot en met drie deeltjes hebben de matrixproduct-toestanden een exacte vorm waarbij elke index dimensie twee heeft. Eender welke niet triviale uitbreiding van de standaard matrixproduct-toestand zal het aantal parameters vergroten wat juist tegengesteld is aan het doel van deze thesis. Onderzoek naar grotere systemen wordt dan weer beperkt wegens de ondermaatse efficiëntie van de implementatie. Als proof of concept kunnen de resultaten misschien wel tellen maar daarbuiten is hun nut beperkt.

5.1. Verbeteringen en vooruitzichten

Ondanks de tegenvallende resultaten is het nog geen verloren zaak. Het idee heeft potentieel maar moet dan eerst gestroomlijnd worden om goed tot uiting te komen. De code moet nagekeken worden op kostelijke implementaties of eventueel volledig herzien worden. Eens de code efficiënt werkt kan er gekeken worden naar meer ingewikkeldere experimenten waar de kracht achter het idee meer tot zijn recht komt.

De Heilige Graal blijft nog steeds de implementatie van een volledig dynamisch algoritme dat in staat is om zowel de parameters als de topologie van een netwerk te variëren. Dit vereist echter meer onderzoek naar het onder controle houden van het algoritme om problematische overfitting situaties te vermijden. Mogelijks kan er inspiratie gehaald worden uit het onderzoek naar neurale netwerken waarin ze experimenteren met zulke dynamische systemen[48].

A. Appendices

A.1. Verwachtingswaarde Bell-ongelijkheid

Ter illustratie wordt de verwachtingswaarde $\mathbb{E}[QS] = \langle \psi | QS | \psi \rangle$ met $|\psi\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$ uitgewerkt. De anderen verlopen analoog. Het gemakkelijkste is om het Kronecker-product tussen de operatoren in hun matrixvorm uit te werken waarbij

$$Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \quad X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

zodat

$$\begin{array}{c} | \\ \text{Q} \\ | \end{array} \begin{array}{c} | \\ \text{S} \\ | \end{array} \Rightarrow \begin{array}{c} | \quad | \\ \text{QS} \\ | \quad | \end{array} \Rightarrow \begin{array}{c} \text{---} \\ \text{QS} \\ \text{---} \end{array} \quad \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \otimes \frac{1}{\sqrt{2}} \begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & -1 \end{bmatrix}.$$

De verwachtingswaarde is dan gegeven door

$$\langle \psi | QS | \psi \rangle = \left(\frac{1}{\sqrt{2}} \right)^3 \begin{bmatrix} 0 & 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} -1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \end{bmatrix} = \frac{1}{\sqrt{2}}.$$

Bibliografie

- [1] Hirotogu Akaike. “Information Theory and an Extension of the Maximum Likelihood Principle”. In: *Selected Papers of Hirotogu Akaike*. Red. door Emanuel Parzen, Kunio Tanabe en Genshiro Kitagawa. Springer Series in Statistics. New York, NY: Springer New York, 1998, p. 199–213. ISBN: 978-1-4612-7248-9 978-1-4612-1694-0. DOI: [10.1007/978-1-4612-1694-0_15](https://doi.org/10.1007/978-1-4612-1694-0_15).
- [2] Hirotogu Akaike. “A New Look at the Statistical Model Identification”. In: *IEEE Transactions on Automatic Control* 19 (1974), p. 716–723.
- [3] John Aldrich. “R.A. Fisher and the Making of Maximum Likelihood 1912-1922”. In: *Statistical Science* 12.3 (sep 1997), p. 162–176. ISSN: 0883-4237, 2168-8745. DOI: [10.1214/ss/1030037906](https://doi.org/10.1214/ss/1030037906).
- [4] J.B. Altepeter, E.R. Jeffrey en P.G. Kwiat. “Photonic State Tomography”. In: *Advances In Atomic, Molecular, and Optical Physics*. Deel 52. Elsevier, 2005, p. 105–159. ISBN: 978-0-12-003852-7. DOI: [10.1016/S1049-250X\(05\)52003-2](https://doi.org/10.1016/S1049-250X(05)52003-2).
- [5] Alain Aspect, Jean Dalibard en Gérard Roger. “Experimental Test of Bell’s Inequalities Using Time-Varying Analyzers”. In: *Physical Review Letters* 49.25 (20 dec 1982), p. 1804–1807. DOI: [10.1103/PhysRevLett.49.1804](https://doi.org/10.1103/PhysRevLett.49.1804).
- [6] J. S. Bell. “On the Einstein Podolsky Rosen Paradox”. In: *Physics Physique Fizika* 1.3 (1 nov 1964), p. 195–200. DOI: [10.1103/PhysicsPhysiqueFizika.1.195](https://doi.org/10.1103/PhysicsPhysiqueFizika.1.195).
- [7] Jean Pierre Briand, red. *Atoms in Unusual Situations*. Deel 143. Nato ASI Series. Boston, MA: Springer US, 1986. ISBN: 978-1-4757-9339-0 978-1-4757-9337-6. DOI: [10.1007/978-1-4757-9337-6](https://doi.org/10.1007/978-1-4757-9337-6).
- [8] Jacob C. Bridgeman en Christopher T. Chubb. “Hand-Waving and Interpretive Dance: An Introductory Course on Tensor Networks”. In: *Journal of Physics A: Mathematical and Theoretical* 50.22 (2 jun 2017), p. 223001. ISSN: 1751-8113, 1751-8121. DOI: [10.1088/1751-8121/aa6dc3](https://doi.org/10.1088/1751-8121/aa6dc3). arXiv: [1603.03039](https://arxiv.org/abs/1603.03039).
- [9] Kenneth P. Burnham en David R. Anderson. “Multimodel Inference: Understanding AIC and BIC in Model Selection”. In: *Sociological Methods & Research* 33.2 (nov 2004), p. 261–304. ISSN: 0049-1241, 1552-8294. DOI: [10.1177/0049124104268644](https://doi.org/10.1177/0049124104268644).
- [10] Bille Chandler Carlson. *Special Functions of Applied Mathematics*. New York: Academic Press, 1977. 335 p. ISBN: 978-0-12-160150-8.
- [11] V. Černý. “Thermodynamical Approach to the Traveling Salesman Problem: An Efficient Simulation Algorithm”. In: *Journal of Optimization Theory and Applications* 45.1 (1 jan 1985), p. 41–51. ISSN: 1573-2878. DOI: [10.1007/BF00940812](https://doi.org/10.1007/BF00940812).

Bibliografie

- [12] Andrzej Cichocki. “Era of Big Data Processing: A New Approach via Tensor Networks and Tensor Decompositions”. 9 mrt 2014. arXiv: [1403.2048 \[cs\]](#).
- [13] B. S. Cirel’son. “Quantum Generalizations of Bell’s Inequality”. In: *Letters in Mathematical Physics* 4.2 (1 mrt 1980), p. 93–100. ISSN: 1573-0530. DOI: [10.1007/BF00417500](#).
- [14] John F. Clauser e.a. “Proposed Experiment to Test Local Hidden-Variable Theories”. In: *Physical Review Letters* 23.15 (13 okt 1969), p. 880–884. DOI: [10.1103/PhysRevLett.23.880](#).
- [15] Ivan-Jago Coric. “Werknota: Zoek de Correlaties”. Task report. 11 sep 2020.
- [16] Marcus Cramer e.a. “Efficient Quantum State Tomography”. In: *Nature Communications* 1.1 (dec 2010), p. 149. ISSN: 2041-1723. DOI: [10.1038/ncomms1147](#).
- [17] A. Einstein, B. Podolsky en N. Rosen. “Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?” In: *Physical Review* 47.10 (15 mei 1935), p. 777–780. DOI: [10.1103/PhysRev.47.777](#).
- [18] A. W. Glaetzle e.a. “A Coherent Quantum Annealer with Rydberg Atoms”. In: *Nature Communications* 8.1 (1 22 jun 2017), p. 15813. ISSN: 2041-1723. DOI: [10.1038/ncomms15813](#).
- [19] Johnnie Gray. *Jcmgray/Quimb*. 5 jul 2022.
- [20] Daniel M. Greenberger, Michael A. Horne en Anton Zeilinger. “Going Beyond Bell’s Theorem”. In: *Bell’s Theorem, Quantum Theory and Conceptions of the Universe*. Red. door Menas Kafatos. Fundamental Theories of Physics. Dordrecht: Springer Netherlands, 1989, p. 69–72. ISBN: 978-94-017-0849-4. DOI: [10.1007/978-94-017-0849-4_10](#).
- [21] Christian Gross en Waseem S. Bakr. “Quantum Gas Microscopy for Single Atom and Spin Detection”. In: *Nature Physics* 17.12 (12 dec 2021), p. 1316–1323. ISSN: 1745-2481. DOI: [10.1038/s41567-021-01370-5](#).
- [22] B. Hensen e.a. “Loophole-Free Bell Inequality Violation Using Electron Spins Separated by 1.3 Kilometres”. In: *Nature* 526.7575 (7575 okt 2015), p. 682–686. ISSN: 1476-4687. DOI: [10.1038/nature15759](#).
- [23] Zdeněk Hradil e.a. “3 Maximum-Likelihood Methods in Quantum Mechanics”. In: *Quantum State Estimation*. Red. door Matteo Paris en Jaroslav Řeháček. Deel 649. Lecture Notes in Physics. Berlin, Heidelberg: Springer Berlin Heidelberg, 18 aug 2004, p. 59–112. ISBN: 978-3-540-22329-0 978-3-540-44481-7. DOI: [10.1007/978-3-540-44481-7_3](#).
- [24] E. T Jaynes en G. Larry Bretthorst. *Probability Theory the Logic of Science*. Cambridge, UK; New York, NY: Cambridge University Press, 2003. ISBN: 978-0-511-06589-7.
- [25] Michael I Jordan en Alan Malek. “Reference Priors and Maximum Entropy”. In: (), p. 5.

- [26] Branimir K. Hackenberger. “Bayes or Not Bayes, Is This the Question?” In: *Croatian Medical Journal* 60.1 (feb 2019), p. 50–52. ISSN: 0353-9504. DOI: [10.3325/cmj.2019.60.50](https://doi.org/10.3325/cmj.2019.60.50). pmid: [30825279](https://pubmed.ncbi.nlm.nih.gov/30825279/).
- [27] S. Kirkpatrick, C. D. Gelatt en M. P. Vecchi. “Optimization by Simulated Annealing”. In: *Science* 220.4598 (13 mei 1983), p. 671–680. ISSN: 0036-8075, 1095-9203. DOI: [10.1126/science.220.4598.671](https://doi.org/10.1126/science.220.4598.671). pmid: [17813860](https://pubmed.ncbi.nlm.nih.gov/17813860/).
- [28] Ken Kreutz-Delgado. “The Complex Gradient Operator and the CR-Calculus”. 25 jun 2009. arXiv: [0906.4835 \[math\]](https://arxiv.org/abs/0906.4835).
- [29] S. Kullback en R. A. Leibler. “On Information and Sufficiency”. In: *Annals of Mathematical Statistics* 22 (1951), p. 79–86.
- [30] B. P. Lanyon e.a. “Efficient Tomography of a Quantum Many-Body System”. In: *Nature Physics* 13.12 (dec 2017), p. 1158–1162. ISSN: 1745-2473, 1745-2481. DOI: [10.1038/nphys4244](https://doi.org/10.1038/nphys4244).
- [31] N. David Mermin. “Hidden Variables and the Two Theorems of John Bell”. In: *Reviews of Modern Physics* 65.3 (1 jul 1993), p. 803–815. ISSN: 0034-6861, 1539-0756. DOI: [10.1103/RevModPhys.65.803](https://doi.org/10.1103/RevModPhys.65.803).
- [32] N. David Mermin. “What’s Wrong with These Elements of Reality?” In: *Physics Today* 43.6 (jun 1990), p. 9–11. ISSN: 0031-9228, 1945-0699. DOI: [10.1063/1.2810588](https://doi.org/10.1063/1.2810588).
- [33] Nicholas Metropolis e.a. “Equation of State Calculations by Fast Computing Machines”. In: *The Journal of Chemical Physics* 21.6 (1 jun 1953), p. 1087–1092. ISSN: 0021-9606. DOI: [10.1063/1.1699114](https://doi.org/10.1063/1.1699114).
- [34] Milorad Milošević. “Numerical Methods”. Lecture Notes. Universiteit Antwerpen, 2015–2016.
- [35] M. Morgado en S. Whitlock. “Quantum Simulation and Computing with Rydberg-interacting Qubits”. In: *AVS Quantum Science* 3.2 (jun 2021), p. 023501. DOI: [10.1116/5.0036562](https://doi.org/10.1116/5.0036562).
- [36] In Jae Myung. “Tutorial on Maximum Likelihood Estimation”. In: *Journal of Mathematical Psychology* 47.1 (feb 2003), p. 90–100. ISSN: 00222496. DOI: [10.1016/S0022-2496\(02\)00028-7](https://doi.org/10.1016/S0022-2496(02)00028-7).
- [37] Andrew A. Neath en Joseph E. cavanaugh. “Regression and Time Series Model Selection Using Variants of the Schwarz Information Criterion”. In: *Communications in Statistics - Theory and Methods* 26.3 (1 jan 1997), p. 559–580. ISSN: 0361-0926. DOI: [10.1080/03610929708831934](https://doi.org/10.1080/03610929708831934).
- [38] Michael A. Nielsen en Isaac L. Chuang. *Quantum Computation and Quantum Information*. 10th anniversary ed. Cambridge ; New York: Cambridge University Press, 2010. 676 p. ISBN: 978-1-107-00217-3.
- [39] Roman Orus. “A Practical Introduction to Tensor Networks: Matrix Product States and Projected Entangled Pair States”. In: *Annals of Physics* 349 (okt 2014), p. 117–158. ISSN: 00034916. DOI: [10.1016/j.aop.2014.06.013](https://doi.org/10.1016/j.aop.2014.06.013). arXiv: [1306.2164](https://arxiv.org/abs/1306.2164).

- [40] Roman Orus. “Tensor Networks for Complex Quantum Systems”. In: *Nature Reviews Physics* 1.9 (sep 2019), p. 538–550. ISSN: 2522-5820. DOI: [10.1038/s42254-019-0086-7](https://doi.org/10.1038/s42254-019-0086-7). arXiv: [1812.04011](https://arxiv.org/abs/1812.04011).
- [41] Jian-Wei Pan e.a. “Experimental Test of Quantum Nonlocality in Three-Photon Greenberger–Horne–Zeilinger Entanglement”. In: *Nature* 403.6769 (6769 feb 2000), p. 515–519. ISSN: 1476-4687. DOI: [10.1038/35000514](https://doi.org/10.1038/35000514).
- [42] D. Perez-Garcia e.a. “Matrix Product State Representations”. 14 mei 2007. arXiv: [quant-ph/0608197](https://arxiv.org/abs/quant-ph/0608197).
- [43] *Quantum Entanglement Offers Unprecedented Precision for GPS, Imaging and Beyond*. ScienceDaily. URL: <https://www.sciencedaily.com/releases/2020/04/200420104856.htm> (bezocht op 22-08-2022).
- [44] Dominik Rauch e.a. “Cosmic Bell Test Using Random Measurement Settings from High-Redshift Quasars”. In: *Physical Review Letters* 121.8 (20 aug 2018), p. 080403. DOI: [10.1103/PhysRevLett.121.080403](https://doi.org/10.1103/PhysRevLett.121.080403).
- [45] Lulu Ricketts. *Maximum Likelihood vs. Bayesian Estimation*. Medium. 23 apr 2021. URL: <https://towardsdatascience.com/maximum-likelihood-vs-bayesian-estimation-dd2eb4dfda8a> (bezocht op 16-08-2022).
- [46] Sandra Van Aert. “Kansrekening En Statistiek: Parameterschatten”. Lecture presentation (Universiteit Antwerpen).
- [47] Sandra Van Aert en P. Goos. “Kansrekening en Statistiek”. Lecture Notes. Universiteit Antwerpen.
- [48] Kenneth O. Stanley en Risto Miikkulainen. “Evolving Neural Networks through Augmenting Topologies”. In: *Evolutionary Computation* 10.2 (jun 2002), p. 99–127. ISSN: 1063-6560, 1530-9304. DOI: [10.1162/106365602320169811](https://doi.org/10.1162/106365602320169811).
- [49] William Stuckey e.a. “Why the Tsirelson Bound? Bub’s Question and Fuchs’ Desideratum”. In: *Entropy* 21.7 (15 jul 2019), p. 692. ISSN: 1099-4300. DOI: [10.3390/e21070692](https://doi.org/10.3390/e21070692).
- [50] Tai-Danae Bradley. *Understanding Entanglement With SVD*. Math3ma. URL: <https://www.math3ma.com/blog/understanding-entanglement-with-svd> (bezocht op 14-08-2022).
- [51] Jacques Tempere. “Kwantummechanica”. Lecture Notes. Universiteit Antwerpen.
- [52] Jacques Tempere. “Reconstructing Wave Functions from Measured Data”. Private communication.
- [53] *The Tensor Network*. Tensor Network. URL: <https://www.tensornetwork.org> (bezocht op 20-08-2022).
- [54] B. S. Tsirel’son. “Quantum Analogues of the Bell Inequalities. The Case of Two Spatially Separated Domains”. In: *Journal of Soviet Mathematics* 36.4 (1 feb 1987), p. 557–570. ISSN: 1573-8795. DOI: [10.1007/BF01663472](https://doi.org/10.1007/BF01663472).

Bibliografie

- [55] A. Utreras-Alarcón e.a. “Stochastic Optimization on Complex Variables and Pure-State Quantum Tomography”. In: *Scientific Reports* 9.1 (1 6 nov 2019), p. 16143. ISSN: 2045-2322. DOI: [10.1038/s41598-019-52289-0](https://doi.org/10.1038/s41598-019-52289-0).
- [56] F. Verstraete, J. I. Cirac en V. Murg. “Matrix Product States, Projected Entangled Pair States, and Variational Renormalization Group Methods for Quantum Spin Systems”. In: *Advances in Physics* 57.2 (mrt 2008), p. 143–224. ISSN: 0001-8732, 1460-6976. DOI: [10.1080/14789940801912366](https://doi.org/10.1080/14789940801912366). arXiv: [0907.2796](https://arxiv.org/abs/0907.2796).
- [57] Gregor Weihs e.a. “Violation of Bell’s Inequality under Strict Einstein Locality Conditions”. In: *Physical Review Letters* 81.23 (7 dec 1998), p. 5039–5043. DOI: [10.1103/PhysRevLett.81.5039](https://doi.org/10.1103/PhysRevLett.81.5039).
- [58] Mark M. Wilde. *Quantum Information Theory*. 2de ed. Cambridge: Cambridge University Press, 2017. ISBN: 978-1-107-17616-4. DOI: [10.1017/9781316809976](https://doi.org/10.1017/9781316809976).
- [59] Ronald J Williams. “Maximum Likelihood vs. Bayesian Parameter Estimation”. Lecture Notes.
- [60] Eberhard Zeidler e.a., red. *Oxford User’s Guide to Mathematics*. Oxford ; New York: Oxford University Press, 2004. 1284 p. ISBN: 978-0-19-850763-5.