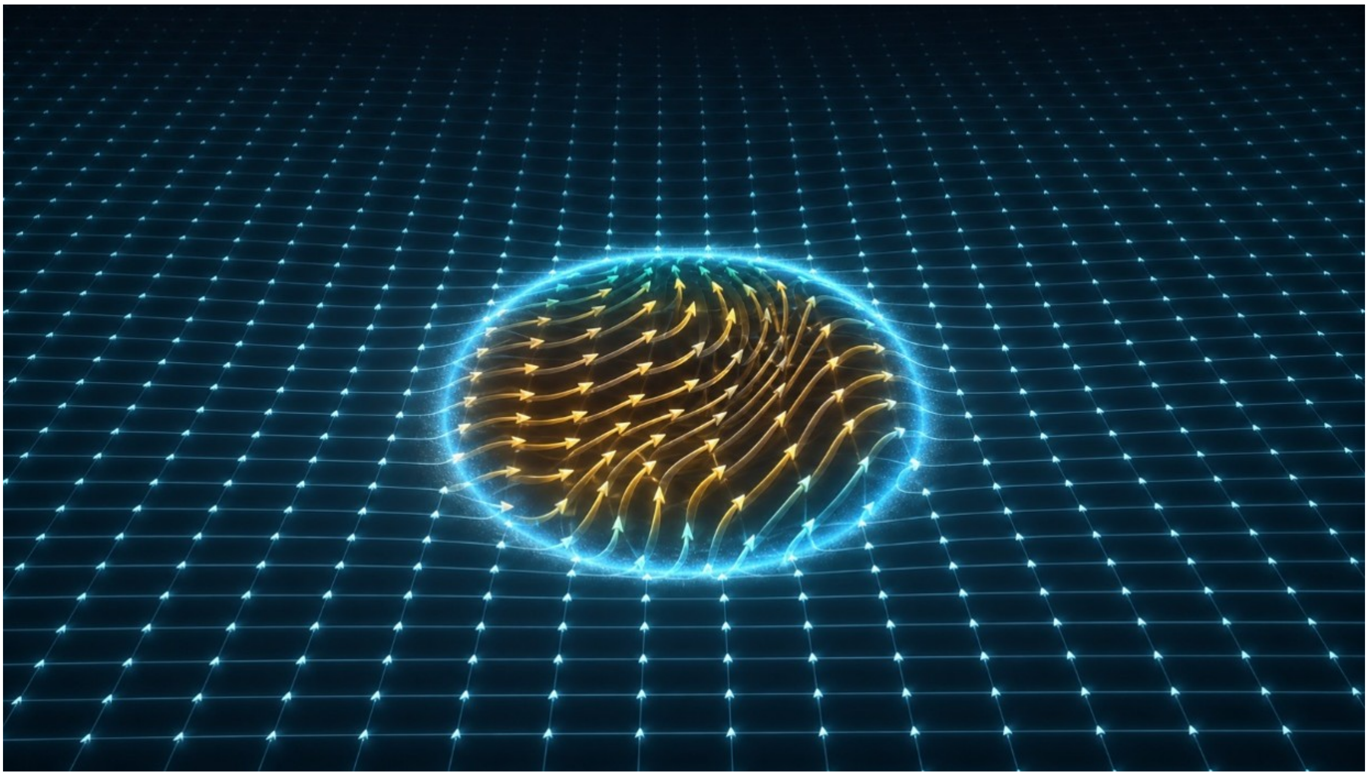
De tweede stelling van Noether

Het door Emmy Noether ontdekte verband tussen symmetrieën en behoudswetten is talloze keren op de Quantum Universe-website beschreven en wordt dikwijls door natuurkundigen aangehaald als een van de belangrijkste resultaten uit de twintigste eeuw. Daarbij gaat men echter vaak voorbij aan het feit dat er in haar oorspronkelijke artikel uit 1918, *Invariante Variationsprobleme*, niet één maar twee hoofdstellingen werden bewezen. De eerste is wat we normaal "dé stelling van Noether'' noemen, maar de tweede is minstens zo belangrijk!

Lokaliseerbare symmetrieën

In een vorig artikel over de symmetrie van de zwaartekracht beschreef ik al het cruciale onderscheid tussen *globale* en *lokale* symmetrieën. De twee stellingen van Noether gaan respectievelijk over zulke globale en lokale symmetrieën, dus het is goed om even te herhalen wat het verschil is. Simpel gezegd doet een globale symmetrietransformatie overal in de ruimte en/of tijd hetzelfde: denk bijvoorbeeld aan een constante verschuiving (translatie) van alle objecten in de ruimte met een bepaalde afstand, wat dan via de eerste Noetherstelling leidt tot impulsbehoud. Een lokale symmetrie kan daarentegen op ieder punt van de ruimte en/of tijd anders werken. Zo staat Maxwells theorie van elektromagnetisme erom bekend dat je aan de zogeheten *potentiaal* van het magnetisch veld een afgeleide van een willekeurige functie kan toevoegen. Die functie kan van punt tot punt verschillen, zonder dat daarmee de energie van het magnetisch veld verandert. Wel moet je tegelijkertijd ook een andere potentiaal, die van het elektrisch veld, op de juiste manier meetransformeren – daar kom ik zo op terug. Lokale symmetrieën zoals die in Maxwells theorie worden ook wel *ijksymmetrieën* genoemd en zijn van wezenlijk belang voor ons begrip van alle fundamentele natuurkrachten.



Afbeelding 1. IJksymmetrie. Een voorstelling van een symmetrietransformatie die gelokaliseerd is in een regio van de ruimte. Bron: GPT 5.5.

De tweede stelling van Noether geeft aan dat er iets heel bijzonders gebeurt wanneer er een lokale symmetrie aanwezig is in een veldentheorie. Bij het bewijs van de stelling wordt gebruik gemaakt van een specifieke eigenschap van lokale symmetrieën: hun *lokaliseerbaarheid*. Een symmetrietransformatie heet ‘lokaliseerbaar’ als het voor iedere twee niet-overlappende, maar verder willekeurige regio’s van de ruimtetijd, noem ze R_1 en R_2, mogelijk is om een ándere symmetrietransformatie te kiezen die hetzelfde doet als de eerste op R_1, maar niets doet op R_2. Je kunt zo iedere symmetrietransformatie dus ‘lokaliseren’ in een bepaald gebiedje, waarbuiten er niets gebeurt. De essentiële vorm van lokaliseerbaarheid van ijktheorieën blijkt nu *tijdslokaliseerbaarheid* te zijn, waarbij de hierboven beschreven regio’s op andere tijdstippen in de ruimtetijd moeten liggen.

Oneindig veel behoudswetten

Om te snappen hoe tijdslokaliseerbaarheid de motor vormt van de tweede stelling van Noether, is het van belang om nog eens wat beter naar de eerste stelling te kijken. Je kunt de tweede stelling namelijk zien als een speciaal gevolg van de eerste dat geldt wanneer de symmetriegroep tijdslokaliseerbaar is. Daarvoor moet je echter wel bereid zijn om de eerste stelling toe te passen in oneindig veel dimensies, hetgeen wat voorstellingsvermogen vereist.

De eerste stelling van Noether legt een verband tussen symmetrieën en behoudswetten. Dat is eigenlijk helemaal niet zo verrassend als we dieper over het begrip ‘energie’ nadenken. In de natuurkunde is energie een waarde die je toekent aan iedere mogelijke toestand van een systeem. Die energiefunctie legt vervolgens volledig vast hoe het systeem zich ontwikkelt. Je kunt van iedere grootheid die het systeem beschrijft bepalen hoe deze verandert door simpelweg te kijken naar de energie die hoort bij de verschillende waarden van die grootheid. Denk hierbij aan een bal op een heuvel, waarbij de hoogte van de heuvel de waarde van de energie moet voorstellen. Als de heuvel schuin is, dan zal de bal naar beneden gaan rollen. De positie van de bal verandert dan omdat de (potentiële) energie onderaan de heuvel lager is dan bovenaan. In andere woorden: de bal verkrijgt een snelheid of *impuls* omdat de potentiële energie niet constant is als functie van de positie. Als het landschap echter vlak was geweest, dan was de bal niet sneller of langzamer gaan rollen, maar dan was de snelheid/impuls constant gebleven (aangenomen dat er geen wrijving is). Bij een vlak landschap, ofwel constante potentiële energie, blijft de impuls dus gelijk: de impuls is behouden! De eerste stelling van Noether verklaart dit impulsbehoud dan vanuit de symmetrie van het landschap / de energie: het maakt niet uit of ik de de bal op het vlakke terrein naar links of rechts verplaats – de potentiële energie blijft gelijk. Dat heet translatiesymmetrie. Als het terrein echter niet vlak maar een steile heuvel was geweest, dan had het wél uit gemaakt of ik de bal verplaats of niet, en dan was de impuls niet behouden geweest.

Zo werkt het bij iedere denkbare symmetrietransformatie: er is een ‘verschuiving’ van een veralgemeniseerde coördinaat (dit kan nu bijvoorbeeld ook een verschuiving van een hoek zijn, dus een rotatie) die de energie niet verandert. Dan blijkt dat er een bijbehorende veralgemeniseerde ‘impuls’ bestaat die behouden is, waarbij de waarde van die impuls aangeeft hoe snel die veralgemeniseerde coördinaat verandert. Dit principe is zo algemeen dat het eigenlijk op iedere natuurkundige theorie toegepast kan worden. Het kan dus ook op veldentheorieën zoals die van het elektromagnetisme toegepast worden. Sterker nog: het kan toegepast worden op de *lokale* ijksymmetrieën van zulke veldentheorieën.

Een veldentheorie heeft een oneindig aantal vrijheidsgraden, omdat er op ieder punt van de ruimte een waarde van het veld gespecificeerd moet worden. De toestandsruimte van een veldentheorie is dus oneindigdimensionaal: deze ruimte heeft oneindig veel coördinaten. Het simpelste type veld kan op ieder punt van de ruimte een reële waarde aannemen – een temperatuurveld dat op ieder punt de temperatuur aangeeft is een voorbeeld. Zo'n veld heeft dus op ieder punt van de ruimte één coördinaat (de temperatuur op die plek). Omdat er oneindig veel punten in de ruimte zijn levert dat in totaal oneindig veel cöordinaten op. Ingewikkeldere veldentheorieën hebben op ieder punt niet één maar meerdere cöordinaten: het veld van de sterke kernkracht, bijvoorbeeld, heeft op ieder punt acht coördinaten.

Welnu, de eerste Noetherstelling geeft voor *iedere* coördinaat waarop de symmetrie werkt een behouden grootheid – de bijbehorende 'impuls'. Als we dit dus toepassen op de lokale symmetrie van een ijktheorie zoals elektromagnetisme, dan verwachten we oneindig veel behoudswetten, namelijk een behoudswet op *ieder* punt van de ruimte. In de elektromagnetische theorie valt eenvoudig te zien dat die verwachting klopt. Het onderliggende veld is daarbij de magnetische vectorpotentiaal $\vec{A}$. Zoals eerder gezegd kent dit veld een ijksymmetrie waarbij de energie hetzelfde blijft als we een afgeleide van een functie λ toevoegen, dus

$$\large \vec{A}\to \vec{A} + \nabla \lambda(x).$$

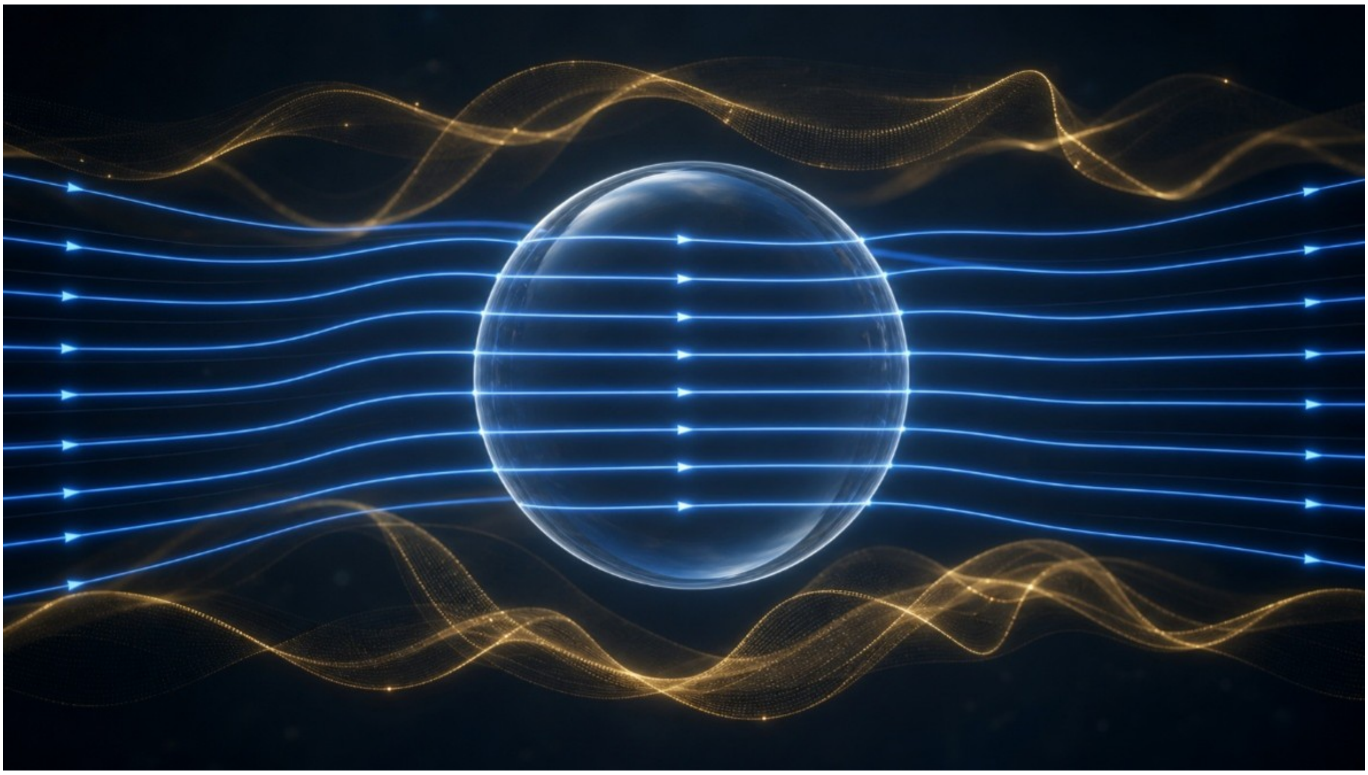
De afgeleide geven we in die formule aan met het symbool ∇. De waardes $\vec{A}(x)$ op ieder punt x zijn de coördinaten, en de eerste Noetherstelling vertelt ons dan dat er voor ieder van deze coördinaten een behouden impuls is. De impuls die hoort bij de magnetische vectorpotentiaal op een bepaalde plek is het elektrisch veld op die plek, $\vec{E}(x)$. We zouden dus verwachten dat het hele veld $\vec{E}(x)$ behouden is. Maar we moeten wel in ons achterhoofd houden dat de potentiaal $\vec{A}(x)$ niet zomaar verschoven kan worden, maar alleen met een afgeleide. Om die reden moet er ook een afgeleide bij de bijbehorende impuls komen. Met andere woorden: de behouden grootheid voor deze lokale ijksymmetrie is de afgeleide van het elektrisch veld:

$$\large \nabla\cdot\vec{E}(x).$$

Deze afgeleide wordt ook wel de divergentie van het elektrisch veld genoemd, en is dus behouden dankzij de eerste stelling van Noether, toegepast op de lokale symmetrie die afzonderlijk werkt op ieder punt x van de ruimte. Omdat de divergentie op ieder punt behouden is, hebben we oneindig veel behoudswetten, eentje per punt van de ruimte.

Van behoudswet naar beperking

Toch kan dit niet het volledige verhaal zijn. Wie voldoende kennis heeft van elektromagnetisme weet namelijk dat één van de vier maxwellvergelijkingen zoals die in het vacuüm gelden, de zogenaamde *wet van Gauss*, stelt dat de divergentie van het elektrisch veld overal nul is: $\nabla\cdot\vec{E}(x)=0$. De wet van Gauss vertelt ons dus dat $\nabla\cdot\vec{E}$ niet zomaar een behouden grootheid is, maar dat deze ‘impuls’, behorend bij de lokale ijksymmetrie, in het bijzonder altijd *nul* is. Vanuit de eerste stelling van Noether zouden we alleen maar verwachten dat $\nabla\cdot\vec{E}$ gelijk is aan één of andere functie die in de loop van de tijd niet verandert, net zoals de gewone impuls van een deeltje willekeurig welke constante waarde kan hebben. Het is echter precies de tweede wet van Noether die uitlegt dat bij een ijksymmetrie als constante waarde alleen de waarde nul is toegestaan.



Afbeelding 2. De wet van Gauss. Een voorsteling van de wet van Gauss: de elektrische veldlijnen die een bepaald gebied in gaan moeten daar ook weer uit komen, omdat de wet van Gauss stelt dat er in het vacuüm geen bronnen voor het elektrisch veld zijn. Bron: GPT 5.5.

Het feit dat de behouden waarde alleen nul kan zijn, heeft alles te maken met de tijdslokaliseerbaarheid waar ik het eerder over had, alsook met mijn eerdere opmerking dat de potentiaal V van het elektrisch veld óók op de juiste manier moet meetransformeren met de magnetische vectorpotentiaal $\vec{A}$. De elektromagnetische theorie is namelijk, zoals vrijwel alle hedendaagse veldentheorieën, relativistisch. Dat betekent dat er geen voorkeurskeuze is voor wat je ruimte en tijd noemt – die begrippen kunnen in elkaar overlopen. Voor het elektromagnetisch veld heeft dat ook gevolgen: wat je elektrisch veld en magnetisch veld noemt, blijkt afhankelijk van hoe je ruimte en tijd opsplitst. De ene waarnemer kan een elektrisch veld zien, terwijl dat er voor een andere waarnemer uitziet als een magnetisch veld. De elektrische en magnetische potentialen moeten dus verenigd worden in één veld, dat waardes op de vierdimensionale ruimtetijd aanneemt. Dit gecombineerde veld schrijven we als A_μ. De index μ kan waarden 0,1,2 en 3 aannemen, die voor de ene tijdrichting en de drie ruimterichtingen staan. Ook deze zogenaamde 4-potentiaal kan door lokale ijksymmetrieën getransformeerd worden zonder dat de energie daarbij verandert:

$$\large A_\mu\to A_\mu +\partial_\mu\lambda.$$

Hier is λ een functie op de gehele ruimtetijd, en geeft ∂_μ weer aan dat er een afgeleide in de betreffende richting wordt genomen. De 0-component van A_μ, dus A_0, wordt ook wel als V geschreven en heet de elektrische potentiaal. De andere drie componenten vormen de eerdergenoemde magnetische vectorpotentiaal $\vec{A}$. Uit bovenstaande formule blijkt dat, als de magnetische potentiaal met $\nabla\lambda$ (wat hetzelfde is als $\partial_i\lambda$ voor $i=1,2,3$), wordt veranderd door een ijktransformatie, de elektrische potentiaal dan ook met $\partial_0\lambda$ moet veranderen. Dit wordt domweg opgelegd door de relativiteit van ruimte en tijd. Maar $\partial_0\lambda$ is niets anders dan de tijdsafgeleide, want de 0-index geeft de tijdsrichting aan. Kortom: de elektrische potentiaal transformeert onder een ijktransformatie mee, maar dan met een tijdsafgeleide van de functie λ, in plaats van met een ruimtelijke afgeleide zoals de magnetische potentiaal. Een gevolg hiervan is dat het elektrische veld kinetische energie draagt, en het magnetische veld potentiële energie.

Dit gegeven leert een natuurkundestudent in het eerste jaar van de bachelor en lijkt misschien niet zo bijzonder. Maar schijn bedriegt: dit is één van de spectaculairste verschijnselen in de moderne fysica. Wat we zien is namelijk een veldentheorie die een lokale ijksymmetrie heeft op de gehele ruimte*tijd*. Dat maakt die symmetrie tijdslokaliseerbaar: we kunnen ervoor kiezen om de functie λ op het ene tijdstip een andere waarde te geven dan op een ander tijdstip. Maar we moeten in ons achterhoofd houden dat de tijdsafgeleide van een veld de “snelheid” van dat veld is, en kinetische energie wordt altijd gegeven door het kwadraat van de snelheid. Zo ook bij de elektromagnetische theorie: de kinetische energie is daar het kwadraat van het elektrisch veld. Als we aan een veld dus een tijdsafhankelijke functie kunnen toevoegen, zoals in dit geval de tijdsafgeleide van λ, dan veranderen we met die transformatie de snelheid van het veld. Maar daarmee zouden we de kinetische energie veranderen, wat de transformatie niet langer tot een symmetrie maakt… De enige uitweg is dat de snelheid van het veld in kwestie niet mag voorkomen in de energie! In het geval van elektromagnetisme zien we dat heel duidelijk wanneer we de formule voor de energie inspecteren: daar komt wél de tijdsafgeleide van de magnetische potentiaal in voor, maar niét de tijdsafgeleide van de elektrische potentiaal.

Dat is een hele rare situatie: we hebben een veld, de elektrische potentiaal V, waarvan de snelheid géén kinetische energie draagt. Dit stelt ons in staat om dat veld naar hartenlust gedurende de tijd te transformeren, zonder consequenties. Althans, de energie verandert niet door zulke transformaties, maar er is juist wel een groot gevolg voor de behouden impuls: die moet overal gelijk aan nul zijn! De vrijheid van tijdslokaliseerbaarheid maakt het namelijk mogelijk om de impuls, die op het ene moment een willekeurige waarde heeft, op een ander moment op nul te zetten door de transformatieparameter λ op dat tijdstip op nul te zetten. De eerste Noetherstelling zegt dat de impuls behouden is, dus als deze op één moment nul is, dan moet de impuls dat op alle momenten zijn. Voor elektromagnetisme betekent dit dat de divergentie van het elektrisch veld niet zomaar een willekeurige constante waarde kan aannemen, maar aan de wet van Gauss $\nabla\cdot\vec{E}=0$ moet voldoen.

Zo’n eis die verder gaat dan een behoudswet en stelt dat de impuls die hoort bij een lokale symmetrie overal nul moet zijn heet een *constraint*, of ‘beperking’. De ruimte van toegestane grootheden is immers beperkt in de zin dat de impuls slechts één waarde in plaats van alle mogelijke waarden kan aannemen. De tweede noetherstelling kan dan als volgt samengevat worden: *een tijdslokaliseerbare symmetrie leidt tot een beperking op de toestanden*.

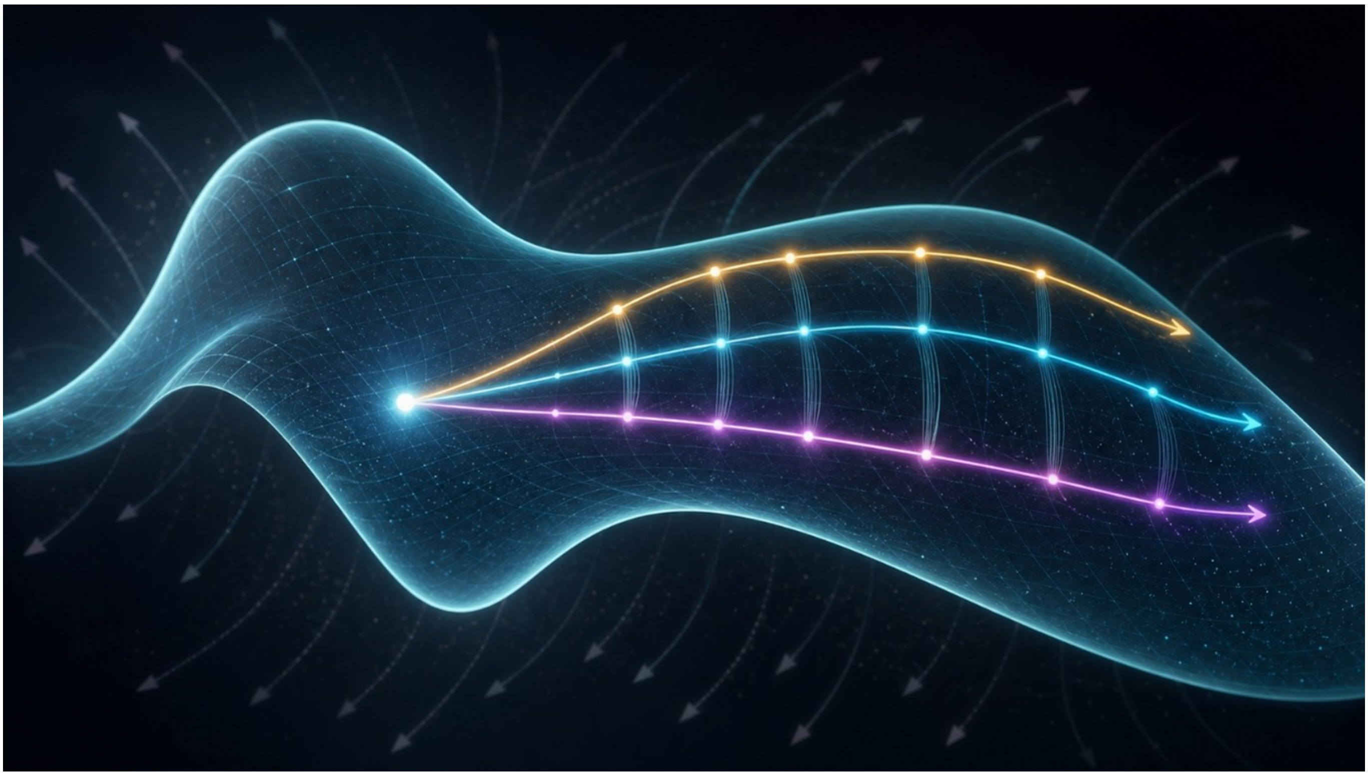
Er is misschien één stap in dit relaas die niet helemaal duidelijk is. Voor de geëngageerde lezer met voldoende affiniteit met wiskunde wil ik die tot besluit van deze sectie nog wat nader toelichten. Het gaat om de vraag waarom de tijdslokaliseerbaarheid van de symmetrietransformatie, dus bij elektromagnetisme het feit dat λ ook een willekeurige functie van de *tijd* is, ertoe moet leiden dat de behouden impuls nul is. Wat hebben die met elkaar te maken? Welnu, de wiskundig precieze beschrijving van impuls gebeurt aan de hand van de *duale Lie-algebra* van de symmetriegroep. Dit klinkt chic maar is eigenlijk eenvoudig: de impuls geeft voor iedere symmetrieparameter een getal. De waarde van de impuls beschrijft immers hoe snel de bijbehorende coördinaat verandert: een heel hoge impuls betekent dat die coördinaat, die door de symmetrieparameter verschoven wordt, heel snel verandert. De impuls wordt daarom wiskundig beschreven als een functie van de symmetrieparameters naar de reële getallen die *lineair* blijkt te zijn. De ruimte van zulke lineaire functies heet de ‘duale Lie-algebra’. Nu zegt de eerste stelling van Noether in precieze vorm dat de impuls voor *iedere* symmetrieparameter in de Lie-algebra behouden is. In elektromagnetisme betekent dat dat de integraal

$$\large \int \lambda \nabla\cdot\vec{E}$$

behouden moet zijn voor *iedere* functie λ, waar de integraal over de *ruimte* wordt gedaan. Voor iedere willekeurige waarde van de functie λ *over de ruimte*, kunnen we nu λ als functie over de gehele ruimte*tijd* zo kiezen dat λ in de toekomst overal nul is. Dit maakt bovenstaande integraal in de toekomst nul. Maar omdat die integraal behouden is (vanwege de eerste stelling), moet deze dus ook op dat eerdere moment al nul zijn. Omdat λ een willekeurige functie over de ruimte was moeten we concluderen dat het de impuls $\nabla\cdot\vec{E}$ zelf is die overal nul moet zijn.

Indeterminisme

Nu we begrijpen dat de tweede stelling van Noether zegt dat een tijdslokaliseerbare symmetrie tot een *constraint* leidt, is het leuk om nog kort iets te zeggen over het verband tussen ijksymmetrie en indeterminisme. Er wordt door natuurkundigen heel vaak gezegd dat een ijksymmetrie geen echte symmetrie is, maar een “redundantie” of “overbodigheid” in de wiskundige beschrijving van een fysisch systeem. Verschillende, door de ijksymmetrie aan elkaar gerelateerde toestanden zouden dus niet écht verschillen, maar alleen in onze beschrijving. Een ijksymmetrie wordt dan vergeleken met een keuze van coördinaten: dat is een keuze van hoe je iets beschrijft, maar er zit niets “echts”, niets “fysisch” in die keuze.



Afbeelding 3. Een constraint surface. Op de constraint surface zijn er vanuit één toestand meerdere toekomstige paden mogelijk. Bron: GPT 5.5.

Uit de eerste stelling van Noether wordt echter helemaal niet duidelijk wat het verschil tussen “gewone” symmetrieën en ijksymmetrieën dan wel zou moeten zijn. Daarvoor heb je echt de tweede stelling nodig. De *constraint* die uit de tweede Noetherstelling volgt zorgt namelijk voor *indeterminisme*: een gebrek aan voorspelbaarheid van het systeem in kwestie. Om determinisme of voorspelbaarheid te waarborgen, moet je voor iedere coördinaat een impuls hebben, en voor iedere impuls een coördinaat. Het maakt daarbij niets uit of die impulsen behouden zijn door een of andere symmetrie via de eerste Noetherstelling. Maar als zo’n impuls *beperkt* is, dan blijkt dat de tijdsevolutie op die beperkte ruimte (de zogenaamde *constraint surface*) niet meer eenduidig is: vanuit één begintoestand zijn er meerdere toekomstige toestanden mogelijk. Het blijkt tevens dat deze verschillende mogelijkheden precies aan elkaar gerelateerd zijn door de ijksymmetrie! Helaas kan ik dit hier niet écht goed uitleggen, want dat zou enige kennis van technische onderwerpen als de symplectische meetkunde en de Hamiltoniaanse mechanica vereisen.

Een andere manier om in te zien dat een ijksymmetrie verband houdt met indeterminisme is als volgt. Eerder zei ik dat een ijkveld altijd een component kent waarvan de snelheid niet in de kinetische energie voortkomt. Maar als die component geen energie draagt, dan bepaalt de energie dus ook niets over hoe die component zich gedurende de tijd zal veranderen, en zijn allerlei tijdsevoluties mogelijk.

Slotbeschouwing

Het werkt van Emmy Noether uit 1918 was geniaal en speelt nog altijd een enorme rol in vrijwel ieder deelgebied van de fysica. Maar natuurkundigen moeten zich er terdege van bewust zijn dat zij vaak onprecies over het werk van Noether spreken. Zij bewees twee stellingen. De eerste geeft het verband tussen symmetrieën en behoudswetten en is toepasbaar op *alle* symmetrieën, inclusief ijksymmetrieën. De tweede stelling geeft aan dat er in het geval van een relativistische ijksymmetrie daarenboven iets heel bijzonders gebeurt: vanwege de tijdslokaliseerbaarheid van die symmetrie krijgen we niet alleen oneindig veel behoudswetten, één op ieder punt, maar is de behouden impuls ook nog eens op elk punt nul. Die impuls is dus niet alleen *behouden*, maar ook *beperkt.* Dit leidt tot indeterminisme en is de wortel van het idee dat een ijksymmetriegeen fysische transformatie van de toestanden van een systeem is. Dirac doorzag dit als eerste, maar vandaag de dag begrijpen we dat er een uitweg bestaat uit deze gevolgtrekking: als een systeem een rand heeft, dan kan een ijksymmetrie wel degelijk fysisch zijn door haar werking op die rand. Een voorbeeld hiervan zijn de asymptotische symmetrieën waar Ameya Kadhe recentelijk over schreef. Ook het fysisch zijn van ijksymmetrieën op de rand van een systeem kan je perfect vanuit de tweede stelling van Noether begrijpen, maar daar moeten we het een andere keer nog maar eens over hebben!