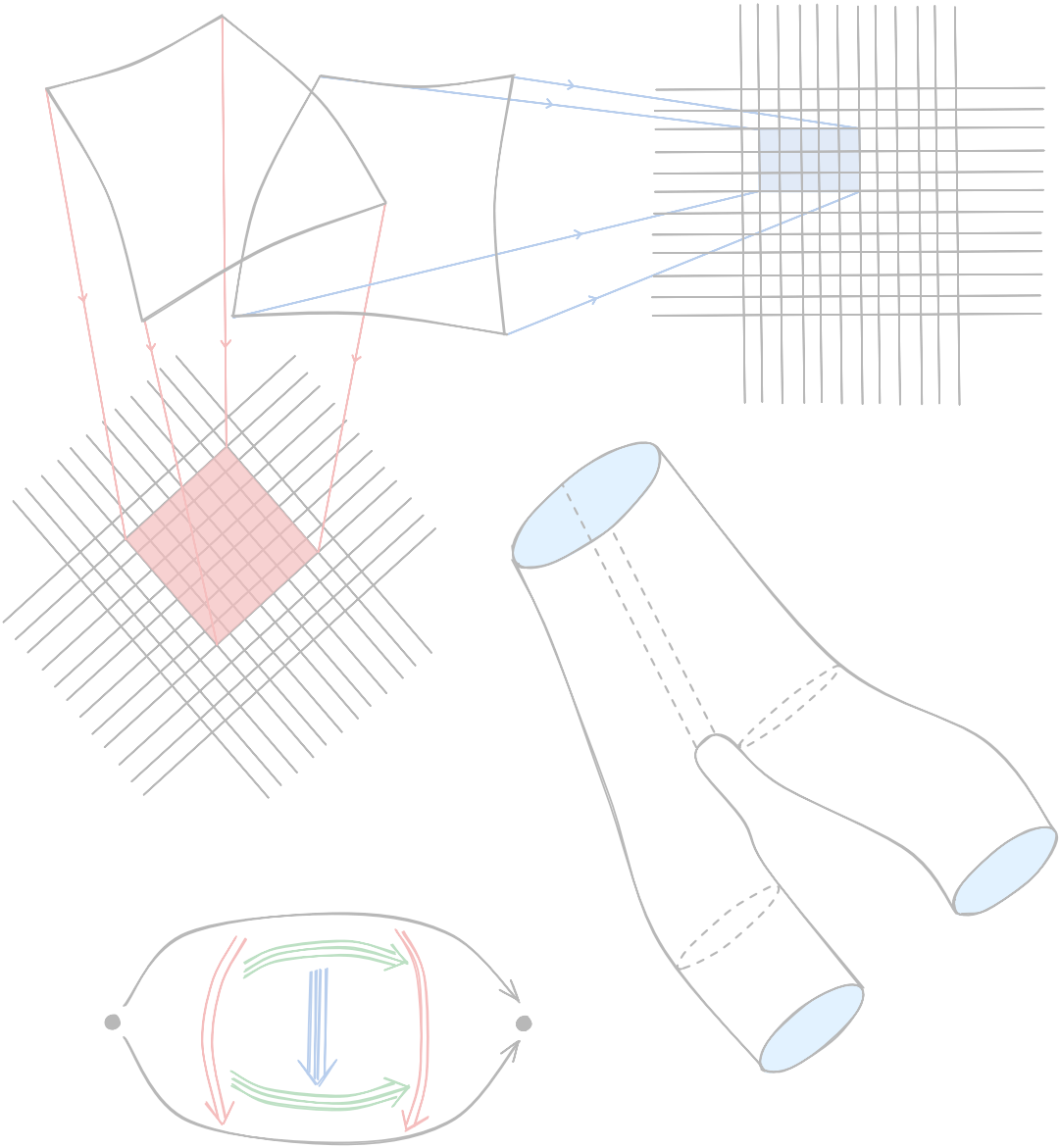
(Theoretische) fysica: Selected Topics

N. Dewolf



Inhoudsopgave

Inleiding

Begin 2026 was het idee naar boven gekomen om een boek te schrijven waarin de evolutie van de theoretische fysica werd beschreven, gaande van de klassieke mechanica van Newton tot de categorietheorie uit de hedendaagse natuurkunde. Een groot probleem bij het schrijven van dat boek was het kiezen van welke concepten al dan niet besproken konden worden. Er bestaan zo veel interessante concepten en ideeën in de natuurkunde, maar een klein boekje biedt maar plaats aan enkele, met als gevolg dat heel wat zaken moesten worden overgeslagen.

Het doel van dit boek is om hier gedeeltelijk soelaas te bieden. Enerzijds wordt een bredere waaier aan onderwerpen behandeld, maar anderzijds zullen de hoofdstukken ook wel (of eerder heel) wat technischer zijn. Waar het eerste boek echt gericht was aan mensen met wat gezonde motivatie en kennis van wiskunde en fysica uit het middelbaar, is dit boek eerder gericht aan mensen die op zoek zijn naar een stevige uitdaging en die niet bang zijn voor wat complexe formules.

Concepten die in het eerste boek "(Theoretische) fysica: van vroeger tot nu" — waar we in het vervolg als Deel 1 naar zullen verwijzen — werden geïntroduceerd, zullen in dit boek in het grijs worden aangeduid wanneer we ze (opnieuw) voor de eerste keer tegenkomen. Dit laat je toe om even terug te keren naar het vorige boekje (of eender welke andere informatiebron) en het concept even op te frissen.

Opfrisser

Net zoals in Deel 1, bespreken we hier enkele noties en notaties die belangrijk zullen zijn doorheen het boek en die mogelijks niet zo vertrouwd (meer) zijn.

- De cirkel S^1 bestaat uit alle punten die op een gelijke afstand (gebruikelijk is deze 1) van de oorsprong in het vlak $\mathbb{R}^n$ liggen. Dit wordt veralgemeend naar hogere dimensies door de n-sfeer S^n te definiëren als de verzameling van alle punten die op gelijke afstand van de oorsprong in $\mathbb{R}^{n+1}$ liggen. Zo is het boloppervlak gelijk aan S^2.
- In de kansrekening moeten we vaak $k \in \mathbb{N}$ elementen uit een verzameling van $n \in \mathbb{N}$ elementen kiezen en berekenen op hoeveel verschillende manieren dit mogelijk is. Een vraag die aanleiding geeft tot het deelgebied van de *combinatoriek*. Het aantal mogelijke manieren wordt gegeven door de binomiale coëfficiënt:

$$\binom{n}{k} := \frac{n!}{k!(n-k)!}\,.$$

- (Dit is nieuw!) Gegeven twee vectorruimtes V, W kunnen we de vectorruimte beschouwen die bestaat uit sommen van een vector in V en een in W:

$$V \oplus W := \{\vec{v} + \vec{w} \mid \vec{v} \in V, \vec{w} \in W\}\,.$$

  Dit wordt de directe som van vectorruimtes genoemd.

- De formule van Euler, een van de zovele, geeft een fundamentele relatie tussen de belangrijkste constanten in de wiskunde:

$$e^{i\pi} + 1 = 0\,.$$

- De formule voor een meetkundige reeks is

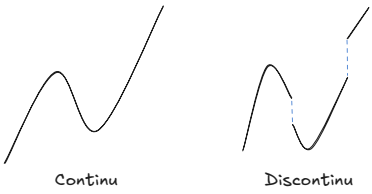
$$\sum_{n=1}^{+\infty} x^n = \frac{1}{1-x}$$

Hoofdstuk 1

Topologie: Van donuts tot algebra

We beginnen dit boek met de notie van continuïteit, wat het betekent om iets 'continu' te vervormen en waarom er zo vaak gezegd wordt dat een wiskundige het verschil niet kent tussen een donut en een koffiekop. In de middelbare school wordt continuïteit van functies meestal gedefinieerd als de eigenschap waarbij je jouw potlood niet hoeft op te tillen om de grafiek te tekenen.

Dit geeft natuurlijk al een idee van wat dit betekent, maar woorden zijn niet voldoende in de wiskunde. We hebben een meer exacte formulering nodig die ons toelaat om, gegeven een functie, te bepalen of die al dan niet continu is. Om het concept continuïteit te formaliseren wordt hierom vaak teruggegrepen naar de be-



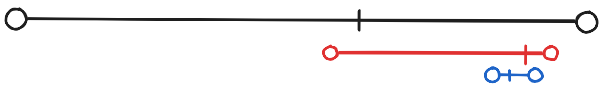
Figuur 1.1: Continuïteit.

ruchte en abstracte ε-δ-definitie:

$$\forall \varepsilon > 0 : \exists \delta > 0 : |x - x_0| < \delta \implies |f(x) - f(x_0)| < \varepsilon \,.$$

Voor elke grootte van afwijking δ, kunnen we een interval rond het punt $x_0 \in \mathbb{R}$ vinden, zodat de functie f minder dan δ varieert. De rechter functie in de figuur op voorgaande pagina voldoet niet aan deze voorwaarde, want als je een heel klein interval rond een van de twee 'discontinuïteiten' neemt, dan is de verticale sprong veel groter dan de breedte van dat interval.

Deze definitie gaat er echter van uit dat de verzamelingen waarop we functies beschouwen beschikt over zowel een notie van optelling als een notie van absolute waardes. Niet ideaal dus als we dit willen veralgemenen naar verzamelingen die geen Euclidische ruimtes $\mathbb{R}^n$ zijn (en zelfs voor $n > 1$ moeten we al de notie van absolute waarde veralgemenen opdat bovenstaande definitie steek zou blijven houden).



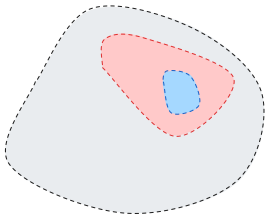
Figuur 1.2: Genestelde open intervallen.

We moeten overgaan op een totaal andere aanpak, een die meer synthetisch is van aard. Als eerste beginnen we met de absolute waardes. Alle punten $x \in \mathbb{R}$ die voldoen aan $|x - x_0| < \delta$ vormen samen het open interval

$$]x_0 - \delta, x_0 + \delta[\,.$$

Het feit dat het hier gaat om een open interval, waarbij de randpunten geen deel uitmaken van de verzameling, is van essentieel belang! Voor elk punt $y \in I_1$ in een open interval I_1, bestaat er een (strikt kleiner) open interval I_2 dat dit punt bevat en zelf volledig bevat zit in I_1. In symbolen wordt dit: $y \in I_2$ en $I_2 \subsetneq I_1$. Je kan altijd een oneindige reeks aan genestelde deelintervallen vinden zoals in de figuur hierboven. Voor gesloten intervallen is dit niet meer het geval, de randpunten doen moeilijk.

Dit concept zullen we nu veralgemenen. We moeten een definitie van open verzameling vinden zodat er voor elk punt $y \in U_1$ van de open verzameling U_1 een (strikt kleinere) open verzameling U_2 bestaat die het punt bevat en zelf volledig bevat zit in de grotere verzameling: $y \in U_2$ en $U_2 \subsetneq U_1$. (Woord voor woord overgenomen uit vorige paragraaf.)

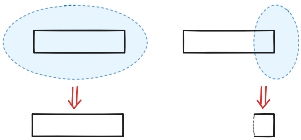


Figuur 1.3: Genestelde open verzamelingen.

Beschouw een verzameling X met een collectie aan deelverzamelingen $\{U \subseteq X\}$. Dit wordt een topologische ruimte genoemd (en de deelverzamelingen worden op hun beurt open genoemd) als deze aan de volgende voorwaarden voldoen:

1. Zowel X zelfs als de ledige verzameling $\emptyset$ zijn open.
2. De doorsnede van twee open verzamelingen is open.
3. Elke unie (eindig of oneindig) van open verzamelingen is open.

De eerste voorwaarde is een soort consistentievoorwaarde. Dat dit niet botst met het voorbeeld van het het gesloten interval hierboven komt omdat randpunten van de ruimte X een soort absolute status hebben. Het zijn geen artificiële randpunten door onze keuze van deelverzameling, het is gewoonweg hun natuur.



Figuur 1.4: Open deelverzamelingen van een rechthoek.

Als we bijvoorbeeld het gesloten interval $[0, 1]$ als een opzichzelfstaande entiteit beschouwen, dan is het interval wel degelijk open. Het is pas wanneer we het als een deelverzameling van de reële lijn zien dat het niet open is. In het algemeen is de topologie op een deelruimte $Y \subset X$ gedefinieerd door voor alle open verzamelingen U de doorsnede $U \cap Y$ te nemen. Hiernaast is een ander voor-

beeld, met een (gesloten) rechthoek in het vlak $\mathbb{R}^2$, afgebeeld om de verwarring wat tegen te gaan. Met de geïnduceerde deelruimtetopologie zijn zowel de volledige rechthoek als een deelverzameling ervan, inclusief een deel van de rand, ook open.

Dat de definitie niet symmetrisch is onder het verwisselen van doorsnede en unie is reeds te zien in het voorbeeld van de reële lijn. Neem de (open) intervallen van de vorm $]-1/n, 1/n[$ voor alle natuurlijke getallen $n \in \mathbb{N}$. Het is niet moeilijk in te zien dat dit een dalende (krimpende) rij van genestelde intervallen voorstelt. De doorsnede van alle deze verzamelingen wordt gegeven door

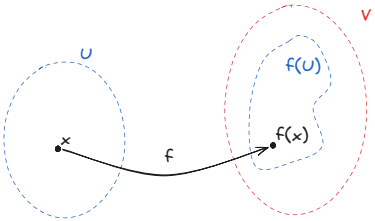
$$\bigcap_{n\in\mathbb{N}} \left]-\tfrac{1}{n}, \tfrac{1}{n}\right[= \{0\}\,,$$

wat helemaal geen open verzameling meer is.

De definitie van een topologische ruimte spreekt over een keuze van deelverzamelingen, dit impliceert dus dat er in het algemeen meerdere topologische structuren op eenzelfde verzameling mogelijk zijn. We zullen nu als voorbeeld zelfs zien dat zodra de verzameling meer dan één punt bevat, de topologische structuur niet meer uniek is. (De lezers van Deel 1 hebben hier een kleine voorsprong.) Daar waren we stiekem al wat van deze topologische noties tegengekomen. Jullie herinneren zich misschien nog de passage over discrete en codiscrete structuren. Wel, de discrete en codiscrete topologieën zijn de meest extreme topologische structuren die men op een verzameling kan plaatsen. De discrete topologie, die waarbij de elementen van de verzameling niet samenhangen, is die waarvoor elk deelverzameling open is. In het andere uiterste, de codiscrete topologie, zijn enkel de ledige verzameling en de totale verzameling open. Alle elementen kleven dus figuurlijk aan elkaar.

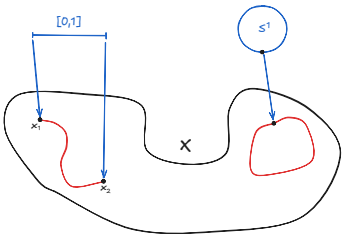
Laten we nu echter terugkeren naar de definitie van een continue functie, want dat was uiteindelijk wat we wouden veralgemenen. De ε-δ-definitie veralgemenen we als volgt. Een functie $f : X \to Y$ tussen topologische ruimtes is continu in een punt $x \in X$ als voor elke open verzameling $V \subseteq Y$ met $f(x) \in V$ er een open verzameling $U \subseteq X$ met $x \in U$ bestaat zodat

$f(U) \subseteq V$, zoals hieronder uitgebeeld.



Figuur 1.5: Continue functie tussen topologische ruimtes.

Een andere manier om hiernaar te kijken is dat voor elke open verzameling V in Y, het inverse beeld $f^1(V)$ ook open moet zijn. (Herinner je dat $f^{-1}(V)$ bestaat uit alle punten $x \in X$ zodat $f(x) \in Y$.) Continue functies zijn dus exact die functies die de notie van een topologische ruimte bewaren, het zijn dus de correcte morfismen in de categorie van topologische ruimtes. (Om het even helemaal abstract te plaatsen.) De continue functies die ook een continue inverse hebben, worden ook wel homeomorfismen genoemd en zij vormen dus de isomorfismen in deze categorie.



Figuur 1.6: Een open en een gesloten pad.

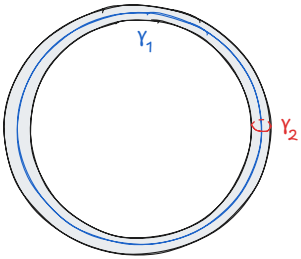
Nu we dit concept hebben, kunnen we onze zoektocht naar de donuts voortzetten. Als eerste beschouwen we paden doorheen een topologische ruimte X. Een (continu) pad tussen twee punten $x_1, x_2 \in X$ kan gezien worden als een continue functie $\gamma : [0,1] \to X$ zodat

$$\gamma(0) = x_1 \qquad \text{en} \qquad \gamma(1) = x_2\,.$$

Het zal je dan ook niet verbazen dat een continue lus, een pad dat eindigt waar het begonnen is, in een topologische ruimte beschreven kan worden als een continue functie $\gamma : S^1 \to X$ met als domein de cirkel. (Deze krijgt de deelruimtetopologie geïnduceerd door $\mathbb{R}^2$.)

Goed, we zitten al aan de lussen en lussen vinden wel degelijk hun toepassing in de studie van donuts en koffiemokken. Neem dus die eerste, de donut. Hier kunnen we twee 'fundamentele' lussen beschouwen. De ene gaat rond de opening van de donut heen en de andere gaat rond de

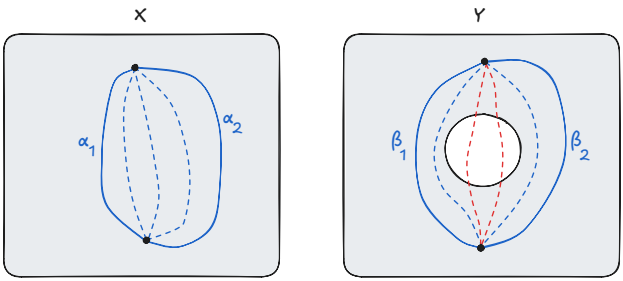
dikte van de donut heen. Een schets van deze situatie kan je terugvinden bovenaan de volgende pagina. @@ check locatie van figuur @@



Figuur 1.7: Fundamentele paden op een donut.

Het speciale aan die lussen is dat je ze onmogelijk in elkaar kan vervormen zonder de lus te breken. Hetzelfde effect krijg je met een koffiemok. Hang een slot of knoop rond het oor van de mok. Zonder het slot of de knoop open te maken, ga je deze niet van rond het oor krijgen. De lussen detecteren dus in zekere zin de gaten die in een object (of, meer specifiek, topologische ruimte) zitten.

Dat beschrijven van de 'gaten' in een topologische ruimte is hetgeen waar we in dit hoofdstuk naartoe zullen werken. Heel veel andere problemen kunnen herleid worden tot vragen binnen dit vakgebied. Omdat topologische problemen echter vaak complex en moeilijk te berekenen vallen, zullen we echter overgaan op een eenvoudigere taal, die van de algebra. Het vervormen van paden of lussen zal voor ons de sleutel vormen om de wereld van de algebraïsche topologie te openen.



Figuur 1.8: (Het ontbreken van) een homotopie in topologische ruimtes.

In Deel 1 zagen we dat er een notie van pad tussen paden bestaat, een

homotopie. Met de definities uit dit hoofdstuk kunnen we dit wat formeler definiëren. Een homotopie tussen twee continue functies $f, g : X \to Y$ is een continue functie $H : [0,1] \times X \to Y$ zodat

$$H(0,x) = f(x) \qquad \text{en} \qquad H(1,x) = g(x)\,.$$

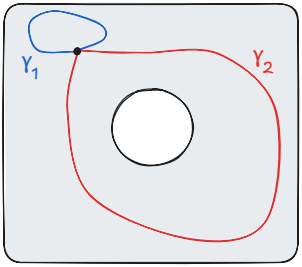
Voor elk waarde van $t \in [0,1]$ dient de functie $X \to Y : x \mapsto H(t,x)$ continu te zijn en, analoog, dient voor elke waarde van $x \in X$ de functie $t \mapsto H(t,x)$ continue te zijn. De parameter t laat ons toe om de ene functie in de andere te 'vervormen' op een continue manier.

In de figuur onderaan voorgaande pagina zien we echter dat dergelijke homotopieën niet altijd bestaan. Indien de topologische ruimte gaten bevat, dan kan het zijn dat we twee paden niet in elkaar kunnen vervormen. In het linkerdeel van die figuur zien we een deel van het vlak en in het vlak kunnen alle continue paden in elkaar vervormd worden. In het rechterdeel van die figuur zien we echter een deel van het vlak waaruit een schijf is verwijderd. Indien we nu op dezelfde wijze als in het linkerdeel het ene blauwe pad in het andere proberen te vervormen, dan komt er een moment waarop het vervormde pad de opening doorkruist (de rode stippellijnen). We kunnen de verzameling van paden in een topologische ruimte dus opdelen in sectoren, waarbij alle paden die tot dezelfde sector behoren in elkaar vervormd kan worden. Wiskundigen zeggen ook wel dat het bestaan van een homotopie tussen twee functies een equivalentierelatie definieert.

Als we nog eens naar Figuur 1.8 kijken, dan kunnen we zien dat we de situatie ook in termen van lussen kunnen uitdrukken. In zowel het linker- als rechterdeel kan je de twee paden aan elkaar plakken om een lus te bekomen. Links leg je als het waar eerst het pad α_1 af en daarna keer je terug langsheen α_2. Zo krijg je in het linkerdeel een soort ongelukkige cirkel die samen te trekken valt tot het startpunt. De lus waarvan je begon is dus homotopie-equivalent aan een punt. Rechts daarentegen is dit niet het geval, het gat zit in de weg. De lus is niet homotopie-equivalent aan een punt.

De verzameling van equivalentieklassen van continue lussen in een to-

pologische ruimte X wordt aangeduid met $\pi_1(X)$. Meer algemeen, wanneer je geen lussen maar functies $S^n \to X$ vanuit hoger-dimensionale sferen beschouwt, krijgen we de verzamelingen $\pi_n(X)$ voor alle $n \in \mathbb{N}$. Zo stellen elementen van $\pi_2(X)$ bijvoorbeeld equivalentieklassen van sferen in X voor.



Figuur 1.9: Product van paden.

Het enige wat nu nog ontbreekt is een algebraïsche structuur op de verzamelingen $\pi_n(X)$. Deze structuur wordt geïnduceerd door de samenstelling van paden die we hierboven reeds gebruikten. Beschouw de twee paden in de figuur hiernaast. Zoals in de voorgaande paragrafen besproken, behoren γ_1 en γ_2 tot twee verschillende klassen in $\pi_1(X)$. We kunnen echter aan de hand van samenstelling een derde pad bekomen $\gamma_2 \circ \gamma_1$. We lopen eerst langsheen γ_1 en daarna langsheen γ_2. «*Kan je bedenken tot welke klasse in $\pi_1(X)$ dit nieuwe pad behoort?*»

Aangezien we het blauwe pad γ_1 kunnen vervormen tot een punt, is het samengestelde pad vervormbaar tot het rode pad γ_2, het windt ook een keer rond de opening in de ruimte. Wat als we nu twee keer langsheen het rode pad lopen? Dan bekomen we een pad dat twee keer rond de opening windt en noch vervormbaar tot γ_1, noch vervormbaar tot γ_2 is. We bekomen dus een element van een klasse in $\pi_1(X)$ die we nog niet waren tegengekomen. Anderzijds kunnen we ook eens langs γ_2 in de ene richting lopen en terugkeren langs γ_2 in de andere richting. In dit geval windt het uiteindelijke pad nul keer om de opening en is het vervormbaar tot het blauwe pad.

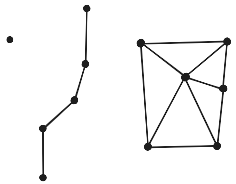
Bij sommige lezers begint het nu misschien te dagen dat we een sterk verband tussen $\pi_1(X)$ en de gehele getallen $\mathbb{Z}$ ontdekt hebben. Als je twee lussen neemt die respectievelijk $m \in \mathbb{N}$ en $n \in \mathbb{N}$ keer in dezelfde zin om de

opening winden en je stelt deze samen, dan bekom je een nieuwe lus die $m+n$ keer om de opening windt. Als je de oriëntatie van een lus omkeert, dan hoef je simpelweg met negatieve getallen te werken.

De samenstelling van paden definieert de structuur van een groep op de verzameling $\pi_1(X)$ (en meer algemeen op $\pi_n(X)$). De vermenigvuldiging komt neer op de samenstelling, het neutraal element wordt gegeven door het constante pad op een punt (of, equivalent, een pad dat hiernaar vervormt kan worden) en inversen worden gegeven door paden met de omgekeerde oriëntatie. Als we dan zoals hierboven bijvoorbeeld kijken naar het vlak met een gat erin, wiskundig aangeduid met $\mathbb{R}^2\backslash\{0\}$, dan geldt $\pi_1\big(\mathbb{R}^2\backslash\{0\}\big) \cong \mathbb{Z}$, deze verzameling is als groep isomorf aan de (additieve) groep van de gehele getallen. De groepen $\pi_n(X)$ worden ook wel de homotopiegroepen van X genoemd en de groep $\pi_1(X)$ in het bijzonder wordt de fundamentele groep van X genoemd.

De belangrijkste eigenschap van de homotopiegroepen is dat ze topologische invarianten zijn. Twee isomorfe topologische ruimtes hebben isomorfe homotopiegroepen (al is dit geen als-en-slechts-alsrelatie). Om aan te tonen dat twee ruimtes fundamenteel verschillend zijn, volstaat het dus vaak om naar een van deze invarianten te kijken.@@ geef meer uitleg @@

Naast de homotopiegroepen, bestaan er nog ander (deels gerelateerde) algebraïsche invarianten van topologische ruimtes. Deze dragen de naam van cohomologiegroepen (zie ook Deel 1) en de eerste sporen van cohomologie dateren uit de tijd van Euler met zijn werk rond het probleem van de *zeven bruggen van Königsberg* uit de 18e eeuw. Het voordeel van deze groepen tegenover de homotopiegroepen is dat ze makkelijker te berekenen vallen. Om dit te doen, gaan we volop aan de slag met driehoeken of, zoals de wiskundigen ze noemen: simplices.

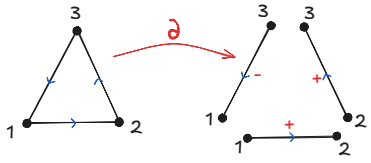


Figuur 1.10: Simpliciële complexen in dimensies 0, 1 en 2.

We beginnen onderaan de ladder, in dimensie 0. De 0-simplices zijn

simpelweg punten. Een trede hoger, in dimensie 1, krijgen we lijnen. (We zouden rechte lijnen kunnen eisen, maar topologie kijkt niet naar het verschil tussen krom en recht.) Elke 1-simplex geeft op zijn beurt aanleiding tot twee 0-simplices: zijn randpunten. Als we nog een trede hoger komen, dan bereiken we de wereld van de driehoeken. Elke 2-simplex geeft aanleiding tot drie 1-simplices: zijn zijdes. In de figuur op voorgaande pagina kan je ook zien hoe je complexere meetkundige figuren kan benaderen door eenvoudige simplices aaneen te plakken.

Objecten die uit simplices opgebouwd zijn door de simplices langsheen hun 'rand' aaneen te plakken, worden simpliciële complexen genoemd. Wij zullen geïnteresseerd zijn in topologische ruimtes die homeomorf zijn aan een simplicieel complex, aangezien we dan de topologische ruimte kunnen bespreken door hem op te breken in zijn simplices. (Dit is onder meer mogelijk voor alle gladde variëteiten, maar niet voor louter *topologische* variëteiten.) Beschouw nu de verzameling van alle k-simplices in zo een simpliciële ruimte X. Hiermee kunnen we, zuiver formeel, een optelling definiëren en de resulterende verzameling wordt aangeduid met $C_k(X)$. Nemen we twee k-simplices ω_1, ω_2, dan geldt zowel $\omega_1, \omega_2 \in C_k(X)$ als $\omega_1 + \omega_2, \omega_1 - \omega_2 \in C_k(X)$. Dergelijke formele sommen en verschillen worden ook wel k-ketens genoemd.



Figuur 1.11: Differentiaal van een 2-simplex.

De functie die een k-simplex naar zijn rand stuurt, bestaande uit $(k-1)$-simplices, induceert ook een afbeelding op ketens. Neem bijvoorbeeld de 2-simplex uit de figuur hiernaast, die we voor de eenvoud noteren als [1 2 3] (het wordt zo meteen duidelijk waarom). De rand bestaat uit de drie 1-simplices [1 2], [2 3] en [3 1]. De afbeelding $\partial : C_k(X) \to C_{k-1}(X)$ beeldt [1 2 3] af op de alter-

nerende som van zijn rand:

$$\partial[1\,2\,3] := [2\,3] - [1\,3] + [1\,2]\,.$$

Waarom we die alternerende definitie met het minteken kiezen in plaats van enkel de som, lijkt misschien wat vreemd. Deze keuze wordt echter duidelijk als we nogmaals de randoperator ∂ toepassen:

$$\partial^2[1\,2\,3] = \partial\big([2\,3] - [1\,3] + [1\,2]\big) = [3] - [2] - [3] + [1] + [2] - [1] = 0\,.$$

De algemene formule voor de randoperator op k-simplices leest als volgt:

$$\boxed{\partial[i_1\, i_2\, \ldots\, i_k]} := \sum_{j=1}^{k} (-1)^{j+1} [i_1\, \ldots \hat{i_j} \ldots\, i_k]\,,$$

waar het hoedje aangeeft dat die index wordt weggelaten. Deze definitie kunnen we verder nog van simplices naar ketens uitbreiden door middel van lineariteit op te leggen aan ∂. «*Kijk maar eens na dat dit ook geldt voor bijvoorbeeld de 4-simplex* $[1\,2\,3\,4]$ (*de piramide*).»

We zijn bijna klaar om onze invarianten te definiëren. In de context van homotopieën hadden we dit gedaan door equivalentieklassen van paden (of lussen) te beschouwen en ook hier zullen we een geschikte equivalentieklasse op de groep $C_k(X)$ construeren. Elementen van de vorm $\partial\omega$ worden toepasselijk randen genoemd en elementen die door ∂ op nul worden afgebeeld worden cycli genoemd. De verklaring voor deze laatste is dat ze gesloten ketens voorstellen. Neem hiervoor opnieuw het voorbeeld van de driehoek. De rand van de driehoek wordt gegeven door de aaneenschakeling van de paden $1 \to 2$, $2 \to 3$ en $3 \to 1$. Dit komt als een keten in $C_1(X)$ exact overeen met de uitdrukking $[2\,3] - [1\,3] + [1\,2]$ wanneer we opmerken dat we een minteken krijgen bij het omdraaien van de oriëntatie van een simplex: $[1\,3] = -[3\,1]$. Het gesloten zijn van dit pad komt dus overeen met de voorwaarde dat de rand nul is (net zoals een cirkel geen meetkundige rand heeft).

Nu, aangezien ∂ voldoet aan de voorwaarde $\partial^2 = 0$, stellen randen eigenlijk triviale cycli voor. Voor eender welke k-keten $\sigma \in C_k(X)$ en $(k+1)$-keten $\omega \in C_{k+1}(X)$ geldt dat

$$\partial(\sigma + \partial\omega) = \partial\sigma\,.$$

We kunnen $C_k(X)$ opdelen in sectoren door middel van volgende equivalentierelatie. Twee ketens of randen zijn equivalent als en slechts als ze verschillen op een rand na, want dan hebben ze ook dezelfde rand:

$$\sigma_1 \sim \sigma_2 \iff \sigma_1 = \sigma_2 + \partial\omega\,.$$

De resulterende quotiëntgroepen worden aangeduid met $H_k(X)$ en worden de (simpliciële) homologiegroepen van X genoemd.

Laten we als laatste eens kijken hoe de homologiegroepen, net zoals de homotopiegroepen, in staat zijn de de gaten in een topologische ruimte te detecteren. Beschouw hiervoor de cirkel S^1. Topologisch kunnen we deze benaderen door de rand van een 1-simplex (een driehoek en een cirkel zien er hetzelfde uit voor een wiskundige). Alle randen in $C_1(S^1)$ zijn dan ook een veelvoud van deze keten:

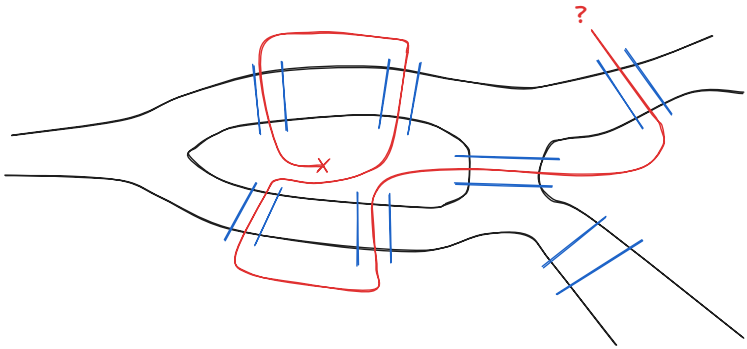
$$\partial\omega = 0 \implies \omega = n\big([2\,3] - [1\,3] + [1\,2]\big)$$

voor een zekere $n \in \mathbb{Z}$. Om $H_1(S^1)$ te vinden moeten we kijken of deze genererende keten de rand is van een 2-keten. Het antwoord hier is echter onmiddellijk nee, want de cirkel is eendimensionaal en bevat dus geen 2-simplices. Het volgt dat de keten $[2\,3]-[1\,3]+[1\,2]$ de gehele groep $H_1(S^1)$ genereert en dat deze groep isomorf is aan $\mathbb{Z}$ (alle elementen ervan zijn gehele veelvouden van de generator). Ook hier vinden we het windingsgetal terug. Voor algemene topologische ruimtes geldt echter niet dat homotopiegroepen en homologiegroepen overeenkomen. De homotopiegroepen zijn in het algemeen veel complexer.

Wanneer we overgaan van eendimensionale naar tweedimensionale ruimtes, ztten we echter een kleine domper op de stelling dat donuts en koffiemokken voor een wiskundige equivalent zijn. Een donut (of torus

voor de specialisten) $\mathbb{T}$ is een verzameling van de vorm $S^1 \times S^1$ zoals in Figuur 1.7. De ene cirkel kan je langs de buitenkant zien, dat is de O-vorm van de donut. De andere cirkel daarentegen zegt dat de donut zelf buisvormig is, hij is niet gevuld. Enkel wanneer we een koffiemok die van binnen hol is, zouden beschouwen, is de stelling correct! We geven ook nog mee dat zowel de fundamentele groep als de eerste homologiegroep van $\mathbb{T}$ isomorf zijn aan $\mathbb{Z}^2$. Voor productruimtes kan je de homotopie- en homologiegroepen opbouwen aan de hand van die van de factoren. (Voor complexere constructies is dit soms ook nog mogelijk, maar dat is een vak op zich.)

Extra's: De puzzel van Euler



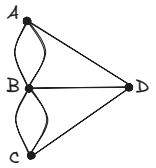
Figuur 1.12: De zeven bruggen van Königsberg.

Laat ons gewapend met de kennis over simpliciële homologie eens naar het probleem van de bruggen van Königsberg kijken en zien hoe dit ons snel tot dezelfde oplossing als Euler kan leiden. Beschouw de prachtige schets van Königsberg[1] hierboven. Deze stad bestaat uit drie landmassa's en een eiland, gescheiden door rivieren. Over deze rivieren liggen zeven

[1] Dit is het hedendaagse Kaliningrad, een Russische enclave tussen Polen en Litouwen.

bruggen die de landmassa's met elkaar verbinden. De vraag die aan Euler gesteld werd, is of het mogelijk is om een pad te vinden dat exact één keer over alle bruggen loopt (al dan niet waarbij het eindpunt samenvalt met het beginpunt).

Om deze vraag te beantwoorden, zullen we eerst de kaart van Königsberg herleiden tot een simplicieel complex. Voor elke landmassa nemen we een punt en voor elke brug een lijn. Hiernaast is de resulterende *graaf* getekend. (Een graaf is een simplicieel complex dat louter bestaat uit 0- en 1-simplices. Grafentheorie is een uiterst goed bestudeerd deelgebied met toepassingen in zowel cartografie als computerwetenschappen.)



Figuur 1.13: De graaf van Königsberg.

De rand van een pad op een simplicieel complex is altijd gelijk aan het verschil van zijn eind- en beginpunt. Een pad kan voorgesteld worden als een keten bestaande uit 1-simplices waarbij het eindpunt van de ene simplex gelijk is aan het startpunt van de volgende (zoals in Figuur 1.10). Als je hier de rand van berekent, dan vallen alle interne punten weg en krijg je enkel het start- en eindpunt, bijvoorbeeld:

$$\begin{aligned}\partial\big([1\,2] + [2\,3] + [3\,4] + [4\,5]\big) &= [2] - [1] + [3] - [2] + [4] - [3] + [5] - [4] \\ &= [5] - [1]\,.\end{aligned}$$

Kijken we echter naar de rand van de graaf K uit de figuur hierboven, dan krijgen we

$$\partial K = \pm[A] \pm [B] \pm [C] \pm [D]\,,$$

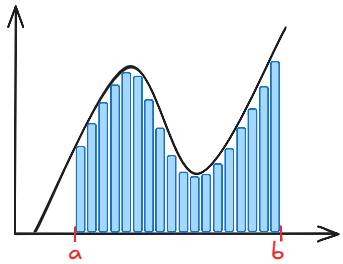
waar de tekens afhangen van welke oriëntaties we beschouwen. Het belangrijkste in dit resultaat is dus dat de rand van de graaf gevormd door de zeven bruggen niet overeenkomt met de rand van eender welk pad. Het is dus fundamenteel onmogelijk dat we een pad vinden zoals gewenst!

Hoofdstuk 2

Calculus: Integraalrekening

In Deel 1 hadden we bewust de notie van integraal zo goed als volledig vermeden. Toch is het niet mogelijk om aan fysica, of enige andere vorm van kwantitatieve wetenschap, te doen zonder het concept van een integraal te kennen. Om deze reden zullen we in dit hoofdstuk niet enkel de meest gekende variant zien, de Riemannintegraal, maar ook nog drie andere uitbreidingen die ons toelaten om integraalrekening op exotischere ruimtes toe te passen.

Het klassieke beeld van de integraal van functies (in één veranderlijke) is het berekenen van de oppervlakte onder hun grafiek zoals hiernaast afgebeeld. Het idee van Riemann was om deze oppervlakte te benaderen als een reeks rechthoeken naast elkaar. Je hebt dan simpel weg de oppervlaktes van de rechthoeken, die een eenvoudige formule hebben, op te tellen om de totale oppervlakte



Figuur 2.1: Riemannintegraal.

te vinden. Hoe smaller we de rechthoekjes nemen, hoe beter de benadering en, in de limiet van oneindig smalle (en oneindig veel) rechthoekjes bekomen we de exacte oppervlakte. Men kan hier dan heel wat eigenschappen voor aantonen, zoals de onafhankelijkheid van de precieze keuze van rechthoekjes, etc. De enige eigenschap die voor ons echt van belang is, is de de fundamentele stelling van de calculus, wiens naam reeds doet vermoeden dat het een van de belangrijkste stellingen binnen de calculus is. Op school leer je de volgende vorm:

$$\int_a^b \frac{\mathrm{d}f}{\mathrm{d}x}\, dx = f(b) - f(a)\,.$$

Dit zegt dat de integraal van een afgeleide over het (gesloten) interval $[a, b]$ is gelijk aan het verschil van zijn waarden op de eindpunten van het interval.

Deze stelling impliceert onder meer al dat integreren en afleiden wel degelijk elkaars omgekeerde zijn. Neem een integreerbare functie $f(x)$ en definieer

$$F(x) := \int_a^x \frac{\mathrm{d}f}{\mathrm{d}x}\, dx\,.$$

Door gebruik te maken van de fundamentele stelling zien we onmiddellijk dat $F(x) = f(x) - f(a)$. Integreren is dus het omgekeerde van afleiden op een constante na, waarbij de constante enkel afhangt van het beginpunt a dat we hier arbitrair gekozen hebben. (Dit komt overeen met de eigenschap dat de afgeleide van een eender welke constante nul is.) Om deze reden rekenen leerkrachten op de middelbare school zo graag minpunten aan wanneer de leerlingen hun "plus constante" vergeten bij integratieoefeningen. Merk wel op dat voor bepaalde integralen, waarbij zowel de onder- als bovengrens vastliggen, deze constante geen rol speelt. Ze komt eigenlijk neer op een keuze van nulpunt van de y-as.

Het is hier misschien wel gepast om enkele voorbeelden te bekijken. Om integralen uit te rekenen, zullen we gretig gebruik van de fundamentele stelling en de kennis van veelvoorkomende afgeleiden.

- Veeltermen x^n:

$$\int_0^a x^k\,dx = \int_0^a \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{x^{n+1}}{n+1}\right)dx = \frac{a^{n+1}}{n+1}\,.$$

- Exponentiële functies:

$$\int_0^a \exp(\lambda x)\,dx = \int_0^a \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\exp(\lambda x)}{\lambda}\right)dx = \frac{\exp(\lambda a)}{\lambda}\,.$$

- Product van een veelterm en een exponentiële functie:

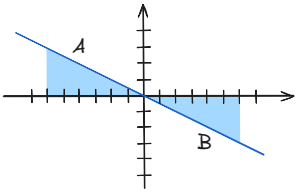
$$\begin{aligned}\int_0^a x^n \exp(\lambda x)\,dx &= \int_0^a \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{x^n \exp(\lambda x)}{\lambda}\right)dx - \int_0^a nx^{n-1}\frac{\exp(\lambda x)}{\lambda}\,dx\\ &= \frac{n}{\lambda}a^{n-1}\exp(\lambda a)\\ &\quad - \int_0^a \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{nx^{n-1}\exp(\lambda x)}{\lambda^2}\right)dx\\ &\quad + \int_0^a n(n-1)x^{n-2}\frac{\exp(\lambda x)}{\lambda^2}\,dx\\ &= \cdots\end{aligned}$$

In deze uitwerking hebben we meerdere malen ($n \in \mathbb{N}$ keer om precies te zijn) de Leibnizregel voor afgeleiden gebruikt om de afgeleide van de exponentiële functie naar de veelterm te verplaatsen en zo de graad van de veelterm te verlagen tot die 0 wordt. Deze techniek wordt ook wel partiële integratie genoemd:

$$\int_a^b f(x)\frac{\mathrm{d}g}{\mathrm{d}x}\,dx = f(b)g(b) - f(a)g(a) - \int_a^b \frac{\mathrm{d}f}{\mathrm{d}x}g(x)\,dx\,.$$

Je hebt nu wel al een idee van wat integralen voorstellen in één dimensie, maar dit is zeker niet voldoende in de praktijk. Want wat als je, in plaats van de oppervlakte onder een grafiek, het volume wil berekenen?

In dit geval moeten we overgaan van 'lijnintegralen' naar oppervlakteintegralen of dubbele integralen. (Uiteraard kan je nog verder gaan en bijvoorbeeld volumeintegralen beschouwen.) De definitie hiervan is heel gelijkaardig aan die van de gewone Riemannintegraal. In plaats van het integratiedomein op te delen in kleine intervalletjes, deel je het nu op in kleine rechthoekjes. *That's it!*



Figuur 2.2: Integraal van een oneven functie.

Om de integralen praktisch uit te rekenen, wordt er in hogere dimensies vaak minder rechtstreeks gebruik gemaakt van de fundamentele stelling. In de plaats gebruiken we twee vuistregels en een reductiemethode om meervoudige integralen te herleiden tot eendimensionale integralen. (Deze eigenschappen zijn geldig in eender welk aantal dimensies.)

- Constante functies: Voor constante functies is de oppervlakteintegraal simpelweg gelijk aan het product van de constante en de oppervlakte van het domein waarover we integreren:

$$\iint_A \lambda \, d^2x = \lambda \, \mathrm{opp}(A) \, .$$

- Symmetrie: Indien het domein symmetrisch is rond de oorsprong en de functie oneven is, dat wil zeggen dat $f(x,y) = -f(-x,-y)$, dan is het resultaat exact gelijk aan nul. Kijk bijvoorbeeld naar de figuur hierboven. De blauwe grafiek voldoet aan de eigenschap $f(x) = -f(-x)$. Waar de integraal A gelijk is aan $\frac{1}{2} \cdot 7 \cdot 3$, is B gelijk aan $\frac{1}{2} \cdot 7 \cdot (-3) = -A$. Het totaal $A + B$ is dus gelijk aan nul.

- Stelling van Fubini: Voor integreerbare functies kan je de dubbelintegraal berekenen aan de hand van twee enkelvoudige integralen:

$$\iint_A f(x,y) \, d^2x = \int_a^b \left(\int_{\alpha(x)}^{\beta(x)} f(x,y) \, dy \right) dx \, ,$$

waarin de functies $\alpha, \beta : \mathbb{R} \to \mathbb{R}$ de onder- en bovengrenzen voor de waarde van y aangeven voor een gegeven waarde van x.

Stel je bijvoorbeeld voor dat we een functie $g(x, y)$ willen integreren over de driehoek A in Figuur 2.2. In dit geval loopt x tussen -7 en 0, maar de mogelijke waarden van y hangen echter af van die van x. (Je kan de logica ook omdraaien en in functie van y werken.) Voor $x = -7$ krijgen we bijvoorbeeld $y \in [0, 3]$ terwijl voor $x = -1$ geldt dat $y \in [0, \frac{6}{7}]$. Die variabele grenzen worden vertaald naar de functies $\alpha, \beta : \mathbb{R} \to \mathbb{R}$. In het bijzonder geldt hier dat $\alpha(x) \equiv 0$ en $\beta(x) = -\frac{3}{7}x$.

Als een praktisch voorbeeld van integralen zullen we eens het volume van een bol berekenen. Slimmeriken zullen misschien wel al gemerkt hebben dat het volume $\frac{4}{3}\pi r^3$ simpelweg de integraal van de oppervlakte πr^2 van een schijf is, maar waarom is dit zo? Wel, plaats de bol mooi gecentreerd op de oorsprong van een Cartesiaans assenstelsel. Doorsnedes van deze bol langsheen de z-as zijn allemaal schijven met straal gegeven door $r^2 = x^2 + y^2 = R^2 - z^2$, waarbij R de straal van de bol is. Om het volume te bepalen moeten we al die schijven op elkaar leggen of, met andere woorden, we moeten de oppervlakte van de schijven integreren over de z-as:

$$\begin{aligned} V_{\text{bol}} &= \int_{-R}^{R} \pi(R^2 - z^2)\, dz \\ &= \pi R^2\big(R - (-R)\big) - \tfrac{1}{3}\pi\big(R^3 - (-R)^3\big) \\ &= 2\pi R^3 - \tfrac{2}{3}\pi R^3 \\ &= \tfrac{4}{3}\pi R^3\,. \end{aligned}$$

«*Probeer dit eens na te rekenen in Cartesiaanse coördinaten in plaats van bolcoördinaten.*»

De Riemannintegraal is echter niet altijd voldoende. Als voorbeeld blijven nog steeds binnen het eendimensionale domein en beschouwen we kansverdelingen. Voor continue verdelingen op $\mathbb{R}$ die een kansdichtheid $p(x)$ bezitten, is de situatie heel eenvoudig. De kans om een waarde in het

interval $[a, b]$ te vinden is gelijk aan de bijhorende integraal:

$$P\big(x \in [a,b]\big) = \int_a^b p(x)\,dx\,.$$

Maar wat in het geval van discrete verdelingen? Beschouw het eenvoudigste geval, dat van een zogenaamde Diracverdeling, waar alle kans aan één punt $x_0 \in \mathbb{R}$ wordt toegewezen. In dat geval krijgen we

$$P\big(x \in I\big) = \begin{cases} 1 & x_0 \in I\,, \\ 0 & x_0 \notin I\,. \end{cases}$$

Als we maar één mogelijke waarde kunnen waarnemen, dan zijn voorspellingen heel eenvoudig. Ofwel zal de waarde met zekerheid in een bepaalde verzameling zitten, ofwel zal hij met zekerheid er niet in zitten. Het probleem met dergelijke kansverdelingen is dat ze niet overeen komen met een 'gewone' functies op $\mathbb{R}$. Er is geen enkele functie $f : \mathbb{R} \to \mathbb{R}$ waarvan de (Riemann)integraal overeenkomt met het binaire resultaat van hierboven en de reden hiervoor is heel simpel. De Riemannintegraal benadert de oppervlakte onder een grafiek met rechthoeken, maar in dit geval is die grafiek overal, behalve in het punt x_0, gelijk aan nul. We zouden hem dus moeten benaderen door een enkele rechthoek met breedte nul... We zitten schijnbaar met een probleem.

Mede om deze reden introduceerden wiskundigen rond de 20e eeuwwisseling een nieuw deelgebied: de matentheorie. (Andere gekke problemen zoals de *Banach–Tarskiparadox* werden hiermee ook opgelost.) Waar de Riemannintegraal vertrekt van een functie op $\mathbb{R}$ om een functie op intervallen te construeren, gaat men in de matentheorie in de andere richting te werk. We vertrekken van twee abstracte concepten. Enerzijds een collectie van meetbare verzamelingen Σ, die je kan zien als een synthetische veralgemening[1] van intervallen of verzamelingen waar we intuïtief noties van lengte, oppervlakte of volume kunnen aan toekennen. Anderzijds hebben

[1] Denk hierbij terug aan de bespreking van analytische versus synthetische constructies uit Deel 1.

de notie van een maat $\mu : \Sigma \to \mathbb{R}$, een soort veralgemening van kansverdelingen die ons net toelaat om de lengtes, oppervlaktes enzovoort van deze nieuwe soort verzamelingen te berekenen. Deze objecten voldoen aan de volgende voorwaarden («*waarvan je jezelf zou moeten kunnen overtuigen dat ze steek houden*»):

1. Positiviteit: $\mu(A) \geq 0$ voor alle $A \in \Sigma$.
2. Compleetheid: $\mu(\emptyset) = 0$.
3. Disjuncte unies: $\mu\left(\bigsqcup_{n=1}^{+\infty} A_n\right) = \sum_{n=1}^{+\infty} \mu(A_n)$ voor alle disjuncte verzamelingen $A_n \in \Sigma$.

Deze aanpak geeft ons veel meer vrijheid. Aangezien we nu van functies op verzamelingen vertrekken, kunnen we zonder problemen een waarde toekennen aan verzamelingen die maar uit één punt bestaan. Anderzijds kunnen we ook verzamelingen met meerdere punten beschouwen waarvan de niet nul is, maar waarvoor wel elke deelverzameling maat nul heeft. Alle gekke combinaties zijn mogelijk.

Bovendien laat de matentheorie ons ook toe om meer algemene integralen te definiëren, de zogenaamde Lebesgue-integralen. Meer bepaald associëren we met elke maat μ een Lebesgue-integraal $\int_X d\mu$. We vertrekken hiervoor van de 'indicatorfuncties', dat zijn functies die constant zijn (met waarde 1) op een meetbare verzameling:

$$\mathbb{1}_A(x) = \begin{cases} 1 & x \in A\,, \\ 0 & x \notin A\,. \end{cases}$$

De integraal hiervan is heel eenvoudig, het is de maat van de verzameling A:

$$\int_X \mathbb{1}_A \, d\mu := \mu(A)\,.$$

Daarna veralgemenen we de integraal naar lineaire combinaties van indicatorfuncties (we nemen dus als axioma dat de integraal lineair is, zoals

de Riemannintegraal). Als laatste gebruiken we een convergentieresultaat, de zogenaamde *monotone convergentiestelling* om algemenere functies te benaderen door dergelijke lineaire combinaties. In zekere zin benaderen we opnieuw de grafieken van onze functies door 'rechthoekjes', maar deze keer hebben ze als basis een meer algemene verzameling dan enkel (gesloten) intervallen.

Een belangrijke eigenschap van de Lebesgue-integraal is dat alle Riemannintegreerbare functies — dit zijn functies waarvoor de Riemannintegraal eindig is — ook Lebesgue-integreerbaar zijn en dat de integralen gelijk zijn. Het is echter de uitbreiding naar andere ruimtes die deze integraal (en matentheorie in het algemeen) zo aantrekkelijk maakt. Deze theorie heeft het mogelijk gemaakt de hele kansrekening en statistiek van een formele fundering te voorzien. Complexe stochastische processes zoals de *Brownse beweging* van Einstein, die uiteindelijk mee aan de basis lagen van het atoommodel en van economische modellen zoals het *Black–Scholes model*, zouden niet te beschrijven vallen zonder matentheorie.

Net die veralgemening naar algemenere ruimtes zal ook voor ons van belang zijn in de theoretische fysica wanneer we bijvoorbeeld naar gekromde ruimtes in de algemene relativiteitstheorie of naar de padintegraal van ijktheorieën kijken. In het algemeen wordt de ruimte(tijd) waar fysica zich op of in afspeelt, gegeven door een (gladde) variëteit M. Dit was een verzameling, een topologische ruimte zelfs met Hoofdstuk 1 in het achterhoofd, die je lokaal kan beschrijven aan de hand van Cartesiaanse coördinaten. Globaal is er echter geen garantie dat je een Cartesiaans assenstelsel kan vinden en de gewone Riemannintegraal is dus ook uitgesloten. Wat we wel kunnen doen, net zoals met de meeste concepten die we veralgemeend hebben van $\mathbb{R}^n$ naar M, is de integraal lokaal definiëren (door middel van de Riemann- of Lebesgue-integraal) en kijken of we de lokale definities kunnen aaneenplakken op een consistent manier. (Uiteraard is het antwoord ja.) Wat blijkt nu, onze geliefde differentiaalvormen duiken ook hier weer op.

Het blijkt dat de juiste objecten om te integreren op een d-dimensionale

variëteit niet de gladde functies zijn, maar wel de d-vormen! (Dit heeft te maken met hun transformatiegedrag die gegeven wordt door de inverse Jacobiaan.) Bij de standaardintegralen uit het middelbaar schrijf je achteraan de integraal altijd een dx, het zogenaamde lijnelement, bijvoorbeeld $\int_a^b x^2\,dx$. Wanneer je overgaat naar oppervlakte- en volumeintegralen, wordt dit dan bijvoorbeeld $\iiint (x^2 + y^2 + z^2)\,d^3x$. Wel, die $d^n x$ die je in een algemene n-voudige integraal zou tegenkomen, is eigenlijk exact zo een n-vorm op $\mathbb{R}^n$. Deze specifieke keuze wordt ook wel een volumevorm of volume-element genoemd, want het komt neer op een keuze van eenheidsvolume (in n dimensies). Voor algemene variëteiten is die keuze echter niet altijd zo voor de hand liggend en is ze zeker niet globaal gedefinieerd, net zoals de coördinaten zelf.

Binnen deze setting van integralen en differentiaalvormen laat ook de fundamentele stelling van de calculus een prachtige veralgemening toe. Voor elke differentiaalvorm ω kan men zijn de Rhamafgeleide definiëren, die lokaal gegeven wordt door de afgeleide van de coëfficiënten te nemen:

$$\mathrm{d}\omega_{ji_1\ldots i_k} := \partial_j \omega_{i_1\ldots i_k}\,.$$

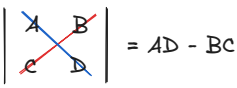
Deze zal de rol van de afgeleide in de fundamentele stelling spelen. Verder evalueren we in het eendimensionale geval de functie in de randpunten $a, b \in \mathbb{R}$ en die randpunten, zo zegge hun naam, vormen de rand van het interval $[a, b]$ waarover we integreren.

Hou je nu vast voor de volgende stap. Het rechterlid van de fundamentele stelling is niet de som van de twee waarden, maar het verschil. We kunnen dus niet zomaar zeggen dat het neerkomt op een 'discrete' integraal (zoals bij discrete kansverdelingen) over de rand $\{a, b\}$ van het interval $[a, b]$. Of toch? Een klassieke eigenschap van enkelvoudige integralen is dat als je de randpunten omwisselt van volgorde, er een minteken opduikt. De oriëntatie van ons interval speelt dus een rol en een oriëntatie van het interval induceert ook een oriëntatie van de randpunten. Het punt b krijgt een positieve oriëntatie en het punt a krijgt een negatieve oriëntatie, en net dit verschil in oriëntatie zorgt voor dat minteken in de fundamentele stelling.

Zonder verdere technische discussies veralgemenen we deze ideeën naar hogere dimensies. De stelling van Stokes lees als volgt:

$$\int_M \mathrm{d}\omega = \int_{\partial M} \omega\,,$$

waar ω nu een differentiaalvorm op M is en ∂M de rand van M voorstelt. «*Aligneren de dimensies en graden van deze objecten wel?*» Een prachtig gevolg hiervan is dat de integraal van een exacte differentiaalvorm, een van de vorm $\mathrm{d}\alpha$, over een variëteit zonder rand exact nul is, want $\mathrm{d}^2 = 0$. Dit zorgt ervoor dat de meetkundige integratietheorie uiterst goed samenwerkt met (de Rham)cohomologie, nog een voorbeeld van dualiteit.



Figuur 2.3: Determinant van een 2 × 2-matrix.

Om dit hoofdstuk af te sluiten, bekijken we nog eens hoe we de transformatieregel voor Riemannintegralen kunnen afleiden uit de transformatieregel voor differentiaalvormen. In Deel 1 zagen we dat bij de overgang van een kaart U op een kaart V de differentiaalvormen transformeerden aan de hand van de Jacobiaan $J(g_{VU})$. Voor hogere k-vormen kan deze regel uitgebreid worden door een Jacobiaanse factor per 1-vorm in rekening te brengen. Neem bijvoorbeeld het voorbeeld van een 2-vorm op een tweedimensionale variëteit:

$$\begin{aligned}
\mathrm{d}v^1 \wedge \mathrm{d}v^2 &= \sum_{i,j=1}^{2} J(g_{VU})^1_i J(g_{VU})^2_j \,\mathrm{d}u^i \wedge \mathrm{d}u^j \\
&= J(g_{VU})^1_1 J(g_{VU})^2_2 \,\mathrm{d}u^1 \wedge \mathrm{d}u^2 + J(g_{VU})^1_2 J(g_{VU})^2_1 \,\mathrm{d}u^2 \wedge \mathrm{d}u^1 \\
&= \Big[J(g_{VU})^1_1 J(g_{VU})^2_2 - J(g_{VU})^1_2 J(g_{VU})^2_1 \Big] \mathrm{d}u^1 \wedge \mathrm{d}u^2\,,
\end{aligned}$$

waar we de antisymmetrie van differentiaalvormen gebruikt hebben om enkel termen $i \neq j$ te behouden. Sommigen zullen zich misschien nog uit de lessen over matrices herinneren dat de factor op de laatste lijn gelijk is aan de determinant $\det J(g_{VU})$. De berekening voor 2 × 2-matrices is in Figuur 2.3 nog eens schematisch weergegeven.

Men kan op een gelijkaardige manier aantonen dat volumevormen altijd via de determinant van de transitiefunctie transformeren. (De vectorbundel van volumevormen wordt om die reden soms ook de determinantbundel genoemd.) Aangezien integralen niet van de keuze van coördinaten afhangen, krijgen we dus

$$\int_M dy^1 dy^2 \cdots dy^n = \int_M \det J(g_{VU})\, dx^1 dx^2 \cdots dx^n \,.$$

De integratiespecialisten onder ons zullen dit onmiddellijk herkennen als de klassieke transformatieregel van Riemannintegralen. Wat in de klassieke calculus soms ad hoc verschijnt, is eigenlijk een rechtstreeks gevolg van de meetkundige natuur van integralen en differentiaalvormen.

Nu we toch terug bij de Riemannintegralen zitten, laat ons eens naar een andere belangrijke toepassing kijken. De rode draad doorheen dit boek is hoe (bijna) alle deelgebieden van de wiskunde en fysica met elkaar verweven zijn. Ook hier is dit niet anders. In de voorgaande paragrafen werd integraalrekening aan (differentiaal)meetkunde gerelateerd en nu gaan we er de lineaire algebra bijhalen.

Herinner je je nog het inproduct van vectoren? Voor een geschikte Cartesiaanse basis werd het inproduct als volgt bepaald:

$$\vec{v} \cdot \vec{w} = \sum_{k=1}^{n} v_k w_k \,.$$

Je vermenigvuldigt de componenten paarsgewijs en telt alles op. Voor rijen kunnen we exact dezelfde formule gebruiken, aangezien rijen een soort (aftelbaar) oneindige vectoren zijn:

$$(v_n)_{n\in\mathbb{N}} \cdot (w_n)_{n\in\mathbb{N}} = \sum_{k=1}^{+\infty} v_k w_k \,.$$

Het belangrijke verschil hier is dat we ons nu moeten beperken tot die rijen

die sommeerbaar zijn, dat wil zeggen dat de reeksen eindig blijven:

$$\sum_{k=1}^{+\infty} v_k^2 < +\infty\,.$$

Dit houdt onder meer in dat de rijen noodzakelijk naar 0 moeten convergeren:

$$v_n \longrightarrow 0 \qquad \text{als} \qquad n \longrightarrow +\infty\,.$$

Dat dit echter nog geen voldoende voorwaarde is, wordt bijvoorbeeld aangetoond door de *harmonische reeks*:

$$\sum_{k=1}^{+\infty} \frac{1}{k} \longrightarrow +\infty\,.$$

Nu we al oneindige sommen hebben, is de overgang naar integraal bijna triviaal. We vervangen het sommatieteken door een integraalteken en de rijen door functies:

$$\langle f \,|\, g \rangle := \int_X f(x) g(x)\, d\mu\,.$$

(We hebben opnieuw de Diracnotatie voor het inproduct ingevoerd door de relatie met kwantummechanica.) Voor elke keuze van maat μ krijg je een ander inproduct op de vectorruimte van (integreerbare) functies.

Aan de hand van het inproduct kunnen we projecties definiëren, zowel voor eindig-dimensionale vectoren als voor functies. Neem eerst een eindige, Cartesiaanse basis:

$$\left\langle e_i \,\middle|\, e_j \right\rangle = \delta_{ij} := \begin{cases} 1 & \text{als } i = j \\ 0 & \text{als } i \neq j\,. \end{cases}$$

Zo een basis wordt in het algemeen een orthonormale basis genoemd. De 'ortho' wijst op de loodrechte stand van de basisvectoren en het 'normale'

wijst op het feit dat de basisvectoren lengte 1 hebben. Als we nu het inproduct van een vector met zo een basisvector nemen, dan krijgen we:

$$\left\langle e_i \,\middle|\, \vec{v} \right\rangle = \sum_{k=1}^{n} v_k \left\langle e_i \,\middle|\, e_k \right\rangle = \sum_{k=1}^{n} v_k \delta_{ik} = v_i \,.$$

Het filtert de coëfficiënt horende bij de gegeven basisvector uit de lineaire combinatie.

Voor functies is dit echter minder voor de hand liggend. Wat zijn hier de gepaste basisfuncties? Hiervoor gaan we op bezoek bij de ingenieurs. In de signaalanalyse leert men ons dat vlakke golven een heel goede keuze vormen (en verderop zullen we snel zien dat ze nogal speciaal zijn):

$$\sin(kx) \qquad \text{en} \qquad \cos(kx)$$

met $k = \frac{2\pi}{\lambda}$ het golfgetal of inverse golflengte, of via de formule van Euler,

$$\exp(ikx) \qquad \text{en} \qquad \exp(-ikx)\,.$$

Deze 'basisfuncties' voldoen aan een soort functionele veralgemening van de orthonormalisatievoorwaarde:

$$\left\langle e^{ikx} \,\middle|\, e^{ilx} \right\rangle = \delta(k-l)\,,$$

waar $\delta(k-l)$ de Diracverdeling in het punt $k-l$ voorstelt. Het rechterlid is dus geen getal, maar een grootheid die enkel in een integraal steek houdt. De reden hiervoor is dat vlakke golven eigenlijk totaal niet integreerbaar zijn. Zowel $\sin^2$ als $\cos^2$ blijven voor eeuwig en altijd oscilleren en gaan niet naar 0, hun integraal kan nooit eindig zijn. Zoals iedere goede fysicus en ingenieur doen we echter even alsof en kijken we hoever we geraken.

Aan de hand van de vlakke golven kunnen we elke (geschikte) functie uitdrukken als een soort oneindige lineaire combinatie:

$$f(x) = \int_{\mathbb{R}} \tilde{f}(k) e^{ikx}\, dk\,,$$

waar de coëfficiënten gegeven worden door middel van de projectie op de vlakke golf:

$$\tilde{f}(k) := \left\langle e^{ikx} \,\middle|\, f \right\rangle = \int_{\mathbb{R}} f(x) e^{-ikx}\, dx\,.$$

(Zie je de gelijkenis tussen beide formules? In de kwantummechanica wordt hier gretig gebruik van gemaakt.) Voor periodieke functies blijkt trouwens dat je maar een discreet (of soms zelfs eindig) aantal golflengtes of golfgetallen nodig hebt. In dit geval wordt de integraal gereduceerd tot een reeks: de Fourierreeks. Dit resultaat dat aan de basis van de *Fourieranalyse* ligt, was al veel langer gekend onder muzikanten.

We zullen hier eens kort aantonen dat bovenstaande Fouriertransformatie wel echt steek houdt:

$$\begin{aligned}
\left\langle e^{ilx} \,\middle|\, f \right\rangle &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \tilde{f}(k) e^{ikx}\, dk \right) e^{-ilx}\, dx \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} \tilde{f}(k) e^{i(k-l)x}\, dk\, dx \\
&= \int_{\mathbb{R}} \tilde{f}(k) \left(\int_{\mathbb{R}} e^{i(k-l)x}\, dx \right) dk \\
&= \int_{\mathbb{R}} \tilde{f}(k) \delta(k-l)\, dk \\
&= \tilde{f}(l)\,.
\end{aligned}$$

Dat we de twee integralen mogen omwisselen is een toepassing van de stelling van Fubini (of toch een uitbreiding ervan).

De eigenschappen van exponentiële functies zorgen er voor dat de Fouriertransformatie bijvoorbeeld heel nuttig is bij het oplossen van differentiaalvergelijkingen. Dit zijn vergelijkingen die we niet oplossen naar een waarde voor de onbekende x (of y of welke letter dan ook), maar naar een functie van de veranderlijken. Het zijn vergelijkingen die naast de functie ook zijn afgeleiden bevat. Een belangrijk voorbeeld is de tweede

wet van Newton:

$$\vec{F}(\vec{x}) = m\frac{\mathrm{d}^2\vec{x}}{\mathrm{d}t^2}\,.$$

We zoeken een functie $\vec{x} : \mathbb{R} \to \mathbb{R}^3$ die voldoet aan bovenstaande vergelijking. (De integratieconstanten die nodig zijn om de oplossing uniek te bepalen worden vaak gegeven door geschikte *randvoorwaarden*).

Beschouw een afleidbare functie $f : \mathbb{R} \to \mathbb{R}$. De afgeleide van f kan ook uitgedrukt worden door middel van de Fouriertransformatie:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}\int_{\mathbb{R}} \tilde{f}(k)e^{ikx}\,dk &= \int_{\mathbb{R}} \tilde{f}(k)\frac{\mathrm{d}e^{ikx}}{\mathrm{d}x}\,dk \\ &= \int_{\mathbb{R}} \tilde{f}(k)ike^{ikx}\,dk \\ &= i\int_{\mathbb{R}} k\tilde{f}(k)e^{ikx}\,dk\,.\end{aligned}$$

Waar de Fouriercoëfficiënten van f gegeven worden door $\tilde{f}(k)$, worden die van zijn afgeleide f' gegeven door $k\tilde{f}(k)$. Afleiden, een analytische operatie, wordt door de Fouriertransformatie omgezet in het vermenigvuldigen met een variabele (of veelterm in het algemeen), een algebraïsche operatie.

Om deze methode aan het werk te zien, beschouwen we de Helmholtz-vergelijking:

$$\frac{\mathrm{d}^2 f}{\mathrm{d}x^2} + \lambda^2 f(x) = 0\,.$$

Door middel van de Fouriertransformatie wordt dit dan

$$\int_{\mathbb{R}} (\lambda^2 - k^2)\tilde{f}(k)e^{ikx}\,dk = 0\,.$$

Aangezien dit moet gelden voor alle waarden van $x \in \mathbb{R}$, geldt noodzakelijk dat

$$(\lambda^2 - k^2)\tilde{f}(k) = 0$$

voor alle waarden $k \in \mathbb{R}$. Daar $\tilde{f}$ niet exact nul kan zijn, aangezien het de coëfficiënten voorstelt in de Fouriertransformatie van f, moet $k^2 = \lambda^2$ of, met andere woorden,

$$k = \pm\lambda\,.$$

De oplossing van de harmonische vergelijking kan, voor een gegeven waarde van $\lambda \in \mathbb{R}$, maar twee Fouriercomponenten met golfgetal λ bevatten, een golft die naar rechts beweegt en een golf die naar links beweegt:

$$f(x) = c_+ e^{i\lambda x} + c_- e^{-i\lambda x}\,.$$

(Merk wel op dat deze golven niet echt bewegen aangezien er geen tijdsafhankelijkheid is. Daarvoor zou een factor $e^{i\omega t}$ moeten worden toegevoegd.)

Als tweede voorbeeld beschouwen we een differentiaalvergelijking voor veeltermen. De meest algemene is van de vorm

$$\frac{\mathrm{d}^l f}{\mathrm{d}x^l} = 0\,.$$

Als we hier de Fouriertransformatie op loslaten, ofwel via de algebraïsche methode zoals hiervoor ofwel via partiële integratie, bekomen we:

$$\int_{\mathbb{R}} k^l \tilde{f}(k) e^{ikx}\, dk = 0\,,$$

waar we de constante factor i^l hebben laten vallen. Net zoals in de voorgaande paragraaf zou dit moeten impliceren dat

$$k^l \tilde{f}(k) = 0$$

voor alle waarden van $k \in \mathbb{R}$, wat uiteraard niet kan. Er is hier iets vreemds aan de hand. De reden hiervoor is dat de integraal

$$\int_{\mathbb{R}} k^l \tilde{f}(k) e^{ikx}\, dk$$

eigenlijk niet goed gedefinieerd is. Zowel k^l als e^{ikx} zijn niet integreerbaar over de hele reële lijn $\mathbb{R}$. Als een goede fysicus doen we toch nog even alsof dit wel kan. We zoeken een veralgemeende functie $\tilde{f}(k)$ die in een integraal voldoet aan de bovenstaande voorwaarde:

$$\int_{\mathbb{R}} k^l \tilde{f}(k)\, dk = 0\,.$$

- Voor $l = 1$: We zoeken een functie $\tilde{f}$ zodat

  $$\int_{\mathbb{R}} k\tilde{f}(k)\, dk = 0\,.$$

  De oplossing bestaat als een maat: de Diracverdeling $\delta(k)$. In dit geval filtert hij de waarde voor $k = 0$ uit andere functies:

  $$\int_{\mathbb{R}} g(k)\delta(k)\, dk = g(0)\,.$$

- Voor $l = 2$: In dit geval dient

  $$\int_{\mathbb{R}} k^2 \tilde{f}(k)\, dk = 0$$

  te gelden. Als eerste zien we direct dat $\delta(k)$ opnieuw een mogelijke oplossing is, de Diracverdeling voldoet aan de vergelijking. De aanwezigheid van het kwadraat zorgt er echter voor dat er nog een oplossing is. Via een (formele) toepassing van partiële integratie kan je nagaan dat ook de afgeleide $\delta'(k)$ voldoet aan deze vergelijking.

Voor algemene $l \in \mathbb{N}_0$ is de oplossing van de differentiaalvergelijking dus van de vorm

$$\tilde{f}(k) = c_{l-1}\frac{\mathrm{d}^{l-1}\delta}{\mathrm{d}k^{l-1}} + \cdots + c_1\delta'(k) + c_0\delta(k)\,.$$

De omgekeerde Fouriertransformatie, samen met nog wat partiële integraties, zorgen er dan voor dat de oplossing in functie van x gegeven wordt door veeltermen van de vorm

$$f(x) = c_{l-1}x^{l-1} + \cdots + c_1 x + c_0\,.$$

De fundamentele oplossing $\frac{\mathrm{d}^l \delta}{\mathrm{d}k^l}$ lijken wel op functies, zeker met de notatie die hier gebruikt werd, maar het zijn absoluut geen gewone functies. Wat zou een afgeleide van een maat zelfs moeten voorstellen? (Wiskundigen huiveren als ze fysici dit naïef zien doen.) Hier komt het concept van dualiteit weer naar boven. Wat we hier tegenkwamen zijn zogenaamde distributies, lineaire afbeeldingen van de vectorruimte van functies naar de reële getallen.[2]

Er zijn strikt meer distributies dan er functies zijn, maar we hebben wel het geluk dat elke integreerbare functie ook een distributie definieert, namelijk de integraal met deze functie:

$$g : f \mapsto \int_{\mathbb{R}} f(x)g(x)\,dx\,.$$

Doordat de ruimte van distributies groter is dan die van functies, is hij ook een stuk rijker. Overgang op een grotere context zorgt er alweer voor dat we toegang krijgen tot meer technieken en dat bepaalde argumenten die voorheen louter formeel konden worden toegepast, nu ook daadwerkelijk hard kunnen worden gemaakt. De correcte setting voor de Fouriertransformatie en voor differentiaalvergelijkingen is dus eigenlijk de wereld van de distributies (tenzij we bepaalde extra eisen leggen op de ruimte van functies die we beschouwen, wat bijvoorbeeld tot de studie van *Sobolevruimtes* leidt).

Extra's: Fermionische integralen

Voor deze extra's beschouwen we een totaal andere integratietheorie, een die absoluut niets te maken heeft met de andere integralen uit dit hoofdstuk. We gaan integreren met fermionen! Als je je nu afvraagt wat dit te betekenen heeft, dat is normaal, het is een buitenbeentje. Bosonische deeltjes zijn fysische deeltjes die wanneer je ze omwisselt, de kwantum-

[2] Dat het woord distributie verdacht goed lijkt op de letterlijke vertaling van het Engelse woord *distribution* voor verdeling, is geen toeval. Er is een zeer nauw verband.

mechanische golffunctie onveranderd blijft. Je mag ze dus commuteren zoals gewone getallen: $xy = yx$.

Fermionen daarentegen anticommuteren: $\psi\xi = -\xi\psi$. Ze gedragen zich dus totaal anders dan de getallen die we gewend zijn. Toch willen we hier ook integralen over kunnen uitrekenen, want hoe gaan we anders een padintegraal definiëren voor fermionische theorieën? We hebben zelfs behoorlijk wat geluk, want fermionische integralen, ook wel Berezin-integralen genoemd, zijn heel wat eenvoudiger te berekenen dan de integralen van hierboven.

Om hun definitie te motiveren, kijken we even naar de basiseigenschappen van de gewone, reële integralen:

- Lineariteit:

$$\int_{\mathbb{R}} \big(f(x) + \lambda g(x)\big)\, dx = \int_{\mathbb{R}} f(x)\, dx + \lambda \int_{\mathbb{R}} g(x)\, dx\,.$$

  De integraal van een lineaire combinatie is de lineaire combinatie van de integralen.

- Invariantie:

$$\int_{\mathbb{R}} f(x + x_0)\, dx = \int_{\mathbb{R}} f(x)\, dx\,.$$

  De functie verschuiven, verandert niks aan zijn integraal. (Merk op dat dit anders ligt voor integralen over eindige intervallen.)

- (On)afhankelijkheid: De integraal van een functie met betrekking tot zijn argument x hangt niet meer af van dat argument. Integralen verminderen dus het aantal vrije parameters.

Deze eigenschappen gaan we gebruiken als definiërende eigenschappen van de fermionische integraal.

Aangezien fermionen anticommuteren, vallen alle functies te benaderen als veeltermen. Neem voor de eenvoud functies van één enkele Gras-

smannvariabele:

$$f(\theta) = a + b\theta\,.$$

Er kunnen geen hoger-ordetermen in de Taylorbenadering voorkomen, want $\theta^2 = 0$ wegens de anticommutativiteit. Zo geldt bijvoorbeeld dat $e^\theta = 1 + \theta$. Volgens de invariantievoorwaarde zou de integraal van $a + b\theta$ dus gelijk moeten zijn aan die van $a + b(\theta + \theta_0) = (a + b\theta_0) + b\theta$. De enige lineaire operatie die daaraan voldoet is

$$\int_{\mathcal{B}} : a + b\theta \mapsto \lambda b$$

voor een (reële) constante $\lambda \in \mathbb{R}$ en als normalisatievoorwaarde kiezen we $\lambda = 1$. Maar wacht eens even, dit betekent dat voor functies van Grassmannvariabelen, de integraal en de afgeleide samenvallen!

De uitbreiding naar meerdere dimensies is vrij vanzelfsprekend. Stel dat we $n \in \mathbb{N}$ Grassmannvariabelen θ_i hebben. Elke functie $f(\theta_1, \dots, \theta_n)$ kan uitgedrukt worden als een (eindige) som:

$$f(\theta_1, \dots, \theta_n) = a + \sum_{i=1}^{n} b_i\theta_i + \cdots + b_{12\dots n}\theta_1\theta_2\cdots\theta_n\,.$$

De Berezinintegraal wordt opnieuw gedefinieerd als de projectie op de hoogste component:

$$\int_{\mathcal{B}} f(\theta_1, \dots, \theta_n)\, d\theta_n \cdots d\theta_1 = b_{12\dots n}\,.$$

Let goed op de volgorde van het volume-element $d\theta_n \cdots d\theta_1$. De variabelen staan in de omgekeerde volgorde van die in de reeksontwikkeling van f. Hierdoor is deze definitie consistent met de stelling van Fubini.

Om even voorop te lopen op Hoofdstuk 5 geven we hier meer dat de Grassmannvariabelen een soort uitbreiding vormen op wat daar een *algebra* genoemd zal worden. We hebben een optelling, een scalaire vermenigvuldiging en een echte vermenigvuldiging tussen de elementen van

de algebra. In tegenstelling tot de inhoud van Hoofdstuk 5 is de algebra $\mathbb{R}^{0|n}$ gegenereerd door de Grassmannvariabelen echter anticommutatief aangezien $\theta_i\theta_j = -\theta_j\theta_i$. Dergelijke structuren worden ook wel superalgebra's genoemd. Meer algemeen wordt de (super)algebra $\mathbb{R}^{m|n}$ gegenereerd door $m \in \mathbb{N}$ 'gewone' coördinaten x^i en $n \in \mathbb{N}$ anticommutatieve Grassmannvariabelen zodat

$$x^i x^j = x^j x^i \qquad \theta^i\theta^j = -\theta^j\theta^i \qquad x^i\theta^j = \theta^j x^i\,.$$

De integraal op $\mathbb{R}^{m|n}$ is vrij eenvoudig gedefinieerd door te stellen dat de commutatieve en anticommutatieve coördinaten onafhankelijk zijn:

$$\int_{\mathbb{R}^{m|n}} f(x,\theta)\, d^n\,\theta d^m x := \int_{\mathbb{R}^m}\left(\int_{\mathcal{B}} f(x,\theta)\, d^n\theta\right) d^m x\,.$$

De transformatieregels zijn echter iets ingewikkelders. We beginnen met de gewone Berezinintegraal en een lineaire transformatie gegeven door de matrix A. In één dimensie is dit simpelweg de transformatie

$$\theta = \lambda\eta$$

voor een reële $\lambda \in \mathbb{R}_0$. Neem dus een functie $f(\theta) = a + b\theta$ en bereken de Berezinintegraal via η:

$$\int_{\mathcal{B}} f(\eta)\, d\eta = \int_{\mathcal{B}} (a + b\lambda\eta)\, d\eta = \lambda b\,.$$

De Berezinintegraal ten opzichte van θ is echter b. We zien dus dat het volume-element op de volgende manier dient te transformeren:

$$d\theta = \frac{1}{\lambda} d\eta\,,$$

exact de omgekeerde transformatieregel van de gewone integraal op $\mathbb{R}$. Voor meerdere variabelen wordt de transformatieregel gegeven door

$$\int_{\mathcal{B}} f(\theta)\, d^n\theta = \int_{\mathcal{B}} f(\eta)\, \mathrm{det}(A)^{-1}\, d^n\eta\,.$$

Waar in de gewone integralen de Jacobiaan optreedt bij de overgang op nieuwe coördinaten, verschijnt bij Berezinintegralen de inverse Jacobiaan.

We kunnen dit nu ook veralgemenen naar de superalgebra $\mathbb{R}^{m|n}$. Als eerste laten we de even coördinaten x^i transformeren onder de matrix A en de oneven coördinaten θ^j onder de matrix D. De resultaten die we tot nu toe gevonden hebben, zeggen dat de integraal als volgt transformeert:

$$\int_{\mathbb{R}^{m|n}} f(x,\theta)\, d^n\theta\, d^m x = \int_{\mathbb{R}^{m|n}} f(y,\eta)\, \mathrm{det}(A)\, \mathrm{det}(D)^{-1}\, d^n\eta\, d^m y\,.$$

We hoeven hier echter niet te stoppen. We hebben toegang tot de superalgebra $\mathbb{R}^{m|n}$, dus we kunnen ook transformaties beschouwen waarvan de coëfficiënten element zijn van deze superalgebra. Een algemene lineaire transformatie kan dus geschreven worden als een supermatrix:

$$M = \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right).$$

A en D zijn even en laten de even (resp. oneven) coördinaten onderling transformeren. B en C zijn daarentegen oneven. Zo zet B de oneven (resp. even) coördinaten om in even (resp. oneven) coördinaten. De veralgemeende transformatie regels is als volgt:

$$\int_{\mathbb{R}^{m|n}} f(x,\theta)\, d^n\theta\, d^m x = \int_{\mathbb{R}^{m|n}} f(y,\eta)\, \mathrm{det}(A)\, \mathrm{det}\big(D - CA^{-1}B\big)^{-1}\, d^n\eta\, d^m y\,.$$

De factor in het rechterlid wordt de superdeterminant van de supermatrix M genoemd:

$$\begin{aligned} \mathrm{sdet}\left(\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right) &= \mathrm{det}(A)\, \mathrm{det}\big(D - CA^{-1}B\big)^{-1} \\ &= \mathrm{det}\big(A - BD^{-1}C\big)\, \mathrm{det}(D)^{-1}\,. \end{aligned}$$

Voor de absolute puristen, de uitdrukkingen $D - CA^{-1}B$ en $A - BD^{-1}C$ worden de Schurcomplementen van D en A genoemd. Indien M volledig uit even elementen bestond, werd de determinant gegeven door

$$\mathrm{det}\left(\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right) = \mathrm{det}(A)\, \mathrm{det}\big(D - CA^{-1}B\big) = \mathrm{det}\big(A - BD^{-1}C\big)\, \mathrm{det}(D)\,.$$

We zien dus dat de superdeterminant heel nauw verwant is aan de gebruikelijke determinant.

Indien we nu veralgemenen naar transformaties $g : \mathbb{R}^{m|n} \to \mathbb{R}^{m|n}$ die niet lineair, maar wel nog afleidbaar zijn, dan krijgen we superdeterminant van de Jacobiaan, de zogenaamde Bereziniaan:

$$\mathsf{Ber}(g) := \mathsf{sdet}\left(\frac{\partial g(y,\eta)}{\partial(y,\eta)}\right).$$

Het veralgemeent de Jacobiaanse determinant uit het begin van dit hoofdstuk naar superalgebra's. Deze eigenschap is van groot belang in de studie naar supersymmetrie waar elk bosonisch (resp. fermionisch) deeltje een fermionische (resp. bosonische) partner heeft.

Deze eigenschappen hebben echter ook hun nut ver buiten de natuurkunde. Beschouw bijvoorbeeld de standaard normale verdeling:

$$\mathcal{N}(0,1) := \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Laten we eens kijken of deze verdeling wel degelijk genormeerd is. De integraal

$$\int_{\mathbb{R}} e^{-\frac{x^2}{2}}\, dx$$

is echter notoir moeilijk te berekenen door middel van de gebruikelijke integratietechnieken. Indien we van 1D overgaan op 2D bestaat er een heel elegante methode die reeds door wiskundigen zoals Poisson gekend

was.

$$\begin{aligned}\left(\int_{\mathbb{R}} e^{-\frac{x^2}{2}}\,dx\right)^2 &= \iint_{\mathbb{R}^2} e^{-\frac{1}{2}(x^2+y^2)}\,dx\,dy \\ &= \iint_{\mathbb{R}\times S^1} e^{-\frac{r^2}{2}}\,r\,d\theta\,dr \\ &= 2\pi\int_{\mathbb{R}} e^{-\frac{r^2}{2}}\,r\,dr \\ &= 2\pi\int_{\mathbb{R}} e^{-u}\,du \\ &= 2\pi\,.\end{aligned}$$

Hier hebben we gebruik gemaakt van een overgang op poolcoördinaten $(x, y) \longrightarrow (r, \theta)$ en een transformatie $u = r^2$. Een mooie integratie-oefening. Dat de dubbelintegraal gelijk is aan het kwadraat in het linkerlid is een gevolg van de stelling van Fubini en vergt normaal gezien nog enkele tussenstappen, maar laat het voldoende zijn te zeggen dat dit geoorloofd was. Alles samen bekomen we dus

$$\int_{\mathbb{R}} e^{-\frac{x^2}{2}}\,dx = \sqrt{2\pi}$$

de normalisatiefactor uit de uitdrukking van de normale verdeling.

We kunnen het voorgaande resultaat ook veralgemenen naar hogere dimensies. Beschouw een (n-dimensionale) multinormale verdeling met covariantiematrix Σ. Deze matrix beschrijft de correlaties tussen de variabelen. Op een normalisatiefactor na wordt deze verdeling door de volgende formule gegeven:

$$\mathcal{N}(0, \Sigma) = e^{-\frac{1}{2}\langle\vec{x}|\Sigma^{-1}|\vec{x}\rangle}\,.$$

Wegens de transformatieregels voor integralen weten we dat de integraal van deze verdeling kan uitgedrukt worden in termen van de ongecorre-

leerde variabelen $\vec{y} := \sqrt{\Sigma}\vec{x}$ door de Jacobiaan te introduceren:

$$\int_{\mathbb{R}^n} e^{-\frac{1}{2}\langle \vec{x}|\Sigma^{-1}|\vec{x}\rangle}\, d^n x = \int_{\mathbb{R}^n} \frac{e^{-\frac{1}{2}\langle \vec{y}\,|\,\vec{y}\rangle}}{\sqrt{\det(\Sigma)}}\, d^n y = \sqrt{\frac{(2\pi)^n}{\det(\Sigma)}}\,.$$

Hoe sterker de correlaties, hoe 'gepiekter' de verdeling zal zijn.

Met deze *Gaussische integralen* krijgen we een uitdrukking waarbij de determinant van de matrix in de noemer staat. We weten echter dat Berezinintegralen op de omgekeerde manier transformeren. Laten we daar dus ook eens naar kijken. Neem een algemene (even) $n \times n$-matrix A en twee n-dimensionale Grassmannvariabelen θ, η. Voor $n = 1$ komt A neer op vermenigvuldiging met een scalair $\lambda \in \mathbb{R}$ (of in $\mathbb{C}$).

$$\begin{aligned}\int_{\mathcal{B}} e^{-\langle \theta|A|\eta\rangle}\, d\theta\, d\eta = \int_{\mathcal{B}} e^{-\lambda\theta\eta}\, d\theta\, d\eta &= \int_{\mathcal{B}} e^{\lambda\eta\theta}\, d\theta\, d\eta \\ &= \int_{\mathcal{B}} (1 - \lambda\eta\theta)\, d\theta\, d\eta \\ &= \lambda\,,\end{aligned}$$

waar we gebruik gemaakt hebben van de Grassmanneigenschappen van θ en η en de reeksontwikkeling van de exponentiële functie. Voor vectoren gebruiken we de transformatieregel van de Berezinintegraal:

$$\int_{\mathcal{B}} e^{-\langle \theta|A|\eta\rangle}\, d^n\theta\, d^n\eta = \det(A) \int_{\mathcal{B}} e^{-\langle \theta\,|\,\tilde{\eta}\rangle}\, d^n\theta\, d^n\tilde{\eta} = \det(A)\,,$$

waar $\tilde{\eta} := A\eta$. «*In de reeksontwikkeling van de exponentiële functie staat in de ne term normaal nog een factor n!. Kan jij bedenken waarom deze wegvalt?*

We kunnen dus in het algemeen determinanten berekenen aan de hand van fermionische, Gaussische integralen. Dit is trouwens een van oorspronkelijke manieren waarop de spookdeeltjes werden ingevoerd in de veldentheorie. Er moesten determinanten van ijktransformaties berekend worden en Faddeev en Popov bewerkstelligden dit door de padintegraal uit te breiden met fermionische vrijheidsgraden.

De Berezinintegraal heeft ook nog een bijkomend voordeel over de gebruikelijke integralen in het padintegraalformalisme van de kwantumveldentheorie. De bosonische padintegraal is notoir slecht gedefinieerd. Slechts in enkele Gaussische gevallen kan de integraal daadwerkelijk als een integraal gedefinieerd worden (en Gaussische padintegralen zijn de eenvoudigste gevallen waarbij er geen interacties voorkomen tussen de deeltjes). De fermionische padintegralen zijn daarentegen altijd te definiëren als een soort oneindig-dimensionale Berezinintegralen. In het algemeen zal de actie van onze kwantumveldentheorie kunnen geschreven worden als de som van een bosonische term en een fermionische term:

$$S[\phi, \psi] := S[\phi] + S_\phi[\psi]\,,$$

waar de index ϕ aangeeft dat de fermionische actie kan afhangen van de bosonische vrijheidsgraden ϕ wanneer er interacties mogelijk zijn tussen fermionen en bosonen. Zoals wanneer we ijkvelden of zwaartekracht toevoegen. De padintegraal krijgt dan de (formele) uitdrukking

$$\int e^{iS[\phi,\psi]}\, D\psi\, D\phi\,.$$

Zoals voorheen berekenen we eerst de binnenste, fermionische integraal:

$$\mathsf{Pfaff}(\phi) := \int e^{iS_\phi[\psi]}\, D\psi\,,$$

De *Pfaffiaan* $\mathsf{Pfaff}(\phi)$ is echter niet altijd een mooie functie van de achtergrondsvelden ϕ. Om even voorop te lopen op het volgende hoofdstuk, we kunnen zeggen dat het mogelijk is dat de Pfaffiaan slechts een *doorsnede* is van een *lijnbundel*. Lokaal is het een functie, maar globaal kunnen er obstructies zijn om de verschillende lokale uitdrukking samen te 'plakken'. Dit is een voorbeeld van een (kwantum)anomalie.

Stel dat de Pfaffiaan wel een echte functie is van de bosonische parameters. Wat stelt deze functie dan voor? In de meeste gevallen is de fermionische actie van de vorm

$$S_\phi[\psi] = \left\langle \psi \,\middle|\, \nabla_\phi \psi \right\rangle .$$

Merk op dat dit bijzonder goed lijkt op de Gaussische integralen die we eerder besproken hebben. (De operator ∇_ϕ staat trouwens gekend als de *Diracoperator* en is een soort covariante afgeleide voor fermionische velden.) Het grote verschil met de fermionische integraaluitdrukking van de determinant van een matrix is dat we hier twee keer dezelfde Grassmannvariabele ψ in het inproduct staan hebben. Als gevolg van de anticommutativiteit, moet de Diracoperator dus antisymmetrisch zijn:

$$\left\langle \psi \middle| \nabla_\phi \eta \right\rangle = -\left\langle \nabla_\phi \psi \middle| \eta \right\rangle .$$

Als we dan dezelfde afleiding volgen als in de voorgaande paragrafen, dan bekomen we

$$\int_{\mathcal{B}} e^{-\left\langle \theta \middle| \nabla_\phi \psi \right\rangle} d^n\psi = \sqrt{\det(\nabla_\phi)} ,$$

We krijgen de vierkantswortel van de determinant. «*Probeer eens te beredeneren waarom de vierkantswortel hier tevoorschijn komt.*» In de lineaire algebra staat de vierkantswortel van de determinant (van een antisymmetrische) matrix gekend als, je raadt het waarschijnlijk al, de Pfaffiaan.

Hoofdstuk 3

Meetkunde: Bundels en instantonen

Dit is een hoofdstuk waar we een brug gaan bouwen tussen de topologische wereld uit het eerste hoofdstuk en de meetkundige wereld uit Deel 1. De (gladde) variëteiten uit Deel 1 zijn uiterst belangrijke voorbeelden van topologische ruimtes in de natuurkunde. In dit hoofdstuk gaan we wat verder voortbouwen op die meetkundige constructies en zien hoe de vectorvelden en differentiaalvormen gerelateerd zijn aan zogenaamde 'bundels'.

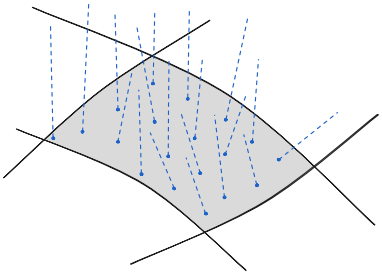
Variëteiten werden gedefinieerd als verzamelingen die lokaal beschreven konden worden aan de hand van Cartesiaanse coördinaten. We kunnen echter een stapje verder gaan. Zoals in de figuur op onderstaande pagina, kunnen we een variëteit beschouw waarop op elk punt een 'vezel' is aangebracht. Lokaal is zo een vezelbundel dus niet meer van de vorm $\mathbb{R}^n$, maar eerder van de vorm $\mathbb{R}^n \times V$ voor een geschikte verzameling V (we zorgen er dus voor dat de vezels in elk punt dezelfde structuur hebben). Bij variëteiten moesten de coördinaten in overlappende kaarten op een consistente manier in elkaar kunnen worden omgezet via transitie-

functies φ_{UV} en hier is dit niet zozeer het geval. Op overlappende kaarten van de basisvariëteit dient er een transitiefunctie g_{UV} te bestaan die de vezelcoördinaten op een consistente manier in elkaar transformeert. Op een drievoudige overlap moet bovendien de voorwaarde

$$g_{UW} = g_{UV} \circ g_{VW}$$

te gelden. Dit was de regel die we onder andere voor de raakbundel in Deel 1 gezien hadden.

In dit hoofdstuk zullen we twee grote klassen van vezelbundels bespreken: die waarvoor V een vectorruimte is en die waarvoor V een groep is. (Het zal ook blijken dat die groepen uit de tweede groep vaak overeenkomen met de symmetriegroepen van de vectorruimtes uit de eerste groep.) Het 'enige' wat alle bundels gemeenschappelijk hebben, is het bestaan van een projectie $\pi : E \to M$ van de bundel E naar de basisvariëteit M, die lokaal de informatie over de vezel V weggooit.



Figuur 3.1: Lokale voorstelling van een vezelbundel.

Een van de essentiële ingrediënten voor meetkundige structuren was hun lokale transformatieregel. Voor variëteiten waren dit de coördinatentransformaties g_{UV} tussen twee kaarten V en U, voor vectorvelden waren dit de Jacobianen $J(g_{UV})$ van die transformaties en voor differentiaalvormen waren het de inverse Jacobianen $J^{-1}(g_{UV})$. Die (inverse) Jacobiaan is de functie of operator die inwerkt op de tweede factor van zo een lokale 'kaart' $\mathbb{R}^n \times V$. Dit verklaart (deels) waarom de transformatieregel zo sterk gerelateerd is aan de coördinatentransformatie, ze werken samen in op de lokale kaarten. De coördinatentransformatie werkt in op de 'horizontale' richtingen langsheen de variëteit en de andere, geïnduceerde transformatie op de 'verticale' richtingen van de vezel.

Er bestaat ook een algemene constructie voor vezelbundels die we de koppelingsconstructie noemen.[1] We starten met een collectie open verzamelingen $\{U_i\}_{i\in I}$ die de basisvariëteit M bedekken en een modelvezel V. Op deze manier bekomen we een verzameling lokale productvariëteiten $\{U_i \times V\}_{i\in I}$. Stel nu dat we een manier hebben om punten te identificeren

$$(x_i, v_i) \sim (x_j, v_j)\,,$$

bijvoorbeeld door middel van een inverteerbare functie $f : V \to V$ die zegt dat $(x_j, v_j) = (x_i, f(v_i))$, en dat deze relaties bovendien consistent zijn

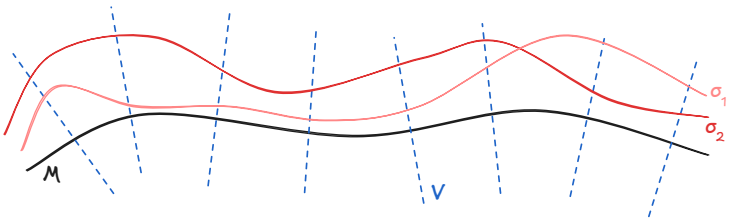
$$(x_i, v_i) \sim (x_j, v_j) \text{ en } (x_j, v_j) \sim (x_k, v_k) \implies (x_i, v_i) \sim (x_k, v_k)\,,$$

dan kunnen we een quotiënt $E := \bigcup_{i\in I} U_i \times V/_\sim$ construeren. We gooien alle koppels samen in een verzameling en identificeren overeenkomstige koppels door gebruik te maken van de consistente relaties. (Deze komen uiteraard overeen met wat we hierboven de transitiefuncties noemden.) De ruimte E is nu een vezelbundel over M waarbij de projectie $\pi : E \to M$ een koppel (x_i, v_i) simpelweg afbeeldt op x_i. Lokaal is de vezelbundel duidelijk triviaal, boven elke kaart U_i hebben we een productstructuur, maar globaal is dit niet meer het geval ten gevolge van de identificaties.

Een van de belangrijkste klassen van bundels bestaat uit de vezelbundels waarbij de modelvezel V gegeven wordt door een vectorruimte, de klasse van vectorbundels. Deze worden geconstrueerd aan de hand van transitiefuncties die niet enkel inverteerbaar zijn, ze moeten ook lineair zijn. In dit geval kunnen we op elk punt $x \in M$ van de basisvariëteit de elementen van de bijhorende vezel optellen en vermenigvuldigen met een reëel (of complex) getal. Let wel goed op, elke vezel vormt een vectorruimte op zich, maar je kan niet zomaar twee elementen uit verschillende vezels optellen, ook al zijn de vezels isomorf. Willen we dit wel kunnen doen, dan hebben we de notie van een connectie of covariante afgeleide nodig zoals in Deel 1. We geven ook nog even mee dat indien elke vezel een eendimensionale vectorruimte vormt, we ook wel van een lijnbundel spreken.

[1] In het Engels wordt dit de *clutching construction* genoemd.

De overeenkomsten tussen vezelbundels en de meetkundige constructies uit Deel 1 — ze transformeren lokaal op een mooi voorgeschreven manier, ze kunnen globaal geconstrueerd worden door lokale objecen samen te plakken en we kunnen elementen op verschillende punten niet zo maar vergelijken zonder rekening te houden met topologische eigenschappen — zijn uiteraard geen toeval. Vectorvelden en differentiaalvormen zijn twee klassieke voorbeelden van doorsnedes $\phi \in \Gamma(E)$ van vectorbundels. Dit zijn continue of zelfs gladde keuzes van een element uit de vezel boven elk punt zoals hieronder afgebeeld. In het geval van een triviale vectorbundel $E \cong M \times V$ komt dit neer op een continue of gladde functie, maar voor algemene vectorbundels kunnen er 'draaien' in zitten. We zullen hier verder in dit hoofdstuk nog voorbeelden van zien wanneer we het hebben over materiaalfysica.



Figuur 3.2: Twee (lokale) doorsnedes van een eendimensionale vectorbundel.

Deze meetkundige doorsnedes hebben ook fysische realisaties in de vorm van de velden die (fundamentele) deeltjes voorstellen. Elk deeltje, zij het een elektron, gluon of proton, komt overeen met een doorsnede van een geschikte vectorbundel, waarbij de basisvariëteit van die bundel gegeven wordt door de ruimtetijd. Die overeenkomst tussen de deeltjesfysica en de differentiaalmeetkunde, waar we een deel van gezien hebben bij de bespreking van ijktheorieën in Deel 1 en die we in dit hoofdstuk verder gaan uitwerken, wordt ook wel de Wu–Yang-woordenboek genoemd.[2]

[2]Vernoemd naar de theoretici Wu en Yang wiens werk aan de basis ligt van ontelbare relaties tussen de theoretische natuurkunde en de wiskunde.

De golffuncties in de kwantummechanica, zeker wanneer ze beschreven worden aan de hand van geometrische kwantisatie zijn ook doorsnedes van een eendimensionale vectorbundel, waarbij hun waarde of amplitude op een punt overeenkomt met hun verticale component. Vanuit dit perspectief is het dan ook logisch dat zowel bij geometrische kwantisatie als in de deeltjesfysica de kromming opdook. De meetkundige structuur van deze theorieën is nagenoeg identiek.

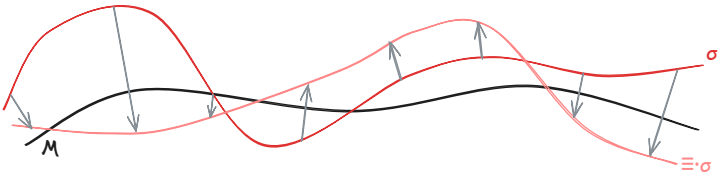
De klasse van vezelbundels waarvoor V een groep is, krijgt de naam hoofdvezelbundels (op wat technische details na). Om het onderscheid te maken met de vectorbundels, zullen we vanaf hier de vezel aanduiden met hetzelfde symbool G als de structuurgroep waarop ze gemodelleerd zijn. Dergelijke hoofdvezelbundels zijn van belang in de studie van ijktheorieën, waar G overeenkomt met de symmetriegroep of ijkgroep van de theorie. Lokaal is zo een hoofdvezelbundel van de vorm $\mathbb{R}^n \times G$. We zien dus dat de groep G op elk punt $x \in M$ van de basisvariëteit kan inwerken op de lokale vezel (de actie werkt langs rechts om vrij technische redenen):

$$(x, h) \lhd g = (x, hg)\,.$$

Door de keuze van g op een continue (of gladde) manier te laten afhangen van het punt x, krijgen we een positie-afhankelijke transformatie die de structuur van de vezelbundel behoudt. Het is een soort isomorfisme van hoofdvezelbundels waarbij punten in de vezels in elkaar worden omgezet, maar nooit de vezel zelf verlaten. Deze transformaties worden ook wel ijktransformaties genoemd en niet zonder goede reden. Het zijn exact de ijktransformaties uit de fysica!

Laten we nog wat op hetzelfde elan verdergaan. Neem een vectorbundel $E \to M$. Op elk punt $x \in M$ kunnen we symmetrietransformaties van de vectorruimte 'boven' dat punt beschouwen. Dergelijke transformaties vormen een groep daar ze inverteerbaar en samenstelbaar moeten zijn. Op deze manier krijgen we een groep boven elk punt en bovendien zijn al deze groepen ook isomorf, aangezien de vezels zelf isomorf zijn. We bekomen dus een soort meetkundige structuur die wel verdacht veel lijkt op

een hoofdvezelbundel. Zoals gewoonlijk in de wiskunde bestaat er niet zoiets als toeval en lijkt deze structuur dus niet enkel op een hoofdvezelbundel, het is er ook echt een: de framebundel van E. In zekere zin bestaat de vezel van de framebundel uit alle basissen van de vectorruimte waarop E gemodelleerd is.



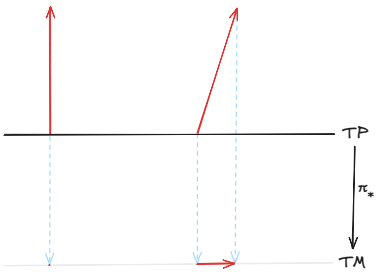
Figuur 3.3: (Lokale) ijktransformatie van een vezelbundel.

Je kan trouwens ook in de andere richting te werk gaan. Startende van een hoofdvezelbundel kan je voor elke vectorruimte waarop de modelgroep inwerkt een bijhorende vectorbundel bouwen. Op deze manier kan je een overvloed aan vectorbundels creëren en dat is nu net hoe het Standaardmodel van de deeltjesfysica aanleiding geeft tot het grote aantal deeltjes dat we in de natuur waarnemen. De totale ijkgroep van het Standaardmodel is $G_{\text{SM}} = \mathsf{U}(1) \times \mathsf{SU}(2) \times \mathsf{SU}(3)$, waarbij de factoren respectievelijk voor het elektromagnetisme, de zwakke wisselwerking en de sterke wisselwerking zorgen. Met de groep G_{SM} kunnen we nu een hoofdvezelbundel P_{SM} construeren. Alle deeltjes (of velden) in de natuurkunde worden op hun beurt bekomen als doorsnedes van vectorbundels geassocieerd aan P_{SM} door een geschikte representatie van G_{SM}, een vectorruimte waarop G_{SM} op een lineaire manier inwerkt. Zo heb je bijvoorbeeld de fotonen die enkel onder de eerste factor $\mathsf{U}(1)$ transformeren en de elektronen die onder de twee factoren $\mathsf{U}(1) \times \mathsf{SU}(2)$ transformeren.

Die hoofdvezelbundels zijn ook de natuurlijke territoria van covariante afgeleiden en territoria. Door hun (lokale) productstructuur krijgen ook de raakruimtes in elk punt een productstructuur. Een deel van de raakrichtingen is ‘evenwijdig’ aan de basisvariëteit M en een deel is evenwijdig

aan de vezels, maar slechts een van deze twee is uniek gedefinieerd. Alle raakvectoren die op nul worden afgebeeld onder de projectie $\pi : P \to M$ zijn verticaal, ze liggen volledig langsheen de vezel. Voor de horizontale component is de situatie echter minder voor de hand liggend, want hoe definiëren we horizontaal als we enkel de projectie π hebben?

We moeten hier echter een keuze maken, met als gevolg dat covariante afgeleiden en connecties niet uniek zijn, wat we trouwens in Deel 1 reeds waren tegengekomen toen we een keuze moesten maken van welke raakvectoren we als evenwijdig beschouwden. Het is hier dat de connectie A of ω — de notatie hangt af van of je een wiskundige of een fysicus bent — opduikt aangezien de connectie bepaalt of een raakvector horizontaal is of niet. Waar π de verticale vectoren op nul afbeeldt, is het de connectie die de horizontale vectoren op nul afbeeldt. In deze context kunnen we nu ook zeggen dat de kromming van een variëteit M eigenlijk de kromming van zijn raakbundel TM is.



Figuur 3.4: Links: een verticale vector. Rechts: Een vector met een 'horizontale' component.

De voorgaande paragrafen waren heel meetkundig van aard, waarbij we gebruikt maakten van de lokale, gladde structuur van de variëteiten en vezelbundels. De constructies van connecties en hun bijhorende covariante afgeleiden zeggen echter niet veel over de globale, topologische structuur. Het was voor wiskundigen dan ook een enorme verrassing toen bleek dat de lokale structuur gegeven door de connecties, een structuur die zelfs niet uniek is voor een gegeven vezelbundel, toch heel wat topologische informatie bevat. De Chern–Weil-theorie die deze lokale en globale aspecten met elkaar verbindt, bleek later ook van belang te zijn in de theoretische fysica waar het toelaat om onderscheid te maken tussen

zogenaamde topologische sectoren.

Zoals we in het vorige hoofdstuk zagen, kunnen we voor elke topologische ruimte een aantal algebraïsche invarianten berekenen. Twee topologische ruimtes met verschillende invarianten kunnen nooit of te nimmer isomorf zijn en Chern–Weil-theorie zal ons in staat stellen om (sommige van deze) invarianten te berekenen aan de hand van de kromming.

Beschouw een (hoofd)vezelbundel $\pi : P \to M$ en kies een connectie A. De structuurvergelijking van Cartan zegt ons hoe we de geassocieerde kromming kunnen berekenen:

$$F_A := D_A A = \mathrm{d}A + A \wedge A \,.$$

Eén vergelijking en twee nieuwe notaties, dat verdient wel wat uitleg, ook al zijn we beide objecten stiekem al tegengekomen in Deel 1. De laatste term in deze uitdrukking is het wigproduct[3]. Waar we vectoren combineren met het tensorproduct, combineren we differentiaalvormen let het wigproduct. In lokale coördinaten zijn differentiaalvormen antisymmetrisch in hun indices, bijvoorbeeld

$$\omega_{ij} = -\omega_{ji}$$

voor een 2-vorm ω. Het wigproduct laat ons toe om een product van differentiaalvormen te berekenen op een zodanige manier dat de antisymmetrie behouden blijft. In lokale coördinaten nemen we het product van de coëfficiënten en 'antisymmetrizeren' we. Voor twee 1-vormen wordt dit bijvoorbeeld

$$(\kappa \wedge \rho)_{ij} = \kappa_i \rho_j - \kappa_j \rho_i \,.$$

De tweede nieuwigheid, de operator D_A in de structuurvergelijking, stelt de covariante afgeleide op de bundel voor. Deze vergelijking zegt ons dus dat de kromming simpelweg de covariante afgeleide van de connectie

[3] De Engelse, meer frequente term is *wedge product*.

is! Als je de structuurvergelijking van Cartan naast de uitdrukking voor de veldsterkte van (niet-Abelse) ijktheorieën uit Deel 1 zou leggen, dan kan je zien dat beide vergelijkingen wel degelijk identiek zijn:

$$F_{ij} = \partial_i A_j - \partial_j A_i + A_i A_j - A_j A_i \,.$$

Gewapend met deze kennis over meetkunde en topologie is het tijd om onze geliefde kromming aan het werk te zetten voor de berekening van topologische invarianten van variëteiten en hun geassocieerde vezelbundels. Geëvalueerd in een punt $x \in M$ kan de kromming in het algemeen uitgedrukt worden als een matrix (voor U(1) is het zelfs een getal, maar getallen zijn 1×1-matrices). Stel nu dat we op de verzameling van matrices een functie

$$\langle -, -, \cdots, - \rangle : A \mapsto \langle A, A, \cdots, A \rangle$$

kunnen definiëren die invariant is onder de actie van de ijkgroep, hij verandert dus niet onder ijktransformaties:

$$\langle g^{-1}Ag, g^{-1}Ag, \cdots, g^{-1}Ag \rangle = \langle A, A, \cdots, A \rangle \,.$$

Op deze manier kunnen we een reeks $2k$-vormen maken door de algemene matrix te vervangen door de kromming: $\langle F_A, F_A, \cdots, F_A \rangle$. Deze differentiaalvormen worden ook wel karakteristieke klassen genoemd. Enkele belangrijke voorbeelden van karakteristieke klassen, aangeduid met $c_k(P)$, worden de Chernklassen van de vezelbundel P genoemd. Ze spelen onder meer een rol in *Chern–Simonstheorieën* die in zowel hoge-energiefysica, algemene relativiteit als materiaalfysica gebruikt worden. Deze worden als volgt berekend. Beschouw de determinant

$$\det\left(\mathbb{1} + \frac{i}{2\pi} tA\right)$$

voor een matrix A en schrijf dit uit als een Taylorreeks:

$$\det\left(\mathbb{1} + \frac{i}{2\pi} tA\right) = 1 + c_1(A)t + c_2(A)t^2 + \cdots + c_n(A)t^2 \,.$$

Als we A opnieuw vervangen door de kromming F_A, dan komen de coëfficiënten $c_i(A)$ in bovenstaande uitdrukking exact overeen met de Chernklassen van P. Daar de determinant wel degelijk invariant is onder de matrixtransformatie $A \mapsto g^{-1}Ag$, zijn de Chernklassen ook invariant onder ijktransformaties en keuzes van connecties, ze hangen enkel af van de kromming.

Nu hebben deze klassen nog enkele uiterst interessante eigenschappen. Een ervan is dat ze maar een discreet aantal waarden kunnen aannemen (dit zal duidelijker worden in het volgende hoofdstuk eens we integralen gezien hebben). Dit impliceert onder meer dat we voor elke variëteit maar een discreet aantal verschillende hoofdvezelbundels met een gegeven ijkgroep kunnen construeren. Bovendien geldt de volgende essentiële eigenschap:

$$\mathrm{d}c_k(P) = 0\,.$$

De Chernklassen zijn gesloten, de meetkundige voorwaarde opdat een differentiaalvorm een klasse in (de Rham)cohomologie zou definiëren. We hebben dus (een deel van) de topologische invarianten teruggevonden. Een echte fysische oplossing voor een ijktheorie is dus niet enkel een kromming of veldsterkte die uit de bewegingsvergelijkingen volgt, het is ook een keuze van hoofdvezelbundel. De Chernklassen die deze bundel karakteriseren worden meestal gegeven door de keuze van randvoorwaarden.

Voor $k = 1$ hebben we trouwens een heel mooie interpretatie waar we geen connecties of krommingen voor nodig hebben (toch zeker niet in het geval van lijnbundels). Als we twee kaarten U_i en U_j beschouwen, hebben we op hun overlap de transitiefunctie g_{ij}. Voor (complexe) lijnbundels is dit een functie $g_{ij} : U_i \cap U_j \to \mathbb{C}\backslash\{0\}$ en kunnen we dus een logaritme nemen:

$$f_{ij} := \frac{1}{2\pi i}\ln(g_{ij})\,.$$

De eerste Chernklasse is dan simpelweg (lokaal) gelijk aan de (uitwen-

dige) afgeleide van deze functie:

$$c_1(L)|_{U_i \cap U_j} = \frac{1}{2\pi i} \mathrm{d} \ln(f_{ij}) \,.$$

Dat de eerste Chernklasse topologische en dus globale informatie draagt, volgt onmiddellijk uit deze uitdrukking ook al is die slechts lokaal gedefinieerd. De transitiefuncties hangen helemaal niet af van lokale data, zoals een connectie, die we op een bundel kunnen definiëren. Chern–Weiltheorie is een uiterst krachtig hulpmiddel in de studie van topologie en differentiaalmeetkunde. Bovendien geeft het ook een relatie tussen 'gewone' cohomologie en differentiële cohomologie.

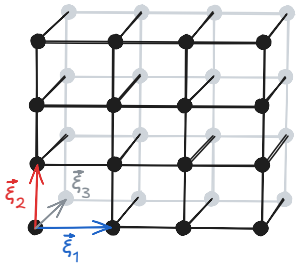
De differentiële cohomologie is de wereld waarin het exotische object $\mathbf{B}G_{\mathrm{conn}}$ uit Deel 1 leefde. De gewone cohomologiegroep $H^1(M;G)$ classificeert vezelbundels met structuurgroep G. Je kan dit zien door te kijken naar de definities van cocycli en transitiefuncties:

$$\omega_{jk} - \omega_{ik} + \omega_{ij} = 0 \qquad \text{en} \qquad g_{ij} g_{jk} g_{ik}^{-1} = 0 \,.$$

Op het omwisselen van optelling en vermenigvuldiging na zijn deze uitdrukkingen nagenoeg identiek. Dit is uiteraard exact dezelfde reden waarom de eerste Chernklasse gegeven wordt door de transitiefuncties, waar de logaritme de vermenigvuldiging in een optelling omzette. Dat er slechts een discreet aantal mogelijke hoofdvezelbundels met ijkgroep G bestaan op M is het gevolg van de structuur van $H^1(M;G)$. Deze is altijd discreet.

Dat eenzelfde vezelbundel kan worden uitgerust met verschillende niet-equivalente connecties is jammer genoeg een eigenschap die niet door globale, topologische data kan worden gedetecteerd. De cohomologie $H^1(M;G)$, die altijd discreet is, is dus niet voldoende. Er bestaat echter een *lift* van de gewone cohomologie naar de differentiële cohomologie die ook rekening houdt met lokale data. Onder meer de Belgische wiskundige Deligne heeft hier grote bijdragen aan geleverd. Waar in synthetische, categorische termen de groep $H^1(M;G)$ gegeven wordt door morfismen $M \to \mathbf{B}G$ wordt de differentiële cohomologiegroep $\check{H}^1(M;G)$ gegeven door morfismen $M \to \mathbf{B}G_{\mathrm{con}}$. $\mathbf{B}G$ weet niks over de topologische of gladde structuur van de groep G, terwijl $\mathbf{B}G_{\mathrm{con}}$ deze informatie wel onthoudt.

Vooraleer we overgaan naar een totaal andere toepassing van vezelbundels, bekijken we nog een stukje ijktheorie. @@ instantonen @@



Figuur 3.5: Een kristal in 3D.

Bundels duiken trouwens niet enkel op in ijktheorieën, ze spelen ook een belangrijke rol in de vastestoffysica. Beschouw een kristal Γ in de vlakke ruimte $\mathbb{R}^n$ zoals hiernaast. Dit bestaat uit een verzameling punten die op een regelmatige manier doorheen verdeeld zijn. In het bijzonder hebben we $n \in \mathbb{N}$ richtingen of vectoren $\vec{\xi}_i$ zodat we een roosterpunt zullen terugvinden op de posities $\vec{\xi}_i$, $2\vec{\xi}_i$, $3\vec{\xi}_i$, $\vec{\xi}_i + \vec{\xi}_j$, etc. De lineaire combinaties met gehele coëfficiënten bepalen het gehele kristalrooster. Aangezien deze punten, gegeven de verzameling van basisvectoren, worden bepaald door koppels van n gehele getallen, is het rooster Γ equivalent aan de groep $\mathbb{Z}^n$. We hoeven dus geen speciale formules te gebruiken om te rekenen met deze groep, de optelling volstaat.

Aangezien het rooster zich met regelmaat herhaalt, is het ook logisch om aan te nemen dat de Hamiltoniaan van een systeem dat op dit rooster leeft, met name zijn potentiaalterm, dezelfde periodiciteit als het rooster heeft. Naïef zouden we dan verwachten dat elke oplossing van de Schrödingervergelijking ook diezelfde periodiciteit heeft, maar daar vergeten we een kleine nuance. Zoals we in Hoofdstuk 5 in meer detail gaan zien, zorgt de regel van Born ervoor dat de oplossingen slechts quasiperiodiek moeten zijn:

$$\psi(\vec{r} + \vec{\xi}) = \lambda(\vec{\xi})\psi(\vec{r})\,.$$

De functie $\lambda : \Gamma \to \mathrm{U}(1)$ is echter niet arbitrair. We kunnen deze voorwaarde herhalen voor $\vec{\xi} + \vec{\xi}'$ op twee manieren en die horen uiteraard gelijk te zijn aan elkaar:

$$\lambda(\vec{\xi} + \vec{\xi}')\psi(\vec{r}) = \lambda(\vec{\xi})\lambda(\vec{\xi}')\psi(\vec{r})\,.$$

De functie λ dient de groepsstructuur van Γ te behouden of, met andere woorden, het is een groepsmorfisme $\Gamma \to \mathsf{U}(1)$. Alle groepsmorfismen vormen samen een nieuwe groep $\hat{\Gamma}$, de Pontryagin-duale van Γ, waarin de vermenigvuldiging gegeven wordt door puntsgewijze vermenigvuldiging op Γ:

$$(\lambda \cdot \chi)(\vec{\xi}) := \lambda(\vec{\xi})\chi(\vec{\xi})\,.$$

Deze duale groep speelt onder meer in de analyse en groepentheorie een heel belangrijke rol.

Hier stelt de Pontryagin-duale ons in staat om de Fouriertransformatie uit het vorige hoofdstuk te veralgemenen. Daar zagen we dat elke periodieke functie op $\mathbb{R}$ kan geschreven worden als een som van exponentiële functies

$$f(x) = \sum_{k=-\infty}^{+\infty} \tilde{f}_k e^{ikx}$$

en, meer algemeen, dat elke integreerbare functie geschreven kan worden als een integraal van exponentiële functies

$$g(x) = \int_{\mathbb{R}} \tilde{g}(k) e^{ikx}\, dk\,.$$

Het blijkt dat de Pontryagin-duale van $\mathbb{R}$ opnieuw $\mathbb{R}$ is en het is eigenlijk deze duale reële lijn waarover we integreren. De veralgemening die we zoeken, zal dus ook een som of integraal over de Pontryagin-duale $\hat{\Gamma}$ bevatten. Zo bekomen we een transformatie die integreerbare functies op $\mathbb{R}^n$ met quasiperiodieke functies op $\mathbb{R}^n \times \hat{\Gamma}$ identificeert:

$$\tilde{f}(\lambda, \vec{r}) = \sum_{\vec{\xi} \in \Gamma} \lambda(\vec{\xi})^{-1} f(\vec{r} + \vec{\xi})\,.$$

«*Heb je zin om eens na te rekenen dat dit steek houdt?*»

In de natuurkunde staat deze relatie gekend als de stelling van Bloch(–Floquet). Elke oplossing van de Schrödingervergelijking op een rooster

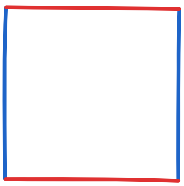
wordt gegeven door een functie van de vorm

$$\psi_{\vec{k}}(\vec{r}) = e^{i\vec{k}\cdot\vec{r}} u_{\vec{k}}(\vec{r})$$

waar $u_{\vec{k}} : \mathbb{R}^n \to \mathbb{C}$ periodiek is. «*Het is eenvoudig om na te gaan dat deze functie quasiperiodiek is.* De vector $\vec{k} \in \hat{\Gamma}$ in deze Blochgolf wordt ook wel de kristalimpuls genoemd (maar verwar deze niet met de echte impuls van deeltjes).

Merk op dat de Blochgolven verdacht goed lijken op de vlakke golven die we in de Fouriertransformatie gebruiken. De reden hiervoor is dat vlakke golven de exacte oplossingen zijn voor 'vrije' systemen waarbij de potentiaal nul is (of ten minste constant is). Een constante functie is periodiek voor eender welke periode $a \in \mathbb{R}$. Aangezien de functie $u_{\vec{k}}$ periodiek moet zijn met dezelfde periode als de potentiaal, moet ook deze een constante zijn, met als gevolg dat de Blochgolven voor vrije deeltjes gegeven worden door vlakke golven.

De quasiperiodieke functies kunnen we ook op een andere manier interpreteren als doorsnedes van een vectorbundel. Voor het gemak kiezen we een vast element $\lambda \in \hat{\Gamma}$ — deze elementen worden ook wel karakters genoemd — en beschouwen we een λ-quasiperiodieke functie $f_\lambda : \mathbb{R}^n \to \mathbb{R}$. Het quasiperiodieke karakter laat ons toe om de koppelingsconstructie van vectorbundels toe te passen waarbij het quotiënt bekomen wordt door $f_\lambda(\vec{r} + \vec{\xi})$ en $\lambda(\vec{\xi}) f_\lambda(\vec{r})$ te identificeren. Op deze manier bekomen we voor elk karakter een vezelbundel $\mathcal{L}_\lambda$ over het quotiënt $\mathbb{R}^n/\Gamma$.



Figuur 3.6: Een opengeknipte torus.

Je vraagt je nu misschien af waarom we als basisvariëteit niet meer de gehele ruimte $\mathbb{R}^n$ gebruiken, maar het quotiënt $\mathbb{R}^n/\Gamma$. Dit is een gevolg van de periodiciteit van het rooster. Aangezien de ruimte zich keer op keer herhaalt, dienen we slechts eenmalig te beschrijven wat er in een fundamenteel gebied tussen de roosterpunten gebeurt. We hebben trouwens geluk want de quotiënten die we hier tegenkomen, hebben een heel eenvoudige vorm.

In één dimensie komt het neer op het identificeren van alle punten die op een gelijke afstand van elkaar liggen. Je neemt een interval met als lengte de roosterconstante en plakt de eindpunten aan elkaar, met als resultaat de cirkel S^1. In twee dimensies neem je een rechthoek waarvan de zijdes gegeven worden door de twee roosterconstanten en je plakt de tegenoverliggende zijdes aan elkaar zoals in de figuur op de voorgaande pagina. In dit geval is het resultaat een donut of torus. «*Knutsel dit zelf maar eens in elkaar.*» De veralgemening naar n dimensies noemen we de n-torus $\mathbb{T}^n$, een product van n cirkels.

Het fundamentele ruimtelijke object in de bundels van quasiperiodieke functies is trouwens niet de enige n-torus $\mathbb{T}^n$ in het verhaal. De kristalimpuls is ook een element van een torus, want wiskundig blijkt dat $\hat{\Gamma} \cong \mathbb{T}^n$. De Pontryagin-duale van $\mathbb{Z}^n$ is de torus $\mathbb{T}^n$. Deze duale torus wordt verder ook vaak de Brillouintorus genoemd.

De stelling van Bloch–Floquet kan dus geherformuleerd worden door te zeggen dat elke oplossing gegeven wordt door de keuze van een element in de Brillouintorus $\hat{\Gamma}$ en een bijhorende functie op de ruimtelijke torus $\mathbb{R}^n/\Gamma$. Aangezien de keuze van kristalimpuls de bepalende factor is, kunnen we onze doorsnedes van de bundels $\mathcal{L}_\lambda$ samennemen en het geheel interpreteren als een doorsnede van een bundel $\mathcal{E} \to \hat{\Gamma}$. Kwantummechanica van kristalsystemen kan volledig worden vertaald naar de studie van vectorbundels over de Brillouintorus [4].

Aangezien het hele probleem opgesplitst worden als een reeks Schrödingervergelijkingen voor elke waarde van de kristalimpuls $\vec{k}$, kunnen we nog meer zeggen over de structuur van de Blochbundel $\mathcal{E}$. We weten dat doorsnedes van de vezelbundel $\mathcal{E} \to \hat{\Gamma}$ continu zijn (per definitie), wat impliceert dat alle resultaten op een continue manier van de impuls zullen afhangen. (Er zit hier heel wat wiskunde achter moest je dit nog niet geraden hebben.) Als we een grafiek zouden maken van de energieniveaus in functie van de kristalimpuls in de Brillouintorus, dan zouden we merken dat deze banden vormen op het diagram. Het concept band zal nu bij de

materiaalfysici onder ons een fel lampje doen branden. Banddiagrammen zijn uiterst belangrijke hulpmiddelen in de studie van materialen zoals halfgeleiders. Ze stellen wetenschappers in staat om op een eenvoudige manier het gedrag van materialen te voorspellen.

Alhoewel de totale Blochbundel zelf topologisch triviaal is (ook al is dit absoluut geen triviale uitspraak), toch gebeurt er in sommige gevallen iets bijzonders. Er ontstaat een opening in het banddiagram waarbij een reeks energiewaarden niet kunnen voorkomen. Als we de energie zodanig verschuiven (herinner je dat de energie op een constante na bepaald is) zodat de maximale energie van de bezette toestanden in deze opening valt, dan valt de bundel $\mathcal{E}$ uiteen in twee componenten:

$$\mathcal{E} \cong \mathcal{E}^- \oplus \mathcal{E}^+ .$$

Deze twee componenten worden de valentie- en conductieband genoemd en zijn zelf niet noodzakelijk triviaal. De valentieband bevat toestanden waar de elektronen stevig rond een atoom vastzitten en is altijd eindig-dimensionaal, terwijl de conductieband toestanden beschrijft die vrij doorheen het rooster kunnen bewegen en oneindig-dimensionaal kan zijn (dit hangt af of we alle mogelijke toestanden beschouwen of ons beperken tot effectieve realisaties zoals *orbitalen*).

Indien de opening tussen de banden op het banddiagram overal voldoende groot is, spreken we van een isolator. Toestanden, in de praktijk meestal elektronen, in de valentieband kunnen niet naar de conductieband springen en er is geen geleiding mogelijk. Indien de valentiebundel topologisch niet triviaal is en dus geen productstructuur bezit, kunnen er wel bijzondere toestanden optreden die we topologische isolatoren noemen. Alhoewel deze geen geleidbaarheid vertonen binnen het materiaal, laten dergelijke exotische materialen wel stromen toe op hun oppervlak. Die *randmodes* zijn een typisch verschijnsel voor zogenaamde *topologische fasen*. (We komen hier nog op terug in Hoofdstuk 8.)

Soms is de 'valentieband' echter niet volledig gevuld en kunnen de elektronen vrij naar een naburig energieniveau in dezelfde band springen. Dit is de situatie van een metaal. De toestanden kunnen zonder proble-

men doorheen het materiaal bewegen. De situatie waar de twee banden dicht bij elkaar in de buurt komen en er maar een klein beetje energie nodig is om van de ene band naar de andere over te gaan, beschrijft halfgeleiders zoals in zonnepanelen of transistoren.

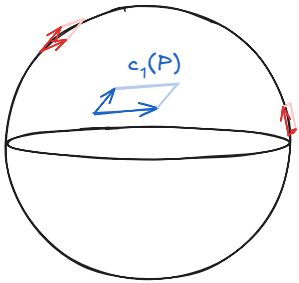
Extra's: Ladingskwantisatie

In het voorgaande hoofdstuk zagen we hoe we differentiaalvormen kunnen integreren over (geschikte) variëteiten. Samen met de bundels en karakteristieke klassen uit dit hoofdstuk, kunnen we die technieken gebruiken om ladingskwantisatie te abstraheren en in een differentiaalmeetkundig jasje te gieten.

Voor elke hoofdvezelbundel $P \to M$ kregen we een reeks Chernklassen $c_k(P)$ gegeven door $2k$-vormen op de basisvariëteit M. Voor elke deelvariëteit N in M van dimensie $2k$ kunnen we dus de k-de Chernklasse integreren:

$$\int_N c_k(P)\,.$$

Zodra N gesloten is, dat wil zeggen dat de rand ∂N ledig is, is het verrassende nu dat deze integraal een geheel getal is (op wat constanten na). Voor $k = 1$ moeten we dus bijvoorbeeld naar een sfeer kijken zoals in de figuur hiernaast. Op elk punt van de sfeer geeft de eerste Chernklasse een oppervlakte-element en wanneer we al die oppervlakte-elementen samennemen om een oppervlakteintegraal uit te rekenen, bekomen we als resultaat een geheel getal.



Figuur 3.7: De eerste Chernklasse op een sfeer.

Dat we altijd een geheel getal krijgen

is trouwens niet enkel een wiskundig merkwaardigheidje, het is ook gerelateerd aan een uiterst belangrijke waarneming in de natuurkunde. Als we terugkijken naar de formules in dit hoofdstuk, dan zien we dat $c_1(P)$ eigenlijk niets meer is dan de veldsterkte F_A. Dit resultaat zal aanleiding geven tot de kwantisatie van de magnetische lading!

In Deel 1 zagen we dat de Maxwellvergelijkingen in het vacuüm er als volgt uitzien (we hebben enkel de twee *divergentie*-vergelijkingen nodig):

$$\vec{\nabla} \cdot \vec{E} = 0 \qquad \text{en} \qquad \vec{\nabla} \cdot \vec{B} = 0\,.$$

In het bijzijn van materie moeten we deze vergelijkingen echter wat aanpassen:

$$\vec{\nabla} \cdot \vec{E} = \rho_E \qquad \text{en} \qquad \vec{\nabla} \cdot \vec{B} = \rho_B\,.$$

waar ρ_E en ρ_B de elektrische en magnetische ladingsdichtheden voorstellen. (Op het moment van schrijven is er nog steeds geen enkele aanwijzing voor het bestaan van magnetische lading, maar we schrijven deze formules in volledige algemeenheid.) Een ander resultaat uit datzelfde hoofdstuk in Deel 1 zegt verder nog hoe de veldsterkte is opgebouwd uit het elektrische veld en het magnetische veld.

Beschouw een integraal van de eerste Chernklasse of veldsterkte over een sfeer in de 3D-ruimte $\mathbb{R}^3$. Aangezien deze deelvariëteit geen componenten in de 'tijdsrichting' heeft, spelen enkel de coëfficiënten $F_{ij} = -B_k$ een rol ($i \neq j \neq k$). We krijgen dus

$$\int_{S^2} c_1(P) = \int_{S^2} \vec{B} \cdot d\vec{S} = \int_B \vec{\nabla} \cdot \vec{B}\, dV = \int_B \rho_B dV = Q_B\,,$$

waar we voor de tweede gelijkheid de stelling van Stokes toegepast hebben om van een oppervlakteintegraal over de sfeer S^2 over te gaan op een volumeintegraal over de ingesloten bol B. Alles samen zien we dat de integraal van de eerste Chernklasse gelijk is aan de magnetische lading!

Uit klassiek elektromagnetisme en differentiaalmeetkunde volgt dus dat de magnetische lading, indien hij zou bestaan, gekwantiseerd is. Voor

de kwantisatie van elektrische lading hebben we echter kwantummechanica nodig (en eventueel het bestaan van magnetische ladingen), met dank aan Dirac. Neem aan dat we een magnetische monopool hebben. Analoog aan het elektrische veld van een elektrische lading zou deze monopool een magnetisch veld van de volgende vorm genereren:

$$\vec{B}(\vec{r}) = \frac{g\vec{r}}{4\pi r^3}\,.$$

Stel nu dat we een elektrische lading rondom deze monopool laten bewegen. Wanneer een kwantummechanisch deeltje doorheen de lege ruimte beweegt, accumuleert het een fase van de vorm

$$\exp\left(i\int_C \hat{p}\cdot d\vec{r}\right).$$

(In Hoofdstuk 4 zal je zien dat het integrandum een belangrijke rol speelt in de theoretische fysica.) Om over te gaan van een lege ruimte naar een ruimte gevuld met een elektromagnetisch veld, stelt het principe van minimale koppeling dat we maar één dingen horen te doen. We moeten de impulsoperator, die in de lege ruimte gegeven wordt door de afgeleiden $\hat{p}_i = -i\hbar\partial_i$, vervangen door de covariante afgeleide $-i\hbar\nabla_i = -i\hbar(\partial_i + qA_i)$. Dit betekent dat we in de fase die de golffunctie oppikt een extra term krijgen:

$$\exp\left(iq\int_C \vec{A}\cdot d\vec{r}\right).$$

Via de stelling van Stokes kunnen we de integraal over het pad C ook herschrijven als een oppervlakteintegraal over eender welk oppervlak begrensd door C. Er zijn echter twee topologisch niet-equivalent keuzes. Een oppervlak dat de monopool omvat en een die dat niet doet. Aangezien het fysische resultaat niet mag afhangen van onze keuze, krijgen we de volgende gelijkheid voor een geschikt getal $n \in \mathbb{Z}$:

$$q\int_{S_1} (\nabla\times\vec{A})\cdot d\vec{S} - q\int_{S_2} (\nabla\times\vec{A})\cdot d\vec{S} = 2\pi n\,.$$

Maar wacht eens even, de twee oppervlakken samen vormen een gesloten oppervlak rondom de monopool. De integraal in het linkerlid stelt dus de volledige magnetische flux rondom de monopool voor, wat gelijk is aan de lading. We krijgen met andere woorden de volgende voorwaarde:

$$qg = 2\pi n\,.$$

De Dirac-kwantisatievoorwaarde impliceert dat, zodra er een magnetische monopool bestaat, de elektrische lading ook gekwantiseerd moet zijn!

Deze voorwaarde kan echter nog uitgebreid worden. Wat als deeltjes zowel een elektrische als een magnetische lading hebben? Dergelijke deeltjes worden ook wel dyonen genoemd (de naamgeving is niet altijd even origineel) en de kwantisatievoorwaarde neemt in dit geval de volgende vorm aan:

$$q_1 g_2 - q_2 g_1 = 2\pi n\,.$$

Het interessante aan deze vergelijking is dat het bestaan van klassieke elektrische deeltjes, zonder magnetische lading, geen enkele voorwaarde oplegt aan de elektrische lading van dyonen. Het is dus perfect mogelijk dat, indien ze zouden bestaan, alle magnetische monopolen ook een elektrische lading hebben.

Het is te zeggen, de Schwinger–Zwanzigervoorwaarde zegt niks over de absolute grootte van de elektrische lading van monopolen. We kunnen echter wel een relatief verband afleiden. Laat e zoals gewoonlijk de elementaire elektrische lading voorstellen, welke noodzakelijk dient te bestaan als de elektrische lading gekwantiseerd is. Hiermee komt de minimale magnetische lading $g = \frac{2\pi}{e}$ overeen. Als we nu twee dyonen (q_1, g) en (q_2, g) beschouwen, dan krijgen we de volgende voorwaarde:

$$q_1 - q_2 = ne\,.$$

De elektrische lading van alle dyonen, of ze nu geheel, rationeel of irrationeel zijn, verschillen een geheel aantal keer de elementaire lading van elkaar.

Wat Witten, een naam die onmogelijk kan ontbreken in een boek over hedendaagse theoretische fysica, aantoonde, is dat we meer kunnen zeggen over deze deeltjes indien we verder kijken dan elektromagnetisme. Vroeger dacht men dat het universum drie fundamentele symmetrieën had: C, P en T. De eerste zegt dat de natuurwetten invariant zijn onder het verwisselen van deeltjes en antideeltjes, de twee zegt dat de wetten invariant zijn onder spiegelsymmetrie en de derde zegt dat de wetten invariant zijn als we de richting van de tijd omdraaien.

Dat P-symmetrie gebroken is in de praktijk was onder meer zichtbaar bij radioactief verval. De zwakke wisselwerking behandelt *linkshandige* en *rechtshandige* deeltjes op een totaal verschillende manier. Daarna hoopte men dat de gecombineerde CP-symmetrie wel een symmetrie van het universum zou zijn. Naar het einde van de 20e eeuw bleek jammer genoeg dat ook deze symmetrie gebroken werd door de zwakke wisselwerking. (Dat dit niet het geval is voor de sterke interactie is trouwens een voorbeeld van een *finetuningprobleem*, want er is geen theoretische reden waarom het in die sector wel behouden zou zijn.) Enkel CPT-symmetrie bleef nog over en deze stelling houdt wiskundig stand in zeer sterke algemeenheid, men gelooft dus stellig dat dit de laatste stap is.

Als we echter aannemen dat CP-symmetrie geldig is (of slechts zwak gebroken wordt), dan bestaat er voor elk dyon ook een dyon met de tegengestelde elektrische lading. De Schwinger–Zwanzigervoorwaarde zegt dan dat

$$q = ne \qquad \text{of} \qquad q = \left(n + \frac{1}{2}\right)e\,.$$

Samen met het vorige resultaat voor dyonen betekent dit dat de elektrische lading van alle dyonen ofwel een geheel aantal keer de elementaire lading is, of een 'half-geheel' aantal keer. Het zogenaamde Witteneffect, uiteraard vernoemd naar onze meneer Witten, impliceert verder nog hoe deze relaties een kleine correctieterm verkrijgen wanneer CP-symmetrie gebroken wordt, het verandert echter niks aan het kwalitatieve resultaat.

Hoofdstuk 4

Calculus: Variatierekening

Variatierekening, wat zou dit nu weer kunnen zijn? Het eerste deel, variatie, is al een eerste indicatie voor hetgeen waarover we in dit hoofdstuk gaan praten. Variatie leunt heel dicht aan bij verandering en verandering komt neer op afgeleiden! Het soort afgeleiden die we hier zullen tegenkomen zijn echter weer van een andere aard dan degene die we tot nu toe zagen (ook al zaten ze stiekem wel verborgen op de achtergrond). Dit hoofdstuk is waarschijnlijk het meest technische in dit boek. Laat je hier echter niet door afschrikken, het is bedoeld om je een meer realistisch beeld te geven van wat theoretische fysica tegenwoordig (kan) inhouden.

De fysica zit vol met allerhande vergelijkingen, maar waar komen die nu eigenlijk vandaag? Initieel zijn de meeste van die vergelijkingen op een eerder experimentele manier of ad hoc wijze afgeleid. Maxwell en voorgangers lieten zich leiden door het gedrag van elektrische stromen en magneten, Einstein kwam tot een (filosofisch) inzicht inzake de fysische — of is het wiskundige? — natuur van ruimte en tijd, en Bohr en kompanen werden dan weer geïnspireerd door de vele eigenschappen van licht en atomen. Het zou echter mooi zijn moesten we in staat zijn een unificerend paradigma te vinden om de fundamentele vergelijkingen van de

natuurkunde op een gestroomlijnde manier af te leiden.

Uiteraard is het mogelijk om zo een paradigma te vinden, anders hadden de inleidende paragrafen natuurlijk niet zo veel zin. Het fysische principe achter wat we in de wiskunde de variatierekening zouden noemen, is het principe van de minste actie. Een goede analogie — en eigenlijk zelfs een heel goed voorbeeld — is het kronkelen van een rivier doorheen het landschap. Afgezien van door de mens aangelegde kanalen zijn waterwegen uitzonderlijk recht. Wanneer water stroomt zal het proberen om zo min mogelijk weerstand te ondervinden, het zal dus hardere stukken grond vermijden en, na enige tijd (op geologische schaal), zullen zo de meanders in een rivier ontstaan.

Dit principe van minste weerstand dat uiteraard ook in andere situaties van toepassing is, zij het in kwalitatieve of kwantitatieve vorm (denk bijvoorbeeld aan sociale interacties), wordt in de wiskunde beschreven door een zogenaamd optimalisatieprobleem:

$$x_{\text{opt}} = \arg\min_{x \in X} F(x)\,,$$

we zoeken de waarde van x die de functie $F : X \to \mathbb{R}$ minimaliseert. De functie F in dit verhaal wordt vaak de kostfunctie genoemd, door zijn geschiedenis in de economische wetenschappen. Voor ons zal die functie echter gegeven worden door S, de actie die we onder meer al in de kwantumveldentheorie waren tegengekomen, vandaar de naam van het principe.

Laten we dus nog eens kijken naar wat die actie precies was. In het algemeen hebben we enkel dynamische variabelen zoals de coördinaten q^i, golftoestanden $|\psi\rangle$ of velden ϕ, en een (reële) functie van die variabelen die we de Lagrangiaan L noemden. In de klassieke mechanica werd deze bijvoorbeeld gegeven door de uitdrukking

$$L(q, \dot{q}) = T(\dot{q}) - V(q)\,,$$

het verschil van de kinetische en de potentiële energie. De actie wordt dan in het algemeen gegeven door de *integraal* van deze Lagrangiaan over zijn

domein:

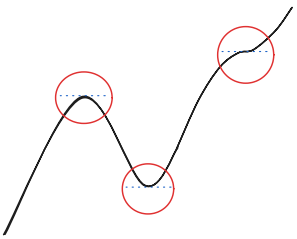
$$S[\phi] := \int_M L[\phi]\, \mathrm{dVol}_M\,,$$

waar de notatie met de rechte haakjes aangeeft dat zowel de actie S als de Lagrangian L niet enkel van ϕ, maar mogelijks ook van zijn afgeleiden kunnen afhangen.

De vorm van deze uitdrukking, die uiteindelijk de rol zal spelen van kostfunctie in ons optimalisatieprobleem, impliceert reeds twee belangrijke zaken:

- Het is een lokaal actieprincipe. De Lagrangiaan hangt enkel af van de variabelen en hun afgeleiden op één bepaald punt. Veraf gelegen punten kunnen elkaar niet instantaan beïnvloeden.
- De actie hangt niet enkel af van de variabelen zelf, maar ook van de meetkunde van de ruimte waarop ze gedefinieerd zijn. Dit is onder meer belangrijk voor ruimtes met een exotische structuur.

Maar hoe leidt zo een en het bijhorende optimalisatieprobleem nu tot de wetten van de fysica? Hiervoor beschouwen we even een functie van één veranderlijke zoals in de figuur hiernaast. Bij extrema — de verzamelnaam voor minima en maxima — liggen waarden aan beide kanten van het extremum ofwel hoger (voor minima) ofwel lager (voor maxima). Als we dus naar de raaklijn kijken in een omgeving van een extremum, dan zal de lijn overgaan van dalend naar stijgend (of omgekeerd) wanneer we het extremum passeren. In het extremum zelf zal de raaklijn dus horizontaal liggen of, met andere woorden, de afgeleide daar is nul.



Figuur 4.1: Extrema in één veranderlijke.

Dergelijke punten, waar de afgeleide van een functie nul is, worden ook wel kritische punten genoemd, daar ze algemener zijn dan extrema. Zo heb je bijvoorbeeld ook buigpunten, waar je een lokaal plateau hebt, zoals in het derde geval hierboven. Ook al is $F'(x) = 0$ dus geen voldoende voorwaarde om een minimum te vinden, toch wordt deze vergelijking in de praktijk als de voornaamste vergelijking voor (gladde) optimalisatieproblemen gezien. We krijgen dus de volgende vergelijking voor het vinden van fysieke oplossingen:

$$\frac{\mathrm{d}}{\mathrm{d}\phi}\int_M L[\phi]\,\mathrm{dVol}_M = 0\,.$$

Aangezien het domein M niet afhangt van onze dynamische variabele ϕ, mogen we de integraal en de afgeleide verwisselen van plaats (al zouden sommige wiskundigen ons hier halt toeroepen). We kunnen echter niet zomaar de afgeleide van L berekenen, aangezien de Lagrangiaan mogelijks ook van hogere afgeleiden van ϕ afhangt.

Dit leidt ons tot het concept van functionele afgeleide. In het algemeen is ϕ een functie op de ruimte M, we variëren dus niet zomaar een parameter of coördinaat om de afgeleide te berekenen, we variëren een hele functie:

$$\phi \longrightarrow \phi + \varepsilon\xi\,.$$

De functionele afgeleide van L wordt dan gedefinieerd als

$$\frac{\mathrm{d}L}{\mathrm{d}\phi}[\phi;\xi] := \frac{\mathrm{d}}{\mathrm{d}\varepsilon}L[\phi+\varepsilon\xi]\,,$$

met als gevolg dat deze stiekem altijd bepaald is in een bepaalde richting. (Het is een voorbeeld van de richtingsafhankelijke *Gateaux-afgeleide*.) De voorwaarde dat de (directionele) afgeleide van de integraal verdwijnt moet dus meer specifiek geïnterpreteerd worden als het verdwijnen van deze afgeleide voor elke 'richting' $\xi : M \to \mathbb{R}$.

Als gevolg krijgen we dus de volgende voorwaarde:

$$\int_{\mathbb{R}}\left(\frac{\partial L}{\partial\phi}\xi + \frac{\partial L}{\partial\dot{\phi}}\dot{\xi} + \cdots\right)\mathrm{d}t = 0\,,$$

waar we voor het gemak $M = \mathbb{R}$ nemen, we kijken enkel naar functies van de tijd zoals in de klassieke mechanica. Uiteraard is deze uitdrukking er niet eenvoudiger op geworden, integendeel. Maar wat als we het integrandum, de uitdrukking onder de integraal, nu eens herschrijven als een product $g(\phi, \dot{\phi}, \ldots)\xi$. De voorwaarde dat de functionele afgeleide moet verdwijnen voor alle ξ komt dan overeen met de voorwaarde dat $g(\phi, \dot{\phi}, \ldots)$ zelf nul moet zijn. We moeten dus enkel het integrandum nog in deze vorm krijgen.

Hiervoor gebruiken we de fundamentele stelling van de calculus om de integraal in het actieprincipe in de volgende vorm te gieten «*nogmaals een leuke oefening op de Leibnizregel*»:

$$\begin{aligned} 0 &= \int_{\mathbb{R}} \left(\frac{\partial L}{\partial \phi} \xi + \frac{\partial L}{\partial \dot{\phi}} \dot{\xi} + \cdots \right) \mathrm{d}t \\ &= \int_{\mathbb{R}} \left[\frac{\partial L}{\partial \phi} - \frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{\partial L}{\partial \dot{\phi}} \right) + \cdots \right] \xi \, \mathrm{d}t + \int_{\mathbb{R}} \frac{\mathrm{d}}{\mathrm{d}t} (\cdots) \, \mathrm{d}t \,. \end{aligned}$$

De tweede term kunnen we uitwerken aan de hand van de fundamentele stelling als het verschil van twee waarden 'op oneindig'. Indien we bijvoorbeeld enkel (mogelijke) oplossingen beschouwen die nul zijn op oneindig, dan valt deze term weg. Dit is een veelvoorkomende aanname die fysisch volledig steek houdt, aangezien er op oneindig meestal niks te beleven valt (al is dit niet volledig correct in bijvoorbeeld de algemene relativiteitstheorie). Aangezien de overblijvende integraal nul moet zijn voor alle mogelijke keuzes van ξ, volgt dus dat de uitdrukking tussen rechte haakjes ook gelijk moet zijn aan nul, en hopelijk komt die uitdrukking jou nog bekend voor. Het is exact de uitdrukking die ook voorkomt in de Euler–Lagrangevergelijking voor ϕ!

Die uitdrukking is zelfs zo belangrijk dat we er een eigen naam aan geven: de functionele of Euler–Lagrange-afgeleide (kwestie van het echt niet nog verwarrender te maken):

$$\frac{\delta L}{\delta \phi} := \frac{\partial L}{\partial \phi} - \frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{\partial L}{\partial \dot{\phi}} \right) + \cdots \,.$$

Voor hoger-dimensionale ruimtes dienen we trouwens enkel de tijdsafgeleiden te vervangen door partiële afgeleiden naar de coördinaten, maar afgezien daarvan is dit de finale uitdrukking. Willen we de bewegingsvergelijkingen voor een bepaald fysisch systeem vinden, dan dienen we enkel de Lagrangiaan op te schrijven en zijn functionele afgeleide gelijk te stellen aan nul, klaar is Kees. Dit is de reden waarom het principe van *extremale* actie, ook wel het variationeel principe genoemd, een van de basisprincipes van de theoretische fysica vormt. Dit was reeds het geval in de 18e eeuw en zal het geval blijven in de nabije toekomst. Voor de compleetheid kan je in Tabel 4.1 de Lagrangianen voor enkele belangrijke theorieën zien.

Theorie	Lagrangiaan
Newton	$T(\dot{q}) - V(q)$
Relativiteit	R
Yang–Mills	$F_{\mu\nu}F^{\mu\nu}$

Tabel 4.1: Bekende Lagrangianen.

Op dit punt in het verhaal lijkt het misschien alsof de titel van het hoofdstuk toepasselijk is, variatierekening is een deelgebied van de calculus. Toch is dit niet het einde van het verhaal. Variatierekening wordt een stuk eleganter en zelfs krachtiger als we het integreren in de wereld van de (differentiaal)meetkunde. Om dit te kunnen bewerkstelligen, moeten we echter ons beeld van de variëteit waar we op werken wat bijstellen. Of we nu werken op de ruimtetijd voor de mechanica van deeltjes of op de covariante faseruimte voor veldentheorie, dit is niet voldoende om de functies en objecten die we hierboven besproken hebben accuraat te beschrijven.

De Lagrangiaan uit het variationeel principe hangt niet enkel af van ϕ, maar ook van zijn afgeleiden. L is dus geen goed-gedefinieerde functie op M. Om de juiste leefwereld voor deze functies te vinden, hebben we de inhoud van Hoofdstuk 3 nodig. In de uitdrukking van de functionele afgeleide leiden we expliciet af naar de afgeleide van ϕ, we beschouwen deze afgeleide $\dot{\phi}$ als een soort onafhankelijke variabele.

Kies een vectorbundel $E \to M$ van rang $k \in \mathbb{N}$, dat wil zeggen dat de dimensie van een vezel gelijk is aan k, en noteer de dimensie van M als $n \in \mathbb{N}$. We definiëren de straalbundel $J^l(E)$ — in het Engels klinkt dit toch iets serieuzer: *jet bundle* — als de bundel die er lokaal als volgt uitziet:

$$\varphi : U \to \mathbb{R}^{n+k\binom{k+l}{l}} : p \mapsto \left(x^i, \phi^\alpha, \phi^\alpha_{,i_1}, \phi^\alpha_{,i_1 i_2}, \cdots, \phi^\alpha_{,i_1 \ldots i_l}\right),$$

waar we de verkorte notatie

$$\phi^\alpha_{,i_1 \ldots i_l} := \frac{\partial}{\partial x^{i_1}} \cdots \frac{\partial}{\partial x^{i_l}} (\phi^\alpha)$$

voor meervoudige afgeleiden hebben ingevoerd. We hebben als horizontale coördinaten dus nog steeds dezelfde als voor E, $J^l(E)$ is ook een (vector)bundel over de basisvariëteit M. Als coördinaten op de vezels hebben we echter niet enkel die van E meer. We krijgen er ook alle mogelijke afgeleiden tot en met die van orde $l \in \mathbb{N}$ bij. Functies op een straalbundel kunnen de afgeleiden van functies of doorsnedes als onafhankelijke, algebraïsche gebruiken.

De volgende stap voor ons is de introductie van het variationele bicomplex. Het voorvoegsel 'bi' wijst op de ontbinding in horizontale en verticale componenten van de meetkundige objecten die we zullen tegenkomen, in het bijzonder van differentiaalvormen. Beschouw onze nieuwe bundel $J^\infty(E)$, waar we dus alle ordes van afgeleiden toelaten (lokaal hebben we echter altijd maar een eindig aantal afgeleiden nodig, dus we moeten ons geen zorgen maken over een oneindig aantal coördinaten). We gaan niet enkel naar differentiaalvormen op M kijken, we gaan naar differentiaalvormen op deze bundel kijken, wat uiteraard mogelijk is aangezien elke vectorbundel over een gladde variëteit zelf ook een gladde variëteit is.

Op M hadden we voor elke coördinaat x^i een bijhorende 1-vorm $\mathrm{d}x^i$ en op E zouden we voor elke vezelcoördinaat ϕ^α een bijhorende 1-vorm $\delta\phi^\alpha$ krijgen (de gekke notatie wordt verderop duidelijk). Volledig analoog krijgen we op $J^\infty(E)$ voor elke coördinaat $\phi^\alpha_{,i_1 \ldots i_l}$ een 1-vorm $\delta\phi^\alpha_{,i_1 \ldots i_l}$. Als we dan de volledige vectorruimte van 1-vormen $\Omega^1(J^\infty(E))$ beschou-

wen, dan laat de opdeling in horizontale en verticale componenten het de Rhamcomplex uiteenvallen als een directe som:

$$\Omega^1(J^\infty(E)) \cong \Omega^{1,0}(J^\infty(E)) \oplus \Omega^{0,1}(J^\infty(E))\,.$$

Voor algemene k-vormen krijgen we

$$\Omega^k(J^\infty(E)) \cong \bigoplus_{i=0}^{k} \Omega^{i,k-i}(J^\infty(E))\,.$$

De volledige vectorruimte van differentiaalvormen op de bundel $J^\infty(E)$ is het variationele bicomplex. Naast de differentiaalvormen hadden we ook de de Rhamafgeleide $\mathsf{d} : \Omega^k(M) \to \Omega^{k+1}(M)$. Hier krijgen we er echter twee, een horizontale d en een verticale δ:

$$\begin{aligned}
\mathsf{d}f(x,\phi,\ldots) &:= \frac{\mathsf{d}f}{\mathsf{d}x^i}\mathsf{d}x^i = \left(\frac{\partial f}{\partial x^i} + \frac{\partial f}{\partial \phi^\alpha}\phi^\alpha_{,i} + \cdots\right)\mathsf{d}x^i\,,\\
\delta f(x,\phi,\ldots) &:= \frac{\partial f}{\partial \phi^\alpha}\delta\phi^\alpha + \frac{\partial f}{\partial \phi^\alpha_{,i}}\delta\phi^\alpha_{,i} + \cdots\,.
\end{aligned}$$

Merk op hoe de kettingregel hier essentieel gebruikt werd in de vorm van totale afgeleiden zoals in Deel 1. De totale de Rhamafgeleide op het variationele bicomplex wordt gegeven door de som van de horizontale en verticale afgeleiden:

$$\mathbf{d} := \mathsf{d} + \delta\,.$$

We kunnen de drie uitwendige afgeleiden ook relateren. Enerzijds krijgen we de eenvoudige relatie

$$\mathbf{d}x^i = \mathsf{d}x^i$$

voor de horizontale coördinaten. Voor de vezelcoördinaten krijgen we

$$\mathbf{d}\phi^\alpha_{,i_1\ldots i_l} = \phi^\alpha_{,ii_1\ldots i_l}\mathsf{d}x^i + \delta\phi^\alpha_{,i_1\ldots i_l}\,.$$

De totale afgeleide van een vezelcoördinaat hangt zowel af van de horizontale verandering als de expliciete verticale verandering.

Gewapend met deze objecten kunnen we aan de slag om de variationele calculus in zijn meetkundige vorm te gieten. Hiervoor hanteren we de volgende (onvolledige) woordenboek:

- Een fysische configuratie wordt gegeven door een doorsnede $\phi \in \Gamma(E)$ of, bij uitbreiding, door een doorsnede $j^\infty\phi \in \Gamma(J^\infty(E))$.
- De Lagrangiaan (of Lagrangiaanse dichtheid) is een n-vorm $L \in \Omega^{n,0}(J^\infty(E))$. Het is een horizontale differentiaalvorm van de hoogste graad, want $n = \dim(M)$ per definitie, dus we kunnen deze integreren over M.
- De actie is een lineaire functie $S : \Gamma(E) \to \mathbb{R}$ van de vorm

$$S(\phi) = \int_M L(j^\infty\phi)\ \mathrm{Vol}_M\ .$$

We hebben opnieuw kinematica, maar het ontbreekt ons nog aan dynamica. Laten we dus de Euler–Lagrangevergelijkingen ook vertalen.

Om de Euler–Lagrangevergelijkingen in dit hoofdstuk af te leiden, hebben we partiële integratie zoals in Hoofdstuk 2 toegepast. Dit was een combinatie van de Leibnizregel $f\mathrm{d}g = g\mathrm{d}f - \mathrm{d}(fg)$ en de stelling van Stokes. In gunstige gevallen valt de integraal van de 'exacte' term $\mathrm{d}(fg)$ weg en kunnen we dus de afgeleide van plaats wisselen. Meetkundig (en ook fysisch) is het echter nuttiger om deze exacte term te behouden. Laten we dus nog eens die afleiding erbij halen (hier uitgedrukt voor veldentheorie en dus niet enkel met tijdsafgeleiden):

$$\begin{aligned} 0 &= \int_M \left(\frac{\partial L}{\partial \phi^\alpha}\xi + \frac{\partial L}{\partial \phi^\alpha_{,i}}\xi_{,i} + \cdots \right) \mathrm{Vol}_M \\ &= \int_M \left[\frac{\partial L}{\partial \phi^\alpha} - \frac{\partial}{\partial x^i}\left(\frac{\partial L}{\partial \phi^\alpha_{,i}} \right) + \cdots \right] \xi\ \mathrm{Vol}_M + \int_M \mathrm{d}\big(\cdots\big)\,. \end{aligned}$$

Als je goed naar de eerste lijn kijkt en dit vergelijkt met de uitdrukking voor de verticale afgeleide, dan zie je dat die verdacht goed op elkaar lijken. Op de tweede lijn bevat de eerste term de Euler–Lagrange afgeleide,

die we nu zullen aanduiden met δ_{EL}, en de tweede term bevat de horizontale afgeleide.

Toch staan de integralen hier nog wat in de weg. De afleiding lijkt hier toch wel fundamenteel gebruik van te maken. Wat blijkt nu, we hebben die integralen helemaal niet nodig. Lokaal geldt de volgende gelijkheid op het niveau van differentiaalvormen:

$$\delta L = \delta_{\mathrm{EL}} L + \mathsf{d}\theta_L \,,$$

waar θ_L een geschikte differentiaalvorm in $\Omega^{n-1,1}(J^\infty(E))$ is die vaak de geleerde naam presymplectische potentiaal krijgt, aangezien $\mathsf{d}\theta_L$ een symplectische structuur definieert. Men kan zelfs aantonen dat δ_{EL} de unieke operator is zodat δL op deze manier geschreven kan worden! De Euler–Lagrangevergelijkingen spelen dus een heel speciale rol in de meetkunde!

Het voordeel van deze uitdrukking, afgezien van het feit dat we nu de volledige gereedschapskist van de differentiaalmeetkunde kunnen gebruiken, is dat het ons toelaat om heel wat nuttige eigenschappen eenvoudig af te lezen. Beide leden van de vergelijking zijn elementen van $\Omega^{n,1}(J^\infty(E))$ en kunnen dus een vectorveld ξ langsheen de verticale dimensies 'opeten'. Op deze manier kunnen we kijken hoe de Lagrangiaan en bijgevolg de bewegingsvergelijkingen zouden wijzigen wanneer we de velden variëren in de richting van het vectorveld ξ. Het rechterlid zegt dat deze variatie bestaat uit een term die evenredig is met de Euler–Lagrangevergelijkingen en een totale afgeleide.

Om de kracht van deze uitdrukking te verduidelijken, zullen we als voorbeeld eens naar fysische symmetrieën kijken. Als eerste zullen we eens aantonen dat de Euler–Lagrange-operator δ_{EL} nul is op exacte differentiaalvormen: $L = \mathsf{d}K$. In dit geval krijgen we

$$\delta \mathsf{d}K = \delta_{\mathrm{EL}} \mathsf{d}K + \mathsf{d}\theta_{\mathsf{d}K} \,.$$

In het linkerlid kunnen we de twee differentiaaloperatoren omwisselen: $\delta \mathsf{d}K = -\mathsf{d}\delta K$. De reden hiervoor is dat zowel d, δ als hun som $(\mathsf{d} + \delta)$ een

de Rhamafgeleide voorstellen (op geschikte ruimtes van differentiaalvormen). Aangezien dit betekent dat hun kwadraten exact gelijk zijn aan 0, volgt de anticommutatie-eigenschap van d en δ. «*Reken dit maar eens na (denk aan distributiviteit, maar let op dat operatoren niet altijd commuteren).*» Voor de Lagrangiaan $\mathrm{d}K$ volgt dus dat het linkerlid ook een exacte differentiaalvorm is:

$$\text{exact} = \delta_{\mathrm{EL}}\mathrm{d}K + \text{exact}\,,$$

maar we hadden reeds gezegd dat de Euler–Lagrange-operator de unieke operator was die zorgde voor een afsplitsing van exacte termen. Dit is enkel mogelijk indien $\delta_{\mathrm{EL}}\mathrm{d}K = 0$.

Hier maken we nu gretig gebruik van. Een symmetrie is een infinitesimale transformatie voorgesteld door een vectorveld ξ langsheen de vezeldimensies waarvoor geldt dat $\delta L(\xi) = \mathrm{d}\eta$ voor een zekere differentiaalvorm $\eta \in \Omega^{n-1,0}(J^\infty(E))$. Uit bovenstaande paragraaf volgt rechtstreeks dat

$$\delta_{\mathrm{EL}}\big(L + \delta L(\xi)\big) = \delta_{\mathrm{EL}}L\,.$$

Een kleine verandering in de richting van ξ laat de bewegingsvergelijkingen, en dus in het bijzonder de oplossingen, invariant.

Als je aan symmetrieën denkt, dan herinner je je misschien ook nog de stelling van Noether die zei dat elke symmetrie (of toch bijna elke) overeen komt met een behouden grootheid. We zouden hier dus in staat moeten zijn om een behouden grootheid af te leiden uit bovenstaande formules. Beschouw een (lokale) symmetrie ξ en definieer de bijhorende Noetherlading (of eerder Noetherstroom) in $\Omega^{n-1,0}(J^\infty(E))$ als volgt:

$$J_\xi := \eta + \theta(\xi)\,.$$

«*Neem hiervan de (horizontale) uitwendige afgeleide en gebruik de relaties uit voorgaande paragrafen.*» De 'ruimtelijke' variatie $\mathrm{d}J_\xi$ is evenredig met de bewegingsvergelijkingen en voor oplossingen van die vergelijkingen geldt dus dat $\mathrm{d}J_\xi = 0$. We hebben een behouden grootheid gevonden!

Als laatste zullen we deze formule nog even in een meer gebruikelijke vorm gieten. Die uit de klassieke mechanica. In dat geval zagen we dat Lagrangiaan gegeven werd door het verschil van de kinetische en potentiële energie. Als we nu aannemen dat er geen krachten zijn, zodat de potentiële energie nul is (of ten minste constant is), dan krijgen we:

$$L(q, \dot{q}) = \frac{1}{2} m \dot{q}^2 \,.$$

In dit geval kan je aantonen (zoals in de afleiding met integralen) dat de presymplectische potentiaal gegeven wordt door

$$\theta_L = p \mathsf{d} q \,.$$

Voor translaties bijvoorbeeld, is de generator gegeven door het constante vectorveld $\xi = \partial_q$. Samen geeft dit al de volgende uitdrukking: $\theta_L(\xi) = p\mathsf{d}q(\partial_q) = p$. (Je ziet waarschijnlijk al waar we naartoe gaan, of dat had je eigenlijk al moeten raden.) Voor de andere term η kunnen we makkelijk beredeneren dat die nul is. L hangt niet af van de coördinaat q, dus de variationele afgeleide δL ook niet. Deze laten inwerken op het vectorveld ∂_q is dus nul. We vinden dat de behouden grootheid die overeenkomt met infinitesimale translaties gegeven wordt door de impuls p, zoals verwacht!

Hoofdstuk 5

Algebra: Projectieve representaties

Groepen, vectorruimtes, matrices... Natuurkundigen hanteren een hele voorraad aan algebraïsche concepten om de natuur te beschrijven. In het algemeen kan men deze concepten opdelen in twee klassen: gestructureerde verzamelingen en operaties tussen verzamelingen. (Vandaar dat categorieën zo nuttig zijn om de boekhouding wat op orde te krijgen.) In dit hoofdstuk zullen we de algebraïsche voorraadkast nog wat verder uitbreiden en aanvullen.

In dit hoofdstuk hebben we als doel om wat algebraïsche structuren en subtiliteiten in de natuurkunde, in het bijzonder de kwantummechanica, te bestuderen. Zo zullen we zien dat symmetriegroepen op een soms nogal speciale manier kunnen inwerken op de wiskundige beschrijving van een fysisch systeem. Als eerste zetten we echter een hele grote stap terug en bouwen we een hiërarchie van algebraïsche structuren op. Hiervoor vertrekken we van een verzameling zonder enige structuur waar we stapsgewijs eigenschappen of operaties zullen aan toevoegen.

1. Stel dat je een verzameling A uitrust met een binaire operatie $* : A \times A \to A$ die associatief is en waarvoor er een eenheidselement 1_A bestaat. Dan krijg je een zogenaamde monoïde. De natuurlijke getallen $\mathbb{N}$ zijn het prototypische voorbeeld van een monoïde.

2. Als er voor elk element van de monoïde A een inverse bestaat, dat wil zeggen dat $x * x^{-1} = 1_A$, dan is A een groep. Een goed voorbeeld van een groep zijn de gehele getallen $\mathbb{Z}$ onder de optelling.

3. Soms zijn we echter niet tevreden met één operatie. Stel nu dat A twee operaties $*$ en $+$ heeft zodat $(A, +)$ een commutatieve groep is en $(A, *)$ een monoïde is. Dan wordt $(A, +, *)$ een ring genoemd als $*$ distributief is over $+$:

$$x * (y + y') = x * y + x * y' .$$

   Als we de gehele getallen nemen met zowel hun optelling als vermenigvuldiging dan krijgen we een voorbeeld van een ring.

4. Als we een ring $(A, +, *)$ hebben waarvoor $(A, *)$ zelf ook een commutatieve groep is, dan spreken we over een veld. Let wel op dat dit niks te maken heeft met de fysische velden, het is een ongelukkig toeval. Belangrijke voorbeelden zijn de rationale, reële en complexe getallen.

5. In sommige gevallen heeft de ring $(A, +, *)$ zelfs nog meer structuur. Vaak is $(A, +)$ eigenlijk ook een vectorruimte en hebben we dus drie operaties op A: optelling, vermenigvuldiging en scalaire vermenigvuldiging. In dit geval wordt A ook wel een algebra genoemd.

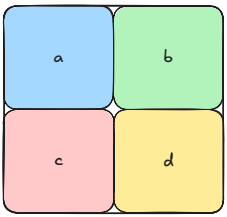
6. Als alle elementen in A ook nog eens een graad hebben zodat

$$\mathsf{deg}(x * y) = \mathsf{deg}(x) + \mathsf{deg}(y) \bmod 2 ,$$

   dan spreken we over superstructuren, zoals supergroepen of superalgebra's.

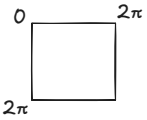
Als oefening op abstracte algebra tonen we hier nog een leuk resultaat: het Eckmann–Hiltonargument[1]. Stel dat we een verzameling A hebben met twee operaties $*$ en $+$ zodat zowel $(A, *)$ als $(A, +)$ een monoïde vormen. Indien de volgende uitwisselingswet geldt, dan zijn beide operaties gelijk aan elkaar en is de monoïde zelfs commutatief:

$$(a + b) * (c + d) = (a * c) + (b * d)\,.$$



Figuur 5.1: Een 2D voorstelling van het Eckmann–Hiltonargument.

Een verzameling laat dus niet noodzakelijk onbeperkt veel structuur toe. Deze voorwaarde wordt trouwens duidelijker wanneer we ze in 2D voorstellen zoals in de figuur hiernaast. We kunnen de vier kleine vierkanten op twee manier paarsgewijs samenkleven om het grote vierkant te vormen: eerst horizontaal of eerst verticaal. Dat beide manier meetkundig op hetzelfde neerkomen, dat is exact wat het Eckmann–Hilton argument zegt.



Figuur 5.2: Een opengeknipte sfeer.

Dit argument toont onder meer aan dat de hogere homotopiegroepen $\pi_{n\geq 2}(X)$ van een topologische ruimte X, die we in Hoofdstuk 1 ingevoerd hadden, altijd commutatief zijn. De cirkel S^1 kan je openknippen om het interval $[0, 2\pi] \cong [0, 1]$. Op die manier kan je een functie $S^2 \to X$ op de sfeer ook zien als een functie op het vierkant dat hiernaast is weergegeven. Waar je dus maar een zinvolle manier hebt om twee functies $f, g : S^1 \to X$ samen te stellen, zijn er twee manieren om functies $F, G : S^2 \to X$ samen te stellen: de horizontale en verticale samenstellingen uit Figuur 5.1.

Dit fenomeen is trouwens niet uitzonderlijk. De eendimensionale wereld is een heel rigide wereld, waar er niet veel mogelijkheden zijn. Zodra je uit die ene dimensie breekt, heb je toegang tot een waaier aan nieuwe

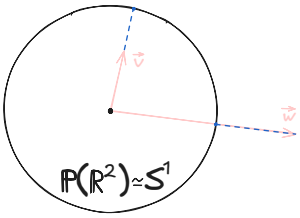
[1] DIt heeft absoluut niks te maken met de gelijknamige hotelketen.

opties. Een ander voorbeeld wordt gegeven door knopen of, zoals wiskundig ze zouden bekijken, eendimensionale deelvariëteiten die homeomorf zijn (als topologische ruimtes) aan de cirkel. In twee dimensies kunnen ze niet bestaan (op de cirkel na) omdat ze zichzelf noodgedwongen zouden moeten snijden, in drie dimensies zijn het de knopen zoals we ze kennen, en in vier dimensies of meer gebeurt iets bijzonders. Elke knoop in dimensie vier of hoger is volledig triviaal, je kan ze altijd ontwarren om de gewone cirkel te bekomen!

Laat ons echter terugkeren naar het oorspronkelijke verhaal. In Deel 1 hebben we reeds gezien hoe de verzameling van alle mogelijke toestanden in de kwantummechanica de structuur van een vectorruimte $\mathcal{H}$ heeft. De regel van Born zei dan hoe de kans om een fysisch systeem $|\psi\rangle$ waar te nemen in de (basis)toestand $|e_i\rangle$ gerelateerd was aan het inproduct van de twee vectoren:

$$p(e_i) = \langle\psi\,|\,e_i\rangle\langle e_i\,|\,\psi\rangle = |\langle e_i\,|\,\psi\rangle|^2 .$$

In het bijzonder geldt $|\langle\psi\,|\,\psi\rangle|^2 = 1$, de kans om een toestand waar te nemen, gegeven dat het systeem zich in die toestand bevindt, moet natuurlijk 1 zijn. Dit toont eigenlijk al hoe we in Deel 1 wat te kort door de bocht gegaan zijn.



Figuur 5.3: Projectieve ruimte van de complexe getallen.

De echte toestandsruimte in de kwantummechanica is niet de vectorruimte $\mathcal{H}$, maar de projectieve ruimte $\mathbb{P}(\mathcal{H})$ waarin we twee vectoren met elkaar identificeren als ze gelijk zijn op een scalaire vermenigvuldiging na (een zoveelste voorbeeld van een quotiëntconstructie). Hiernaast geven we het voorbeeld van het vlak $\mathbb{R}^2$. Alle vectoren op dezelfde lijn worden met elkaar geïdentificeerd en we krijgen dus dat de projectieve ruimte van het vlak equivalent is aan $[0, \pi]$ waarbij de eindpunten ook met elkaar geïdentificeerd dienen

te worden. Zo bekomen we dus de cirkel S^1.

Dat we met de projectieve ruimte $\mathbb{P}(\mathcal{H})$ moeten werken en niet met $\mathcal{H}$ zelf, heeft trouwens heel wat gevolgen. Enerzijds is $\mathbb{P}(\mathcal{H})$ geen vectorruimte meer. We moeten dus altijd eerst de vectoren optellen in $\mathcal{H}$ en daarna hun projectie op $\mathbb{P}(\mathcal{H})$ berekenen. Anderzijds gebeurt er iets heel interessants met de symmetrieën en dat is waarvoor we gekomen zijn.

Alle symmetrieën van een verzameling X (of die nu extra structuur heeft of niet) vormen een groep $\mathrm{Aut}(X)$. «*Probeer dit eens te beredeneren*!» (De reden voor die notatie is trouwens dat isomorfismen tussen een verzameling en zichzelf ook wel automorfismen genoemd worden.) In het specifieke geval van een vectorruimte met een inproduct, bestaat er een heel belangrijke deelgroep van alle symmetrieën, namelijk die die het inproduct behouden. Als we werken over de complexe getallen, dan spreekt met over de deelgroep $\mathrm{U}(\mathcal{H})$ van de unitaire transformaties:

$$\langle U\psi \,|\, U\varphi \rangle = \langle \psi \,|\, \varphi \rangle$$

voor alle $U \in \mathrm{U}(\mathcal{H})$. Deze eigenschap is van fysisch belang omdat de totale kans in de realiteit uiteraard behouden moet zijn. De kans om iets te meten is altijd gelijk aan 1.

Het is echter niet de vectorruimte $\mathcal{H}$ met zijn inproduct die van fysisch belang zijn, maar de uiteindelijke fysische toestanden en de kans gegeven door de regel van Born, en die functie is invariant onder meer dan enkel de unitaire transformaties. (Je kan trouwens aantonen dat de kansfunctie p een geschikte notie van afstand op de projectieve ruimte $\mathbb{P}(\mathcal{H})$ definieert.) Een unitaire functie is lineair, hij werkt dus als volgt in op lineaire combinaties:

$$U \sum_{i=1}^{n} \lambda_i |e_i\rangle = \sum_{i=1}^{n} \lambda_i U|e_i\rangle \,.$$

De coëfficiënten λ_i in de kwantummechanica zijn complexe getallen, dus we beschikken over nog een heel speciale operator: complexe toevoeging.

Deze operator $\mathcal{C}$ is echter niet lineair:

$$\mathcal{C} \sum_{i=1}^{n} \lambda_i |e_i\rangle = \sum_{i=1}^{n} \overline{\lambda_i} \mathcal{C}|e_i\rangle \,.$$

De coëfficiënten worden vervangen door hun complex toegevoegde. Dit zorgt er ook voor dat $\mathcal{C}$ geen symmetrie is van het inproduct, al is het bijna een symmetrie. Door gebruik te maken van de eigenschap $\mathcal{C}^2 = \mathbb{1}_{\mathcal{H}}$ kan je aantonen dat $\mathcal{C}$ ook het complex toegevoegde van het inproduct neemt:

$$\langle C\psi \,|\, C\varphi \rangle = \overline{\langle \psi \,|\, \varphi \rangle} = \langle \varphi \,|\, \psi \rangle \,.$$

Als we nu naar de kansfunctie p kijken, dan zien we dat die volledig invariant is onder zo een omwisseling van de twee vectoren. De complexe toevoeging $\mathcal{C}$ is dus een fysische symmetrie!

De stelling van Wigner zegt nu dat dit alles is wat er te vinden valt. Elke fysische symmetrie kan op de hoogte van de vectorruimte $\mathcal{H}$ geïmplementeerd worden als een unitaire transformatie U, eventueel gevolg door de complexe toevoeging $\mathcal{C}$. In zekere zin bestaat de volledige automorfismegroep dus uit twee kopieën van $\mathsf{U}(\mathcal{H})$.

Stel nu dat we een fysisch systeem krijgen waarvan we weten dat het symmetrisch is onder een groep G. Wiskundig betekent dit dat we voor elk element $g \in G$ een symmetrietransformatie $\rho(g)$ van het fysische systeem. Indien $\rho(g)$ een lineaire operator was, dan spraken we over een representatie $\rho : G \to \mathsf{Aut}(\mathcal{H})$ zoals in Hoofdstuk 3. Hier kan $\rho(g)$ echter ook antilineair zijn, dat wil zeggen dat het ook de complexe toevoeging neemt van de coëfficiënten. Om deze reden spreken we vaak van een projectieve representatie. (Dit duidt nogmaals op het feit dat de projectieve ruimte $\mathbb{P}(\mathcal{H})$ eigenlijk de juiste context vormt.)

Zoals in de vorige paragraaf aangegeven is een representatie een groepsmorfisme $\rho : G \to \mathsf{Aut}(\mathcal{H})$ tussen de gegeven groep G en de automorfismegroep $\mathsf{Aut}(\mathcal{H})$. Dit betekent dat we niet alleen een toewijzing hebben van een lineaire (of unitaire) operator aan elk groepselement $g \in G$, maar ook dat deze toewijzing de groepsvermenigvuldiging respecteert:

$$\rho(gh) = \rho(g) \circ \rho(g) \,.$$

Voor projectieve representaties moeten we deze vergelijking echter aanpassen. Daar kwantummechanische toestanden slechts op een factor na bepaald zijn, door hun projectieve aard, is het toegelaten dat bovenstaande vergelijking slechts geldt op een factor na:

$$\rho(g) \circ \rho(g) = \omega(g,h)\rho(gh)\,.$$

De factor $\omega(g,h)$ is echter niet volledig arbitrair. Laten we daar dus eens in detail naar kijken.

De projectieve aard van de kwantummechanica is al voldoende voor bovenstaande wijziging, zelfs zonder de antilineaire operatoren van Wigner, dus laten we die nog even achterwege laten. Associativiteit van G zegt dat we het drievoudig product ghk op twee manier kunnen berekenen: $(gh)k$ of $g(hk)$. Als we dit in de bovenstaande formules invullen, krijgen we een voorwaarde op de functie $\omega : G \times G \to \mathbb{C}$:

$$\big(\rho(g) \circ \rho(h)\big) \circ \rho(k) = \omega(g,h)\rho(gh) \circ \rho(k) = \omega(g,h)\omega(gh,k)\rho(ghk)$$

en

$$\rho(g) \circ \big(\rho(h) \circ \rho(k)\big) = \omega(h,k)\rho(g) \circ \rho(hk) = \omega(g,hk)\omega(h,k)\rho(ghk)\,.$$

Beide vergelijking geven een uitdrukking voor dezelfde operator, dus we krijgen de volgende gelijkheid voor de coëfficiënten:

$$\omega(g,h)\omega(gh,k) = \omega(g,hk)\omega(h,k)\,.$$

Een functie $\omega : G \times G \to \mathbb{C}$ die hieraan voldoet wordt een groep-2-cocyclus genoemd.

Voor we verdergaan met hoe de stelling van Wigner deze voorwaarde verder aanpast, vraag je je misschien af of er een verband is tussen deze groepscycli en de (co)cycli uit de voorgaande hoofdstukken in dit boek. Er zijn drie mogelijke antwoorden die we hieronder zullen bespreken: een oppervlakkige ja, een naïeve nee en een vrij technische zeker en vast. Het technische antwoord zal alweer een mooi voorbeeld geven van hoe uiterst verschillende deelgebieden toch intiem gerelateerd kunnen worden.

De oppervlakkige gelijkenis is dat groepcohomologie, net zoals de cohomologie van topologische ruimtes of gladde variëteiten, bepaalde algebraïsche invarianten ter beschikking stelt die ons in staat stellen om de algemene structuur van een groep te efficiënt te beschrijven. Voor lage graden zijn de voorbeelden uiterst eenvoudig. In graad 0 gaat het letterlijk over invarianten (hier komen we later op terug). In graad 1 worden de groepcocycli gegeven door de groepsmorfismen:

$$\varphi(gh) = \varphi(g)\varphi(h)\,.$$

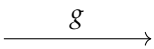
In graad 2 krijgen we exact de coëfficiënten die projectieve representaties definiëren. Voor hogere graden is de beschrijving minder voor de hand liggend en zijn de toepassingen eerder te vinden in de zuivere wiskunde, dus we stoppen in graad 2. Als we echter de formules voor bijvoorbeeld groepcohomologie en de Rhamcohomologie vergelijken, dan is de overeenkomst toch vrij oppervlakkig.

De (uiterst grote) waaier aan (co)homologietheorieën die in de wiskunde te vinden vallen, zijn uiteraard sterk verschillend, wat niet zo verwonderlijk is aangezien bijna elk deelgebied zijn eigen verzameling van dergelijke constructies heeft. Ze bestaan in de algebra, de topologie, de meetkunde, de grafentheorie, de categorietheorie, ... Al bestaan er soms wel connecties tussen die onderwerpen, toch lijden ze in het algemeen een onafhankelijk leven. We hoeven niet achter elk toeval meer te zoeken. Of toch?

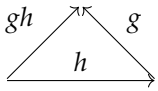
Wiskunde zou geen wiskunde zijn als toeval slechts toeval was. Uiteraard is er hier meer aan de hand. Voor elk groep G kunnen we een topologische ruimte BG, zijn classificerende ruimte, construeren die ons meer inzicht gaat geven. De constructie bouwt een simplicieel complex zoals in Hoofdstuk 1 op de volgende manier:

1. De elementen van BG worden niet gegeven door die van G. We laten ze onbepaald.

2. Voor elk element $g \in G$ hebben we een 1-simplex



3. Voor elk relatie $g \cdot h = gh$ in G hebben we een 2-simplex



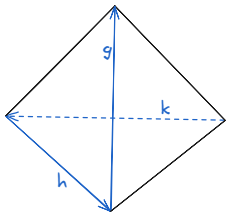
4. Meervoudige vermenigvuldigingen geven dan op een volledig analoge manier de hogere simplices in BG. Al die simplices dienen we dan aan elkaar te lijmen langsheen randen, vlakken, etc. met hetzelfde label.

Om verder te kunnen, hebben we de cohomologie van BG nodig. Voor gladde variëteiten hadden we de Rhamcohomologie, maar voor topologische ruimtes kunnen we in het algemeen geen differentiaalvormen gebruiken. Voor de correcte definitie hanteren we nogmaals het concept dualiteit. De cohomologiegroep $H^k(BG; A)$ is de duale van de (simpliciële) homologiegroep $H_k(BG)$. Meer precies krijgen we enerzijds dat de elementen van $H^k(BG; A)$ lineaire functies op $H_k(BG)$ zijn met als bereik de (commutatieve) groep A en anderzijds krijgen we een duale (co)randoperator:

$$\mathrm{d}\omega(c) := \omega(\partial c)$$

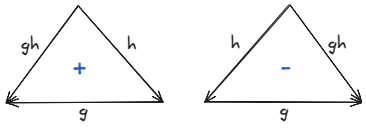
voor alle $c \in H_k(BG)$.

Met deze constructies bekomen we de volgende equivalentie tussen groepcohomologie en cohomologie van de classificerende ruimte: $H^k(G; A) \cong H^k(BG; A)$. We moeten enkel nog aantonen dat dit wel echt de juiste relaties geeft. Het drievoudig product ghk komt overeen met de 3-simplex in de figuur hiernaast. «*Reken maar*



Figuur 5.4: Een 3-simplex in BG.

eens na dat de vier vlakken overeenkomen met de vier producten uit de cocyclusvoorwaarden: $gh, (gh)k, g(hk)$ en hk.» De cocyclus *to be* ω wijst aan elk van die vlakken een coëfficiënt toe. Nu zullen we aantonen dat de cocyclusvoorwaarde exact overeenkomt met randvoorwaarde $\mathrm{d}\omega(\Delta^3) = 0$ waarbij Δ^3 de 3-simplex uit de figuur is.



Figuur 5.5: Oriëntaties van 2-simplices.

Door de definitie van d krijgen we dat $\mathrm{d}\omega(\Delta^3) = \omega(\partial\Delta^3)$ waarbij $\partial\Delta^3$ gelijk is aan de som van de vlakken met gepaste tekens die door de oriëntatie van de vlakken bepaald worden. Als we de conventie vanuit de figuur hiernaast hanteren, waarbij samenstelling in wijzerzin overeenkomt met een positieve oriëntatie, dan bekomen we de volgende voorwaarde:

$$\omega(h,k) + \omega(g,hk) - \omega(gh,k) - \omega(g,h) = 0\,,$$

exact de voorwaarde voor groepscycli wanneer we de sommaties vervangen door vermenigvuldigingen. (Voor commutatieve groepen is het gebruikelijk om de vermenigvuldiging als een optelling te noteren.)

Nu is het tijd om terug te keren naar de projectieve representaties uit de kwantummechanica. De stelling van Wigner zegt dat de operatoren op de toestandsruimte $\mathcal{H}$ zowel lineair als antilineair kunnen zijn en die tweede optie zorgt ervoor dat we de cocyclusvoorwaarde een beetje moeten wijzigen. We starten opnieuw met het drievoudig product ghk:

$$\omega(g,h)\rho(gh)\rho(k) = \rho(g)\big(\omega(h,k)\rho(hk)\big)\,.$$

In het linkerlid gebeurt er niks speciaals. In het rechterlid daarentegen verschijnt de factor ω rechts van de operator $\rho(g)$. Indien $\rho(g)$ antilineair is, krijgen we dus

$$\rho(g)\omega(h,k) = \overline{\omega(h,k)}\rho(g)\,.$$

De volledige voorwaarde wordt dan:

$$g \rhd \omega(h,k) + \omega(g,hk) = \omega(gh,k) = \omega(g,h) = 0\,,$$

waar de actie $\rhd$: $G \to \mathrm{Aut}(\mathbb{C})$ bepaalt of we de complex toegevoegde moeten nemen of niet.

Dat de algemene voorwaarde een actie van de groep G op de coëfficiëntengroep A bevat, was onder de wiskundigen al lang gekend. Voor hen is dit de normale situatie. Spijtig genoeg zit die actie wat verstopt in de notatie $H^\bullet(G;A)$ voor groepcohomologie. Er wordt impliciet verondersteld dat A gegeven is als een commutatieve groep met een actie van G. Wigner heeft trouwens sterk bijgedragen aan de studie naar de relatie tussen groepcohomologie en de algemene structuur van groepentheorie.

Een deel van de taak van de (theoretische) kwantumfysicus bestaat er dus uit om, gegeven een fysisch systeem, de juiste oplossing te vinden van de cocyclusvoorwaarde. Voor eindige groepen is dit een vrij eenvoudig probleem omdat het te herleiden valt tot een stelsel van vergelijkingen. Voor oneindige groepen ligt de situatie echter anders. Computers kunnen niet goed overweg met oneindige aantallen, dus we moeten overgaan op een andere formulering. Een optie is om de transformaties te vervangen door hun generatoren, toch zeker in het geval waarbij de groep ook de structuur van een gladde variëteit heeft (zogenaamde Liegroepen), wat het probleem opnieuw tot een eindig-dimensionaal probleem zou reduceren: *Lie-algebracohomologie*. Er is echter een goede reden om ook naar een ander soort cohomologie te kijken. In de kwantummechanica en kwantumveldentheorie wordt vaak gebruik gemaakt van de padintegraal. Als we even aannemen dat deze kan geformuleerd worden als een echte integraal zoals in Hoofdstuk 2, dan moeten de functies die erin voorkomen integreerbaar en dus in het bijzonder *meetbaar* zijn. In de definitie van de cohomologie van de classificerende ruimte BG zouden we ons dus moeten beperken tot die functies die meetbaar zijn, wat aanleiding geeft tot (niet noodzakelijk eindig-dimensionale) *Borelcohomologie*. Bovendien is er geen algemene relatie tussen deze verschillende cohomologietheorieën. De zoektocht gaat voorlopig verder.

Klasse	C^2	P^2	T^2
A	/	/	/
AIII	/	/	1
AI	1	/	/
BDI	1	1	1
D	/	1	/
DIII	-1	1	1
AII	-1	1	1
CII	-1	-1	1
C	/	-1	/
CI	1	-1	1

Tabel 5.1: Altland–Zirnbauerclassificatie van topologische isolatoren.

Vooraleer we overgaan van groepen naar categorieën in de extra's, beschouwen we nog een toepassing van groepcohomologie en projectieve representaties. In Hoofdstuk 3 haalden we zowel het concept topologische isolator als CPT-symmetrie aan. Wanneer we deze combineren bekomen we de zogenaamde periodieke tabel van topologische isolatoren[2]. Hiervoor beschouwen we de drie operatoren C, P en T, die elk een groep $\mathbb{Z}_2$ genereren, afzonderlijk. Door de specifieke structuur van de operatoren, ze zijn hun eigen inversen en zowel C als T zijn antilineair, is er maar een beperkt aantal mogelijkheden. P heeft maar één projectieve representatie en dat is de triviale representatie $\mathbb{1}$. Cr en T^2 daarentegen kunnen ook als $-\mathbb{1}$ geïmplementeerd worden. We hebben zo al vijf mogelijkheden. Het is echter ook mogelijk, zoals bij in het standaardmodel, dat de operatoren afzonderlijk geen symmetrie vormen, maar hun product wel. Op die manier komen er nog eens vijf klassen bij. Deze Altland–Zirnbauerclassificatie wordt hiernaast weergegeven in een tabel.

Extra's: Categorische algebra

In Deel 1 hadden we als voorbeeld van categorieën voor elke groep G de groepoïde $\mathbf{B}G$ ingevoerd. Dit was een categorie met slechts één object $*$ en

[2] In het Engels staat deze ook gekend als de *tenfold way* (wat een verwijzing is naar eerdere classificaties zoals de *threefold way* en *eightfold way*).

voor elk element van G een isomorfisme $g : * \to *$. In dit hoofdstuk kregen we een topologische ruimte, opgebouwd als een simplicieel complex, waarvoor er eigenlijk ook maar één punt bestond met voor elk element van G een 1-simplex tussen dit punt en zichzelf. Zowel de notatie als de gelijkaardige structuur zouden dus kunnen doen vermoeden dat er ook hier meer aan de hand is. Beide constructies stellen een en dezelfde entiteit voor. De ene is een topologische of meetkundige voorstelling, de andere een algebraïsche. (Al kan kan men stellen dat de algebraïsche iets fundamenteler is.)

Dit is een voorbeeld van hoe algebra (en meetkunde) natuurlijk aanleiding geven tot categorische constructies, ook al is het in dit geval een gedegenereerd geval. Het lijkt een beetje alsof we de groep G dwingen om een categorie te worden. We kunnen echter proberen om de categorische 'ontlussing' $\mathbf{B}G$ te veralgemenen om zo een echte 'categorische groep' te bekomen met meerdere objecten die zich gedragen als de elementen van een gewone groep.

Als eerste hebben we dus een notie van vermenigvuldiging nodig:

$$X \otimes Y$$

voor alle $X, Y \in \mathsf{ob}(\mathbf{G})$ zodanig dat er voor elk object X een invers object X^{-1}. Dat we hier dezelfde notatie als voor het tensorproduct van vectorruimtes gebruiken, is trouwens niet geheel toevallig. Dat tensorproduct is, net zoals het (Cartesiaans) product van verzamelingen, een voorbeeld van wat wiskundigen een monoïdaal product. Net zoals een groepoïde een categorische veralgemening van een groep is, is een monoïdale categorie een categorische veralgemening van een monoïde. Voor elke twee objecten X, Y bestaat er een product $X \otimes Y$ en voor elke twee morfismen $f : X \to Y, g : V \to W$ bestaat er een product $f \otimes g : X \otimes V \to Y \otimes W$. Uiteraard betekent deze veralgemening ook dat er een eenheidsobject $\mathbf{1_G}$ bestaat zodat

$$X \otimes \mathbf{1_G} \cong X$$

voor alle objecten $X \in \mathsf{ob}(\mathbf{G})$.

Bovenop deze productstructuur dienen we nog twee extra eigenschappen te eisen. Enerzijds het bestaan van inversen en anderzijds de associativiteit. We beginnen met de eerste. Voor elk object $X \in \mathrm{ob}(\mathbf{G})$ bestaat er een invers object X^{-1} zodat

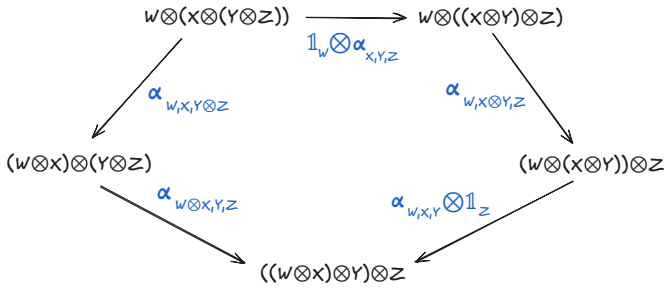
$$X \otimes X^{-1} \cong \mathbf{1}_{\mathbf{G}} \cong X^{-1} \otimes X\,.$$

Merk op dat voorgaande voorwaarden, in tegenstelling tot het geval van groepen, geen gelijkheid meer bevatten. Zoals bij alle constructies in de categorietheorie is het beter om vergelijkingen op een isomorfisme na te vragen (dit staat ook gekend als het principe van equivalentie [1]).

Waar het bestaan van een invers element extra structuur van een monoïdale categorie vergt, is associativiteit (of de veralgemening ervan) een inherente eigenschap, daar deze categorieën de klassieke monoïdes dienen te veralgemenen. Het principe van equivalentie zegt ons echter dat associativiteit niet meer strikt kan gelden, we vervangen het door een isomorfisme:

$$\alpha_{X,Y,Z} : X \otimes (Y \otimes Z) \cong (X \otimes Y) \otimes Z\,.$$

We noemen dit isomorfisme de associator van de monoïdale categorie. Net zoals bij gewone monoïdes de associativiteit een coherentievoorwaarde met drie elementen voor het binaire product geeft, bestaat er hier een coherentievoorwaarde met vier objecten voor de ternaire associator. Deze pentagonvergelijking staat hieronder afgebeeld.



Alles samen, krijgen we de volgende definitie voor een categorische

groep of 2-groep: een monoïdale groepoïde waarin elk object een inverse heeft. We introduceren deze definitie uiteraard niet zomaar. Als eerste geeft het aanleiding tot een cocyclus in groepcohomologie, een 3-cocyclus zelfs. Waar de associativiteit van groepen de 2-cocyclusvoorwaarde gaf, zal de pentagonvergelijking hier de 3-cocyclusvoorwaarde geven. Maar de groepcohomologie van welke groep? En met welke coëfficiënten?

Gegeven een 2-groep $\mathbf{G}$ krijgen we gratis en voor niets twee gewone groepen $\pi_0(\mathbf{G})$ en $\pi_1(\mathbf{G})$. De eerste, $\pi_0(\mathbf{G})$, wordt de Picardgroep van $\mathbf{G}$ genoemd en bestaat uit isomorfismeklassen van objecten uit $\mathbf{G}$. We identificeren dus alle objecten die isomorf zijn. Het is een soort 'decategorificatie' van $\mathbf{G}$. Hierdoor worden de isomorfismen in de definitie van de zwakke inversen vervangen door echte gelijkheden en bekomen we een 'normale' groep. Als tweede groep, $\pi_1(\mathbf{G})$, nemen we de automorfismegroep van het eenheidsobject $\mathbf{1}_{\mathbf{G}}$. Door middel van het Eckmann–Hiltonargument kan je zelfs aantonen dat deze groep commutatief is!

Nu kunnen overgaan naar het afleiden van de cocyclusvoorwaarde uit de pentagonvergelijking. Elk van de hoekpunten van de vijfhoek komt overeen met hetzelfde element in $\pi_0(\mathbf{G})$. Elk van de pijlen geeft een dus automorfisme van dit object $W \otimes X \otimes Y \otimes Z$. Wegens de eigenschappen van een 2-groep kan je trouwens aantonen dat alle automorfismegroepen van de objecten in $\mathbf{G}$ equivalent zijn en dus overeenkomen met de fundamentele groep $\pi_1(\mathbf{G})$. De pijlen in de pentagonvergelijking geven dus elementen in $\pi_1(\mathbf{G})$. We hebben de coëfficiënten voor onze cocyclus gevonden. Laat ons de twee richtingen in het diagram eens gelijkstellen aan elkaar:

$$W \triangleright \alpha(X,Y,Z) + \alpha(W, X \otimes Y, Z) + \alpha(W,X,Y) \\ = \alpha(W,X,Y \otimes Z) + \alpha(W \otimes X, Y, Z)\,.$$

Merk op dat de factor $\mathbb{1}_W$ in de pentagonvergelijking hier is omgezet in een actie van $\pi_0(\mathbf{G})$ op $\pi_1(\mathbf{G})$, net zoals de groep G op de coëfficiëntengroep A inwerkte in de 2-cocyclusvoorwaarde. Als je alles tot dusver volgde, dan zal je waarschijnlijk wel overtuigd zijn dat deze vergelijking ook wel echt de 3-cocyclusvoorwaarde is. «*Als je toch wat wantrouwig bent, gebruik*

de cohomologie van een 4-simplex in BG om dezelfde voorwaarde af te leiden. (Let wel op dat je in 4D zal moeten tekenen.)» Men kan trouwens aantonen dat (equivalentieklassen van) 2-groepen volledig bepaald worden door de keuze van $\pi_0(\mathbf{G})$ en $\pi_1(\mathbf{G})$, de actie van $\pi_0(\mathbf{G})$ op $\pi_1(\mathbf{G})$ en een cocyclus $\alpha \in H^3\big(\pi_0(\mathbf{G}), \pi_1(\mathbf{G})\big)$, die de Postnikovklasse van $\mathbf{G}$ genoemd wordt.

Na al die zuivere wiskunde maken we nu terug wat tijd vrij voor de natuurkunde, waarbij we deels voortbouwen op de extra's in Hoofdstuk 3. Beschouw een fysisch systeem met een (globale) symmetriegroep G, zoals translaties in de klassieke mechanica of fasetransformaties in elektromagnetisme. De geassocieerde ladingen, de Noetherladingen uit Hoofdstuk 4, worden hier volledig gedragen door puntmassa's: impuls van een deeltje, lading van een deeltje, enz. In Hoofdstuk 3 zagen we echter hoe we de totale lading in een deel van de ruimte kunnen bepalen door een integraal over de rand van dat gebied te berekenen. Verder werken de symmetrieën ook in op elk punt van de ruimte, en dat op eenzelfde manier aangezien we over globale symmetrieën spreken. Samengevat krijgen we dus de volgende structuren:

- Symmetrie: (al dan niet commutatieve) groep G.
- Ladingsdragers: punten.
- Ladingsdetectie: (gesloten) oppervlakken.
- Lokale operatoren: punten.

Dat de ladingsdragers puntdeeltjes zijn, is eigenlijk irrelevant. Het is het waarnemen van lading dat van fysisch belang is. We mogen het tweede puntje dus schrappen. In de moderne natuurkunde spreken we enkel nog over de topologische operatoren geassocieerd aan symmetrieën.

Neem bijvoorbeeld elektromagnetisme waarbij de symmetriegenerator overeenkomt met de stroom $\mathbf{J} \in \Omega^3(M)$, een 3-vorm op de ruimtetijd M. Als we nu een gesloten driedimensionale deelvariëteit $\Sigma \subset M$ nemen,

zegt Hoofdstuk 2 dat we de operator

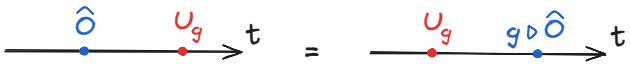
$$U(\Sigma) := \exp\left(i \int_\Sigma \mathbf{J}\right)$$

kunnen definiëren. Door de stelling van Stokes en het behoud van lading, volgt dat deze operator invariant is onder vervormingen van het (hyper)oppervlak Σ zolang de totale lading die bevat zit in Σ gelijk blijft. De defectoperator $U(\Sigma)$ is wel degelijk een topologische grootheid. (Hier zit nog wat cohomologie achter verstopt.) In het algemene geval, waarbij de groep G meer dan één generator heeft zoals bij de zwakke of sterke wisselwerking, hangt de operator ook af van de gekozen generator:

$$U_g(\Sigma) := \exp\left(i \int_\Sigma \mathbf{J}_g\right).$$

Ons lijstje van aanwezige structuren wordt dus:

- Symmetrie: (al dan niet commutatieve) groep G.
- Symmetrie-operatoren: Topologische operatoren $U_g(\Sigma)$ die leven op deelvariëteiten van dimensie $n-1$ (of codimensie 1).
- Lokale operatoren: operatoren $\widehat{O}$ die leven op deelvariëteiten van dimensie 0.



Figuur 5.6: Fysische actie van een symmetrie.

Wanneer de lokale operatoren $\widehat{O}$ zich in een representatie van de groep G bevinden, zullen de symmetrie-operatoren ook op hen inwerken. In de figuur hierboven wordt afgebeeld wat er gebeurt als we een lokale operator doorheen de tijd volgen wanneer hij het oppervlak Σ doorkruist. In termen van toestanden vertrekken we van de situatie

$$U_g \widehat{O} |\psi\rangle\,.$$

Aangezien de symmetrie inverteerbaar is, bestaat er dus een inverse operator U_g^{-1}. We kunnen voorgaande formule dus herschrijven als

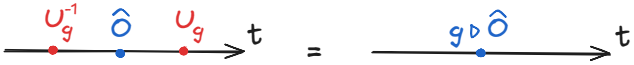
$$\left(U_g \widehat{O} U_g^{-1}\right) U_g |\psi\rangle \,.$$

De identiteitsoperator $\mathbb{1} = U^{-1}U$ tussen twee andere operatoren invullen is een typische tussenstap in veel theoretische berekeningen. Het grootste deel van de tijd besteden theoretici aan het herschrijven van eenvoudige zaken op een complexe manier.

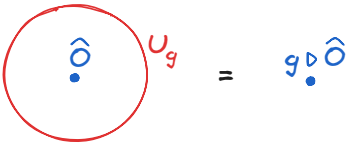
Enerzijds hebben we nu de fysische operator doorheen de symmetrie getrokken, anderzijds is dit wel ten koste van een actie van de symmetrie op die operator gegaan: $U_g \widehat{O} U_g^{-1}$. We zien bovendien ook dat deze uitdrukking de groepsvermenigvuldiging respecteert:

$$U_g U_h \widehat{O} U_h^{-1} U_g^{-1} = U_g U_h \widehat{O} \left(U_g U_h\right)^{-1} = U_{gh} \widehat{O} U_{gh}^{-1} \,.$$

We bekomen een echte representatie, ook wel de toegevoegde representatie genoemd. Deze representatie kan ook anders voorgesteld worden.



Figuur 5.7: Fysische actie van een symmetrie (herschikt).

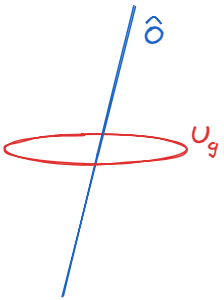


Figuur 5.8: Covariante voorstelling van een symmetrie.

In een algemeen covariante beschrijving horen we echter niet enkel rekening te houden met de tijd. Gelukkig helpt de topologische beschrijving hier ook. Nergens hebben we geëist dat het oppervlak Σ ruimtelijk moest zijn. Het mag ook een component doorheen de tijd bevatten. Meer algemeen kunnen we de situatie voorstellen zoals in de figuur hier-

naast. Deze situatie kunnen we nu ook zonder enig probleem veralgemenen naar symmetrieën die op deelvariëteiten in andere dimensies leven. Symmetrieën in codimensie $p+1 \leq n$ zullen interageren met lokale operatoren in dimensie p. In 3D komt $p = 1$ overeen met een symmetrie-operator die leeft op een deelvariëteit van dimensie 1. Zo zal een lijnoperator interageren met symmetrie-operatoren die op gesloten lijnen leven, cirkels dus. We spreken in deze gevallen ook over p-vormsymmetrieën (daar de lokale operatoren kunnen uitgedrukt worden aan de hand van p-vormen).

Hier treedt echter een heel belangrijk verschil op met de situatie voor $p = 0$. Kijk bijvoorbeeld naar Figuur 5.8. Indien er twee symmetrie-operatoren waren, de ene levend op de cirkel X_1 en de andere op X_2 waarbij X_1 volledig bevat zit in X_2, dan is er geen enkele mogelijkheid om de volgorde van de operatoren om te wisselen zonder dat X_1 en X_2 elkaar snijden. Als je een punt neemt binnenin X_1 en dit naar buiten beweegt, zal je altijd eerst X_1 snijden vooraleer je X_2 tegenkomt.



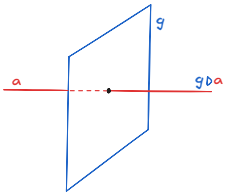
Figuur 5.9: *Linken* van een lijnoperator en een eendimensionale symmetrie-operator.

Voor $p \geq 1$ is dit echter niet meer het geval. Je kan de deelvariëteiten zodanig vervormen zodat de volgorde wordt omgewisseld. We krijgen dus noodgedwongen dat de symmetriegroep voor deze operatoren commutatief is:

$$U_{g_1}(X_1)U_{g_2}(X_2) = U_{g_2}(X_2)U_{g_1}(X_1)\,.$$

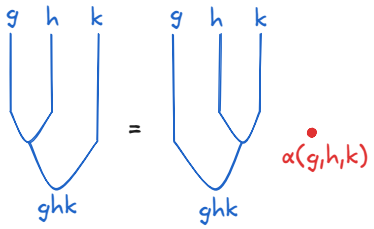
Dit resultaat is volledig analoog aan de resultaten eerder in dit hoofdstuk waar bleek dat de hogere homotopiegroepen van een topologische ruimte of de automorfismegroepen in een 2-group altijd commutatief zijn. Dat de symmetriegroepen in de 'klassieke' fysica (hier bedoelen we dus de fysica zoals de meeste mensen hem kennen) niet noodzakelijk commutatief zijn, is dus een heel unieke situatie!

Het verhaal is echter nog niet gedaan. Voor twee dimensies $p, q \leq n$, kunnen p-vormsymmetrieën en q-vormsymmetrieën elkaar mogelijks ook beïnvloeden. In het bijzonder kunnen symmetrieën in lagere dimensies die in hogere dimensies vervormen en kunnen symmetrie-operatoren in hogere dimensies, wanneer ze fuseren, symmetrieën in lagere dimensies produceren. Voor de eenvoud zullen we ons beperken tot 0- en 1-vormsymmetrieën in een driedimensionale ruimte.



Figuur 5.10: Interactie tussen symmetrie-operatoren.

Neem een 0-vormsymmetrie $g \in G_0$ (we vereenvoudigen even de notatie), voorgesteld door het blauwe vlak in de figuur hiernaast, en een 1-vormsymmetrie $a \in G_1$, voorgesteld door de rode lijn, die g transversaal snijdt. Dit wil dus zeggen dat er meetkundig maar één snijpunt is, nergens ligt de lijn in het vlak. Aan de ene kant van het vlak hebben mooi het symmetrie-element a, maar zodra de operator het vlak snijdt, moeten we een actie van de symmetrie-groep G_0 op de groep G_1 in rekening brengen: $g \triangleright a$. We zien hier reeds de eerste handtekening van een 2-groep opduiken. Het tweede deel duikt op wanneer we de fusie van drie 0-vormsymmetrieën $g, h, k \in G_0$ beschouwen zoals hieronder afgebeeld. Op de plaats waar de drie fuseren, begint een 1-vormsymmetrie $\alpha(g, h, k)$. Ook meetkundig komen de dimensies mooi overeen. Twee oppervlakken die (transversaal) snijden, snijden elkaar in een lijn. Als hier nu nog een derde oppervlak doorloopt, krijgen we een enkel punt. Dit punt kan dus de oorsprong voor een lijn vormen, de meetkundige structuur van een 1-vormsymmetrie in drie dimensies. Volg je nog?



Figuur 5.11: Fusie van symmetrie-

De fusie van drie 0-vormsymmetrieën speelt als het ware de rol van een bron voor een 1-vormsymmetrie. Algebraïsch

kan dit worden voorgesteld als

$$\mathrm{d}B = \alpha(A)\,.$$

De gebruikelijke Gausswet voor elektromagnetisme (in het vacuüm) stelt dat de kromming of veldsterkte globaal verdwijnt:

$$F_A = \mathrm{d}A = 0\,.$$

Als we naar de ijktheorie van een 2-groep kijken dan zien we dat, zelfs in het krommingsloze geval (dus zonder fysieke bronnen), de ene component van de 2-groep een bron kan vormen voor de andere component. De Postnikovklasse $\alpha \in H^3(G_0; G_1)$ van onze 2-groep zorgt voor een intrinsieke kromming!

Vooraleer we dit hoofdstuk afsluiten, bespreken we nog even een verdere veralgemening van het concept symmetrie. Tot nu toe zijn we enkel zogenaamde inverteerbare symmetrieën tegengekomen, daar ze worden gegeven door groepen (of veralgemening daarvan). Na een transformatie kunnen we altijd nog terugkeren naar de initiële situatie door een inverse transformatie uit te voeren. In recente jaren is echter gebleken dat deze niet volstaan om alle mogelijke symmetrieën in de natuur te beschrijven, er bestaan ook niet-inverteerbare symmetrieën. We zullen de fusieregels van n-groepsymmetrieën, die bepaald werden door hogere analogieën van groepen, vervangen door regels die plaatsvinden in de wereld van de categorieën.

Herinner je je nog dat een 2-groep kon worden gedefinieerd als een soort monoïdale categorie. Dit is ook nu het uitgangspunt, aangezien we zonder product geen fusieregels kunnen definiëren. De typische groepsvermenigvuldiging

$$G \times G \to G : (g, h) \mapsto gh\,,$$

moet nu echter plaats geven aan een regel die eerder door de lineaire al-

gebra geïnspireerd lijkt:

$$U_g \otimes U_h \cong \bigoplus_k F_{gh}^k U_k \,.$$

Hierin zijn de componenten F_{gh}^k van de F-symbolen simpelweg natuurlijke getallen. De objecten U_g kunnen als een soort basis gezien worden en, bij uitbreiding, vormen deze fusiecategorieën een categorische veralgemening van algebra's.

In dergelijke categorieën hebben we bijvoorbeeld geen invers object meer. Ook dit moet verzwakt worden tot een 'duale' operator:

$$U_g \otimes \overline{U_g} \cong \mathbf{1} \oplus \cdots \,.$$

Als we deze objecten zouden implementeren op meetkundige manier, zoals de elementen uit een n-groepsymmetrie, dan zou de duale operator $\overline{U_g}$ overeenkomen met het meetkundige object dat bij U_g hoort, maar met de omgekeerde oriëntatie.

Het onderzoek naar n-groepen en (hogere) fusiecategorieën is nog steeds een bijzonder actief vakgebied met toepassingen in zowel de zuivere wiskunde, hoge-energiefysica en vastestoffysica. Zo zijn er bijzonder sterke relaties tussen bijzondere (hogere) veldentheorieën en (hogere) fusiecategorieën.

@@ OPVULLEN (simple objects? ...) @@

Hoofdstuk 6

Algebraïsche meetkunde: meetkunde of algebra?

Misschien heb je wel het gevoel gekregen dat we in de theoretische fysica uitsluitend met gladde functies en variëteiten werken. Toch is dit niet helemaal correct, zeker niet in het theoretische onderzoek vandaag de dag. In dit hoofdstuk zullen we geen ruimtes beschouwen die lokaal door gladde functies of coördinaten beschreven worden, maar door veeltermen (of algemene ringen).

De algebraïsche meetkunde staat gekend als een notoir complex deelgebied van de wiskunde, dus je zou je misschien al zorgen kunnen maken.[1] Toch is dit best gek, want in het middelbaar zijn het net de veeltermen waar we volop mee rekenen en die we gebruiken om meetkundige figuren te beschrijven. Rechten werden beschreven door middel van lineaire vergelijkingen, parabolen door tweedegraadsveeltermen in één veranderlijke, cirkels door tweedegraadsveeltermen in één of twee veranderlijke (afhankelijk van de gekozen coördinaten), het snijpunt van grafieken

[1] Dit is veruit het meest technische hoofdstuk in dit boek, dus het is je vergeven als je het volledig overslaat of onmiddellijk overgaat naar de extra's die wat meer fysica bevatten.

door stelsels van veeltermvergelijkingen, enzovoort. Stiekem traden jullie reeds in de voetstappen van enkele legendarische namen uit de wiskunde.

We beginnen ook hier bij het begin en vertrekken van de verzameling $\mathbb{R}[x_1, \ldots, x_n]$ van veeltermen (ook wel polynomen genoemd) in $n \in \mathbb{N}$ variabelen. Op school kom je deze genoeg tegen, bijvoorbeeld $x+5$, x^2-3 of $-7x^3+5x^2$. (Vaak willen we in de praktijk ook veeltermen met complexe coëfficiënten beschouwen, maar dit is hier nog niet van belang.) Deze verzameling heeft twee voor de hand liggende bewerkingen. Enerzijds de optelling, waardoor deze verzameling de structuur van een vectorruimte krijgt, en anderzijds de vermenigvuldiging, waardoor deze verzameling ook de structuur van een algebra krijgt zoals in Hoofdstuk 5. In dit hoofdstuk zullen we zien dat al dee structuur ons toelaat om alle meetkundige objecten die je in de (klassieke) algebraïsche meetkunde kan tegenkomen te beschrijven. Zo goed als meetkundige figuren die je je kan bedenken, kunnen lokaal beschreven worden door geschikte deelverzamelingen van $\mathbb{R}[x_1, \ldots, x_n]$. De vraag is natuurlijk wat geschikt hier betekent.

Elke veelterm $P \in \mathbb{R}[x_1, \ldots, x_n]$ beschrijft een meetkundig object V_P door middel van de volgende vergelijking:

$$P(x_1, \ldots, x_n) = 0\,.$$

Het gezochte object bestaat uit alle koppels van punten $(x_1, \ldots, x_n) \in \mathbb{R}^n$ waarvoor wanneer je ze invult in P, het resultaat nul is. Klassieke voorbeelden zijn de rechte $x-y$ met richtingscoëfficiënt 1 en de cirkel x^2+y^2-1 met straal 1. Een ander voorbeeld waar we later op terugkomen is de figuur bepaald door de veelterm x^2-y^2. «*Probeer zelf eens uit te werken hoe deze figuur eruit zou kunnen zien.*»

Als gevolg van bovenstaande formule lijk het voor wezens die op V_P leven alsof P exact nul is: $P \equiv 0$. Er is voor hen geen enkele manier om het onderscheid te maken tussen de (extrinsieke) veelterm P en nul. De verzameling van relevante functies is voor hen dus kleiner dan wezens in de omringende ruimte $\mathbb{R}^n$. Bovendien geldt ook nog eens dat $PQ \equiv 0$ voor alle $Q \in \mathbb{R}[x_1, \ldots, x_n]$ voor de inwoners V_P. Zodra je een veelterm neemt die vanuit het perspectief van V_P nul is, mag je hem met eender

welke andere veelterm vermenigvuldigen, het resultaat blijft equivalent aan nul.

Als $\mathbb{R}[x_1, \ldots, x_n]$ alle veeltermen op de Euclidische ruimte $\mathbb{R}^n$ voorstelt, dan is de ring van veeltermen op V_P strikt kleiner. We moeten alle veelvouden van P gelijkstellen aan nul. Twee veeltermen $R, R' \in \mathbb{R}[x_1, \ldots, x_n]$ waarvoor geldt dat

$$R = R + PQ$$

voor een geschikte $Q \in \mathbb{R}[x_1, \ldots, x_n]$ kunnen (of moeten) we identificeren als we vanuit het perspectief van de V_P kijken: $R \equiv R'$. Net zoals in Hoofdstuk 1 nemen we dus een algebraïsch quotiënt, wat in dit geval tot de quotiëntring

$$\mathbb{R}[V_P] := \mathbb{R}[x_1, \ldots, x_n]/I(P)$$

leidt. Hierin is $I(P)$ de verzameling van alle veelvouden van P. Het wordt ook wel het ideaal gegenereerd door P genoemd. Voor alle $Q \in I(P)$ en $R \in \mathbb{R}[x_1, \ldots, x_n]$ geldt dat $QR \in I(P)$.

Voor meetkundige objecten die bepaald worden door niet één veelterm maar door een stelsel van veeltermvergelijkingen, krijgen we een gelijkaardige constructie. Hierbij wordt het ideaal I niet meer door één veelterm gegenereerd, maar door alle veeltermen in het stelsel. Het meetkundige object $V(I)$ bestaat dan uit alle oplossingen in $\mathbb{R}^n$ van alle veeltermen in I. Dergelijke objecten noemen we affiene (algebraïsche) variëteiten. De voorbeelden uit de inleiding, en bij uitbreiding alle meetkundige objecten die je in de middelbare school tegenkwam, zijn voorbeelden van affiene variëteiten.

De *Nullstellensatz*[2] van Hilbert zegt trouwens dat er een equivalentie bestaat tussen de affiene variëteiten en geschikte minimale idealen, de zogenaamde radicalen, toch zeker wanneer we over de complexe getallen

[2]Dit is Duits voor 'nulpuntenstelling', wat toepasselijk is aangezien het gaat om de nulpunten van vergelijkingen.

werken waar elke veelterm kan opgesplitst worden in lineaire factoren wegens de *fundamentele stelling van de algebra*. Als $I \subseteq \mathbb{R}[x_1, \ldots, x_n]$ een ideaal is, dan bestaat de radicaal $\sqrt{I}$ uit alle elementen $r \in \mathbb{R}[x_1, \ldots, x_n]$ zodat $r^n \in I$ voor minstens een $n \in \mathbb{N}$. De radicaal bestaat dus uit alle wortels van elementen in I.[3] Dit is een van de vele voorbeelden van hoe algebra en meetkunde twee zijden van dezelfde munt zijn. Hier komt de geest van Lawvere weer naar boven.

Door over te gaan van de meetkundige objecten op hun (lokale) functies, in dit geval hun veeltermen, kunnen we heel wat (lokale) meetkundige eigenschappen vertalen naar de algebraïsche eigenschappen van die (lokale) ringen van functies. Let wel op, in het algemeen zijn de meetkundige structuren die op deze manier bekomen worden van een totaal ander karakter dan de gladde variëteiten die we tot nu toe tegengekomen zijn. Zoals we in de laatste twee hoofdstukken van Deel 1 zagen is dit maar een kleine opoffering. Het laat ons toe om het merendeel van de gekende concepten uit de differentiaalmeetkunde te veralgemenen naar de algebraïsche meetkunde. Denk maar aan afleidbaarheid, raakruimtes, differentiaalvormen, etc.

Rond het midden van de 20e eeuw onderging de algebraïsche meetkunde een ware revolutie. Het werk van Grothendieck, Serre en anderen had een fundamentele impact op de wiskunde, waarbij ze voor zowel een enorme uitbreiding van de wiskundige gereedschapskist zorgden met concepten zoals *topoï, schoven* en *spectrale rijen* die tegenwoordig niet meer weg te denken zijn, als voor een nieuwe, revolutionaire kijk op heel wat gekende fenomenen. Het bestuderen van algemene meetkundige objecten, niet aan de hand van hun lokale structuur, maar aan de hand van de functies die erop leven, was een enorme uitbreiding van de algebraïsche meetkunde uit de tijd van Hilbert (deze laatste maakte dit nieuwe tijdperk jammer genoeg niet meer mee). Als je je geroepen voelt om hieraan bij te dragen, er bestaan twee vertalingsprojecten om de oorspronkelijke Franstalige manuscripten naar het Engels te vertalen [5, 7].

[3] Het symbool $\sqrt{\ }$ staat ook gekend als het 'radicaalteken'.

De manier waarop Grothendieck en Serre de klassieke algebraïsche meetkunde veralgemeenden, was uiterst elegant. Zoals hierboven vermeld, komen in de klassieke algebraïsche meetkunde algebraïsche variëteiten overeen met idealen van de ring $\mathbb{R}[x_1, \dots, x_n]$. Er is echter geen enkele reden waarom we ons moeten beperken tot de ring van veeltermen. We kunnen de meeste concepten veralgemenen voor eender welke (commutatieve) ring R. Als eerste moeten we dus bepalen welk soort ideaal geschikt is om de notie van een punt te beschrijven.

De juiste soorten idealen blijken diegenen te zijn die het lemma van Euclides veralgemenen: als het priemgetal $p \in \mathbb{N}$ een deler is van $ab \in \mathbb{N}$, dan is het ook een deler van a of b. De idealen die dit lemma veralgemenen, worden dan ook geheel toepasselijk priemidealen genoemd. Als $ab \in I$ dan geldt ofwel $a \in I$ of $b \in I$. De verzameling van alle priemidealen van een ring R wordt het spectrum $\mathsf{Spec}(R)$ genoemd en geeft een directe veralgemening van affiene variëteiten. Spectra en idealen zullen de basisconcepten zijn waarmee we ideeën uit de vertrouwde differentiaalmeetkunde overdragen naar de nieuwe algebraïsche meetkunde. Om dit wat meer invulling te geven, voor affiene variëteiten komt het priemideaal $\mathfrak{m}_x$ horende bij het punt $x = (a_1, \dots, a_n)$ neer op het ideaal $(x_1 - a_1, \dots, x_n - a_n)$. «*Reken maar eens na dat deze verzameling inderdaad bestaat uit alle veeltermen die nul zijn wanneer je dit punt invult.*»

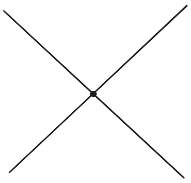
We kunnen ons wereldbeeld echter nog wat verder verbreden. In de differentiaalmeetkunde zijn niet alle ruimtes van de vorm $\mathbb{R}^n$, maar ze zijn wel altijd lokaal van die vorm (per definitie). We plakken kopieën van de Euclidische ruimte samen om zo een meer algemeen object te bekomen. In de klassieke algebraïsche meetkunde wordt exact hetzelfde gedaan. Men neemt een hoop affiene variëteiten en plakt die samen. Het resultaat wordt een algebraïsche variëteit genoemd.

Hetzelfde kunnen we nu doen met spectra. Elk spectrum kunnen we omtoveren tot een topologische ruimte door te stellen dat de gesloten verzamelingen van de vorm

$$V(I) = \{\mathfrak{p} \in \mathsf{Spec}(R) \mid I \subseteq \mathfrak{p}\}$$

zijn, voor elk ideaal $I \subseteq R$. Wanneer we dan een topologische ruimte beschouwen die lokaal op een spectrum $\mathsf{Spec}(R)$ lijkt, spreken we van een schema (gemodelleerd op R). Dit was de grote vernieuwing van Grothendieck. Het is wel belangrijk om nogmaals op te merken dat de resulterende structuur uitermate sterk verschilt van de gekende ruimtes. De Zariskitopologie op variëteiten is alles behalve fijn. Elke open deelverzameling bevat zo goed als alle punten in de ruimte. In tegenstelling tot in de gebruikelijke topologie op de reële lijn $\mathbb{R}$, zullen de 'open' intervallen $]a, b[$ dus niet meer open zijn in de Zariskitopologie.

In de definitie van een schema hoeft R zelfs geen 'brave' ring te zijn. Het kan bijvoorbeeld een ring met infinitesimale of fermionische elementen zijn. Op deze manier krijgen we bijvoorbeeld *superschema's* die onder andere toepassing vinden in de snaartheorie of synthetische meetkunde (alweer in de geest van Lawvere). De gladde variëteiten uit de differentiaalmeetkunde zijn dan weer voorbeelden van schema's gemodelleerd op de *gladde ringen* $C^\infty(\mathbb{R}^n)$.



Figuur 6.1: Een singuliere variëteit in het vlak.

Tot nu toe hebben we heel wat uitleg en definities gelezen, maar relatief weinig concrete voorbeelden gezien. Laten we daar dus eens verandering in brengen. Om het meteen wat interessanter te maken, beschouwen we een algebraïsche variëteit die niet glad is: $\mathsf{Spec}\left(\frac{\mathbb{R}[x,y]}{(x^2-y^2)}\right)$. (Het omgekeerde van glad noemen we trouwens singulier.) Hiernaast kan je de ruimtelijke voorstelling zien van dit spectrum.

De veelterm $x^2 - y^2$ kunnen we ontbinden als $(x + y)(x - y)$ en het overeenkomstige meetkundige object is de unie van de objecten die met de afzonderlijke factoren overeenkomen. We krijgen dus twee rechten die in de oorsprong snijden. Dit is trouwens de reden waarom we normaal over een veld zoals de complexe getallen werken. Voor dergelijke algebraïsch gesloten velden kan elke veelterm ontbonden worden in lineaire factoren. De oorsprong in

onze figuur is jammer genoeg een vervelend punt. Net daar wijkt de figuur af van een gladde variëteit. Maar hoe kunnen we dat nu karakteriseren?

Denken we aan gladde structuren, dan denken we ook aan afgeleiden. Denken we aan afgeleiden, dan denken we dankzij Deel 1 ook aan raakvectoren. Een niet-singulier punt zullen we dan ook definiëren als een punt waar de raakvectoren zich 'goed' gedragen. Hiervoor zullen we opnieuw de relatie met Taylorontwikkelingen gebruiken, waar de afgeleide van een functie overeenkwam met de eerste-orde benadering van die functie.

Voor elke definiërende veelterm $f \in I$ van een affiene variëteit zullen we met f_1 de termen aanduiden die van de eerste graad zijn. Bijvoorbeeld: Voor $f = x^2 + x - 3y$ hebben we dat $f_1 = x - 3y$. Twee functies f, g hebben dezelfde afgeleide in het punt $x_0 \in \mathbb{R}$ als en slechts als geldt dat

$$f(x) = f(x_0) + \xi(x_0)(x - x_0) + f''(x_0)(x - x_0)^2 + \cdots$$

en

$$g(x) = g(x_0) + \xi(x_0)(x - x_0) + g''(x_0)(x - x_0)^2 + \cdots$$

in een omgeving van x_0, waarbij $\xi(x_0)$ uiteraard hun gemeenschappelijke afgeleide voorstelt. Met andere woorden, ze hebben dezelfde afgeleide als en slechts als ze hetzelfde eerstegraadsdeel hebben: $f_1 = g_1$. Heel lokaal rondom een punt wordt de variëteit dan ook volledig bepaald door het eerstegraadsdeel van de definiërende veeltermen in I. Laten we dus als eerste gok de raakruimte als de volgende affiene variëteit definiëren:

$$T_{x_0}X := \mathsf{Spec}\left(\frac{R}{I_1}\right).$$

Bij deze definitie moeten we echter heel goed opletten! (Want het was nog niet complex genoeg.) De notie van eerstegraadsdeel hangt sterk af van de gebruikte coördinaten, we zondigen dus tegen een van de belangrijkste leidraden ons verhaal. We moeten naar de eerstegraadstermen kijken in de Taylorontwikkeling in een omgeving van het punt dat we beschouwen. We mogen dus niet zomaar stellen dat onze veelterm $x^2 - y^2$

geen termen van de eerste orde heeft. We dienen deze veelterm eerst uit te drukken in termen van $x - x_0$ en $y - y_0$ (als het ware verschuiven de oorsprong naar het punt (x_0, y_0)), die toevallig[4] ook het priemideaal van dat punt bepalen:

$$\begin{aligned} x^2 - y^2 &= x^2 - y^2 - (x - x_0)^2 + (y - y_0)^2 + (x - x_0)^2 - (y - y_0)^2 \\ &= x^2 - y^2 - x^2 + 2xx_0 - x_0^2 + y^2 - 2yy_0 + y_0^2 \\ &\qquad + (x - x_0)^2 - (y - y_0)^2 \\ &= 2x_0(x - x_0) - 2y_0(y - y_0) + (x - x_0)^2 - (y - y_0)^2 \end{aligned}$$

We hebben wel degelijk een lineair stuk.

Onze raakruimte krijgt bijgevolg de vorm

$$T_{(x_0,y_0)}X := \mathsf{Spec}\left(\frac{\mathbb{R}[x,y]}{\big(x_0(x - x_0) - y_0(y - y_0)\big)}\right)$$

die er heel wat meer intimiderend uitziet dan het eigenlijk is. De uiteindelijke structuur van deze verzameling hangt nu wel af van de specifieke waarde van x_0 en y_0. We kunnen een onderscheid maken tussen twee gevallen: $x_0 \neq 0$ (en dus $y_0 \neq 0$) en $x_0 = y_0 = 0$. Weg van de oorsprong is

$$x_0(x - x_0) - y_0(y - y_0) = 0$$

een eerstegraadsvergelijking die we kunnen oplossen naar y. Het quotiënt zorgt er dus voor dat we de afhankelijkheid van y volledig annihileren en enkel $\mathbb{R}[x]$ overblijft. Met andere woorden krijgen we:

$$T_{(x_0,y_0)}X := \mathsf{Spec}\left(\frac{\mathbb{R}[x,y]}{\big(x_0(x - x_0) - y_0(y - y_0)\big)}\right) \cong \mathsf{Spec}\big(\mathbb{R}[x]\big) \cong \mathbb{R}\,.$$

Het resultaat is de reële lijn, een eendimensionale vectorruimte, wat ook intuïtief klopt aangezien het meetkundige object weg van de oorsprong

[4] Toeval bestaat uiteraard niet.

inderdaad lokaal op een lijn lijkt. Voor $x_0 = y_0 = 0$ is het resultaat echter anders. In dit geval zakt de vergelijking

$$x_0(x - x_0) - y_0(y - y_0) = 0$$

ineen, daar het linkerlid ook exact nul is. Het quotiënt is in dit geval geen echt quotiënt, er verandert helemaal niks:

$$T_{(0,0)}X := \mathsf{Spec}\left(\frac{\mathbb{R}[x,y]}{(0)}\right) = \mathsf{Spec}\big(\mathbb{R}[x,y]\big) \cong \mathbb{R}^2\,.$$

Het resultaat is een tweedimensionale vectorruimte die weerspiegelt dat we rond de oorsprong in twee richting kunnen bewegen.

Het singuliere punt in Figuur 6.1, het snijpunt in de oorsprong (we spreken ook wel van een node), kan worden gedetecteerd aan de hand van de dimensie van zijn raakruimte. In het algemeen komen singuliere punten overeen met die punten waar de dimensie van de raakruimte groter is dan de dimensie van het meetkundige object. De definitie die we hier hanteren is echter niet ideaal. Ze gaat uit van een extrinsieke notie van een ruimte waar we onze figuur in kunnen onderbrengen, gebruikt expliciete coördinaten en gaat ook specifiek uit van het geval van variëteiten. Voor spectra van algemene ringen zal dit jammer genoeg niet werken.

Er bestaat gelukkig een andere, intrinsieke constructie die zonder problemen kan worden veralgemeend naar de context van schema's. In het punt $x \in X$ beschouwen we de evaluatie

$$\mathsf{ev}_x : f \mapsto f(x)\,.$$

De verzameling functies die op nul worden afgebeeld door de evaluatie in x komen overeen met priemideaal $\mathfrak{m}_x$. Als $f \in \mathfrak{m}_x$ en $g \in \mathbb{R}[X]$, dan geldt:

$$\mathsf{ev}_x(fg) = f(x)g(x) = 0g(x) = 0\,.$$

Zoals we ondertussen weten zijn raakvectoren (lineaire) operatoren die inwerken op functies en voldoen aan de Leibnizregel:

$$\delta_x(fg) = \delta_x(f)g(x) + f(x)\delta_x(g)\,.$$

Als we deze voorwaarde nu naast die van de evaluatie leggen, kunnen we zien dat een raakvector δ_x in x altijd exact nul zal geven voor functies in het ideal $\mathfrak{m}_x^2$ dat wordt gegenereerd door producten van elementen in $\mathfrak{m}_x$:

$$\delta_x(fg) = \delta_x(f)g(x) + f(x)\delta_x(g) = \delta_x(f)0 + 0\delta_x(g) = 0$$

voor alle $f, g \in \mathfrak{m}_x$. Met elke raakvector $\delta_x : \mathbb{R}[X] \to \mathbb{R}$ krijgen we dus een operator $D_x : \mathfrak{m}_x/\mathfrak{m}_x^2 \to \mathbb{R}$. Met het oog op de volgende paragraaf is het ook belangrijk om op te merken dat het quotiënt $\mathfrak{m}_x/\mathfrak{m}_x^2$ een vectorruimte is. De basis wordt gegeven door de generatoren $x_i - a_i$ van $\mathfrak{m}_x$. «*Kan je jezelf hiervan overtuigen?*»

Omgekeerd krijgen we voor elke operator $D_x : \mathfrak{m}_x/\mathfrak{m}_x^2 \to \mathbb{R}$ een raakvector $\delta_x : \mathbb{R}[X] \to \mathbb{R}$. Eerst zorgen we ervoor dat we de functie projecteren op het ideaal $\mathfrak{m}_x$: $f \mapsto f - f(x)$. Dit resultaat moeten we dan verder projecteren door alle hogere-orde bijdragen weg te gooien: $f - f(x) \mapsto \big(f - f(x)\big)_1$. Als laatste passen we de operator D_x toe:

$$\delta_x(f) = D_x\big(f - f(x)\big)_1 .$$

De raakruimte T_xX is dus isomorf aan de vectorruimte van lineaire operatoren van $\mathfrak{m}_x/\mathfrak{m}_x^2$ naar de reële getallen. Deze ruimte is echter, per definitie, de duale ruimte van $\mathfrak{m}_x/\mathfrak{m}_x^2$:

$$T_xX \cong \left(\mathfrak{m}_x/\mathfrak{m}_x^2\right)^* .$$

Met andere woorden, $\mathfrak{m}_x/\mathfrak{m}_x^2$ is de coraakruimte aan X in x. Deze twee vectorruimtes worden in de literatuur vaak de Zariski-(co)raakruimtes genoemd om aan te geven dat het de algebraïsche varianten zijn van de gekende noties uit differentiaalmeetkunde.

Het mooie is nu dat we het ideaal $\mathfrak{m}_x$ kunnen definiëren voor algemene ringen R, zonder te moeten verwijzen naar coördinaten of omringende ruimtes. (Hoe we dat precies doen, zou ons echter veel te ver leiden.) De Zariski-(co)raakruimte kan dus zonder problemen veralgemeend worden naar de categorie van schema's, superschema's, enzovoort. Heel wat concepten uit de differentiaalmeetkunde kunnen op die manier ook veralgemeend worden naar de niet zo gladde context van schema's. Denk

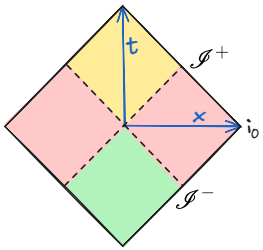
bijvoorbeeld aan differentiaalvergelijkingen. De bundels uit Hoofdstuk 3 kunnen trouwens ook op een gelijkaardige manier veralgemeend worden. Hiervoor hebben we echter het concept *schoof* nodig en ook dat zou ons te ver leiden. (Het is bijzonder moeilijk om de hedendaagse theoretische natuurkunde te volgen zonder eerst een diploma zuivere wiskunde te behalen.)

Het lot wil nu dat een gelijkaardig idee ook voor vernieuwing zorgde in de natuurkunde. Rond het midden van vorige eeuw probeerden enkele theoretici de kwantumveldentheorie van een rigoreuze wiskundige basis te voorzien. De eerste die hieraan begon was Wightman die de velden trachtte te karakteriseren als (kans)verdelingen die aan elke gebeurtenis een lineaire operator toewijzen en niet zomaar een getal zoals in de klassieke kansrekening.

Een van de technische en conceptuele problemen met deze aanpak is echter de keuze van toestandsruimte waar die lineaire operatoren moeten op inwerken. In kwantummechanica is deze keuze eigenlijk uniek bepaald, een resultaat dat gekend staat als de *stelling van Stone–von Neumann* (dezelfde von Neumann die mee aan de basis lag van de computerwetenschap, *game theory*, wiskundige economie, ...). In de kwantumveldentheorie daarentegen, waar we een oneindig, zelf onaftelbaar, aantal vrijheidsgraden hebben, geldt deze stelling echter niet. In tegendeel, Haag toonde aan dat er *überhaupt* geen unieke keuze mogelijk is. Het leek dus beter om de zoektocht naar de juiste toestandsruimte even links te laten liggen. (Dit wil echter niet zeggen dat Haag ervoor zorgde dat de ideeën van Wightman ten dode opgeschreven waren, integendeel, Wightman had net gewacht met zijn publicatie totdat Haag zijn werk afrondde.)

Reeds eerder, rond het begin van de 20e eeuw, waren enkele vooraanstaande wetenschappers zoals Einstein aanhangers van het zogenaamde *operationalisme*. In deze filosofische stroming poneert men dat grootheden maar een echte fysische betekenis krijgen als je kan zeggen hoe ze gemeten kunnen worden. (De golffunctie in kwantummechanica is dus een voorbeeld van een niet-fysische grootheid voor Einstein.) Deze filoso-

fie motiveerde Haag en Kastler om kwantumveldentheorie te definiëren aan de hand van de fysische observabelen, zonder enige verwijzing naar een vectorruimte van toestandsvectoren. Ze omzeilden dus helemaal het probleem van de illustere toestandsruimte. De theorie wordt volledig gekarakteriseerd door lokale ringen van observabelen.



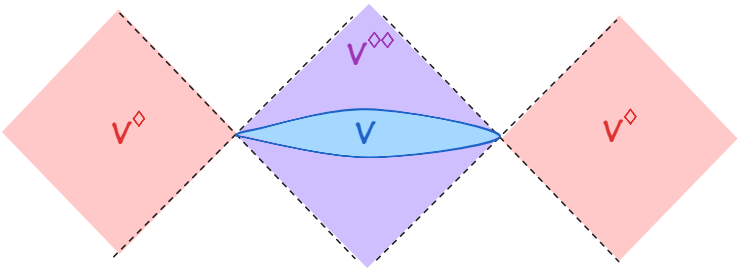
Figuur 6.2: Een Penrosediagram in 2D.

Om de axioma's van Haag en Kastler te kunnen begrijpen, dienen we de structuur van de ruimtetijd in wat meer detail te bespreken. Hiervoor nemen we even aan dat er maar één ruimtelijk dimensie is en zetten we de lichtsnelheid gelijk aan $c = 1$, de absolute favoriet van elke natuurkundige. Lokaal kunnen we dan loodrechte assen kiezen zoals in de figuur hiernaast. (In de 3D-ruimte vervang je x door de afstand r tot het middelpunt.)

Een dergelijk diagram wordt een Penrosediagram genoemd. Het laat ons toe om de Minkowskiruimte op een compacte manier te analyseren. Als we ons in de oorsprong bevinden en een lichtsignaal verzenden, dan zal dit even snel doorheen de ruimte als doorheen de tijd bewegen (aangezien $c = 1$). Het beweegt zich dus langs een van de twee diagonale stippellijnen boven de x-as. Op een geheel analoge manier zullen lichtsignalen die in de oorsprong toekomen, zich langsheen de diagonale stippellijnen onder de x-as voortbewogen hebben. In een Penrosediagram worden paden van deeltjes die zich aan de lichtsnelheid voortbewegen — dit zijn dus alle deeltjes zonder massa — voorgesteld door lijnen die een hoek van 45° maken met de x-as.

Met deze informatie kan je dan ook onmiddellijk zien wat de twee rode ruiten in het Penrosediagram voorstellen. Dit zijn alle punten in de ruimtetijd die op geen enkele manier met een causaal signaal —dit is een signaal dat zich aan de lichtsnelheid (of trager) voortbeweegt — kunnen communiceren met de oorsprong. Om dezelfde reden bestaan de groene

en gele ruiten respectievelijk uit alle punten die de oorsprong causaal kunnen beïnvloeden of door een gebeurtenis in de oorsprong causaal beïnvloedt kunnen worden. De twee diagonale lijnen $\mathscr{I}^{\pm}$ stellen de rand van de ruimtetijd op oneindig voor. Waar de twee lijnen samenkomen in i_0, vinden we ruimtelijk oneindig terug. Enkel lichtsignalen die bij het ontstaan van het universum, het onderste punt van het diagram, verzonden zijn, kunnen ruimtelijk oneindig bereiken.



Figuur 6.3: Causale complementen.

Voor een gegeven deelverzameling V van de ruimtetijd kunnen we ook kijken naar alle punten die er causaal mee verbonden zijn of, en dit blijkt nuttiger te zijn, welke punten net niet causaal verbonden zijn met V. In de figuur hierboven tonen we de verzameling V in blauw. Zijn causaal complement $V^{\diamond}$ is aangeduid in rood. In het geval dat V uit slechts een punt bestaat, zouden we opnieuw de situatie uit Figuur 6.2 krijgen. Als we nu nogmaals het causale complement zouden nemen, krijgen we de paarse ruit $V^{\diamond\diamond}$. Uit deze figuur valt duidelijk af te lezen dat V en $V^{\diamond\diamond}$ niet gelijk zijn aan elkaar, maar er wel (altijd) geldt dat $V \subseteq V^{\diamond\diamond}$. Indien de gelijkheid wel zou gelden voor een specifieke V, spreken we van een causaal gesloten verzameling.

Met onze kennis over de causale structuur van de ruimtetijd kunnen we de Haag–Kastleraxioma's voor kwantumveldentheorie op de vlakke Minkowskiruimte M^4 (in deels vereenvoudigde vorm) introduceren:

1. Voor elke causaal gesloten verzameling $U \subseteq M^4$ hebben we een geschikte ring (of algebra) van observabelen $\mathcal{O}(U)$.

2. Indien $U \subseteq U'$, dan geldt $\mathcal{O}(U) \subseteq \mathcal{O}(U')$. De orderelatie wordt behouden, in tegenstelling tot bij variëteiten en schema's waar de orderelatie wordt omgedraaid. «*Kan je inzien waarom dit zo is voor variëteiten?*»

3. Indien U en V ruimtelijk gescheiden zijn, er kan dus geen enkel lichtsignaal tussen U en V verstuurd worden (ze liggen te ver van elkaar in de ruimtetijd), dan commuteren de observabelen in $\mathcal{O}(U)$ en $\mathcal{O}(V)$.

Het laatste axioma wordt vaak (causale) lokaliteit genoemd. Wanneer twee observabelen met elkaar commuteren, hebben ze geen (causale) invloed op elkaars metingen. Het onzekerheidsprincipe van Heisenberg in zijn meest gekende vorm zei bijvoorbeeld dat dit niet het geval is voor positie en impuls (gemeten op dezelfde plek). Als je de ene meet, verstoor je ontegenzeggelijk de meting van de andere. Wanneer twee stukken ruimtetijd voldoende ver van elkaar verwijderd zijn (zoals de twee rode ruiten in Figuur 6.3), zodat er geen informatieoverdracht tussen de twee mogelijk is, kunnen metingen in die omgevingen elkaar logischerwijze onmogelijk beïnvloeden.

De eerste twee axioma's worden tegenwoordig trouwens vaak samengenomen. In Deel 1 werd de notie van een functor geïntroduceerd. Dit was de correcte veralgemening van een structuurbehoudende afbeelding tussen twee categorieën. Als je nu enerzijds de categorie **Loc** van causaal gesloten verzamelingen neemt en anderzijds die van (geschikte) ringen **Alg**, dan zeggen voorwaarden 1 en 2 dat een kwantumveldentheorie een functor $\mathcal{O} : \mathbf{Loc} \to \mathbf{Alg}$ is. (Schema's zijn voorbeelden van contravariante functoren die de richting van morfismen omdraaien.)

De afgelopen dertig jaar is de theorie van Haag en Kastler onder meer uitgebreid naar algemene ruimtetijden [8], maar waar de oorspronkelijke axioma's een niet-perturbatieve theorie geven, werkt men tegenwoordig met een perturbatieve versie, waar de ringen van operatoren vervangen

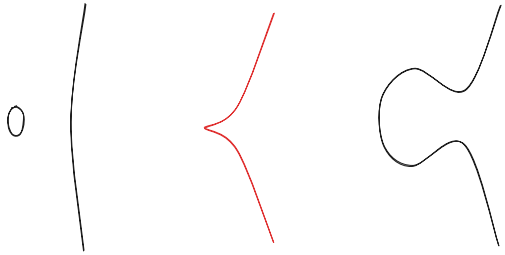
zijn door ringen van oneindige reeksen in een 'kleine' parameter $\hbar$ zoals in Hoofdstuk 4 van Deel 1. Dit toont onder meer hoe deze theorie nauw verbonden is aan deformatiekwantisatie.

Extra's: Elliptische krommen

We sluiten dit hoofdstuk af met een kleine omweg langsheen de algebraïsche getaltheorie, waar men problemen zoals de *laatste stelling van Fermat* bestudeerd aan de hand van algebraïsche technieken. We beginnen hiervoor met de definitie van een elliptische kromme:

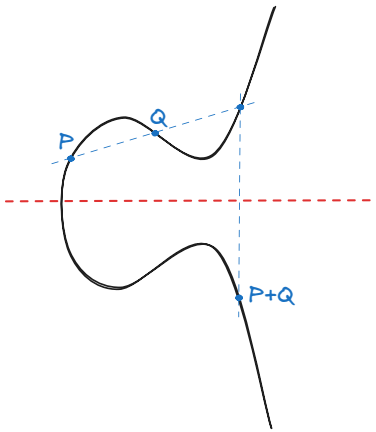
$$y^2 = x^3 + ax + b$$

voor $a, b \in \mathbb{R}$. Hieronder staan drie voorbeelden afgebeeld. In het midden staat een kromme die singulier is met een zogenaamde spits[5] in de oorsprong. Dergelijke singuliere voorbeelden sluiten we uit in de definitie van elliptische krommen, we laten enkel gladde krommen toe. Merk op dat, zoals het linker voorbeeld duidelijk maakt, algebraïsche krommen niet noodzakelijk uit één deel bestaan.



Figuur 6.4: Drie elliptische krommen met parameters $(-2,-1)$, $(0,0)$ en $(-2,2)$.

[5] In het Engels spreekt men over een *cusp*.



Figuur 6.5: Optelling op een elliptische kromme.

Als eerste beschouwen we een heel belangrijke, doch onverwachte eigenschap van elliptische krommen: ze vormen een groep! (Dit kan zelfs als een van de definiërende eigenschappen gebruikt worden.) Nu vraag je je waarschijnlijk af hoe een gekke kromme ooit een groep kan vormen. Hiervoor gaan we opnieuw naar de schoolbanken en halen we onze lat boven.

Hiernaast is de generieke grafische constructie weergegeven. Neem twee punten P, Q op de kromme en trek er een rechte door. Tenzij P en Q spiegelbeelden om de x-as zijn, zal de rechte een derde snijpunt met de kromme hebben. In dat geval spiegelen we dit derde snijpunt om de x-as en bekomen we de som $P + Q$. De som van de drie snijpunten P, Q, R van een rechte met een elliptische kromme voldoen dus altijd aan

$$P + Q + R = 0\,.$$

Er zijn echter een aantal situaties waarbij dit niet zo vanzelfsprekend is, dus laten we de constructie in wat meer detail bespreken.

1. Het neutrale element O wordt gegeven door 'het punt op oneindig'. (Om dit concreet te maken zouden we nog wat projectieve meetkunde boven moeten halen, maar dat maakt het enkel technischer.) Het neutrale element O van de groep is dus eigenlijk een punt dat we nooit gaan kunnen construeren.

2. Indien $P \neq Q \neq O$, dan werkt de constructie zoals in de figuur hierboven zonder problemen.

3. Indien $P = Q$, dienen we de raaklijn in P te gebruiken. Aangezien we ervan uitgaan dat de kromme altijd glad is, is deze raaklijn goed gedefinieerd.

4. Indien P en Q spiegelbeelden zijn, stellen ze elkaars inverse voor. Het derde snijpunt is hierbij dus het punt op oneindig.

In de praktijk is het uiteraard nuttig om ook expliciete formules te hebben, dus laten we eens de twee situaties $P \neq Q$ en $P = Q$ bekijken. Voor twee verschillende punten kunnen we de cultuurliefhebbers blij maken met de formules van Vieta[6] De richtingscoëfficiënt van de rechte door P en Q wordt gegeven door

$$\lambda = \frac{y_Q - y_P}{x_Q - x_P}$$

en met deze coëfficiënt kunnen we de vergelijking voor de elliptische kromme herschrijven als

$$\big(\lambda x + (y_P - \lambda x_P)\big) = x^3 + ax + b\,.$$

Indien we nu alle termen naar dezelfde zijde brengen, bekomen we een derdegraadsvergelijking in x:

$$x^3 - \lambda^2 x^2 + \big(a - 2\lambda(y_P - \lambda x_P)\big) + \big(b - (y_P - \lambda x_P)^2\big) = 0\,.$$

We weten dat x_P en x_Q twee oplossingen zijn, dus we moeten enkel nog de derde vinden. De eenvoudigste van Vieta's formules zegt ons nu dat de som van de nulpunten bepaald wordt door de twee eerste coëfficiënten in de vergelijking:

$$x_P + x_Q + x_R = -\frac{a_2}{a_3} = \lambda^2$$

of met andere woorden:

$$x_R = \lambda^2 - x_P - x_Q\,.$$

[6]Vernoemd naar de Franse wiskundige Viète, maar in die tijd (16e eeuw) waren gelatiniseerde namen behoorlijk hip.

Dit kunnen we nu invullen in de vergelijking van de rechte om de ordinaat van R te vinden:

$$y_R = \lambda x_R + (y_P - \lambda x_P) = \lambda(x_R - x_P) + y_P\,.$$

Het enige wat we dan nog moeten doen is hier een minteken voor plaatsen.

Voor twee samenvallende punten — we verdubbelen het punt P — moeten we de richtingscoëfficiënt anders bepalen. Om een uitdrukking te vinden voor de afgeleide in P, starten we van de vergelijking voor de elliptische kromme en passen we een techniek genaamd impliciete afleiding toe, waarbij we y als een functie van x zien:

$$2yy' = 3x^2 + a$$

of met andere woorden:

$$y' = \frac{3x^2 + a}{2y}\,.$$

Vanaf hier verloopt de constructie zoals in voorgaande paragraaf waar de formule van Vieta ons uitdrukkingen voor x_R en y_R bezorgt.

Er zijn enkele redenen om de groepsstructuur van elliptische krommen te bestuderen. Enerzijds is er de zuiver wiskundige motivatie om wiskundige objecten zo goed mogelijk te begrijpen. Zo is er de gevierde stelling van Mordell–Weil die stelt dat de groep $X(\mathbb{Q})$ — de verzameling van punten op een algebraïsche variëteit X met coördinaten in een veld K wordt vaak genoteerd als $X(K)$ — geassocieerd aan een elliptische kromme X over de rationale getallen altijd van de volgende vorm is:

$$X(\mathbb{Q}) \cong \mathbb{Z}^r \oplus \mathbb{Z}_{q_1} \oplus \cdots \oplus \mathbb{Z}_{q_n}$$

waarbij alle q_i gehele machten van priemgetallen zijn. Als we deze *torsiegroepen* even vergeten en enkel naar de groep $\mathbb{Z}^r$ kijken, dan zien we dat de Mordell–Weilgroep volledig bepaald wordt door een basis met r elementen (zoals bijvoorbeeld een n-dimensionaal kristalrooster). Een van de op dit moment nog openstaande Milleniumproblemen, het *vermoeden van*

Birch en Swinnerton-Dyer, probeert de algebraïsche dimensie r te relateren aan analytische informatie van de elliptische kromme.

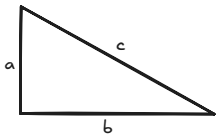
Daarnaast zijn er uiteraard ook meer praktische toepassingen, waaronder cryptografie. Misschien heb je wel al eens van cryptografische algoritmen zoals RSA of Diffie–Hellman gehoord. Een centraal idee achter de sterkte van deze algoritmes is dat priemfactorisatie van grote getallen tot op heden niet efficiënt kan gedaan worden met (klassieke) computers. De fundamentele stelling van de rekenkunde (zie de inleiding) zegt dat elk natuurlijk getal $n \geq 2$ kan ontbonden worden in priemfactoren:

$$n = p_1^{m_1} p_2^{m_2} \cdots p_k^{m_k}.$$

Het vinden van de getallen p_i is echter uitermate traag (ook al is er tot nu toe nog nooit aangetoond dat er geen efficiënt klassiek algoritme bestaat). Als twee partijen elk een heel groot priemgetal kiezen dan kunnen ze makkelijk het product berekenen, maar iemand die hen afluistert en enkel het product van de getallen krijgt, zal heel moeilijk de oorspronkelijke getallen kunnen achterhalen.

Elliptische krommen bieden hier een alternatief, gerelateerd aan het probleem van de *discrete logaritmen*. Startende van een punt $P \in X(\mathbb{Q})$ is het heel eenvoudig om een veelvoud nP te berekenen voor grote $n \in \mathbb{N}$. Omgekeerd, gegeven P en zijn veelvoud nP, is het bijzonder moeilijk om de factor n te vinden. Het voordeel van cryptografische algoritmen gebaseerd op elliptische krommen is dat de benodigde getallen in het algemeen kleiner zijn en we dus minder bandbreedte nodig hebben.

Het nadeel is dat het onderliggende concept, het *discrete logaritme*, voor al deze methodes hetzelfde is. Klassiek mag er dan misschien (nog) geen efficiënt algoritme zijn om dit probleem te kraken, kwantumcomputers hebben er minder moeite mee. Het *algoritme van Shor* zorgt er ook voor dat algoritmes met elliptische krommen efficiënt gekraakt kunnen worden. In de (nabije?) toekomst zullen we dus ook hier een oplossing moeten voor zoeken.



Figuur 6.6: Een rechthoekige driehoek.

Om af te sluiten, gaan we eens heel ambitieus doen en nemen we de laatste stelling van Fermat onder handen. Als we wat eenvoudiger starten, dan kennen we allemaal nog wel de stelling van Pythagoras, althans dat zou toch moeten:

$$a^2 + b^2 = c^2$$

geldt voor de zijden van een rechthoekige driehoek zoals in de figuur hiernaast, waarbij c de schuine zijde voorstelt. Het is relatief makkelijk om in te zien dat er oneindig veel Pythagorese drietallen zijn door gebruik te maken van een algoritme dat reeds door Euclides gekend was. Vergelijk de twee merkwaardige producten

$$(x-y)^2 = x^2 - 2xy + y^2 \qquad \text{en} \qquad (x+y)^2 = x^2 + 2xy + y^2\,.$$

Als je deze van elkaar aftrekt, dan blijft $4xy$ over. Indien we nu $x = m^2$ en $y = n^2$ voor eender welke twee natuurlijke getallen $m > n \in \mathbb{N}$ nemen, dan zie je eenvoudig dat we een Pythagorees drietal $(2mn, m^2 - n^2, m^2 + n^2)$ krijgen.

Fermat stelde zich de vraag of er ook dergelijke drietallen van natuurlijke getallen bestonden voor de vergelijking

$$a^3 + b^3 = c^3$$

of, meer algemeen, voor

$$a^n + b^n = c^n$$

met $n \geq 3$. Hij beweerde dat hij een bewijs had dat er geen enkele oplossing bestond (behalve 0). De woorden[7] "*Hanc marginis exiguitas non caperet.*" zijn ondertussen legendarisch geworden [9]. Ook al had Fermat klaarblijkelijk te weinig plaats om een algemeen bewijs op te schrijven (men is trouwens nog steeds niet zeker of hij het al dan niet echt had), hij

[7] Latijn voor "Deze kantlijn is te nauw om het te bevatten."

gaf wel een bewijs voor het geval $n = 4$. Enkele decennia later gaf Euler een uiterst elegant bewijs voor $n = 3$, met als gevolg dat men enkel nog de gevallen voor (oneven) priemgetallen moest aantonen.

Eeuwenlang hebben wiskundigen hier bijdragen aan geleverd, waarbij beetje bij beetje specifieke situaties werden aangetoond. Het probleem is echter dat er oneindig veel priemgetallen zijn, dus die specifieke situaties waren onvoldoende, en het kostte meerdere eeuwen om tot een sluitend resultaat te komen. Het uiteindelijke bewijs bevat dermate complexe wiskunde dat we het hier uiteraard niet in zijn volledige vorm kunnen geven, maar we zullen wel enkele aspecten toelichten. Ons stappenplan ziet er als volgt uit: we relateren de stelling aan elliptische krommen, we introduceren het concept 'modulariteit', we relateren elliptische krommen aan modulariteit en we komen tot een conclusie.

Stap 1: Elliptische krommen) In de tweede helft van de 20e eeuw kwamen enkele slimme wiskundigen op het lumineuze idee om een elliptische kromme te construeren op basis van de vergelijking in de laatste stelling van Fermat. Laten we aannemen dat er een tegenvoorbeeld (a, b, c) van de stelling bestaat. De Frey-kromme voor dit drietal wordt gegeven door de volgende vergelijking over de rationale getallen:

$$y^2 = x(x - a^p)(x + b^p)\,.$$

Indien het tegenvoorbeeld echt kon bestaan, dan zou dit de vergelijking van een elliptische kromme over $\mathbb{Q}$ zijn. Het blijkt echter dat dit een zeer bijzondere kromme zou zijn, met dank aan de exponenten die in de factoren voorkomen.

Wanneer we een elliptische kromme met gehele coëfficiënten hebben, is het vaak nuttig om deze te reduceren. Voor elk priemgetal $p \in \mathbb{N}$ kunnen we de vergelijking modulo p beschouwen (voor $p = 2, 3$ moeten we iets meer opletten, maar dat detail laten we hier terzijde):

$$y^2 = x^3 + (a \bmod p)x + (b \bmod p) \bmod p\,.$$

De reden hiervoor is dat de groep van natuurlijke getallen modulo p een eindig veld $\mathbb{F}_p$ vormen en we dus nog steeds de gebruikelijke technieken

uit de algebraïsche meetkunde kunnen hanteren. «*Reken maar eens na dat* $\mathbb{F}_p$ *aan alle voorwaarden voldoet.*»

Het meetkundige object X_p dat we overhouden na de reductie modulo p is nog steeds een algebraïsche variëteit, maar het is niet noodzakelijk een variëteit zonder singulariteiten en dus ook niet noodzakelijk een elliptische kromme. Er zijn drie mogelijkheden:

- Goede reductie: X_p is glad.
- Semistabiele slechte reductie: X_p heeft een node.
- Instabiele slechte reductie: X_p heeft een spits.

Gelukkig kan men aantonen dat het aantal priemgetallen waarvoor de reductie slecht is, ook eindig is. In het algemeen is de reductie goed en krijgen we een mooi resultaat.

Op basis van deze classificatie kunnen we nu een zogenaamde L-reeks construeren:

$$L_X(s) := \prod_{p \text{ priem}} \alpha(p, s)$$

waarbij de expliciete vorm van α afhangt van het reductietype van X_p. (Jullie blijven gespaard van de volledige uitdrukking.) Het eerste dat je misschien opmerkt is dat dit helemaal geen reeks is, maar een oneindig product (dit wordt ook wel een *Eulerproduct* genoemd). Het toeval wil nu dat elke factor $\alpha(p, s)$ van de vorm

$$\alpha(p, s) = \frac{1}{1 - f(p)p^{-s}}$$

is voor een geschikte keuze van de coëfficiënt $f(p)$. We kunnen dus de formule voor meetkundige reeksen gebruiken om elke factor als een reeks te schrijven. Al deze reeksen kunnen we dan (formeel) vermenigvuldigen en de termen groeperen, en door gebruik te maken van de fundamentele

stelling van de rekenkunde, kunnen we producten van priemgetallen vervangen door het bijhorende natuurlijke getal. De L-reeks kan dus als een echte reeks herschreven worden:

$$L_X(s) = \sum_{n=1}^{+\infty} \frac{a_n}{n^s}\,.$$

Als je op dit moment geen enkel idee hebt waar we naartoe willen, dat is volledig begrijpelijk! Deze uitdrukking heeft voor wiskundigen heel wat betekenis, hij is onder meer uniek voor elke elliptische kromme, maar zegt niet zo veel voor een onschuldige fysicus. Wat we wel kunnen doen, is de coëfficiënten a_n hergebruiken om een ander object te construeren.

Stap 2: Modulaire vormen) In Hoofdstuk 8 zullen we het hebben over de fasen van de materie (en een veralgemening van de klassieke betekenis). Bij klassieke fasetransities, zoals wanneer je overgaat van vloeibaar water naar ijs, gebeurt er iets vreemds. Zolang de volledige vloeistof niet bevroren is, zal de temperatuur niet onder nul dalen. Dit worden ook wel eerste-orde fasetransities genoemd. Tegenwoordig zijn er echter heel wat fasetransities gekend van hogere ordes, in het bijzonder van de tweede orde. Op de kritische punten, waar twee fasen niet meer te onderscheiden vallen, worden deze beschreven door een speciaal soort veldentheorie: een conforme veldentheorie.

Dit zijn theorieën die invariant zijn onder herschaling. Een beetje zoals in een *fractaal* kunnen we inzoomen en uitzoomen zonder een merkbaar verschil te zien in de microscopische beschrijving. De correcte beschrijving heeft zelfs nog meer symmetrie en wordt beschreven door de modulaire groep:

$$\mathsf{SL}(2,\mathbb{Z}) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \;\middle|\; ad - bc = 1 \text{ en } a,b,c,d \in \mathbb{Z} \right\}.$$

Deze groep bestaat uit alle 2×2-matrices met gehele coëfficiënten en waarvan de determinant gelijk is aan 1. We hebben nu enkel nog een vectorruimte nodig waarop deze matrices kunnen inwerken. Gelukkig komen

de complexe getallen tot onze redding:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} z = \frac{az+b}{cz+d}\,.$$

Als we nu functies van complexe getallen beschouwen, dan kunnen we ook eens kijken welke functies invariant zijn onder modulaire transformaties. Omdat strikte invariantie te sterk blijkt, relaxeren we de voorwaarde wat:

$$f\left(\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) z\right) = (cz+d)^k f(z)$$

voor een zekere exponent $k \in \mathbb{N}$. Dit zijn de modulaire functies of modulaire vormen (toch zeker als ze overal *complex afleidbaar* zijn). Als we ons nu beperken tot een van de belangrijkste deelgroepen van $\mathsf{SL}(2,\mathbb{Z})$, de translatiegroep, gegenereerd door $a = b = d = 1$ en $c = 0$, dan bekomen we de volgende relatie:

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} z = z + 1\,.$$

Als we dit in de voorwaarde voor modulaire vormen gieten, dan zien we dat deze in het bijzonder voldoen aan

$$f(z+1) = f(z)\,.$$

Alle modulaire functies zijn periodiek!

Als je je Hoofdstuk 2 nog herinnert, dan herinner je je misschien ook nog dat elke periodieke functie een Fourierreeks heeft:

$$f_X(z) = \sum_{n=1}^{+\infty} a_n e^{2\pi i n z}\,,$$

met $z \in \mathbb{C}$. We kunnen modulaire vormen dus ook karakteriseren aan de hand van hun Fourierreeks (waarbij er relaties bestaan dus de coëfficiënten a_n die onder meer afhangen van de exponent k in de modulariteitsvoorwaarde).

Stap 3: Modulariteitsstelling) In Stap 1 hebben we gezien dat we met elke elliptische kromme een L-reeks kunnen associëren en in Stap 2 hebben we gezien dat we met elke modulaire vorm een Fourierreeks kunnen associëren. Twee objecten, twee reeksen. Kunnen we hier echter een verband tussen vinden?

Deze vraag stond centraal in de studie van de laatste stelling van Fermat. De cruciale doorbraak kwam er toen Wiles de modulariteitsstelling[8] aantoonde. Het idee is simpel. Neem de coëfficiënten in de L-reeks van een elliptische kromme en gebruik deze om een Fourierreeks te bouwen. Het resultaat is altijd een modulaire vorm (een speciale zelfs). Omgekeerd, neem de coëfficiënten in de Fourierreeks van een (geschikte) modulaire vorm en gebruik deze om een L-reeks te bouwen. Het resultaat is altijd een elliptische kromme.

Dit verband lijkt op het eerste gezicht bijzonder toevallig en uiterst abstract, en dat laatste is het zeker. Het verband tussen twee ogenschijnlijk verschillende deelgebieden zoals getaltheorie, algebraïsche meetkunde en complexe analyse is echter van uitermate groot belang in de wiskunde. Deze modulariteitsstelling is trouwens nog maar het begin van een veel rijker onderwerp dat gekend staat als de *overeenkomst van Langlands*. (De snaartheorie probeert ook hierin een rol te spelen.)

Stap 4: De contradictie) Alles lijkt uitermate goed te gaan voor de elliptische krommen, maar tot nu toe lijken we nog geen vooruitgang te boeken richting een bewijs van de laatste stelling van Fermat. Serre had echter eerder voorgesteld dat de modulariteitsstelling samen met een klein extraatje voldoende zou moeten zijn voor een bewijs:

$$\text{modulariteit} + \varepsilon = \text{Fermat}\,.$$

In de fysica en wiskunde worden infinitesimale grootheden vaak met ε aangeduid en om die reden werd het ontbrekende deel ook wel de 'ε-vermoeden' genoemd.

Dit vermoeden zei dat elliptische krommen zoals de Frey-kromme (in-

[8] Deze staat ook wel bekend als de stelling van Shimura–Taniyama(–Weil).

dien deze kon bestaan) een heel bijzonder eigenschap hebben. Voor de Frey-kromme in het bijzonder zou dit resultaat aantonen dat de exponent in de modulariteitseigenschap van de bijhorende modulaire vorm relatief klein zou zijn: 2. Men wist echter reeds dat de kleinste exponent voor het soort modulaire vorm dat uit de modulariteitsstelling kan komen minstens 11 moet zijn. De Frey-kromme kon dus helemaal niet modulair zijn, in tegenspraak met de modulariteitsstelling. Indien de kromme bestond, zouden we een contradictie hebben en dat is niet mogelijk in de wiskunde. De stelling van Fermat was aangetoond!

Hoofdstuk 7

Statistiek: Een gevaarlijke wereld

Tot nu toe hebben we voor klassieke systemen hoogstens een handvol deeltjes beschouwd. Indien het, net zoals in Hoofdstuk 8, zou gaan over systemen met een uiterst groot aantal deeltjes (van de orde van het getal van Avogadro), dan moeten we ook in de theoretische context overgaan op een ander formalisme. De wetten van Newton laten ons in dergelijke gevallen in de steek.

Het idee van Boltzmann en anderen in de 19e eeuw was om dergelijke fysische systemen op een statistische manier te beschrijven. In plaats van een microscopische beschrijving, waarbij de beweging en eigenschap van de individuele deeltjes gemodelleerd worden, hanteren we een macroscopische beschrijving van de gemiddelde eigenschappen van de deeltjes. De belangrijkste aanname hierbij is dat het systeem zich in evenwicht bevindt (al zijn ook hier ondertussen veralgemening voor gevonden). Er mogen geen externe krachten op het systeem inwerken en we moeten lang genoeg wachten zodat de deeltjes zich voldoende verspreid hebben en er geen informatie over de initiële toestand meer beschikbaar is. Zo zullen onder

meer de energie en de impuls van de deeltjes zich mooi verdelen door de botsingen die tussen de deeltjes plaatsvinden.

Een van de kansverdelingen die elke natuurkundige hoort te kennen, is de Maxwell–Boltzmannverdeling. Deze beschrijft hoeveel deeltjes zich in een gegeven 'macrotoestand' bevinden (zonder informatie over de exacte 'microtoestand'). Het verschil tussen micro- en macrotoestand is hier van groot belang. Als je bijvoorbeeld de energie E kent, weet je eigenlijk niks over de impuls $\vec{p}$ van het deeltje. De snelheid kan volledig langs de x-as liggen, uniform verdeeld zijn over de drie richtingen of totaal willekeurig verdeeld zijn. Dit is exact waarom de statistische formulering zo interessant is, de specifieke details doen er niet toe.

Als eerste zullen we trachten om deze verdeling vanuit enkele belangrijke basisprincipes af te leiden. Het belangrijkste principe gaat terug op het werk van Jaynes en Shannon, twee reuzen uit de informatietheorie. Het principe van maximale entropie zegt dat, gegeven enkele voorwaarden, we die verdeling moeten kiezen die de entropie (of ook wel wanorde genoemd) maximaliseert. De fysische tegenhanger hiervan is de tweede wet van de thermodynamica, die zegt dat de entropie in het universum (of eender welk gesloten systeem) nooit afneemt:

$$\Delta S \geq 0\,.$$

Als het systeem zich in evenwicht bevindt, zal de entropie dus maximaal zijn.

Als maat voor de wanorde zou de entropie groot moeten zijn wanneer de kans om de gegeven toestand waar te nemen klein is. Een goed basisingrediënt is dus de informatie van de kans p_ω:

$$I(p_\omega) := -\ln(p_\omega)\,.$$

Daar $p_\omega \in [0,1]$ voor eender welke toestand $\omega \in \Omega$, zal de logaritme altijd negatief zijn. De informatie wordt dus groter, naarmate de kans kleiner wordt. Indien we de informatie uitmiddelen over alle toestanden, beko-

men we de Shannonentropie (of Gibbsentropie):

$$S(p) := -\sum_{\omega\in\Omega} p_\omega \ln(p_\omega)\,.$$

Deze functie, of zijn continue variant waarbij we de som door een integraal vervangen, zullen we nu trachten te maximaliseren. Hiervoor hanteren we de inhoud van Hoofdstuk 4 en zullen we dus de functie S variëren met betrekking tot de kansverdeling p. Om fysische correct te zijn, dienen we echter nog een voorwaarde op te leggen. De gemiddelde energie E van de deeltjes, daar we een afgesloten systeem beschouwen, moet constant zijn: $\langle E\rangle = \mu$.

Dergelijke voorwaarden in rekening brengen is gelukkig makkelijker dan je zou kunnen verwachten. We dienen enkel de Lagrangiaan wat uit te breiden. De oorspronkelijke Lagrangiaan voor het principe van maximale entropie is

$$L(p) := -S(p)\,.$$

Hier voegen we nu een term evenredig met de voorwaarde aan toe:

$$\tilde{L}(p,\beta) := -S(p) + \beta\big(\langle E\rangle - \mu\big)$$

waar $\langle E\rangle$ de verwachtingswaarde of het gemiddelde voorstelt:

$$\langle E\rangle = \int_\Omega E(\omega)\,dp_\omega\,.$$

De Euler–Lagrangevergelijkingen voor deze Lagrangiaan worden dan gegeven door:

$$\frac{\partial\tilde{L}}{\partial p} = 0 \qquad \text{en} \qquad \frac{\partial\tilde{L}}{\partial\beta} = 0\,.$$

Daar de nieuwe Lagrangiaan lineair is in de parameter β, die we ook wel een Lagrangevermenigvuldiger noemen, is de tweede vergelijking heel eenvoudig. Deze reduceert immers tot de voorwaarde

$$\langle E\rangle = \mu\,.$$

De Lagrangevermenigvuldigers zijn trouwens gerelateerd aan de (anti)spookdeeltjes uit de veldentheorie. De eerste vergelijking kunnen we nu ook uitwerken:

$$\begin{aligned}\frac{\delta\tilde{L}}{\delta p} &= \frac{\partial}{\partial p_\omega}\big(p_\omega \ln(p_\omega) + \beta E(\omega) p_\omega - \beta\mu\big) \\ &= \ln(p_\omega) + \beta E(\omega) - 1 \\ &= 0\,.\end{aligned}$$

Dit kunnen we dus nog herschrijven als

$$\ln(p_\omega) = 1 - \beta E(\omega)$$

of nog als

$$p_\omega = \exp\big(1 - \beta E(\omega)\big)\,.$$

De betekenis van de factor β komen we later nog op terug. Waar we nu echter wel al naar gaan kijken is een niet zo onbelangrijk detail dat we in deze afleiding over het hoofd hebben gezien.

De grootheid p hoort een kansverdeling voor te stellen. Toen we de variatie namen, hebben we echter geen rekening gehouden met deze voorwaarde. (Kansverdelingen vormen helemaal geen vectorruimte, dus het hele formalisme zoals we het in Hoofdstuk 4 hadden gepresenteerd, stond al op wankele fundamenten.) Er is dus geen zekerheid dat de voorwaarde

$$\int_\Omega dp_\omega = 1$$

behouden is gebleven. Om dit op te lossen, passen we een standaardtechniek uit te fysica toe: wat niet genormaliseerd is, normaliseren we handmatig:

$$\tilde{p}_\omega := \frac{p_\omega}{\displaystyle\int_\Omega dp_\omega} = \frac{\exp\big(-\beta E(\omega)\big)}{\displaystyle\int_\Omega \exp\big(-\beta E(\omega)\big)\, d\omega}\,.$$

Als we nu de partitiefunctie

$$Z := \int_\Omega \exp\big(-\beta E(\omega)\big)\, d\omega$$

definiëren, dan krijgen we de volgende compact uitdrukking

$$\tilde{p}_\omega \equiv \frac{\exp\bigl(-\beta E(\omega)\bigr)}{Z}$$

voor de gelauwerde Boltzmannverdeling (nog niet te verwarren met de Maxwell–Boltzmannverdeling die we later zien). Dit is een prachtig voorbeeld van wat in de kansrekening en statistiek gekend staat als een 'exponentiële familie' (waarvan de *exponentiële verdelingen* een ander voorbeeld vormen). In de *machine learning* zijn dergelijke verdeling ook enorm gegeerd door hun talloze aantrekkelijke eigenschappen.

In de klassieke fysica is de Boltzmannverdeling echter beter gekend onder een meer specifieke vorm, namelijk in termen van de snelheid van de deeltjes. De kinetische energie kan geschreven worden als

$$E(\vec{v}) = \frac{1}{2} m \vec{v} \cdot \vec{v}\,.$$

Elke microtoestand beschreven door een snelheidsvector $\vec{v}$ bepaalt dus een macrotoestand met energie $E(\vec{v})$, maar deze relatie is niet één-op-één. Om de verdeling van toestanden in termen van de grootte v te kennen, moeten we een verandering van variabelen doorvoeren. Laten we beginnen met de partitiefunctie:

$$Z := \int_\Omega \exp\bigl(-\beta E(\omega)\bigr)\, d\omega = \int_{\mathbb{R}^3} \exp\bigl(-\tfrac{1}{2}\beta m v^2\bigr)\, d^3v\,.$$

Om dit uit te rekenen, kunnen we in bolcoördinaten werken, aangezien de energie niet richtingsafhankelijk is:

$$Z = 4\pi \int_{\mathbb{R}^3} v^2 \exp\bigl(-\tfrac{1}{2}\beta m v^2\bigr)\, dv = \left(\frac{2\pi}{\beta m}\right)^{\frac{3}{2}}.$$

«*Eigenlijk zou je nu al voldoende moeten weten over integralen om dit te kunnen*

narekenen.» De Boltzmannverdeling voor de snelheid wordt dan:

$$\tilde{p}_v = \left(\frac{\beta m}{2\pi}\right)^{\frac{3}{2}} \exp\left(-\frac{1}{2}\beta m v^2\right).$$

Dit is de Maxwell–Boltzmannverdeling die we eerder hadden aangekondigd. (De veralgemening naar hogere dimensies is vrij voor de hand liggend.)

Laten we hiermee nu eens de gemiddelde energie $\langle E\rangle = \mu$ berekenen:

$$\langle E\rangle = \int_{\mathbb{R}^3} \frac{1}{2} m v^2 \, dp_v = \frac{3}{2\beta}.$$

De Lagrangevermenigvuldiger β is dus rechtstreeks gerelateerd aan de gemiddelde energie. We kunnen de betekenis van β echter nog duidelijker maken. Daarvoor dienen we nog de entropie te berekenen:

$$\begin{aligned} S(\tilde{p}) &= -k_B \int_{\mathbb{R}^3} \ln(\tilde{p}_v)\, d\tilde{p}_v \\ &= -k_B \int_{\mathbb{R}^3} \big(-\beta E(v)\big)\, d\tilde{p}_v + \int_{\mathbb{R}^3} \ln(Z)\, d\tilde{p}_v \\ &= k_B \beta \langle E\rangle + k_B \ln(Z), \end{aligned}$$

waar k_B de constante van Boltzmann genoemd wordt. (Deze conventiefactor geeft het verschil tussen de Shannon- en Gibbsentropie.) De lezers die wat thermodynamica kennen, in het bijzonder de eerste wet, zullen hierin (met wat geluk) misschien de volgende vergelijking herkennen:

$$E = ST + PV,$$

waar T de temperatuur voorstelt, P de druk en V het volume van het systeem. Als we

De tweede term in het rechterlid kunnen we associëren met de term $k_B \ln(Z)$. Dit wordt soms ook wel de vrije energie F genoemd. als we de

andere termen vergelijken, dan zien we dat

$$\beta = \frac{1}{k_B T} .$$

De Lagrangevermenigvuldiger die we al een hele tijd meeslepen, is (op een constante na) gelijk aan de inverse temperatuur! Als we deze relatie nu in de waarde van de gemiddelde energie invullen, krijgen we nog een belangrijkere relatie:

$$\langle E \rangle = \frac{3}{2} k_B T .$$

Dit equipartitietheorem stelt dat elke ruimtelijke dimensie, elke dimensie waarin de deeltjes zich kunnen bewegen, een bijdrage van $\frac{1}{2} k_B T$ leveren aan de gemiddelde energie. We kunnen dit resultaat natuurlijk ook omdraaien en stellen dat de energie van een systeem volledig bepaald wordt door de gemiddelde energie van de deeltjes!

Dit zorgt er ook voor dat het concept temperatuur in de moderne fysica een probleem kan vormen, want hoe definiëren we bijvoorbeeld de temperatuur van een systeem dat bestaat uit één deeltje? De gemiddelde energie houdt misschien wel nog steek, maar de verdeling is absoluut geen Maxwell-Boltzmannverdeling meer, de afleiding hierboven valt volledig uiteen!

Als je nu wat hoofdpijn hebt gekregen van alle formules en tussenstappen, dat is niet zo verwonderlijk. Statistische fysica is allesbehalve een eenvoudig onderwerp. Zoals de titel van het hoofdstuk al wat aangaf, rust er volgens sommigen zelfs een vloek op dit vakgebied. Aan het begin van de 20e eeuw pleegde Boltzmann zelfmoord na een lange depressie die deels te wijten was aan het gebrek aan geloof dat anderen in zijn werk hadden. Om het verhaal nog erger te maken was er voor zijn student Ehrenfest hetzelfde lot weggelegd. De natuurkunde had betere tijden gekend.

Laten we echter niet te lang bij die donkere momenten stilstaan. Ondertussen heeft Boltzmann zijn werk keer op keer zijn nut bewezen, zelfs

in zo een mate dat men het ook in de kwantummechanica toepast. Hiervoor starten we opnieuw van de partitiefunctie

$$Z = \int_\Omega \exp\bigl(-\beta E(\omega)\bigr)\, d\omega\,.$$

In de kwantummechanica is de Hamiltoniaan $\widehat{H}$ de tegenhanger van de energie, dus we moeten de exponentiële functie kunnen samenstellen met operatoren. Gelukkig is dit geen probleem, al zeker niet in het eindigdimensionale geval waar de Hamiltoniaan een matrix is, want de exponentiële functie is hier op dezelfde manier gedefinieerd als voor gewone getallen:

$$\exp(X) := \sum_{n=0}^{+\infty} \frac{X^n}{n!}\,.$$

Wat wel nog overblijft is de integraal en ook hier worden we wat door de wiskunde geholpen. Meetbare waarden worden gegeven door de eigenwaarden van observabelen en voor fysisch relevante observabelen is het aantal eigenwaarden (en hun bijhorende eigenvectoren) aftelbaar. We kunnen de integraal dus vervangen door een som over eigenvectoren:

$$Z = \int_\Omega \exp\bigl(-\beta E(\omega)\bigr)\, d\omega = \sum_{n=1}^{+\infty} \langle \psi_i | \exp(-\beta \widehat{H}) | \psi_i \rangle\,.$$

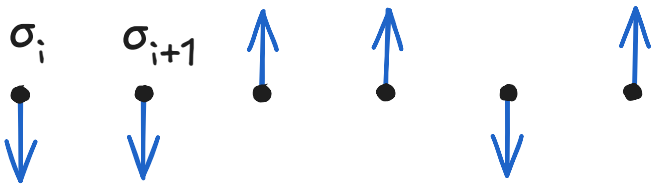
De som in de uitdrukking van de partitiefunctie wordt ook wel het spoor genoemd. Als we een gegeven matrixvoorstelling X van een operator hebben, dan wordt het spoor simpelweg gegeven door de som van de diagonaalelementen:

$$X = \begin{pmatrix} X_{11} & \cdots & X_{1n} \\ \vdots & \ddots & \vdots \\ X_{n1} & \cdots & X_{nn} \end{pmatrix} \qquad \longrightarrow \qquad \mathrm{tr}(X) := \sum_{i=1}^{n} X_{ii}\,.$$

De (kwantum)partitiefunctie kan dus beknopt als volgt genoteerd worden:

$$Z = \mathrm{tr}\Bigl(e^{-\beta \widehat{H}}\Bigr)\,.$$

Om dit hoofdstuk af te sluiten, zullen we echter terugkeren naar de klassieke fysica (de kwantummechanische tegenhanger wordt in de extra's verder besproken). Desondanks zullen we wel het discrete karakter van de kwantummechanica behouden door naar systemen te kijken met maar een eindig aantal mogelijke toestanden: kristalroosters. In het bijzonder zullen we magneten in 1D bestuderen. Net zoals hun kwan-



Figuur 7.1: Een keten van klassieke spins.

tum tegenhangers, kunnen klassieke spins maar twee waarden aannemen: $\sigma_i = \pm 1$. Het meest bestudeerde systeem van dergelijke spins is het Isingmodel, waarbij enkel naburige spins met elkaar interageren:

$$E_\omega = \sum_{i=-\infty}^{+\infty} -J\sigma_i\sigma_{i+1} - h\sigma_i\,,$$

waar $\omega \equiv \{\sigma_i\}$ een spinconfiguratie voorstelt. De eerste term zorgt ervoor dat naburige spins het liefst in dezelfde richting wijzen. De tweede term zorgt er dan weer voor dat de spins het liefst in dezelfde richting als het externe magnetische veld h wijzen.

Dit model is bijvoorbeeld al een goede eerste stap om ferromagneten, het soort magneten die je kent van op de koelkast, te beschrijven. Wanneer je een stuk ijzer neemt, werkt dit in het algemeen niet spontaan als een magneet. Lokaal zullen de elektronen wel in dezelfde richting wijzen (ze vormen zogenaamde domeinen), maar die domeinen zijn willekeurig verspreid. Wanneer we het stuk ijzer echter opwarmen zodat de elektronen wat los kunnen komen en het in een extern magnetisch veld brengen (hoe zwak het ook mag zijn), zullen alle elektronen zich spontaan aligne-

ren.

In het geval waarbij er geen extern veld aanwezig is ($h = 0$), bestaat er een eenvoudige beschrijving van dit systeem. Definieer als eerste de gekoppelde spintoestanden: $S_i := \sigma_i \sigma_{i+1}$. De energie kan dan als volgt herschreven worden:

$$E_\omega = \sum_{i=-\infty}^{+\infty} -JS_i \,.$$

Bijgevolg krijgen we ook voor de partitiefunctie een eenvoudigere uitdrukking:

$$Z = \sum_{\omega \in \Omega} \exp\left(-\beta \sum_{i=-\infty}^{+\infty} -JS_i \right) = 2 \prod_{i=-\infty}^{+\infty} \sum_{S_i = \pm 1} e^{\beta J S_i} \,.$$

In de tweede stap mogen we de volgorde van de twee bewerkingen omwisselen omdat we door het invoeren van de variabelen S_i onafhankelijke variabelen bekomen zijn. Daar S_i ook maar twee waarden kan aannemen, krijgen we

$$\sum_{S_i = \pm 1} e^{\beta J S_i} = e^{\beta J} + e^{-\beta J}$$

en dus, als we op een gesloten keten van lengte L werken (voor oneindige ketens moeten we iets subtieler te werk gaan, maar het resultaat is hetzelfde),

$$Z = 2\left(e^{\beta J} + e^{-\beta J}\right)^L \,.$$

Bovenstaande uitdrukkingen laten ons toe om heel eenvoudig eigenschappen van het systeem uit te rekenen.

Extra's: Open systemen

Aangezien de inhoud van dit hoofdstuk betrekkelijk lichte materie was, gaan we hier het formalisme hier nog wat uitbreiden, we betrekken er de

relativiteitstheorie en kwantumveldentheorie ook bij. In Deel 1 zagen we hoe de Hamiltoniaan zowel een rol speelde in de (symplectische) meetkundige beschrijving van de klassieke mechanica aan de hand van het Poissonhaakje als in de kwantummechanica aan de hand van operatoren. In beide gevallen was de Hamiltoniaan de generator van tijdsevolutie. Als we naar de totale transformatie kijken en niet de infinitesimale generator, dan bekomen we de evolutie-operator:

$$\widehat{U}(t) := \exp(-i\widehat{H}t)\,.$$

Als we deze uitdrukking nu naast de formule van de Boltzmannverdeling leggen, dan lijken die wel verdacht veel op elkaar (op de normalisatie na):

$$\exp(-i\widehat{H}t) \qquad \text{versus} \qquad \exp(-\beta\widehat{H})\,.$$

Om dit verband concreet te maken, zouden we echter tijd en temperatuur aan elkaar moeten relateren en dat klinkt toch wel echt te gek voor woorden. Toch is een dergelijke transformatie geen uitzondering. Bij de bespreking van kwantumveldentheorie in Deel 1 werd reeds kort de Wick-transformatie aangehaald.

Om deze beter te begrijpen maken we een omweg langs de relativiteitstheorie. Om relativistische problemen op te lossen werken we in de Minkowskiruimte M^4, en het enige verschil tussen deze ruimte en de Euclidische ruimte $\mathbb{R}^4$ is hun notie van afstand:

$$\|\vec{x}\|^2_{M^4} = -|t|^2 + |x|^2 + |y|^2 + |z|^2$$

en

$$\|\vec{x}\|^2_{\mathbb{R}^4} = |t|^2 + |x|^2 + |y|^2 + |z|^2\,.$$

Op het teken van de eerste term na zijn deze twee uitdrukkingen gelijk aan elkaar. Om van de tijdsdimensie in M^4 een ruimtelijke dimensie in $\mathbb{R}^4$ te maken, kunnen we de volgende transformatie uitvoeren, aangezien $i^2 = -1$:

$$t \longrightarrow it\,.$$

Door de tijd 'complexificeren' kunnen we tussen de twee formalismen wisselen en deze transformatie laat ons om heel wat problemen te vereenvoudigen.

Laten we de Wicktransformatie dus eens toepassen op de Boltzmann-verdeling. Als we in de klassieke verdeling de energie vervangen door de Hamiltoniaanse operator, dan bekomen we volgende operator:

$$\hat{\rho}_T := \frac{e^{-\beta\widehat{H}}}{\mathrm{tr}\left(e^{-\beta\widehat{H}}\right)}\,.$$

(Dat deze operator ook een kwantumtoestand kan beschrijven, bespreken we later wanneer we open systemen introduceren.) Door een Wicktransformatie uit te voeren, wordt de teller omgezet in de evolutie-operator $\widehat{U}(t)$, terwijl de noemer wordt omgezet in het spoor van de evolutie-operator. Het fascinerende is nu dat het spoor ook kan worden geschreven als een padintegraal.

Eerst schrijven we het spoor formeel als een gewone integraal:

$$\mathrm{tr}\left(e^{-\beta\widehat{H}}\right) = \int_\Omega \langle\omega|e^{-\beta\widehat{H}}|\omega\rangle\, d\omega\,.$$

Door de Wicktransformatie is de operator in de integraal de evolutie-operator voor een tijdstap $-i\beta$. De kans om van een toestand φ op tijd 0 naar de toestand ψ op tijd t te evolueren wordt gegeven door de volgende padintegraal:

$$\langle\psi|e^{-it\widehat{H}}|\varphi\rangle = \int_{\omega(0)=\varphi}^{\omega(t)=\psi} e^{iS[\omega]}\, D\omega\,.$$

We integreren over alle paden die voldoen aan de gegeven randvoorwaarden.

Om onze twee uitdrukkingen aan elkaar te relateren moeten we echter twee aanpassingen aanbrengen. Allereerst de Wicktransformatie:

$$t \longrightarrow -i\tau \qquad \text{en} \qquad dt \longrightarrow -id\tau\,.$$

Dit geeft

$$\begin{aligned} iS[\omega] &= i\int_{M^4} L[\omega]\, d^3x\, dt \\ &= \int_{\mathbb{R}^4} L[\omega]\, d^3x\, d\tau\,. \end{aligned}$$

Op dit moment dienen we een aanname te maken over de Lagrangiaan van onze theorie. Zoals we reeds eerder in de klassieke mechanica zagen, kan deze geschreven worden als de som van een kinetische term en een (tijdsonafhankelijke) potentiële term:

$$L[\omega] = \frac{1}{2}m\dot{\omega}^2 - V(\omega)\,.$$

De potentiële energie is volledig ongevoelig voor de Wicktransformatie aangezien deze enkel van de ruimtelijke dimensies afhangt. De kinetische term daarentegen is kwadratisch is de tijdsdimensie en zal dus een minteken bekomen onder een Wicktransformatie:

$$L[\omega] = \frac{1}{2}m\dot{\omega}^2 - V(\omega) \xrightarrow{t\longrightarrow -i\tau} -\left(\frac{1}{2}m\dot{\omega}^2 + V(\omega)\right) = -L_E[\omega]\,.$$

Voor de acties krijgen we dus:

$$iS[\omega] = -S_E[\omega]\,.$$

De Euclidische padintegraal bevat dus een minteken in plaats van de imaginaire eenheid.

Naast de Wicktransformatie, dienen we ook te eisen dat de begintoestand dezelfde is als de eindtoestand, aangezien in de formule van het spoor zowel links als rechts dezelfde toestand staat. In de Euclidische padintegraal moeten we ons dus beperken tot periodieke toestanden, waarbij de periode gegeven wordt door de (inverse) temperatuur:

$$\mathrm{tr}\left(e^{-\beta\widehat{H}}\right) = \int_{\omega(0)=\omega(\beta)} e^{-S_E[\omega]}\, D\omega\,.$$

We hebben hier de partitiefunctie van een systeem in thermisch evenwicht geschreven als de padintegraal van een kwantumsysteem dat 'leeft' op een cilinder $\mathbb{R}^3 \times S^1$ met straal omgekeerd evenredig aan de temperatuur. Kan je het zelf zo gek bedenken? Sommige wetenschappers konden dat zeker wel. Zo gebruikten sommigen dit verband bijvoorbeeld om de zogenaamde Hawkingstraling van een zwart gat af te leiden. (Hawking zelf gebruikte een andere aanpak.)

Vooraleer we verdergaan, vraag je je misschien nog af wat er gebeurt met acties die niet van de vorm $T(\dot{\omega}) - V(\omega)$ zijn. Typische voorbeelden zijn tijdsafhankelijke potentialen of hogere-orde interactietermen. Dergelijke termen zullen niet zo mooi transformeren als die in de actie zoals hierboven en geven aanleiding tot een imaginaire term in de Euclidische actie:

$$S_E[\omega] \sim S_E^{(1)}[\omega] - iS_E^{(2)}[\omega]\,.$$

Bijgevolg zal de padintegraal zowel een exponentiële factor als een oscillerende factor bevatten:

$$\mathrm{tr}\left(e^{-\beta\widehat{H}}\right) = \int_{\omega(0)=\omega(\beta)} e^{-S_E^{(1)}[\omega]} e^{iS_E^{(2)}[\omega]}\, D\omega\,.$$

We behouden dus een deel van het kwantummechanische karakter waarbij verschillende paden ω met elkaar kunnen interfereren. Dit geeft echter aanleiding tot het zogenaamde *tekenprobleem* in numerieke simulaties. Door het sterk oscillerende karakter van de fasefactor $e^{iS_E^{(2)}[\omega]}$ moeten de simulaties tot uiterst hoge precisie uitgevoerd worden opdat de paden niet allemaal destructief zouden interfereren.

Ondanks dit belangrijk probleem is de relatie tussen statistische mechanica en (kwantum)veldentheorie fundamenteel voor vele toepassingen. We halen er nog eens Hoofdstuk 2 bij. De Fouriertransformatie stelde ons in staat om functies te ontbinden in frequenties. In het bijzonder konden periodieke functies geschreven worden als sommen over discrete frequenties. Nu, hierboven zagen we hoe de statistische partitiefunctie kan herleid worden tot een padintegraal over periodieke paden. Deze twee

resultaten samen zorgen er dus voor dat de partitiefunctie kan uitgedrukt worden als een som over discrete frequenties, de Matsubarafrequenties $\frac{2\pi n}{\beta}$ met $n \in \mathbb{N}$. 'Thermische veldentheorie' is echter nog een relatief jong vakgebied waar mensen nog volop werken aan nieuwe technieken en problemen.

Voor het laatste deel van dit hoofdstuk zeggen we vaarwel tegen de veldentheorie en keren we terug naar de gewone kwantummechanica. In Deel 1 hadden we de volgende onderdelen ingevoerd:

Toestand	vector $\lvert\psi\rangle$
Observabele	operator (of matrix) $\widehat{O}$
Meting	duale vector $\langle\psi\rvert$
Kans	inproduct $\langle\varphi \,\vert\, \psi\rangle$
Verwachtingswaarde	$\langle\psi\rvert\widehat{O}\lvert\psi\rangle$

Het inproduct laat ons toe om van twee vectoren naar een getal te gaan. We starten van twee eerste-orde tensoren en bekomen een nulde-orde tensor. We kunnen echter ook de omgekeerde richting uitgaan door middel van het tensorproduct. We starten van twee eerste-orde tensoren en bekomen een tweede-orde tensor:

$$\lvert\psi\rangle\langle\varphi\rvert\,.$$

Zoals de notatie doet vermoeden, werkt deze operator in op vectoren door middel van het inproduct. We namen als eenvoudig voorbeeld de operator bepaald door een toestandsvector $\lvert\psi\rangle$:

$$\hat{\rho}_\psi := \lvert\psi\rangle\langle\psi\rvert\,.$$

Als eerste kunnen we eenvoudig zien dat het spoor van deze operator gelijk is aan 1:

$$\mathrm{tr}(\hat{\rho}_\psi) = \sum_i \langle e_i \,\vert\, \psi\rangle\langle\psi \,\vert\, e_i\rangle = \sum_i |c_i|^2 = 1$$

Dit is eenvoudigweg de normalisatievoorwaarde van toestandsvectoren.

Op een gelijkaardige manier kunnen we aantonen dat deze operator *positief-definiet* is:

$$\langle\varphi|\hat{\rho}_\psi|\varphi\rangle \geq 0$$

voor eender welke toestandsvector $|\varphi\rangle$. «*Probeer dit maar eens aan te tonen.*» Als derde eigenschap kunnen we ook aantonen dat de operator symmetrisch is. Op een complex toegevoegde na, kunnen we de input en output omwisselen zonder het resultaat te wijzigen:

$$\langle\varphi|\hat{\rho}_\psi|\xi\rangle = \overline{\langle\xi|\hat{\rho}_\psi|\varphi\rangle}\,.$$

Deze drie eigenschappen — normalisatie, positiviteit en symmetrie — definiëren een brede klasse aan operatoren, de dichtheidsoperatoren.

Deze operatoren, zoals hun naam doet vermoeden, zijn een soort kwantummechanische veralgemening van kansverdelingen of kans*dichtheden*. Ze stellen verdelingen van kwantumtoestanden voor aangezien ze in het algemeen als volgt kunnen geschreven worden voor een geschikte verzameling vectoren $|\psi_i\rangle$:

$$\hat{\rho} = \sum_i \lambda_i |\psi_i\rangle\langle\psi_i|\,.$$

Dit kan geïnterpreteerd worden als een stochastisch ensemble van 'pure' kwantumtoestanden. In tegenstelling tot de inherente stochasticiteit uit de kwantummechanica, stelt dit ensemble eerder een vorm van *epistemische onzekerheid* voor. Het is de onzekerheid die voortkomt uit ons gebrek aan informatie. Dit treedt bijvoorbeeld op wanneer twee kwantumsystemen interageren. Aangezien we in het algemeen niet elk detail van de systemen en de interactie kennen, moeten we die onwetendheid ook modelleren en daarvoor hebben we dichtheidsoperatoren nodig. We combineren dus kwantummechanica en statistiek in één formalisme.

Waar we voor pure toestanden het inproduct gebruiken om kansen en verwachtingswaarden te berekenen, dienen we hier volledig over te gaan op het spoor van operatoren. En dat spoor, dat waren we reeds tegenge-

komen in de Boltzmannverdeling:

$$\hat{\rho}_T := \frac{e^{-\beta\widehat{H}}}{\mathrm{tr}\left(e^{-\beta\widehat{H}}\right)}\,.$$

Dat we hier hetzelfde symbool als voor dichtheidsoperatoren gebruikt hadden, was geen toeval. De operator $\hat{\rho}_T$ is namelijk een dichtheidsoperator waarbij de coëfficiënten volledig door de Boltzmannverdeling (en dus de temperatuur) bepaald worden. Dit wordt ook wel een thermische toestand genoemd en is, zoals reeds hiervoor bij thermische veldentheorie besproken, een uiterst interessant object.

Om af te sluiten, bekijken we eens hoe dit formalisme ons nu echt toelaat om open systemen te bestuderen, waarbij interactie met de omgeving is toegelaten. Stel dat we vertrekken vanuit een pure toestand $\hat{\rho}_{AB}$ voor een gecombineerd systeem AB. In de praktijk hebben we echter vaak enkel toegang tot een van de twee deelsystemen. Een goed voorbeeld is het geval van twee verstrengelde deeltjes waarvan een zich op Aarde bevindt en het andere door het heelal vliegt. Wat is de correcte beschrijving van het deelsysteem A dat zich op Aarde bevindt? Ook al is $\hat{\rho}_{AB}$ puur, de toestand van A zal in het algemeen 'gemengd' zijn doordat we moeten uitmiddelen over de informatie over B:

$$\begin{aligned}\hat{\rho}_A = \mathrm{tr}_B(\hat{\rho}_{AB}) &= \sum_i \langle B_i \,|\, \psi\rangle_{AB}\langle\psi \,|\, B_i\rangle \\ &= \sum_{i,j} \big\langle B_i \big| B_j\big\rangle \big| A_j\big\rangle\big\langle A_j\big|\big\langle B_j \big| B_i\big\rangle \\ &= \sum_i |c_i|^2 |A_i\rangle\langle A_i|\,.\end{aligned}$$

Deze dichtheidsoperator die we bekomen door alle informatie over B weg te gooien, wordt ook wel de gereduceerde dichtheidsoperator van A genoemd. Indien de gereduceerde dichtheidsoperator uit meerdere termen bestaat en dus een echt ensemble voorstelt, dan spreken we over een verstrengelde toestand. We hebben informatie van B nodig om met zekerheid uitspraken over A te doen.

De studie naar dichtheidsoperatoren heeft ook geleid tot veralgemeningen van de Schrödingervergelijking. Om interacties met de omgeving te modelleren, zoals dissipatie, dient men extra termen in beschouwing te nemen. Dit geeft aanleiding tot de zogenaamde *meestervergelijkingen* zoals de Lindbladvergelijking:

$$\frac{\mathrm{d}\hat{\rho}}{\mathrm{d}t} = \mathcal{L}(\hat{\rho})\,.$$

We krijgen dus operatoren die inwerken op operatoren! Fysici hebben de neiging om dit dan superoperatoren te noemen, maar voor een getrainde wiskundige is er geen verschil.

Daarnaast heeft deze studie ook meer inzicht gegeven in de abstractie theorie van operatoren zoals geïnitieerd door von Neumann. De overgang van kwantumtoestanden naar algebra's van observabelen als primaire ingrediënten van de kwantummechanica is mede door bovenstaande inzichten gedreven. Als je bijvoorbeeld zwarte gaten bestudeerd, waar je als waarnemer (per definitie) enkel toegang hebt tot informatie buiten de horizon, kan je niet anders dan dichtheidsoperatoren gebruiken. Het is ondertussen een essentieel hulpmiddel bij het bestuderen van hedendaagse problemen, denk bijvoorbeeld aan de informatieparadox.

Hoofdstuk 8

Numeriek: Turnoefeningen

Een van de grootste problemen met kwantummechanica voor praktische doeleinden is de vloek van de dimensionaliteit. Als je N deeltjes hebt die elk beschreven worden door een d-dimensionale toestandsruimte, dan telt de toestandsruimte van het volledige systeem niet Nd maar d^N dimensies, de dimensie schaalt exponentieel. Met een gewone computer zullen we dus nooit in staat zijn om de toestandsruimte van gewone materie exact te kunnen beschrijven, aangezien het aantal deeltjes in een gebruikelijk stuk materie van de grootteorde 10^{23} is (denk maar aan de constante van Avogadro). We moeten dus een andere oplossing zoeken.

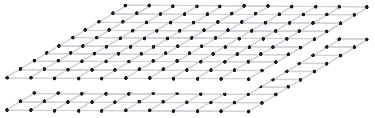
Om op te warmen zullen echter eerst nog wat klassieke fysica behandelen en hiervoor gaan we zelfs weer onze goede vriend Newton opzoeken. De tweede wet van Newton, de wet die de dynamica van klassieke objecten beschrijft, las als volgt:

$$\vec{F}(\vec{x}) = m\frac{\mathrm{d}^2\vec{x}}{\mathrm{d}t^2}\,.$$

Laten we nu voor de eenvoud aannemen dat we maar in één dimensie werken zodat we het vectorsymbool kunnen laten wegvallen. We krijgen een tweede-orde differentiaalvergelijking. Zoals we in Hoofdstuk 2 zagen

is zijn deze vergelijkingen totaal anders dan de algebraïsche vergelijkingen die je op school ziet. (Er is een manier om dergelijke vergelijkingen te interpreteren als algebraïsche vergelijkingen aan de hand van zogenaamde *straalbundels*, maar hier gaan we spijtig genoeg niet verder op ingaan.)

Aangezien dit hoofdstuk over numerieke methoden gaat, moeten we er uiteraard als eerste voor zorgen dat dergelijke differentiaalvergelijkingen op te lossen vallen door een computer. Ook al zijn er tegenwoordig methodes zoals *neural ODEs* & *neural operators*, die rechtstreeks in de ruimte van functies en lineaire operatoren proberen te werken, vanuit een fysisch standpunt is het veel nuttiger om de klassieke methoden eens te bekijken.



Figuur 8.1: Discretisatie van driedimensionale (Euclidische) ruimtetijd.

In het algemeen kunnen we zowel ruimte als tijd discretiseren, we vervangen dus de continue ruimtetijd door een roostersysteem, waarbij we voor de eenvoud punten op een gelijke afstand kiezen. Er geldt dus dat $x_{i+1} - x_i$ gelijk is voor alle roosterpunten $i \in \mathbb{Z}^n$ (en hetzelfde voor de tijd). Voor gewone differentiaalvergelijkingen zoals die uit de tweede wet van Newton, waar geen partiële afgeleiden in voorkomen, hoeven we zelfs enkel de tijd te discretiseren, de ruimtelijke coördinaten mogen we continu laten. Op deze manier kunnen we de tijdsafgeleiden nu ook benaderen door eindige quotiënten:[1]

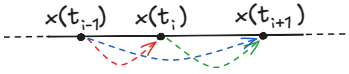
$$\frac{\mathrm{d}x}{\mathrm{d}t} \approx \frac{\Delta x}{\Delta t}\,.$$

De vraag is echter hoe we eindige verschillen Δt en Δx definiëren, of eerder kiezen. Aangezien we nu met eindige verschillen werken, moeten we

[1] De calculus die bekomen wordt door te vertrekken van een eindig quotiënt in de definitie van afgeleide wordt soms ook wel *kwantumcalculus* genoemd (ook al is de relatie met kwantummechanica niet vanzelfsprekend).

de punten in tijd kiezen waartussen we het verschil in positie willen berekenen.

Er zijn twee voor de hand liggende keuzes voor Δx: de (eerste-orde) voorwaartse en achterwaartse schema's. Deze zijn op de figuur hiernaast in respectievelijk groen en rood aangeduid. In het voorwaartse schema benaderen we de snelheid van een deeltje op basis van de afstand die hij tussen de huidige tijdstap en de volgende tijdstap zal afleggen:



Figuur 8.2: Eerste- en tweede-orde differentieschema's.

$$\frac{\mathrm{d}x}{\mathrm{d}t}(t_i) \approx \frac{x(t_{i+1}) - x(t_i)}{t_{i+1} - t_i}\,.$$

Voor het achterwaartse schema kijken we dan analoog naar de afgelegde afstand in de voorgaande tijdstap. Deze twee keuzes kunnen echter nogal asymmetrisch aanvoelen. Waarom zouden we enkel naar de toekomst of naar het verleden kijken? Het gemiddelde van deze twee geeft het gecentreerde (tweede-orde) differentieschema:

$$\frac{\mathrm{d}x}{\mathrm{d}t}(t_i) \approx \frac{1}{2}\left(\frac{x(t_{i+1}) - x(t_i)}{t_{i+1} - t_i} + \frac{x(t_i) - x(t_{i-1})}{t_i - t_{i-1}}\right).$$

Aangezien we equidistante punten in de tijd beschouwen, zijn beide noemers gelijk — laten we deze 'fundamentele' roosterconstante aanduiden met Δt — en krijgen we finaal

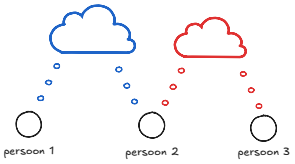
$$\frac{\mathrm{d}x}{\mathrm{d}t}(t_i) \approx \frac{x(t_{i+1}) - x(t_{i-1})}{2\Delta t}\,.$$

«*Schrijf $x(t_{i+1})$ eens uit in zijn Taylorbenadering (rond t_i) en verifieer dat het tweede-orde schema een betere benadering geeft.*»

@@ OOK WAT PDE's DOEN, INTEGRATIESCHEMAS (Euler), etc. @@

Nu we wat klassieke mechanica op een rooster gesmeten hebben, is het tijd om de kwantummechanische roostersystemen te bestuderen. In

de kwantummechanica zijn roostersystemen zelfs niet noodzakelijk benaderingen. Denk bijvoorbeeld aan kristalroosters in vaste stoffen zoals magneten. Om de eigenschappen die deze materialen hebben te bestuderen, kunnen we dus maar beter de kwantummechanica op dergelijke roosters doorgronden, maar aan het begin van dit hoofdstuk zagen we hoe dit zonder benaderingen fysisch (en niet alleen technologisch) onmogelijk is met 'gewone' computers.

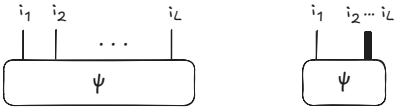


Figuur 8.3: Dichtstenabuurinteractie tussen mensen.

Om de complexiteit wat te beperken, zullen we ons ook hier beperken tot één ruimtelijke dimensie zoals in de figuur hiernaast afgebeeld. Of de roosterpunten bestaan uit elementaire deeltjes, atomen of volledige moleculen is voor ons van ondergeschikt belang. Het enige wat telt is dat ze beschreven worden door een kwantummechanische toestandsvector en dat interacties tussen roosterpunten lokaal zijn, dat wil zeggen dat twee punten op arbitraire afstand geen invloed kunnen hebben op elkaar. De analogie hiervoor, die hiernaast is afgebeeld, is een groep mensen die op een rij naast elkaar staan. Om een boodschap van de ene kant naar de andere kant door te geven, laten we niet toe dat mensen roepen en er moeten dus paarsgewijze interacties plaatsvinden. In het verdere verloop van dit hoofdstuk zullen we methodes bespreken die exact dit soort kwantumsystemen beschrijven.

Als de toestandsruimte van een roostersite gegeven wordt door $\mathcal{H}$, dan zal de toestandsruimte van $L \in \mathbb{N}$ sites gegeven wordt door het L-voudige tensorproduct $\mathcal{H}^L$. Zoals reeds eerder aangehaald schaalt de dimensie van deze ruimte exponentieel in L en krijgen we dus problemen als we dit numeriek willen benaderen. Een oplossing hiervoor bekomen we door eens wat beter na te denken over de de structuur van een algemene toestandsvector $|\psi\rangle$. Voor een eindig-dimensionale toestandsruimte $\mathcal{H}$ kunnen we $|\psi\rangle$ beschrijven door coëfficiënten $\psi_{i_1 \ldots i_L}$. Samen vormen deze een

'tensor' met L beentjes die we grafisch kunnen voorstellen zoals links in onderstaande afbeelding:



Die L beentjes kunnen we nu in twee groepen opdelen: $\{i_1\}$ en $\{i_2, \ldots, i_L\}$. Op die manier bekomen we een tensor met twee beentjes zoals rechts in de bovenstaande afbeelding.

In de lineaire algebra is een tensor met twee beentjes niets anders dan een matrix, waardoor we veel makkelijker aan de slag kunnen met dit object dan de oorspronkelijke tensor. Onze goede vrienden uit de ingenieurswetenschappen schieten ons te hulp en leren ons dat elke matrix kan geschreven worden aan de hand van twee unitaire transformaties:

$$A = XDY\,.$$

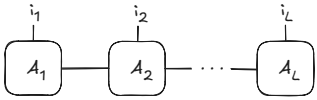
Hierin zijn X en Y unitair en is D een matrix met enkel elementen op de diagonaal:

$$D = \begin{pmatrix} \lambda_1 & 0 & 0 & \cdots \\ 0 & \lambda_2 & 0 & \cdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 \end{pmatrix}.$$

Deze matrix in de singulierewaardendecompositie veralgemeent de notie van eigenwaarde. Door het aantal waarden λ_i verschillend van nul te beperken, kunnen we de oorspronkelijke matrix A benaderen door matrices met minder informatie, wat de computationele complexiteit doet afnemen. Deze methode voor de beperking van het aantal vrijheidsgraden is ook uiterst populair in de datawetenschappen, waar het statistici toelaat om enkel te focussen op de belangrijkste variabelen in een probleem.

Door D en Y samen te nemen, hebben we de matrix A herschreven als een product van twee nieuwe matrices, waarvan we dimensies deels onder

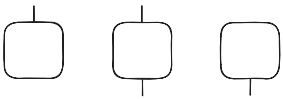
controle hebben. Dit proces kunnen we nu herhalen om DY te ontbinden als een product van twee matrices, enzovoort. Diagrammatisch krijgen we een uitdrukking van de volgende vorm: De bindingsdimensie $n \in \mathbb{N}$, het



Figuur 8.4: Matrixproducttoestand voor L sites.

aantal waarden in de matrices D die we niet gelijkstelden aan nul, bepaalt het aantal vrijheidsgraden op de virtuele beentjes tussen de tensoren A_i. Als we verder de (fysische) dimensie van $\mathcal{H}$ noteren als $d \in \mathbb{N}$, dan is de totale dimensie van de ruimte van dergelijke toestanden gelijk aan $Ld + (L-1)n$, wat slechts lineair schaalt in L en niet meer exponentieel!

Toestanden van de bovenstaande vorm staan gekend als matrixproducttoestanden (MPS) aangezien ze bekomen worden door matrices met elkaar te vermenigvuldigen. Deze toestanden liggen aan de basis van enkele van de meest succesvolle algoritmen in de studie van de kwantummechanische eigenschappen van materialen. We zullen hieronder enkele interessante eigenschappen bespreken.

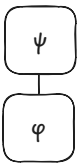


Figuur 8.5: Diagrammatische notatie.

Als eerste dienen we misschien wel eerst een woordje uitleg te geven over de diagrammatische notaties die we tot nu toe gebruikt hebben. Dit zijn niet zo maar analogieën van formules en dragen wel degelijk wiskundige betekenis. Zo stelt Figuur 8.4 echt wel een product van matrices of tensoren voor en is de naam MPS verdiend. Een vector noteren we als een tensor met één beentje, zoals hiernaast afgebeeld, links in de figuur. Analoog zou een tensor van orde n gegeven worden door een tensor met n beentjes. De tensor in het midden van de figuur hiernaast stelt bijvoorbeeld een matrix voor, er is één inko-

mend beentje en één uitgaand beentje. Een duale vector wordt simpelweg bekomen door een object om te draaien, net zoals rechts in Figuur 8.5.

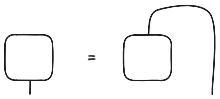
We kunnen ook rekenen met deze objecten. Neem het beentje van een tensor en verbindt het met het beentje van een andere. Een dergelijke contractie stelt een sommatie voor:



$$= \quad \sum_{i=1}^{n} \overline{\psi}_i \varphi_i \quad = \quad \langle \psi \,|\, \varphi \rangle \,.$$

Voor vectoren krijgen we dus bijvoorbeeld het inproduct. Figuur 8.4 stelt dus de contractie van de virtuele beentjes van de tensoren A_i voor en, indien we de fysische beentjes vasthouden, krijgen we de contractie (of het product) van matrices zoals beloofd.

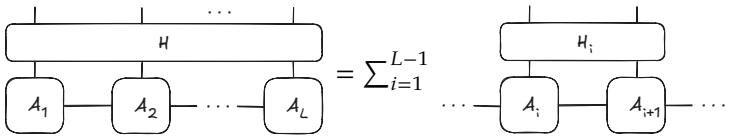
We kunnen zo goed als alle formules uit de lineaire algebra vertalen in de vorm van diagrammen. Dat dit mogelijk is en bovendien consistent is met intuïtieve manipulaties zoals in de figuur hiernaast was een belangrijk resultaat in de categorietheorie. Dit resultaat is ondertussen zelfs uitgebreid naar heel wat andere contexten en vakgebieden, zoals de kansrekening, zodat ook daar een diagrammatische calculus mogelijk is. Dit helpt zowel om het (theoretische) rekenwerk te versnellen als om argumenten makkelijker over te brengen aan andere wetenschappers.



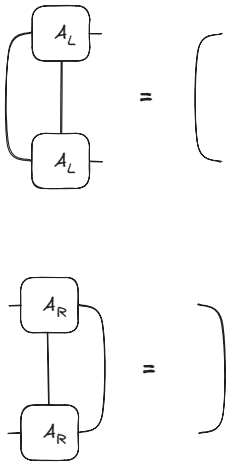
Figuur 8.6: Coherentie van diagrammen.

In het verdere verloop van onze bespreking zullen we aannemen dat het systeem invariant is onder translaties. Dit zorgt ervoor dat we zonder verlies van algemeenheid op elke site dezelfde tensor A mogen plaatsen. Van zo een toestand zullen we nu eens de energie berekenen. We nemen verder ook aan dat de Hamiltoniaan $\widehat{H}$ lokaal is, dat wil zeggen dat hij opgebouwd is uit termen die op een eindig aantal aangrenzende sites inwerkt. (In de praktijk is dit meestal zelfs maar twee.) Diagrammatisch

ziet dit er dan als volgt uit:



Om de verwachtingswaarde van de energie te berekenen, dienen we nog eens de toestand $|\psi[A]\rangle$ langs boven te contraheren. Jammer genoeg is deze volledig uitdrukking behoorlijk complex. Er zijn $2L$ verticale contracties en $2(L-1)$ horizontale contracties.

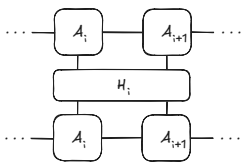


Figuur 8.7: Canonische vorm van een MPS.

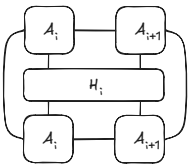
Gelukkig is er een heel belangrijke algebraïsche eigenschap waar we gebruik kunnen van maken. Om de MPS-vorm te bekomen, hebben we gebruik gemaakt van de singulierewaardendecompositie die een matrix herschrijft als het product van twee (of drie) andere matrices. Er bestaan echter nog heel wat andere matrixdecomposities — van de details blijven jullie gespaard — en een ervan laat ons toe om de tensoren A in een 'canonische' vorm te schrijven. Deze canonische vormen voldoen aan de eigenschap die in de figuur hiernaast voorgesteld wordt en zorgen ervoor dat contracties van tensoren waar geen operator op inwerkt kunnen weggelaten worden.

Als we deze uitdrukkingen in de verwachtingswaarde van de energie invullen, krijgen we de volgende uitdrukking:

$$\langle\psi[A]|\widehat{H}|\psi[A]\rangle = \sum_{i=1}^{L-1}$$



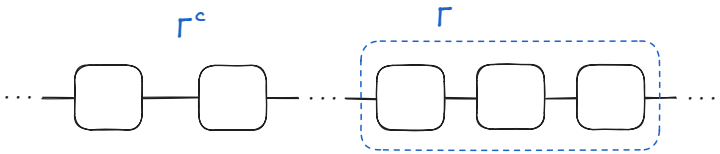
$$= \sum_{i=1}^{L-1}$$



.

De uitdrukking valt uiteen in termen die bestaan uit contracties over twee sites. We kunnen de energie efficiënt berekenen aangezien moderne computeralgoritmen geen enkel probleem hebben met dergelijke beperkte contracties, en we kunnen de grondtoestand van deze Hamiltoniaan, dat is de eigenvector met de laagste energie, bepalen door iteratief de coëfficiënten in A aan te passen om de verwachtingswaarde te minimaliseren.

Naast een uiterst geschikt startpunt voor variationele problemen, bieden MPS'en ook nog inzicht in andere theoretische vraagstukken. Het concept van verstrengeling is een van de basisbouwstenen van de huidige natuurkunde, met toepassingen gaande van materiaalfysica tot zwarte gaten, en hierin spelen MPS'en (en meer algemeen tensornetwerken) een bijzondere rol. Zoals in het volgende hoofdstuk meer in detail zal besproken worden, wordt verstrengeling altijd tussen twee (of meer) deelsystemen gedefinieerd. We moeten dus onze keten in twee stukken (Γ en Γ^c) opdelen en kijken hoe de vrijheidsgraden met elkaar gecorreleerd zijn.

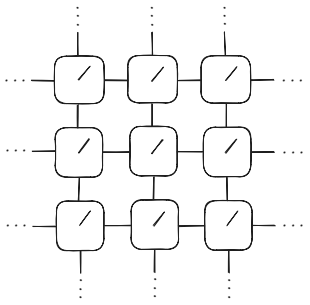


Hoe groot we Γ ook maken, de grens tussen Γ en zijn complement Γ^c zal echter altijd even groot zijn. Aangezien de sites in een MPS enkel met de dichtste naburen communiceren, impliceert dit dat de verstrengeling tussen Γ en Γ^c niet afhangt van de grootte van Γ. Dergelijke toestanden voldoen aan wat we een oppervlaktewet noemen, aangezien de verstrengeling schaalt volgens de grootte van de rand van Γ (en in 1D is die rand constant).

Het gevolg is nu ook dat MPS'en enkel fysische toestanden kunnen beschrijven die aan zo een oppervlaktewet voldoen. Een voorwaarde die misschien wel wat restrictief. Wat blijkt echter, als de Hamiltoniaan lokaal is en er een energie-opening bestaat, dat wil zeggen dat het verschil tussen de laagste twee energie-eigenwaarden strikt positief blijft ongeacht de

grootte van het systeem, dan zal de grondtoestand noodzakelijk aan een oppervlaktewet voldoen. Eender welke grondtoestand van zo een Hamiltoniaan kan tot op arbitraire nauwkeurigheid benaderd worden door een matrixproducttoestand! En indien je je zou afvragen of Hamiltonianen met deze eigenschappen fysisch relevant zijn, bijna alle interessante systemen vallen onder deze categorie.

De benadering of veronderstelling dat niet alle vrijheidsgraden nodig zijn om onze berekeningen tot op hoge nauwkeurigheid uit te voeren, is trouwens niet uniek voor de natuurkunde. In bijvoorbeeld de statistiek en datawetenschappen worden ook heel vaak dimensiereductiemethodes gebruikt om de essentiële informatie beter te weerspiegelen. Een gangbare hypothese in de datawetenschappen is dat zo goed als alle problemen afhangen van een laag- of lager-dimensionale deelruimte: de 'variëteithypothese'.

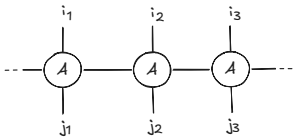


Figuur 8.8: 2D PEPS.

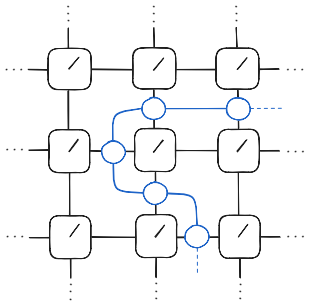
Indien we overgaan van 1D naar 2D of meer, ligt de situatie wel wat anders. De rechtstreekse veralgemening van MPS'en zijn de PEPS: *projected-entangled pair states*. Hiernaast kan je een PEPS op een vierkant rooster zien. We hebben opnieuw op elke roostersite een tensor met één fysiek beentje, maar nu lopen er (gecontraheerde) virtuele beentjes in zowel de horizontale als verticale richting. Net zoals in 1D voldoen deze toestanden aan een oppervlaktewet en kunnen ze de grondtoestand van heel wat Hamiltonianen uiterst nauwkeurig benaderen. Wat wel anders is, is de efficiëntie waarmee we alles kunnen uitrekenen. Er zijn heel wat meer geometrische configuraties mogelijk, zo zou je bijvoorbeeld ook een driehoekig rooster kunnen beschouwen, en de contractievolgordes spelen hier een enorm belangrijke rol in hoe efficiënt de algoritmen zijn. Contraheren we eerst alle horizontale beentjes? Of eerst de verticale? Of wisselen we

af? Dit probleem behoort tot de klasse van '#P-complete' problemen. Dit houdt in dat het minstens zo moeilijk is om een oplossing te vinden als eender welk 'NP-compleet' probleem, die problemen die eenvoudig zijn om te verifiëren maar uiterst moeilijk zijn om op te lossen. (Je hebt misschien wel al eens van '$P = NP$' gehoord, een van de Milleniumproblemen.)

De notie van MPS kan niet enkel uitgebreid worden naar hogere dimensies, we kunnen naast vectoren ook operatoren beschrijven. In dit geval spreken we over matrixproductoperatoren (MPO). Operatoren zoals de Hamiltoniaan kunnen op deze manier ook in de taal van tensornetwerken bestudeerd worden. Zo wordt de actie van de Hamiltoniaan $\widehat{H}$ op een toestandsvector $|\psi[A]\rangle$ simpelweg gegeven door de Hamiltoniaan in MPO-vorm te contraheren met de MPS met tensoren A, we verbinden de onderste beentjes van $\widehat{H}$ met de fysische beentjes van A.



Figuur 8.9: Matrixproductoperator (MPO).



Figuur 8.10: Virtuele MPO.

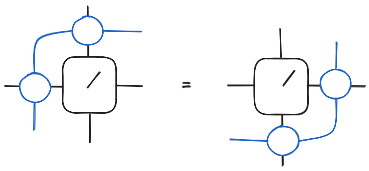
We kunnen echter ook iets exotische proberen. Beschouw een 2D PEPS zoals in Figuur 8.8. Matrixproductoperatoren kunnen eender welke soort operator voorstellen, niet enkel fysische zoals de Hamiltoniaan. We kunnen bijvoorbeeld een operator op virtueel niveau in een PEPS introduceren. In de afbeelding hiernaast is zo een MPO in het blauw aangeduid. Deze hoeft geen rechte lijn te vormen en mag doorheen het tensornetwerk slingeren. Dergelijke operatoren kunnen we relateren aan de defectoperatoren uit Hoofdstuk 5 die symmetrieën implementeerden in een (kwantum)veldentheorie. Indien de fysische toestand die overeenkomt

met een tensornetwerk onveranderd blijft na toevoeging van een virtuele MPO, dan hebben we een symmetrie gevonden. Omgekeerd kunnen we elke symmetrie (die lokaal op sites inwerkt) ook vertalen naar een virtuele MPO.

Om dit hoofdstuk af te sluiten gebruiken we de virtuele MPO's nog even om de notie van fasen te veralgemenen. In het middelbaar leer je dat er drie fasen van de materie zijn: vast, vloeibaar en gasvormig. Met wat geluk spreken ze ook over plasma. Er is echter al lang geweten dat dit een heel naïeve en rudimentaire opdeling is. De natuur is heel wat subtieler en verfijnder dan wat we soms geleerd hebben tijdens de lessen fysica. Denk bijvoorbeeld aan supergeleiding, of aan hoe materialen soms wel en soms niet magnetisch zijn.

Reeds in de eerste helft van de twintigste eeuw beseften natuurkundigen dat de gekende fasen van de materie konden worden opgedeeld op basis van hun symmetrie [6]. Zo heeft een kristalrooster beduidend minder symmetrie dan een gas aangezien de atomen in een kristal vastzitten in een stug patroon. Het Landau(–Ginzburg)paradigma werkt uiterst goed voor de klassieke fasen, maar voor de meer exotische fasen die sinds de jaren '80 werden ontdekt, volstond het schijnbaar niet. Er ontbrak een puzzelstukje.

Er bestaat een klasse van fasen onder de noemer topologische orde. (Het woordje 'topologisch' is bijzonder populair in de theoretische fysica.) Dit zijn toestanden die niet worden gekarakteriseerd door de lokale patronen en symmetrieën in een materiaal, maar door globale eigenschappen (vandaar de naam). Wat blijkt nu, dergelijke exotische systemen komen overeen met PEPS-



Figuur 8.11: De *pulling-through*voorwaarde.

tensoren die aan de zogenaamde *pulling-through*voorwaarde[2] voldoen die onderaan de voorgaande pagina is afgebeeld. De MPO die overeenkomt met de symmetrie kan 'doorheen' een PEPS-tensor getrokken worden. Waar topologische orde ook theoretisch kan beschreven worden door de (hogere) fusiecategorieën uit Hoofdstuk 5, biedt de tensornetwerkbeschrijving ook een enorm computationeel voordeel. We kunnen de toestanden niet enkel meer classificeren, we kunnen ze nu ook simuleren.

De taal van tensornetwerken heeft ons enkele grote voordelen geboden. Enerzijds staat het ons toe om via diagrammen te rekenen, net zoals de Feynmandiagrammen in de hoge-energiefysica. Zo kunnen complexe problemen eenvoudig worden uitgelegd aan collega's en theoretische berekeningen vaak ook overzichtelijker en met minder fouten worden uitgewerkt. Daarnaast zijn ze ook sterk gerelateerd aan de algebraïsche taal van symmetrieën, waardoor ze ons toelaten om ook problemen uit dat deelgebied te onderzoeken. (Dan zwijgen we nog over hoe tensornetwerken ook een vezelbundel vormen wiens raakvectorvelden overeenkomen met *excitaties* van de grondtoestand. Zo veel interessante eigenschappen, zo weinig tijd.) Anderzijds kunnen moderne computers uiterst efficiënt met vectoren, matrices en tensoren rekenen. Fysici kunnen tegenwoordig optimaal gebruik maken van de moderne hardware zoals GPU's om problemen in de materiaalfysica, zoals supergeleiding, te kraken.

Extra's

Hier zullen we een drietal specifieke algoritmen bespreken die zowel in de praktijk heel nuttig zijn als enkele interessante connecties met de wiskunde hebben. Hiervoor zullen we eens langsgaan bij zowel de klassieke fysica, kwantummechanica als statistiek (of statistische mechanica uit het volgende hoofdstuk).

We beginnen met een alternatief integratieschema voor de bewegings-

[2] Een geschikte Nederlandstalige term ontbreekt nog.

vergelijkingen van Newton. Eerder in dit hoofdstuk hadden we onder meer het Eulerschema gezien dat het eenvoudigste algoritme geeft om de bewegingsvergelijkingen op te lossen:

$$v(t_{i+1}) = v(t_i) + a(t_i)\,\Delta t \qquad \text{en} \qquad x(t_{i+1}) = x(t_i) + v(t_i)\,\Delta t\,.$$

Afgezien van het feit dat dit schema niet zo nauwkeurig is, de fouten stapelen zich nogal snel op, heeft het mogelijks nog een groter probleem, een dat veel nauwkeurigere algoritmen (zoals de hogere-orde *Runge–Kutta-methodes*) ook vaak hebben. De energie (en andere behouden grootheden) blijven niet behouden gedurende de numerieke simulaties.

Daar behouden grootheden een van de basisbouwstenen van de natuurkunde zijn, is dit een bijzonder belangrijk probleem. We moeten op zoeken naar alternatieve methodes die we symplectische integratoren zullen noemen. (De naam wijst erop dat ze de symplectische structuur van de klassieke mechanica behouden.) Het eenvoudigste voorbeeld is de *leap-frog*integrator[3], een methode die toevallig ook heel intuïtief is. Bovendien is hij ook nog eens nauwkeuriger dan de klassieke Eulerschema, het is een tweede-orde methode.

We beginnen met de definitie. Waar de Eulermethode zowel de positie als snelheid op hetzelfde tijdstip evalueert of updatet, gaan we bij de *leapfrog*integrator te werk zoals bij haasje over, we wisselen de updates af. Gegeven de positie $x(t_i)$ op tijdstip t_i en de snelheid $v(t_{i-1/2})$ een halve tijdstap eerder, krijgen we de volgende twee regels:

$$v(t_{i+1/2}) = v(t_{i-1/2}) + a(t_i)\,\Delta t \qquad \text{en} \qquad x(t_{i+1}) = x(t_i) + v(t_{i+1/2})\,\Delta t\,.$$

@@ afbeelding invoegen @@ We hebben dus evenveel (mogelijks kostelijke) evaluaties van de versnelling a nodig als bij de Eulermethode, maar we krijgen wel een veel nauwkeurigere benadering van het eindresultaat.

Het symplectische aspect van deze integrator verdient echter wel een extra woordje uitleg, voornamelijk omdat het verleidelijk is om deze eigenschap verkeerd te interpreteren. Vaak wordt gezegd dat dit neerkomt

[3] Engels voor 'haasje over'.

op het behoud van energie. Dit is jammer genoeg niet correct. We werken nog steeds met eindige benaderingen, dus we kunnen niet zo maar verwachten dat de behoudswetten exact blijven gelden. Het is niet de symplectische structuur geïnduceerd door de (klassieke) Hamiltoniaan $H(q,p)$ via het Poissonhaakje die behouden blijft, maar wat dan wel?

Voor kleine tijdstappen Δt verwachten we dat de resultaten van het integratieschema dicht bij de echte waarden zullen liggen, we kunnen dus proberen om storingsrekening toe te passen. We beschouwen een 'schaduw-Hamiltoniaan' van de vorm

$$H'(q,p) = H(q,p) + \Delta t\, H_1(q,p) + (\Delta t)^2 H_2(q,p) + \cdots .$$

Hoe kleiner de tijdstap, hoe dichter deze functie bij de echte Hamiltoniaan H zal aansluiten. Eigenlijk kunnen we hier al de eerste-orde term (en alle hogere termen van oneven orde) laten vallen. Voor tijd-symmetrische schema's, zoals de *leapfrog*integrator, zijn enkel de even ordes van belang.

Het blijkt nu dat we voor eender welke tijdstap Δt (zolang die klein genoeg is om alle dynamica in het systeem te beschrijven) en voor eender welke order van benadering, geschikte termen kunnen vinden zodat de (discrete) tijdsevolutie van de *leapfrog*integrator samenvalt met de (continue) tijdsevolutie van de verstoorde Hamiltoniaan H'. Bovendien, uit de bovenstaande storingsreeks volgt ook dat het verschilt tussen de ware energie $H(q,p)$ en die van onze benaderende oplossing begrensd is. Wat de juiste keuze is van de hogere-orde termen zou ons wat te ver leiden, maar laat het volstaan om te zeggen dat heel wat mensen hier heel veel tijd in gestoken hebben, en dat het vinden van geschikte en optimale integratieschema's voor een gegeven probleem nog steeds heel belangrijk is.

Nu gaan we even over naar de statistiek. In de fysica starten we vaak met een hoop datapunten en proberen aan de hand daarvan een fundamentele beschrijving van het achterliggende proces te vinden. In de statistiek bestaat er een gelijkaardig probleem waarbij we, gegeven datapunten, een kansverdeling proberen te vinden die de data beschrijft. Er bestaat echter ook een omgekeerd probleem: uit een gegeven kansverdeling data

genereren. De vraag die we nu zullen behandelen, is hoe men dat in de praktijk vaak doet.

Voor eenvoudige kansverdeling, zoals de uniforme verdeling of normale verdeling uit Hoofdstuk 2, bestaan er eenvoudige algoritmen die zonder al te veel berekeningen een steekproef genereren. Voor meer complexe verdelingen is dit echter niet altijd zo eenvoudig. De meest bekende klasse van methodes zijn de Monte Carlo-algoritmen die in de eerste helft van de 20e eeuw werden geïntroduceerd door enkele grote namen in de nucleaire fysica, waaronder Fermi.[4] Het algemene idee achter Monte Carlo-algoritmen is om te starten met een (naïeve) initiële gok en hierna iteratief deze waarde te laten evolueren zodat de opeenvolgende datapunten uit de gewenste verdeling getrokken lijken te worden.

Uiteraard kunnen we dit niet zomaar op een deterministische manier doen, anders kunnen we onmogelijk een kansverdeling simuleren. We zullen dus in elke stap ook wat willekeur moeten introduceren. De techniek die we hier zullen bespreken wordt Hamiltoniaanse Monte Carlo (HMC) genoemd omdat het de oorspronkelijke ideeën uit Monte Carlotheorie combineert met ideeën uit de Hamiltoniaanse mechanica. Eerst moeten we echter wat meer (basis)statistiek achter de kiezen hebben.

Stel je voor dat je twee grootheden X en Y meet die onafhankelijk van elkaar gegenereerd worden, zoals de hoeveelheid water die je vandaag gedronken hebt en het aantal baby's dat vandaag geboren is. Deze zijn in principe totaal ongerelateerd aan elkaar en dus zal de kansverdeling die beide grootheden beschrijft een eenvoudige vorm aannemen. Als X door de kansdichtheid f_X beschreven wordt en Y door f_Y, dan wordt het koppel (X, Y) beschreven door het product van de dichtheden:

$$f_{X,Y}(x, y) = f_X(x) f_Y(y) .$$

In ons geval zal f_X bij de verdeling horen waaruit we graag een steek-

[4] Monte Carlo verwijst hier wel degelijk naar het bekende casino in Monaco, maar niet door de link tussen kansrekening en gokken. De oom van Ulam, een van de grondleggers van deze methoden, gokte vaak in dit casino.

proef willen simuleren en f_Y zal een kansdichteid zijn waaruit we heel eenvoudig waarden kunnen genereren (een normale verdeling om precies te zijn):

$$f_{X,Y}(x,y) = f_X(x)\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{y^2}{2\sigma^2}\right).$$

Kansdichteden zijn altijd positief, dus we kunnen er een logaritme van nemen. Dit laat ons toe om bovenstaande formule als volgt te herschrijven:

$$f_{X,X}(x,y) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\ln f_X(x) - \frac{y^2}{2\sigma^2}\right).$$

Laat ons nu eens goed naar de functie in de exponent kijken:

$$\ln f_X(x) - \frac{y^2}{2\sigma^2}.$$

Als we de, op het eerste gezicht nogal uit de lucht gegrepen, substituties

$$V(x) := -\ln f_X(x) \qquad \text{en} \qquad m := \sigma^{-2}$$

doorvoeren, krijgen we de uitdrukking van de Hamiltoniaan van een eenvoudig klassiek systeem waarbij x de positie is en y de impuls:

$$H(x,y) := V(x) + \frac{y^2}{2m}.$$

Daar zien we al de relatie die zijn naam leent aan de methode! Als we x en y nu via de klassieke bewegingsvergelijkingen evolueren, dan krijgen we een nieuw koppel van waarden waarbij x beschreven wordt door de potentiaal V of, met andere woorden, waarbij x uit de kansdichtheid f_X getrokken is.

Met deze kennis kunnen we het naïeve HMC-algoritme als volgt samenvatten:

1. Kies een initiële waarde voor x.

2. Genereer een waarde voor y uit een normale verdeling met gemiddelde 0 en *variantie* (een maat voor de spreiding van de data) σ^2.

3. Los de bewegingsvergelijkingen voor de Hamiltoniaan $H(x, y)$ op voor enkele tijdstappen om een tweede koppel punten (x', y') te bekomen. Dit is jouw nieuwe datapunt.

4. Start opnieuw bij punt 2.

Indien we deze procedure zouden toepassen, zouden we echter geen al te goede resultaten krijgen. Ten eerste, wat kiezen we voor de variantie σ^2 of voor het aantal tijdstappen? Dit is een onderzoeksgebied op zich en hangt uiterst sterk af van de dichtheid f_X. Hoe grilliger de vorm van deze functie, hoe nauwkeuriger we dienen te simuleren. Bovendien is het mogelijk dat de potentiaal V die door f_X wordt geïnduceerd aanleiding geeft tot periodieke banen. In dit geval mogen we niet te lang in de tijd evolueren, want dan bestaat de kans dat we eigenlijk terug op ons beginpunt uitkomen.

Bovendien is het algoritme zelf ook nog niet volledig of optimaal. Men kan aantonen dat de uiteindelijke verdeling van de data, indien we het bovenstaande schema hanteren, niet correct zal zijn. In fysische termen dienen we ons systeem aan een *warmtebad* te koppelen zodat het deeltje beschreven door x en y af en toe een duw krijgt. In statistische termen betekent dit dat we de voorgestelde waarde (x', y') niet altijd gaan accepteren. Voor elk voorstel zullen we een (oneerlijke) munt opwerpen en we zullen het voorstel aanvaarden met de volgende kans:

$$\min\left(1, \frac{\exp\big(H(x', y')\big)}{\exp\big(H(x, y)\big)}\right).$$

Indien de kans om de voorgestelde waarde waar te nemen groter is dan de huidige, dan accepteren we zonder meer het voorstel. In het andere geval bestaat er echter een kans dat we bij de huidige waarde blijven. Dit zorgt ervoor dat heel uitzonderlijke waarden niet te vaak zullen voorkomen.

Als laatste zitten we nog met het oplossen van de bewegingsvergelijkingen. Om de kansdichtheid f_X goed te simuleren, is het uiterst belangrijk dat de Hamiltoniaan H correct gevolgd wordt, aangezien deze rechtstreeks van f_X afhangt. Gelukkig hebben we aan het begin van deze sectie een relatief efficiënt algoritme gezien dat de Hamiltoniaan (bijna) exact behoudt: de *leapfrog*-integrator.

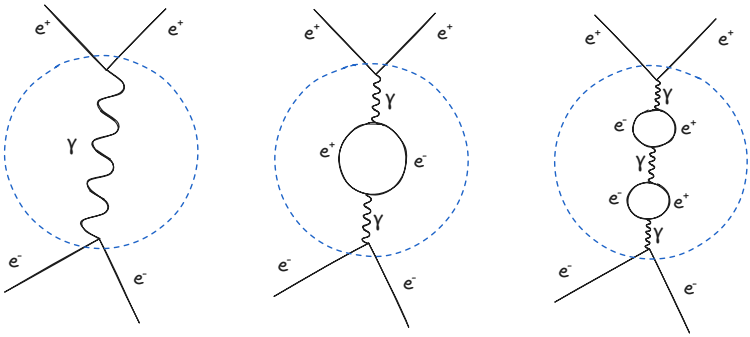
De uiteindelijke HMC-procedure wordt dus de volgende:

1. Kies een initiële waarde voor x.
2. Bepaal de tijdstap Δt en het aantal tijdstappen $k \in \mathbb{N}$ door jaren aan ervaring.
3. Bepaal de variantie σ^2 als de inverse van de variantie van f_X. Dit zorgt ervoor dat de totale spreiding van (x, y) genormaliseerd wordt. (Wat gouden raad uit de statistiek.)
4. Genereer een waarde voor y uit een normale verdeling met gemiddelde 0 en *variantie* (een maat voor de spreiding van de data) σ^2.
5. Los de bewegingsvergelijkingen voor de Hamiltoniaan $H(x, y)$ op met de *leapfrog*-integrator voor k tijdstappen Δt om een tweede koppel punten $\big(x(k\Delta t), y(k\Delta t)\big)$ te bekomen.
6. Accepteer dit voorstel met kans
$$\min\left(1, \frac{\exp\big(H(x', y')\big)}{\exp\big(H(x, y)\big)}\right).$$
7. Start opnieuw bij punt 4.

Als laatste keren we terug naar de kwantummechanica, in het bijzonder naar de tensornetwerken en hoe ze ons toestaan om grondtoestanden te vinden. We zullen hier echter wat meer de historische ontwikkelingen volgen en starten bij het werk van Wilson over renormalisatie. In de jaren '50, bij het uitwerken van problemen in kwantumveldentheorieën

zoals kwantumelektrodynamica, beseften natuurkundigen dat er schijnbaar grote problemen waren met hun vergelijkingen en resultaten. Her en der doken oneindigheden op voor waarden waarvan men experimenteel wist dat ze eindig waren. Bovendien, zelfs wanneer de oneindigheden konden worden weggewerkt, leken de voorspellingen af te hangen van de energieschaal van het probleem. Zo leek de lading van het elektron af te hangen van of we op lange afstand kijken, zoals in de Coulombwet, of op heel korte afstand, zoals in CERN.

Een grote doorbraak kwam er in de jaren '70 met het werk van Wilson. Hij werkte het idee van effectieve theorieën in detail uit en toonde aan waarom fysische grootheden schaalafhankelijk zijn. Het centrale idee is dat het aantal vrijheidsgraden in een theorie afhangt van de schaal waarop we werken, microscopische theorieën komen overeen met hoge energie en macroscopische met lage energie. Wanneer we starten met een microscopische theorie en hieruit een macroscopische afleiden, dienen we vrijheidsgraden uit te integreren.



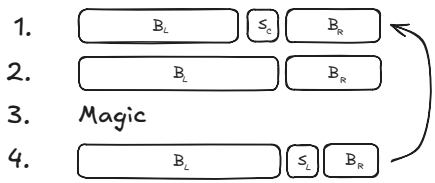
Figuur 8.12: Feynmandiagrammen voor een elektron-positroninteractie.

In kwantumveldentheorie komt dit neer op het samentellen van een hele hoop Feynmandiagrammen met interne (virtuele) deeltjes om zo een effectieve interactie te bekomen. We kunnen die virtuele deeltjes niet waarnemen bij lage energie, maar hun invloed is er wel nog steeds. Voor de

lading van een elektron komt dit erop neer dat we niet enkel het 'naakte' elektron moeten beschouwen, maar ook de zwerm van virtuele elektron-positronparen die erom heen kunnen vliegen. Ze schermen als het ware de echte lading af waardoor we op lange afstand een andere waarde waarnemen: de elementaire lading $q_e \approx 1,602 \times 10^{-19}$ C. In bovenstaande figuur is dit diagrammatische voorgesteld. Wanneer we de virtuele deeltjes niet kunnen waarnemen, is er geen echt onderscheid te maken tussen de inhoud van de drie gearceerde cirkels. Alle drie deze diagrammen (en alle andere Feynmandiagrammen die we hier kunnen bedenken) dragen bij aan de elektromagnetische koppeling.

De overgang van hoge energie naar lage energie, waarbij de vrijheidsgraden en fysische grootheden veranderen, wordt de renormalisatiestroming genoemd. Men kan voor de natuurconstanten onder meer vergelijkingen neerpennen die deze evolutie beschrijven. Deze ideeën hebben natuurkundigen ook verder inzicht gegeven in de soorten vrijheidsgraden die in een theorie aanwezig kunnen zijn. Zo maken we tegenwoordig het verschil tussen relevante en irrelevante vrijheidsgraden. De eerste zijn degene die onder de renormalisatiestroming intact blijven. De tweede zijn degene die bij lage energie niet meer meetbaar zijn. (Er bestaat nog een derde soort, de marginale vrijheidsgraden, en deze doen behoorlijk lastig.) Renormalisatie is echter niet enkel nuttig in de theoretische natuurkunde, ook computationeel is het interessant. (Anders zouden we dit hier natuurlijk niet introduceren.)

Stel dat we opnieuw een (eindige) keten van deeltjes beschouwen en de grondtoestand van een Hamiltoniaan $\widehat{H}$ willen bepalen. De kleinste eigenwaarde en bijhorende eigenvector van de Hamiltoniaan uitrekenen is onmogelijk door de vloek van de dimensionaliteit. We moeten de dimensie op een slimme manier beperken. Gemotiveerd door de renormalisatietheorie zullen we stapsgewijs vrijheidsgraden 'weggooien' en enkel de meest relevante behouden. Dit geeft uiteindelijk aanleiding tot het *density matrix renormalisation group*-algoritme (DMRG) van White [2].



Figuur 8.13: Het eindige DMRG-algoritme.

We starten met een initiële gok $|\psi\rangle$ voor het systeem op $L \in \mathbb{N}$ sites. Dit kan een heel eenvoudige onverstrengelde toestand zijn, dit kan een totaal willekeurige toestand zijn of dit kan een toestand zijn die reeds met een ander algoritme geoptimaliseerd is. De keuze is vrij, maar heeft wel een invloed op de snelheid waarmee ons algoritme zijn eindresultaat zal bereiken. Met voorkeur construeren we de initiële *ansatz* op zo een manier dat we het systeem eenvoudig kunnen opdelen in subsystemen, want we willen een toestand die we kunnen schrijven als in Stap 1 van het algoritme dat hiernaast is afgebeeld.

We beginnen met twee 'omgevingen' B_L, B_R en een centrale site S_C ertussenin, waarbij de omgevingen normaal gezien worden opgebouwd door alle sites links van S_C en rechts van S_C samen te nemen. S_C is hier de site die we in de huidige iteratie optimaliseren. Met deze opdeling krijgen we een basis voor de volledige toestandsruimte die bestaat uit het tensorproduct van de ruimtes $\mathcal{H}_{B_L}$, $\mathcal{H}_{S_C}$ en $\mathcal{H}_{S_R}$. In deze basis bepalen we nu (benaderend) de grondtoestand van de Hamiltoniaan $\widehat{H}$: $|\psi_0\rangle$. Dit is de inhoud van Stappen 2 en 3.

Hiervoor dienen we de actie van de totale Hamiltoniaan $\widehat{H}$ te bepalen. Deze bijdragen bestaan ook uit drie delen. Als eerste de lokale termen die inwerken op de linker en rechter omgeving: E_L en E_R. Hier is het belangrijk dat we deze bijdragen opslaan voor verdere berekeningen. Daarnaast hebben we ook nog de bijdrage E_C van termen die zowel op B_L, S_C als B_R inwerken. Om de totale energie $E_L + E_C + E_R$ te minimaliseren, bestaan heel wat algoritmes zoals gradiënt-gebaseerde algoritmes, die de energie minimaliseren zoals in een standaard optimalisatieprobleem, of *Krylovmethodes*, die een efficiënte zoekruimte opbouwen door de Hamiltoniaan meerdere keren na elkaar toe te passen op onze initiële vector. Er is keuze genoeg voor dit onderdeel van Stap 3.

Nu we de nieuwe benadering $|\psi_0\rangle$ hebben, dienen we het systeem opnieuw op te delen in een linker- en rechterdeel. Als we de dimensie van B_L met D aanduiden en de lokale dimensie van een site met d, dan zien we dat de nieuwe dimensie van het linkerdeel gegroeid zal zijn van D naar Dd, zoals de exponentiële groei van de dimensie vereist. Hier moeten we dus het idee van renormalisatie toepassen en de dimensie beperken. Oorspronkelijk had Wilson geprobeerd om een numeriek algoritme te ontwikkelen waarin we de toestand uitschreven in termen van de energie-eigenvectoren van de Hamiltoniaan en enkel de D belangrijkste toestanden meenamen. Het bleek echter dat dit in de praktijk niet goed werkte, ondanks dat het intuïtief wel steek hield aangezien we op zoek zijn naar een eigenvector van de Hamiltoniaan.

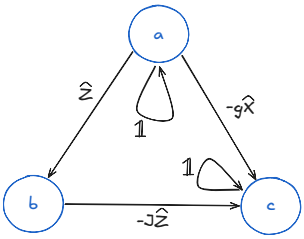
White zijn idee was om naar de verstrengeling te kijken. We beginnen dus met de constructie van de dichtheidsmatrix uit het vorige hoofdstuk: $\hat{\rho} = |\psi_0\rangle\langle\psi_0|$. Hierna kunnen we de linker en rechter partiële dichtheidsoperatoren construeren door het spoor te nemen over, respectievelijk, $\{S_R, B_R\}$ en $\{B_L, S_L\}$. Aangezien we hier voornamelijk in het nieuwe blok B_L geïnteresseerd zijn, focussen we op $\hat{\rho}_L$. Hier kunnen we nu de D belangrijkste eigenvectoren van berekenen, zelfs met een exact algoritme aangezien de dimensie beperkt is. Als laatste projecteren we het linkerdeel van $|\psi_0\rangle$ op de deelruimte van deze D vectoren om de dimensie van de totale ruimte constant te houden. (Indien gewenst kunnen we hier ook beslissen om de dimensie te verkleinen of te laten toenemen.) De keuze om in termen van verstrengeling en niet energie te werken, heeft ook een theoretische verklaring. Men kan aantonen dat de optimale benadering van een toestand, in termen van de dimensionaliteit, gegeven wordt door de (eigen)basis van verstrengelde toestanden. Dit sluit Punt 3 in Figuur 8.13 af.

Vooraleer we verdergaan naar Punt 4 en het patroon herhalen, waarbij we nu een site naar rechts zijn opgeschoven, is het belangrijk dat we de huidige toestand van B_L opslaan in het geheugen. Wanneer we de rechterzijde van het systeem bereikt hebben, gaan we naar links beginnen bewegen waarbij dan B_R zal vergroten en B_L zal verkleinen. Dat verkleinen

zullen we bekomen door de oude, reeds opgeslagen toestanden te hergebruiken.

Een volledige 'veeg' van het algoritme bestaat eruit naar rechts en links te bewegen doorheen het systeem tot we weer bij de initiële configuratie uitkomen. Hierbij zullen alle sites geüpdatet zijn en zal het totale systeem de echte grondtoestand beter benaderen dan voorheen. We blijven dan doorheen het systeem 'vegen' zolang we merken dat de updates de energie significant verlagen.

Het zal nu misschien niet als een verrassing aankomen als we zeggen dat dit algoritme heel wat eenvoudiger wordt met, of zelfs draait rond, matrixproducttoestanden. Start met een initiële MPS $|\psi[A]\rangle$. De linker en rechter blokken kunnen eenvoudig bekomen worden door de MPS geschikt te contraheren. Dit vereenvoudigt al heel wat werk, maar we kunnen nog beter doen. Hiervoor dienen we de Hamiltoniaan echter in MPO vorm te gieten. Voor veel eenvoudige Hamiltoniaan bestaat er een procedure die gebruik maakt van *cellulaire automata*, in het bijzonder de *eindige-toestandsautomaten*[5].



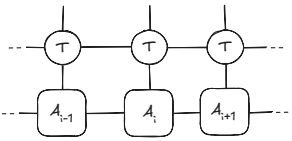
Figuur 8.14: Automaatrepresentatie voor het Isingmodel.

Als voorbeeld nemen we nog eens het Isingmodel, maar dan in zijn kwantummechanische gedaante:

$$\widehat{H}_{i,i+1} = -J\widehat{Z}_i\widehat{Z}_{i+1} - g\widehat{X}_i\,.$$

Hier stelt $\widehat{Z}$ een meting van de spin langsheen de z-as voor en, analoog, stelt $\widehat{X}$ de meting van de spin langsheen de x-as voor. (Trouwens, deze voldoen aan een onzekerheidsrelatie zoals positie en momentum: $[\widehat{Z}, \widehat{X}] = i\hbar\widehat{Y}$.) In de totale Hamiltoniaan hebben we dus drie operatoren $\mathbb{1}, \widehat{X}$ en $\widehat{Z}$. Alle mogelijke paden in de *gerichte graaf* hiernaast geven aanleiding tot alle mogelijke termen in de Hamiltoniaan $\widehat{H}$. «*Reken dit maar eens na.*»

[5] In het Engels spreken we van *finite state machines*.

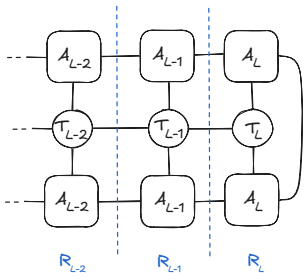


Figuur 8.15: De tensornetwerkvariant van DMRG.

In het Isingmodel zijn werken de termen in de Hamiltoniaan op maximaal twee termen tegelijk in, maar dit kan makkelijk veralgemeend worden.Voor lokale operatoren op drie sites zou de maximale lengte van een pad in de graaf drie zijn, enzovoort. We kunnen voor alle operatoren die bestaan uit lokale termen een dergelijke representatie opstellen. Voor gerichte grafen zoals die in de figuur bestaat er echter ook een matrixvoorstelling die ons zegt welke transformatie we nodig hebben om van de ene toestand naar de andere te gaan (deze matrix wordt ook wel de tranfermatrix genoemd). Voor het Isingmodel wordt dit:

$$T = \begin{pmatrix} \mathbb{1} & 0 & 0 \\ \widehat{Z} & 0 & 0 \\ -g\widehat{X} & -J\widehat{Z} & \mathbb{1} \end{pmatrix}.$$

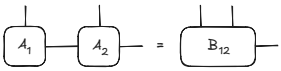
En zie hier, de vierde-orde tensor, een matrix van matrices, die zoeken om onze MPO te definiëren. «*Nog een leuke rekenoefening op matrixvermenigvuldiging. Kijk je ook eens na hoe we dit moeten doen aan de rand van onze keten (of op een periodieke keten)?*



Figuur 8.16: Contracties in de DMRG-omgevingen.

Laat ons nu maar terugkeren naar hoe MPS'en en MPO's ons kunnen helpen bij de implementatie van het DMRG-algoritme [3]. Als eerste brengen we de initiële MPS in een canonische vorm, bijvoorbeeld de rechts-canonische vorm, en berekenen we de rechter omgeving. De tensornetwerkuitdrukking laat ons toe om dit heel eenvoudig en gestructureerd, site per site, uit te rekenen en op te slaan. We beginnen bijvoorbeeld rechts in de figuur hiernaast en contraheren eerst A_L, T_L en de duale A_L (van

onder naar boven). De resulterende tensor R_L stockeren we in het geheugen voor verder gebruik. Vervolgens voegen we hier de tensoren op site $L-1$ aan toe om zo de tensor R_{L-1} te bekomen, enzovoort.



Figuur 8.17: Contracties van de linker omgeving.

Vervolgens starten we de eerste iteratie aan de linkerkant. We contraheren de eerste twee MPS-tensoren om de gecombineerde tensor B_{12} te bekomen (met een dimensie nd^2). De actie $\widehat{H} \cdot B_{12}$ van de Hamiltoniaan op deze gecombineerde tensor kunnen we nu berekenen op basis van de kennis van B_{12}, R_3, T_1 en T_2. «*Als je dit tekent en vergelijkt met de formule, zal je zien hoe efficiënt diagrammen zijn ten opzichte van expliciete formules.*» Ook hier weer kunnen we kiezen uit een reeks algoritmes om B_{12} te optimaliseren.

Eens we een nieuwe (betere) tensor B'_{12} gevonden hebben, kunnen we dezelfde methode als bij de introductie van matrixproducttoestanden eerder in dit hoofdstuk gebruiken. Door middel van een singuliere waardendecompositie kunnen we B'_{12} opsplitsen als de contractie van twee kleinere tensoren, waarbij de keuze van bindingsdimensie de dimensionaliteit zal beperken, in lijn met het idee van renormalisatie. (Door de structuur van de singuliere waardendecompositie zal de resulterende linker tensor ook automatisch in links-canonische vorm staan.) Men kan aantonen dat deze operatie volledig equivalent is aan het vinden van de eigenvectoren van de geassocieerde dichtheidsoperator.

Vervolgens gaan we over naar de volgende site en berekenen we ondertussen een nieuwe linker omgeving L_1 met de tensor A'_1 die we in de vorige paragraaf gevonden hebben. Wanneer we aan het einde van onze keten gekomen zijn, keren we de richting om en beginnen we opnieuw. Zo blijven we heen en weer bewegen totdat de updates een bepaalde, vooraf gekozen precisie bereikt hebben. Ook al had White het niet door, zijn algoritme optimaliseerde impliciet de grondtoestand binnen de variëteit opgespannen door matrixproducttoestanden.

Hoofdstuk 9

Epiloog

Bibliografie

[1] B. Ahrens and P. R. North. Univalent Foundations and the Equivalence Principle. In *Reflections on the Foundations of Mathematics*, volume 407, pages 137–150. Springer International Publishing, 2019.

[2] G. Catarina and B. Murta. Density-matrix renormalization group: a pedagogical introduction. *The European Physical Journal B*, 2023.

[3] Flatiron Institute. The Tensor Network, 2026.

[4] D. S. Freed and G. W. Moore. Twisted Equivariant Matter. *Annales Henri Poincaré*, pages 1927–2023, 2013.

[5] R. Keleti. Translation project of Grothendieck's EGA, 2026.

[6] J. Kepler. *The Six-Cornered Snowflake*. 1611.

[7] J. Moeller, T. Hosgood, J. Bian, and M. Pierce. An English Translation of Grothendieck's SGA, 2018.

[8] K. Rejzner. *Perturbative Algebraic Quantum Field Theory*. Mathematical Physics Studies. Springer International Publishing, 2016.

[9] S. Singh. *Fermat's Last Theorem*. Fourth Estate, 1997.