GRAVITATIE

van Zwaartekracht tot Kosmologie

J.W. van Holten

Nationaal Instituut voor Kern- en Hoge-Energiefysica (NIKHEF)

Bijzonder hoogleraar theoretische natuurkunde aan de Vrije Universiteit te Amsterdam vanwege de Stichting Hoge-Energiefysica

Amsterdam, 2000

Inhoudsopgave

Er zijn meer dingen in de hemel en op aarde, Horatio, dan wij in onze dromen kunnen bedenken

Hoofdstuk 1

Gravitatie

1.1 Zwaartekracht en de wetten van Kepler

Newtons wet van de zwaartekracht stelt dat twee massa's een aantrekkende kracht op elkaar uit oefenen die evenredig is met beide massa's, en omgekeerd evenredig met het kwadraat van de afstand:

$$\mathbf{F} = -G\frac{m_1 m_2}{r^2}\,\hat{\mathbf{r}}. \tag{1.1}$$

De eenheidsvector $\hat{\mathbf{r}} = \mathbf{r}/r$ geeft de richting van de kracht aan, centraal langs de as die de massamiddelpunten van de lichamen verbindt.

Vanuit een historisch gezichtspunt leverde Newtons wet de eerste kwantitatieve beschrijving van een fundamentele natuurkracht. Zij leidt tot een correcte beschrijving van zwaartekrachtsverschijnselen zoals de vrije val, de worp van een bal of ander projectiel, en de slingerbeweging; maar de belangrijkste toepassing was een verklaring van de beweging van de maan om de aarde en van de planeten om de zon, zoals vastgelegd in de drie empirische wetten van Kepler:

1. De banen van de planeten om de zon zijn kegelsneden (ellipsen), met de zon in een van de brandpunten.
2. In gelijke tijden doorloopt de voerstraal –het lijnstuk dat een planeet met de zon verbindt– gelijke oppervlakken.
3. Het kwadraat van de omlooptijd van een planeet om de zon is evenredig met de derde macht van de (halve) lange as.

Ook de getijden in de oceaan konden als samenspel tussen zwaartekracht en centrifugale (schijn)kracht begrepen worden. Daarmee leverde de fysica van de zwaartekracht een bewijs, dat dezelfde natuurwetten die gelden voor verschijnselen op aarde ook van toepassing zijn op de beweging van hemellichamen. Met andere woorden, Newtons wet leidde tot een unificatie van de aardse en de hemelmechanica. Het natuurwetenschappelijk en filosofisch belang van deze vaststelling kan in historisch perspektief nauwelijks worden overschat: Newtons beschrijving van de zwaartekracht was en is een paradigma dat vele ontwikkelingen in de natuurwetenschap tot op de dag van vandaag beinvloedt.

Om die reden, en om later de door Einstein ingevoerde verbetering van deze beschrijving beter te kunnen appreciëren, zullen we hier de afleiding van Keplers wetten uit Newtons zwaartekrachtwet geven.

De versnellingen van de twee massa's ten gevolge van de kracht (1.1) worden gegeven door

$$\ddot{\mathbf{r}}_1 = G\frac{m_2}{r^2}\hat{\mathbf{r}}, \qquad \ddot{\mathbf{r}}_2 = -G\frac{m_1}{r^2}\hat{\mathbf{r}}. \tag{1.2}$$

In deze vergelijkingen is $\hat{\mathbf{r}}$ de eenheidsvector gericht van massa 1 naar massa 2. Het massamiddelpunt wordt gegeven door de vector $\mathbf{R}$ die voldoet aan

$$M\mathbf{R} = m_1\mathbf{r}_2 + m_2\mathbf{r}_2, \qquad M = m_1 + m_2, \tag{1.3}$$

en overeenkomstig de hierboven ingevoerde conventies definiëren we de relatieve positievector $\mathbf{r}$ als

$$\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1 = r\hat{\mathbf{r}}. \tag{1.4}$$

Voor de versnellingen van deze beide positievectoren vinden we dan door optellen en aftrekken van geschikte combinaties van vgln.(1.2) dat het massamiddelpunt niet versneld wordt:

$$\ddot{\mathbf{R}} = 0, \tag{1.5}$$

terwijl de relatieve versnelling gegeven wordt door

$$\ddot{\mathbf{r}} = -\frac{GM}{r^2}\,\hat{\mathbf{r}} = -\frac{GM}{r^3}\,\mathbf{r}. \tag{1.6}$$

Het massamiddelpunt beweegt daarom eenparig en rechtlijnig:

$$\mathbf{R}(t) = \mathbf{R}(0) + \mathbf{V}t, \tag{1.7}$$

met een snelheid $\mathbf{V}$ die wordt vastgelegd door de beginvoorwaarden. De relatieve positie $\mathbf{r}$ gedraagt zich daarentegen effectief als de positie van een enkel deeltje in het stationair centraal krachtveld van een massa M. Nemen we in het linker- en rechterlid van vgl.(1.6) het inwendig product met de relatieve snelheid $\mathbf{v} = \dot{\mathbf{r}}$, dan krijgen we na herschrijving

$$\frac{d}{dt}\left(\frac{1}{2}\dot{\mathbf{r}}^2\right) = \frac{d}{dt}\left(\frac{GM}{r}\right), \tag{1.8}$$

waarbij $r = \sqrt{\mathbf{r}^2}$. De bovenstaande vergelijking is equivalent met het behoud van energie, en kan ook geschreven worden in de vorm

$$\frac{1}{2}\,\dot{\mathbf{r}}^2 - \frac{GM}{r} = \epsilon, \tag{1.9}$$

met $\epsilon = E_{CM}/\mu$, een constante met de waarde van de energie in het massamidelpuntsysteem gedeeld door de gereduceerde massa:

$$\mu = \frac{m_1 m_2}{M}. \tag{1.10}$$

Behalve de energie is ook het baanimpulsmoment behouden. Dit is een gevolg van de bolsymmetrie van het krachtveld. Het bewijs is simpel: met

$$\mathbf{L} = \mu \mathbf{r} \times \mathbf{v} \tag{1.11}$$

volgt direct dat

$$\begin{aligned} \dot{\mathbf{L}} &= \mu \dot{\mathbf{r}} \times \dot{\mathbf{r}} + \mu \mathbf{r} \times \ddot{\mathbf{r}} \\ &= 0 - G \frac{\mu M}{r^3} \mathbf{r} \times \mathbf{r} \, = \, 0. \end{aligned} \tag{1.12}$$

Omdat zowel de grootte als de richting van het impulsmoment behouden zijn, ligt de baan in een plat vlak loodrecht op $\mathbf{L}$; dit volgt direct uit de vergelijking $\mathbf{r} \cdot \mathbf{L} = 0$. We kunnen de wiskundige analyse verder vereenvoudigen door de z-as in de richting van het impulsmoment te kiezen, en het baanvlak als het x-y-vlak te nemen. In bolcoördinaten (r, θ, φ), gedefiniëerd door

$$x = r \sin\theta \cos\varphi, \quad y = r \sin\theta \sin\varphi, \quad z = r \cos\theta, \tag{1.13}$$

levert dit $\theta = \pi/2 =$ constant, en

$$\begin{aligned} L_z &= \mu \left(x\dot{y} - y\dot{x} \right) \\ &= \mu r^2 \dot{\varphi} \, \equiv \, \mu \ell. \end{aligned} \tag{1.14}$$

Aangezien dit in onze conventies de enige component van het impulsmoment is, krijgen we $\mathbf{L} = (0, 0, \mu\ell)$.

Het behoud van impulsmoment leidt direct tot Keplers perkenwet: het oppervlak A dat de voerstraal r per tijdseenheid doorloopt is constant:

$$\frac{dA}{dt} = \frac{1}{2} r^2 \dot{\varphi} = \frac{\ell}{2}. \tag{1.15}$$

Hiermee ligt de hoekbeweging van het twee-lichamensysteem onder invloed van een centrale kracht vast. Nu moet alleen het gedrag van de radiële coördinaat als functie van de tijd nog worden opgelost. Hiertoe gebruiken we het behoud van energie, vgl.(1.9). Eerst herschrijven we deze in bolcoördinaten, waarin de kinetische term de volgende vorm aanneemt:

$$\dot{\mathbf{r}}^2 = \dot{r}^2 + r^2 \dot{\theta}^2 + r^2 \sin^2\theta \, \dot{\varphi}^2. \tag{1.16}$$

Door gebruik te maken van het behoud van impulsmoment (1.14), met het bijkomend gevolg dat de baan in een plat vlak ligt: $\sin\theta = 1$ en $\dot{\theta} = 0$, krijgen we dan:

$$\frac{1}{2}\,\dot{r}^2 + \frac{\ell^2}{2r^2} - \frac{GM}{r} = \epsilon. \tag{1.17}$$

Dit is een 1-dimensionaal potentiaal-probleem voor een deeltje op de halflijn $r \geq 0$, met een effectieve potentiaal

$$U_{eff} = \frac{\ell^2}{2r^2} - \frac{GM}{r}. \tag{1.18}$$

Voor radiële beweging: $\ell = 0$, bereikt een inkomend deeltje altijd de $1/r$-singulariteit in de oorsprong, en is er een centrale botsing. Echter, voor $\ell \neq 0$ gaat de effectieve potentiaal bij kleine afstanden als ℓ^2/r^2 naar $+\infty$, zodat de oorsprong nooit bereikt wordt. De afgeleide van de potentiaal

$$\frac{dU_{eff}}{dr} = -\frac{\ell^2}{r^3} + \frac{GM}{r^2}, \tag{1.19}$$

heeft nulpunten voor $r \to \infty$, waar de potentiaal zelf naar nul toegaat en het deeltje in essentie vrij is, en voor de eindige waarde

$$r_0 = \frac{\ell^2}{GM}, \qquad U_{eff}(r_0) = -\frac{GM}{2\ell^2}. \tag{1.20}$$

Er zijn dus gebonden toestanden mogelijk met $-GM/2\ell^2 \leq \epsilon < 0$, waarbij de stabielste toestand een cirkelbeweging is, met als straal $r = r_0$. Meer in het algemeen geldt voor zulke banen met negatieve bindingsenergie ϵ dat de radiële beweging wordt gekenmerkt door een minimale afstand r_- (het perihelium of periastron) en een maximale afstand r_+ (het apohelium of apastron). Deze afstanden worden gevonden uit de oplossing van de kwadratische vergelijking

$$2\epsilon r_\pm^2 + 2GMr_\pm - \ell^2 = 0. \tag{1.21}$$

De vorm van de gesloten banen is een ellips. Dit is het makkelijkst na te gaan door de straal r als functie van de hoek φ op te lossen. Hiertoe combineren we de vergelijking voor energiebehoud (1.17) met die voor het impulsmoment (1.14), via

$$\dot{r} = \dot{\varphi}\,\frac{dr}{d\varphi} = \frac{\ell}{r^2}\,\frac{dr}{d\varphi} = -\,\ell^2\frac{d}{d\varphi}\frac{1}{r}. \tag{1.22}$$

We kunnen nu vgl.(1.17) herschrijven in de vorm

$$\left(\frac{d}{d\varphi}\frac{1}{r}\right)^2 = -\left(\frac{1}{r} - \frac{1}{r_0}\right)^2 + \frac{2\epsilon}{\ell^2} + \frac{1}{r_0^2}. \tag{1.23}$$

Hierin is r_0 de eerder gevonden waarde van r in het minimum van U_{eff}. De oplossing van deze vergelijking is

$$r = \frac{r_0}{1 + e\cos(\varphi - \varphi_0)}, \qquad e^2 = 1 + \frac{2r_0^2}{\ell^2}\,\epsilon. \tag{1.24}$$

Zonder verlies aan algemeenheid kunnen we de x-as zo richten dat $\varphi_0 = 0$. Voor negatieve ϵ is $e^2 < 1$ en de vergelijking beschrijft een ellips met eccentriciteit e; r_0 is de lengte van de voerstraal vanuit het brandpunt loodrecht op de x-as. De halve lange en korte as zijn in termen hiervan gegeven door

$$a = \frac{r_0}{1 - e^2}, \qquad b = \frac{r_0}{\sqrt{1 - e^2}}. \tag{1.25}$$

In termen van euclidische coördinaten

$$x = r\cos\varphi, \qquad y = r\sin\varphi, \tag{1.26}$$

kan de baanvergelijking (1.24) ook herschreven worden als

$$\frac{(x + ea)^2}{a^2} + \frac{y^2}{b^2} = 1. \tag{1.27}$$

Dit is de standaardvergelijking voor een ellips met middelpunt $x = -ea$, $y = 0$, waarmee Keplers kegelsnedenwet bewezen is. Uit vergelijking (1.24) volgt verder, dat voor $\epsilon \geq 0$ de eccentriciteit in het gebied $e \geq 1$ ligt, zodat dezelfde vergelijking dan een parabolische of hyperbolische baan beschrijft.

Tenslotte leiden we de derde wet van Kepler af, die de omlooptijd in verband brengt met de lange as van de ellips. Eerst merken we op, dat in een periode T de voerstraal precies eenmaal het hele ellipsoppervlak doorloopt. Dus met vgl.(1.15)

$$\begin{aligned} T &= \frac{2A}{\ell} = 2\pi\,\frac{ab}{\ell} \\ &= 2\pi\,\frac{ar_0}{\sqrt{GMr_0(1 - e^2)}} = 2\pi\,\frac{a^{3/2}}{\sqrt{GM}}. \end{aligned} \tag{1.28}$$

Kwadrateren van deze vergelijking geeft dan

$$T^2 = \frac{4\pi^2}{GM}\,a^3. \tag{1.29}$$

In de bijgaande tabel hebben we de waarde van a^3/T^2 uitgezet voor de planeten in ons zonnestelsel. Daaruit blijkt dat tot op zeer goede benadering voor bijna alle planeten

$$\frac{GM}{4\pi^2} = \frac{a^3}{T^2} = 3.366 \times 10^{18}\,\mathrm{m^3 s^{-2}}. \tag{1.30}$$

De enige uitzondering is Pluto, met een afwijking van ongeveer 3%. We trekken hieruit de volgende conclusies:
- De baan van bijna alle planeten wordt alleen bepaald door de zwaartekracht van de zon; slechts Pluto, de lichtste planeet en het verst van de zon, ondergaat enige andere invloeden (zoals die van Neptunus); Jupiter en Saturnus beinvloeden elkaar zwak (afwijkingen van minder dan 1%).
- De totale massa in ieder twee-lichamen systeem (zon + planeet) is dezelfde, en mag worden geidentificeerd met de massa van de zon; de planeten zijn zoveel lichter dat hun eigen massa op het niveau van 1% nauwkeurigheid verwaarloosbaar is.
- Voor de zon vinden we dat

$$R_{S\odot} \equiv \frac{2GM_\odot}{c^2} = 2.953\,\mathrm{km}. \tag{1.31}$$

Deze karakteristieke dimensie wordt de Schwarzschildstraal genoemd. Bij een waarde van $G = 6.673 \times 10^{-11}$ kg^{-1} m^3 s^{-2} voor Newtons constante volgt dan voor de massa van de zon dat $M_\odot = 1.989 \times 10^{30}$ kg. Keplers wetten stellen ons dus in staat de zon te wegen, met als enige kennis de waarde van Newtons constante en de belangrijkste parameters van een of meer planetenbanen. Evenzo kan men uit de beweging van aardsatellieten de massa van de aarde bepalen. Deze bedraagt (bij dezelfde waarde van G): $M_\oplus = 5.97 \times 10^{24}$ kg. Dit is inderdaad in eerste benadering verwaarloosbaar t.o.v. de massa van de zon in de berekening van de aardbaan; zozeer zelfs, dat het massamiddelpunt van het stelsel aarde-zon diep in het binnenste van de zon ligt, op ongeveer 450 km van het middelpunt van de zon zelf.

planeet	a (10^{11} m)	T (10^7 sec)	a^3/T^2 (10^{18} m^3 s^{-2})
Mercurius	0.5791	0.7595	3.366
Venus	1.082	1.940	3.366
Aarde	1.496	3.154	3.365
Mars	2.279	5.932	3.366
Jupiter	7.783	37.41	3.369
Saturnus	14.28	92.90	3.370
Uranus	28.69	264.9	3.366
Neptunus	44.98	519.7	3.370
Pluto	58.65	783.3	3.287

1.2 Zware en trage massa

In de hoofdwet van de newtonse mechanica wordt de massa ingevoerd als de evenredigheidsconstante die vastlegt welke kracht nodig is om een lichaam een bepaalde versnelling te geven. Deze grootheid wordt daarom de *trage* massa genoemd. Uit de vorm van de zwaartekrachtwet (1.1) blijkt de massa van een lichaam nog een andere rol te spelen: als de grootheid die de sterkte van de zwaartekracht bepaalt die op een ander lichaam wordt uitgeoefend. Deze grootheid wordt daarom de *zware* massa genoemd. Dat de zware en trage massa evenredig met elkaar zijn —en bij een voor de hand liggende keuze van eenheden aan elkaar gelijkgesteld kunnen worden— is een empirisch feit; het vertelt ons iets over het karakter van de zwaartekracht.

Aangezien de actie en reactiekracht tussen twee lichamen in grootte aan elkaar gelijk zijn, is niet alleen de *door* een lichaam uitgeoefende zwaartekracht evenredig met zijn massa, maar ook de door de andere massa *op* dit lichaam uitgeoefende zwaartekracht. Bijgevolg is de versnelling die een lichaam met massa m krijgt onder invloed van de zwaartekracht van een ander lichaam met massa M alleen afhankelijk van die andere massa:

$$F = mg = -G\frac{mM}{r^2} \quad \Rightarrow \quad g = -\frac{GM}{r^2}. \tag{1.32}$$

De versnelling g t.g.v. de zwaartekracht is onafhankelijk van de eigen massa m. Merk op dat dit een groot verschil is met andere krachten, zoals de elektrische kracht waarbij de versnelling die een lichaam ondergaat evenredig is met zowel het elektrisch veld als met de eigen elektrische lading.

De gravitationele veldsterkte, uitgedrukt door de versnelling g, is evenredig met het produkt GM. De zware massa van de bron kan dus niet absoluut uit de valversnelling worden bepaald. Door M gelijk te stellen aan de trage massa van de bron leggen we de numerieke waarde van Newtons constante G in een gekozen systeem van eenheden vast. De geaccepteerde waarde van deze constante in mks-eenheden, gemeten in het laboratorium, is[1]

$$G = (6.672\,6 \pm 0.000\,9) \times 10^{-11}\,\mathrm{kg}^{-1}\ \mathrm{m}^3\ \mathrm{s}^{-2}. \tag{1.33}$$

De experimentele bepaling van G wordt bemoeilijkt door de zwakte van de zwaartekracht in vergelijking met andere krachten (de kleine waarde van G), en doordat —in tegenstelling tot het geval van elektrische krachtmetingen— uitwendige zwaartekrachtvelden niet afgeschermd kunnen worden: er bestaat niet zoiets als een gravitationele 'kooi van Faraday'. Dat blijkt niet alleen uit het relatief kleine aantal decimalen van G dat bekend is, maar ook uit de grote spreiding in gemeten waarden van G. De standaardwaarde (1.33) is gemeten in moderne torsie-balans experimenten, gebaseerd op hetzelfde principe als dat van de experimenten van Cavendish aan het eind van de 18e eeuw. Recentere experimenten met een andere aanpak geven soms sterk afwijkende resultaten. Een voorbeeld is het experiment van Meyer c.s. aan de universiteit van Wuppertal[2],

[1] Codata, 1996

[2] A. Schumacher et al., XXXII Rencontres de Moriond, 1997

waarin twee massa's worden gebruikt die de wanden van een trilholte voor microgolfstraling vormen. Veranderingen in de afstand tussen de massa's worden gemeten als verschuivingen in de resonantiefrequentie van de trilholte. De beste waarde van deze metingen tot heden (najaar 1998) gepubliceerd is $6.663\,7 \times 10^{-11}$ in mks-eenheden, met een statistische fout die veel kleiner is dan het verschil met de Codata waarde (1.33).

De equivalentie van zware en trage massa is proefondervindelijk tot op veel meer decimalen vastgesteld dan de waarde van G. Het eerste experiment dat in de literatuur bekend is, werd uitgevoerd door Simon Stevin en Jan de Groot[3]. Zij lieten twee loden kogels van verschillende massa (de een tien keer zo zwaar als de andere) samen van 30 voet hoog op een houten bord vallen, en stelden op grond van het geluid vast dat de twee kogels gelijktijdig het bord bereikten. Soorgelijke experimenten zijn later vele malen en in verbeterde versies herhaald. Newton experimenteerde zelf met slingers van verschillend materiaal, waarvan hij de slingertijd bepaalde, die binnen de (niet geweldige) meetnauwkeurigheid voor verschillende slingers onafhankelijk van de massa bleek te zijn. Beroemd zijn ook de proeven van baron von Eötvös, die met behulp van een torsiebalans probeerde differentiële variaties in de verhouding van de verticale zwaartekracht en horizontale centrifugale (schijn)kracht op verschillende materialen vast te stellen. Modernere versies van dit experiment[4] laten zien dat de verschillen in valversnelling voor verschillende materialen kleiner zijn dan 1 op 10^{11}.

Samenvatting van hoofdstuk 1

In het eerste hoofdstuk hebben we laten zien, dat Newtons wet van de zwaartekracht de drie wetten van Kepler voor de banen van planeten (en andere hemellichamen) verklaart. Met behulp van de in het laboratorium gemeten constante van Newton (G) kunnen we uit de parameters van de planetenbanen de massa van de zon bepalen. Voorts hebben we vastgesteld dat er twee begrippen massa zijn: *trage* massa en *zware* massa; Newtons wet van de zwaartekracht is in overeenstemming met het beginsel dat deze twee evenredig met elkaar zijn.

[3]S. Stevin, *Beginselen der Weeghconst*, 1586.

[4]P. Roll, R. Krotkov en R. Dicke, Ann. Phys. 26 (1964), 442
V. Braginsky en V. Panov, JETP 34 (1971), 464

Hoofdstuk 2

Relativiteit

2.1 Inertiaalstelsels

Een inertiaalstelsel is een (cartesisch te kiezen en fysisch construeerbaar) stelsel van ruimte- en tijdcoördinaten waarin het traagheidsbeginsel geldt: *alle* vrije deeltjes zijn t.o.v. elkaar in rust of bewegen eenparig langs een rechte lijn. Eenparig wil dan zeggen: in gelijke tijden worden gelijke afstanden afgelegd. De mogelijkheid zulke inertiaalstelsels te construeren staat niet bij voorbaat vast, maar zij blijken in de praktijk op zijn minst in eindige gebieden van de lege ruimte door het gebruik van b.v. meetlatten en klokken goed te realiseren. Daarmee is dus iets vastgesteld over de mechanische kwaliteiten van de lege ruimte.

Natuurlijk is het altijd mogelijk een coördinatenstelsel te kiezen waarin een gegeven deeltje in rust is. Ook kan de tijdcoördinaat altijd zo gekozen worden, dat de rechtlijnige beweging van een tweede deeltje t.o.v. het eerste eenparig is. Dat in de afwezigheid van krachtvelden alle andere deeltjes in zo'n coördinatenstelsel ook in rust zijn of eenparig rechtlijnig bewegen is echter een *fysisch* inzicht, dat ten grondslag ligt aan de klassieke mechanica.

Inertiaalstelsels zijn niet uniek: als twee cartesische stelsels van ruimte en tijdcoördinaten door een lineaire transformatie verbonden zijn, en een van de twee is een inertiaalstelsel, dan is de ander dat ook. We vervangen eerst de tijdcoördinaat t door een lengtecoördinaat $x^0 = ct$, met c de lichtsnelheid; x^0 heeft dus de dimensie van een lengte. Zij nu $\{x^\mu\}_{\mu=0}^3 = (ct, x, y, z)$ een inertiaalstelsel. Alle vrije deeltjes, te onderscheiden door een label ($i = 1, ..., N$), beschrijven in dit systeem banen van de vorm

$$\mathbf{r}_i(t) = \mathbf{r}_i(0) + \mathbf{v}_i t. \tag{2.1}$$

De snelheden $\mathbf{v}_i = d\mathbf{r}_i/dt$ zijn constant in grootte en richting. Zij nu $\{x'^\mu\}_{\mu=0}^3$ een tweede stelsel coördinaten, die zijn uit te drukken in de oorspronkelijke coördinaten d.m.v. een stel functies $x'^\mu = f^\mu(x)$. In de nieuwe coördinaten zijn de snelheden van de deeltjes, uitgedrukt als fractie van de lichtsnelheid:

$$\begin{aligned}\frac{\mathbf{v}'_i}{c} &= \frac{d\mathbf{r}'_i}{cdt'} = \frac{(d\mathbf{r}_i \cdot \nabla)\mathbf{f}_i + dt(d\mathbf{f}/dt)_i}{(d\mathbf{r}_i \cdot \nabla)f_i^0 + dt(df^0/dt)_i} \\ &= \frac{(\mathbf{v}_i \cdot \nabla)\mathbf{f}_i + (d\mathbf{f}/dt)_i}{(\mathbf{v}_i \cdot \nabla)f_i^0 + (df^0/dt)_i}\end{aligned} \tag{2.2}$$

Als de nieuwe coördinaten ook een inertiaalstelsel definiëren, moeten de snelheden $\mathbf{v}'_i$ ook allemaal constant zijn, ongeacht het aantal deeltjes (N). Dat is alleen mogelijk als de afgeleiden $\partial f^\mu/\partial x^\nu$, $(\mu,\nu) = 0,...,3$, zelf alle constant zijn, en de $f^\mu(x)$ dus lineaire functies van de oorspronkelijke coördinaten x^μ zijn:

$$x'^\mu = \sum_{\nu=0}^{3} \Lambda^\mu{}_\nu x^\nu + a^\mu, \tag{2.3}$$

met constante coëfficienten $\Lambda^\mu{}_\nu$ en verschuivingen a^μ.

2.2 De lichtsnelheid

De lichtsnelheid is een universele constante, die in alle inertiaalstelsels in dezelfde eenheden dezelfde waarde heeft. In mks-eenheden is dit $c = 2.9979 \times 10^8$ m/s; in eenheden waarin tijd wordt gemeten in seconden en afstanden in lichtseconden, is dit $c = 1$. De universaliteit van de lichtsnelheid is de grondslag voor de relativiteitstheorie zoals die door Einstein is opgesteld, voortbouwend op eerdere analyses van de electrodynamica door Lorentz en Fitzgerald. Een experimentele aanwijzing voor dit beginsel komt o.a. uit het beroemde experiment van Michelson en Morley, dat er niet in slaagde de etherdrift van het licht aan te tonen, d.w.z. het verschil in snelheid van het licht dat met de aardse baanbeweging meebeweegt en van licht dat zich in tegengestelde richting voortplant.

We definiëren het Minkowski-interval tussen twee gebeurtenissen met tijdruimtecoördinaten $\mathcal{P}_1 = (t_1, x_1, y_1, z_1)$ en $\mathcal{P}_2 = (t_2, x_2, y_2, z_2)$ als

$$(\Delta s)^2 = (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 - c^2(\Delta t)^2 = \sum_{\mu,\nu=0}^{3} \eta_{\mu\nu} \Delta x^\mu \Delta x^\nu, \tag{2.4}$$

waarin $\Delta x^\mu = x_2^\mu - x_1^\mu$, en waarin $\eta_{\mu\nu}$ de Minkowski-metriek is, met cartesische componenten $\eta_{\mu\nu} = \mathrm{diag}(-1,+1,+1,+1)$. In het vervolg zullen we de expliciete vermelding van de sommatie over de indices μ en ν achterwege laten, en de sommatieconventie gebruiken die zegt dat —tenzij anders vermeld— over iedere index die twee keer voorkomt in een produkt van vector of tensorcomponenten automatisch wordt gesommeerd.

Als twee punten $\mathcal{P}_1$ en $\mathcal{P}_2$ door een lichtstraal verbonden kunnen worden, moet gelden dat

$$(\Delta \mathbf{r})^2 = c^2(\Delta t)^2, \tag{2.5}$$

en dus is in dat geval $(\Delta s)^2 = 0$. Alle punten $\mathcal{P}_2$ in de tijd-ruimte die met een gegeven punt $\mathcal{P}_1$ door lichtstralen verbonden zijn, vormen samen de lichtkegel van het punt $\mathcal{P}_1$; de lichtkegel bestaat dus uit de verzameling punten die voldoet aan $(\Delta s)^2 = 0$.

Aangezien de lichtsnelheid hetzelfde is in alle inertiaalstelsels, moet de vergelijking (2.5) voor de lichtkegel in al die inertiaalstelsels gelden. Inertiaalstelsels zijn echter met elkaar verbonden door coördinatentransformaties van de vorm (2.3). Daarom houdt het postulaat van de universele lichtsnelheid in, dat de transformaties (2.3) tussen inertiaalstelsels beperkt moeten worden tot die transformaties die de lichtkegel invariant laten. Merk op, dat de translaties a^μ automatisch ieder interval Δx^μ invariant laten, omdat x_1^μ en x_2^μ met hetzelfde bedrag veranderen. De universaliteit van de lichtsnelheid levert dus alleen een beperking op de matrix van coëfficienten $\Lambda^\mu{}_\nu$:

$$(\Delta s')^2 \;=\; \eta_{\mu\nu}\Delta x'^\mu \Delta x'^\nu \;=\; \eta_{\mu\nu}\Lambda^\mu{}_\kappa \Lambda^\nu{}_\lambda \Delta x^\kappa \Delta x^\lambda. \tag{2.6}$$

Dit vergelijken we nu met het oorspronkelijke interval

$$(\Delta s)^2 \;=\; \eta_{\kappa\lambda}\Delta x^\kappa \Delta x^\lambda \;=\; 0. \tag{2.7}$$

De universaliteit van de lichtsnelheid houdt dan in, dat

$$\eta_{\mu\nu}\Lambda^\mu{}_\kappa \Lambda^\nu{}_\lambda = a^2 \eta_{\kappa\lambda}, \tag{2.8}$$

met a^2 een willekeurige constante. Voor het speciale geval $a^2 = 1$ is niet alleen de lichtkegel invariant, maar ieder Minkowski-interval: $(\Delta s')^2 = (\Delta s)^2$, ook voor $\Delta s \neq 0$. Zulke speciale lineaire transformaties noemt men Lorentztransformaties. Preciezer gezegd, de transformaties die de oorsprong op zijn plaats laten (dus met $a^\mu = 0$) heten de eigenlijke of homogene Lorentztransformaties; de transformaties die de oorsprong niet invariant laten ($a^\mu \neq 0$) worden inhomogene Lorentztransformaties genoemd. Naast de Lorentztransformaties omvat (2.8) ook de schaaltransformaties $x^\mu \to a x^\mu$. Echter, waar de Lorentztransformaties ook met de dynamica van deeltjes verenigbaar zijn (het totale kwadraat van de vierimpuls is behouden; zie hierna), geldt dit niet voor schaaltransformaties (het totale kwadraat van de vier-impuls is niet nul). Daarom beschouwen we verder alleen de Lorentztransformaties, en dan voornamelijk de homogene Lorentztransformaties.

2.3 Lorentztransformaties

Uit vergelijking (2.8) met $a^2 = 1$ volgt dat

$$\eta_{\mu\nu}\Lambda^\mu{}_\kappa \Lambda^\nu{}_\lambda = \eta_{\kappa\lambda}. \tag{2.9}$$

De lineaire transformaties die hieraan voldoen zijn de homogene Lorentztransformaties. We kunnen dit ook anders schrijven, door op te merken dat de Minkowski-metriek als 4×4 matrix in cartesische coördinaten zijn eigen inverse is. We noteren de inverse metriek algemeen door de indices boven te schrijven i.p.v. onder; dus

$$\eta^{\mu\lambda}\eta_{\lambda\nu} = \delta^{\mu}_{\nu}, \tag{2.10}$$

hetgeen in cartesische coördinaten leidt tot $\eta^{\mu\nu} = \mathrm{diag}(-1, +1, +1, +1)$. Met deze notatie kan vgl.(2.9) worden herschreven als

$$\eta_{\mu\nu}\Lambda^{\nu}{}_{\lambda}\eta^{\lambda\tau}\Lambda^{\mu}{}_{\kappa} = \delta^{\tau}_{\kappa}. \tag{2.11}$$

Daaruit vinden we voor de inverse van de transformatie $\Lambda^{\mu}{}_{\nu}$:

$$(\Lambda^{-1})^{\tau}{}_{\mu} = \eta_{\mu\nu}\Lambda^{\nu}{}_{\lambda}\eta^{\lambda\tau} \equiv \Lambda_{\mu}{}^{\tau}. \tag{2.12}$$

In de laatste stap hebben we de conventie ingevoerd dat de Minkowski-metriek en zijn inverse worden gebruikt om indices van boven naar beneden te verplaatsen en omgekeerd. Vgl.(2.12) kan worden opgevat als de wiskundige definitie van de homogene Lorentztransformaties. We bekijken nu twee bijzondere gevallen om enig idee te krijgen van het fysisch karakter van Lorentztransformaties.

Rotaties. Een onderklasse van de Lorentztransformaties zijn de gewone rotaties. Deze transformeren alleen de ruimtecoördinaten, en niet de tijd. Zij zijn dus van de algemene vorm

$$\Lambda = \begin{pmatrix} 1 & 0 \\ 0 & R_{ij} \end{pmatrix}, \tag{2.13}$$

met R_{ij} de componenten van een 3×3 matrix $\mathbf{R}$. Zulke transformaties laten het Minkowski-interval $(\Delta s)^2$ invariant als ze het gewone 3-dimensionale interval $(\Delta \mathbf{r})^2$ invariant laten. De meest algemene transformatie die dat doet is een combinatie van gewone rotaties en spiegelingen.

Wiskundig zien we dit als volgt: vgln.(2.10) en (2.12) gaan voor transformaties van de vorm (2.13) over in

$$R_{ik}R_{j}{}^{k} = \delta_{ij} \qquad \Leftrightarrow \qquad \mathbf{R}^{-1} = \mathbf{R}^{T}. \tag{2.14}$$

In de indexvrije matrixnotatie staat het superscript T voor transpositie. Vergelijking (2.14) is de definitie van een orthogonale matrix, d.w.z. een matrix die de lengte van een vector invariant laat: als $\mathbf{r}' = \mathbf{R} \cdot \mathbf{r}$, dan is

$$\mathbf{r}'^{\,2} = \mathbf{r}^{T} \cdot \mathbf{R}^{T} \cdot \mathbf{R} \cdot \mathbf{r} = \mathbf{r}^{T} \cdot \mathbf{R}^{-1} \cdot \mathbf{R} \cdot \mathbf{r} = \mathbf{r}^{2}. \tag{2.15}$$

Voor transformaties met $\det \mathbf{R} = 1$ zijn dit de gewone 3-dimensionale rotaties van een vector in de euclidische ruimte; als $\det \mathbf{R} = -1$, dan is het een spiegeling, al dan niet met een draaiing eraan toegevoegd.

Speciale Lorentztransformaties. De complementaire klasse van Lorentz-transformaties zijn de lineaire transformaties tussen ruimte- en tijdcoördinaten. Aangezien we altijd eerst een rotatie kunnen uitvoeren (deze laat immers $(\Delta s)^2$ invariant) kunnen we ons beperken tot de transformaties tussen t en x; we beschouwen daarom het geval $\Delta y' = \Delta y$ en $\Delta z' = \Delta z$, zodat we als eis overhouden

$$(\Delta x')^2 - c^2(\Delta t')^2 = (\Delta x)^2 - c^2(\Delta t)^2. \tag{2.16}$$

De algemene vorm van homogene lineaire transformaties met deze eigenschap is

$$\begin{aligned} x' &= \frac{x - vt}{\sqrt{1 - v^2/c^2}}, \\ t' &= \frac{t - vx/c^2}{\sqrt{1 - v^2/c^2}}. \end{aligned} \tag{2.17}$$

Deze transformaties zijn fysisch gerealiseerd door twee gelijkgeoriënteerde inertiaalstelsels waarvan de een met een constante snelheid in de x-richting beweegt t.o.v. de ander: $\mathbf{v} = (v, 0, 0)$. Dit is eenvoudig te zien door op te merken dat de oorsprong van het tweede stelsel, $x' = 0$, overeenkomt met het eenparig bewegende punt $x = vt$ in het eerste stelsel. Voor een klok meebewegend met dit punt geldt, dat

$$t' = t\sqrt{1 - v^2/c^2}. \tag{2.18}$$

Anderzijds geldt voor een klok in rust in de oorsprong van het eerste stelsel, dat $x = 0$, zodat

$$t = t'\sqrt{1 - v^2/c^2}, \tag{2.19}$$

en

$$x' = -vt'. \tag{2.20}$$

De symmetrie tussen de stelsels is dus compleet. In het bijzonder ziet men in beide stelsels een klok in het andere systeem vertraagd lopen, met dezelfde vertraging.

2.4 Viervectoren

De componenten van een vier-dimensionale ruimte-tijd vector Δx^μ veranderen in een ander inertiaalstelsel in componenten $\Delta x'^\mu$, die worden verkregen door een Lorentztransformatie

$$\Delta x'^\mu = \Lambda^\mu{}_\nu \Delta x^\nu. \tag{2.21}$$

De grootheid

$$(\Delta s)^2 = \eta_{\mu\nu} \Delta x^\mu \Delta x^\nu \tag{2.22}$$

is echter onder deze overgang invariant. We definiëren nu in overeenstemming met eerdere afspraken een andere vier-dimensionale vector door $\Delta x_\mu \equiv \eta_{\mu\nu}\Delta x^\nu$; dan is

$$(\Delta s)^2 = \Delta x_\mu \Delta x^\mu. \tag{2.23}$$

Uit de eigenschap (2.12) van de Lorentztransformaties volgt, dat

$$\begin{aligned} \Delta x'_\mu &= \eta_{\mu\nu}\Lambda^\nu{}_\lambda \Delta x^\lambda \\ &= \eta_{\mu\nu}\Lambda^\nu{}_\lambda \eta^{\lambda\kappa}\Delta x_\kappa = \Lambda_\mu{}^\kappa \Delta x_\kappa \\ &= \Delta x_\kappa (\Lambda^{-1})^\kappa{}_\mu. \end{aligned} \tag{2.24}$$

De invariantie van $(\Delta s)^2$ als in vgl.(2.23) is daarmee manifest.

We voeren nu enige nieuwe terminologie in.
- Een grootheid zoals $(\Delta s)^2$, die niet verandert onder de overgang van een inertiaalstelsel naar een ander inertiaalstelsel, noemen we een *scalar*.
- Een grootheid met vier componenten a^μ, die onder de overgang tussen inertiaalstelsels veranderen volgens de Lorentztransformatie

$$a'^\mu = \Lambda^\mu{}_\nu a^\nu, \tag{2.25}$$

heet een *contravariante viervector*.
- Een grootheid met vier componenten a_μ, die onder de overgang tussen inertiaalstelsels veranderen volgens de inverse Lorentztransformatie

$$a'_\mu = \Lambda_\mu{}^\nu a_\nu = a_\nu (\Lambda^{-1})^\nu{}_\mu, \tag{2.26}$$

heet een *covariante viervector*.
Het verschil tussen contravariante en covariante viervectoren zit dus in de manier waarop ze zich onder Lorentztransformaties gedragen. Dit wordt uitgedrukt door de notatie met onder- en bovenindices voor, respectievelijk, de co- en contravariante componenten. Het volgt uit deze definities onmiddellijk, dat de *contractie* van covariante en contravariante viervectoren a en b, gedefiniëerd door

$$a \cdot b = a_\mu b^\mu = \eta_{\mu\nu} a^\nu b^\mu, \tag{2.27}$$

invariant is; het inproduct van twee viervectoren, gedefiniëerd door $a \cdot b$, is dus een scalar.

Behalve covariante en contravariante vectoren bestaan er ook covariante en contravariante tensoren. Dit zijn grootheden waarvan de componenten transformeren als directe produkten van viervectoren. Bij voorbeeld, een contravariante tensor van rang 2 is een grootheid met componenten $T^{\mu\nu}$ die onder de overgang tussen inertiaalstelsels transformeren in

$$T'^{\mu\nu} = \Lambda^\mu{}_\kappa \Lambda^\nu{}_\lambda T^{\kappa\lambda}. \tag{2.28}$$

Op vergelijkbare manier heeft een covariante tensor van rang 2 componenten $T_{\mu\nu}$ met de transformatieëigenschap

$$T'_{\mu\nu} = \Lambda_\mu{}^\kappa \Lambda_\nu{}^\lambda T_{\kappa\lambda}. \tag{2.29}$$

Ook tensoren met gemengd covariante en contravariante componenten zijn mogelijk, b.v. door de Minkowski-metriek te gebruiken: als $T_\mu{}^\nu = \eta_{\mu\lambda} T^{\lambda\nu} = \eta^{\nu\lambda} T_{\mu\lambda}$, dan transformeren de componenten als

$$T'_\mu{}^\nu = \Lambda_\mu{}^\kappa \Lambda^\nu{}_\lambda T_\kappa{}^\lambda. \tag{2.30}$$

Met deze conventies zijn contracties als b.v.

$$a \cdot T \cdot b \;=\; a^\mu \, T_\mu{}^\nu \, b_\nu, \tag{2.31}$$

manifest invariant onder Lorentztransformaties.

2.5 Eigentijd en viersnelheid

Het Minkowski-interval

$$(\Delta s)^2 = (\Delta \mathbf{r})^2 - c^2 (\Delta t)^2 \tag{2.32}$$

is invariant onder Lorentztransformaties. Er zijn drie soorten intervallen:
- Ruimteachtige intervallen met $(\Delta s)^2 > 0$; voor zulke intervallen is het altijd mogelijk een coördinatenstelsel te vinden waarin $(\Delta s)^2 = (\Delta \mathbf{r})^2$, en dus $\Delta t = 0$. In dit coördinatenstelsel zijn de gebeurtenissen $\mathcal{P}_1$ en $\mathcal{P}_2$ gelijktijdig.
- Lichtachtige intervallen met $(\Delta s)^2 = 0$. Deze verbinden gebeurtenissen die op elkaars voorwaartse/achterwaartse lichtkegel liggen, met $(\Delta \mathbf{r})^2 = c^2 (\Delta t)^2$.
- Tijdachtige intervallen met $(\Delta s)^2 < 0$; Voor zulke intervallen kunnen we altijd een coördinatenstelsel vinden waarin $(\Delta s)^2 = -c^2 (\Delta t)^2$, en dus $\Delta \mathbf{r} = 0$. Dit is het rustsysteem, waarin de gebeurtenissen $\mathcal{P}_1$ en $\mathcal{P}_2$ zich afspelen in hetzelfde punt van de ruimte. In dit laatse geval komt het interval dus overeen met c maal de tijd in het rustsysteem. Dit invariante tijdinterval noemen we de *eigentijd* $\Delta\tau$. Voor tijdachtige intervallen definiëren we dus

$$c^2 (\Delta\tau)^2 = -(\Delta s)^2 = c^2 (\Delta t)^2 - (\Delta \mathbf{r})^2. \tag{2.33}$$

Wanneer $\mathbf{r}$ de positie van een deeltje is, dan is $\Delta\mathbf{r}$ de door dit deeltje in Δt afgelegde afstand, en we krijgen

$$\Delta\tau = \Delta t \sqrt{1 - \frac{\mathbf{v^2}}{c^2}}, \tag{2.34}$$

met $\mathbf{v} = d\mathbf{r}/dt$ de snelheid van het deeltje in het coördinatenstelsel $(t, \mathbf{r})$.

De beweging van een puntdeeltje wordt beschreven door de baan die het deeltje aflegt in ruimte en tijd; in de context van de relativistische beschrijving heet deze baan de wereldlijn van het deeltje. De punten van de wereldlijn kunnen worden gemerkt door de tijd te noteren waarop het deeltje op dat punt is, gemeten door een klok ter plekke van het deeltje die momentaan t.o.v. het deeltje in rust is. Zo kan de wereldlijn geparameteriseerd worden m.b.v. de eigentijd, en de baan van het deeltje wordt dan vastgelegd door de functies $x^\mu(\tau)$ te specificeren die de positie van het deeltje in de Minkowskiruimte geven als functie van de eigentijd.

Aangezien $\Delta\tau$ invariant is onder Lorentztransformaties, terwijl Δx^μ transformeert als een contravariante viervector, is de raakvector aan de wereldlijn van een deeltje

$$u^\mu = \frac{dx^\mu}{d\tau} \tag{2.35}$$

een contravariante vector. Het is de snelheid van het deeltje in de Minkoswki-ruimte, gemeten m.b.v. een klok die t.o.v. het deeltje momentaan in rust is. Deze snelheid wordt gemeenlijk de viersnelheid van het deeltje genoemd. Omdat de viersnelheid (2.35) contravariant is, zijn de componenten van de viersnelheid in een ander inertiaalstelsel te vinden door de Lorentztransformatie

$$u'^\mu = \Lambda^\mu{}_\nu u^\nu. \tag{2.36}$$

Uit de definitie van eigentijd volgt nu onmiddellijk, dat de viersnelheid invariant genormeerd is op de lichtsnelheid:

$$u_\mu u^\mu = \eta_{\mu\nu}\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau} = -c^2. \tag{2.37}$$

Deze conditie legt de vierde (=tijd) component van de viersnelheid vast in termen van de drie gewone snelheidscomponenten. Dit wordt duidelijk door gebruik te maken van de relatie (2.34) en de componenten van de viersnelheid te schrijven als

$$u^\mu = \left(c\frac{dt}{d\tau}, \frac{d\mathbf{r}}{d\tau}\right) = \left(\frac{c}{\sqrt{1-\mathbf{v^2}/c^2}}, \frac{\mathbf{v}}{\sqrt{1-\mathbf{v^2}/c^2}}\right). \tag{2.38}$$

Het is niet moeilijk te zien, dat dit een oplossing van de conditie (2.37) is.

De vierimpuls is te definiëren als het produkt van de invariante massa m (de rustmassa) en de viersnelheid van een deeltje; in contravariante componenten:

$$p^\mu = mu^\mu = \left(\frac{mc}{\sqrt{1-\mathbf{v^2}/c^2}}, \frac{m\mathbf{v}}{\sqrt{1-\mathbf{v^2}/c^2}}\right). \tag{2.39}$$

Uit de relaties

$$E = p^0 c = \frac{mc^2}{\sqrt{1-\mathbf{v^2}/c^2}}, \qquad \mathbf{p} = \frac{m\mathbf{v}}{\sqrt{1-\mathbf{v^2}/c^2}}, \tag{2.40}$$

volgt dan dat

$$E^2 = \mathbf{p}^2 c^2 + m^2 c^4. \tag{2.41}$$

Voor een vrij deeltje is de snelheid constant (een versnelling zou een kracht impliceren); dan is ook de vierimpuls constant, wat op een covariante manier geschreven kan worden als

$$\frac{dp^\mu}{d\tau} = 0. \tag{2.42}$$

Onder een Lorentztransformatie veranderen de componenten van de vierimpuls als

$$p'^{\,\mu} = \Lambda^\mu{}_\nu p^\nu. \tag{2.43}$$

De matrix van coëfficienten $\Lambda^\mu{}_\nu$ is zelf tijdonafhankelijk, zodat

$$\frac{dp'^{\,\mu}}{d\tau} = \Lambda^\mu{}_\nu \frac{dp^\nu}{d\tau} = 0. \tag{2.44}$$

Dus de bewegingsvergelijking (2.42) voor een vrij deeltje is geldig in ieder inertiaalstelsel. Wanneer er een kracht op het deeltje wordt uitgeoefend en er een versnelling optreedt, zal deze bewegingsvergelijking gewijzigd moeten worden naar een vergelijking van de vorm

$$\frac{dp^\mu}{d\tau} = m a^\mu = f^\mu. \tag{2.45}$$

In deze vergeljking is $a^\mu = du^\mu/d\tau$ de covariante vierversnelling, en f^μ een contravariante vier-vector die een generalisatie is van de Newtonse kracht in de niet-relativistische beschrijving. Als een viervector voor de kracht gevonden is, hebben we in vgl.(2.45) wederom een vergelijking die in ieder inertiaalsysteem van toepassing is, omdat het linker- en rechterlid van de vergelijking zich beide op dezelfde manier onder Lorentztransformaties gedragen. Een hulpmiddel om een uitdrukking voor de vierkracht f^μ te vinden is op te merken, dat

$$u_\mu a^\mu = \frac{1}{2}\frac{d}{d\tau}\left(u_\mu u^\mu\right) = \frac{1}{2}\frac{d}{d\tau}\left(-c^2\right) = 0. \tag{2.46}$$

Als viervectoren zijn de vierversnelling en de viersnelheid dus orthogonaal. Dat zelfde geldt dan ook voor de vierkracht:

$$u_\mu f^\mu = 0 \qquad \Rightarrow \qquad f^0 = \frac{\mathbf{u}\cdot\mathbf{f}}{u^0} = \frac{\mathbf{v}\cdot\mathbf{f}}{c}. \tag{2.47}$$

De tijdcomponent van de vierkracht ligt dus vast wanneer we de ruimtecomponenten $\mathbf{f}$ kennen.

Samenvatting van hoofdstuk 2

In dit hoofdstuk is het begrip *inertiaalstelsel* ingevoerd. In een inertiaalstelsel bewegen alle vrije deeltjes rechtlijnig en eenparig. Het speciale relativiteitsprincipe stelt, dat de lichtsnelheid een universele natuurconstante is. Hieruit volgt dat alle inertiaalstelsels met elkaar zijn verbonden door Lorentztransformaties. Deze transformaties tussen coördinatenstelsels in de tijd-ruimte (de vierdimensionale Minkowskiruimte) kunnen worden onderverdeeld in drie klassen: *translaties*, verschuivingen waarbij de orientatie van het stelsel niet verandert; *rotaties*, waaronder we zowel ruimtelijke draaiingen met als zonder spiegeling verstaan; en de speciale Lorentztransformaties die de ruimte- en tijd-coördinaten onderling mengen. Co- en contravariante viervectoren, waaronder de viersnelheid en vierimpuls, worden gekenmerkt door de manier waarop hun componenten onder Lorentztransformaties veranderen. Tenslotte bleek het begrip eigentijd een handig hulpmiddel bij het formuleren van covariante relativistische bewegingsvergelijkingen.

Hoofdstuk 3

Elektrodynamica

3.1 De Maxwell-vergelijkingen

De relativistische beschrijving van de beweging van deeltjes in wisselwerking met een krachtveld is in het bijzonder van toepassing op de elektrodynamica. In dit hoofdstuk geven we de covariante formulering van de Maxwell-vergelijkingen voor het elektromagnetische veld, en van de Lorentzkracht die de beweging van een geladen deeltje in zo'n veld bepaalt.

De inhomogene Maxwell-vergelijkingen leggen het verband tussen de ladingsdichtheid $\rho(x)$ en de stroomdichtheid $\mathbf{j}(x)$ als bronnen voor het elektrisch veld met veldsterkte $\mathbf{E}$, en het magnetisch veld met veldsterkte $\mathbf{B}$. Ook beschrijven ze de veranderingen van die velden in ruimte en tijd. De standaardvorm van de vergelijkingen is

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0}, \qquad \nabla \times \mathbf{B} - \frac{1}{c^2}\frac{\partial \mathbf{E}}{\partial t} = \frac{1}{\epsilon_0 c^2}\,\mathbf{j}. \tag{3.1}$$

De constante $\epsilon_0 = 8.854 \times 10^{-12}$ F m^{-1} is de dielectische constante van de lege ruimte. De homogene Maxwell-vergelijkingen leggen voorwaarden op aan de velden zelf; deze zijn van de vorm

$$\nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0, \qquad \nabla \cdot \mathbf{B} = 0. \tag{3.2}$$

De homogene vergelijkingen houden in, dat de velden uitgedrukt kunnen worden in termen van vier potentialen, een scalarpotentiaal ϕ en een vectorpotentiaal $\mathbf{A}$:

$$\mathbf{E} = -\nabla\phi - \frac{\partial \mathbf{A}}{\partial t}, \qquad \mathbf{B} = \nabla \times \mathbf{A}. \tag{3.3}$$

Deze vergelijkingen kunnen in een vier-dimensionale covariante vorm gegoten worden. Om dat te bereiken definiëren we de covariante gradiëntoperator

$$\partial_\mu = (\partial_0, \nabla) = (\frac{1}{c}\frac{\partial}{\partial t}, \nabla). \tag{3.4}$$

Verder definiëren we de vierstroom

$$j^{\mu} = (\rho c, \mathbf{j}) \,, \tag{3.5}$$

en de anti-symmetrische Maxwell-tensor

$$F^{\mu\nu} = -F^{\nu\mu} = \begin{pmatrix} 0 & \frac{1}{c}E_x & \frac{1}{c}E_y & \frac{1}{c}E_z \\ -\frac{1}{c}E_x & 0 & B_z & -B_y \\ -\frac{1}{c}E_y & -B_z & 0 & B_x \\ -\frac{1}{c}E_z & B_y & -B_x & 0 \end{pmatrix}. \tag{3.6}$$

Met deze definities kunnen de inhomogene Maxwell-vergelijkingen worden herschreven in de vorm

$$\partial_{\mu} F^{\mu\nu} = -\frac{1}{\epsilon_0 c^2}\, j^{\,\nu}, \tag{3.7}$$

en de homogene Maxwell-vergelijkingen kunnen worden samengevat door

$$\varepsilon^{\mu\nu\kappa\lambda}\, \partial_{\nu} F_{\kappa\lambda} = 0, \tag{3.8}$$

waarbij het vier-dimensionale permutatiesymbool als epsilon-tensor is geschreven, met componenten

$$\varepsilon^{\mu\nu\kappa\lambda} = \begin{cases} +1, & \text{voor } (\mu\nu\kappa\lambda) = \text{even permutatie van } (0123), \\ -1, & \text{voor } (\mu\nu\kappa\lambda) = \text{oneven permutatie van } (0123), \\ \;\;0, & \text{voor alle andere componenten.} \end{cases} \tag{3.9}$$

De epsilon-tensor is pseudo-invariant, d.w.z. hij heeft dezelfde componenten in alle rechtshandige of linkshandige inertiaalstelsels, maar wisselt van teken onder overgang van een rechts- naar een linkshandig stelsel. Als we namelijk een Lorentztransformatie uitvoeren, krijgen we

$$\varepsilon'^{\,\mu\nu\kappa\lambda} = \Lambda^{\mu}{}_{\rho} \Lambda^{\nu}{}_{\sigma} \Lambda^{\kappa}{}_{\tau} \Lambda^{\lambda}{}_{\theta}\, \varepsilon^{\rho\sigma\tau\theta} = \varepsilon^{\mu\nu\kappa\lambda} \det \Lambda. \tag{3.10}$$

Voor Lorentztransformaties zonder spiegelingen, die rechtshandige in rechtshandige coödinatenstelsels laten overgaan, is $\det \Lambda \equiv \det\,[\Lambda^{\mu}{}_{\nu}] = +1$, terwijl voor een transformatie met spiegeling —waarbij een rechtshandig in een linkshandig stelsel overgaat— $\det \Lambda = -1$. Dit volgt uit vergelijking (2.9), die na het nemen van de determinant aan beide zijden van de gelijkheid leidt tot

$$(\det \Lambda)^2 = 1. \tag{3.11}$$

De transformaties met $\det \Lambda = +1$ zijn continu met de eenheidstransformatie $\Lambda^{\mu}{}_{\nu} = \delta^{\mu}_{\nu}$ verbonden, waarvoor $t' = t$, $\mathbf{r}' = \mathbf{r}$. Die met $\det \Lambda = -1$ zijn continu verbonden met de ruimtelijke spiegeling $\mathbf{r}' = -\mathbf{r}$, of de tijdomkeer $t' = -t$.

Een oplossing van de homogene Maxwellvergelijking (3.8) is lokaal altijd te schrijven in termen van een viervector-potentiaal A_μ als

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu. \tag{3.12}$$

Na invulling in (3.8) levert dit immers, rekening houdend met de anti-symmetrie van de epsilon-tensor, waardoor beide termen in (3.12) hetzelfde resultaat leveren:

$$\varepsilon^{\mu\nu\kappa\lambda}\partial_\nu\partial_\kappa A_\lambda = 0. \tag{3.13}$$

Het resultaat is gegarandeerd door de eigenschap $\partial_\nu\partial_\kappa = \partial_\kappa\partial_\nu$.

De covariante vergelijking (3.12) is equivalent met de vergelijkingen (3.3), wanneer we de ruimtecomponenten van de viervector A_μ ($\mu = 1, 2, 3$) identificeren met $\mathbf{A}$, terwijl $A_0 = -\phi/c$. Voor de contravariante componenten is dan

$$A^\mu = (\frac{\phi}{c}, \mathbf{A}). \tag{3.14}$$

Hieruit volgt, dat bij de overgang tussen inertiaalstelsels de potentialen als de componenten van een viervector transformeren:

$$A'^\mu = \Lambda^\mu{}_\nu A^\nu. \tag{3.15}$$

Hoewel de Lorentztransformatie van een gegeven vectorpotentiaal dus eenduidig is, zijn de uitdrukkingen voor de componenten van de potentiaal zelf dit niet. Uit de algemene vorm (3.12) van de covariante veldsterktetensor volgt namelijk, dat deze invariant is onder een herdefinitie van de potentiaal van de vorm

$$A_\mu \to A'_\mu = A_\mu + \partial_\mu\Lambda, \tag{3.16}$$

voor willekeurige scalaire functies $\Lambda(x)$; immers,

$$\partial_\mu\partial_\nu\Lambda - \partial_\nu\partial_\mu\Lambda = 0. \tag{3.17}$$

De herdefinitie (3.16) heet een ijktransformatie. De eigenschap dat de veldsterktetensor onder een ijktransformatie niet verandert, staat bekend als de ijkinvariantie van deze tensor. De betekenis van de ijkinvariantie van de veldsterkten is erin gelegen, dat het aantal fysisch relevante componenten van de vectorpotentiaal wordt teruggebracht. In het vervolg komen we hier nog nader op terug.

3.2 De Lorentzkracht

De niet-relativistische uitdrukking voor de Lorentzkracht op een geladen deeltje is

$$\mathbf{F} = q\left(\mathbf{E} + \mathbf{v} \times \mathbf{B}\right). \tag{3.18}$$

Deze vergelijking beschrijft de invloed van elektrische en magnetische velden op de beweging van geladen deeltjes goed zolang ze een snelheid hebben die klein is t.o.v. de lichtsnelheid: $v/c \ll 1$. Een covariante vergelijking die de Lorentzkracht generaliseert, en in de niet-relativistische limiet naar vgl.(3.18) terugvoert, is

$$f^{\mu} = qF^{\mu}{}_{\nu}\, \dot{x}^{\nu}, \tag{3.19}$$

waarin $\dot{x}^{\mu} \equiv dx^{\mu}/d\tau$. Wanneer deze uitdrukking wordt ingevuld in vergelijking (2.45), en deze expliciet wordt uitgedrukt in componenten van de verschillende viervectoren en tensoren, vinden we

$$\begin{aligned} \frac{d}{dt}\left(\frac{mc^2}{\sqrt{1-\mathbf{v}^2/c^2}}\right) &= q\mathbf{E}\cdot\mathbf{v}, \\ \frac{d}{dt}\left(\frac{m\mathbf{v}}{\sqrt{1-\mathbf{v}^2/c^2}}\right) &= q\left(\mathbf{E}+\mathbf{v}\times\mathbf{B}\right). \end{aligned} \tag{3.20}$$

Uit de eerste vergelijking zien we, dat ook in de covariante versie van de theorie magneetvelden geen arbeid verrichten. Voor de beschrijving van het gedrag van geladen deeltjes die grote versnellingen ondergaan —in elektromagnetische velden in de kosmos, dan wel in versnellers op aarde— zijn de relativistische correcties in deze vergelijkingen van groot belang.

Tenslotte zij hier opgemerkt, dat in de uitdrukking (3.19) voor de relativistische Lorentzkracht alleen de veldsterktetensor voorkomt; de componenten van de vectorpotentiaal zelf spelen niet rechtstreeks een rol. Hieruit volgt dat de vergelijking voor de Lorentzkracht ijkinvariant is.

3.3 Elektromagnetische golven

De homogene Maxwell-vergelijkingen zijn opgelost door de veldsterktetensor uit te drukken in termen van de vectorpotentiaal A_{μ}. De inhomogene vergelijkingen krijgen dan de vorm

$$\Box A_{\mu} - \partial_{\mu}\partial\cdot A = -\frac{1}{\epsilon_0 c^2}\, j_{\mu}. \tag{3.21}$$

De vergelijking kan nog verder vereenvoudigd worden door gebruik te maken van de ijktransformaties (3.16): we kunnen de ijktransformatie zo kiezen dat A_{μ} daarna voldoet aan

$$\partial\cdot A = 0. \tag{3.22}$$

Hiermee verdwijnt de tweede term aan de linkerzijde van vgl.(3.21).

Beschouwen we deze vergelijkingen nu in vacuum, d.w.z. ver van elektrische ladingen en stromen, dan is de rechterzijde nul, en de vergelijking wordt

$$\Box A_{\mu} = \left(-\frac{1}{c^2}\frac{\partial^2}{\partial t^2} + \Delta\right) A_{\mu} = 0. \tag{3.23}$$

Deze vergelijking heeft oneindig veel oplossingen van de vorm

$$A_\mu(x) = a_\mu(k)\cos(\mathbf{k}\cdot\mathbf{r} - \omega_k t + \phi_k), \tag{3.24}$$

waarin de frequentie ω_k is vastgelegd in termen van de golfvector $\mathbf{k}$ door

$$\omega_k^2 = \mathbf{k}^2 c^2, \tag{3.25}$$

terwijl de amplitude $a_\mu(k)$ en de fase ϕ_k vrije intergratieconstanten zijn. Uit de leariteit van de Maxwellvergelijkingen volgt, dat de algemene oplossing een superpositie van zulke monochromatische vlakke golven is.

Wiskundig blijkt het relativistische karakter van de oplossing wanneer we een golf-viervector definiëren als

$$k^\mu = (k^0, \mathbf{k}) = \left(\frac{\omega_k}{c}, \mathbf{k}\right). \tag{3.26}$$

De eigenschap (3.25) krijgt dan de invariante vorm

$$k^2 \,=\, k_\mu k^\mu \,=\, 0. \tag{3.27}$$

Ook wordt het plaatsafhankelijke deel van het argument van de cosinusfunctie $\mathbf{k}\cdot\mathbf{r} - k^0 ct \,=\, k\cdot x$, een invariant inprodukt van viervectoren. Het argument van de vlakke golfoplossingen is dus manifest Lorentzinvariant, en de oplossing (3.24) voor $A_\mu(x)$ transformeert in zijn geheel als een viervector wanneer we de amplitude $a_\mu(k)$ zelf als een covariante viervector laten transformeren.

Het relativistische karakter van de oplossingen blijkt ook uit twee belangrijke fysische eigenschappen van de golven:
1. De golven planten zich voort met de lichtsnelheid.
2. De golven zijn altijd transversaal gepolariseerd.
We zullen deze twee eigenschappen nu expliciet laten zien.

1. *Voortplantingssnelheid.* Beschouw het front van punten $\mathbf{r}$ waar op tijdstip t de amplitude van het veld de waarde $a_\mu(k)$ heeft. Dit front is een vlak gedefiniëerd door de vergelijking

$$\mathbf{k}\cdot\mathbf{r} = \omega_k t - \phi_k. \tag{3.28}$$

Deze vergelijking bepaalt alle punten met dezelfde waarde van de $\mathbf{r}$-component langs de golfvector $\mathbf{k}$ op tijd t, maar laat de transversale component van $\mathbf{r}$ loodrecht op $\mathbf{k}$ volledig vrij. In de loop van de tijd verplaatst dit front zich: de posities $\mathbf{r}'$ waar de amplitude ten tijde t' weer dezelfde waarde $a_\mu(k)$ heeft worden bepaald door de eis dat het argument van de cosinus niet is veranderd: $k\cdot x' = k\cdot x$; in aparte ruimte en tijd componenten:

$$\mathbf{k}\cdot\mathbf{r}' - \omega_k t' = \mathbf{k}\cdot\mathbf{r} - \omega_k t. \tag{3.29}$$

Nu is volgens vgl.(3.25) $\omega_k = |\mathbf{k}|\,c$; dus

$$\mathbf{k}\cdot(\mathbf{r}' - \mathbf{r}) = |\mathbf{k}|\,c(t' - t). \tag{3.30}$$

Zij nu $\Delta t = t' - t$, en zij Δr de afstand $\mathbf{r}' - \mathbf{r}$ gemeten langs de golfvector $\mathbf{k}$: $\mathbf{k} \cdot (\mathbf{r}' - \mathbf{r}) = |\mathbf{k}|\, \Delta r$; dan is vgl.(3.30) tenslotte te herschrijven als

$$\frac{\Delta r}{\Delta t} = c. \tag{3.31}$$

Hiermee is aangetoond dat het golffront zich met de lichtsnelheid in de richting van $\mathbf{k}$ voortplant.

2. *Polarisatie.* De elektrische en magnetische velden $(\mathbf{E}, \mathbf{B})$ horend bij de golfoplossingen (3.24) liggen in het vlak loodrecht op de golfvector $\mathbf{k}$:

$$\mathbf{k} \cdot \mathbf{E} = 0, \qquad \mathbf{k} \cdot \mathbf{B} = 0. \tag{3.32}$$

Om dit aan te tonen kiezen we een basis van viervectoren om de amplitude $a_\mu(k)$ en de bijbehorende veldsterkten in uit te drukken. Om te beginnen kiezen we een basis van twee orthonormale vectoren in het transversale vlak: $\varepsilon^i_\mu(k) = (0, \boldsymbol{\varepsilon}^i(\mathbf{k}))$, $i = 1, 2$, met de eigenschappen dat in drie-dimensionale zin

$$\mathbf{k} \cdot \boldsymbol{\varepsilon}^i(\mathbf{k}) \;=\; 0, \qquad \boldsymbol{\varepsilon}^i(\mathbf{k}) \cdot \boldsymbol{\varepsilon}^j(\mathbf{k}) \;=\; \delta^{ij}. \tag{3.33}$$

Dan is ook in vier-dimensionale zin

$$\varepsilon^i(k) \cdot k = 0, \qquad \varepsilon^i \cdot \varepsilon^j = \delta^{ij}. \tag{3.34}$$

Aangezien de Minkowski-ruimte vier-dimensionaal is moeten we aan de transversale vectoren ε^i_μ nog twee niet-transversale viervectoren toevoegen, met een longitudinale en een tijdcomponent. Als een daarvan kunnen we de viervector $k^\mu = (k^0, \mathbf{k})$ zelf kiezen. De keuze van de vierde is niet zo belangrijk, mits deze lineair onafhankelijk is van de drie eerdere basisvectoren. Een veelgemaakte keuze is de tijd-omgekeerde golfvector $\bar{k}^\mu = (-k^0, \mathbf{k})$. Deze heeft de eigenschappen

$$\bar{k}^2 = 0, \qquad \bar{k} \cdot \varepsilon^i(k) = 0, \tag{3.35}$$

en

$$\bar{k} \cdot k = k_0^2 + \mathbf{k}^2 = 2\mathbf{k}^2 \geq 0. \tag{3.36}$$

Ontwikkelen we nu de amplitude $a_\mu(k)$ in deze basis, dan krijgen we een uitdrukking van de vorm

$$a_\mu(k) = \sum_{i=1,2} a^i \varepsilon^i_\mu(k) + a k_\mu + \bar{a} \bar{k}_\mu. \tag{3.37}$$

Echter, de eis $\partial \cdot A = 0$, die we hebben opgelegd om de vrijheid van ijktransformaties in te perken, is voor de golfoplossingen (3.24) equivalent met $k \cdot a(k) = 0$. Vullen we de expliciete vorm (3.37) van de amplitude in, dan volgt uit $k^2 = 0$ en vgln.(3.34) en (3.36), dat

$$k \cdot a(k) = 2\mathbf{k}^2 \bar{a} = 0 \quad \Rightarrow \quad \bar{a} = 0. \tag{3.38}$$

Derhalve is bij keuze van de ijkvoorwaarde (3.22) de amplitude te schrijven als

$$a_\mu(k) = \sum_{i=1,2} a^i \varepsilon^i_\mu(k) + a k_\mu. \tag{3.39}$$

In de uitdrukking (3.12) voor de veldsterktetensor $F_{\mu\nu}$ levert deze oplossing dan

$$\begin{aligned} F_{\mu\nu}(k) &= (k_\nu a_\mu - k_\mu a_\nu) \sin(k \cdot x + \phi_k) \\ &= \sum_{i=1,2} a^i \left(k_\nu \varepsilon^i_\mu(k) - k_\mu \varepsilon^i_\nu(k) \right) \sin(k \cdot x + \phi_k) \end{aligned} \tag{3.40}$$

Voor de elektrische componenten van de veldsterkte vinden we dus

$$\begin{aligned} E_j(k) &= c F_{j0}(k) = c \sum_{i=1,2} a^i \left(k_0 \varepsilon^i_j - k_j \varepsilon^i_0 \right) \sin(k \cdot x + \phi_k) \\ &= \omega_k \sum_{i=1,2} a^i \varepsilon^i_j(k) \sin(k \cdot x + \phi_k). \end{aligned} \tag{3.41}$$

In drie-dimensionale vectornotatie wordt dit

$$\mathbf{E}(k) = \mathbf{E}_k \sin(k \cdot x + \phi_k), \qquad \mathbf{E}_k = \omega_k \sum_{i=1,2} a^i \boldsymbol{\varepsilon}^i(\mathbf{k}). \tag{3.42}$$

De elektrische vector $\mathbf{E}(k)$ ligt dus geheel in het transversale vlak.

Voor de magnetische vector $\mathbf{B}(k)$ kunnen we iets soortgelijks doen. Aangezien $F_{mn} = \nabla_m A_n - \nabla_n A_m = \varepsilon_{mnj} B_j$, levert vgl.(3.40)

$$\begin{aligned} B_j(k) &= (\nabla \times \mathbf{A})_j = -\frac{1}{2} \varepsilon_{jmn} (k_m a_n - k_n a_m) \sin(k \cdot x + \phi_k) \\ &= -\sum_{i=1,2} a^i \left(\mathbf{k} \times \boldsymbol{\varepsilon}^i(\mathbf{k}) \right)_j \sin(k \cdot x + \phi_k). \end{aligned} \tag{3.43}$$

In drie-dimensionale vectornotatie wordt dit

$$\mathbf{B}(k) = \mathbf{B}_k \sin(k \cdot x + \phi_k), \qquad \mathbf{B}_k = \sum_{i=1,2} a^i \boldsymbol{\varepsilon}^i(\mathbf{k}) \times \mathbf{k} = \frac{1}{\omega_k} \mathbf{E_k} \times \mathbf{k}. \tag{3.44}$$

$\mathbf{B}_k$ staat dus niet alleen loodrecht op de golfvector $\mathbf{k}$, maar ook loodrecht op het elektrische veld: $\mathbf{E_k} \cdot \mathbf{B_k} = 0$.

Tenslotte merken we op, dat de elektrische en magnetische veldsterkten in de golf volledig onafhankelijk zijn van de component $a k_\mu$; dit was te verwachten op grond van de ijkonafhankelijkheid van de veldsterkten.

3.4 Het Coulombveld

De golfoplossingen van de Maxwellvergelijkingen voor vrije velden laten zien dat het elektromagnetisch veld zelf fysische vrijheidsgraden bezit: ook in de afwezigheid van ladingen en stromen is de dynamica van het veld niet triviaal. Deze uit zich b.v. in het gedrag van geladen deeltjes (testladingen) onder invloed van een golfveld.

In deze paragraaf beschouwen we een ander soort oplossing van de Maxwellvergelijkingen: het veld van een statische, bolsymmetrische ladingverdeling. De oplossing zegt niets over hoe deze ladingsverdeling tot stand komt; zoals we zullen zien, kan het gaan om het veld van een structuurloze puntlading, maar dat is niet noodzakelijk.

We nemen aan dat de ladingsverdeling is gecentreerd om de oorsprong en een eindige uitgebreidheid heeft met een karakteristieke straal r_0: $\rho(r) = 0$ voor $r > r_0$. Dan moeten de velden op een afstand $r > r_0$ van de oorsprong (buiten de ladingsverdeling zelf) voldoen aan de veldvergelijkingen in de lege ruimte; aangezien ze ook nog statisch moeten zijn, krijgen we de vergelijkingen

$$\begin{aligned} \frac{\partial \mathbf{E}}{\partial t} &= 0, \quad \nabla \times \mathbf{E} = 0, \quad \nabla \cdot \mathbf{E} = 0; \\ \frac{\partial \mathbf{B}}{\partial t} &= 0, \quad \nabla \times \mathbf{B} = 0, \quad \nabla \cdot \mathbf{B} = 0. \end{aligned} \tag{3.45}$$

Deze vergelijkingen hebben bolsymmetrische oplossingen van de vorm

$$\mathbf{E} = e\frac{\mathbf{r}}{r^3}, \qquad \mathbf{B} = b\frac{\mathbf{r}}{r^3}. \tag{3.46}$$

Om dit te controleren gebruiken we de volgende identiteiten voor $r \neq 0$:

$$\nabla r = \nabla\sqrt{\mathbf{r}^2} = \frac{\mathbf{r}}{r}, \qquad \nabla \cdot \mathbf{r} = \frac{\partial x}{\partial x} + \frac{\partial y}{\partial y} + \frac{\partial z}{\partial z} = 3. \tag{3.47}$$

Dan is b.v.

$$\nabla \cdot \mathbf{E} = e\left(\frac{\nabla \cdot \mathbf{r}}{r^3} - \frac{3}{r^4}\,\mathbf{r}\cdot\nabla r\right) = 0, \qquad r \neq 0. \tag{3.48}$$

Als we nu de wet van Gauß in integraalvorm schrijven voor een bol om de oorsprong met straal $R > r_0$, dan is

$$\int_{r<R} \nabla \cdot \mathbf{E}\, d^3\mathbf{r} \;=\; \oint_{r=R} E_r(R) R^2 \sin\theta d\theta d\varphi. \tag{3.49}$$

In deze uitdrukking is $E_r(R)$ de radiële component van het elektrisch veld op de bol $r = R$. Voor het integratiegebied $0 \leq \theta \leq \pi$ en $0 \leq \varphi < 2\pi$, corresponderend met het volledige boloppervlak, en met $E_r(R) = e/R^2$ voor $R > r_0$, krijgen we dan

$$\frac{q}{\epsilon_0} \;=\; \int_{r<R} \frac{\rho}{\epsilon_0}\, d^3\mathbf{r} \;=\; \int_{r<R} \nabla \cdot \mathbf{E}\, d^3\mathbf{r} \;=\; 4\pi e. \tag{3.50}$$

Hierin is q de totale lading binnen de bol met straal R. We vinden zo

$$e = \frac{q}{4\pi\epsilon_0}, \qquad \mathbf{E} = \frac{q}{4\pi\epsilon_0 r^2}\hat{\mathbf{r}}. \tag{3.51}$$

Dit is het Coulombveld. Merk op, dat de constante e die de elektrische veldsterkte bepaalt alleen ongelijk aan nul kan zijn als de totale lading binnen de bol ongelijk aan nul is. I.h.b., als we de straal van de ladingsverdeling naar nul laten gaan, maar de totale lading q eindig houden, moet de ladingsverdeling een delta-functie worden:

$$\nabla \cdot \mathbf{E} = \frac{q}{\epsilon_0}\,\delta^3(\mathbf{r}). \tag{3.52}$$

Dit is immers de oplossing van vgl.(3.50) voor $r_0 \to 0$. Dit argument laat zien, dat niet-triviale oplossingen van de vorm (3.46) bestaan mits de divergentie van de velden niet overal nul is. Het magneetveld $\mathbf{B}$ in (3.46) met $b \neq 0$ correspondeert dan op dezelfde manier met een magnetische monopool. We zien dus, dat de homogene Maxwellvergelijking die zegt dat $\nabla\cdot\mathbf{B} = 0$ *overal* geldt, equivalent is met de conditie dat er geen magnetische monopolen bestaan.

In de afwezigheid van magnetische velden, is de statische oplossing (3.51) uit te drukken in termen van de scalaire potentiaal

$$\mathbf{E} \,=\, -\boldsymbol{\nabla}\phi, \qquad \phi \,=\, \frac{q}{4\pi\varepsilon_0 r}. \tag{3.53}$$

Uit de hierboven besproken limiet van de puntlading leren we nu, dat de Coulombpotentiaal de oplossing is van de inhomogene Laplacevergelijking

$$-\Delta\phi \,=\, \frac{q}{\varepsilon_0}\,\delta^3(\mathbf{r}). \tag{3.54}$$

We vinden hieruit door lineaire superpositie de oplossing voor de potentiaal van een algemene statische ladingsverdeling $\rho(\mathbf{r})$:

$$-\Delta\phi \,=\, \frac{\rho(\mathbf{r})}{\varepsilon_0} \quad\to\quad \phi \,=\, \frac{1}{4\pi\varepsilon_0}\int d^3r'\,\frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}. \tag{3.55}$$

Dit resultaat kunnen we direkt verifiëren:

$$\begin{aligned} -\Delta\,\frac{1}{4\pi\varepsilon_0}\int d^3r'\,\frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \;&=\; \frac{1}{4\pi\varepsilon_0}\int d^3r'\,\rho(\mathbf{r}')\,\Delta\frac{1}{|\mathbf{r}-\mathbf{r}'|} \\ &=\; \frac{1}{4\pi\varepsilon_0}\int d^3r'\,\rho(\mathbf{r}')\,4\pi\delta^3(\mathbf{r}-\mathbf{r}') \;=\; \frac{\rho(\mathbf{r})}{\varepsilon_0}. \end{aligned} \tag{3.56}$$

Het is nu ook mogelijk de scalaire potentiaal te geven in het geval van een willekeurige tijdafhankelijke ladingsverdeling. Dit is niets anders dan de oplossing van vergelijking (3.21) voor $\mu = 0$:

$$\left(\frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \Delta\right)\phi \,=\, \frac{\rho(\mathbf{r},t)}{\varepsilon_0} \quad\to\quad \phi(\mathbf{r},t) = \frac{1}{4\pi\varepsilon_0}\int d^3r'\frac{\rho(\mathbf{r}',t' = t - |\mathbf{r}-\mathbf{r}'|/c)}{|\mathbf{r}-\mathbf{r}'|}. \tag{3.57}$$

De potentiaal is dus een superpositie van Coulombpotentialen, maar niet gegenereerd op hetzelfde tijdstip als waarop $\phi(\mathbf{r}, t)$ wordt bepaald: de superpositie wordt gemaakt met de waarde van de ladingsdichtheid ter plaatse van $\mathbf{r}'$ op een tijd t' die zoveel eerder is dat een lichtsignaal de afstand $|\mathbf{r} - \mathbf{r}'|$ in de tussentijd precies overbruggen kan. We zien dus dat de oplossingen van de Maxwell-vergelijkingen voor tijdafhankelijke ladingsdichtheden aan het relativistisch causaliteitsprincipe voldoen; namelijk, de invloed van een verandering in de ladingsverdeling $\rho(\mathbf{r}, t)$ is op een andere plaats merkbaar na een tijdsinterval $\Delta t = \Delta r/c$.

Samenvatting van hoofdstuk 3

De wetten van de elektrodyamica kunnen worden geschreven op een manier die laat zien dat de theorie manifest Lorenztcovariant is. Dit geldt zowel voor de Maxwellvergelijkingen, die de veldsterkten bepalen als functie van de aanwezige ladingen en stromen, als voor de Lorentzvergelijking, die de kracht op een geladen deeltje geeft in een gegeven elektrisch en magnetisch veld. We hebben afgeleid, dat vrije velden in de afwezigheid van ladingen en stromen kunnen voorkomen in de vorm van elektromagnetische golven, die zich met de lichtsnelheid door de lege ruimte voortplanten en transversaal zijn gepolariseerd. Daarnaast hebben we laten zien, dat de statische bolsymmetrische oplossing van de Maxwell vergelijkingen, het Coulombveld, het veld van een enkele puntlading in rust beschrijft.

Hoofdstuk 4

De meetkunde van ruimte en tijd

4.1 Een veldentheorie van de zwaartekracht

In het voorgaande hoofdstuk hebben we laten zien, dat de Maxwellvergelijkingen voor het elektromagnetische veld, en de Lorentzvergelijking voor de beweging van geladen deeltjes in zulke velden, in een manifest relativistische vorm gegoten kunnen worden. Deze vergelijkingen zijn dus consistent met het principe dat de velden zich met de lichtsnelheid uitbreiden *en* dat deze snelheid een universele constante is, met dezelfde waarde in alle inertiaalstelsels. In het bijzonder bestaat er niet een bevoorrecht inertiaalsysteem waarin de Maxwellvergelijkingen van toepassing zijn, al dan niet geassocieërd met een medium —de *ether*— waarin het licht zich voortplant: alle inertiaalstelsels zijn in elektromagnetisch opzicht equivalent.

Om een consistente beschrijving van alle krachten en bewegingen te krijgen moet het principe van de equivalentie van inertiaalstelsels en de universaliteit van de lichtsnelheid ook tot andere wisselwerkingen worden uitgebreid. Wiskundig betekent dit dat alle veldvergelijkingen en bewegingsvergelijkingen in een Lorentzcovariante vorm geschreven moeten kunnen worden. Moderne theorieën van de kleurkrachten —die quarks en gluonen in een nucleon of pion bij elkaar houden— voldoen hieraan, evenals het standaardmodel van de elektromagnetische en zwakke wisselwerkingen, die verantwoordelijk zijn voor transmutatieprocessen als het β-verval van neutronen en muonen. Ook voor de zwaartekracht is een veldentheoretische beschrijving nodig: aangezien het begrip 'gelijktijdigheid' geen universele betekenis heeft, voldoet een instantane werking op afstand, zoals in de wet van Newton, niet meer. Een veldentheorie van de zwaartekracht moet in overeenstemming zijn met het localiteitsbeginsel: de kracht op een lichaam wordt alleen bepaald door het veld ter plaatse; en aan het causaliteitsbeginsel: veranderingen in het veld planten zich met eindige uitbreidingsnelheid (niet sneller dan het licht) voort. De combinatie van deze twee principes houdt dan in, dat veranderingen in het veld geen invloed kunnen hebben op de materie voordat deze verandering het veld ter plekke van de materie bereikt, wat altijd later is dan het ogenblik waarop de oorspronkelijke verandering plaats vindt. Nauwkeuriger, maar abstracter: een gebeurtenis die het veld op een bepaalde plek op een bepaalde tijd verandert kan alleen invloed hebben op de materie die zich binnen de voorwaartse lichtkegel van de gebeurtenis bevindt.

De directe werking op afstand in Newtons wet van de zwaartekracht (1.1) kan in de context van een veldentheorie hooguit van toepassing zijn op statische situaties, waarin de inequivalentie van het begrip gelijktijdigheid voor verschillende waarnemers geen rol speelt. We verwachten daarom, dat in een relativistische veldentheorie van de zwaartekracht afwijkingen van Newtons wet optreden wanneer de massa's die de bron van het zwaartekrachtveld zijn —en er ook de effecten van ondergaan— t.o.v. elkaar in beweging zijn.

Hoewel Newtons wet zelf niet algemeen geldig kan zijn, kan het principe van de equivalentie van zware en trage massa, dat er aan ten grondslag ligt, wel in een lokale veldentheorie worden opgenomen. De veldentheorie van de zwaartekracht die dit principe en de universaliteit van de lichtsnelheid als uitgangspunt heeft, is de algemene relativiteitstheorie (ART), die in 1915 door Einstein werd geformuleerd. Deze theorie gaat uit van het beginsel dat alle lichamen bij gelijke begincondities (plaats en snelheid), en in zover hun interne structuur bij de beweging geen rol speelt, in een zwaartekrachtveld dezelfde baan volgen – een uitbreiding van de observatie dat de valversnelling van een object onafhankelijk is van de massa. Dit betekent dat de banen van testdeeltjes een fysisch universele betekenis krijgen, die te vertalen is in meetkundige taal: we kunnen een meetkunde van ruimte en tijd construeren, door metingen van afstanden en de veranderingen daarvan uit te voeren m.b.v. velden (licht, radar) of m.b.v. bewegende deeltjes (metingen van snelheden en vluchttijden) die alleen onder invloed van de zwaartekracht staan. Het gegeneraliseerde equivalentieprincipe stelt dan, dat een zo verkregen meetkundige beschrijving van de tijd-ruimte universele geldigheid bezit in de zin dat ze onafhankelijk is van de aard van het meetinstrument (de aard van de testdeeltjes).

In deze meetkundige beschrijving van de zwaartekracht treedt een wijziging op in de status van inertiaalstelsels. We definiëren inertiaalstelsels nog steeds als coördinatenstelsels waarin vrije deeltjes ten opzichte van elkaar in rust zijn of eenparig rechtlijnig bewegen. In de aanwezigheid van krachtvelden, zoals b.v. een elektromagnetisch veld, kunnen inertiaalstelsels geconstrueerd worden m.b.v. lichamen (testdeeltjes) die geen invloed van het krachtveld ondergaan, omdat ze niet aan het veld koppelen: deeltjes zonder lading of magnetisch moment. Zulke vrije deeltjes bestaan en dienen als referentie om het gedrag van deeltjes die wel onder invloed van het krachtveld staan mee te vergelijken. De geladen deeltjes oefenen onderling krachten op elkaar uit, en ondergaan in een veld versnellingen t.o.v. de ongeladen vrije deeltjes. De veld- en bewegingsvergelijkingen —zoals de Maxwell- en Lorentzvergelijkingen in het geval van elektromagnetische krachten— beschrijven het gedrag van geladen objecten en hun velden in een inertiaalstelsel van neutrale klokken en meetlatten. Het zwaartekrachtveld is daarentegen een universeel krachtveld: alle lichamen (alle testdeeltjes) koppelen eraan. In een zwaartekrachtsveld zijn er geen vrije deeltjes die als referentie kunnen dienen en waarmee een inertiaalstelsel geconstrueerd zou kunnen worden. De enige manier om een inertiaalstelsel praktisch te realiseren is een deel van de tijd-ruimte te zoeken waar geen zwaartekrachtveld aanwezig is, zodat in ieder geval lokaal (d.w.z. in een eindig volume en gedurende een eindige periode) vrije testdeeltjes gevonden kunnen worden die zich als in een inertiaalstelsel gedragen.

Daarbij dient nog een belangrijke opmerking gemaakt te worden: beschouwen we een gebied in de ruimte waarin gedurende een zekere tijd een constant zwaartekrachtveld aanwezig is, dan worden alle voorwerpen in dit veld in gelijk mate versneld. Er is dus geen relatieve versnelling tussen verschillende objecten. T.o.v. een meebewegende waarnemer (een waarnemer in vrije val) zijn dan alle deeltjes in rust of bewegen met een snelheid die constant is in grootte en richting. Uit de beweging van locale objecten kan een waarnemer in vrije val zijn omgeving dus niet onderscheiden van een inertiaalstelsel. Natuurlijk kan hij of zij ontdekken dat fysische objecten die zich op grote afstand bevinden —zoals hemellichamen— versneld bewegen, b.v. door metingen te doen van de Dopplerverschuiving van spectraallijnen van sterren als functie van de tijd. Onze waarnemer kan die versnelling toeschrijven aan een zwaartekrachtveld ter plaatse van deze sterren. Maar omgekeerd kan dat ook: een stelsel in vrije val met een van de sterren is lokaal ook een inertiaalstelsel; een waarnemer in dat stelsel schrijft de relatieve versnelling van ons systeem t.o.v. van hem of haar toe aan een zwaartekrachtveld hier. De gevolgtrekking die op grond van deze voorstelling van zaken gemaakt kan worden is, dat zwaartekracht en versnelling geen absolute begrippen zijn, maar alleen relatief tussen lokale inertiaalstelsels op verschillende plaatsen en/of tijden gedefiniëerd kunnen worden. Hoe ver weg je moet gaan om afwijkingen van het ideaal gedrag van een inertiaalsysteem te vinden hangt natuurlijk af van de gradienten in het veld: hoe meer het veld in de omgeving van een waarnemer in vrije val verschilt van een constant veld, hoe kleiner het gebied dat in de praktijk (binnen de meetnauwkeurigheid) als lokaal inertiaalstelsel fungeert. Zolang de verandering van de velden echter continue is en zonder sprongen verloopt (als de afgeleiden van de velden naar tijd en ruimte ook continu zijn) kan in ieder punt van de tijd-ruimte een coördinatenstelsel geconstrueerd worden dat een inertiaalstelsel willekeurig dicht benadert als we maar dicht genoeg in de buurt van dat punt blijven. Om precies te zijn: we kunnen in ieder punt natuurlijk een hele schaar van zulke lokale inertiaalstelsels construeren, die onderling allemaal door Lorentztransformaties verbonden zijn.

Wat we echter niet meer kunnen is inertiaalstelsels op verschillende punten in ruimte en tijd door een Lorentztransformatie in elkaar over laten gaan: Lorentztransformaties hebben alleen een lokale betekenis op een vast punt, tussen coördinatenstelsels die daar lokaal inertiaalstelsels zijn. Inertiaalstelsels op verschillende punten in tijd en ruimte kunnen verbonden zijn door algemene coördinatentransformaties, waarvan de expliciete vorm afhangt van de geometrie van de tijd-ruimte.

Hoewel de formulering van de ART gebruik maakt van de taal en inzichten van de meetkunde, willen we hier benadrukken dat de beschrijving van de zwaartekracht die de ART levert vooral steunt op *fysische* inzichten, zoals de universaliteit van de zwaartekracht, en de equivalentie van zware en trage massa. Als uit waarnemingen zou blijken dat deze inzichten niet juist of slechts beperkt geldig zijn, dan moet de theoretische beschrijving van de zwaartekracht worden aangepast. Met name merken we op, dat de fysische uitgangspunten van de ART uitsluitend voortkomen uit waarnemingen van verschijnselen op macroscopische schaal: laboratoriumproeven, zoals die van Cavendish of aan boord van satellieten, bewegingen van projectielen en hemellichamen, en de reconstructie van de evolutie van het zichtbare heelal. Op microscopische schaal kunnen we verwijzen naar statistisch-thermodynamische effecten, zoals de barometrische hoogteverdeling van de lucht, die laat zien dat Newtons wet ook op moleculair niveau werkzaam is. Maar echt dynamische effecten op atomaire schaal zijn experimenteel niet bekend. Er is dus ruimte genoeg voor aanpassingen van de ART, in het bijzonder in het regime van de quantumverschijnselen. De beschrijving van de macroscopische zwaartekrachtsverschijnelen is echter op zich al interessant genoeg; we zullen zien dat deze vanzelf leidt tot vragen waarbij de quantumtheorie niet meer buiten beschouwing gelaten kan worden.

4.2 Meetkunde en velden

De opmerkingen over de fysische uitgangspunten van de ART laten onverlet, dat de theorie vereist een aantal wiskundige begrippen in te voeren die voor de lokaal-meetkundige interpretatie onmisbaar zijn. Om tegelijkertijd het fysische aspect niet uit het oog te verliezen, zullen we hier de belangrijkste ingredienten van de ART parallel invoeren in de terminologie van twee complementaire zienswijzen: de meetkundige, en de veldentheoretische. De meetkundige interpretatie van de zwaartekracht houdt in dat in ieder punt van de tijd-ruimte lokale inertiaalstelsels geconstrueerd kunnen worden, en dat de onderlinge versnelling van naburige inertiaalsystemen ten opzichte van elkaar het gevolg zijn van de kromming van de tijd-ruimte. Ieder object in vrije val volgt in de tijd-ruimte een baan die zuiver geometrisch bepaald is, en die een extreme waarde vertegenwoordigt van de eigentijd langs de wereldlijn; in meetkundige taal is de baan van het object een geodetische lijn. De zwaartekracht is vanuit dit gezichtspunt in zekere zin een schijnkracht: de onderlinge versnellingen van objecten zijn geen gevolg van krachten maar van de meetkundige eigenschappen van de tijd-ruimte. De meetkundige structuur zelf wordt bepaald door de verdeling van materie en straling in ruimte en tijd.

De veldentheoretische beschrijving van de zwaartekracht begint met het invoeren van potentialen en veldsterkten, die door de gradiënten van de potentiaal worden gegeven. De veldsterkte bepaalt de kracht die het veld op bepaalde plaats op een gegeven deeltje uitoefent. Anderzijds is de verdeling van materie en straling in ruimte en tijd een bron van zwaartekrachtvelden, of preciezer: van variaties in de veldsterkten.

We geven nu eerst een korte samenvatting van de wiskundige begrippen en fysische grootheden die een rol spelen in de ART. Een uitwerking van de concepten en interpretatie volgt daarna. De lokale meetkundige eigenschappen van de tijdruimte worden vastgelegd door de metriek, die bepaalt hoe het invariante interval ds^2 in een omgeving van het punt $\mathcal{P}(x)$ wordt uitgedrukt in termen van coördinatendifferentialen dx^μ in een lokaal coördinatenstelsel in die omgeving; voor tijdachtige intervallen langs de wereldlijn van een deeltje is dit de eigentijd[1]:

$$ds^2 = -c^2 d\tau^2 = g_{\mu\nu}(x)\, dx^\mu dx^\nu. \tag{4.1}$$

In de veldentheoretische beschrijving spelen de tien componenten van de metriek $g_{\mu\nu}(x) = g_{\nu\mu}(x)$ de rol van potentialen voor de zwaartekracht. De inverse van de metriek is ook een belangrijke grootheid; de componenten worden genoteerd als $g^{\mu\nu}(x)$, zodat

$$g^{\mu\lambda}(x)\, g_{\lambda\nu}(x) = \delta^\mu_\nu. \tag{4.2}$$

De rechterzijde van deze vergelijking is per definitie onafhankelijk van het punt $\mathcal{P}(x)$. De baan van een testdeeltje $x^\mu(\tau)$ in een gegeven metriek is een geodetische kromme. Deze wordt bepaald door het Riemann-Christoffel symbool, ook de connectie genoemd:

$$\Gamma_{\mu\nu}^{\;\;\lambda} = \Gamma_{\nu\mu}^{\;\;\lambda} \equiv \frac{1}{2}\, g^{\lambda\kappa} \left(\partial_\mu g_{\kappa\nu} + \partial_\nu g_{\kappa\mu} - \partial_\kappa g_{\mu\nu} \right). \tag{4.3}$$

In deze uitdrukking noteren we partiële afgeleiden als

$$\partial_\kappa g_{\mu\nu} = \frac{\partial g_{\mu\nu}}{\partial x^\kappa}. \tag{4.4}$$

Fysisch spelen de componenten van de connectie de rol van de veldsterkten: deze bepalen de versnelling van een testdeeltje, en dus de kracht die door het zwaartekrachtveld wordt uitgeoefend. De vergelijking voor geodetische krommen $x^\mu(\tau)$ maakt dit duidelijk: geodeten zijn oplossingen van de gewone differentiaalvergelijking

$$\frac{d^2 x^\mu}{d\tau^2} + \Gamma_{\lambda\nu}^{\;\;\mu} \frac{dx^\lambda}{d\tau} \frac{dx^\nu}{d\tau} = 0. \tag{4.5}$$

Geinterpreteerd als bewegingsvergelijking voor een testdeeltje, stelt deze vergelijking dat de eigenversnelling $d^2x^\mu/d\tau^2$ (de versnelling gemeten in termen van

[1] Ook hier gebruiken we weer de conventie van impliciete sommatie over indices die in een product herhaald worden.

de eigentijd τ) wordt bepaald door de connectie en de eigensnelheid $dx^\mu/d\tau$. Deze vergelijking speelt als bewegingsvergelijking voor testdeeltjes in een zwaartekrachtveld de rol die de Lorenztkracht speelt in de theorie van het elektromagnetische veld.

Tenslotte moeten we aangeven hoe de meetkundige eigenschappen van ruimte en tijd samenhangen met de verdeling van materie en straling. De door Einstein gepostuleerde vergelijkingen beschrijven de variaties in de veldsterkten in termen van de Riemanntensor, waarvan de componenten worden gedefiniëerd door

$$R_{\mu\nu\kappa}{}^{\lambda} = \partial_\mu \Gamma_{\nu\kappa}{}^{\lambda} - \partial_\nu \Gamma_{\mu\kappa}{}^{\lambda} - \Gamma_{\mu\kappa}{}^{\sigma}\Gamma_{\nu\sigma}{}^{\lambda} + \Gamma_{\nu\kappa}{}^{\sigma}\Gamma_{\mu\sigma}{}^{\lambda}. \tag{4.6}$$

Het spoor van de Riemanntensor wordt de Riccitensor genoemd:

$$R_{\mu\nu} = R_{\mu\lambda\nu}{}^{\lambda}. \tag{4.7}$$

Hieruit kan nog weer een scalar worden gevormd door contractie van de tensor met de inverse metriek, de Riemannscalar:

$$R = g^{\mu\nu} R_{\mu\nu}. \tag{4.8}$$

De Einsteintensor $G_{\mu\nu}$ is een speciale combinatie van Riccitensor en Riemannscalar:

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}\, g_{\mu\nu} R. \tag{4.9}$$

Einsteins vergelijkingen voor het zwaartekrachtveld stellen nu, dat deze grootheid direct wordt bepaald door de verdeling van de energie en de impuls van straling en materie in ruimte en tijd. Deze verdeling wordt beschreven door een nader te specificeren energie-impuls tensor $T_{\mu\nu}$. Volgens de ART is de Einsteintensor evenredig met de energie-impuls tensor

$$G_{\mu\nu} = -\frac{8\pi G}{c^4}\, T_{\mu\nu}. \tag{4.10}$$

Hierin is G de gravitatieconstante van Newton en c de lichtsnelheid. In de volgende paragrafen zullen we de hier ingevoerde grootheden nader toelichten. Eerst nog een opmerking over dimensies. De metrische tensor $g_{\mu\nu}$ en zijn inverse zijn dimensieloos; dus de connectie $\Gamma_{\mu\nu}{}^{\lambda}$ heeft dimensie ℓ^{-1}, en de Riemann- en Riccitensor, evenals de Riemannscalar en Einsteintensor, hebben dimensie ℓ^{-2}. Aangezien de constante van Newton de dimensie $m^{-1}\ell^3 t^{-2}$ heeft, en gedeeld wordt door vierde macht van de lichtsnelheid, heeft de energie-impulstensor $T_{\mu\nu}$ de dimensie $m\ell^{-1}t^{-2} = m\ell^2 t^{-2}/\ell^3$. Dit is precies de dimensie van een energie per volumeëenheid, d.w.z. een energiedichtheid. De constante G, voortgekomen uit Newtons wet voor de zwaartekracht (1.1), speelt in de ART een vergelijkbare rol: die van evenredigheidsconstante tussen kromming en energiedichtheid in de Einsteinvergelijking.

4.3 De metriek

De tien componenten van de symmetrische tensor $g_{\mu\nu}(x)$ spelen in de veldentheoretische interpretatie van de ART de rol van potentialen: hun afgeleiden bepalen de veldsterkte, gedefiniëerd als de eigensversnelling of kracht per eenheid van massa op een testdeeltje; zie vgl.(4.5). De meetkundige interpretatie van deze tensor als metriek is vastgelegd in de relatie (4.1) die bepaalt hoe een infinitesimaal interval van eigentijd $d\tau$ gemeten langs de wereldlijn van een deeltje kan worden uitgedrukt in de infinitesimale coördinatenintervallen dx^μ ter plaatse. Nu is het altijd mogelijk in een gegeven vast punt $\mathcal{P}$ lokaal een coördinatenstelsel $\bar{x}^\mu$ te kiezen dat in vrije val is, en waarin de metriek de Minkowskivorm heeft:

$$\bar{g}_{\mu\nu}(\bar{x})|_{\mathcal{P}} = \eta_{\mu\nu} = \mathrm{diag}(-1, +1, +1, +1). \tag{4.11}$$

In dat stelsel is de relatie tussen interval ds of eigentijd $d\tau$ en de coördinatenintervallen in $\mathcal{P}$ dezelfde als in een inertiaalstelsel:

$$c^2 d\tau^2|_{\mathcal{P}} = -\bar{g}_{\mu\nu}(\bar{x}) d\bar{x}^\mu d\bar{x}^\nu|_{\mathcal{P}} = -\eta_{\mu\nu} d\bar{x}^\mu d\bar{x}^\nu. \tag{4.12}$$

Er zijn zelfs oneindig veel coördinatenstelsels $\bar{x}^\mu$ met deze eigenschap, alle verbonden via Lorentztransformaties (2.9). Het verschil met de inertiaalstelsels die we eerder gebruikt hebben, is echter dat eenzelfde eigentijd-interval $d\tau$ langs de baan van een deeltje in een ander punt $\mathcal{P}'$ in dit coördinatenstelsel niet meer de Minkowskivorm (4.12) zal hebben, maar de algemene vorm (4.1). Het coördinatenstelsel $\bar{x}^\mu$ is dus een lokaal inertiaalstelsel in het punt $\mathcal{P}$, maar niet noodzakelijk in naburige punten $\mathcal{P}'$. De relatie tussen de coördinaten $\bar{x}^\mu$ en x^μ legt vanzelf ook de relatie tussen de componenten van de metriek $\bar{g}_{\mu\nu}(\bar{x})$ en $g_{\mu\nu}(x)$ in de twee stelsels vast; voor tijdachtige intervallen is

$$c^2 d\tau^2 = -g_{\mu\nu}(x) dx^\mu dx^\nu = -\bar{g}_{\mu\nu}(\bar{x}) d\bar{x}^\mu d\bar{x}^\nu. \tag{4.13}$$

Aangezien $d\bar{x}^\mu = (\partial \bar{x}^\mu / \partial x^\nu)\, dx^\nu$, volgt hieruit dat

$$g_{\mu\nu}(x) = \bar{g}_{\kappa\lambda}(\bar{x}(x)) \, \frac{\partial \bar{x}^\kappa}{\partial x^\mu} \frac{\partial \bar{x}^\lambda}{\partial x^\nu}. \tag{4.14}$$

Dit is een lokale generalisatie van de vergelijking (2.9) die de transformatie tussen twee inertiaalstelsels vastlegt, waarbij de metriek niet van vorm verandert. Hier mag de metriek wel van vorm veranderen en zijn de transformaties niet noodzakelijk lineair. Analoog aan de conventies voor Lorentztransformaties in de Minkowskiruimte, definiëren we nu algemeen covariante en contravariante vectoren V als objecten waarvan de componenten onder een algemene coördinatentransformatie veranderen op dezelfde wijze als een afgeleide $\partial / \partial x^\mu$, dan wel als een differentiaal dx^μ:

$$V_\mu(x) = \bar{V}_\nu(\bar{x}) \frac{\partial \bar{x}^\nu}{\partial x^\mu}, \qquad V^\mu(x) = \bar{V}^\nu(\bar{x}) \frac{\partial x^\mu}{\partial \bar{x}^\nu}. \tag{4.15}$$

Covariante en contravariante tensoren zijn hierop voortbouwend gedefiniëerd als objecten die transformeren als directe produkten van differentialen en differentiaaloperatoren; zo laat vgl.(4.14) voor de transformatie van de metriek zien dat $g_{\mu\nu}(x)$ onder algemene coördinatentransformaties een covariante tensor van rang twee is. Evenzo is de inverse metriek een contravariante tensor van rang twee:

$$g^{\mu\nu}(x) = \bar{g}^{\kappa\lambda}(\bar{x}) \frac{\partial x^\mu}{\partial \bar{x}^\kappa} \frac{\partial x^\nu}{\partial \bar{x}^\lambda}. \tag{4.16}$$

Dit is consistent met de inverse relatie (4.2) tussen deze tensoren als matrices, omdat

$$\frac{\partial x^\mu}{\partial \bar{x}^\lambda} \frac{\partial \bar{x}^\lambda}{\partial x^\nu} = \delta^\mu_\nu \tag{4.17}$$

Tenslotte is een scalar een grootheid die niet verandert onder een algemene coördinatentransformatie:

$$S(x) = \bar{S}(\bar{x}). \tag{4.18}$$

Zulke scalaire grootheden zijn o.a. te verkrijgen door een invariante vermenigvuldiging van covariante en contravariante vectoren en tensoren; voorbeelden hiervan zijn de contractie van een covariante en een contravariante vector: $V_\mu W^\mu$, of de contractie van twee contravariante vectoren met de covariante metrische tensor: $g_{\mu\nu} V^\mu W^\nu$; de definitie van het invariante interval (4.1) heeft deze vorm. Zoals uit dit voorbeeld blijkt, kunnen contravariante vectoren tot covariante vectoren worden getransformeerd, en omgekeerd, door contractie met de metriek of de inverse metriek:

$$V_\mu = g_{\mu\nu} V^\nu, \qquad V^\mu = g^{\mu\nu} V_\nu. \tag{4.19}$$

4.4 De connectie

We hebben de connectie ${\Gamma_{\mu\nu}}^\lambda$ ingevoerd als de speciale combinatie van afgeleiden van de metriek gegeven in vgl.(4.3):

$${\Gamma_{\mu\nu}}^\lambda = {\Gamma_{\nu\mu}}^\lambda = \frac{1}{2} g^{\lambda\kappa} \left(\partial_\mu g_{\lambda\nu} + \partial_\nu g_{\mu\lambda} - \partial_\lambda g_{\mu\nu} \right).$$

We zullen nu laten zien, dat deze combinatie de juiste is voor het bepalen van geodetische krommen m.b.v. vgl.(4.5). Vanwege de fysische interpretatie hiervan als bewegingsvergelijking van testdeeltjes beschouwen we een tijdachtige kromme $x^\mu(\lambda)$, waarin λ een reële parameter is voor de afstand afgelegd langs de baan; voor ruimte-achtige geodeten gaat de afleiding geheel analoog. Het totale interval (eigentijd) langs een willekeurige differentiëerbare en tijdachtige baan tussen twee punten $\mathcal{P}_1$ en $\mathcal{P}_2$ is

$$c\Delta\tau = \int_{\mathcal{P}_1}^{\mathcal{P}_2} \sqrt{-g_{\mu\nu}(x) dx^\mu dx^\nu} = \int_{\mathcal{P}_1}^{\mathcal{P}_2} d\lambda \sqrt{-g_{\mu\nu}(x) \dot{x}^\mu \dot{x}^\nu}. \tag{4.20}$$

Net als voorheen hebben we hier de notatie $\dot{x} \equiv dx/d\lambda$ ingevoerd voor de viersnelheid, d.w.z. de raakvector aan de baan. We stellen nu dat de baan van een testdeeltje tussen $\mathcal{P}_1$ en $\mathcal{P}_2$ een extreme waarde voor het totale eigentijdinterval oplevert, zodat onder lineaire variaties langs het pad $x^\mu(\lambda) \rightarrow \bar{x}^\mu(\lambda) = x^\mu(\lambda) + \delta x^\mu(\lambda)$:

$$\delta\Delta\tau = 0. \tag{4.21}$$

De variatie van de integrand in (4.20) levert

$$\delta\sqrt{-g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu} = \frac{-1}{2\sqrt{-g_{\rho\sigma}\dot{x}^\rho\dot{x}^\sigma}}\left(2g_{\mu\nu}\dot{x}^\nu\delta\dot{x}^\mu + \delta x^\lambda\partial_\lambda g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu\right). \tag{4.22}$$

De factor 2 in de laatste uitdrukking is verkregen door gebruik te maken van de symmetrie van de metrische tensor om de twee termen die komen van de variatie van de eigentijd-afgeleiden aan elkaar gelijk te stellen. Omdat de variatie en de eigentijd-afgeleide onafhankelijke operaties zijn, kunnen we de variatie van de afgeleide vervangen door de afgeleide van de variatie:

$$\delta\dot{x}^\nu = \frac{d}{d\lambda}\left(\delta x^\nu\right). \tag{4.23}$$

We integreren nu partiëel onder de integraal en maken gebruik van de voorwaarde dat de variatie aan het begin- en eindpunt nul is, omdat $\mathcal{P}_1$ en $\mathcal{P}_2$ vastliggen; aangezien

$$\frac{d}{d\tau} = \frac{1}{\sqrt{-g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu}}\frac{d}{d\lambda}, \tag{4.24}$$

krijgen we dan

$$\delta\sqrt{-g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu} \simeq \sqrt{-g_{\rho\sigma}\dot{x}^\rho\dot{x}^\sigma}\,\delta x^\kappa\left(g_{\kappa\mu}\frac{d^2x^\mu}{d\tau^2} + \frac{dx^\lambda}{d\tau}\frac{dx^\nu}{d\tau}\left(\partial_\lambda g_{\kappa\nu} - \frac{1}{2}\partial_\kappa g_{\nu\lambda}\right)\right). \tag{4.25}$$

Het teken $\simeq$ duidt hier op gelijkheid na partiële integratie. De combinatie van afgeleiden van de metriek tussen de binnenste haakjes komt precies overeen met de uitdrukking voor $g_{\kappa\mu}\Gamma_{\nu\lambda}{}^\mu$, als we de symmetrisatie over $(\nu\lambda)$ gebruiken. We zien dus, dat de voorwaarde van een extreme waarde van de eigentijd leidt tot vergelijking (4.5):

$$\frac{d^2x^\mu}{d\tau^2} + \Gamma^\mu_{\lambda\nu}\frac{dx^\lambda}{d\tau}\frac{dx^\nu}{d\tau} = 0.$$

Merk op dat in deze vergelijking de parameter λ niet meer voorkomt; i.h.b. zijn de afgeleiden naar λ vervangen door afgeleiden naar de eigentijd τ. De oplossing van het variatieprobleem (4.21) hangt dus niet af van de speciale parametrisatie die voor de wereldlijn gekozen is.

Het postulaat dat testdeeltjes in de tijd-ruimte langs geodetische banen bewegen leidt nu inderdaad tot een identificatie van de veldsterkte met de connectie als bron van de versnelling die het deeltje ondergaat. In het bijzonder geldt in de niet-relativistische limiet $v/c \ll 1$ dat $\dot{x}^0 \approx c$ en $d\tau \approx dt$, zodat een deeltje met massa m een kracht ondervindt ter grootte

$$m\frac{d^2x^i}{dt^2} \approx -mc^2\Gamma^i_{00} + \mathcal{O}\left(\frac{v}{c}\right). \tag{4.26}$$

De grootheid aan de rechterkant bepaalt dus de sterkte van een zwaartekrachtveld in een niet-relativistische situatie. Wat de transformatieëigenschappen onder algemene coördinatentransformaties betreft is de connectie *geen* tensor: de connectie transformeert niet homogeen. Dit volgt uit de transformatieregel van de metriek, gegeven in vgl.(4.14), en de transformatie van de afgeleiden volgens de kettingregel

$$\frac{\partial}{\partial x^\mu} = \frac{\partial \bar{x}^\nu}{\partial x^\mu}\frac{\partial}{\partial \bar{x}^\nu}. \tag{4.27}$$

Na invullen van deze transformaties in de definitie van de connectie vinden we dat:

$$\Gamma_{\mu\nu}{}^\lambda(x) = \bar{\Gamma}_{\rho\sigma}{}^\tau(\bar{x})\frac{\partial \bar{x}^\rho}{\partial x^\mu}\frac{\partial \bar{x}^\sigma}{\partial x^\nu}\frac{\partial x^\lambda}{\partial \bar{x}^\tau} + \frac{\partial^2 \bar{x}^\sigma}{\partial x^\mu \partial x^\nu}\frac{\partial x^\lambda}{\partial \bar{x}^\sigma}. \tag{4.28}$$

In deze uitdrukking zit een inhomogene term die de tweede partiële afgeleide van de coördinatentransformatie bevat. Deze extra term zorgt ervoor dat de vergelijking voor de geodetische paden *als geheel* covariant transformeert, ook al komt er een tweede afgeleide naar de eigentijd τ in voor. Om dit te zien, schrijven we de vergelijking eerst in een andere vorm. Zij $u^\mu = \dot{x}^\mu$ de viersnelheid, de raakvector aan de wereldlijn $x^\mu(\tau)$. De viersnelheid is een contravariante vector, zoals volgt uit de kettingregel:

$$\bar{u}^\mu = u^\nu \frac{\partial \bar{x}^\mu}{\partial x^\nu}. \tag{4.29}$$

De geodetische vergelijking kan nu worden herschreven als

$$u^\nu D_\nu u^\mu = \frac{dx^\nu}{d\tau}\left(\frac{\partial u^\mu}{\partial x^\nu} + \Gamma_{\nu\lambda}{}^\mu u^\lambda\right) = 0, \tag{4.30}$$

waarin D_μ de covariante afgeleide is; deze lineaire operatie is op een contravariante vector gedefiniëerd als

$$D_\mu V^\lambda \equiv \partial_\mu V^\lambda + \Gamma_{\mu\nu}{}^\lambda V^\nu. \tag{4.31}$$

De covariante afgeleide heeft de bijzondere eigenschap dat hij onder een algemene coördinatentransformatie vectoren en tensoren in tensoren van een hogere rang omzet. In het bovenstaande geval is $D_\mu V^\lambda$ een gemengde tensor van rang (1,1):

$$D_\mu V^\lambda(x) = \frac{\partial \bar{x}^\nu}{\partial x^\mu}\frac{\partial x^\lambda}{\partial \bar{x}^\kappa}\bar{D}_\nu \bar{V}^\kappa(\bar{x}). \tag{4.32}$$

Het bewijs van deze bewering volgt hieronder. Eerst merken we op, dat vgl.(4.30) dan geldt in ieder coördinatenstelsel, omdat

$$\bar{u}^\nu \bar{D}_\nu \bar{u}^\mu \;=\; \frac{\partial \bar{x}^\mu}{\partial x^\lambda}\, u^\nu D_\nu u^\lambda \;=\; 0. \tag{4.33}$$

Het bewijs van vgl.(4.32) gaat als volgt:

$$\begin{aligned} \frac{\partial V^\lambda}{\partial x^\mu}(x) \;&=\; \frac{\partial \bar{x}^\nu}{\partial x^\mu}\frac{\partial}{\partial \bar{x}^\nu}\left(\frac{\partial x^\lambda}{\partial \bar{x}^\kappa}\bar{V}^\kappa(\bar{x})\right) \\ &=\; \frac{\partial \bar{x}^\nu}{\partial x^\mu}\frac{\partial x^\lambda}{\partial \bar{x}^\kappa}\frac{\partial \bar{V}^\kappa}{\partial \bar{x}^\nu}(\bar{x}) \;+\; \frac{\partial \bar{x}^\nu}{\partial x^\mu}\frac{\partial^2 x^\lambda}{\partial \bar{x}^\nu \partial \bar{x}^\kappa}\,\bar{V}^\kappa(\bar{x}). \end{aligned} \tag{4.34}$$

De eerste term in de laatste uitdrukking volgt precies de transformatieregel voor een gemengd co- en contravariante tensor. De laatste term met de tweede afgeleide van de coördinatentransformatie erin gooit echter roet in het eten. Deze wordt nu gecompenseerd door de inhomogene term in de transformatie van de connectie; de laatste kan namelijk geschreven worden als

$$\Gamma_{\mu\nu}^{\;\;\;\lambda} V^\nu(x) \;=\; \left(\frac{\partial \bar{x}^\sigma}{\partial x^\mu}\frac{\partial x^\lambda}{\partial \bar{x}^\nu}\,\bar{\Gamma}_{\sigma\kappa}^{\;\;\;\nu} + \frac{\partial^2 \bar{x}^\sigma}{\partial x^\mu \partial x^\nu}\frac{\partial x^\lambda}{\partial \bar{x}^\sigma}\frac{\partial x^\nu}{\partial \bar{x}^\kappa}\right)\bar{V}^\kappa(\bar{x}). \tag{4.35}$$

De eerste term binnen de haakjes is weer precies wat we verwachten; de tweede valt weg tegen de extra term uit de vorige vergelijking:

$$\begin{aligned} \frac{\partial^2 \bar{x}^\sigma}{\partial x^\nu \partial x^\mu}\frac{\partial x^\nu}{\partial \bar{x}^\kappa}\frac{\partial x^\lambda}{\partial \bar{x}^\sigma}\,\bar{V}^\kappa(\bar{x}) \;&=\; \frac{\partial^2 \bar{x}^\sigma}{\partial \bar{x}^\kappa \partial x^\mu}\frac{\partial x^\lambda}{\partial \bar{x}^\sigma}\,\bar{V}^\kappa(\bar{x}) \\ &=\; \left[\frac{\partial}{\partial \bar{x}^\kappa}\left(\frac{\partial \bar{x}^\sigma}{\partial x^\mu}\frac{\partial x^\lambda}{\partial \bar{x}^\sigma}\right) - \frac{\partial \bar{x}^\sigma}{\partial x^\mu}\frac{\partial^2 x^\lambda}{\partial \bar{x}^\kappa \partial \bar{x}^\sigma}\right]\bar{V}^\kappa(\bar{x}) \\ &=\; -\frac{\partial \bar{x}^\sigma}{\partial x^\mu}\frac{\partial^2 x^\lambda}{\partial \bar{x}^\kappa \partial \bar{x}^\sigma}\,\bar{V}^\kappa(\bar{x}) \end{aligned} \tag{4.36}$$

We gebruiken hier, dat de Kronecker delta constant is:

$$\partial_\kappa \delta^\lambda_\mu = 0. \tag{4.37}$$

De rechterzijde van (4.36) is inderdaad precies tegengesteld aan de extra term met de tweede afgeleide in vgl.(4.34). Combinatie van de vgln.(4.34), (4.35) en (4.36) leidt dan tot het resultaat (4.32).Dit bewijs is makkelijk te generaliseren voor contravariante tensoren, als we de covariante afgeleide definiëren door

$$D_\mu T^{\nu_1\nu_2\ldots} \;=\; \partial_\mu T^{\nu_1\nu_2\ldots} \;+\; \Gamma_{\mu\lambda}^{\;\;\;\nu_1} T^{\lambda\nu_2\ldots} \;+\; \Gamma_{\mu\lambda}^{\;\;\;\nu_2} T^{\nu_1\lambda\ldots} \;+\; \ldots \tag{4.38}$$

Evenzo kunnen we een covariante afgeleide van een covariante vector of tensor definiëren door de uitdrukking

$$D_\mu T_{\nu_1\nu_2\ldots} = \partial_\mu T_{\nu_1\nu_2\ldots} - \Gamma_{\mu\nu_1}{}^{\lambda} T_{\lambda\nu_2\ldots} - \Gamma_{\mu\nu_2}{}^{\lambda} T_{\nu_1\lambda\ldots} - \ldots \tag{4.39}$$

We kunnen dus vaststellen: als $T_{\nu_1\ldots\nu_r}$ een covariante tensor van rang r is, dan is de covariante afgeleide $(DT)_{\mu\nu_1\ldots\nu_r} \equiv D_\mu T_{\nu_1\ldots\nu_r}$ een covariante tensor van rang $r+1$. Voor contravariante en gemengde tensoren geldt iets soortgelijks.

Een bijzondere eigenschap van de covariante afgeleide is, dat de metrische tensor en zijn inverse covariant constant zijn:

$$D_\lambda g_{\mu\nu} = 0, \qquad D_\lambda g^{\mu\nu} = 0. \tag{4.40}$$

Om dit te zien, schrijven we de covariante afgeleide uit:

$$D_\lambda g_{\mu\nu} = \partial_\lambda g_{\mu\nu} - \Gamma_{\lambda\mu}{}^{\kappa} g_{\kappa\nu} - \Gamma_{\lambda\nu}{}^{\kappa} g_{\mu\kappa}. \tag{4.41}$$

Gebruiken we nu de notatie

$$\Gamma_{\lambda\mu\nu} = \Gamma_{\lambda\mu}{}^{\kappa} g_{\kappa\nu} = \frac{1}{2}\left(\partial_\lambda g_{\mu\nu} + \partial_\mu g_{\lambda\nu} - \partial_\nu g_{\lambda\mu}\right), \tag{4.42}$$

dan volgt hieruit dat

$$\Gamma_{\lambda\mu\nu} + \Gamma_{\lambda\nu\mu} = \partial_\lambda g_{\mu\nu} \tag{4.43}$$

Vergelijking van dit resultaat met de uitdrukking (4.41) leidt dan onmiddellijk tot de conclusie dat de metrische tensor inderdaad covariant constant is. Voor de inverse metriek geldt dan natuurlijk hetzelfde. Formeel kunnen we dit b.v. als volgt bewijzen; merk eerst op dat de Kronecker delta niet alleen een constante tensor is, maar tevens covariant constant:

$$D_\mu \delta_\nu^\lambda = \partial_\mu \delta_\nu^\lambda + \Gamma_{\mu\kappa}{}^{\lambda} \delta_\nu^\kappa - \Gamma_{\mu\nu}{}^{\kappa} \delta_\kappa^\lambda = \Gamma_{\mu\nu}{}^{\lambda} - \Gamma_{\mu\nu}{}^{\lambda} = 0. \tag{4.44}$$

Dit betekent, dat

$$\begin{aligned} D_\mu\left(g_{\nu\kappa} g^{\kappa\lambda}\right) &= D_\mu \delta_\nu^\lambda \\ &= \left(D_\mu g_{\nu\kappa}\right) g^{\kappa\lambda} + g_{\nu\kappa} D_\mu g^{\kappa\lambda} = g_{\nu\kappa} D_\mu g^{\kappa\lambda} = 0. \end{aligned} \tag{4.45}$$

Het rechterlid van de laatste regel volgt uit vgl.(4.44), terwijl het middelste lid uit het linkerlid volgt, omdat $g_{\nu\kappa}$ zelf covariant constant is. Tenslotte staan in het linkerlid impliciet twee termen die tegen elkaar wegvallen: een term $\Gamma_{\mu\kappa}{}^{\sigma}\, g_{\nu\sigma}\, g^{\kappa\lambda}$ die uit de covariante afgeleide van $g_{\nu\kappa}$ komt met een minteken, terwijl dezelfde term uit de covariante afgeleide van $g^{\kappa\lambda}$ komt met een plusteken.

In het voorgaande hebben we de vorm van de connectie afgeleid uit de vergelijking voor geodetische krommen, en laten zien dat met deze uitdrukking voor de connectie de metriek covariant constant is. We zouden natuurlijk ook de omgekeerde weg kunnen bewandelen, en de connectie bepalen door te eisen dat de metriek covariant constant is, waarna afgeleid wordt dat deze connectie dan ook de geodeten bepaalt. In deze aanpak is de covariante onveranderlijkheid van de metriek het uitgangspunt; vergelijking (4.40) die dit vastlegt heet dan het *metrisch postulaat*. Tenslotte merken we op dat de definitie van de covariante afgeleide op co- en contravariante componenten van een vector of tensor zo gekozen zijn, dat de Leibnizregel ook voor vermenigvuldiging van vectoren en tensoren blijft gelden; zo geeft het invullen van de definities direct dat

$$D_\lambda \left(A_\mu B_\nu \right) \,=\, \left(D_\lambda A_\mu \right) \, B_\nu \,+\, A_\mu D_\lambda B_\nu. \tag{4.46}$$

Het bewijs hiervan laten we aan de lezer over.

4.5 Invariantie van de totale viersnelheid

De definitie (4.13) van de eigentijd:

$$c^2 d\tau^2 = -g_{\mu\nu} dx^\mu dx^\nu$$

houdt in dat het kwadraat van de viersnelheid een invariante constante is:

$$u_\mu u^\mu \,=\, g_{\mu\nu} \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} = -c^2. \tag{4.47}$$

Fysisch is deze relatie direct duidelijk uit de interpretatie van de eigentijd als de tijd gemeten in het momentane rustsstelsel van het bewegende punt, waarin $dx^0/d\tau = cdt/d\tau = c$ terwijl alle andere componenten van de viersnelheid nul zijn. Omdat beide zijden van vergelijking (4.47) een scalaire grootheid vormen, heeft het kwadraat van de viersnelheid deze waarde echter in alle coördinatenstelsels.

De geodetische vergelijking en het metrisch postulaat samen garanderen nu de consistentie van dit resultaat met de bewegingsvergelijking (4.5):

$$\frac{d}{d\tau} \left(g_{\mu\nu} u^\mu u^\nu \right) \,=\, 2 u_\mu \frac{Du^\mu}{D\tau} \,=\, 0. \tag{4.48}$$

In de laatste uitdrukking gebruiken we de geodetische vergelijking in de notatie

$$\frac{Du^\mu}{D\tau} \,\equiv\, \frac{du^\mu}{d\tau} \,+\, \Gamma_{\lambda\nu}{}^{\mu} u^\lambda u^\nu \,=\, 0. \tag{4.49}$$

Ook gebruiken we de covariante componenten van de viersnelheid $u_\mu = g_{\mu\nu} u^\nu$. Om de covariante afgeleide $Du^\mu/D\tau$ in het rechterlid van vgl.(4.48) te krijgen hebben we het metrisch postulaat gebruikt in de vorm

$$\frac{dg_{\mu\nu}}{d\tau} \,=\, u^\lambda \partial_\lambda g_{\mu\nu} \,=\, u^\lambda \left(\Gamma_{\lambda\mu}{}^{\kappa} g_{\kappa\nu} + \Gamma_{\lambda\nu}{}^{\kappa} g_{\mu\kappa} \right). \tag{4.50}$$

Vgl.(4.48) leert ons dat $u_\mu u^\mu$ constant is, in overeenstemming met vgl.(4.47).

4.6 Kromming

In het voorgaande is gebleken, dat alleen relatieve versnellingen objectief vastgesteld kunnen worden; een constant zwaartekrachtveld, een veld waarin de versnelling overal gelijk is in grootte en richting, bestaat niet voor een waarnemer in vrije val. Het zijn dan ook de gradienten in het zwaartekrachtveld die een waarnemeronafhankelijke betekenis hebben. In de meetkundige interpretatie van de zwaartekracht bepaalt de kromming van de tijd-ruimte de sterkte van zulke relatieve versnellingen. De Riemanntensor is een maat voor die kromming. Daarom zegt deze tensor iets over de sterkte van de gradienten in het zwaartekrachtveld. We zullen dit eerst illustreren aan de hand van een eenvoudig voorbeeld; daarna bespreken we de algemene theorie.

Beschouw het tweedimensionale oppervlak van een bol; op dit oppervlak zijn de grote cirkels de lijnen die de kortste verbinding tussen twee punten vormen. Neem drie grote cirkels, b.v. de evenaar en twee meridianen door de noordpool. Twee deeltjes die vanaf de noordpool elk met een constante snelheid langs een meridiaan bewegen, dus langs een geodetische lijn, zullen zich eerst van elkaar verwijderen met een snelheid die langzaam afneemt, tot bij de evenaar de deeltjes op gelijke afstand blijven; daarna komen ze weer steeds dichter bij elkaar. Hieruit blijkt dat hun onderlinge snelheid verandert, en halverwege zelfs van teken omkeert. De deeltjes ondergaan dus een relatieve versnelling, die ze objectief aan een zwaartekrachtveld zouden kunnen toeschrijven; immers, de baan en de versnelling zijn onafhankelijk van de aard van de deeltjes. Deze versnelling is niet constant: de relatieve snelheid van de deeltjes verandert bij de evenaar van teken, maar is bij de zuidpool weer maximaal, om daarna weer af te nemen, enz. De versnelling wisselt dus ook van positief naar negatief en terug. Het is duidelijk dat bij gegeven beginsnelheden (grootte en richting) de relatieve versnelling afhangt van de van de straal van de bol. Bij een grote straal neemt de afstand tussen de deeltjes vrijwel lineair in de tijd toe, dus is de snelheid bijna constant en de versnellingen zijn klein. Naarmate de straal kleiner wordt is de relatieve snelheidsverandering tussen de deeltjes groter en ook de verandering in de versnelling; de relatieve versnelling zowel als de gradient van de relatieve versnelling tussen de banen van de deeltjes hangen in dit voorbeeld af van de kromming.

Een manier om de mate van kromming te karakteriseren met meetkundige gegevens van het boloppervlak zelf, onafhankelijk van de inbedding van het oppervlak in de drie-dimensionale ruimte, is de volgende. Beschouw de driehoek gevormd door de cirkelbogen van de meridianen tussen de noordpool en de evenaar, en de boog van de evenaar tussen de meridianen. Beschouw een vector op de noordpool, bij voorbeeld de raakvector aan een van de meridianen. Verplaats deze evenwijdig aan zichzelf langs de meridiaan tot de evenaar. Hij staat dan loodrecht op de evenaar. Verschuif hem nu evenwijdig aan zichzelf langs de evenaar tot aan de andere meridiaan. Hij staat dan nog steeds loodrecht op de evenaar; daarom is hij ook weer een raakvector aan de tweede meridiaan.

Verschuif deze raakvector tenslotte weer terug naar de noordpool, altijd evenwijdig aan de meridiaan, om te ontdekken dat hij t.o.v. de oorspronkelijke vector gedraaid is over een hoek corresponderend met de booglengte afgelegd langs de evenaar. De conclusie is, dat vectoren die altijd evenwijdig aan zichzelf langs een gesloten baan worden verplaatst niet evenwijdig aan zich zelf zijn bij terugkomst. Merk op, dat dit in een plat vlak nooit zou kunnen gebeuren. Voor kleine hoeken $\Delta\theta$ kunnen we de grootte van de verandering in de vector berekenen door de lineaire benadering

$$\Delta V = V \Delta\theta. \tag{4.51}$$

Tegelijk is het oppervlak $\Delta\Sigma$ dat door de drie bogen wordt omsloten gelijk aan

$$\Delta\Sigma = R^2 \Delta\theta, \tag{4.52}$$

waarin R de straal van de bol is. Combinatie van deze twee vergelijkingen geeft

$$\Delta V = K V \Delta\Sigma, \tag{4.53}$$

met $K = 1/R^2$, de zgn. Gaußische kromming van het oppervlak. Deze is groter naarmate de straal kleiner is.

De generalisatie van het begrip kromming naar hoger-dimensionale ruimten is gebaseerd op hetzelfde idee: wanneer een vector evenwijdig wordt verplaatst langs een gesloten lijn, is de verandering in de vector na terugkeer evenredig met de lengte van de vector en de grootte van het omsloten oppervlak:

$$\Delta V^\mu = -\frac{1}{2}\, R_{\kappa\lambda\nu}{}^{\mu}\, V^\nu\, \Delta\Sigma^{\kappa\lambda}. \tag{4.54}$$

In de uitdrukking aan de rechterkant van deze vergelijking is $\Delta\Sigma^{\kappa\lambda} = -\Delta\Sigma^{\lambda\kappa}$ de projectie van het georiënteerde oppervlakteëlement $\Delta\Sigma$ in het x^κ-x^λ-vlak, en de factor $1/2$ compenseert de dubbeltelling in de sommatie over alle projecties, die komt op rekening van de antisymmetrie in de indices van $\Delta\Sigma^{\kappa\lambda}$. De evenredigheidsconstanten $R_{\kappa\lambda\nu}{}^{\mu}$ zijn de componenten van de krommingstensor. We zullen laten zien dat deze identiek is met de Riemanntensor die we eerder hebben ingevoerd. De huidige definitie benadrukt de zuiver meetkundige betekenis van deze grootheid.

We beginnen ermee het begrip evenwijdige verschuiving een precieze betekenis te geven. Beschouw een contravariant vectorveld $V^\mu(x)$ en een kromme $x^\mu(\lambda)$. De raakvector aan de kromme is $\dot{x}^\mu(\lambda) = dx^\mu/d\lambda$. De covariante verandering van de vector onder verplaatsing langs de kromme is dan

$$\dot{x}^\nu D_\nu V^\mu = \dot{V}^\mu + \dot{x}^\nu \Gamma_{\nu\lambda}{}^{\mu} V^\lambda. \tag{4.55}$$

Zij $\mathcal{P}(\lambda)$ het punt op de kromme dat hoort bij een geven waarde van de parameter λ. $V^\mu(\lambda)$ zijn dan de componenten van het vectorveld in dit punt. De componenten van het vectorveld in een naburig punt $\mathcal{P}(\lambda')$, met $\lambda' = \lambda + \Delta\lambda$, zijn $V^\mu(\lambda + \Delta\lambda) = V^\mu(\lambda) + \Delta V^\mu(\lambda)$, met

$$\Delta V^\mu(\lambda) \approx \Delta\lambda \frac{dV^\mu}{d\lambda} = \Delta\lambda\, \dot{V}^\mu. \tag{4.56}$$

We zeggen nu dat de vector in het naburige punt $\mathcal{P}(\lambda')$ is verkregen uit de oorspronkelijke vector in $\mathcal{P}(\lambda)$ door evenwijdige verplaatsing, indien

$$\Delta V^\mu = -\Delta\lambda\, \dot{x}^\nu\, \Gamma_{\nu\lambda}{}^\mu V^\lambda = -\Delta x^\nu\, \Gamma_{\nu\lambda}{}^\mu V^\lambda. \tag{4.57}$$

Deze definitie is equivalent met de vaststelling dat het vectorveld in het punt $\mathcal{P}$ covariant constant is:

$$\dot{x}^\nu D_\nu V^\mu(\mathcal{P}) = 0. \tag{4.58}$$

De covariantie houdt in, dat deze vergelijking dezelfde vorm houdt in ieder coördinatenstelsel. I.h.b., in lokale Minkowski-coördinaten (een stelsel in vrije val in $\mathcal{P}$) is $\Gamma_{\nu\lambda}{}^\mu(\mathcal{P}) = 0$; in dat stelsel is dus $\Delta V^\mu(\mathcal{P}) = 0$. Hieruit blijkt, dat onze definitie van evenwijdige verplaatsing zo is gekozen, dat de componenten van V^μ onder zo'n verplaatsing in een omgeving van $\mathcal{P}$ constant zijn in het speciale coördinatenstelsel gevormd door een Minkowskistelsel in vrije val in $\mathcal{P}$.

We gaan nu deze definitie van evenwijdige verplaatsing gebruiken om de verandering van een vector onder parallelverschuiving langs een gesloten baan te berekenen. We beschouwen voor de eenvoud een lus gevormd door de vierhoek bestaande uit lijnelementen Δx^κ en Δx^λ tussen een punt $\mathcal{P}_0$ met coördinaten x_0^μ, en $\mathcal{P}_1$ het punt met coördinaten $x_1^\kappa = x_0^\kappa + \Delta x^\kappa$ en $x_1^\lambda = x_0^\lambda + \Delta x^\lambda$, terwijl alle ander coördinaten onveranderd zijn; de vierhoek wordt verkregen door hieraan het punt $\mathcal{P}_2$ toe te voegen met $x_2^\kappa = x_0^\kappa + \Delta x^\kappa$ en alle andere coördinaten gelijk aan die van $\mathcal{P}_0$, en evenzo het punt $\mathcal{P}_3$ met $x_3^\mu = x_0^\mu + \Delta x^\mu$ waarbij $\Delta x^\mu = 0$, $\mu \neq \lambda$. We berekenen nu de verandering in de componenten van een vector bij evenwijdige verplaatsing van $\mathcal{P}_0$ naar $\mathcal{P}_1$ via $\mathcal{P}_2$, zowel als de verandering onder verplaatsing via $\mathcal{P}_3$. Volgens vgl.(4.57) is de verandering van de vector tussen $\mathcal{P}_0$ en $\mathcal{P}_2$ bij parallelverplaatsing gegeven door

$$\Delta_1 V^\mu = -\Delta_1 x^\kappa \Gamma_{\kappa\nu}{}^\mu V^\nu. \tag{4.59}$$

Hierin zijn alle componenten van de vector $\Delta_1 x$ gelijk aan nul, behalve $\Delta_1 x^\kappa$. Evenzo nemen we $\Delta_2 x$ als de vector met alle componenten gelijk aan nul behalve $\Delta_2 x^\lambda$. Wanneer we de vector V^μ vervolgens verder evenwijdig verschuiven van $\mathcal{P}_2$ naar $\mathcal{P}_1$ verandert de variatie $\Delta_1 V^\mu$ met een bedrag

$$\Delta_2\left(\Delta_1 V^\mu\right) = -\Delta_2 x^\lambda \Delta_1 x^\kappa \left(\partial_\lambda \Gamma_{\kappa\nu}{}^\mu V^\nu - \Gamma_{\kappa\sigma}{}^\mu \Gamma_{\lambda\nu}{}^\sigma V^\nu\right). \tag{4.60}$$

Wanneer we nu het alternatieve pad volgen, van $\mathcal{P}_0$ naar $\mathcal{P}_1$ via $\mathcal{P}_3$, dan krijgen we een soortgelijke uitdrukking met nu de rol van $\Delta_1 x$ en $\Delta_2 x$ verwisseld. Het verschil in de verandering van de componenten van V^μ langs het ene en langs het andere pad is nu

$$\begin{aligned}(\Delta_2\Delta_1 - \Delta_1\Delta_2)\, V^\mu &= -\Delta_2 x^\lambda \Delta_1 x^\kappa \left(\partial_\lambda \Gamma_{\kappa\nu}{}^\mu - \partial_\kappa \Gamma_{\lambda\nu}{}^\mu - \Gamma_{\lambda\nu}{}^\sigma \Gamma_{\kappa\sigma}{}^\mu + \Gamma_{\kappa\nu}{}^\sigma \Gamma_{\lambda\sigma}{}^\mu\right) V^\nu \\ &= -\frac{1}{2}\, \Delta\Sigma^{\kappa\lambda}\, R_{\kappa\lambda\nu}{}^\mu V^\nu,\end{aligned} \tag{4.61}$$

waarin we de notatie

$$\Delta\Sigma^{\lambda\kappa} = \Delta_1 x^\kappa \Delta_2 x^\lambda - \Delta_1 x^\lambda \Delta_2 x^\kappa \tag{4.62}$$

hebben gebruikt. Deze uitdrukking is de determinant van de vectoren $\Delta_1 x$ en $\Delta_2 x$ in het x^κ-x^λ-vlak, en vertegenwoordigt het geörienteerde oppervlakteëlement dat door deze vectoren wordt opgespannen. Hiermee ligt de meetkundige interpretatie van de Riemanntensor vast. Merk op, dat de afleiding van deze gelijkheid ook het tensorkarakter van de Riemanntensor bevestigt, omdat de andere grootheden ter linker- en ter rechterzijde van het gelijkteken in vgl.(4.54) covariant transformeren onder coördinatentransformaties. Dat is niet vanzelfsprekend, omdat de Riemanntensor uit de connectie en zijn afgeleiden is gevormd, en de connectie zelf zich niet als een tensor onder algemene coördinatentranformaties gedraagt. Een directe manier om te bewijzen dat in de transformatie van de Riemanntensor alle inhomogene termen met tweede afgeleiden uit de transformatie van de connectie tegen elkaar wegvallen is door de transformatie volledig uit te schrijven. Dit is een goede oefening in indexmanipulatie. Een tweede algebraisch argument, dat nauw aansluit bij de meetkundige interpretatie van de Riemanntensor, is het volgende. Beschouw voor de eenvoud een covariant vectorveld $V_\lambda(x)$. De covariante afgeleide $D_\nu V_\lambda$ is dan een covariante tensor van de rang twee. Vervolgens is de covariante afgeleide hiervan weer een covariante tensor van de rang drie:

$$D_\mu D_\nu V_\lambda = \partial_\mu\left(D_\nu V_\lambda\right) - \Gamma_{\mu\nu}{}^\kappa D_\kappa V_\lambda - \Gamma_{\mu\lambda}{}^\kappa D_\nu V_\kappa. \tag{4.63}$$

We gaan deze uitdrukking nu antisymmetriseren zodat we de commutator van twee covariante afgeleiden krijgen:

$$\begin{aligned}\left[D_\mu, D_\nu\right] V_\lambda &= D_\mu D_\nu V_\lambda - D_\nu D_\mu V_\lambda \\ &= \partial_\mu\left(D_\nu V_\lambda\right) - \partial_\nu\left(D_\mu V_\lambda\right) - \Gamma_{\mu\lambda}{}^\kappa D_\nu V_\kappa + \Gamma_{\nu\lambda}{}^\kappa D_\mu V_\kappa.\end{aligned} \tag{4.64}$$

In de laatste uitdrukking is al gebruik gemaakt van de symmetrie van de connectie $\Gamma_{\mu\nu}{}^\lambda = \Gamma_{\nu\mu}{}^\lambda$ om enige termen te elimineren. Als we nu de covariante afgeleiden van V_λ zelf uitschrijven: $D_\nu V_\lambda = \partial_\nu V_\lambda - \Gamma_{\nu\lambda}{}^\kappa V_\kappa$, en rekening houden met de antisymmetrisatie waardoor o.a. de termen met twee gewone partiële afgeleiden tegen elkaar wegvallen, dan krijgen we tenslotte

$$\begin{aligned}\left[D_\mu, D_\nu\right] V_\lambda &= -\left(\partial_\mu \Gamma_{\nu\lambda}{}^\kappa - \partial_\nu \Gamma_{\mu\lambda}{}^\kappa - \Gamma_{\mu\lambda}{}^\sigma \Gamma_{\sigma\nu}{}^\kappa + \Gamma_{\nu\lambda}{}^\sigma \Gamma_{\sigma\mu}{}^\kappa\right) V_\kappa \\ &= -R_{\mu\nu\lambda}{}^\kappa V_\kappa.\end{aligned} \tag{4.65}$$

Deze identiteit, die vastlegt dat de commutator van twee covariante afgeleiden een contractie met de Riemanntensor oplevert, staat bekend als de Ricci-identiteit. Deze geldt niet alleen voor covariante vectoren, maar in aangepaste vorm voor covariante en contravariante tensoren van iedere rang ≥ 1. We kunnen nu opmerken, dat de linkerzijde van de identiteit (4.65) transformeert als een tensor, in dit geval een van rang drie, terwijl V_κ de componenten van een covariante vector zijn. Dan moet ook de Riemanntensor correct als een tensor transformeren. Daarmee is een algebraisch bewijs van het tensorkarakter van $R_{\mu\nu\kappa}{}^{\lambda}$ geleverd.

Behalve de Ricci-identiteit (4.65) die is gebaseerd op de commutator van twee covariante afgeleiden, zijn er nog twee identiteiten die bekend staan als Bianchi-identiteiten, gebaseerd op de dubbele commutator van drie covariante afgeleiden. De eigenlijke onderliggende identiteit is de Jacobi-identiteit voor lineaire operatoren

$$[D_\mu, [D_\nu, D_\lambda]] + [D_\nu, [D_\lambda, D_\mu]] + [D_\lambda, [D_\mu, D_\nu]] = 0. \tag{4.66}$$

Passen we deze identiteit weer toe op b.v. een covariant vectorveld, en scheiden we na toepassen van de Ricci-identiteit de termen die alleen het vectorveld vermenigvuldigen van de termen die de afgeleide van het vectorveld bevatten, dan krijgen we de volgende twee identiteiten:

$$\begin{aligned} R_{\mu\nu\lambda}{}^{\kappa} + R_{\nu\lambda\mu}{}^{\kappa} + R_{\lambda\mu\nu}{}^{\kappa} &= 0, \\ D_\mu R_{\nu\lambda\kappa}{}^{\sigma} + D_\nu R_{\lambda\mu\kappa}{}^{\sigma} + D_\lambda R_{\mu\nu\kappa}{}^{\sigma} &= 0. \end{aligned} \tag{4.67}$$

Wanneer we van de laatste identiteit een dubbel spoor nemen, vinden we

$$2D_\nu R^\nu{}_\mu = \partial_\mu R \quad \Leftrightarrow \quad D_\nu \left(R^\nu{}_\mu - \frac{1}{2}\, \delta^\nu_\mu R \right) = D_\nu G^\nu{}_\mu = 0. \tag{4.68}$$

Vergelijking van dit resultaat met de Einsteinvergelijking (4.9), (4.10) leidt dan tot de conclusie dat de energie-impulstensor die als bron voor de gravitatievelden fungeert, moet voldoen aan de consistentieëis dat

$$D_\nu T^\nu{}_\mu = 0. \tag{4.69}$$

4.7 Symmetrieën van de Riemanntensor

Uit de definitie van de Riemanntensor volgt direct, dat

$$R_{\mu\nu\kappa}{}^{\lambda} = -R_{\nu\mu\kappa}{}^{\lambda}. \tag{4.70}$$

Tevens voldoet de Riemanntensor aan de Bianchi-identiteiten (4.67), in het bijzonder

$$R_{\mu\nu\kappa}{}^{\lambda} + R_{\nu\kappa\mu}{}^{\lambda} + R_{\kappa\mu\nu}{}^{\lambda} = 0. \tag{4.71}$$

Beschouwen we nu de volledig covariante componenten van de Riemanntensor: $R_{\mu\nu\kappa\lambda} = R_{\mu\nu\kappa}{}^{\sigma} g_{\sigma\lambda}$, dan gelden voor deze componenten de volgende symmetrie-eigenschappen:

$$R_{\mu\nu\kappa\lambda} = -R_{\nu\mu\kappa\lambda} = -R_{\mu\nu\lambda\kappa} = R_{\kappa\lambda\mu\nu}. \tag{4.72}$$

Het belangrijkste bestanddeel van het bewijs van deze identiteiten is de tweede stap; samen met de eerste gelijkheid, en de Bianchi-identiteit (4.71) twee maal toegepast, volgt daaruit dan de laatste gelijkheid. Om de tweede stap te bewijzen beginnen we met de expliciete uitdrukking voor de covariante vorm van de Riemanntensor:

$$\begin{aligned} R_{\mu\nu\kappa\lambda} &= \partial_\mu \Gamma_{\nu\kappa\lambda} - \Gamma_{\nu\kappa}{}^{\sigma} \left(\Gamma_{\mu\lambda\sigma} + \Gamma_{\mu\sigma\lambda}\right) - \Gamma_{\mu\kappa}{}^{\sigma} \Gamma_{\nu\sigma\lambda} - (\mu \leftrightarrow \nu) \\ &= \partial_\mu \Gamma_{\nu\kappa\lambda} - \Gamma_{\nu\kappa}{}^{\sigma} \Gamma_{\mu\lambda\sigma} - (\mu \leftrightarrow \nu) \end{aligned} \tag{4.73}$$

Om de uitdrukking ter rechterzijde op de eerste regel van (4.73) te krijgen hebben we gebruik gemaakt van het metrisch postulaat in de vorm (4.43), om termen met de afgeleide van de metrische tensor te vervangen door een product van connecties. Uit de laatste regel zien we dat de termen met produkten van connecties manifest antisymmetrisch zijn onder verwisseling van hetzij $[\mu\nu]$, hetzij $[\kappa\lambda]$. Dezelfde eigenschap geldt ook voor de termen met afgeleiden van $\Gamma_{\nu\kappa\lambda}$, maar om dat te zien moet hiervoor de expliciete uitdrukking in termen van afgeleiden van de metriek worden ingevuld:

$$\Gamma_{\nu\kappa\lambda} = \frac{1}{2} \left(\partial_\nu g_{\kappa\lambda} + \partial_\kappa g_{\nu\lambda} - \partial_\lambda g_{\nu\kappa}\right).$$

Bij antisymmetrisatie in $[\mu\nu]$ valt de eerste term in de uitdrukking voor $\partial_\mu \Gamma_{\nu\kappa\lambda}$ nu weg omdat gewone partiële afgeleiden verwisselbaar zijn. De resterende termen zijn weer manifest antisymmetrisch in zowel $[\mu\nu]$ als $[\kappa\lambda]$.

4.8 Variatieprincipe

Met behulp van het tensorformalisme dat in de voorgaande paragrafen is besproken kunnen de vergelijkingen van de algemene relativiteitstheorie geschreven worden op een manier die manifest covariant is onder algemene coördinaten transformaties; d.w.z., het linker- en rechterlid van een vergelijking transformeren onder een coördinatentransformatie op dezelfde manier. We geven nu een voorschrift hoe de integratie van functies over het ruimte-tijdcontinuüm te definiëren op een manier die invariant is onder coördinatentransformaties. Daartoe merken we eerst op dat onder een algemene coördinatentransformatie $x^\mu \to \bar{x}^\mu$ het volumeëlement overgaat in

$$d^4x \;\to\; d^4\bar{x} = d^4x \left|\frac{\partial \bar{x}}{\partial x}\right|. \tag{4.74}$$

Om een invariante integratie te krijgen voeren we in de integratie een schaalfactor voor het volumeëlement in. Deze moet zo worden gekozen, dat de jacobiaan van de transformatie die in de laatste uitdrukking optreedt, kan worden gecompenseerd door een inverse jacobiaan uit de transformatie van de schaalfactor. Een grootheid die het geschikte transformatiegedrag heeft om als schaalfactor op te kunnen treden, is te vormen uit de determinant van de metriek, $g \equiv \det g_{\mu\nu}$. Uit de transformatieëigenschap (4.14) van de metriek volgt namelijk dat

$$\det g_{\mu\nu} = (\det \bar{g}_{\kappa\lambda})(\det \frac{\partial \bar{x}^{\kappa}}{\partial x^{\mu}})(\det \frac{\partial \bar{x}^{\lambda}}{\partial x^{\nu}}), \tag{4.75}$$

ofwel $g = \bar{g}\, |\partial \bar{x}/\partial x|^2$. We moeten rekening houden met het lorentz-karakter van de metriek, dat bepaalt dat $g_{\mu\nu}$ als matrix drie positieve en één negatieve eigenwaarde heeft, zodat $g = \det g_{\mu\nu} < 0$. Dan is door omkering van de vergelijking (4.75)

$$\sqrt{-\bar{g}} = \left| \frac{\partial x}{\partial \bar{x}} \right| \sqrt{-g} = \left| \frac{\partial \bar{x}}{\partial x} \right|^{-1} \sqrt{-g}. \tag{4.76}$$

Uit de combinatie van vgl.(4.74), (4.76) volgt nu, dat $\sqrt{-g}$ een geschikte schaalfactor voor het volumeëlement is:

$$d^4\bar{x}\sqrt{-\bar{g}} = d^4x\sqrt{-g}. \tag{4.77}$$

Als $\Phi(x)$ onder coördinatentransformaties een scalaire grootheid is: $\overline{\Phi}(\bar{x}) = \Phi(x)$, dan is een invariante integraal van $\Phi(x)$ dus te definiëren als

$$I \;=\; \int d^4x\, \sqrt{-g}\, \Phi(x) \;=\; \int d^4\bar{x}\, \sqrt{-\bar{g}}\, \overline{\Phi}(x). \tag{4.78}$$

Na deze inleiding over invariante integratie, keren we terug naar de Einsteinvergelijkingen (4.9), (4.10). Deze houden in dat de ruimte-tijd metriek wordt bepaald uit

$$G_{\mu\nu} \;=\; R_{\mu\nu} - \frac{1}{2}\, g_{\mu\nu} R \;=\; -\,\frac{8\pi G}{c^4}\, T_{\mu\nu}.$$

We zullen nu laten zien dat deze vergelijkingen kunnen worden afgeleid uit een eenvoudig variatieprincipe. De actie voor dit variatievoorschrift is een invariante integraal in de bovenbeschreven zin; i.h.b volgt de Einsteintensor uit de variatie van de invariante integraal van de krommingsscalar:

$$\delta \int d^4x\, \sqrt{-g}\, R \;=\; \int d^4x\, \sqrt{-g}\; \delta g^{\mu\nu} \left(R_{\mu\nu} - \frac{1}{2}\, g_{\mu\nu} R \right). \tag{4.79}$$

De oorsprong van de eerste term in het rechterlid is de variatie van de krommingsscalar $R = g^{\mu\nu} R_{\mu\nu}$; de volledige variatie van deze factor is

$$\delta R = \delta g^{\mu\nu} R_{\mu\nu} + g^{\mu\nu} \delta R_{\mu\nu}. \tag{4.80}$$

Als we de tweede term weglaten krijgen we de gewenste bijdrage aan het variatieprincipe. De rechtvaardiging van het weglaten van de term met de variatie van de Riccitensor $\delta R_{\mu\nu}$ bespreken we later. De tweede term in het rechterlid van (4.79) komt van de variatie van $\sqrt{-g}$. Dit leiden we als volgt af:

$$\delta\sqrt{-g} = \sqrt{-g}\,\frac{1}{\sqrt{-g}}\,\delta\sqrt{-g} \;=\; \frac{1}{2}\,\sqrt{-g}\,\delta\ln(-g). \tag{4.81}$$

Nu is $g_{\mu\nu}$ een symmetrische matrix, die we in elk willekeurig punt kunnen diagonaliseren door een lokale 4-dimensionale rotatie uit te voeren. Als de vier lokale eigenwaarden $(\lambda_0(x), \lambda_1(x), \lambda_2(x), \lambda_3(x))$ zijn, dan is[2]

$$\begin{aligned} \ln\{-g(x)\} \;&=\; \ln\left\{\prod_{i=0}^{3}|\lambda_i(x)|\right\} \;=\; \sum_{i=0}^{3}\ln|\lambda_i(x)| \\ &=\; \mathrm{Sp}\left\{\ln|g_{\mu\nu}(x)|\right\}. \end{aligned} \tag{4.82}$$

We kunnen deze redenering als volgt verder voeren:

$$\ln(-g) \;=\; \frac{1}{2}\sum_{i=0}^{3}\ln\lambda_i^2 \quad\Rightarrow\quad \delta\ln(-g) \;=\; \sum_{i=0}^{3}\frac{\delta\lambda_i}{\lambda_i}.$$

Op deze wijze, en met gebruik van de inverse relatie tussen $g_{\mu\nu}$ en $g^{\mu\nu}$, volgt uit het voorgaande dat

$$\delta\ln(-g) \;=\; g^{\mu\nu}\delta g_{\mu\nu} \;=\; -g_{\mu\nu}\delta g^{\mu\nu}. \tag{4.83}$$

Bij elkaar geven de vergelijkingen (4.81)-(4.83) dat

$$\delta\sqrt{-g} \;=\; -\,\frac{1}{2}\,\delta g^{\mu\nu}g_{\mu\nu}\sqrt{-g}. \tag{4.84}$$

Hiermee is de tweede term in de variatie (4.79) verklaard. Om deze discussie af te ronden moeten we nog laten zien, dat de term

$$\int d^4x\,\sqrt{-g}\,g^{\mu\nu}\delta R_{\mu\nu}. \tag{4.85}$$

in de variatie (4.79) geen rol speelt. Een rechtstreekse berekening laat zien dat deze term omgevormd kan worden tot een totale afgeleide onder de integraal, d.w.z. een randterm die bij een verstandige keuze van de variaties $\delta g^{\mu\nu}$ verdwijnt. Zonder deze berekening in detail te reproduceren, kunnen we hier wel aangeven hoe dit resultaat tot stand komt.

[2] De logaritme van een diagonale matrix M is de diagonale matrix met de logaritmen van de eigenwaarden van M als diagonaalelementen; het spoor van die matrix is dan de som van de logaritmen op de diagonaal.

Eerst merken we op dat de Riccitensor $R_{\mu\nu} = R_{\mu\lambda\nu}{}^{\lambda}$, evenals de Riemann-tensor zelf, volledig uitgedrukt kan worden in termen van de connectie. Dit is manifestin vergelijking (4.6). De variatie van de Riccitensor is dus eveneens volledig uit te drukken in variaties van de connectie. Uit de expliciete uitdrukking voor de Riccitensor volgt dat

$$\begin{aligned} \delta R_{\mu\nu} &= \partial_\mu \delta\Gamma_{\lambda\nu}{}^{\lambda} - \partial_\lambda \delta\Gamma_{\mu\nu}{}^{\lambda} \\ &\quad - \Gamma_{\mu\nu}{}^{\kappa}\delta\Gamma_{\lambda\kappa}{}^{\lambda} - \Gamma_{\lambda\kappa}{}^{\lambda}\delta\Gamma_{\mu\nu}{}^{\kappa} + \Gamma_{\lambda\nu}{}^{\kappa}\delta\Gamma_{\mu\kappa}{}^{\lambda} + \Gamma_{\mu\kappa}{}^{\lambda}\delta\Gamma_{\lambda\nu}{}^{\kappa} \\ &= D_\mu \delta\Gamma_{\lambda\nu}{}^{\lambda} - D_\lambda \delta\Gamma_{\mu\nu}{}^{\lambda}. \end{aligned} \tag{4.86}$$

De laatste uitdrukking wordt verkregen door de variatie $\delta\Gamma_{\mu\nu}{}^{\lambda}$ als een tensor te behandelen en aan de middelste regel nul toe te voegen in de vorm:

$$-\Gamma_{\mu\lambda}{}^{\kappa}\delta\Gamma_{\kappa\nu}{}^{\lambda} + \Gamma_{\mu\lambda}{}^{\kappa}\delta\Gamma_{\kappa\nu}{}^{\lambda}. \tag{4.87}$$

Met de uitdrukking (4.86) voor $\delta R_{\mu\nu}$ in de integraal (4.85), en de eigenschap van de metriek en de inverse metriek dat deze covariant constant zijn, volgt dat de integraal na partiële integratie nul is; immers, we krijgen dan covariante afgeleiden die werken op $\sqrt{-g}g^{\mu\nu}$, wat identiek nul levert. De randtermen die optreden bij de partiële integratie verdwijnen vanwege de randvoorwaarden op de variatie $\delta g^{\mu\nu}(x)$.

Uit het voorgaande leiden we af, dat de Einsteinvergelijkingen in de lege ruimte: $G_{\mu\nu} = 0$, kunnen worden opgevat als de voorwaarde waaraan de metriek moet voldoen om de actie $\int \sqrt{-g}R$ onder infinitesimale variaties van de metriek stationair te laten zijn.

Om de Einsteinvergelijkingen af te leiden voor gravitatie in wisselwerking met materie —algemeen op te vatten als een verdeling van energie en impuls in de ruimte, zoals massa's, gassen of straling— moeten we aan deze invariante actie een term toevoegen die de dynamica van de materie in een tijd-ruimte met metriek $g_{\mu\nu}$ beschrijft. Zij $\{q(x)\}$ een stel grootheden die deze dynamica van de materie beschrijven. Zij $\mathcal{L}(q(x), g_{\mu\nu}(x))$ een scalaire functie die de actie levert voor de grootheden $q(x)$:

$$S_m = \int d^4x \sqrt{-g}\, \mathcal{L}(q, g_{\mu\nu}), \tag{4.88}$$

zodat de bewegingsvergelijkingen voor de materiële vrijheidsgraden kunnen worden geschreven als

$$\frac{\delta S_m}{\delta q(x)} = 0. \tag{4.89}$$

Dan is de totale actie voor gravitatie plus materie de combinatie van deze twee bijdragen:

$$S_{tot}[q, g_{\mu\nu}] = \int d^4x \sqrt{-g} \left(-\frac{c^4}{16\pi G} R + \mathcal{L}(q, g_{\mu\nu}) \right). \tag{4.90}$$

Natuurlijk volgen de bewegingsvergelijkingen voor de materie opnieuw uit

$$\frac{\delta S_{tot}}{\delta q(x)} = 0. \tag{4.91}$$

Aangezien R niet van de $q(x)$ afhangt, is deze vergelijking immers identiek met (4.89). Wanneer we nu de totale actie naar $g^{\mu\nu}$ variëren krijgen we echter een algemene vorm van de Einsteinvergelijkingen:

$$G_{\mu\nu} = -\frac{8\pi G}{c^4} T_{\mu\nu}, \tag{4.92}$$

met de energie-impulstensor gedefiniëerd door

$$T_{\mu\nu}(q, g_{\mu\nu}) = -\frac{2}{\sqrt{-g}} \frac{\delta S_m}{\delta g^{\mu\nu}}. \tag{4.93}$$

Dit is in de praktijk een veelgebruikte methode om energie-impulstensoren voor verschillende vormen van materie af te leiden. We zien hieruit dat actieprincipes ook in de ART een nuttige rol spelen.

Samenvatting van hoofdstuk 4

Dit hoofdstuk is gewijd aan de meetkunde van de tijd-ruimte. De begrippen metriek, connectie en kromming zijn ingevoerd, met een uiteenzetting hoe deze grootheden kunnen worden gebruikt om een veldentheorie van de zwaartekracht te formuleren die voldoet aan relativistisch-fysische eisen, zoals causaliteit en de universaliteit van de lichtsnelheid; dit is de algemene relativiteitstheorie (ART). Het begrip inertiaalstelsel kan daarin nog slechts een lokale betekenis hebben. Het beginsel van de equivalentie van zware en trage massa is ook in de ART van toepassing: alle lichamen met dezelfde beginvoorwaarden volgen in een gravitatieveld dezelfde baan, een geodetisch pad, ongeacht hun massa. De tensorrekening en het begrip covariante afgeleide passeren in dit hoofdstuk de revue, en worden toegepast in de constructie van de Einsteinvergelijkingen voor het gravitatieveld van een gegeven energie-impuls verdeling.

Hoofdstuk 5

Vrije gravitatievelden

5.1 Zwaartekrachtvelden in de lege ruimte

In de afwezigheid van materie en straling nemen de Einsteinvergelijkingen een eenvoudige vorm aan:

$$R_{\mu\nu} = R_{\mu\lambda\nu}{}^{\lambda} = 0. \tag{5.1}$$

Omdat de Riccitensor het spoor van de Riemanntensor is, kan de Riccitensor nul zijn zonder dat de componenten van de Riemanntensor zelf nul hoeven te zijn. Einsteins vergelijkingen laten daarom toe dat ook de lege tijd-ruimte een dynamische geometrie heeft, en dat er gravitatievelden voorkomen in gebieden waar geen materie of straling aanwezig zijn.

In dit hoofdstuk zullen we twee soorten van zulke vacuümoplossingen bespreken. De eerste soort zijn veldconfiguraties waarin een locale kromming zich met de lichtsnelheid door de lege ruimte voortplant; deze veldconfiguraties zijn te interpreteren als gravitatiegolven. De tweede soort vacuümoplossing die we bespreken is de statische bolsymmetrische oplossing, die het veld van een stationaire puntmassa of homogene bolsymmetrische massaverdeling beschrijft. Deze tegenhanger van het Coulombveld in de zwaartekrachtstheorie staat bekend als de Schwarzschildoplossing, en is van grote praktische betekenis in de astrofysica.

5.2 Gravitatiegolven

Een groot aantal oplossingen van de Einsteinvergelijkingen van verscheidene aard is bekend. Het is niet mogelijk in kort bestek een volledig overzicht en classificatie te geven; hiervoor verwijzen we naar de literatuur[1].

Onze aanpak hier is er op gericht inzicht te geven in de fysica van het zwaartekrachtveld; om de wiskundige uitwerking kort en overzichtelijk te houden zullen

[1] Zie b.v.: D. Kramer, H. Stephani, M. MacCallum en E. Herlt: *Exact solutions of Einstein's field equations*, Cambridge Univ. Press (1980).

we de vorm van de oplossingen zonder afleiding poneren, en laten zien dat het oplossingen zijn door ze eenvoudig in de veldvergelijkingen in te vullen. Een nadeel van deze methode is, dat het niet duidelijk is hoe algemeen de oplossing is; daar staat tegenover dat we uitgebreider op de interpretatie van de oplossingen zullen ingaan.

Om gravitatiegolven te beschrijven, gebruiken we in deze sectie de vier coördinaten $\{x^\mu\}_{\mu=1}^4 = (u, v, x, y)$. De coördinaten (x, y) beschrijven twee ruimtelijke dimensies, terwijl de lineaire combinaties $u = ct - z$ en $v = ct + z$ lichtkegelcoördinaten zijn: de vergelijkingen $u = constant$ en $v = constant$ beschrijven x-y-vlakken die zich met de lichtsnelheid in de positieve, repectievelijk negatieve, z-richting verplaatsen.

De oplossingen die we zoeken hebben in deze coördinaten een lijnelement gekarakteriseerd door een enkele functie $\Phi(u, x, y)$:

$$-c^2 d\tau^2 = g_{\mu\nu} dx^\mu dx^\nu = -du dv - \Phi(u, x, y)\, du^2 + dx^2 + dy^2. \tag{5.2}$$

Merk op dat voor $\Phi \to 0$ het lijnelement de standaard Minkowskivorm aanneemt, en dus de gewone (vlakke) vierdimensionale ruimte zonder gravitatie beschrijft. Merk voorts op, dat de functie Φ niet van v afhangt, zodat we ons beperken tot oplossingen waarvoor de amplitude op een vast punt in het (x, y)-vlak constant is voor constante u; zoals juist betoogd, betekent dit dat deze amplitude zich met de lichtsnelheid in de positieve z-richting beweegt. Uiteraard zouden we ook oplossingen kunnen zoeken die zich in de negatieve z-richting verplaatsen door de rol van u en v te verwisselen, en metrieken te nemen die alleen afhangen van een functie $\Psi(v, x, y)$.

We kunnen nu de componenten van de connectie $\Gamma_{\mu\nu}{}^\lambda$ en van de krommingstensor $R_{\mu\nu\kappa}{}^\lambda$ berekenen voor metrieken van de vorm (5.2). De matrixuitdrukking voor deze metriek en zijn inverse is

$$g_{\mu\nu} = \begin{pmatrix} -\Phi & -\frac{1}{2} & 0 & 0 \\ -\frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad g^{\mu\nu} = \begin{pmatrix} 0 & -2 & 0 & 0 \\ -2 & 4\Phi & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{5.3}$$

Invullen van deze vorm in vergelijking (4.3) geeft dan een beperkt aantal connectiecomponenten die niet nul zijn. De volledige lijst luidt:

$$\Gamma_{uu}{}^v = \Phi_{,u}, \quad \Gamma_{uu}{}^x = \frac{1}{2}\,\Gamma_{xu}{}^v = \frac{1}{2}\,\Phi_{,x}, \quad \Gamma_{uu}{}^y = \frac{1}{2}\,\Gamma_{yu}{}^v = \frac{1}{2}\,\Phi_{,y}. \tag{5.4}$$

De notatie van een komma gevolgd door een index μ staat hier voor een partiële afgeleide naar de coördinaat x^μ: $\Phi_{,\mu} = \partial_\mu \Phi$.

Gebruiken we deze uitdrukkingen vervolgens in vgl.(4.6) of (4.73) om de componenten van de volledig covariante Riemanntensor te berekenen, dan vinden we drie onafhankelijke componenten verschillend van nul:

$$
\begin{aligned}
R_{uxux} &= -\frac{1}{2}\,\Phi_{,xx}, \qquad R_{uyuy} = -\frac{1}{2}\,\Phi_{,yy}, \\
R_{uxuy} &= R_{uyux} = -\frac{1}{2}\,\Phi_{,xy}.
\end{aligned} \tag{5.5}
$$

Hiermee is het niet moeilijk na te gaan, dat alle componenten van de Ricci tensor $R_{\mu\nu} = g^{\kappa\lambda} R_{\mu\kappa\nu\lambda}$ vanzelf nul zijn, op een na:

$$
R_{uu} = -\frac{1}{2}\,\Delta_{trans}\Phi \equiv -\frac{1}{2}\,\left(\Phi_{,xx} + \Phi_{,yy}\right). \tag{5.6}
$$

Δ_{trans} staat hier voor de twee-dimensionale laplace operator in het transversale x-y-vlak. Deze is nog eenvoudiger te schrijven in complexe coördinaten $\zeta = x + iy$ en $\overline{\zeta} = x - iy$; de Einstein vergelijking (5.1) reduceert dan tot

$$
\Phi_{,\zeta\overline{\zeta}} = 0. \tag{5.7}
$$

De reële oplossingen van deze vergelijking zijn alle te schrijven als het reële deel van een analytische functie van ζ:

$$
\Phi(u, \zeta, \overline{\zeta}) = f(u, \zeta) + \bar{f}(u, \overline{\zeta}). \tag{5.8}
$$

Deze functie kunnen we als de Fourierintegraal van een machtreeks schrijven:

$$
\Phi(u, \zeta, \overline{\zeta}) = \sum_{n=0}^{\infty} \int_{-\infty}^{\infty} \frac{dk}{2\pi}\, \left(\epsilon_n(k) e^{-iku} \zeta^n + \bar{\epsilon}_n(k) e^{iku} \overline{\zeta}^n\right). \tag{5.9}
$$

Iedere Fourier amplitude $\epsilon_n(k)$ beschrijft nu een monochromatische zwaartekrachtsgolf met frequentie $\omega = kc$ in de z-richting. De index n karakteriseert de transversale veldconfiguratie, overeenkomend met een multipoolmoment van de orde 2^n. De eenvoudigste manier om dit te zien, is gebruik te maken van de karakterisering van multipoolmomenten door hun gedrag onder rotaties: wanneer een golfconfiguratie van een veld zich met de lichtsnelheid voortplant, dan is het multipoolkarakter van de veldconfiguratie bepaald door de mate van draaiing van de (lineaire) polarisatievlakken bij een gegeven rotatie van het coördinatenstelsel (of de bron, c.q. stralingsdetector). Bij een draaiing van de coördinaten over een hoek φ ondergaan de polarisatievlakken van een veldconfiguratie met multipoolmoment 2^n een draaiing van $\pm n\varphi$. In het onderhavige geval is een rotatie

$$
x' = x\cos\varphi + y\sin\varphi, \qquad y' = -x\sin\varphi + y\cos\varphi, \tag{5.10}
$$

equivalent met een faseverandering $\zeta' = e^{-i\varphi}\zeta$. Voor de amplitude van de n^e component van Φ leidt dit tot een verandering

$$
\varepsilon_n'\,\zeta'^{\,n} \;=\; \varepsilon_n'\,e^{-in\varphi}\,\zeta^n \;\equiv\; \varepsilon_n\,\zeta^n. \tag{5.11}
$$

Dit is equivalent met een draaiing van het reële en imaginaire deel van de amplitude ε_n over een hoek $n\varphi$:

$$\begin{aligned} \mathrm{Re}\,[\varepsilon'_n] &= \mathrm{Re}\,[\varepsilon_n] \cos n\varphi + \mathrm{Im}\,[\varepsilon_n] \sin n\varphi, \\ \mathrm{Im}\,[\varepsilon'_n] &= -\mathrm{Re}\,[\varepsilon_n] \sin n\varphi + \mathrm{Im}\,[\varepsilon_n] \cos n\varphi. \end{aligned} \tag{5.12}$$

Voor het complex geconjugeerde deel van de amplitude gelden natuurlijk dezelfde relaties, modulo een omkering van de draaiingsrichting $\varphi \rightarrow -\varphi$.

We merken nu op, dat alle krommingscomponenten (5.5) evenredig zijn met een tweede transversale afgeleide van Φ. Daarom kunnen de termen met $n = 0, 1$ in de reeksontwikkeling (5.9) worden weggelaten: zij beschrijven een ruimte zonder kromming, een vlakke ruimte dus, maar in een ongebruikelijk niet-inertiaal coördinatenstelsel. Alleen voor $n \geq 2$ beschrijven de termen een niet-triviale fysisch waarneembare zwaartekrachtgolf, met een quadrupoolkarakter voor de laagste waarde $n = 2$. Er is dus altijd een coördinatenstelsel te vinden waarin de termen met $n = 0, 1$ afwezig zijn, en waarin zwaartekrachtgolven fundamenteel van het quadrupooltype zijn. Dit is een belangrijk onderscheid t.o.v. electromagnetische golven, die een fundamenteel dipoolkarakter hebben.

5.3 Beweging in een gravitatiegolf

De oplossingen beschreven door vergelijking (5.9) hebben het karakter van gravitatiegolven die zich met de lichtsnelheid door de lege ruimte voortplanten. Dat zij niet triviaal zijn, blijkt uit de kromming, die voor $n \geq 2$ niet overal nul is. Om het fysische karakter van deze dynamische zwaartekrachtsvelden beter te begrijpen analyseren we de beweging van een testmassa onder invloed van een lopende gravitatiegolf. Daartoe lossen we de geodetische vergelijking (4.5) op:

$$\frac{d^2 x^\mu}{d\tau^2} + \Gamma_{\lambda\nu}^{\;\;\;\mu} \frac{dx^\lambda}{d\tau} \frac{dx^\nu}{d\tau} = 0.$$

We zijn daarbij vooral geinteresseerd in tijdachtige wereldlijnen, waarvan volgens vgl.

(4.47) de raakvector $u^\mu = dx^\mu / d\tau$ voldoet aan

$$u_\mu u^\mu = g_{\mu\nu} \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} = -c^2. \tag{5.13}$$

We beginnen met op te merken dat er geen versnelling langs de lichtkegelrichting $u = ct - z$ is:

$$\frac{d^2 u}{d\tau^2} = 0. \tag{5.14}$$

Dit volgt onmiddellijk uit het gegeven dat er geen connectiecomponenten van de vorm ${\Gamma_{\mu\nu}}^{u} \neq 0$ zijn. De u-component van de viersnelheid is dus constant:

$$\frac{du}{d\tau} \equiv \gamma = constant. \tag{5.15}$$

Vervolgens leiden we hieruit af, dat de eigentijdbeweging in het transversale vlak een tweedimensionaal potentiaalprobleem is, met een u-afhankelijke (dus tijdafhankelijke) potentiaal gegeven door $\gamma^2\Phi/2$:

$$\begin{aligned} \frac{d^2x}{d\tau^2} &= -\frac{\partial}{\partial x}\left(\frac{\gamma^2\Phi}{2}\right), \\ \frac{d^2y}{d\tau^2} &= -\frac{\partial}{\partial y}\left(\frac{\gamma^2\Phi}{2}\right). \end{aligned} \tag{5.16}$$

De enige relevante componenten van de connectie die niet nul zijn, zijn namelijk ${\Gamma_{uu}}^{(x,y)}$, terwijl $du/d\tau = \gamma$ een constante is, zoals we juist hebben laten zien. Deze kan dus desgewenst achter de (transversale) gradiëntoperator $(\frac{\partial}{\partial x}, \frac{\partial}{\partial y})$ geschreven worden, zoals in vgl.(5.16) is gedaan.

Tenslotte moeten we nog de beweging in de andere lichtkegelrichting v bepalen. De uitdrukking voor de eigenversnelling $d^2v/d\tau^2$ is nogal ingewikkeld, maar we kunnen deze buiten beschouwing laten door gebruik te maken van vergelijking (5.13), die bepaalt dat de absolute waarde van de viersnelheid constant is. Hieruit kunnen we rechtstreeks een uitdrukking voor de v-component van de viersnelheid afleiden:

$$\gamma\frac{dv}{d\tau} + \gamma^2\,\Phi(u,x,y) - \left(\frac{dx}{d\tau}\right)^2 - \left(\frac{dy}{d\tau}\right)^2 = c^2. \tag{5.17}$$

Door te gebruiken dat per definitie van γ en (u,v)

$$\gamma\frac{dv}{d\tau} = \frac{du}{d\tau}\frac{dv}{d\tau} = \left(c\frac{dt}{d\tau}\right)^2 - \left(\frac{dz}{d\tau}\right)^2, \tag{5.18}$$

krijgen we de vergelijking

$$\left(c\frac{dt}{d\tau}\right)^2 + \gamma^2\,\Phi = c^2 + \left(\frac{d\mathbf{r}}{d\tau}\right)^2, \tag{5.19}$$

waarin $\mathbf{r}$ de driedimensionale positievector (x,y,z) is. Tenslotte merken we op, dat $d\mathbf{r}/d\tau = \mathbf{v}dt/d\tau$, waarin $\mathbf{v} = d\mathbf{r}/dt$ de gewone driesnelheid in het coördinatenstelsel van de waarnemer is. Na enige herschikking van de vergelijking (5.19) volgt dan de tijddilatatieformule

$$\frac{dt}{d\tau} = \sqrt{\frac{1-\gamma^2\Phi/c^2}{1-\mathbf{v}^2/c^2}}. \tag{5.20}$$

We zien dat de tijddilatatie uit de combinatie van twee effecten ontstaat:
(i) het gewone kinematische effect uit de speciale relativiteitstheorie, gegeven door de factor $\sqrt{1-\mathbf{v}^2/c^2}$ in de noemer;
(ii) een zuiver gravitationele roodverschuiving t.g.v. de potentiaal Φ in de golf, gegeven door de factor $\sqrt{1-\gamma^2\Phi/c^2}$ in de teller; deze is er ook is voor een deeltje in rust ($\mathbf{v}=\mathbf{0}$).
Uit de definitie van de lichtkegelcoördinaat u en de driesnelheid $\mathbf{v}$ volgt nu, dat

$$\begin{aligned} \gamma &= \frac{du}{d\tau} = c\frac{dt}{d\tau} - \frac{dz}{d\tau} \\ &= c\frac{dt}{d\tau}\left(1-\frac{v_z}{c}\right). \end{aligned} \tag{5.21}$$

Door uit vgl.(5.20) en (5.21) $dt/d\tau$ te elimineren, krijgen we een uitdrukking voor de integratieconstante γ:

$$\frac{\gamma^2}{c^2} = \frac{1}{\Phi + \dfrac{1-\mathbf{v}^2/c^2}{(1-v_z/c)^2}} \tag{5.22}$$

We zien dus, dat als op $t=-\infty$ een testdeeltje in rust is ($\mathbf{v}=0$), en als de tijd-ruimte in het verre verleden een Minkowskikarakter had: $\Phi(t=-\infty)=0$, dan is $\gamma=c$. Ook volgt uit invertering van vgl.(5.22), dat

$$K \equiv \Phi + \frac{1-\mathbf{v}^2/c^2}{(1-v_z/c)^2} = \frac{c^2}{\gamma^2} \tag{5.23}$$

een constante van beweging is. Door voor Φ de uitdrukking (5.9) in te vullen krijgen we een directe relatie tussen de snelheid en positie van het testdeeltje, geheel analoog aan de wet van behoud van energie. Daarmee wordt het lokale potentiaalkarakter van Φ nogmaals benadrukt;

De uitdrukking (5.9) voor Φ geeft ons ook de concrete vorm van het potentiaalprobleem voor de transversale beweging. In het bijzonder geldt voor een zuivere quadrupoolgolf dat

$$\Phi(u,\zeta,\bar{\zeta}) = h(u)\,\zeta^2 + \bar{h}(u)\,\bar{\zeta}^2 \equiv h_+(u)\left(x^2-y^2\right) + 2h_\times(u)\,xy. \tag{5.24}$$

Als we bovendien gebruiken dat $\gamma d\tau = du$, dan krijgen de transversale bewegingsvergelijkingen de vorm

$$\begin{pmatrix} x''(u) \\ y''(u) \end{pmatrix} = -\begin{pmatrix} h_+(u) & h_\times(u) \\ h_\times(u) & -h_+(u) \end{pmatrix}\begin{pmatrix} x(u) \\ y(u) \end{pmatrix}, \tag{5.25}$$

waarbij het accent een afgeleide naar u beduidt. Dit is een stelsel van gewone homogene lineaire tweede-orde differentiaalvergelijkingen.

Als voorbeeld kunnen we het simpele geval beschouwen van een constante diagonale amplitude $h_+ = k^2$, $h_\times = 0$. Dan is de oplossing van de transversale bewegingsvergelijkingen gegeven door

$$x(u) = a_x \cos k(u - u_0), \qquad y(u) = a_y \cosh k(u - u_0). \tag{5.26}$$

Als de golf een eindige lengte L heeft: $h_+(u) = k^2$ in het interval $0 < u < L$, en $h_+(u) = 0$ buiten dit interval, dan moeten we deze oplossingen voor $u \in (0, L)$ glad laten aansluiten op de oplossingen voor vrije deeltjes daarbuiten:

$$\begin{aligned} &x(u) = b_x + p_x u, && y(u) = b_y + p_y u, && u \leq 0; \\ &x(u) = c_x + q_x(u - L), && y(u) = c_y + q_y(u - L), && u \geq L. \end{aligned} \tag{5.27}$$

De positie en richting van deze lijnen moet overeenkomen met die van de oplossing (5.26) op de aansluitingspunten $u = 0, L$. Dit houdt in, dat

$$\begin{aligned} &x(0) = b_x = a_x \cos ku_0, && y(0) = b_y = a_y \cosh ku_0, && \text{op } u = 0; \\ &x(L) = c_x = a_x \cos k(L - u_0), && y(L) = c_y = a_y \cosh k(L - u_0), && \text{op } u = L. \end{aligned} \tag{5.28}$$

En evenzo voor de viersnelheden:

$$\begin{aligned} &x'(0) = p_x = ka_x \sin ku_0, && y'(0) = p_y = -ka_y \sinh ku_0, && \text{op } u = 0; \\ &x'(L) = q_x = -ka_x \sin k(L - u_0), && y'(L) = q_y = ka_y \sinh k(L - u_0), && \text{op } u = L. \end{aligned} \tag{5.29}$$

Wanneer de parameters (b_i, c_i) en (p_i, q_i), $i = (x, y)$, voor en na de beweging gemeten worden, kunnen de waarden van de golfparameters k, L en u_0 hieruit worden afgeleid. Zo is

$$\begin{aligned} \frac{p_x}{b_x} &= k \tan ku_0, \\ \frac{c_x}{b_x} &= \frac{\cos k(L - u_0)}{\cos ku_0} = \cos kL + \tan ku_0 \sin kL, \\ \frac{q_x}{b_x} &= -k\,\frac{\sin k(L - u_0)}{\cos ku_0} = -k \sin kL + k \tan ku_0 \cos kL, \end{aligned} \tag{5.30}$$

met soortgelijke uitdrukkingen in termen van hyperbolische functies voor de y-componenten.

Tenslotte merken we op dat de substitutie $u = ct - z$ in de bovenstaande oplossingen (5.26), (5.27) de positie als functie van de waarnemertijd t geeft, met dien verstande dat de positie in de z-richting nog bepaald moet worden uit $dz/dt = v_z$. Een uitdrukking voor v_z is impliciet gegeven in vgl.(5.23). Voor $v_z \ll c$ is echter $z \approx$ *constant*, en dus is $u - u_0 \approx c(t - t_0)$.

5.4 De Schwarzschild-Droste oplossing

Gravitatiegolven vertegenwoordigen het dynamische aspect van de relativistische gravitatietheorie van Einstein: zij dragen de eigen vrijheidsgraden van het zwaartekrachtveld en vertonen dynamisch gedrag los van de aanwezigheid van enige vorm van materie.

Daarnaast of daartegenover zijn er zwaartekrachtvelden die direct met het bestaan van materie samenhangen: daarin vinden zij hun oorsprong en hun eigenschappen worden rechtstreeks door de structuur en constitutie van de materie bepaald. Het meest simpele voorbeeld van een dergelijk materiegebonden zwaartekrachtveld is het externe veld van een bolsymmetrische massaverdeling, zoals een homogene, niet-roterende ster in perfect mechanische evenwicht. Deze oplossing van de Einsteinvergelijkingen werd in 1916 onafhankelijk van elkaar door Schwarzschild en Droste (een student van Lorentz) gevonden. Hoewel Schwarzschild's publicatie het eerste uitkwam, is de gebruikelijke formulering van deze oplossing die van Droste. Het coördinatensysteem van Droste bestaat uit een tijdcoördinaat t, een radiële coördinaat r en twee hoekcoördinaten (θ, φ). Omdat de oplossing statisch moet zijn, hangen de componenten van de metriek niet af van de tijdcoördinaat t; de bolsymmetrie wordt uitgedrukt door te zorgen dat bij vaste straal r de vorm van een lijnelement in het vlak, verkregen door variatie van de hoekcoördinaten (θ, φ), dezelfde is als die van lijnelement op een boloppervlak met straal r in de vlakke Euclidische ruimte. De volledige vorm van de metriek wordt in deze formulering gegeven door het vierdimensionale lijnelement

$$g_{\mu\nu} dx^\mu dx^\nu = -\left(1 - \frac{\alpha}{r} \right) c^2 dt^2 + \frac{1}{\left(1 - \frac{\alpha}{r} \right)} dr^2 + r^2 \left(d\theta^2 + \sin^2 \theta \, d\varphi^2 \right). \tag{5.31}$$

We zullen later zien dat de parameter α, die de dimensie van lengte heeft, wordt gegeven door de Schwarzschildstraal van de materieverdeling, gedefiniëerd als

$$\alpha = \frac{2GM}{c^2}, \tag{5.32}$$

waarin M de totale massa is. Dat α uitsluitend van M afhangt is niet verwonderlijk: in afwezigheid van een impulsmoment, van multipoolmomenten van de massaverdeling, en van electromagnetische of andere ladingen, is er geen andere parameter die het veld van de massaverdeling kan bepalen dan de totale massa.

We zullen eerst nagaan dat de uitdrukking (5.31) een oplossing is van de Einsteinvergelijking buiten de sferische massaverdeling, d.w.z. in de lege ruimte om de ster heen. Uit de hierboven gegeven expliciete vorm van de metriek leiden we door toepassing van de definitie (4.3) de componenten van de connectie af; de componenten die niet identiek nul zijn hebben de volgende vorm:

$$\begin{aligned}
&\Gamma_{rt}{}^{t} = \Gamma_{tr}{}^{t} = \frac{\alpha}{2r(r-\alpha)}, \\
&\Gamma_{tt}{}^{r} = \frac{\alpha c^2}{2r^3}\,(r-\alpha)\,, \qquad \Gamma_{rr}{}^{r} = -\,\frac{\alpha}{2r(r-\alpha)},
\end{aligned} \tag{5.33}$$

en

$$\begin{aligned}
&\Gamma_{\theta\theta}{}^{r} = \alpha - r, \qquad \Gamma_{\varphi\varphi}{}^{r} = \sin^2\theta\,(\alpha - r)\,, \\
&\Gamma_{r\theta}{}^{\theta} = \Gamma_{\theta r}{}^{\theta} = \frac{1}{r}, \quad \Gamma_{\varphi\varphi}{}^{\theta} = -\sin\theta\,\cos\theta, \\
&\Gamma_{r\varphi}{}^{\varphi} = \Gamma_{\varphi r}{}^{\varphi} = \frac{1}{r}, \quad \Gamma_{\theta\varphi}{}^{\varphi} = \Gamma_{\varphi\theta}{}^{\varphi} = \frac{\cos\theta}{\sin\theta}.
\end{aligned} \tag{5.34}$$

Met behulp van deze uitdrukkingen vinden we een groot aantal niet-triviale componenten van de Riemann-krommingstensor; ten eerste:

$$\begin{aligned}
&R_{trt}{}^{r} = \frac{\alpha c^2}{r^4}\,(r-\alpha), \\
&R_{t\theta t}{}^{\theta} = R_{t\varphi t}{}^{\varphi} = -\,\frac{\alpha c^2}{2r^4}\,(r-\alpha).
\end{aligned} \tag{5.35}$$

Door deze componenten op te tellen vinden we op eenvoudige wijze dat

$$R_{tt} = \sum_{i=(r,\theta,\varphi)} R_{tit}{}^{i} = 0. \tag{5.36}$$

Voorts zijn er krommingscomponenten

$$\begin{aligned}
&R_{rtr}{}^{t} = -\,\frac{\alpha}{r^2(r-\alpha)}, \qquad R_{r\theta r}{}^{\theta} = R_{r\varphi r}{}^{\varphi} = \frac{\alpha}{2r^2(r-\alpha)}, \\
&R_{\theta t\theta}{}^{t} = R_{\theta r\theta}{}^{r} = \frac{\alpha}{2r}, \qquad R_{\theta\varphi\theta}{}^{\varphi} = -\,\frac{\alpha}{r}, \\
&R_{\varphi t\varphi}{}^{t} = R_{\varphi r\varphi}{}^{r} = \frac{\alpha}{2r}\,\sin^2\theta, \quad R_{\varphi\theta\varphi}{}^{\theta} = -\,\frac{\alpha}{r}\,\sin^2\theta.
\end{aligned} \tag{5.37}$$

Hieruit volgt opnieuw door direkte optelling dat

$$R_{\mu\nu} = 0, \tag{5.38}$$

zodat de Schwarzschild-Droste metriek een oplossing geeft van de Einsteinvergelijkingen in de lege ruimte; deze oplossing beschrijft het *externe* gravitatieveld van een statische bolsymmetrische massaverdeling.

Wat betreft de fysische eigenschappen van dit veld merken we op, dat asymptotisch voor $r \to \infty$ de metriek de Minkowskivorm krijgt:

$$ds^2\Big|_{r\to\infty} = -c^2dt^2 + dr^2 + r^2d\theta^2 + r^2\sin^2\theta\, d\varphi^2, \tag{5.39}$$

terwijl tegelijk alle componenten van de krommingstensor naar nul gaan:

$$\lim_{r\to\infty} R_{\mu\nu\kappa}{}^{\lambda} = 0. \tag{5.40}$$

Ver van de massaverdeling verdwijnt het statische gravitatieveld dus, en wordt de tijd-ruimte vlak.

5.5 Beweging in een statisch bolsymmetrisch veld

Met behulp van de in de vorige paragraaf afgeleide connectiecoëfficiënten kunnen we direct de geodetische vergelijkingen (4.5) in de Schwarzschild-Droste metriek opschrijven. De oplossingen van deze vergelijkingen hebben veel gemeen met de Keplerbanen die we in het eerste hoofdstuk beschreven hebben. Dat is ook te verwachten, aangezien de Keplerbanen de oplossingen zijn voor de beweging van een puntdeeltje in een bolsymmetrisch zwaartekrachtveld in het niet-relativistische geval. In de limiet van grote afstand ($r \gg \alpha$) en kleine snelheid ($v \ll c$) gaan de oplossingen van de geodetische vergelijkingen dan ook over in de Keplerbanen.

Om te beginnen impliceert de bolsymmetrie ook hier, dat er een behouden impulsmoment is, waardoor de baan altijd in een plat vlak ligt. Dit vlak kiezen we als het equatoriale vlak $\theta = \pi/2$. Om te laten zien dat dit een goede oplossing is, beschouwen we ten eerste de bewegingsvergelijking voor θ:

$$\frac{d^2\theta}{d\tau^2} + \frac{2}{r}\frac{dr}{d\tau}\frac{d\theta}{d\tau} - \sin\theta\,\cos\theta\left(\frac{d\varphi}{d\tau}\right)^2 = 0. \tag{5.41}$$

We zien onmiddellijk dat de constante waarde $\theta = \pi/2$ inderdaad een oplosing is. Vervolgens beschouwen we de vergelijking voor de hoekcoördinaat φ in het equatoriale vlak $\theta = \pi/2$:

$$\frac{d^2\varphi}{d\tau^2} + \frac{2}{r}\frac{dr}{d\tau}\frac{d\varphi}{d\tau} = \frac{1}{r^2}\frac{d}{d\tau}\left(r^2\frac{d\varphi}{d\tau}\right) = 0. \tag{5.42}$$

De oplossing van de vergelijking is de relativistische versie van de tweede Keplerwet:

$$\frac{d\varphi}{d\tau} = \frac{\ell}{r^2}, \tag{5.43}$$

waarin de constante ℓ het impulsmoment per eenheid van massa m van het testdeeltje is. Uit de vergelijking voor t volgt op soortgelijke wijze dat de energie —gerepresenteerd door de tijdcomponent van de impulsviervector van het testdeeltje— eveneens een constante van beweging is:

$$\varepsilon \equiv \frac{p_0}{mc} = g_{00}\frac{dx^0}{cd\tau} = \left(1 - \frac{\alpha}{r}\right)\frac{dt}{d\tau}. \tag{5.44}$$

ε is hier een dimensieloze constante, de deeltjesenergie als fractie van de rustenergie mc^2. Ook in het geval van een stationair gravitatieveld is er dus een zuiver gravitationele tijddilatatie, gegeven door:

$$\frac{dt}{d\tau} = \frac{\varepsilon}{1 - \dfrac{\alpha}{r}}. \tag{5.45}$$

In het bijzonder volgt uit de definitie van eigentijd, dat als het deeltje op oneindige afstand in rust is, zodat $dt = d\tau$, dan is $\varepsilon = 1$:

$$\lim_{r\to\infty} \mathbf{v} = 0 \quad \Rightarrow \quad \varepsilon = \lim_{r\to\infty}\frac{dt}{d\tau} = 1. \tag{5.46}$$

De enige component van de viersnelheid die nu nog ontbreekt is $dr/d\tau$. Deze kunnen we bepalen uit de behoudswet voor de totale viersnelheid (4.47):

$$\left(\frac{dr}{d\tau}\right)^2 = c^2\left[\varepsilon^2 - \left(1 + \frac{\ell^2}{c^2r^2}\right)\left(1 - \frac{\alpha}{r}\right)\right]. \tag{5.47}$$

Hieruit kunnen we ook een uitdrukking voor de radiële versnelling afleiden, die equivalent is met de geodetische vergelijking voor deze coördinaat:

$$\frac{d^2r}{d\tau^2} = -\frac{\alpha c^2}{2r^2} + \frac{\ell^2}{r^3}\left(1 - \frac{3\alpha}{2r}\right). \tag{5.48}$$

We beschouwen nu eerst twee speciale gevallen: zuiver radiële en zuiver circulaire beweging. Daarna komt het algemene geval ter sprake.
A. *Radiële beweging.* Dit is het geval dat $d\varphi/d\tau = 0$, en dus $\ell = 0$. We zien dan, dat vgl.(5.47) en (5.48) vereenvoudigd worden tot:

$$\frac{dr}{d\tau} = -\,c\sqrt{\varepsilon^2 - 1 + \frac{\alpha}{r}}, \qquad \frac{d^2r}{d\tau^2} = -\frac{\alpha c^2}{2r^2}. \tag{5.49}$$

Het minteken in de eerste vergelijking duidt op beweging naar binnen toe (dus r neemt af). Beginnen we met een deeltje in rust op $r = \infty$ en $t = -\infty$, dan is $\varepsilon = 1$. In dat geval krijgen we

$$\frac{dr}{d\tau} = -c\sqrt{\frac{\alpha}{r}}, \tag{5.50}$$

met als oplossing

$$r^{3/2} \;=\; r_0^{3/2} \;-\; \frac{3}{2}\sqrt{\alpha}\, c\,(\tau - \tau_0). \tag{5.51}$$

In deze oplossing is r_0 de radiële positie van het deeltje op de referentie-eigentijd τ_0. We kunnen b.v. de eigenklok zo instellen dat deze $\tau_0 = 0$ aangeeft op het ogenblik dat het deeltje de Schwarzschildstraal $r_0 = \alpha$ bereikt. Dan is

$$r \;=\; \alpha\left(1 - \frac{3c\tau}{2\alpha}\right)^{2/3}. \tag{5.52}$$

We mogen hierbij niet vergeten dat alle eigentijdwaarden *voor* het bereiken van de Schwarzschildstraal ook *voor* $\tau = 0$ liggen, en dus negatief zijn.

Uit de tweede vergelijking (5.49) volgt dat voor een deeltje op grote afstand ($r \gg \alpha$), dat een niet-relativistische snelheid heeft ($v \ll c$, zodat $d\tau = dt$), de versnelling wordt gegeven door

$$\frac{d^2 r}{dt^2} \;=\; -\frac{\alpha c^2}{2r^2}. \tag{5.53}$$

Dit moet nu overeenkomen met de niet-relativistische uitdrukking van Newton:

$$\left.\frac{d^2 r}{dt^2}\right|_{Newton} \;=\; -\frac{GM}{r^2}, \tag{5.54}$$

waarin M de centrale massa is. Uit deze eis van overeenstemming volgt direkt de eerdere uitdrukking (5.32) voor de Schwarzschildstraal:

$$\alpha = \frac{2GM}{c^2}$$

B. *Cirkelbeweging.* Voor $\ell \neq 0$ hebben de radiële bewegingsvergelijkingen (5.47) en (5.48) oplossingen $r = R =$ constant. De eerste vergelijking geeft ons dan een uitdrukking voor de energieparameter ε als functie van ℓ en R:

$$\varepsilon^2 \;=\; \left(1 + \frac{\ell^2}{c^2 R^2}\right)\left(1 - \frac{\alpha}{R}\right), \tag{5.55}$$

De tweede vergelijking geeft nu het verband tussen ℓ en R:

$$R^2 \;-\; \frac{2\ell^2}{\alpha c^2}\, R \;+\; \frac{3\ell^2}{c^2} \;=\; 0 \;\Rightarrow\; \ell^2 \;=\; GMR\left(\frac{c^2 R}{c^2 R - 3GM}\right). \tag{5.56}$$

Voor cirkelbeweging liggen dus alle dynamische grootheden vast in termen van de radiële coördinaat R. Aangezien de hoeksnelheid op een cirkelbaan constant is, is de omlooptijd T eenvoudig te verkrijgen uit

$$\frac{d\varphi}{dt} = \frac{2\pi}{T} = \frac{\ell}{\varepsilon R^2}\left(1 - \frac{\alpha}{R}\right). \tag{5.57}$$

Combinatie met de uitdrukkingen (5.55) voor ε en (5.56) voor ℓ geeft dan

$$T = \frac{2\pi}{\sqrt{GM}} R^{3/2}. \tag{5.58}$$

Voor dit geval krijgen we het niet-relativistische Kepler resultaat (1.28) dus weer terug, mits we R kunnen identificeren met de afstand tot de oorsprong van het veld. Dit is een extra bevestiging dat de gepostuleerde uitdrukking (5.32) voor α een continue overgang naar de Newtonse zwaartekrachtstheorie mogelijk maakt in de limiet van lage snelheden en grote afstanden, wanneer de relativistische effecten verwaarloosd mogen worden.

C. *Algemene baanbeweging.* Tenslotte beschouwen we het algemene geval van gebonden beweging in een centraal zwaartekrachtveld met specifiek impulsmoment (per eenheid van massa) ℓ. Eerst bepalen we uit de vergelijkingen voor $r(\tau)$en $\varphi(\tau)$ een vergelijking voor de baan $r(\varphi)$ door de τ-afhankelijkheid te elimineren:

$$\frac{dr}{d\varphi} = \frac{dr/d\tau}{d\varphi/d\tau}, \tag{5.59}$$

zodat vergl.(5.47) met behulp van (5.43) herschreven kan worden als

$$\frac{\ell^2}{c^2 r^4} \left(\frac{dr}{d\varphi} \right)^2 = \frac{\ell^2}{c^2} \left(\frac{d}{d\varphi} \frac{1}{r} \right)^2 = \varepsilon^2 - \left(1 - \frac{\alpha}{r} \right) \left(1 + \frac{\ell^2}{c^2 r^2} \right). \tag{5.60}$$

Deze vergelijking heeft oplossingen van de vorm

$$r(\varphi) = \frac{r_0}{1 + e \cos y(\varphi)}, \tag{5.61}$$

voor een nader te bepalen monotoon stijgende functie $y(\varphi)$. De parameters r_0 en $e < 1$ zijn de oplossingen van algemene versies van vergelijkingen (5.55) en (5.56):

$$\varepsilon^2 = \left(1 - \frac{\alpha}{r_0} \right) \left(1 + \frac{\ell^2}{c^2 r_0^2} \right) + e^2 \frac{\ell^2}{c^2 r_0^2} \left(1 - \frac{3\alpha}{r_0} \right), \tag{5.62}$$

en

$$r_0^2 - \frac{2\ell^2}{c^2 \alpha} r_0 + \frac{\ell^2}{c^2} (3 + e^2) = 0. \tag{5.63}$$

Deze zijn te krijgen door vergelijking (5.60) m.b.v. (5.61) te evalueren op de stationaire punten waar r zijn kleinste en grootste waarde bereikt, dus waar $dr/d\varphi = 0$. Het is direkt te zien, dat deze punten corresponderen met $\cos y(\varphi) = 1$ (periastron), en $\cos \varphi = -1$ (apastron), ofwel

$$\begin{array}{llcl} \text{periastron:} & \cos y(\varphi) = 1 & \Rightarrow & y(\varphi) = 2n\pi, \\ \text{apastron:} & \cos y(\varphi) = -1 & \Rightarrow & y(\varphi) = (2n+1)\pi. \end{array} \tag{5.64}$$

Merk op, dat met $\ell^2 \neq 0$ een oplossing van vergelijking (5.63) alleen bestaat voor $r_0 > \alpha\,(3+e^2)/2$. Met behulp van de relaties (5.62) en (5.63) vinden we, door algemeen de baanvergelijking (5.61) in (5.60) te substitueren, dat $y(\varphi)$ moet voldoen aan de vergelijking

$$\left(\frac{dy}{d\varphi}\right)^2 = 1 - \frac{\alpha}{r_0}\,(3 + e\cos y)\,. \tag{5.65}$$

Voor $r_0 > \alpha\,(3+e)$ is dus inderdaad $|dy/d\varphi| \neq 0$, en als we de afgeleide in $\varphi = 0$ positief nemen, dan is $y(\varphi)$ een monotone stijgende functie van de hoekcoördinaat φ voor alle waarden van φ.

De beginwaarde van φ kunnen we zo kiezen, dat de waarde $\varphi = 0$ in het periastron $y = 0$ valt. Het volgende periastron valt dan op $y = 2\pi$. De bijbehorende waarde van de hoek φ vinden we door integreren van vergelijking (5.65):

$$\Delta\varphi = \int_0^{2\pi} \frac{dy}{\sqrt{1 - \frac{\alpha}{r_0}\,(3 + e\cos y)}} \approx 2\pi + \frac{3\alpha\pi}{r_0} + \mathcal{O}[(\alpha/r_0)^2]. \tag{5.66}$$

We zien dus dat het periastron in vergelijking met een gesloten elliptische Keplerbaan verschoven is over een hoek

$$\frac{3\alpha\pi}{r_0} = \frac{6\pi GM}{r_0 c^2}. \tag{5.67}$$

We kunnen nu de verschuiving van het perihelium evalueren voor de planeet Mercurius. Deze planeet beweegt op een gemiddelde afstand van $r_0 = 5.5 \times 10^7$ km van de zon. Voor de zon is

$$\alpha = \frac{2GM}{c^2} = 2.95\,\text{km}. \tag{5.68}$$

Dan is de verschuiving van het perihelium per omloopperiode:

$$\Delta\varphi = 0.48 \times 10^{-6}\,\text{rad} \approx 0.104'' \tag{5.69}$$

Nu maakt Mercurius per eeuw ongeveer 415 omwentelingen om de zon, zodat we voor de geintegreerde periheliumverschuiving per eeuw een bedrag vinden van

$$\Delta\varphi|_{100\,jr} \approx 43''. \tag{5.70}$$

Dit komt zeer goed overeen met de uit observaties bekende waarde van eveneens $43''$ per eeuw. Deze verschuiving valt dus geheel te verklaren uit relativistische correcties op de Newtonse zwaartekrachtwet.

5.6 De dubbelpulsar PSR 1913+16

De verschuiving van het perihelium van Mercurius is een van de klassieke tests van de ART. De precessie van de baan van Mercurius is echter uiterst gering, omdat de omloopsnelheid van deze planeet niet-relativistisch is en het zwaartekrachtveld van de zon ter plaatse vrijwel Newtoniaans. Het verschijnsel is toch nog meetbaar omdat Mercurius van alle planeten het dichtst bij de zon staat, en de beweging door sterrenkundigen gedurende vele jaren is gevolgd, waarbij er een cumulatief effect kon optreden.

Uiteraard is de precessie veel groter in het geval van objecten die relativistisch en/of in sterkere zwaartekrachtvelden bewegen. Zulke objecten kunnen worden gevormd door paren van zeer compacte sterren, zoals witte dwergen, neutronensterren of zwarte gaten, die op niet al te grote afstand om een gemeenschappelijk massamiddelpunt bewegen. Zoals we in het eerste hoofdstuk hebben gezien, beschrijft de relatieve beweging van zo'n systeem volgens de theorie van Newton een ellips met het massamiddelpunt van de sterren in een van de brandpunten. In de relativistische theorie van de beweging van twee massa's onder invloed van elkaars zwaartekrachtveld is deze ellips slechts een eerste benadering, waarop vervolgens een aantal correcties nodig zijn die afhangen van de massa, onderlinge afstand en relatieve snelheid van de objecten. Een van die correcties is de precessie van het periastron, het punt van dichtste nadering. Deze correctie is geheel analoog aan de precessie van het perihelium van Mercurius die we in de vorige paragraaf hebben berekend.

Het bekendste voorbeeld van zo'n systeem waarin de relativistische (zgn. post-newtoniaanse) correcties zeer groot zijn, wordt gevormd door een paar neutronensterren bekend als de dubbelpulsar PSR 1913+16. Deze dubbelster werd in 1974 door Hulse en Taylor ontdekt[2], en bestaat uit twee bijna even massieve neutronensterren, met massa's van 1.387 $M_\odot$ en 1.441 $M_\odot$. De neutronensterren bewegen om elkaar heen met een periode (anno 2000) van iets meer dan 7 uur. De baan is sterk eccentrisch: $e = 0.617$, met een gemiddelde afstand tussen de sterren van ruwweg een miljoen kilometer. De gemiddelde baansnelheid v is daarbij ongeveer $v/c \approx 1.3 \times 10^{-3}$. Dit systeem heeft men zeer nauwkeurig kunnen bestuderen, omdat in ieder geval een van de neutronensterren in dit systeem een pulsar[3] is, waarvan het radiosignaal op aarde waargenomen kan worden.

Vanwege de relatief grote baansnelheid zowel als de sterkte van het zwaartekrachtveld bij een onderlinge afstand van slechts een miljoen kilometer, zijn de post-newtoniaanse correcties groot. Zo is de precessie van het periastron van de

---

[2] R.A. Hulse en J.H. Taylor, Ap. J. Lett. 195 (1975), L51

[3] Een pulsar is een snel roterende neutronenster met een zeer sterk magnetische veld. De radiostraling veroorzaakt door de beweging van geladen deeltjes in het roterende veld is vanaf de aarde waar te nemen als een periodiek terugkerend signaal met een zeer stabiele frequentie; deze kan worden gebruikt om de beweging van de pulsar te bepalen door metingen van de Dopplerverschuiving van het signaal als functie van de tijd.

twee neutronensterren ongeveer 4.2^o *per jaar*, tegen de bovengenoemde $43''$ *per eeuw* voor de precessie van het perihelium van Mercurius. Wat dit stelsel nog interessanter maakt zijn de overige post-newtontiaanse correcties die nodig zijn om waarnemingen en theorie tot onderlinge overeenstemming te brengen. Hiertoe behoort ook het verlies aan energie t.g.v. het uitzenden van zwaartekrachtstraling. Dit energieverlies leidt tot het afnemen van zowel de gemiddelde diameter van de baan als de omlooptijd, hetgeen ook merkbaar is in de precessie van het periastron[4]. Aan de hand van dit zeer interessante object kunnen we dus niet alleen de relativistische correcties op de wet van Newton testen, maar ook het bestaan van zwaartekrachtgolven aantonen.

5.7 De afbuiging van licht

Naast de gebonden en radiële banen is er natuurlijk een schaar van open banen, met verschillende beginvoorwaarden. Wegens de praktische toepassingen bespreken we hier de lichtachtige banen, waarvoor $g_{\mu\nu}\, dx^\mu dx^\nu = -c^2 d\tau^2 = 0$. Deze beschrijven de voortplanting van licht in de geometrisch-optische benadering.

Omdat langs een lichtachtige baan $d\tau = 0$, kan de eigentijd τ niet gebruikt worden om de baan te parametriseren. Om dit probleem te omzeilen voeren we een nieuwe parameter λ in die de rol van $c\tau$ overneemt, en zo is gekozen dat de baan nog steeds door de geodetische vergelijking (4.5) wordt beschreven:

$$\frac{d^2 x^\mu}{d\lambda^2} + \Gamma_{\lambda\nu}{}^{\mu} \frac{dx^\lambda}{d\lambda} \frac{dx^\nu}{d\lambda} = 0, \tag{5.71}$$

maar onder de bijkomende voorwaarde dat

$$g_{\mu\nu} \frac{dx^\mu}{d\lambda} \frac{dx^\nu}{d\lambda} = 0. \tag{5.72}$$

Zo'n parameter λ met de eigenschap dat de geodetische vergelijking de standaardvorm (5.71) heeft, heet een *affine* parameter. We zullen zien dat deze door de conditie (5.72) weer geëlimineerd kan worden.

Nu de geodetische vergelijkingen na de vervanging $c d\tau \to d\lambda$ hetzelfde zijn als voorheen, zijn de oplossingen natuurlijk ook van dezelfde vorm:

$$\begin{aligned} c\,\frac{dt}{d\lambda} &= \frac{\eta}{1 - \dfrac{\alpha}{r}}, \\ \frac{d\varphi}{d\lambda} &= \frac{\kappa}{r^2}, \qquad \theta = \frac{\pi}{2}, \end{aligned} \tag{5.73}$$

[4] J.H. Taylor en J.M. Weisberg, Ap. J. 253 (1982), 908; *id.* 345 (1989), 434.

terwijl de vergelijking voor de radiële coördinaat r wegens de voorwaarde (5.72) nu luidt:

$$\left(\frac{dr}{d\lambda} \right)^2 = \eta^2 - \frac{\kappa^2}{r^2} \left(1 - \frac{\alpha}{r} \right). \tag{5.74}$$

De schaal van de integratieconstanten (η, κ) wordt vastgelegd door de schaal van de affine parameter λ. De verhouding η/κ heeft echter een direct fysische interpretatie; de minimale waarde r_m van de radiële coördinaat wordt namelijk bereikt wanneer $dr/d\lambda = 0$, zodat

$$\frac{\eta}{\kappa} = \frac{1}{r_m} \sqrt{1 - \frac{\alpha}{r_m}}. \tag{5.75}$$

We kunnen nu de affine parameter λ elimineren door de baanvergelijking voor $r(\varphi)$ te beschouwen, die volgt uit toepassing van de kettingregel:

$$\frac{dr}{d\lambda} = \frac{dr}{d\varphi}\frac{d\varphi}{d\lambda} \quad \rightarrow \quad \frac{\kappa^2}{r^4} \left(\frac{dr}{d\varphi} \right)^2 = \kappa^2 \left(\frac{d}{d\varphi} \frac{1}{r} \right)^2 = \eta^2 - \frac{\kappa^2}{r^2} \left(1 - \frac{\alpha}{r} \right). \tag{5.76}$$

Hieruit volgt onmiddelijk, dat

$$\left. \frac{d}{d\varphi} \frac{1}{r} \right|_{r=\infty} = \frac{\eta}{\kappa}, \tag{5.77}$$

hetgeen een maat is voor de verhouding tussen radiële en hoeksnelheid in de asymptotisch vlakke ruimte. Vergelijking (5.75) vertelt ons dan hoe deze grootheid is gerelateerd aan de minimale afstand waarop het licht de ster passeert. Om de baanvergelijking op te lossen definiëren we de dimensieloze parameters

$$R = \frac{\eta r}{\kappa}, \qquad A = \frac{\eta \alpha}{\kappa}, \tag{5.78}$$

zodat $A/R = \alpha/r$. De vergelijking voor $R(\varphi)$ wordt dan

$$\left(\frac{d}{d\varphi} \frac{1}{R} \right)^2 = 1 - \frac{1}{R^2} \left(1 - \frac{A}{R} \right). \tag{5.79}$$

Voor normale sterren ligt de Schwarzschildstraal diep onder het oppervlak, zodat we alleen het geval hoeven te beschouwen waarin $\alpha/r < \alpha/r_m \ll 1$. Zelfs voor compacte objecten als neutronensterren is hieraan voldaan. We kunnen daarom orde bij orde in de dimensieloze verhouding $\alpha/r = A/R$ rekenen. Hier doen we dit tot eerste orde. We definiëren daartoe de functie $y(\varphi)$ door

$$\cos^2 y = \frac{1}{R^2} \left(1 - \frac{A}{R} \right). \tag{5.80}$$

Tot eerste orde in A/R kunnen we dit omschrijven tot

$$\frac{1}{R} = \cos y \left(1 + \frac{A}{2}\cos y + ...\right). \tag{5.81}$$

Invullen in vergelijking (5.79) levert dan

$$\frac{d}{d\varphi}\frac{1}{R} = -\sin y\,(1 + A\cos y + ...)\,\frac{dy}{d\varphi} = -\sin y. \tag{5.82}$$

De baan begint bij $1/R = 0$, bereikt de minimale afstand bij $1/R_m$, en gaat dan verder naar opnieuw $1/R = 0$; dit wordt geparametriseerd door y te laten lopen van $y = -\pi/2$ via $y = 0$ tot $y = \pi/2$. De verandering die de hoek φ daarbij ondergaat, wordt bepaald uit

$$\Delta\varphi = \int_{-\pi/2}^{\pi/2} dy\ (1 + A\cos y + ...) = \pi + 2A + ... \tag{5.83}$$

Zonder afbuiging zou de hoekcoördinaat φ veranderen met een bedrag π; de afbuiging zelf wordt dus gegeven door

$$\pi - \Delta\varphi \approx 2A \approx \frac{2\alpha}{r_m} = \frac{4GM}{c^2\, r_m}. \tag{5.84}$$

We kunnen dit resultaat toepassen op afbuiging van licht door de zon. De straal van de zon in 696 000 km, en de Schwarzschildstraal 2.95 km. Voor licht dat langs het oppervlak van de zon scheert is de afbuiging dan ongeveer $1.8''$. Deze afbuiging wordt daadwerkelijk gemeten, wanneer licht van sterren of sterrenstelsels achter de zon wordt waargenomen tijdens een zonsverduistering.

Een belangrijke toepassing van dit verschijnsel is het gravitatielens-effect: het licht van zeer verre sterrenstelsels kan door concentraties van tussenliggende materie op zo'n manier worden afgebogen dat er verscheidene beelden van het achterliggende sterrenstelsel zichtbaar zijn. Dit effect wordt b.v. gebruikt om naar donkere materie te zoeken; zulke materie geeft zelf geen detecteerbare straling af, maar kan wel als gravitatielens werken.

5.8 Stationaire testmassa's

Het zwaartekrachtveld van een bolsymmetrische massaverdeling hebben we in de voorgaande paragrafen in kaart gebracht door de beweging van testmassa's in vrije val te bestuderen. In vrije val worden testmassa's altijd versneld; hun uiteindelijk lot is of te verdwijnen in het oneindige, of voor eeuwig aan de centrale massa gebonden te blijven, dan wel in te slaan op het oppervlak van de centrale massa.

Een andere invalshoek op de eigenschappen van het veld krijgen we door te vragen welke radiële versnelling je een testmassa —b.v. een door een raketmotor voortgestuwde kunstmaan— moet geven om hem op een stationaire positie op vaste afstand boven het oppervlak te houden.

In dit geval gaat het erom de beweging van een testmassa onder de gezamenlijke invloed van het zwaartekrachtveld en de uitwendige kracht te beschrijven. De eigenversnelling t.g.v. de uitwendige kracht duiden we aan met de viervector $a^\mu(\tau)$. Dan is de bewegingsvergelijking voor een testmassa met viersnelheid $u^\mu = dx^\mu/d\tau$:

$$\frac{Du^\mu}{D\tau} = \frac{du^\mu}{d\tau} + \Gamma_{\lambda\nu}{}^{\mu} u^\lambda u^\nu = a^\mu. \tag{5.85}$$

De voorwaarde dat de testmassa stationair is, laat zich direkt vertalen in de voorwaarde

$$u^i = 0, \qquad i = (r, \theta, \varphi). \tag{5.86}$$

Dus de viersnelheid heeft alleen een tijdcomponent $u^0 = cdt/d\tau$. De grootte daarvan volgt uit

$$g_{\mu\nu} u^\mu u^\nu = -c^2 \left(1 - \frac{\alpha}{r}\right) \left(\frac{dt}{d\tau}\right)^2 = -c^2. \tag{5.87}$$

Het resultaat is

$$u^0 = c\frac{dt}{d\tau} = \frac{c}{\sqrt{1 - \dfrac{\alpha}{r}}}. \tag{5.88}$$

We zien zoals verwacht, dat voor een waarnemer in rust op $r = \infty$ de klok ter plekke van een andere stationaire waarnemer op radiële coördinaat r een factor $\sqrt{(1 - \alpha/r)}$ langzamer loopt.

M.b.v. de expliciete uitdrukkingen voor de connecties gegeven in vgln.(5.33) en (5.34) vinden we vervolgens dat

$$a \cdot u = g_{\mu\nu} a^\mu u^\nu = u \cdot \frac{du}{d\tau} + \Gamma_{\lambda\nu\mu} u^\lambda u^\nu u^\mu = \frac{d}{d\tau} \left(g_{\mu\nu} u^\mu u^\nu\right) = 0. \tag{5.89}$$

Door dit te combineren met (5.88) zien we onmiddelijk dat de tijdcomponent van de vierversnelling nul is: $a^0 = 0$. De andere componenten van de vierversnelling volgen nu uit (5.85) en (5.86):

$$a^i = \Gamma_{00}{}^{i} \left(u^0\right)^2 = \delta^i_r \frac{\alpha c^2}{2r^2}. \tag{5.90}$$

Deze berekening laat zien dat een stationaire testmassa, zoals te verwachten, alleen radiëel versneld hoeft te worden, en dat de radiële component van de uitwendige versnelling die nodig is om op de constante afstand $r = R$ te blijven wordt gegeven door

$$a^r|_{r=R} = \frac{GM}{R^2}. \tag{5.91}$$

Dit is precies het resultaat dat uit de Newtonse beschrijving van het zwaartekrachtveld volgt; deze beschrijft dus correct de statische limiet. Wanneer een testmassa precies boven het oppervlak van een bol met de Schwarzschildstraal $r = \alpha$ zou hangen, dan heeft deze een radiële eigenversnelling

$$a_H = \frac{c^2}{2\alpha} = \frac{c^4}{4GM}. \tag{5.92}$$

Deze grootheid, die alleen van de massa afhangt, staat bekend als de *oppervlakte-zwaartekracht* (Eng.: *surface gravity*). Hoewel wij de uitdrukking (5.92) hier hebben afgeleid in een bepaald coördinatensysteem (Drostecoördinaten), is het mogelijk een manifest cördinaten-invariante definitie van deze grootheid te geven, zodat de waarde (5.92) een universele betekenis heeft.

5.9 Zwarte gaten

Tot nu toe hebben we de nadruk gelegd op de interpretatie van de Schwarzschild-Droste oplossing als extern veld van een bolsymmetrische massaverdeling, zoals een isotrope niet-roterende ster. Daarmee hebben we de precessie van het perihelion van Mercurius en de afbuiging van licht door de zon afgeleid. In al deze gevallen is de oplossing alleen van belang in het gebied ruim buiten de Schwarzschildstraal $r > \alpha = 2GM/c^2$. In het binnengebied, waar de stermaterie zich bevindt, is de Schwarzschild-Droste oplossing niet van toepassing.

Wanneer een ster onder invloed van zijn eigen zwaartekracht zou instorten tot een compact object van de afmetingen van een Schwarzschildstraal of kleiner, of wanneer we simpelweg een statisch bolsymmetrisch object hebben van massa M gelocaliseerd binnen de Schwarzschildstraal, dan moet de oplossing ook voor afstanden binnen de Schwarzschildstraal beschouwd worden. Uit de uitdrukking (5.31) voor het lijnelement is het onmiddellijk duidelijk dat hier een nieuwe interpretatie van de geometrie van de ruimte-tijd, beschreven door de Schwarzschild-Droste metriek, nodig is.

In de eerste plaats is het lijnelement in de Drostevorm op de Schwarzschildstraal $r = \alpha$ singulier. Voor waarden van $r < \alpha$ zien we, dat een lijnelement in de r-richting zuiver tijdachtig is:

$$ds^2 = -\frac{dr^2}{\left(\frac{\alpha}{r} - 1\right)} < 0, \tag{5.93}$$

terwijl een lijnelement in de t-richting ruimteachtig is:

$$ds^2 = \left(\frac{\alpha}{r} - 1\right) c^2 dt^2 > 0. \tag{5.94}$$

De rollen van de radiële en tijdcoördinaat worden dus op de Schwarzschildstraal verwisseld. Dit is ook de reden voor het singulier gedrag van de metriek op het

Schwarzschildoppervlak. Deze singulariteit is in bepaalde opzichten vergelijkbaar met de singulariteit van standaard poolcoördinaten voor $r = 0$,; zij is namelijk een gevolg van de coördinatenkeuze en niet van de meetkunde zelf [5]. Binnen de Schwarzschildstraal is de metriek nu niet langer statisch (tijdonafhankelijk): de componenten van de metriek en van de krommingstensor $R_{\mu\nu\lambda\kappa}$ hangen expliciet van de tijd-coördinaat r af. Daarentegen is de tijd-ruimte wel translatie-invariant in de (ruimtelijke) t-richting.

Van deze vreemde verandering van de meetkundige eigenschappen van het tijd-ruimte continuüm voorbij de Schwarzschildstraal is echter voor een waarnemer in rust op $r = \infty$ niets te merken. Voor deze waarnemer valt het gebied voorbij de Schwarzschildstraal namelijk buiten zijn gezichtsveld. Dit is direkt in te zien door gebruik te maken van de oplossing van de bewegingsvergelijking voor een radiëel bewegende testmassa. Deze oplossing, gegeven in vergelijkingen (5.45) en (5.52), is te schrijven als

$$\frac{dt}{d\tau} = \frac{1}{1 - \dfrac{\alpha}{r}}, \qquad r = \alpha \left(1 - \frac{3c\tau}{2\alpha} \right)^{2/3}. \tag{5.95}$$

Merk op dat voor onze waarnemer $\varepsilon = 1$. Volgens de eerste van deze vergelijkingen wordt de waarnemertijd t die verstrijkt per interval eigentijd τ in de buurt van de Schwarzschildstraal oneindig. Tegelijk volgt uit de tweede van deze vergelijkingen, dat de vrije val vanaf ieder punt op eindige afstand r tot aan de Schwarzschildstraal α een eindige hoeveelheid eigentijd in beslag neemt. Voor de waarnemer in $r = \infty$ heeft de testmassa dus een oneindige hoeveelheid laboratoriumtijd nodig om de Schwarzschildstraal te bereiken.

Een soortgelijk resultaat is ook voor licht af te leiden. Langs een radiële lichtachtige geodetische kromme is $ds = 0$ en $d\theta = d\varphi = 0$. Uit de uitdrukking (5.31) voor het lijnelement volgt dan dat bij een invallende lichstraal ($dr/dt < 0$):

$$cdt = -\frac{dr}{\left(1 - \dfrac{\alpha}{r} \right)}. \tag{5.96}$$

Intergratie van deze lineaire differentiaalvergelijking levert:

$$c(t - t_0) = -r + r_0 + \alpha \log \left(\frac{r_0 - \alpha}{r - \alpha} \right). \tag{5.97}$$

Dus ook voor een lichtstraal geldt, dat er oneindig veel waarnemertijd verstrijkt voor hij de Schwarzschildstraal bereikt: $t \to \infty$ als $r \to \alpha$. Om deze reden noemt men het oppervlak met de Schwarzschildstraal $r = \alpha$ de *horizon* voor een externe waarnemer. We leggen er echter de nadruk op, dat een waarnemer

[5] Er zijn andere keuzen van coördinaten, zoals die van Kruskal en Szekeres, waarin dit manifest is; dit valt buiten het bestek van onze schetsmatige behandeling.

die met de testmassa meevalt in principe niets bijzonders hoeft te merken bij het passeren van de Schwarzschildstraal: dit gebeurt in een eindige hoeveelheid eigentijd, en zowel de viersnelheid als de vierversnelling en de kromming blijven er eindig. Voor een elektromagnetische golf geldt, dat het aantal oscillaties waarin de horizon wordt bereikt eindig is, zodat voor een uitwendige waarnemer de frequentie steeds verder omlaag gaat; tegelijk duurt het oneindig lang voordat deze strict nul wordt.

Een eindig eigentijdinterval na het passeren van de horizon bereikt een radiëel invallende testmassa echter de locus $r = 0$, waar zich een echte singulariteit bevindt: hier wordt de kromming oneindig. Voor de testmassa die vanuit rust op $r = \infty$ begonnen is, wordt de periode tussen het passeren van de Schwarzschildstraal en het bereiken van de singulariteit op $r = 0$ (een ruimteachtig oppervlak) volgens de tweede vergelijking (5.95) gegeven door

$$\Delta\tau = \frac{2\alpha}{3c}. \tag{5.98}$$

Voor een lichtstraal zien we uit (5.97) dat de *afstand* $c(t - t_0)$ die wordt afgelegd tussen een *tijd* $(r_0/c) < (\alpha/c)$ en $r/c = 0$ wordt gegeven door

$$c(t - t_0) \;=\; r_0 \;+\; \alpha \log\left(1 - \frac{r_0}{\alpha}\right). \tag{5.99}$$

In de limiet $r_0 \to \alpha$ wordt dit oneindig, wat aantoont dat de coördinatenafstand tussen de horizon en de singulariteit ruimtelijk oneindig is op dezelfde wijze, als waarop de tijd die de lichtstraal volgens een externe waarnemer nodig heeft om de horizon te bereiken oneindig is. De fysische betekenis van deze berekeningen, vooral die van de krommingssingulariteit in $r = 0$, is echter niet direct te duiden. Voor een externe waarnemer blijven ze achter de horizon verborgen.

Voor een fysische interpretatie van de Schwarzschild-Droste oplossing in het hele coördinatengebied $0 < r < \infty$ is vooral het bestaan van de horizon op $r = \alpha$ van belang. De onmogelijkheid waarnemingen voorbij de horizon te doen, is de aanleiding om de ermee corresponderende fysische lichamen als *zwarte gaten* aan te duiden. Behalve de bolsymmetrische zwarte gaten van het Schwarzschild-Droste type, die geheel gekenmerkt worden door hun massa, zijn er ook soortgelijke stationaire oplossingen met een impulsmoment en/of elektrische lading bekend.

Een interessante observatie is, dat er een sterke analogie bestaat tussen de beschrijving van zwarte gaten en de thermodynamica. Om dit duidelijk te maken beschouwen we eerst de oppervlakte van de horizon. Voor een Schwarzschild-Droste zwart gat is deze gegegeven door

$$A_H \;=\; 4\pi\alpha^2 \;=\; \frac{16\pi G^2 M^2}{c^4}. \tag{5.100}$$

Als er materie in het zwarte gat valt, waardoor de massa toeneemt met δM, dan verandert de oppervlakte van de horizon met een bedrag

$$\delta A_H = \frac{32\pi G^2 M}{c^4}\,\delta M. \tag{5.101}$$

We zien dus dat de oppervlakte toeneemt als de massa toeneemt. Uit het bovenstaande mag bovendien duidelijk zijn, dat de val van een testmassa door de horizon van een zwart gat irreversibel is: als de horizon is gepasseerd valt de massa onherroepelijk verder richting singularioteit en ieder causaal contact tussen het testdeeltje en een externe waarnemer is verbroken. Als we nu de massa identificeren met de eigen rustenergie van het zwarte gat, dan geldt onder quasi-statische veranderingen (waarbij er geen impuls wordt overgedragen) dat

$$dE = c^2 dM = \frac{c^2 a_H}{8\pi G}\,dA_H. \tag{5.102}$$

waarbij we de evenredigheidsfactor in termen van de oppervlaktezwaartekracht a_H uit vergelijking (5.92) hebben geschreven. De irreversibiliteit van de absorptie van materie door een zwart gat suggereert dan een identificatie van deze vergelijking met de eerste hoofdwet van de thermodynamica:

$$dE = TdS, \tag{5.103}$$

waarin we de entropie evenredig laten zijn met oppervlak; we kunnen deze evenredigheid beschrijven door een vergelijking

$$\frac{dS}{k_B} = \eta\,\frac{dA_H}{\ell_{Pl}^2}, \tag{5.104}$$

waarin k_B de Boltzmannconstante is, η een nader te bepalen dimensieloze constante, en ℓ_{Pl} is de Plancklengte[6]:

$$\ell_{Pl} = \sqrt{\frac{2Gh}{c^3}} \approx 0.57 \times 10^{-34}\,\mathrm{m}. \tag{5.105}$$

De temperatuur van het zwarte gat wordt dan gegeven door

$$T = \frac{h}{4\pi\eta c k_B} a_H. \tag{5.106}$$

Deze thermodynamische beschrijving van processen waaraan zwarte gaten deelnemen is ontwikkeld door Bekenstein. Hawking heeft vervolgens laten zien dat de combinatie van quantumveldentheorie met de aanwezigheid van een horizon

[6] We gebruiken de niet-gereduceerde Planckconstante $h = 2\pi\hbar$.

in de tijd-ruimte inderdaad een temperatuur toekent aan de horizon: het zwarte gat straalt als een zwart lichaam met een temperatuur

$$T_H = \frac{h}{4\pi^2 c k_B} a_H. \tag{5.107}$$

De evenredigheidsconstante heeft volgens Hawking dus de waarde

$$\eta = \pi, \tag{5.108}$$

zodat de entropie van het zwarte gat bepaald wordt uit de uitdrukking

$$dS = \frac{\pi k_B}{\ell_{Pl}^2} dA_H. \tag{5.109}$$

Uit de waarde van η volgt, dat

$$4\pi^2 k_B T_H = \ell_{Pl} M_{Pl}\, a_H, \tag{5.110}$$

met M_{Pl} de Planckmassa:

$$M_{Pl} = \sqrt{\frac{hc}{2G}} \approx 3.86 \times 10^{-8}\,\text{kg}. \tag{5.111}$$

De rechterzijde van vergelijking (5.110) komt dus overeen met de hoeveelheid arbeid die het zwarte gat levert door een Planckmassa over een Plancklengte te versnellen met de oppervlaktezwaartekracht op de horizon.

Samenvatting van hoofdstuk 5

De Einsteinvergelijkingen in de lege ruimte hebben niet-triviale oplossingen. Twee soorten vacuumvelden met bijzondere eigenschappen worden hier besproken: vlakke gravitatiegolven en het statische bolsymmetrische Schwarzschild-Droste-veld. Deze oplossingen zijn het gravitatieëquivalent van de elektromagnetische golven en het Coulombveld in de Maxwelltheorie, besproken in hoofdstuk 3. We hebben de banen van testdeeltjes in een bolsymmetrisch veld bepaald, waarmee we o.a. klassieke tests van de ART zoals de precessie van het perihelium van Mercurius en de afbuiging van licht door de zon kunnen narekenen. Nieuwere toepassingen, zoals de analyse van de beweging van de dubbele pulsar PSR 1913+16 en het gravitatielenseffect zijn kort toegelicht. Tenslotte hebben we de interpretatie van de Schwarzschild-Droste oplossing als die van een zwart gat gegeven en de thermodynamica van objecten met een horizon besproken.

Hoofdstuk 6

De inhomogene Einsteinvergelijkingen

6.1 De kosmologische constante

De inhomogene Einsteinvergelijkingen (4.10) veronderstellen een lineair lokaal verband tussen de kromming van de tijd-ruimte en de dichtheid van energie en impuls en de druk ter plaatse. Deze dichtheid wordt beschreven door een symmetrische lokale tensor $T_{\mu\nu}(x)$ met de dimensie van energie per volume, die moet voldoen aan de consistentieëis (4.69) dat de covariante divergentie nul is:

$$D_\nu T^\nu_{\ \mu} = 0. \tag{6.1}$$

De dimensie van energiedichtheid en druk is in overeenstemming met de dimensie van de Einsteintensor: $\dim[G_{\mu\nu}] = l^{-2}$, en die van de evenredigheidsconstante: $\dim[G/c^4] = t^2 m^{-1} l^{-1}$; namelijk

$$\begin{aligned} \dim\left[T_{\mu\nu}\right] = \dim\left[G_{\mu\nu}\right] \dim\left[\frac{c^4}{G}\right] &= \frac{m}{lt^2} = \frac{mv^2}{l^3} = \dim\left[\frac{E}{V}\right] \\ &= \frac{F}{l^2} = \dim\left[p\right]. \end{aligned} \tag{6.2}$$

We zien, dat energiedichtheid (energie/volume) en druk (kracht/oppervlak) dezelfde dimensies hebben.

Een specifieke tensor met deze eigenschappen is eenvoudig te construeren in termen van het gravitatieveld zelf. De tensor

$$T_{\mu\nu} = -\varepsilon_\Lambda \, g_{\mu\nu}, \tag{6.3}$$

met ε_Λ een constante met de dimensie van een energiedichtheid of druk, voldoet automatisch aan de consistentieëis wegens het metrisch postulaat, vgl.(4.40). Er zijn in dit geval dus geen veld- of bewegingsvergelijkingen van straling of materie nodig om er voor te zorgen dat de covariante divergentie van deze tensor nul wordt. Dit betekent dat er geen andere fysische vrijheidsgraden dan die van het gravitatieveld zelf in het spel zijn; in deze zin beschrijft de constante ε_Λ de energiedichtheid van een *lege ruimte* (een ruimte zonder materie of straling).

Een andere interpretatie van het rechterlid van vergelijking (6.3) volgt onmiddellijk uit de Einsteinvergelijking (4.10) door er het spoor van te nemen:

$$R = -\frac{32\pi G}{c^4}\,\varepsilon_\Lambda. \tag{6.4}$$

De kromming van de lege ruimte is dus constant en evenredig met de energiedichtheid ε_Λ. Voor de zo gevonden constante kromming gebruikt men gewoonlijk het symbool Λ; de relatie tussen Λ en ε_Λ is dus

$$\varepsilon_\Lambda = -\frac{\Lambda c^4}{32\pi G}. \tag{6.5}$$

Einstein voerde een dergelijke constante in zijn theorie in wegens de belangrijke invloed die hij kan hebben op de grootschalige evolutie van het heelal. Om deze reden wordt Λ de *kosmologische constante* genoemd.

6.2 Ideale gassen en vloeistoffen

Om de wisselwerking tussen het zwaartekrachtveld en macroscopische materie en straling te beschrijven, gebruiken we het model van een simpele ideale vloeistof. Een ideale homogene en isotrope vloeistof —waaronder we ook een homogeen en isotroop gas verstaan— wordt gekarakteriseerd door twee parameters: de lokale massa- of energiedichtheid $\varepsilon(x) = \rho(x)c^2$, waarin ρ de dimensie van massa per volume heeft, en de lokale druk $p(x)$. We zullen hier geen uitvoerige discussie aan de betekenis van deze grootheden in een relativistische context wijden. We interpreteren ze eenvoudig als de betreffende grootheden in een lokale thermische evenwichtssitutatie in een lokaal rustsysteem. In de klassieke niet-relativistische hydrodynamica, en in de afwezigheid van externe velden zoals dat van de zwaartekracht, voldoen deze grootheden aan twee vergelijkingen: de *continuiteitsvergelijking*:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0, \tag{6.6}$$

die inhoudt, dat de verandering in de dichtheid in evenwicht is met de netto uitstroom/instroom van massa per volumeëenheid; en de *Eulervergelijking*:

$$\rho \left(\frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla \right) v_i = -\nabla_i p. \tag{6.7}$$

Deze laatste vergelijking bepaalt dat meebewegende veranderingen in de stroomsnelheid van een punt in de vloeistof worden bepaald door de drukgradient.

De hydrodynamische vergelijkingen (6.6) en (6.7) hebben een directe relativistische generalisatie, die kan worden uitgedrukt in termen van een energie-impulstensor

$$T_{\mu\nu} \,=\, \rho u_\mu u_\nu \,+\, p\left(g_{\mu\nu} + \frac{u_\mu u_\nu}{c^2}\right). \tag{6.8}$$

De vorm van deze uitdrukking impliceert dat ρ en p scalaire grootheden zijn. In de vlakke Minkowski-ruimte, waarbij gravitatie-effecten buiten beschouwing worden gelaten en $g_{\mu\nu} = \eta_{\mu\nu}$, is de covariante divergentie van deze tensor

$$\partial_\nu T^\nu{}_\mu \,=\, \partial_\mu p \,+\, u_\mu u^\nu \partial_\nu\left(\rho + \frac{p}{c^2}\right) \,+\, \left(\rho + \frac{p}{c^2}\right)\partial_\nu\left(u_\mu u^\nu\right). \tag{6.9}$$

Als we nu de tijd- en ruimtecomponenten van deze uitdrukking afzonderlijk uitwerken, dan krijgen we voor de tijdcomponent:

$$\begin{aligned}
\partial_\nu T^\nu{}_0 \;&=\; \frac{1}{c}\frac{\partial p}{\partial t} - \frac{c}{(1-\mathbf{v}^2/c^2)}\left(\frac{\partial}{\partial t} + \mathbf{v}\cdot\nabla\right)\left(\rho + \frac{p}{c^2}\right) \\
&\quad -c\left(\rho + \frac{p}{c^2}\right)\left[\frac{\partial}{\partial t}\frac{1}{(1-\mathbf{v}^2/c^2)} + \nabla\cdot\frac{\mathbf{v}}{(1-\mathbf{v}^2/c^2)}\right], \qquad (6.10) \\
&=\; \frac{1}{c}\frac{\partial p}{\partial t} - c\left(\frac{\partial}{\partial t} + \nabla\cdot\mathbf{v}\right)\left(\frac{\rho + p/c^2}{1-\mathbf{v}^2/c^2}\right).
\end{aligned}$$

en voor de ruimtecomponenten:

$$\begin{aligned}
\partial_\nu T^\nu{}_i \;&=\; \nabla_i p + \frac{v_i}{(1-\mathbf{v}^2/c^2)}\left(\frac{\partial}{\partial t} + \mathbf{v}\cdot\nabla\right)\left(\rho + \frac{p}{c^2}\right) \\
&\quad +\left(\rho + \frac{p}{c^2}\right)\left(\frac{\partial}{\partial t}\left[\frac{v_i}{(1-\mathbf{v}^2/c^2)}\right] + \nabla\cdot\left[\frac{\mathbf{v}v_i}{(1-\mathbf{v}^2/c^2)}\right]\right).
\end{aligned} \tag{6.11}$$

met $i = (1, 2, 3)$. Door de eerste vergelijking met de tweede te combineren, kan deze laatste worden herschreven in de vorm

$$\partial_\nu T^\nu{}_i \,+\, \frac{v_i}{c}\,\partial_\nu T^\nu{}_0 \,=\, \left(\nabla_i + \frac{v_i}{c^2}\frac{\partial}{\partial t}\right)p \,+\, \frac{(\rho + p/c^2)}{(1-\mathbf{v}^2/c^2)}\left(\frac{\partial}{\partial t} + \mathbf{v}\cdot\nabla\right)v_i. \tag{6.12}$$

Nemen we nu in vergelijking (6.12) en in vergelijking (6.10) —na deling door c— de limiet $c \to \infty$, dan vinden we de uitdrukkingen terug die volgens de continuiteitsvergelijking en de Eulervergelijking nul zijn. Met andere woorden, de enkele conditie

$$\partial_\nu\, T^\nu{}_\mu \,=\, 0 \tag{6.13}$$

voor de energie-impulstensor (6.8) in de Minkowskiruimte levert een relativistische versie van de beide hydrodynamische vergelijkingen op. We interpreteren daarom deze energie-impulstensor als die van een relativistische ideale vloeistof. In een algemene (niet-noodzakelijk vlakke) tijd-ruimte krijgen we voor zo'n vloeistof een consistent stelsel hydrodynamische vergelijkingen uit dezelfde energie-impulstensor —zij het met de bijbehorende uitdrukkingen voor de metrische componenten $g_{\mu\nu}$— door de algemenere eis (6.1) op te leggen. Immers, deze vergelijking is manifest covariant, en als er geen gravitatievelden zijn, zodat $g_{\mu\nu} = \eta_{\mu\nu}$ en ${\Gamma_{\mu\nu}}^{\lambda} = 0$, dan reduceert hij tot de speciaal-relativistische uitdrukking (6.13).

We eindigen deze paragraaf met de opmerking dat de kosmologische constante een speciaal geval van een ideale vloeistof vertegenwoordigt, nl. het geval waarbij

$$\varepsilon_\Lambda = \rho c^2 = -p. \tag{6.14}$$

In dit geval zijn de energiedichtheid en druk constant, waarbij als bijzondere eigenschap moet worden aangetekend dat altijd een van de twee negatief is. Het is dan ook niet mogelijk een materiëel gas te vinden dat deze eigenschappen heeft, hetgeen nogmaals de aparte status van de kosmologische constante bevestigt.

6.3 De toestandsvergelijking

In een ideaal gas of ideale vloeistof zijn de druk en energiedichtheid niet onafhankelijk van elkaar. Indien er lokaal thermisch evenwicht heerst, kunnen ze aan elkaar worden gerelateerd via de thermodynamische toestandsfunctie $U(S, V)$, die de energie U uitdrukt in termen van de entropie S en het volume V:

$$\rho c^2 = \frac{U}{V}, \qquad p = -\left(\frac{\partial U}{\partial V}\right)_S. \tag{6.15}$$

In veel gevallen zijn deze grootheden evenredig met elkaar:

$$p = \eta \rho c^2. \tag{6.16}$$

Een dergelijk verband is in overeenstemming met het scalaire karakter zowel als met de dimensie van beide grootheden. We zullen de toestandvergelijking (6.16) voor een paar simpele gevallen beargumenteren, met de bijbehorende waarde van de dimensieloze parameter η.

We beschouwen eerst een niet-relativistisch ideaal gas van N deeltjes met massa m in een volume V, en temperatuur $T \ll mc^2/k_B$. Als we thermodynamische gemiddelden aanduiden met $\langle ... \rangle$, dan geldt volgens de equipartitiestelling

$$U = N \langle \frac{mc^2}{\sqrt{1 - \mathbf{v}^2/c^2}} \rangle \approx N \langle mc^2 + \frac{1}{2} m \mathbf{v}^2 + ... \rangle \approx Nmc^2 + \frac{3}{2} N k_B T. \tag{6.17}$$

De energiedichtheid is dan

$$\rho c^2 = \frac{U}{V} = \frac{Nmc^2}{V}\left(1 + \frac{3k_BT}{2mc^2}\right), \tag{6.18}$$

waarbij de eerste term veel groter is dan de tweede: $k_BT \ll mc^2$. Op eenzelfde manier kunnen we het verband tussen druk en temperatuur van een gas schrijven als

$$p = \frac{Nk_BT}{V} = \frac{Nmc^2}{V}\frac{k_BT}{mc^2}. \tag{6.19}$$

In eerste benadering vinden we hieruit dat

$$p \approx \rho\frac{k_BT}{m} = \eta\,\rho c^2, \qquad \eta = \frac{k_BT}{mc^2} \ll 1. \tag{6.20}$$

Naarmate de temperatuur stijgt en het gas relativistisch wordt is de bijdrage van de rustenergie mc^2 aan de totale energie relatief onbelangrijker. Voor een gas van massaloze deeltjes –straling in thermisch evenwicht– is de uitdrukking voor de energie als functie van de temperatuur:

$$U = 3Nk_BT \qquad \Rightarrow \qquad \rho c^2 = 3\,\frac{Nk_BT}{V}. \tag{6.21}$$

In combinatie met vergelijking (6.19) voor de druk vinden we nu

$$\rho c^2 = 3p. \tag{6.22}$$

In deze limiet is η niet langer temperatuurafhankelijk, maar heeft de constante waarde

$$\eta = \frac{1}{3}. \tag{6.23}$$

De standaard afleiding van vergelijking (6.21) is gebaseerd op de statistiek van massaloze deeltjes zonder wisselwerking in de Minkowskiruimte. Een eenvoudig heuristisch argument laat echter zien dat de toestandsvergelijking (6.22) voor straling ook van toepassing is bij inachtneming van de koppeling aan het gravitatieveld. Daartoe nemen we het scalaire spoor van de energie-impulstensor en vinden

$$T_\mu{}^\mu = 3p - \rho c^2. \tag{6.24}$$

Het verband (6.22) tussen druk en temperatuur geldt dus altijd wanneer $T_\mu{}^\mu = 0$. We kunnen beargumenteren dat dit voor een gas van massaloze deeltjes inderdaad het geval is.

Daartoe merken we op, dat volgens de Einsteinvergelijking de krommingscalar evenredig is met het spoor van de energie-impulstensor:

$$R = \frac{8\pi G}{c^4}T_\mu{}^\mu = \frac{8\pi G}{c^4}\left(3p - \rho c^2\right). \tag{6.25}$$

De krommingsscalar heeft de dimensie van $1/(\text{lengte})^2$. Voor een niet-relativistisch gas is de drukterm verwaarloosbaar, en legt vergelijking (6.25) een lineair verband tussen twee dimensievolle grootheden: de gemiddelde kromming en de gemiddelde massadichtheid $\rho \approx Nm/V$. In het geval van massaloze deeltjes zoals b.v. fotonen is het niet mogelijk zo'n verband te construeren, en is de Riemannscalar nul; dan geldt dus

$$T_\mu{}^\mu = 0 \qquad \Rightarrow \qquad \rho c^2 \,=\, 3p. \tag{6.26}$$

De toestandvergelijking voor het relativistische gas geldt daarom in ieder gravitatieveld, zolang de energie-impuls tensor wordt gegeven door de uitdrukking (6.8) en het spoor nul is.

Tenslotte beschouwen we het speciale geval van een ruimte zonder materie of straling, maar met een kosmologische constante. Dan is

$$\rho c^2 \,=\, -\,p \,=\, \varepsilon_\Lambda \qquad \Leftrightarrow \qquad \eta = -1. \tag{6.27}$$

In dit geval geldt dus ook automatisch een lineair verband tussen druk en energiedichtheid, zij het van ongebruikelijke aard. Zoals eerder opgemerkt, wordt de krommingscalar dan direct gegeven door de kosmologische constante: $R = \Lambda$. De waarde van de kosmologische constante is a priori onbekend. Als $\Lambda \neq 0$, dan kunnen we een bijbehorende massaschaal M definiëren door de bijbehorende massadichtheid gelijk te stellen aan een massa-eenheid per kubieke Comptongolflengte:

$$|\varepsilon_\Lambda| \,=\, \frac{Mc^2}{\left(\dfrac{h}{Mc}\right)^3} \,=\, \frac{(Mc^2)^4}{(hc)^3}. \tag{6.28}$$

Nemen we voor M de Planckmassa:

$$M_{Pl} \,=\, \sqrt{\frac{hc}{2G}} \,\approx\, 3.86 \times 10^{-8}\,\text{kg}, \tag{6.29}$$

dan is de bijbehorende energiedichtheid die van een Planckmassa per Planckvolume, ofwel

$$\varepsilon_{Pl} \,\approx\, 10^{112}\,\text{Jm}^{-3}. \tag{6.30}$$

De huidige bovenlimiet uit astrofysische waarnemingen op kosmische schaal is echter van de orde

$$|\varepsilon_\Lambda| \leq 10^{-9}\,\text{Jm}^{-3}, \tag{6.31}$$

hetgeen correspondeert met een massa van $M_\Lambda \leq 2 \times 10^{-38}$ kg, met een energie van $M_\Lambda c^2 \leq 10^{-2}$ eV, of met een temperatuur van $T_\Lambda = M_\Lambda c^2/k_B \leq 120$ K. Het opvallende verschil tussen deze dichtheden geeft aan dat ons heelal zeer ver van de Planckschaal verwijderd is.

Samenvatting van hoofdstuk 6

De Einsteinvergelijkingen bepalen het gravitatieveld van een gegeven energie- en impulsdichtheid van materie en straling. In de praktijk is dit een macroscopische verdeling, die in veel gevallen kan worden beschreven als die van een ideaal gas of een ideale vloeistof. In dit hoofdstuk hebben we de bijbehorende uitdrukking voor de energie-impulstensor afgeleid, en laten zien dat deze volledig is bepaald door de druk en de dichtheid van het gas. Deze twee parameters zijn met elkaar verbonden door de thermodynamische toestandsvergelijking, die we expliciet hebben besproken voor relativistische en niet-relativistische gassen en vloeistoffen. Een bijzonder geval is dat van een constante energiedichtheid, die samenhangt met een constante kromming van de tijd-ruimte: de kosmologische constante. Deze valt ook onder het geval van een ideaal gas, maar de toestandsvergelijking is van een ongebruikelijke aard: of de druk of de dichtheid moet negatief zijn.

Hoofdstuk 7

Structuur en evolutie van het heelal

7.1 Homogeniteit en isotropie van het heelal

Op basis van de inhomogene Einsteinvergelijkingen besproken in hoofdstuk 6, kunnen we een eenvoudig model voor de grootschalige structuur en evolutie van het heelal ontwikkelen. Dit model gaat uit van de aanname dat op zeer grote schaal gezien, het waarneembare heelal homogeen en isotroop is. Deze aanname staat soms bekend als het *kosmologisch beginsel*. Waarnemingen van de verdeling van clusters van melkwegstelsels op afstandschalen van 100 megaparsec (Mpc) of meer zijn in goede overeenstemming met dit uitgangspunt. De meest indrukwekkende gegevens komen echter van de kosmische microgolf-achtergrondstraling (de CMBR, wat staat voor *Cosmic Microwave Background Radiation*). Deze reststraling uit een zeer vroege fase van het heelal heeft een vrijwel zuiver zwartlichaamspectrum behorend bij een temperatuur[1] van

$$T_{CMBR} = 2.728\,\mathrm{K}. \tag{7.1}$$

De intrinsieke variaties in temperatuur van de straling uit verschillende richtingen van de hemel zijn buitengewoon klein: van de orde van 1 op 10^5 of kleiner. De isotrope verdeling is een aanwijzing, dat het kosmologisch beginsel ook in het vroege heelal van toepassing is geweest.

Het is zelfs waarschijnlijk dat de verdeling in het vroege heelal homogener was dan nu. Het heelal was toen heter en compacter: de temperatuur, druk en dichtheid van de straling en de materie waren zo hoog dat ze onder het relativistische regime vielen. De elektro-magnetische wisselwerkingen en, in een eerder

[1] Dit komt overeen met een gemiddelde energie per foton van $E = kT = 0.235 \times 10^{-3}$ eV; in dit hoofdstuk zullen we temperaturen regelmatig weergeven in termen van de gemiddelde energie per vrijheidsgraad door impliciete vermenigvuldiging met de Boltzmannconstante; deze heeft de numerieke waarde $k = 8.617 \times 10^{-5}$ eV K^{-1}.

stadium, de sterke en zwakke kernkrachten speelden daarom een veel belangrijker rol op macroscopische schaal; de interacties verliepen snel, en relaxatietijden en vrije weglengten waren kort. Gravitatiekrachten waren daarentegen zwak en onbelangrijk, hun belangrijkste effect was de uitdijing van het heelal als geheel. Pas nadat het heelal zover was uitgedijd en afgekoeld, dat er slechts een geringe hoeveelheid straling en een ijl gas van niet-relativistische materie was overgebleven, kon de zwaartekracht een belangrijkere rol gaan spelen en compactere objecten vormen zoals gas- en stofwolken, en uiteindelijk melkwegstelsels met daarin sterren en planeten. De inhomogeniteit van de materie in het heelal die we nu waarnemen op afstandschalen $\simeq 1$ Mpc en kleiner (veel kleiner dan de straal van het zichtbare heelal), en die zichtbaar is in de vorm van melkwegstelsels, is dus pas relatief laat in de ontwikkeling van het heelal tot stand gekomen. Hoe deze tot stand is gekomen, is een interessante vraag waarover veel theoretische ideeën zijn geopperd. Met de nieuwe gegevens over de fluctuaties in het spectrum van de microgolf-achtergrondstraling kunnen sommige van deze ideeën quantitatief getest worden. Mogelijk weten we in de nabije toekomst meer over dit stuk van de geschiedenis van het heelal.

De evolutiemodellen die we hier behandelen zijn van toepassing op de grootschalige structuur van het heelal, en spelen vooral een rol in de vroegste ontwikkelingsfasen, voor het ontstaan van sterren en melkwegstelsels. Daarbij is het kosmologisch beginsel ons uitgangspunt. Dit moet echter eerst nog wat aangescherpt worden. De uitspraak dat het heelal homogeen en isotroop is kan namelijk niet onvoorwaardelijk voor iedere waarnemer gelden. Hij kan op zijn best gelden voor een waarnemer die ten opzichte van de materie in zijn omgeving (gemiddeld) in rust is, een waarnemer die op kosmische schaal gezien dus passief met de materie meebeweegt op de uitdijing van het heelal. Voor andere waarnemers komen de effecten van de eigen beweging in het spel, zoals de dopplerverschuiving van spectraallijnen die afhankelijk is van de richting van de beweging t.o.v. de richting van de lichtbron — een ster of een melkwegstelsel.

En zelfs voor meebewegende waarnemers geldt het kosmologisch beginsel alleen bij gebruik van een speciale tijdcoördinaat. Immers, gelijktijdigheid is een relatief begrip, en twee waarnemers die ten opzichte van elkaar in rust zijn en die locaal dezelfde tijd gebruiken kunnen nog verschillen in hun conventies voor het vastleggen van de plaats en tijd van gebeurtenissen ver weg in het heelal. De preciezere formulering van het kosmologische beginsel die we nodig hebben luidt daarom, dat de grootschalige structuur van het heelal gekenmerkt wordt door het bestaan van een coördinatenstelsel $(ct, \mathbf{r})$, waarin de drie-dimensionale ruimten $t = constant$ de eigenschap hebben dat ze homogeen en isotroop zijn. Homogeniteit en isotropie houden hier in, dat zulke drie-dimensionale ruimten er hetzelfde uitzien voor meebewegende waarnemers op verschillende plaatsen met verschillende orientaties van hun coördinatensystemen. De speciale tijdcoördinaat waarvoor dit geldt noemen we de *kosmologische tijd*.

7.2 De meetkunde van de ruimte

Het kosmologisch beginsel legt een sterke beperking op aan de meetkundige structuur van de ruimte. Voor een waarnemer in rust t.o.v. de grootschalige materieverdeling in het heelal moet de ruimte in alle richtingen dezelfde eigenschappen hebben. In het bijzonder moet de kromming van de drie-dimensionale ruimte $t = constant$ overal hetzelfde zijn. Er zijn daarom maar drie verschillende klassen van ruimtelijke structuur mogelijk:
1. De kromming van de ruimte heeft overal dezelfde positieve waarde; dan is de ruimte in alle richtingen gesloten en heeft de eigenschappen van het driedimensionale oppervlak van een vier-dimensionale bol.
2. De ruimte is in alle richtingen open en vlak; er is geen kromming in drie-dimensionale zin, en de meetkunde van de driedimensionale ruimte $t = constant$ is euclidisch.
3. De kromming van de ruimte is konstant en negatief; de ruimte is in alle richtingen open en heeft de eigenschappen van het driedimensionale oppervlak van een vier-dimensionale hyperboloide.
We zullen zien dat voor ieder van deze gevallen de metriek van de tijd-ruimte volledig is vastgelegd op een schaalfactor na. We beschouwen een drie-dimensionale ruimte $t = constant$ met vaste kromming. Voor het geval van een gesloten ruimte is dit een drie-dimensionaal boloppervlak. We kunnen dit beschrijven door het in te bedden in een fictieve vier-dimensionale euclidische hulpruimte met behulp van de vergelijking

$$x_1^2 + x_2^2 + x_3^2 + x_4^2 = a^2, \tag{7.2}$$

waarin a de straal van de bol is. Voor een lijnelement dx_a tussen naburige punten op de bol geldt dan

$$\sum_{a=1}^{4} x_a dx_a = 0 \qquad \Rightarrow \qquad dx_4 = -\frac{\mathbf{r} \cdot d\mathbf{r}}{x_4} = -\frac{\mathbf{r} \cdot d\mathbf{r}}{\sqrt{a^2 - \mathbf{r}^2}}. \tag{7.3}$$

We gebruiken de notatie $\mathbf{r}$ voor de drie-dimensionale vector (x_1, x_2, x_3) verkregen door projectie van x_a in de richting van de vierde coördinaat. De vierdimensionale lengte van het lijnelement dx_a is

$$\begin{aligned} dl^2 &= d\mathbf{r}^2 + dx_4^2 = d\mathbf{r}^2 + \frac{(\mathbf{r} \cdot d\mathbf{r})^2}{a^2 - \mathbf{r}^2} \\ &= \frac{(a^2 - \mathbf{r}^2)\, d\mathbf{r}^2 + (\mathbf{r} \cdot d\mathbf{r})^2}{a^2 - \mathbf{r}^2}. \end{aligned} \tag{7.4}$$

We gaan nu over op poolcoördinaten (ρ, θ, φ) gedefiniëerd door

$$\begin{aligned} x_1 &= \rho \sin\theta \cos\varphi, \\ x_2 &= \rho \sin\theta \sin\varphi, \\ x_3 &= \rho \cos\theta, \end{aligned} \tag{7.5}$$

met $\rho^2 = \mathbf{r}^2 = a^2 - x_4^2 \leq a^2$. Voor het lijnelement in het boloppervlak vinden we dan

$$\begin{aligned} dl^2 &= \frac{(a^2 - \rho^2)\left(\rho^2 d\theta^2 + \rho^2 \sin^2\theta\, d\varphi^2\right) + a^2 d\rho^2}{a^2 - \rho^2} \\ &= \frac{a^2 d\rho^2}{a^2 - \rho^2} + \rho^2\, d\theta^2 + \rho^2 \sin^2\theta\, d\varphi^2 \\ &= a^2 \left(\frac{dr^2}{1 - r^2} + r^2 d\theta^2 + r^2 \sin^2\theta\, d\varphi^2 \right). \end{aligned} \tag{7.6}$$

In de laatste regel hebben we de radiële coördinaat herschaald tot $r = \rho/a \leq 1$, zodat we een canonieke parametrisatie in termen van coördinaten op de eenheidsbol krijgen. Dit vereenvoudigt de berekeningen die gebruik maken van deze metriek aanzienlijk. Voor het bepalen van fysische afstanden moeten we deze herschaling natuurlijk ongedaan maken en terugkeren tot de coördinaat ρ.

We kunnen een analoge formulering vinden voor ruimten met constante negatieve kromming, uitgaande van de inbedding in een vier-dimensionale minkowskiruimte via

$$x_4^2 = a^2 + x_1^2 + x_2^2 + x_3^2 \geq a^2, \tag{7.7}$$

waarbij het vier-dimensionale ruimte-achtige interval van een lijnelement in dit oppervlak wordt gegeven door

$$dl^2 = d\mathbf{r}^2 - dx_4^2 = \frac{(a^2 + \mathbf{r}^2)\, d\mathbf{r}^2 - (\mathbf{r} \cdot d\mathbf{r})^2}{a^2 + \mathbf{r^2}}. \tag{7.8}$$

Door over te gaan op poolcoördinaten (7.5) krijgen we nu

$$\begin{aligned} dl^2 &= \frac{a^2 d\rho^2}{a^2 + \rho^2} + \rho^2 d\theta^2 + \rho^2 \sin^2\theta\, d\varphi^2 \\ &= a^2 \left(\frac{dr^2}{1 + r^2} + r^2 d\theta^2 + r^2 \sin^2\theta\, d\varphi^2 \right). \end{aligned} \tag{7.9}$$

In de laatste regel hebben we dezelfde herschaling doorgevoerd die we ook voor het geval van positieve kromming hebben gebruikt om naar canonieke coördinaten gebaseerd op een eenheidshyperboloide over te gaan.

Tenslotte beschouwen we het geval van een vlakke drie-dimensionale ruimte. Hier is geen referentielengte, zoals de straal van een bol of de corresponderende minimale waarde van $|x_4|_{min} = \sqrt{a^2 + \mathbf{r}^2}|_{min} = a$ in de voorgaande gevallen; in poolcoördinaten is

$$dl^2 \,=\, d\rho^2 \,+\, \rho^2 d\theta^2 \,+\, \rho^2 \sin^2\theta \, d\varphi^2. \tag{7.10}$$

In deze zin is de vlakke ruimte schaalinvariant: onder herschaling $r = \rho/a$ behoudt de metriek haar vorm op een universele factor a^2 na:

$$dl^2 \,=\, a^2 \left(dr^2 + r^2 d\theta^2 + \sin^2\theta \, d\varphi^2 \right). \tag{7.11}$$

De factor a^2 hier en in de uitdrukkingen voor de sferische en hyperbolische lijnelementen, kan zo nodig worden geëlimineerd door aanpassing van de eenheden waarin dl wordt gemeten.

Dit wordt echter anders, wanneer we de ontwikkeling van de geometrie in de tijd beschouwen. Eerst merken we op, dat de vergelijkingen (7.6), (7.9) en (7.11) kunnen worden samengevat in de uitdrukking

$$dl^2 \,=\, a^2 \left(\frac{dr^2}{1 - kr^2} + r^2 d\theta^2 + r^2 \sin^2\theta \, d\varphi^2 \right), \tag{7.12}$$

met $k = (1, 0, -1)$ voor respectievelijk een sferische, een vlakke of een hyperbolische ruimte. Volgens het kosmologisch beginsel beschrijven deze uitdrukkingen de metriek van een driedimensionaal hypervlak $t = constant$ in de vier-dimensionale tijd-ruimte; op dit hypervlak is $dt = 0$ en $a = constant$. Met andere woorden, de vier-dimensionale metriek van de tijd-ruimte die een isotroop en homogeen heelal beschrijft moet de eigenschap hebben dat ze voor $dt = 0$ reduceert tot de vorm (7.12). In een niet-roterend coördinatenstelsel betekent dit nu, dat we de tijdcoördinaat zo kunnen kiezen dat

$$ds^2 \,=\, -c^2 dt^2 \,+\, a^2(t) \left(\frac{dr^2}{1 - kr^2} + r^2 d\theta^2 + r^2 \sin^2\theta \, d\varphi^2 \right). \tag{7.13}$$

Het is deze tijdcoördinaat die we met de kosmologische tijd associëren.

We kunnen de evolutie van een tijd-ruimte met gegeven ruimtelijke structuur (een vaste waarde van k) dus volledig beschrijven door aan te geven hoe de schaalfactor $a(t)$ tijdens die ontwikkeling in de tijd verandert. Daarmee heeft $a(t)$ een fysische betekenis gekregen: we kunnen nog steeds de coördinaten zo herschalen dat op een bepaald tijdstip $t = t_0$ de schaalfactor de eenheid is: $a(t_0) = 1$, maar dat is dan in het algemeen niet meer zo op andere tijdstippen $t \neq t_0$. Naar degenen die deze heelalmodellen het eerst hebben bestudeerd, worden tijd-ruimten beschreven door het lijnelement (7.13) ook Friedmann-Robertson-Walker (FRW)-modellen genoemd.

7.3 De kosmologische vergelijkingen

7.3.1 De Einsteinvergelijkingen

De tijd-ruimte metriek (7.13) van een homogeen en isotroop heelal kan in matrix-notatie worden geschreven als

$$g_{\mu\nu} = \begin{pmatrix} -1 & 0 \\ & \\ 0 & a^2(t) g^*_{ij} \end{pmatrix}, \tag{7.14}$$

waarin g^*_{ij} de herschaalde eenheidsmetriek van een de driedimensionale ruimte is:

$$g^*_{ij} = \begin{pmatrix} \frac{1}{1-kr^2} & 0 & 0 \\ & & \\ 0 & r^2 & 0 \\ & & \\ 0 & 0 & r^2 \sin^2\theta \end{pmatrix}. \tag{7.15}$$

In deze notatie is

$$ds^2 = -c^2 dt^2 + a^2(t)\, g^*_{ij} dx^i dx^j, \tag{7.16}$$

met $x^i = (r, \theta, \varphi)$. De niet-triviale componenten van de connectie zijn:

$$\begin{aligned} \Gamma_{ij}{}^{0} &= \frac{a\dot{a}}{c}\, g^*_{ij}, \qquad \Gamma_{0j}{}^{i} = \frac{\dot{a}}{ca}\, \delta^i_j, \\ \Gamma_{kj}{}^{i} &= \Gamma^*_{kj}{}^{i} == \frac{1}{2}\, g^{il}_*\, \left(g^*_{lj,k} + g^*_{kl,j} - g^*_{kj,l} \right). \end{aligned} \tag{7.17}$$

De expliciete uitdrukkingen voor de statische connnecties op de eenheidsruimte zijn

$$\begin{aligned} \Gamma^*_{rr}{}^{r} &= \frac{kr}{1-kr^2}, \\ \Gamma^*_{\theta\theta}{}^{r} &= -r(1-kr^2), \qquad & \Gamma^*_{r\theta}{}^{\theta} &= \frac{1}{r}, \\ \Gamma^*_{\varphi\varphi}{}^{r} &= -r\sin^2\theta\,(1-kr^2), \qquad & \Gamma^*_{r\varphi}{}^{\varphi} &= \frac{1}{r}, \\ \Gamma^*_{\varphi\varphi}{}^{\theta} &= -\sin\theta\,\cos\theta, \qquad & \Gamma^*_{\theta\varphi}{}^{\varphi} &= \frac{\cos\theta}{\sin\theta}. \end{aligned} \tag{7.18}$$

De Riccitensor kan nu in deze coördinaten worden berekend, met als uitkomst

$$R_{\mu\nu} = \begin{pmatrix} \frac{3\ddot{a}}{c^2 a} & 0 \\ & \\ 0 & -\frac{1}{c^2}\, g^*_{ij} \left[a\ddot{a} + 2\dot{a}^2 + 2kc^2 \right] \end{pmatrix}. \tag{7.19}$$

De Riemannscalar wordt dan gegeven door de uitdrukking

$$R = -\frac{6}{c^2}\left(\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 + \frac{kc^2}{a^2}\right). \tag{7.20}$$

Door de laatste twee vergelijkingen te combineren vinden we nu de Einsteintensor

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}\,g_{\mu\nu}R = \begin{pmatrix} -\dfrac{3}{c^2}\left[\left(\dfrac{\dot{a}}{a}\right)^2 + \dfrac{kc^2}{a^2}\right] & 0 \\ & \\ 0 & \dfrac{1}{c^2}\,g^*_{ij}\left[2a\ddot{a} + \dot{a}^2 + kc^2\right] \end{pmatrix}. \tag{7.21}$$

Beschouw nu een ideaal gas in rust t.o.v. dit kosmisch coördinatenstelsel. Volgens vergelijking (6.8) wordt dit gas beschreven door een energie-impulstensor

$$T_{\mu\nu} = \rho u_\mu u_\nu + p\left(g_{\mu\nu} + \frac{u_\mu u_\nu}{c^2}\right),$$

waarin $u_\mu = (-c, 0, 0, 0)$. In matrixnotatie wordt dit

$$T_{\mu\nu} = \begin{pmatrix} \rho c^2 & 0 \\ & \\ 0 & pa^2 g^*_{ij} \end{pmatrix}. \tag{7.22}$$

Vergelijking van de uitdrukkingen (7.21) en (7.22) laat zien dat $G_{\mu\nu}$ en $T_{\mu\nu}$ evenredig zijn, in overeenstemming met de inhomogene Einsteinvergelijkingen (4.10). Deze kunnen we dan gebruiken om verband te leggen tussen de dichtheid ρ en de druk p enerzijds, en de schaalfactor a anderzijds:

$$\begin{aligned} \rho &= \frac{3}{8\pi G}\left[\left(\frac{\dot{a}}{a}\right)^2 + \frac{kc^2}{a^2}\right], \\ p &= -\frac{c^2}{8\pi G}\left[2\,\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 + \frac{kc^2}{a^2}\right]. \end{aligned} \tag{7.23}$$

Door differentatie van de eerste vergelijking krijgen we de tijdafgeleide van de dichtheid:

$$\dot{\rho} = \frac{3}{4\pi G}\left[\frac{\dot{a}\ddot{a}}{a^2} - \left(\frac{\dot{a}}{a}\right)^3 - \frac{kc^2\dot{a}}{a^3}\right]. \tag{7.24}$$

Uit deze vergelijking en de uitdrukking voor p in (7.23) kunnen we nu de tweede afgeleide $\ddot{a}$ elimineren; het elegantst is dit te doen door op te merken dat

$$\frac{d}{dt}\left(\rho c^2 a^3\right) + p\,\frac{da^3}{dt} = \dot{\rho}c^2a^3 + 3\left(\rho c^2 + p\right)a^2\dot{a} = 0. \tag{7.25}$$

Deze vergelijking heeft een eenvoudige fysische interpretatie: ze houdt in, dat voor een gas in thermisch evenwicht de uitdijing van het heelal een adiabatische proces is. Dit is als volgt in te zien. Zij V_0 een referentievolume op de kosmische tijd $t = t_0$ waarop $a_0 = a(t_0) = 1$. Het volume van dit gebied op tijd t is dan

$$V(t) \,=\, a^3(t)V_0. \tag{7.26}$$

De energiedichtheid van het gas is $\varepsilon(t) = \rho(t)c^2$, dus is de totale energie van het gas in het volume $V(t)$:

$$U \,=\, \varepsilon V \,=\, \rho c^2 a^3 V_0. \tag{7.27}$$

Vergelijking (7.25) laat zich dan na vermenigvuldiging met V_0 omschrijven tot

$$T\frac{dS}{dt} \,=\, \frac{dU}{dt} + p\,\frac{dV}{dt} \,=\, 0. \tag{7.28}$$

Het eerste deel van deze vergelijking, die het thermodynamische verband tussen entropie, energie en volume geeft, bevat impliciet de aaname van thermisch evenwicht. Het tweede deel volgt uit de dynamische vergelijking (7.25) voor een ideaal gas in een homogeen en isotroop heelal.

7.3.2 De toestandsvergelijking

Om een volledige beschrijving van een homogeen en isotroop heelal te krijgen moeten we nu nog de toestandsvergelijking specificeren, die het verband legt tussen de dichtheid ρ en de druk p. We gaan ervan uit, dat de energiedichtheid van het heelal bepaald wordt door een combinatie van drie componenten:
- hete, relativistische materie en straling;
- koude, niet-relativistische materie;
- vacuüm energie (een kosmologische constante).

Dan wordt het algemene verband tussen de toestandsvariabelen (ρ, p) gegeven door vgl.(6.16):

$$p \,=\, \eta\rho c^2.$$

Zoals uiteengezet in hoofdstuk 6, heeft voor deze drie componenten afzonderlijk de evenredigheidsparameter η de waarde

$$\eta \,=\, \begin{cases} \frac{1}{3}, & \text{relativistische materie;} \\ 0, & \text{niet-relativistische materie;} \\ -1, & \text{vacuümenergie.} \end{cases} \tag{7.29}$$

Door substitutie van de toestandsvergelijking in de adiabatische evolutieconditie (7.25) vinden we dan een verband tussen druk en volume, en tussen dichtheid en volume, gegeven door

$$pa^{3(1+\eta)} \,=\, constant \quad \Leftrightarrow \quad \rho a^{3(1+\eta)} \,=\, constant. \tag{7.30}$$

Na vermenigvuldiging met de overeenkomstige macht van het referentievolume V_0 kunnen we dit ook schrijven als

$$pV^{(1+\eta)} = constant \quad \Leftrightarrow \quad \rho V^{(1+\eta)} = constant. \tag{7.31}$$

Uitgesplitst naar de verschillende gevallen wordt dit

$$\begin{array}{lll} \rho a^3 = constant, & p = 0, & \text{koude, niet-relativistische materie;} \\ \rho a^4 = constant, & p = \frac{1}{3}\,\rho c^2, & \text{hete, relativistische materie en straling;} \\ \rho \,=\, constant & p = \varepsilon_\Lambda, & \text{vacuümenergie.} \\ \quad = \dfrac{\Lambda c^2}{32\pi G} \,=\, -\dfrac{\varepsilon_\Lambda}{c^2}, & & \end{array} \tag{7.32}$$

Aangezien een relativistisch gas tijdens uitdijing afkoelt en niet-relativistisch kan worden —en omgekeerd bij compressie— is in het algemene geval η niet constant, maar schaalafhankelijk. We kunnen dan bij voorbeeld schrijven:

$$\eta(a) \,=\, \frac{1}{3}\,\frac{\bar{a}^2}{a^2+\bar{a}^2}. \tag{7.33}$$

Dit is relativistisch voor $0 < a \ll \bar{a}$ en niet-relativistisch voor $a \gg \bar{a}$. Vgl.(7.25) wordt dan

$$\frac{d(\rho a^3)}{da} + \eta(a)\rho\,\frac{da^3}{da} \,=\, 0, \tag{7.34}$$

en in geintegreerde vorm

$$\int_{\bar{\rho}}^{\rho} \frac{d\rho}{\rho} \,=\, -3\int_{\bar{a}}^{a} (1+\eta(a))\,\frac{da}{a}. \tag{7.35}$$

De oplossing is

$$\frac{\rho a^4}{\sqrt{a^2+\bar{a}^2}} \,=\, \frac{\bar{\rho}\bar{a}^3}{\sqrt{2}}. \tag{7.36}$$

In de limiet $a \to 0$ geeft dit het relativistische geval en in de limiet $a \to \infty$ het niet-relativistische geval; $a = \bar{a}$ is de schaal waarbij de overgang tussen de twee regimes plaats heeft. Het speciale geval van straling geassociëerd met massaloze deeltjes (zoals fotonen) correspondeert met de limiet $\bar{a} \to \infty$ en $\bar{\rho} \to 0$ op zo'n manier dat $\bar{\rho}\bar{a}^4$ eindig blijft.

7.4 De dynamische ontwikkeling van het heelal

Met behulp van de toestandvergelijking kunnen de dichtheid en druk als functie van de schaalparameter a worden bepaald, onder de aanname van thermisch evenwicht. Uit $\rho(a)$ en de eerste Einsteinvergelijking (7.23) volgt het gedrag van de schaalparameter als functie van de tijd:

$$\left(\frac{\dot{a}}{a}\right)^2 + \frac{kc^2}{a^2} = \frac{8\pi G}{3}\,\rho(a). \tag{7.37}$$

Deze vergelijking bepaalt de dynamische ontwikkeling van het heelal voor alle drie de klassen van modellen: ruimtelijk hyperbolisch ($k = -1$), vlak ($k = 0$) of sferisch ($k = +1$). Beschouwen we hyperbolische ruimten, dan is

$$\frac{1}{c^2}\,\dot{a}^2 = 1 + \frac{8\pi G}{3c^2}\,a^2\rho(a). \tag{7.38}$$

Uit vgl.(7.36) zien we, dat asymptotisch de tweede term naar nul gaat voor grote waarden van a, zodat de krommingsterm, met $k = -1$, altijd overheerst over de materiedichtheid. Bijgevolg wordt de uitdijing in afwezigheid van een kosmologische constante gedreven door de meetkunde van de ruimte (de kromming) en gaat altijd door: $a(t) \sim ct$. De dichtheid wordt daarbij steeds kleiner: $\rho \sim 1/a^3$, in overeenstemming met de verwachting dat het heelal door niet-relativistische materie wordt gedomineerd.

Voor een sferische ruimte met $k = +1$ is daarentegen

$$\frac{1}{c^2}\,\dot{a}^2 + 1 = \frac{8\pi G}{3c^2}\,a^2\rho(a). \tag{7.39}$$

Hieruit blijkt, dat de materiedichtheid niet kleiner kan worden dan

$$\rho(a_m) = \frac{3c^2}{8\pi G}\,\frac{1}{a_m^2}, \tag{7.40}$$

waarbij $a = a_m$ op het moment van maximale expansie, wanneer $\dot{a} = 0$. Als er geen kosmologische constante meespeelt, kunnen we door vergelijking met de uitdrukking (7.36) zien voor welk domein van begincondities, impliciet vastgelegd door $\bar{a}$ en $\bar{\rho}$, deze maximumschaal in het niet-relativistische regime ligt:

$$a_m \geq \bar{a} \quad \Leftrightarrow \quad \frac{8\pi G}{3c^2}\,\bar{\rho}\bar{a}^2 \geq 1. \tag{7.41}$$

Indien er wel een kosmologische constante is, dan verandert de eis voor de maximale waarde van de schaalparameter a_m in

$$\rho_{materie}(a_m) = \frac{3c^2}{8\pi G}\,\frac{1}{a_m^2} + \frac{1}{c^2}\,\varepsilon_\Lambda = \frac{3c^2}{8\pi G}\left(\frac{1}{a_m^2} - \frac{\Lambda}{12}\right), \tag{7.42}$$

waarbij het linkerlid van deze vergelijking staat voor de dichtheid van materie en straling. Aangezien deze altijd positief moet zijn, volgt hieruit een limiet op de kosmologische constante: $\Lambda < 12/a_m^2$.

Tenslotte beschouwen we het geval van een vlakke ruimte, $k = 0$. In dit geval wordt de kosmologische evolutievergelijking

$$\frac{\dot{a}^2}{a^2 \rho(a)} = \frac{8\pi G}{3}. \tag{7.43}$$

We beschouwen eerst een heelal waarvan de dichtheid evolueert volgens vergelijking (7.36). In termen van de dimensieloze parameter $\alpha \equiv a/\bar{a}$ wordt de evolutie dan gegeven door:

$$\frac{\alpha^2 \dot{\alpha}^2}{\sqrt{\alpha^2 + 1}} = \frac{8\pi G \bar{\rho}}{3\sqrt{2}}. \tag{7.44}$$

Deze vergelijking heeft de oplossing

$$\alpha^2 + 1 = 2 \left[1 + \sqrt{\frac{3\pi G \bar{\rho}}{2}} \, (t - \bar{t}) \right]^{\frac{4}{3}}, \tag{7.45}$$

met $\bar{t}$ het tijdstip waarop de overgang van het relativistische naar het niet-relativistische regime plaats heeft: op $t = \bar{t}$ is $a = \bar{a}$, dus $\alpha = 1$. In het bijzonder is voor grote waarden van de schaalparameter a op late tijden $t \gg \bar{t}$:

$$a(t) \approx \left(3\pi \sqrt{2} G \bar{\rho} \right)^{\frac{1}{3}} \bar{a} \, t^{\frac{2}{3}}, \qquad t \gg \bar{t}. \tag{7.46}$$

In het relativistische regime is er een singulariteit $a(t) = 0$ (de 'oerknal') voor

$$t = t_\infty = \bar{t} - \frac{2^{\frac{3}{4}} - 1}{\sqrt{3\pi \sqrt{2} G \bar{\rho}}}. \tag{7.47}$$

De tijdafhankelijke tweede term binnen de haken in het rechterlid van vergelijking (7.45) is daarom altijd kleiner dan de eerste, en we kunnen in het relativistische regime de evolutie van de schaalparameter tot op eerste orde benaderen met

$$a(t) \approx \bar{a} - \sqrt{\frac{8\pi G \bar{\rho}}{3}} \, \bar{a} \, (\bar{t} - t), \qquad t < \bar{t}. \tag{7.48}$$

Voor tijden vlak na de singulariteit krijgen we een betere benadering door $t = t_\infty + \delta t$ te schrijven, wat leidt tot

$$\alpha^2 + 1 = \left[1 + \sqrt{3\pi \sqrt{2} G \bar{\rho}} \, \delta t \right]^{\frac{4}{3}}. \tag{7.49}$$

Voor kleine waarden van δt geeft dit:

$$a^2(t) \approx 2^{\frac{3}{4}} \bar{a}^2 \sqrt{\frac{8\pi G \bar{\rho}}{3}} \, \delta t, \qquad \delta t \ll \bar{t} - t_\infty, \tag{7.50}$$

zodat de schaalparameter in het vroege heelal groeit met de wortel uit de tijd: $a \sim \sqrt{\delta t}$.

Tenslotte beschouwen we het geval van een vlakke ruimte met een kosmologische constante:

$$\rho = \rho_0 = -\frac{\Lambda c^2}{32\pi G}.$$

De evolutievergelijking (7.43) wordt dan

$$\frac{\dot{a}^2}{a^2} = -\frac{\Lambda c^2}{12} \geq 0. \tag{7.51}$$

We zien dat alleen een negatieve kosmologische constante mogelijk is. De oplossing van deze vergelijking is eenvoudig en beschrijft exponentiële expansie of contractie:

$$a(t) = a(0)\, e^{Ht}, \qquad H = \pm c\sqrt{\frac{|\Lambda|}{12}}. \tag{7.52}$$

Om een uitdijend heelal te beschrijven, in overeenstemming met de waarnemingen van astronomen, moeten we hier natuurlijk de positieve wortel kiezen.

7.5 De kosmologische parameters

Om de dynamica van een isotroop en homogeen heelal zo te beschrijven dat de theorie op een eenvoudige wijze met de waarnemingen vergeleken kan worden, is het gebruikelijk een drietal parameters in te voeren die de kromming en uitdijing impliciet vastleggen. De eerste van deze parameters is de Hubbleparameter $H(t)$, gedefiniëerd door

$$H = \frac{\dot{a}}{a} = \frac{d\log a}{dt}. \tag{7.53}$$

De Hubbleparameter beschrijft dus de fractionele verandering in de schaalparameter per tijdseenheid.

We merken op dat de notatie in vergelijking (7.52) hiermee geheel in overeenstemming is: een constante waarde van H komt overeen met een exponentiëel uitdijend heelal:

$$H = constant \quad \Rightarrow \quad a(t) = a(0)\, e^{Ht}. \tag{7.54}$$

Dit is alleen een oplossing van de kosmologische vergelijkingen voor een vlak heelal: $k = 0$. Meer algemeen bepaalt vgl.(7.43) voor een vlak heelal een speciaal functioneel verband tussen de dichtheid ρ en de schaalparameter a, geheel in termen van de Hubbleparameter:

$$\rho(t) = \rho_c(t) = \frac{3H^2(t)}{8\pi G}. \tag{7.55}$$

Deze vorm van de dichtheid als functie van de tijd wordt de *kritische dichtheid* genoemd. Uit vgl.(7.46) volgt dat in afwezigheid van een kosmologische constante $\dot{a}(t) \sim 1/t^{1/3}$, zodat de uitdijingssnelheid in dit geval asymptotisch naar nul gaat voor late tijden. Dit geldt dan dus ook voor de kritische dichtheid.

De kritische dichtheid wordt alleen gerealiseerd in een vlak heelal, maar verschaft een nuttige referentiewaarde in alle gevallen, als grenswaarde tussen een open ($k = -1$) en een gesloten ($k = +1$) heelal. Dit is nog duidelijker in termen van de dimensieloze dichtheidsparameter

$$\Omega(t) = \frac{\rho(t)}{\rho_c(t)}. \tag{7.56}$$

Met behulp van deze grootheid kunnen we de eerste Einsteinvergelijking (7.23) voor de evolutie van een FRW-heelal herschrijven in de volgende vorm:

$$\Omega = 1 + \frac{kc^2}{a^2 H^2}. \tag{7.57}$$

We zien onmiddellijk dat voor een vlak heelal $\Omega = 1$ op alle tijden, terwijl voor een hyperbolisch (open) heelal $\Omega(t) < 1$, en voor een sferisch (gesloten) heelal $\Omega(t) > 1$.

Het blijkt niet eenvoudig om uit waarnemingen van de uitdijing van het heelal de huidige waarde van deze parameters te bepalen. De beste waarde van de Hubbleparameter op dit moment is:

$$H_0 = h_0 \times 100\,\mathrm{km/sec/Mpc}, \qquad h_0 = 0.7 \pm 0.07. \tag{7.58}$$

Voor de kritische dichtheid nu geldt dan een beste waarde

$$\rho_{c0} = 5.3 \left(\frac{h_0}{0.7} \right)^2 \mathrm{GeV/m^3c^2}. \tag{7.59}$$

Over de huidige waarde van Ω valt bijna niets met zekerheid te zeggen, aangezien deze b.v. ook afhangt van de hoeveelheid donkere materie in het heelal, waaronder mogelijk zwarte gaten, massieve neutrino's of schaduwmaterie. Er is alleen een vrij harde ondergrens:

$$\Omega_0 > 0.1. \tag{7.60}$$

Wel kunnen we eenvoudig de bijdrage van de microgolfachtergrondstraling of van een achtergrond van kosmische neutrino's aan de dichtheid bepalen. De energiedichtheid van een fotongas wordt gegeven door de wet van Stefan-Boltzmann:

$$\rho_\gamma = \alpha T_\gamma^4 \approx 0.26 \times 10^6\,\mathrm{eV/m^3c^2}. \tag{7.61}$$

Hierin is

$$\alpha = \frac{\pi^2 k^2}{15\hbar^3 c^3} = 7.56 \times 10^{-16}\,\mathrm{J\ m^{-3}\ K^{-4}}, \tag{7.62}$$

terwijl we als temperatuur $T_\gamma = 2.7\,\mathrm{K}$ hebben genomen. Voor relativistische fermionen zoals neutrino's geldt een Fermi-Dirac versie van dezelfde wet:

$$\rho_\nu = \frac{7}{8}\,\alpha T_\nu^4 \approx 0.6 \times 10^5\,\mathrm{eV/m^3c^2}. \tag{7.63}$$

Zoals we later nog zullen afleiden, nemen we hier voor het bepalen van de numerieke waarde aan dat $T_\nu \approx 2\,\mathrm{K}$. De huidige bijdrage aan de dimensieloze dichtheidsparameter is in deze gevallen dan

$$\Omega_{\gamma 0} = 0.5 \times 10^{-4}, \tag{7.64}$$

voor fotonen met een temperatuur van 2.7 K, en

$$\Omega_{\nu 0} = 0.1 \times 10^{-4}, \tag{7.65}$$

voor een enkele soort neutrino's met een temperatuur van 2 K. Een totale achtergrondstraling van fotonen en drie soorten neutrino's zou dan tesamen een bijdrage leveren ter grootte van

$$\Omega_{(\gamma+3\nu)\,0} \approx 0.8 \times 10^{-4}. \tag{7.66}$$

Dit is vrijwel verwaarloosbaar ten opzichte van de huidige dichtheid van baryonische materie in de vorm van protonen en neutronen, die de belangrijkste bijdrage aan de ondergrens (7.60) levert.

7.6 De roodverschuiving

De eerste aanwijzingen voor de uitdijing van het heelal kwamen uit de waarnemingen van de roodverschuiving van spectraallijnen van vergelegen melkwegstelsels door Hubble in de jaren 1920-30. Deze lieten zien dat alle sterrenstelsels buiten onze lokale groep van ons af bewegen, met een snelheid die gemiddeld groter is naarmate de afstand groter is. Hubble vond een bij benadering lineair statistisch verband tussen de afstand en de roodverschuiving, waaruit hij de waarde van de naar hem genoemde parameter H kon bepalen in het huidige kosmische tijdperk.

Uit ons homogeen en isotroop heelalmodel zijn Hubble's waarnemingen en analyse eenvoudig te begrijpen. Daartoe definiëren we eerst de roodverschuivingsparameter z als de relatieve verandering in de golflengte van het licht tussen het moment van emissie, met golflengte λ_e, en het moment van waarneming, met golflengte λ_0:

$$z = \frac{\lambda_0 - \lambda_e}{\lambda_e} = \frac{\lambda_0}{\lambda_e} - 1. \tag{7.67}$$

Als de golflengte meeverandert met de uitdijing van het heelal: $\lambda(t) \sim a(t)$, en als dit de enige bron van roodverschuiving is, dan volgt hieruit dadelijk dat

$$\frac{a_0}{a_e} = \frac{\lambda_0}{\lambda_e} = 1 + z. \tag{7.68}$$

Hierbij wordt dus geen rekening gehouden met een eventuele eigenbeweging van de bron. Statistisch is dit geen probleem: gemiddeld zijn de sterrenstelsels in rust t.o.v. het kosmisch coördinatensysteem. Uiteraard geldt dit niet voor specifieke stelsels, waarvan de eigenbeweging wel degelijk van belang kan zijn.

Door het maken van een Taylorreeksontwikkeling van $a(t)$ voor tijden niet al te verschillend van de huidige, krijgen we

$$a(t) \,=\, a_0 \,+\, \dot{a}_0\,(t - t_0) \,+\, ... = \, a_0 \,\left(1 + H_0(t - t_0) + ...\right); \qquad (7.69)$$

De index '0' geeft aan dat de waarde op t_0 genomen wordt. In combinatie met vergl. (7.68) vinden we dan, dat in eerste benadering

$$1 \,+\, H_0\,(t_e - t_0) \,=\, \frac{1}{1+z} \,\approx\, 1 - z. \qquad (7.70)$$

Als de sterrenstelsels niet te ver weg staan, dan mag de verandering in de afstand d gedurende de tijd dat het signaal onderweg is in eerste instantie verwaarloosd worden, en vinden we uit $d = c(t_0 - t_e)$ dat

$$z \,=\, \frac{H_0 d}{c}. \qquad (7.71)$$

H_0 wordt in deze benadering dus bepaald door de evenredigheidsfactor in de grafiek van de roodverschuiving z versus de afstand d, of beter nog de daarmee corresponderende kosmische tijd $\Delta t \,=\, d/c$ dat het signaal onderweg is. Voor zeer verre melkwegstelsels, waarvan het licht er zo'n lange tijd over doet om ons te bereiken dat de uitdijing onderweg niet meer verwaarloosbaar is, moeten we uiteraard hogere orde termen in de verschillende reeksontwikkelingen meenemen. We krijgen dan niet meer een eenvoudig lineair verband tussen z en d, maar een lijn met positieve of negatieve kromming, ervan afhankelijk of het heelal steeds sneller of steeds langzamer uitdijt.

In de analyse van Hubble's oorspronkelijke waarnemingen speelden deze overwegingen geen rol, omdat alle door hem bestudeerde sterrenstelsels verhoudingsgewijs dichtbij ons staan. Het grootste probleem bij het bepalen van H_0 is dan het schatten van de afstanden tot diverse kosmische objecten. In de loop van een eeuw zijn de schattingen voor de afstanden tot verre melkwegstelsels door nieuwe gegevens en methoden te gebruiken bijna tien keer groter geworden. In overeenstemming daarmee is de huidige waarde van Hubble's constante H_0 nu bijna tien keer zo klein als de oorspronkelijk door Hubble zelf opgegeven waarde.

Op basis van de Hubblerelatie (7.71) (of een nauwkeuriger vorm daarvan) worden kosmische afstanden en/of tijden vaak uitgedrukt in termen van een roodverschuiving z (voor een bepaalde frequentie). Deze grootheid is in de praktijk beter te bepalen en minder gevoelig voor systematische fouten dan de direkte (gelijktijdige) afstand zelf, met uitzondering van afwijkingen door de invloed van de eigen beweging.

7.7 De thermische ontwikkeling van het heelal

De ontwikkelingsgang van het heelal, zoals we die hierboven hebben geschetst, gaat van een klein compact heelal gevuld met dichte relativistische materie naar een groot heelal, voor zover niet leeg voornamelijk gevuld met koude niet-relativistische materie. Deze ontwikkeling gaat relatief langzaam: de vrije weglengte van de straling en relativistische deeltjes in het vroege heelal was klein vergeleken met de karakteristieke afmetingen van het heelal, en de relaxatietijden die nodig waren om thermisch evenwicht tot stand te brengen over het algemeen veel kleiner dan de tijdschaal waarop de uitdijing van het heelal plaats vond. Het heelal was daarom gedurende de langste tijd van zijn geschiedenis in thermisch evenwicht, en we kunnen het heelal op ieder moment in zijn evolutie karakteriseren door de temperatuur te geven van de materie en straling die op dat ogenblik aanwezig waren en met elkaar in wisselwerking stonden.

In zowel het relativistische als het niet-relativistische regime is de relatie tussen de dichtheid, de temperatuur en de schaalfactor gegeven door

$$\rho a^5 T \, = \, \text{constant}. \tag{7.72}$$

In het relativistische geval zijn namelijk ρa^4 en aT beide constant, terwijl in het niet-relativistische regime zowel ρa^3 als $a^2 T$ constant zijn. De relaties tussen ρ en a volgen rechtstreeks uit vergelijking (7.32). Het verband tussen a en T volgt in het relativistische geval uit de wet van Stefan-Boltzmann, die we hebben gebruikt bij het afleiden van vergl.(7.61) en (7.63):

$$\rho \, = \, \alpha T^4, \tag{7.73}$$

waarbij de evenredigheidsconstante α verschillend is voor Bose-gassen (zoals fotonen) en Fermi-gassen (zoals neutrino's):

$$\begin{aligned} \alpha_{Bose} \, = \, \frac{8}{7}\, \alpha_{Fermi} \, &= \, 0.76 \times 10^{-15}\ \text{J m}^{-3}\ \text{K}^{-4}/c^2 \\ &= \, 4.74 \times 10^{3}\ \text{eV m}^{-3}\ \text{K}^{-4}/c^2. \end{aligned} \tag{7.74}$$

Uit (7.73) en $\rho a^4 =$ constant volgt dan meteen dat $a \sim T^{-1}$.

In het niet-relativistische geval zijn de energie en druk van een ideaal gas gegeven door (6.17) en (6.19):

$$U = Nmc^2 + \frac{3}{2} N k_B T, \qquad pV = N k_B T.$$

Hieruit volgt met $V = a^3 V_0$ dat

$$0 \, = \, \frac{dU}{dt} + p \frac{dV}{dt} \, = \, \frac{3}{2} N k_B \, \frac{1}{a^2} \, \frac{d}{dt} \left(a^2 T \right), \tag{7.75}$$

en dus is a^2T constant.

Met behulp van de relatie (7.36) kan vergelijking (7.72) door te kwadrateren worden omgevormd tot een universele relatie tussen a en T, die zowel geldig is in het relativistische als het niet-relativistische regime:

$$\frac{a^2T^2}{\eta(a)} = \text{constant.} \tag{7.76}$$

Uiteraard geldt deze relatie zoals altijd alleen binnen het model waarin de toestandsvergelijking van de straling en materie kan worden beschreven met de universele functie $\eta(a)$: $p = \eta(a)\rho c^2$.

Het is opmerkelijk, dat een temperatuur die dezelfde schaalwetten volgt ook kan worden toegekend aan materie en straling die ontkoppeld is van de rest van het heelal. Een gas van deeltjes in het heelal is ontkoppeld, als de vrije weglengte van de deeltjes langer is dan de karakteristieke schaal van het heelal, d.w.z. de gemiddelde tijd tussen twee interacties van een deeltje is langer dan de leeftijd van het heelal. Dit geldt b.v. voor de 2.7 K microgolf-achtergrondstraling, waarvan de fotonen statistisch gesproken geen wisselwerking meer hebben gehad met de materie in het heelal sinds het laatste plasma van vrije geladen deeltjes verdween, bij de vorming van neutraal waterstof uit vrije elektronen en protonen. Zo'n ontkoppeld gas is dus niet meer in thermisch contact met de rest van de straling en materie in het heelal.

Het thermisch schaalgedrag voor een ontkoppeld gas volgt in het relativistische geval onmiddellijk uit de Bose-Einstein en Fermi-Dirac-verdelingen. Op het moment van de ontkoppeling van een gas wordt de bezettingsgraad van de verschillende toestanden immers door een van deze evenwichtsverdelingen beschreven:

$$\langle n(\nu)\rangle = \frac{1}{e^{\frac{h\nu}{kT}} \pm 1}. \tag{7.77}$$

Het enige wat in deze distributie verandert tijdens de ontwikkeling van het heelal, is dat het bezettingsgetal voor een mode met frequentie ν_0 bij ontkoppeling op een later tijdstip hoort bij een roodverschoven frequentie ν_1, gegeven door

$$\nu_1 = \frac{a_0}{a_1}\,\nu_0. \tag{7.78}$$

We krijgen dan voor de bezettingsgraad

$$\langle n(\nu_1)\rangle_{t_1} = \langle n(\nu_0)\rangle_{t_0} = \frac{1}{e^{\frac{h\nu_0}{kT_0}} \pm 1} = \frac{1}{e^{\frac{h\nu_1}{kT_1}} \pm 1}, \tag{7.79}$$

waarin de effectieve 'roodverschoven' temperatuur gegeven wordt door

$$T_1 = \frac{a_0}{a_1}\,T_0. \tag{7.80}$$

Dit betekent dat relativistische gassen die in thermisch evenwicht zijn bij ontkoppeling, ook daarna een evenwichtsverdeling blijven volgen met een temperatuur die zich aanpast aan de schaal volgens $a_1 T_1 = a_0 T_0 =$constant.

Voor niet-relativistische gassen in thermisch evenwicht op het moment van ontkoppeling is de Maxwell-Boltzmann-verdeling van toepassing:

$$\langle n(v) \rangle \,=\, C\, e^{\frac{mv^2}{2kT}}. \tag{7.81}$$

De effectieve snelheid van deeltjes tijdens de uitdijing van het heelal is omgekeerd evenredig met de schaalfactor: $a_1 v_1 = a_0 v_0$; om een de afstand tussen twee gegeven meebewegende punten van de ruimte af te leggen is immers meer tijd nodig als het heelal intussen uitdijt. Dan volgt uit de initiële Maxwell-Boltzmann-verdeling dat op een later tijdstip na ontkoppeling het gemiddeld aantal deeltjes met snelheid v gegeven wordt door een aangepaste Maxwell-Boltzmann-verdeling met een effectieve temperatuur

$$T_1 \,=\, \frac{a_0^2}{a_1^2}\, T_0, \tag{7.82}$$

zodat in dit geval $a^2 T =$constant. In alle gevallen geldt, dat de thermische verdeling na ontkoppeling niet wordt gehandhaafd door interactie met de rest van het heelal (dat dus niet fungeert als warmtebad), maar uitsluitend een gevolg is van de thermische begintoestand en de isotropie en homogeniteit van het kosmische gravitatieveld.

Toepassingen

1. De neutrino-achtergrond

Naast de gravitatie, die verantwoordelijk is voor de uitdijing van het heelal, zijn er drie soorten fundamentele wisselwerkingen tussen atomaire en subatomaire deeltjes; deze spelen ook een rol in de kosmologie. De elektromagnetische wisselwerking zorgde voor thermische evenwicht tussen fotonen en elektrisch geladen deeltjes tot het moment van ontkoppelen van de CMBR; de microgolfachtergrond is eigenlijk een restgas van fotonen, overgebleven na de omzetting van het plasma van vrije elektronen en protonen in neutraal waterstof.

Extrapolerend naar eerdere perioden in de geschiedenis van het heelal, komen we bij energiedichtheden en temperaturen waarin ook de zwakke wisselwerkingen —die een rol spelen in β-verval en neutrinoverstrooiing— en de kleurkrachten tussen quarks en gluonen effectief geweest moeten zijn bij het tot stand brengen van een thermisch evenwicht.

De temperaturen waarbij deze krachten relevant worden zijn van de orde van 1 MeV voor de zwakke wisselwerkingen, en 100 MeV voor de kleurkrachten. Bij temperaturen boven 1 MeV kunnen zwakke wisselwerkingen een thermische verdeling van nucleonen (protonen en neutronen), leptonen (elektronen en neutrino's) en antileptonen in stand houden, via de reacties

$$p^+ + e^- \leftrightarrow n + \nu, \qquad n + \bar{e}^+ \leftrightarrow p^+ + \bar{\nu}, \qquad n \to p^+ + e^- + \bar{\nu}. \tag{7.83}$$

Daarnaast is er in principe nog de reactie $e^- + \bar{e}^+ \leftrightarrow \nu + \bar{\nu}$, maar deze is bij de energieën in het MeV-bereik sterk onderdrukt. Toen het heelal afkoelde tot lagere temperaturen verliepen deze zwakke interacties te langzaam om de aantallen deeltjes in evenwicht te houden: de gemiddelde tijd tussen twee zwakke wisselwerkingsprocessen werd van dezelfde orde van grootte als de leeftijd van het heelal, en effectief werden de eerste twee reacties (7.83) stopgezet. Het gevolg daarvan was dat alleen het β-verval van de neutronen nog doorging, zodat de verhouding van aantallen neutronen tot protonen geleidelijk afnam. De oorspronkelijk in thermisch evenwicht geproduceerde concentraties leptonen (elektronen, positronen, neutrino's en antineutrino's) raakten daarmee ontkoppeld van de concentraties van de baryonen (protonen en neutronen). De elektronen en positronen konden door hun elektromagnetische wisselwerkingen in thermisch evenwicht met fotonen blijven, zolang de temperatuur meer dan een paar honderd keV was. Daarna werden ze door foto-annihilatieprocessen $e^-\bar{e}^+ \to \gamma\gamma$ geëlimineerd tot er slechts een klein aantal elektronen over was, nodig om het ladingsevenwicht met de protonen in stand te houden.

De overgebleven protonen en elektronen bleven nog aanwezig als een plasma van vrije geladen deeltjes in wisselwerking met de elektromagnetische straling, tot de temperatuur bereikt werd waarbij stabiel waterstof gevormd kon worden: ongeveer 1 eV.

Daarnaast was en is er volgens dit scenario echter nog altijd een restgas van ontkoppelde neutrino's, dat tot op heden echter nog niet is gedetecteerd. Aangezien deze neutrino's een thermische verdeling hadden tot het ogenblik waarop de zwakke wisselwerkingen niet langer effectief waren, volgt uit onze eerdere beschouwingen dat zij nu nog een thermische verdeling zouden moeten volgen, zij het bij de lagere temperatuur corresponderend met een groter heelal.

Deze temperatuur is echter niet hetzelfde als die van de microgolfachtergrondstraling. De reden daarvoor is, dat bij de annihilatie van de $e^-\bar{e}^+$-paren in fotonparen ($\gamma\gamma$) er extra fotonen bijkomen. Daardoor wordt het fotongas opgewarmd, en de fotontemperatuur stijgt t.o.v. die van het oorspronkelijke foton-elektron-positron plasma. Aangezien dit dezelfde temperatuur had als het neutrinogas, is de verwachting daarom dat de temperatuur van de CMBR wat hoger zal zijn dan die van de neutrino-achtegrond.

We kunnen dit scenario in een quantitatieve vorm gieten en een voorspelling doen voor de neutrinotemperatuur, door gebruik te maken van vergelijkingen (7.25) en (7.28) die ons vertellen dat de uitdijing van het heelal een adiabatisch proces is. Daarbij kunnen we gebruikmaken van de omstandigheid dat bij de temperaturen die we beschouwen alle relevante deeltjes relativistisch zijn. Dan geldt de toestandsvergelijking $p = \varepsilon/3 = \rho c^2/3$, en de totale entropie van een gas van zulke deeltjes in wisselwerking is

$$S = \frac{U + pV}{T} = \frac{4}{3}\frac{U}{T}. \tag{7.84}$$

Tevens geldt voor de thermische energiedichtheid de uitgebreide wet van Stefan-Boltzmann:

$$\varepsilon = \frac{U}{V} = \left(n_B + \frac{7}{8}\,n_F\right)\alpha\, T^4, \tag{7.85}$$

met $n_{(B,F)}$ respectievelijk het aantal soorten bosonische en fermionische deeltjes per volumeëenheid. Door het combineren van (7.84) en (7.85) vinden we

$$S = \frac{4}{3}\left(n_B + \frac{7}{8}\,n_F\right)\alpha\,(aT)^3\,. \tag{7.86}$$

De adiabatische ontwikkeling van het heelal leidt er nu toe, dat de entropie voor en na het verdwijnen van de $e^-\bar{e}^+$-paren hetzelfde moet zijn. We stellen daarom een entropiebalans op, als volgt. *Voor* de annihilatie is er een plasma van fotonen, elektronen en positronen in evenwicht. Daarnaast is er een gas van ontkoppelde neutrino's, met op dat ogenblik nog dezelfde temperatuur $T_\nu = T_\gamma$ als het plasma. *Na* de annihilatie van elektronen en positronen blijft er alleen een fotongas over, met een temperatuur T'_γ, plus de neutrino's. Aangezien de neutrino's ontkoppeld zijn is hun bijdrage aan de entropie voor en na de $e\bar{e}$-annihilatie hetzelfde, en mag deze in de balans worden genegeerd.

De balans ziet er dan zo uit, dat voor annihilatie 1 soort bosonen (γ's) en 2 soorten fermionen (e en $\bar{e}$) aan de entropie bijdragen. Dus

$$S_{voor} = \frac{4}{3}\left(1 + \frac{14}{8}\right)\alpha\,(aT_\gamma)^3\,. \tag{7.87}$$

Na de annihilatie zijn er alleen fotonen, dus $n_B = 1$ en $n_F = 0$; dan is

$$S_{na} = \frac{4}{3}\,\alpha\left(aT'_\gamma\right)^3\,. \tag{7.88}$$

Door deze entropieën aan elkaar gelijk te stellen: $S_{voor} = S_{na}$, krijgen we

$$T'_\gamma = \left(\frac{11}{4}\right)^{1/3} T_\gamma, \tag{7.89}$$

met T_γ de temperatuur die de fotonen zouden hebben als er geen annihilatie had plaats gehad; deze temperatuur is dezelfde als die van de neutrino's. De conclusie is dan, dat de neutrino's een temperatuur zullen hebben gegeven door

$$T_\nu = \left(\frac{4}{11}\right)^{1/3} T'_\gamma. \tag{7.90}$$

Als we voor T'_γ de gemeten waarde van 2.73 K gebruiken, vinden we voor de huidige temperatuur van het neutrinogas $T_\nu = 1.95$ K. Deze getallen liggen ten grondslag aan de schattingen (7.61) en (7.63) van de energiedichtheid van fotonen en neutrino's in het huidige heelal.

2. De ontkoppeling van de achtergrondstraling

Waterstof wordt permanent geioniseerd door interacties met vrije fotonen indien de gemiddelde dichtheid en energie van de fotonen overeenkomen met een temperatuur van $T_{ontk} = 3.7 \times 10^3\,\mathrm{K}$ (een gemiddelde fotonenergie van ~ 0.3 eV). De huidige temperatuur van de microgolfachtergrondstraling (CMBR) is 2.7 K. Aangezien voor ontkoppelde relativistische straling geldt dat $aT =$ constant, was de waarde van $a(t)$ ten tijde van de ontkoppeling

$$\frac{a_{ontk}}{a_0} = \frac{T_0}{T_{ontk}} = 0.73 \times 10^{-3}. \tag{7.91}$$

Volgens vergelijking (7.68) komt de tijd van ontkoppeling dan overeen met een roodverschuiving van $z \approx 1350$.

Uit dit resultaat kunnen we nu bepalen wat de relatieve energiedichtheid in straling en materie was ten tijde van de ontkoppeling van de CMBR, in termen van de huidige verhouding. Namelijk, voor fotonen en neutrino's geldt dat

$$\rho_{r\,ontk} = \frac{a_0^4}{a_{ontk}^4}\,\rho_{r\,0} = z_{ontk}^4\,\rho_{r\,0}. \tag{7.92}$$

Anderzijds geldt voor de niet-relativistische materie (die bij een temperatuur van 3700 kelvin, ongeveer 0.3 eV, nog steeds niet-relativistisch is) dat

$$\rho_{m\,ontk} = \frac{a_0^3}{a_{ontk}^3}\,\rho_{m\,0} = z_{ontk}^3\,\rho_{m\,0}. \tag{7.93}$$

Door deze uitkomsten te combineren vinden we, dat de verhouding van de relatieve bijdragen van straling en koude materie zich ontwikkeld heeft volgens

$$\left(\frac{\Omega_m}{\Omega_r}\right)_0 = \left(\frac{\Omega_m}{\Omega_r}\right)_{ontk} z_{ontk}. \tag{7.94}$$

De relatieve bijdrage van de straling aan de totale dichtheid in verhouding tot die van de materie was ten tijde van de ontkoppeling dus 1350 maal groter dan tegenwoordig. We kunnen dit ook als volgt interpreteren: voor fotonen plus 3 neutrino's (met de standaardaannamen voor de huidige temperatuur) is

$$\Omega_{r\,0}\,z_{ontk} = \Omega_{(\gamma+3\nu)\,0}\,z_{ontk} \approx 0.11. \tag{7.95}$$

Dus als de huidige dichtheid van koude materie $\Omega_{m\,0} > 0.11$, dan werd ten tijde van de ontkoppeling van de CMBR de dichtheid van het heelal al gedomineerd door niet-relativistische materie: $\Omega_{m\,ontk} > \Omega_{r\,ontk}$. Gezien de beste schattingen voor de koude materie nu: $0.1 < \Omega_{m\,0} < 0.3$, is dit met tamelijk grote zekerheid het geval.

Met dit gegeven kunnen we ook een schatting maken van de leeftijd van het heelal; op zijn minst kunnen we er een ondergrens aan stellen. Het uitgangspunt is, dat het heelal voor de ontkoppeling van de microgolfachtergrond effectief vlak was: $\bar{\Omega} = 1$, ofwel $\bar{\rho} = \bar{\rho}_c$. We kunnen dan de evolutievergelijking (7.47) gebruiken:

$$t_\infty = \bar{t} - \frac{2^{\frac{3}{4}} - 1}{\sqrt{3\pi\sqrt{2}G\bar{\rho}}},$$

met $\bar{t}$ en $\bar{\rho}$ de tijd en de dichtheid bij de overgang van een stralings- naar materiegedomineerd heelal. Nu is de energiedichtheid gegeven door vergl.(7.85), zodat

$$\bar{\rho}c^2 = \left(n_B + \frac{7}{8} n_F \right) \alpha \bar{T}^4. \tag{7.96}$$

De overgang naar een materiegedomineerd heelal vond plaats niet al te lang voor de ontkoppeling van de CMBR. Zoals we al hebben vastgesteld werd het heelal op dat moment gedomineerd door fotonen, elektronen en positronen; het effectieve aantal vrijheidsgraden bedroeg dus

$$n_B + \frac{7}{8} n_F = \frac{11}{4}. \tag{7.97}$$

Voor de tijd verstreken tussen de singulariteit (de *oerknal*) en de overgang naar het materie-gedomineerd heelal vinden we dan

$$\begin{aligned} \bar{t} - t_\infty &= \frac{\left(1 - 2^{-3/4}\right)}{\bar{T}^2} \sqrt{\frac{8c^2}{33\pi G\alpha}} \\ &\approx \frac{5 \times 10^5}{\left(\bar{T}/10^4\,\mathrm{K}\right)^2} \text{ jr.} \end{aligned} \tag{7.98}$$

Daar $\bar{T} \geq 10^4$ K, kan het heelal volgens deze berekening niet ouder zijn geweest dan een half miljoen jaar ten tijde van de overgang naar het materiegedomineerde regime. Het grootste deel van de ontwikkeling vond dus plaats na deze overgang en na de ontkoppeling van de CMBR. We kunnen nu (7.46) en $\bar{z} = a_0/\bar{a} = 1350$ gebruiken om te berekenen hoeveel tijd is verstreken na de overgang[2]:

$$t_0 = \frac{\bar{z}^{3/2}}{2^{3/4}\bar{T}^2} \sqrt{\frac{8c^2}{33\pi G\alpha}} \approx \frac{15}{\left(\bar{T}/10^4\,\mathrm{K}\right)^2} \text{ Gjr.} \tag{7.99}$$

[2]We gebruiken de eenheid 1 Gjr = 10^9 jaar.

De enige aanname die we in deze berekening hebben gestopt, is dat het heelal bij de overgang van stralings- naar materiegedomineerd regime effectief vlak was: $\bar{\Omega} = 1$. Deze aanname steunt op het gedrag van Ω als functie van de schaalparamter a volgens vergelijking (7.57):

$$\Omega = 1 + \frac{kc^2}{a^2 H^2} = 1 + \frac{kc^2}{\dot{a}^2}. \tag{7.100}$$

Merk eerst op dat de huidige waarde van Ω niet zo heel veel van 1 afwijkt; alle data wijzen, ruim genomen, op $0.1 < \Omega < 5$, ofwel $\Omega = \mathcal{O}(1)$; dan is ook $kc^2/\dot{a}^2 \leq \mathcal{O}(1)$. In vrijwel alle heelalmodellen groeide het heelal in de allervroegste beginfase echter veel sneller dan nu: $\dot{\bar{a}}^2 > \dot{a}_0^2 \Leftrightarrow \bar{a}\bar{H} > a_0 H_0$. Het gevolg daarvan is, dat

$$\frac{kc^2}{\bar{a}^2 \bar{H}^2} \ll \mathcal{O}(1). \tag{7.101}$$

Voor een vlak heelal met $k = 0$ is $\Omega = 1$ natuurlijk een identiteit. Voor $k = \pm 1$ kunnen we de ongelijkheid (7.101) kwantificeren; daartoe herschrijven we eerst vergelijking (7.100) in de vorm

$$\begin{aligned} \frac{\Omega}{\Omega - 1} &= 1 + \frac{a^2 H^2}{kc^2} = \frac{a^2 H^2}{kc^2}\,\Omega = \frac{8\pi G}{3kc^2}\,\rho a^2 \\ &= \frac{1}{\sqrt{2}}\,\frac{\bar{\Omega}}{\bar{\Omega} - 1}\,\frac{\bar{a}^2}{a^2}\sqrt{1 + \frac{a^2}{\bar{a}^2}}. \end{aligned} \tag{7.102}$$

$\bar{\Omega}$ is de dichtheidsparameter bij de overgang van het stralingsgedomineerde naar het massagedomineerde regime, en voor de laatste gelijkheid hebben we gebruik gemaakt van de relatie (7.36). In termen van de roodverschuiving is $\bar{a}/a = (1 + z)/(1 + \bar{z})$, zodat

$$1 - \frac{1}{\Omega} = \sqrt{2}\left(1 - \frac{1}{\bar{\Omega}}\right)\frac{1 + \bar{z}}{1 + z}\,\frac{1}{\sqrt{1 + (1 + z)^2/(1 + \bar{z})^2}}. \tag{7.103}$$

Vullen we voor de dichtheidsparameter de huidige waarde Ω_0 bij $z = 0$ in, dan vinden we

$$1 - \frac{1}{\bar{\Omega}} \approx \frac{1}{\bar{z}\sqrt{2}}\left(1 - \frac{1}{\Omega_0}\right). \tag{7.104}$$

waarin $\bar{z} + 1 \approx \bar{z} \geq z_{ontk} \approx 1350$, de waarde bij de ontkoppeling van de CMBR. Met $\Omega_0 \geq 0.1$ is dan $\left|\bar{\Omega} - 1\right| \leq 0.5 \times 10^{-3}$. Voor eerdere tijden wordt het verschil $\Omega - 1$ nog veel kleiner. Dat in het vroege heelal $\Omega = 1$ dus een zeer goede benadering is, houdt in dat het heelal vlak is (de bijdrage van k is verwaarloosbaar, dan wel $k = 0$), en dat de groei van de schaalfactor in het verleden aanzienlijk groter was dan nu.

3. De vorming van de lichte elementen

De theorie van de evolutie van de kosmos, zoals we die hier schetsen, is een toepassing van de algemene relativiteitstheorie die we kunnen toetsen aan een aantal astrofysische waarnemingen. Daaronder vallen in de eerste plaats de uitdijing van het heelal, zoals gemeten door Hubble en zijn opvolgers, en de isotropie en homogeniteit, vooral ook van het vroege heelal, zoals blijkt uit de microgolfachtergrondstraling. Deze tests betreffen direct de waargenomen grootschalige structuur van het heelal in ruimte en tijd. Andere waarnemingen vullen het beeld van de evolutie van het heelal op meer indirecte manier aan, en leveren daarmee bijkomend bewijsmateriaal voor de kosmologische toepassingen van de ART. Een van deze aanvullende waarnemingen betreft de verhouding van de hoeveelheden waterstof en helium oorspronkelijk gevormd in het vroege heelal; deze leverden de brandstof voor de eerste sterren, die pas veel later gevormd werden en waarin de vorming van elementen zwaarder dan helium verder ging.

In deze paragraaf beschrijven we de redenering, die de verwachte verhouding van de hoeveelheden oorspronkelijk waterstof en helium levert. Daar de gedetailleerde berekeningen gecompliceerd zijn en alleen numeriek uitvoerbaar, zullen we deze achterwege laten en alleen de resultaten geven.

We hebben al eerder vastgesteld dat bij temperaturen beneden 1 MeV de zwakke wisselwerkingen niet meer in staat zijn om het thermisch evenwicht tussen protonen, neutronen en leptonen te handhaven. Beneden een paar honderd keV zijn er ook geen elektron-positronparen meer. Het heelal is dan slechts gevuld met fotonen en een zeer ijl gas van nucleonen en elektronen met netto ladingsdichheid gelijk aan nul (we verwaarlozen de neutrinoachtergrond; deze speelt hier geen rol).

Bij temperaturen rond 1 MeV zijn nucleonen niet-relativistisch, en wordt de verhouding van aantallen neutronen en protonen bepaalt door de Boltzmannfactor $e^{-\Delta mc^2/kT}$, met $\Delta m = m_n - m_p = 1.3$ MeV/c^2 het massaverschil tussen een neutron en een proton. De zwakke interacties die het thermisch evenwicht handhaven houden op effectief te zijn bij een temperatuur beneden $T_1 \approx 0.8$ MeV. Op het moment dat dit gebeurt is de verhouding van aantallen neutronen en protonen per volumeëenheid dus

$$\left. \frac{n_n}{n_p} \right|_{T_1} = e^{-1.3/0.8} = 0.2. \tag{7.105}$$

De vrije neutronen zijn niet stabiel onder β-verval; ze zouden geheel uit het heelal verdwenen zijn, als ze niet stabiele gebonden toestanden met protonen hadden kunnen vormen, in de eerste plaats deuteriumkernen: $p + n \to d + \gamma$. Dit kan echter pas op grote schaal gebeuren bij een temperatuur van $T_2 \sim 0.06$ MeV. Gebruikmakend van de relatie $a_2 T_2 = a_1 T_1$, vinden we dan voor de toename van de schaal van het heelal $a_2/a_1 = T_1/T_2 \approx 13$. Wanneer we deze temperaturen en schaalfactoren omrekenen naar tijden:

$$\frac{t_2}{t_1} = \frac{a_2^2}{a_1^2} = \frac{T_1^2}{T_2^2},$$

dan verlopen er tot het moment dat de deuteriumvorming plaats heeft ongeveer 2 3/4 minuut = 165 seconden. Bij een levensduur van 887 seconden voor vrije neutronen, is de fractie overgebleven neutronen dan

$$\frac{n_{n\,2}}{n_{n\,1}} = e^{-165/887} = 0.84. \tag{7.106}$$

Hieruit bepalen we de nieuwe verhouding van neutronen tot protonen als volgt:

$$\frac{n_{n\,2}}{n_{p\,2}} = \frac{n_{n\,1} - \delta n_n}{n_{p\,1} + \delta n_n} \approx \frac{n_{n\,1}}{n_{p\,1}} - 2\,\frac{\delta n}{n_{p\,1}} = \frac{n_{n\,1}}{n_{p\,1}} \left(1 - 2\,\frac{\delta n_n}{n_{n\,1}} \right) \approx 0.13. \tag{7.107}$$

Het deuterium dat de overgebleven neutronen heeft gebonden wordt via reacties zoals $d{+}d \rightarrow \alpha$ vrijwel gelijktijdig omgezet in ^{4}He (α-deeltjes); ^{4}He heeft namelijk een veel grotere bindingsenergie dan deuterium, en is dus stabieler. Daarnaast wordt ook ^{3}He gevormd in de reactie $d + n \rightarrow {}^3$He, maar dit wordt ten dele weer omgezet in ^{4}He door invangen van een neutron: ^{3}He + $n \rightarrow \alpha$. Slechts een kleine fractie ^{3}He blijft over. Het is een daarom een zeer goede eerste orde benadering om te veronderstellen dat alle neutronen eindigen in α-deeltjes (^{4}He).

De massa van een α-deeltje is ongeveer gelijk aan $m_\alpha = 4m_p$. Voor de fractie van de massa van baryonen die voorkomt in de vorm van ^{4}He vinden we dan bij de temperatuur T_2 en daarbeneden:

$$\frac{\rho_{He}}{\rho_p + \rho_n} = \frac{4n_\alpha}{n_p + n_n} = \frac{4n_n/2}{n_p + n_n} = 2\,\frac{n_n}{n_p}\,\frac{1}{1 + n_n/n_p} \approx 0.24. \tag{7.108}$$

Dit is in zeer goede overeenstemming met de waargenomen fractie van vrij (niet in sterren gevormd) helium in het heelal.

Soortgelijke berekeningen voor de veel kleinere hoeveelheden deuterium d en ^{3}He die in het vroege heelal zijn gevormd leiden ook tot waarden die overeenkomen met de waarnemingen, voor zover deze met enige betrouwbaarheid kunnen worden verricht. De vorming van lichte elementen blijkt dus uitstekend te verklaren binnen het raamwerk van het kosmologisch model van een homogeen en isotroop uitdijend heelal dat de ART ons levert.

7.8 Inflatie

De voorgaande beschouwingen leren ons, dat de Einsteinvergelijkingen via de FRW-oplossingen een goede beschrijving van de structuur en ontwikkeling van het heelal op grote schaal leveren. Grote schaal betekent in deze context: de schaal waarop het kosmologisch beginsel van isotropie en homogeniteit van toepassing is; het is niet vanzelfsprekend dat zo'n schaal bestaat, maar empirisch blijkt dit zo te zijn. Toch blijven er een aantal vragen over die binnen het tot nu toe beschreven raamwerk nog niet beantwoord zijn. Daartoe behoren drie principekwesties, die bekend staan als het probleem van de vlakheid, het horizonprobleem en het probleem van de structuurvorming. We zullen deze drie vragen een voor een kort bespreken.

1. Op grond van de schattingen voor de huidige waarde van de dichtheidsparameter Ω_0 en vergelijking (7.100) en volgende, hebben we de conclusie getrokken dat in het vroege heelal $|\Omega - 1| \ll 1$. Dit betekent of dat het heelal ruimtelijk vlak is ($k = 0$), of dat de expansiesnelheid in het verleden zo groot was (veel groter dan de lichtsnelheid in de Minkowskiruimte c) dat de ruimtelijke kromming niet bijdroeg aan de uitdijing:

$$\left| \frac{kc^2}{a^2 H^2} \right| = \left| \frac{kc^2}{\dot{a}^2} \right| \ll 1 \quad \Rightarrow \quad \left(\frac{\dot{a}}{a} \right)^2 = H^2 \approx \frac{8\pi G}{3}\, \rho(a). \tag{7.109}$$

De laatste vergelijking, equivalent met (7.43), volgt uit (7.37) onder verwaarlozing van de term met de ruimtelijke kromming k. Het is hierbij van belang te benadrukken, dat de expansiesnelheid $\dot{a}$ geen fysische beweging representeert, maar een uniforme verandering van alle schalen. Een uitbreidingssnelheid groter dan c is dus niet in strijd met de causaliteit van fysische processen: velden en deeltjes kunnen geen afstanden overbruggen met snelheden groter dan de werkelijke lichtsnelheid.

Het vlakheidsprobleem in de kosmologie is de vraag of, en zo ja hoe, de speciale begintoestand $\Omega = 1$ aannemelijk gemaakt kan worden: is er een mechanisme dat het heelal vanzelf naar de kritische dichtheid drijft vanwaaruit het kan beginnen zich naar de huidige toestand te ontwikkelen?

2. Behalve dat het vroege heelal zeer vlak was, was het ook zeer homogeen en isotroop; dit volgt in het bijzonder uit de waarnemingen van de microgolfachtergrondstraling, waarvan het spectrum zuiver thermisch is met fluctuaties in de temperatuurverdeling niet groter dan $\Delta T/T \leq 10^{-5}$. Zo'n homogene verdeling wijst op thermisch contact tussen alle gebieden vanwaar de straling afkomstig is. Zulk thermisch contact lijkt echter onwaarschijnlijk: op het moment dat de CMBR ontstond was het heelal nog zo jong, dat de afstand tussen gebieden die aan de hemel meer dan 10° van elkaar verwijderd zijn niet te overbruggen was door velden of deeltjes die met de lichtsnelheid of langzamer reizen.

Meer algemeen: als er een eindige tijd verstreken is sinds het ontstaan van het heelal, dan eist het causaliteitsbeginsel dat de invloed van fysische verschijnselen zich slechts tot een eindige afstand uitstrekt. Deze afstand wordt bepaald door de lichtkegel met als oorsprong het ogenblik en de plaats van de gebeurtenis die het verschijnsel heeft voortgebracht. Omgekeerd kunnen we de *horizon* van een waarnemer definiëren als het geheel van punten op de grootst mogelijke afstand vanwaar fysische verschijnselen hem kunnen bereiken: zijn achterwaartse lichtkegel tot de begintijd. Het probleem met de achtergrondstraling is, dat de homogeniteit wijst op thermisch evenwicht tussen gebieden die voor de ontkoppeling van de CMBR buiten elkaars horizon lagen. Vandaar dat dit vraagstuk het *horizonprobleem* genoemd wordt.

3. In tegenstelling tot de microgolfachtergrond is de verdeling van materie in het huidige heelal niet homogeen op afstandschalen beneden 200 Mpc. Melkwegstelsels en clusters van melkwegstelsels zijn ontstaan door samentrekking van gaswolken onder invloed van hun eigen zwaartekracht. Daartoe zijn dan wel kleine inhomogeniteiten nodig in de oorspronkelijke verdeling van niet-relativistische materie zoals waterstof en helium. Echter, de buitengewoon homogene verdeling van de microgolfachtergrond legt ook sterke beperkingen op aan de mate van inhomogeniteit van de materieverdeling ten tijde van de ontkoppeling van de CMBR, die toen immers nog met de materie in wisselwerking was. Het blijkt dat deze beperkingen het moeilijk maken om de vorming van melkwegstelsels en clusters zoals we die zien te kunnen verklaren. Dit probleem staat bekend als het probleem van de structuurvorming, een probleem dat zich voordoet juist ten gevolge van de homogene en vlakke begincondities van het heelal.

Om een mogelijke verklaring te geven voor de eerste twee moeilijkheden met het standaard oerknal-scenario gaan we opnieuw terug naar vergelijking (7.100). Met behulp van deze vergelijking zien we, dat in het huidige tijdperk waarin de uitdijing van het heelal steeds langzamer gaat ($\dot{a}$ wordt steeds kleiner), de dichtheid van het heelal weg zal lopen van de kritische waarde $\Omega = 1$, tenzij het heelal vlak is ($k = 0$). Is er echter een fase van het heelal waarin de uitdijingssnelheid toeneemt (waarin $\dot{a}$ steeds groter wordt), dan zal de dichtheid juist naar de kritische waarde worden toegedreven: de krommingsterm $|kc^2/\dot{a}^2|$ wordt dan steeds kleiner. Zo'n mogelijke fase van versnelde uitdijing wordt een fase van *inflatie* van het heelal genoemd. Een natuurlijke verklaring voor een schier kritische dichtheid van het heelal in het verre verleden zou dus kunnen zijn, dat er een fase van inflatie aan vooraf is gegaan.

Deze hypothese zou dan ook tegelijk het horizonprobleem kunnen oplossen. Tijdens een voldoend lange periode van inflatie kan namelijk een gebiedje in het vroege heelal dat volledig causaal samenhangt makkelijk worden opgeblazen tot de grootte van het zichtbare heelal ten tijde van de ontkoppeling van de achtergrondstraling. Op die manier is themisch evenwicht tussen gebieden in alle richtingen van het zichtbare heelal mogelijk.

In het scenario van inflatie van het vroege heelal blijft dan alleen het probleem van de structuurvorming over. Dit probleem heeft twee kanten; enerzijds is de vraag hoe uit de waargenomen fluctuaties in de CMBR het ontstaan van melkwegstelsels verklaard kan worden. Anderzijds ligt er de vraag hoe de fluctuaties in de CMBR zelf ontstaan zijn. Beide problemen zijn op dit moment geheel open. Recente precisiemetingen van de amplitude en het spectrum van de fluctuaties leggen de beginvoorwaarden voor de vorming van melkwegstelsels en clusters met grote nauwkeurigheid vast. Door simulaties kan men dan proberen met welk model van de samenstelling en verdeling van materie de tegenwoordig waargenomen structuur van het heelal het best verklaard kan worden.

Het ontstaan van de fluctuaties in de microgolfachtergrond hangt samen met de manier waarop het heelal de inflatiefase beëindigt. Tijdens de inflatie worden alle bestaande homogeniteiten in de verdeling van straling en materie zo goed als uitgewist, maar het aflopen van de inflatie gaat in de meest gangbare modellen gepaard met een substantiële opwarming van het heelal en de productie van een grote hoeveelheid entropie. Zowel de samenstelling van de straling en materie als de thermodynamische processen die plaats vinden bepalen dan de uiteindelijke verdeling van inhomogeniteiten in de CMBR.

We beëindigen dit hoofdstuk nu met bespreking van enige simpele kwantitatieve modellen voor inflatie. Ter inleiding nemen we eerst het begrip *afstand* in het heelal wat nauwkeuriger onder de loep. De afstand tussen twee punten $\mathcal{P}_1$ en $\mathcal{P}_2$ op tijdstip t wordt vastgelegd door de waarden van de coördinaten van de punten en de metriek (7.13). *De afstand op tijd t* betekent hier: de 3-dimensionale afstand in het hypervlak t = constant. Nu kunnen we de homogeniteit en isotropie van de ruimte gebruiken om de oorsprong van het coördinatenstelsel (r, θ, φ) in $\mathcal{P}_1$ te leggen. De afstand tot $\mathcal{P}_2$ wordt dan gemeten langs de radiële kromme $(r, 0, 0)$, met r lopend van $r(\mathcal{P}_1) = 0$ tot $r(\mathcal{P}_2) = R$. Een infinitesimale radiële afstand is $ds = a(t)dr/\sqrt{1 - kr^2}$, waaruit voor eindige afstanden d volgt dat

$$d(t) \,=\, a(t) \int_0^R \frac{dr}{\sqrt{1 - kr^2}} \,\equiv\, a(t)\, \Delta(R). \tag{7.110}$$

De fysische afstand is dus de gereduceerde coördinatenafstand $\Delta(R)$ in het meebewegende stelsel, vermenigvuldigd met de schaalfactor ten tijde t. Het is niet moeilijk de gereduceerde coördinatenafstand $\Delta(R)$ voor de verschillende waarden van k uit te rekenen:

$$\Delta(R) \,=\, \begin{cases} \arcsin R, & k = +1; \\ R, & k = 0; \\ \operatorname{arcsinh} R, & k = -1. \end{cases} \tag{7.111}$$

Volgens vergelijking (7.110) neemt de afstand tussen twee punten die beide in rust zijn t.o.v. het meebewegende kosmologische coördinatensysteem dus op dezelfde wijze toe als de schaalparameter $a(t)$ zelf:

$$d(t_2) \,=\, d(t_1)\, e^{\int_{t_1}^{t_2} H(\tau) d\tau}. \tag{7.112}$$

In differentiaalvorm wordt dit:

$$\dot{d} = Hd, \tag{7.113}$$

waaruit blijkt dat de karakteristieke tijdschaal voor veranderingen in de afstand tussen 'vaste' punten in het meebewegend coördinatensysteem (punten waarvoor $\Delta(R) = \text{constant}$) de Hubbletijd $1/H$ is. We vinden dus de wet van Hubble terug: de snelheid waarmee de punten uit elkaar gaan is evenredig met de afstand; in het bijzonder is de snelheid $\dot{d} > c$ wanneer $d > c/H$, de Hubble-afstand. Punten in het heelal die verder dan een Hubble-afstand van elkaar liggen, verwijderen zich dus —gemeten in termen van de kosmologische tijdcoördinaat t— van elkaar met een snelheid groter dan de lichtsnelheid in de lege Minkowski-ruimte. Dit betekent echter niet, dat er fysische verschijnselen zijn die zich door de uitdijende ruimte sneller voortplanten dan het licht zelf (mits dat licht niet gehinderd wordt door verstrooiing aan materie).

Om de causale relatie tussen punten te kunnen analyseren, construeren we vervolgens de lichtkegel behorend bij een punt $\mathcal{P}_1$ op tijd t_1. Voor een fysisch verschijnsel dat zich met de lichtsnelheid in radiële richting uitbreidt, geldt dat

$$ds^2 = -c^2dt^2 + a^2(t)\,\frac{dr^2}{1-kr^2} = 0. \tag{7.114}$$

Het tijdinterval $t_2 - t_1$ dat nodig is om de gereduceerde coördinatenafstand $\Delta(R)$ tussen de punten $\mathcal{P}_1$ en $\mathcal{P}_2$ af te leggen met de snelheid van het licht wordt dan gegeven door de relatie

$$\Delta(R) = \int_0^{R(t_1,t_2)} \frac{dr}{\sqrt{1-kr^2}} = \int_{t_1}^{t_2} \frac{cdt}{a(t)}. \tag{7.115}$$

De fysieke afstand tussen $\mathcal{P}_1$ en $\mathcal{P}_2$ op tijdstip t_2 is dus

$$d_{12}(t_2) = a(t_2)\,\Delta(R) = a(t_2)\int_{t_1}^{t_2} \frac{cdt}{a(t)}. \tag{7.116}$$

Alle punten $\mathcal{P}_2$ die op tijdstip t_2 deze afstand tot $\mathcal{P}_1$ hebben, vormen tesamen de lichtkegel behorend bij $\mathcal{P}_1$ op tijd t_1. Omdat het heelal tussen t_1 en t_2 uitdijt, is de lichtkegel groter dan het geval zou zijn zonder uitdijing. De lichtkegel, d.w.z. de afstand d_{12}, groeit dus sneller dan de lichtsnelheid c in het Minkowski-vacuum[3]:

$$\dot{d}_{12}(t_2) = c + H(t_2)d_{12}(t_2). \tag{7.117}$$

Tenslotte kunnen we nu ook het begrip *horizon* een nauwkeurige inhoud geven. Verschijnselen die zich na tijdstip t_1 voordoen, kunnen op of voor het tijdstip t_2 alleen waargenomen worden door waarnemers binnen de lichtkegel met straal $d_{12}(t_2)$. Omgekeerd zijn voor een waarnemer op tijdstip t_2 alle verschijnselen

[3]We zien dus dat in deze coördinaten de lichtsnelheid c niet de snelheid van het licht is.

die zich na t_1 afspelen op afstanden groter dan $d_{12}(t_2)$ onzichtbaar. We noemen de punten op afstand $d_{12}(t_2)$ daarom de horizon voor verschijnselen die zich na t_1 hebben voorgedaan. In het bijzonder geldt natuurlijk, dat als het heelal een beginpunt in de tijd heeft (een oerknal op t_∞), er een absolute horizon is met straal

$$d_\infty(t) = a(t)\,\Delta(R_\infty), \qquad R_\infty = R(t, t_\infty). \tag{7.118}$$

Voorbij deze afstand valt er dan niets waar te nemen.

Om het probleem van de vlakheid en het horizonprobleem nader te analyseren, bestuderen we nu hoe de lichtkegel zich uitbreidt als functie van de tijd voor verschillende modellen van de evolutie van de schaalfactor $a(t)$. Beschouw bij voorbeeld een machtwet van de vorm

$$a(t) = \lambda t^n \quad \rightarrow \quad \dot{a}(t) = n\lambda t^{n-1}. \tag{7.119}$$

Voor $n \neq 1$ kunnen we de evenredigheidsparameter schrijven als $\lambda = c\tau^{1-n}$, waarin τ een karakteristieke tijd is voor de evolutie van de schaalfactor. Een machtwet van deze vorm geldt b.v. voor een uitdijend vlak heelal gedomineerd door koude materie met $n = 2/3$ (zie vergl. 7.50), en voor een heelal gedomineerd door straling met $n = 1/2$ (vergl. 7.46). Een statisch heelal correspondeert daarentegen met $n = 0$, terwijl voor $n < 0$ het heelal krimpt. Merk op, dat voor $n < 1$ de snelheid van de uitdijing $\dot{a}(t)$ afneemt, dat voor $n = 1$ deze constant is (stationaire uitdijing), terwijl voor $n > 1$ de uitdijingssnelheid blijft groeien.

We bepalen nu de lichtkegel in gereduceerde coördinaten:

$$\Delta[R(t_2, t_1)] = \int_{t_1}^{t_2} \frac{cdt}{a(t)} = \begin{cases} \dfrac{1}{(1-n)} \left(\dfrac{t_2}{\tau}\right)^{1-n} \left[1 - \left(\dfrac{t_1}{t_2}\right)^{1-n}\right], & n < 1; \\[2ex] \dfrac{c}{\lambda} \ln \dfrac{t_2}{t_1}, & n = 1; \\[2ex] \dfrac{1}{(n-1)} \left(\dfrac{\tau}{t_1}\right)^{n-1} \left[1 - \left(\dfrac{t_1}{t_2}\right)^{n-1}\right], & n > 1. \end{cases} \tag{7.120}$$

De werkelijke straal van de lichtkegel op t_2 is dan

$$d_{12}(t_2) = a(t_2)\,\Delta(R) = \begin{cases} \dfrac{ct_2}{(1-n)} \left[1 - \left(\dfrac{t_1}{t_2}\right)^{1-n}\right], & n < 1; \\[2ex] ct_2 \ln \dfrac{t_2}{t_1}, & n = 1; \\[2ex] \dfrac{ct_1}{(n-1)} \left(\dfrac{t_2}{t_1}\right)^{n} \left[1 - \left(\dfrac{t_1}{t_2}\right)^{n-1}\right], & n > 1. \end{cases} \tag{7.121}$$

Evenzo kunnen we analoge uitdrukkingen vinden voor een exponentiëel uitdijend heelal, zoals in het geval van een vlak heelal met een kosmologische constante:

$$a(t) = a_0 e^{Ht}, \tag{7.122}$$

met H = constant, zoals in vergl. (7.52). Dan is de lichtkegel in dimensieloze gereduceerde eenheden:

$$\Delta[R(t_2, t_1)] = \frac{c}{a_0 H} e^{-Ht_1} \left(1 - e^{-H(t_2 - t_1)}\right), \tag{7.123}$$

terwijl de fysische straal wordt gegeven door

$$d_{12}(t_2) = \frac{c}{H} e^{H(t_2 - t_1)} \left(1 - e^{-H(t_2 - t_1)}\right). \tag{7.124}$$

Aan de hand van deze modellen zien we, dat verschijnselen die zich met de snelheid van het licht uitbreiden in een snel uitdijend heelal slechts een eindig gebied van naburige punten kunnen bereiken, terwijl in een langzaam uitdijend heelal ieder punt uiteindelijk met ieder ander punt kan communiceren mits er voldoende tijd ter beschikking staat. Volgt de schaalfactor bij voorbeeld de machtwet $a(t) \sim t^n$, en is $n > 1$, dan zien we uit vergelijking (7.120) dat in de asymptotische limiet de horizon in gereduceerde meebewegende coördinaten wordt gegeven door $r = R$, met R de constante waarde die impliciet is gegeven door

$$\lim_{t_2 \to \infty} \Delta[R(t_2, t_1)] = \frac{1}{(n-1)} \left(\frac{\tau}{t_1}\right)^{n-1}. \tag{7.125}$$

Voor een exponentiëel groeiend heelal nadert deze grootheid de constante waarde

$$\lim_{t_2 \to \infty} \Delta[R(t_2, t_1)] = \frac{c}{a_0 H} e^{-Ht_1}. \tag{7.126}$$

In beide gevallen kunnen punten met een radiële coördinaat $r > R(t_2, t_1)$ in het meebewegende systeem voor geen enkele tijd $t_2 < \infty$ met de oorsprong op t_1 of later communiceren. Desondanks kan de fysieke grootte van de lichtkegel $d_{12}(t_2)$ onbeperkt groeien; op deze manier kunnen we toch een zeer groot heelal krijgen uit een gebied dat geheel causaal samenhangt.

Voor een langzaam expanderend heelal, b.v. met een schaalfactor die een machtwet volgt met $n \leq 1$, ligt dit kwalitatief anders: in de limiet $t_2 \to \infty$ blijft $\Delta[R(t_2, t_1)]$ onbeperkt groeien, en daarmee ook $R(t_2, t_1)$ zelf. Wanneer we lang genoeg wachten kan dus ieder punt, met willekeurig welke waarde van de radiële coördinaat r, vanuit de oorsprong bereikt worden; namelijk zodra $R(t_2, t_1) \geq r$.

De laatste stap is nu de verschillende evolutiemodellen voor de schaalfactor in de Einsteinvergelijking (7.57) in te vullen om het gedrag van de dichtheid als functie van de tijd te bepalen voor het geval $k = \pm 1$. Als de schaalfactor de machtwet (7.119) volgt vinden we

$$\Omega(t) \;=\; 1 \;+\; \frac{kc^2}{n^2\lambda^2}\, t^{2(1-n)} \;=\; \begin{cases} 1 \pm \dfrac{1}{n^2}\left(\dfrac{t}{\tau}\right)^{2(1-n)}, & n < 1; \\[2ex] 1 \pm \dfrac{c^2}{\lambda^2} = \text{constant}, & n = 1; \\[2ex] 1 \pm \dfrac{1}{n^2}\left(\dfrac{\tau}{t}\right)^{2(n-1)}, & n > 1. \end{cases} \tag{7.127}$$

Voor een exponentiëel groeiend heelal wordt dit

$$\Omega(t) \;=\; 1 \;\pm\; \frac{c^2}{a_0^2 H^2}\, e^{-2Ht}. \tag{7.128}$$

We zien dus inderdaad, dat voor een langzaam evoluerend heelal (een machtwet met $n < 1$) de dichtheidsparameter wegloopt van de kritische waarde $\Omega = 1$: voor $k = +1$ wordt het heelal sterk superkritisch, terwijl voor $k = -1$ de dichtheid subkritisch is en door nul gaat op $t = n^{1/(1-n)}\tau$. Voor een stationair uitdijend heelal ($n = 1$) is Ω echter constant, terwijl voor een snel groeiend inflatoir heelal (een machtwet met $n > 1$ of exponentiëel met $H =$ constant) de dichtheid naar de kritische waarde gaat:

$$\lim_{t\rightarrow\infty} \Omega(t) \;=\; 1, \qquad n > 1 \quad \text{of} \quad H = \text{constant}. \tag{7.129}$$

De zo verkregen evolutiewet voor de dichtheid kan worden geinterpreteerd in termen van een toestandsvergelijking. Voor het geval van exponentiële inflatie geeft vergelijking (7.128) bij voorbeeld

$$\rho(t) \;=\; \Omega(t)\rho_c \;=\; \frac{3H^2}{8\pi G}\left(1 + \frac{kc^2}{a^2H^2}\right) \;=\; \frac{3H^2}{8\pi G}\left(1 + \frac{kc^2}{a_0^2H^2}e^{-2Ht}\right). \tag{7.130}$$

Als we voldoende lang wachten, zodat a heel groot wordt, dan wordt de dichtheid dus constant:

$$\rho \;=\; \frac{3H^2}{8\pi G}. \tag{7.131}$$

Dit is de vergelijking voor een de Sitter-heelal, een leeg heelal met een kosmologische constante $\Lambda = 12H^2/c^2$ zoals in vergelijking (7.51).

Inflatie met een machtwet geeft daarentegen

$$H(t) \;=\; \frac{n}{t} \;=\; \frac{n}{\tau}\left(\frac{c\tau}{a}\right)^{1/n}. \tag{7.132}$$

In de late fase van de inflatie waarin de krommingsterm verwaarloosd mag worden vinden we dan

$$\rho(a) = \frac{3n^2}{8\pi G\tau^2}\left(\frac{c\tau}{a}\right)^{2/n}, \qquad n > 1. \tag{7.133}$$

Deze toestandsvergelijking is dus van de vorm

$$\rho a^{2/n} = \text{constant}, \tag{7.134}$$

hetgeen na de identificatie

$$3(1+\eta) = \frac{2}{n}, \tag{7.135}$$

terugvoert naar de vergelijking (6.16):

$$p = \eta\rho c^2, \qquad \text{met } \eta = \frac{2}{3n} - 1 < -\frac{1}{3}. \tag{7.136}$$

Het is mogelijk zulke toestandsvergelijkingen te verkrijgen, door modellen te construeren met een vacuümenergie afkomstig van een tijdsafhankelijk scalair veld; deze vacuumenergie speelt dan een rol vergelijkbaar met die van de kosmologische constante, en wordt *quintessentie* genoemd.

Samenvatting van hoofdstuk 7

Uit observaties zijn een aantal aspecten van de grootschalige evolutie van het heelal goed bekend: (a) Het heelal heelal dijt uit, en de snelheid waarmee sterrenstelsels zich van ons verwijderen is groter naarmate ze zich verder weg bevinden. (b) Het heelal is over grote afstanden homogeen en isotroop, en ten tijde van de ontkoppeling van de kosmische microgolf-achtergrondstraling was dat in nog grotere mate het geval dan nu. (c) De verhouding van de massadichtheid van primordiaal helium en waterstof is bij goede benadering 1 : 3. Al deze feiten passen goed in het Friedmann-Robertson-Walker model van de evolutie van het heelal, waarin de kromming van de drie-dimensionale ruimte constant is. We hebben deze aanname voor de vorm van het kosmische gravitatieveld gecombineerd met de aanname van thermisch evenwicht om een beeld te kunnen schetsen van de thermische ontwikkeling van het heelal, oorspronkelijk gedomineerd door een heet plasma van relativistische deeltjes, waaronder nucleonen, elektronen, neutrino's en fotonen, tot een nogal leeg heelal gedomineerd door koude niet-relativistische materie voornamelijk bestaand uit waterstof en helium, vergezeld van enige achtergrondstraling zoals de microgolf-achtergrond, nu. Ter afsluiting hebben we een paar open problemen besproken, zoals de oorsprong van de homogeniteit en isotropie en het ontstaan van structuur in het vroege heelal.

Hoofdstuk 8

Een kritische beschouwing van de algemene relativiteitstheorie

In de zeven voorafgaande hoofdstukken hebben we een overzicht gegeven van een aantal hoofdlijnen van de theorie van de zwaartekrachtverschijnselen, zoals we die nu begrijpen. We hebben gezien dat Newtons succesvolle wet van de zwaartekracht in historisch opzicht slechts een eerste stap was op weg naar een veldentheorie in overeenstemming is met de beginselen van lokale werking en causaliteit, die ten grondslag liggen aan alle moderne opvattingen over krachtvelden en dynamica; Newtons wet is in dat perspectief niet meer en niet minder dan een goede benadering voor verschijnselen waarbij de snelheden van de lichamen die onder invloed van de zwaartekracht bewegen klein zijn in vergelijking met de lichtsnelheid, en hun afmetingen groot in vergelijking met hun Schwarzschild-straal.

Wanneer dat niet het geval is, is een beschrijving van zwaartekrachtverschijnselen in termen van een volledig relativistische veldentheorie noodzakelijk. De algemene relativiteitstheorie voldoet aan alle noodzakelijke voorwaarden, plus een aantal extra condities zoals het inbouwen van de equivalentie van zware en trage massa en daarmee samenhangende equivalentie van een constant gravitatieveld met een constante versnelling in een inertiaalstelsel. Het resultaat is een theorie waarin alle zwaartekrachtverschijnselen vertaald kunnen worden in termen van de meetkundige eigenschappen van de tijd-ruimte. A priori was er geen reden waarom dit zo zou moeten zijn; universeel aantrekkende krachten kunnen bij voorbeeld ook worden overgebracht door scalaire velden, en het is geenszins uitgesloten dat er naast het Einsteinse zwaartekrachtveld —het metrische tensorveld $g_{\mu\nu}(x)$— nog een of meer scalaire velden $\phi(x)$ in de fundamentele fysica een rol spelen, ook op kosmische schaal. Experimenten die speuren naar effecten van mogelijke nieuwe krachten worden wereldwijd op verschillende plaatsen ondernomen, en ook in de kosmologie kunnen scalaire velden een belangrijke en mogelijk noodzakelijke rol spelen, b.v. als bron van inflatie in het vroeg heelal.

We kunnen dus niet uitsluiten dat er naast de Einsteinse zwaartekracht nog andere krachtvelden nodig zijn om een sluitende verklaring te geven van alle dynamica die we in het heelal waarnemen.

Een andere richting waarin de ART zeker uitbreiding behoeft, is die van de microscopische fysica waar de quantumtheorie in het spel is. De zwaartekrachtverschijnselen die met succes door de ART beschreven worden liggen allen geheel in het domein van de macrosopische klassieke[1] fysica. De zwaartekracht is zo extreem zwak in vergelijking met andere natuurkrachten —zoals die overgedragen door het elektromagnetische veld of het kleurveld van de sterke wisselwerkingen— dat zij op atomair of subatomair niveau geen waarneembare invloed uitoefent. Je zou je daarom kunnen afvragen of de zwaartekracht niet een puur macroscopisch klassiek verschijnsel is, dat zijn oorsprong heeft in het lange-dracht residue van andere microscopische wisselwerkingen. Dit is b.v. het geval met van der Waalskrachten, die een elektromagnetische oorsprong hebben.

Het blijkt uiterst moeilijk om zo'n scenario theoretisch te onderbouwen. In de eerste plaats moet ook zo'n 'effectief zwaartekrachtveld' voldoen aan de relativistische uitgangspunten van de fysica, zowel als aan het beginsel van de equivalentie van zware en trage massa. Dit betekent in de praktijk, dat de beschrijving die de ART levert van macroscopische verschijnselen nog steeds geformuleerd moet kunnen worden in termen van het metrische tensorveld $g_{\mu\nu}$ dat voldoet aan de Einsteinvergelijkingen, alleen is dit veld nu niet meer een elementair veld, maar het effectieve lange-dracht residue van andere krachten, beschreven met b.v. scalaire en/of vectorvelden zoals in de electrodynamica. Dat effectieve macroscopische veld moet dan in de Newtonse limiet afvallen als $1/r$, en niet zoals de van der Waalskrachten met een hogere macht van r of, zoals in het Yukawamodel, met een exponent $e^{-\lambda r}$. Vele pogingen zijn in de loop der tijd ondernomen om een theoretisch model dat aan al deze eisen voldoet te construeren. Op één nogal speciale uitzondering na, waarover later meer, is geen van deze pogingen succesvol geweest; sterker, ze hebben voornamelijk tot het inzicht geleid dat er een groot aantal algemene redenen is waarom zulke pogingen bij voorbaat tot mislukken gedoemd zijn als niet aan heel speciale voorwaarden wordt voldaan.

Het ligt daarom in de rede, dat het zwaartekrachtveld een krachtveld is dat even fundamenteel is als het elektromagnetisch veld tussen geladen deeltjes, of het kleurkrachtveld dat de wisselwerkingen beschrijft van quarks en gluonen. Er blijft dan een belangrijke eigenschap die gravitatie onderscheidt van de andere bekende krachtvelden: terwijl de elektromagnetische krachten of de kleurkrachten als krachtvelden in de klassieke limiet geen enkele intrinsieke afstandschaal bevatten, en op alle afstanden even effectief zijn, bevat de ART expliciet een afstandschaal die karakteristiek is voor de theorie: de Planckschaal (5.105)

$$\ell_{Pl} = \sqrt{\frac{2Gh}{c^3}} \approx 0.57 \times 10^{-34}\,\mathrm{m}.$$

[1]Met *klassiek* wordt hier dus uitsluitend bedoeld: *niet quantum-mechanisch*.

In de quantumelektrodynamica of quantumchromodynamica wordt dit beeld in zekere mate genuanceerd, omdat daar toch een schaalparameter opduikt die verband houdt met de schaal waarop de betreffende velden van zwak naar sterk overgaan; maar bij extrapolatie van wat we over deze andere wisselwerkingen weten, vinden we dat deze zich in de buurt van de Planckschaal alle als relatief zwakke interacties gedragen. Gravitatie zelf wordt daarentegen bij dezelfde extrapolatie van de Einsteinvergelijkingen naar de Planckschaal buitengewoon sterk. Als we dus gravitatie als fundamenteel krachtveld kunnen beschouwen, dan *moet* deze bij afstanden die klein genoeg zijn een allesoverheersende rol gaan spelen. En dit is dan zeker relevant voor de kosmologie in het regime van de allereerste quantumfase van het heelal. Maar ook andere regimes van sterke gravitatievelden waarbij quantumverschijnselen een rol spelen zijn denkbaar, zoals bij de Hawkingstraling van kleine (compacte en dus hete) zwarte gaten.

De vraag die zich dan aandient is, hoe gravitatie zich gedraagt in het regime van de quantumverschijnselen, en hoe de theorie tot een quantumtheorie van de zwaartekracht uitgebreid kan worden. Het meest voor de hand ligt de veronderstelling, dat het zwaartekrachtveld net als het elektromagnetische veld of het kleurveld deeltjeseigenschappen zal vertonen; de zwaartekracht wordt in dat beeld overgebracht door gravitatiequanta of *gravitonen*, die net als een foton een energie hebben die evenredig is met de bijbehorende frequentie:

$$E_{graviton} = h\nu. \tag{8.1}$$

Een verschil tussen fotonen en gravitonen is echter, dat fotonen elkaar in eerste benadering niet zien: het foton heeft geen lading, en kan dus zelf geen fotonen uitzenden of absorberen. Gravitatievelden koppelen echter aan alles wat massa, energie of impuls bezit, inclusief het gravitatieveld zelf; dit is inherent aan het niet-lineaire karakter van de Einsteinvergelijkingen. Deze zelfkoppeling zal niet belangrijk zijn in de limiet van laag-energetische gravitonen, omdat hun zelfgravitatie zwak zal zijn; maar voor gravitonen met golflengte in de orde van de Planckschaal zal dit zeker niet meer het geval zijn. Daar zal het beeld van gravitonen dus ongetwijfeld gemodificeerd moeten worden. Er is alle reden om te denken, dat deze modificatie niet klein of geleidelijk is. Dat blijkt bij voorbeeld uit het sterk schaal-afhankelijke gedrag van de verstrooiing van gravitonen onderling en met ander deeltjes met energieën in de buurt van de Planckenergie. Zodra we quantumcorrecties op zulke verstrooiingsprocessen proberen uit te rekenen worden deze in de meeste gevallen onherroepelijk oneindig.

De vraag hoe een uitbreiding van Einsteins algemene relativiteitstheorie te geven tot een quantumtheorie die in de buurt van de Planckschaal een consistente en zinvolle beschrijving van gravitatieverschijnselen levert, houdt theoretisch fysici al meer dan een halve eeuw bezig, zonder tot een definitief antwoord te hebben geleid. Dat wil niet zeggen dat er geen vooruitgang is geboekt; het onderzoek heeft vele verrassende wiskundige en fysische inzichten opgeleverd, die als wegwijzers in het niet eerder in kaartgebrachte gebied van wiskundige modellen kunnen fungeren. Welke weg uiteindelijk de enige juiste is, kan pas blijken als deze modellen op een of andere manier getoetst kunnen worden in experimenten of aan astrofysische waarnemingen.

Een van de wegwijzers die een belangrijke splitsing in de weg markeert is een eigenschap van sommige theorieën die bekend staat als *supersymmetrie*[2]. Supersymmetrie is de eigenschap dat alle deeltjes voorkomen in combinaties van bosonen en fermionen met dezelfde interne quantumgetallen, zoals lading, isospin e.d.. Het foton, een boson met spin $s = 1$ (in eenheden $\hbar$) heeft dan een partner-fermion met spin $s = 1/2$, en het elektron met spin $s = 1/2$ heeft een boson-tegenhanger[3] met dezelfde lading maar spin $s = 0$. Ook het graviton wordt dan vergezeld van een fermion, het *gravitino*, een deeltje met spin $s = 3/2$. Deze fermion/boson symmetrie is enigszins vergelijkbaar met de symmetrie tussen deeltjes en antideeltjes, zij het dat de natuur ons vertelt dat in supersymmetrische praktijk de deeltjes en hun partners niet dezelfde massa kunnen hebben.

Alle quantumtheorieën van de zwaartekracht kunnen worden geklassificeerd als supersymmetrisch of niet-supersymmetrisch. Daarbij blijken supersymmetrische theorieën zich vrijwel altijd veel beter te gedragen dan niet-supersymmetrische. Zo kunnen supersymmetrische theorieën van de zwaartekracht, gezamenlijk bekend onder de naam *supergravitatie*[4], in een aantal gevallen eindige quantumcorrecties op de verstrooiingsprocessen opleveren, zij het niet tot op alle orde van storingsrekening. Onder niet-supersymmetrische modellen van de quantumgravitatie is de eindigheid van 1-lus pure gravitatie zonder koppeling aan de materie[5] het enige voorbeeld van zulk deels acceptabel wiskundig gedrag.

Ondanks de bemoedigende resultaten van de supersymmetrische uitbreidingen van de ART als modellen voor een quantumtheorie van de gravitatie, blijft het natuurlijk waar, dat ook gedeeltelijk consistente theorieën volledig inconsistent zijn. Supersymmetrie is misschien een belangrijke ingredient, maar het is op zich niet voldoende om een volledige welgedefiniëerde theorie mee te construeren. Een mogelijkheid zou kunnen zijn, dat we in de theorieën te gauw ophouden met nog meer wisselwerkingstermen mee te nemen. We hebben b.v. in hoofdstuk 4 laten zien, dat de Einsteinvergelijkingen wat het gravitationele deel betreft zijn af te leiden uit de actie

$$S_{tot}[g_{\mu\nu}] = -\frac{c^4}{16\pi G} \int d^4x \sqrt{-g}\, R. \tag{8.2}$$

Dit is niet de enige scalaire grootheid die we kunnen voorstellen; termen met R^2, of met produkten van drie of meer Riemanntensoren erin, gecontraheerd met

[2] J. Wess and B. Zumino, Phys. Lett. B49 (1974), 52

[3] Deze supersymmetrische partnerdeeltjes worden in de literatuur aangeduid met merkwaardige namen als *fotino* of *selektron*.

[4] S. Ferrara, D.Z. Freedman and P. van Nieuwenhuizen, Phys. Rev. D13 (1976), 3214. Zie ook: P. van Nieuwenhuizen, Phys. Rep. 68 (1981), 1989, en de daarin opgenomen literatuur.

[5] G. 't Hooft en M. Veltman, Ann. Inst. Henri Poincaré, 20 (1974), 69

inverse metriek of epsilon-symbolen, zijn wat hun invariantie onder algemene coördinatentransformaties betreft ook acceptabel. We krijgen dan veldvergelijkingen met meer dan twee afgeleiden per term, maar dat hoeft noodzakelijk tot inconsistenties te leiden.

Dit scenario is denkbaar, maar het is een doos van Pandora: als je eenmaal extra termen toelaat, is er geen stoppen meer; berekeningen in deze richting laten zien dat deze benadering van het probleem waarschijnlijk alleen kan werken als je oneindig veel termen met willekeurig veel afgeleiden meeneemt. Een van de vragen die je dan kunt stellen is, of zo'n theorie nog een echte theorie is in de zin dat zij een voorspellende waarde heeft. Zo'n theorie bevat namelijk in principe een oneindig aantal parameters: de coëfficienten van alle mogelijke invariante termen die je aan de Einsteinactie (8.2) kunt toevoegen. Toch zou het kunnen zijn, dat het vermijden van oneindige uitkomsten voor verstrooiingsamplituden en andere fysische grootheden een verband tussen al deze parameters oplevert die hun aantal zo inperkt, dat er onafhankelijk meetbare grootheden gevonden kunnen worden waarvan de waarde dan volledig vastligt. Dit levert dan een echte voorspelling die door vergelijking met een experimenteel resultaat getoetst kan worden. Zo'n theorie is dan in ieder geval falsifiëerbaar, op zijn minst in beginsel.

Zoals uit de oorspronkelijke doos van Pandora na alle plagen uiteindelijk ook nog enkele deugden tevoorschijn kwamen, zo zou ook deze benadering van de quantumgravitatie tenslotte nog wel iets waardevols kunnen opleveren. Maar alvorens we daar nog wat meer over zeggen, zullen we eerst een vierde idee bespreken om quantumgravitatietheorieën mee te verrijken. Dit is het idee van het bestaan van meer dan drie ruimtedimensies, waarvan de extra dimensies echter miscroscopische afmetingen bezitten en daarom in alledaagse verschijnselen niet waarneembaar zijn. Dit idee, lang geleden geopperd door Kaluza[6] en Klein[7], is rond 1980 opnieuw populair geworden in de context van supergravitatie. Dit idee heeft veel aantrekkelijke aspecten; niet de geringste daarvan is de automatische unificatie van gravitatie met elektromagnetisme en andere wisselwerkingen in één theorie. Het vier-dimensionale gravitatieveld en het elektromagnetische veld zijn in deze theorieën namelijk verschillende componenten van de metriek in de hoger-dimensionale ruimte waarmee men begint. Deze theorieën leveren in principe voor ieder hoger-dimensionaal veld een oneindig aantal velden in de (1+3)-dimensionale tijd-ruimte op: een enkel veld in vier of meer ruimtedimensies kan ontwikkeld worden in een Fourierreeks over de extra dimensie(s), met als Fouriercoëfficienten velden die in 3 ruimtedimensies leven; de meeste daarvan zullen echter deeltjes met een zeer grote massa produceren (van de orde van grootte van de Planckmassa), en dus normaal in waarneembare fysische verschijnselen geen rol van betekenis spelen.

Aangezien de eigenschappen van een veldentheorie, en ook van gravitatie,

[6]T. Kaluza, Verh. Preuss. Ak. der Wiss. (1921), 966

[7]O. Klein, Zeitschr. f. Physik 37 (1926), 895; Nature 118 (1926), 516

sterk afhangen van het aantal dimensies waarin ze worden geformuleerd (denk aan het zojuist genoemde grote aantal velden dat nodig is vanuit lager-dimensionaal gezichtspunt), zou een uitbreiding van het aantal dimensies van de ruimte mogelijk een bijdrage kunnen leveren aan de constructie van een quantumtheorie van de zwaartekracht. Dit vermoeden wordt gevoed door een bijzonder verband tussen het aantal dimensies en de hoeveelheid supersymmetrie die in een gravitatietheorie kan worden ingebouwd: in 1+3 tijd-ruimte dimensies zijn er maximaal 8 fermion-partners voor een graviton mogelijk, maar dit aantal neemt af naarmate de het aantal dimensies toeneemt. In 1+10 tijd-ruimte dimensies wordt tenslotte de limiet bereikt van 1 gravitino per graviton, terwijl er in meer dan 10 ruimte-dimensies geen supersymmetrie mogelijk is.

Op dit punt van onze verhandeling is het nuttig om de ideeën die we hebben besproken als uitbreidingen van de ART om tot een consistente quantumtheorie te komen, op een rijtje te zetten. Dit waren achtereenvolgens:
- een volledig andere microscopische theorie met het gravitatieveld als effectief lange-dracht verschijnsel;
- supersymmetrie;
- uitbreiding van de veldvergelijkingen met hoger-afgeleide termen;
- extra dimensies.

Het zal niet verbazen, dat een theorie die al deze elementen in zich draagt een belangrijke kandidaat zal zijn voor een quantumgravitatietheorie. Wat misschien wel verbazingwekkend is, is dat zo'n theorie inderdaad bestaat: de theorie van de supersnaren, meer algemeen bekend als *superstringtheorie*[8]. Superstringtheorie stelt de wereld voor als opgebouwd uit elementaire snaartjes (strings) die op oneindig veel manieren kunnen trillen. Iedere trillingswijze heeft in een quantummechanische context een eigen energie en impulsmoment, die we kunnen associëren met de massa en spin van een deeltje. Daarvan zijn er een aantal (het multiplet van grondtoestanden van een gesloten string) die massaloze deeltjes kunnen beschrijven zoals het foton en het graviton. Bovendien zijn de snaren supersymmetrisch: ze kunnen ook fermionische trillingstoestanden hebben met halfintegrale impulsmomenten. Tenslotte leven ze in meer dan 3 dimensies, namelijk in (1+9) tijd-ruimtedimensies; dit levert na compactificatie van 6 ruimtedimensies deeltjes op die effectief in 3 ruimtedimensies leven, met verschillende quantumgetallen zoals massa, spin en lading die afhangen van hoe de betreffende string zich in de verder onzichtbare hogere dimensies gedraagt.

Een sterk punt van het superstringscenario is, dat alles erop wijst dat de verstrooiingsamplituden van deeltjes inclusief gravitonen daarin ook bij de Planckschaal eindig zijn. Het gedrag van de lichtste deeltjes in de theorie (met massa kleiner dan de Planckmassa) is effectief te beschrijven met een gewone veldentheorie, waarschijnlijk een vorm van supergravitatie, waarbij de effecten van de

[8]Voor een overzicht van oorspronkelijke literatuur, zie: J. Schwarz, ed.: *Superstrings, the first 15 years of superstring theory*, Vols. I & II (World Scientific, 1985)

andere stringtoestanden effectief kunnen worden geabsorbeerd in het toevoegen van termen met hogere afgeleiden aan de veldvergelijkingen. Stringtheorie lijkt dus alles te hebben wat men van een quantumgravitatie theorie mogelijk zou kunnen verwachten. Wat nog ontbreekt zijn twee zeer belangrijke elementen:
- de stringtheorie is geen theorie van alleen gravitatie, maar een theorie van *alle* deeltjes en hun wisselwerkingen; op stringtheorie rust dus de last om te laten zien dat alle deeltjes en wisselwerkingen die in de natuur voorkomen ook door de theorie beschreven worden, en niet meer dan alleen deze.
- stringtheorie moet vragen kunnen beantwoorden naar de grootte en aard van waarneembare verschijnselen, b.v. de waarde van de bekende parameters zoals de fijnstructuurconstante of de massa van het elektron. Ook een voorspelling van nieuwe effecten zou het vertrouwen in stringtheorie kunnen doen toenemen. Daarbij kun je denken aan een voorspelling voor de massa en koppelingen van het gravitino, of andere nog waar te nemen nieuwe deeltjes die nodig zijn in een consistente lange-dracht limiet van de superstringtheorie. Een eerste aanzet tot een theoretische beschrijving van een waarneembaar kosmologisch verschijnsel, het spectrum van gravitatiegolven in het vroege heelal is recent gegeven[9]. Tot nu toe is er echter nog niet zoveel in deze richting bereikt, dat de superstringtheorie de status van een hypothese of een scenario al achter zich heeft kunnen laten.

Hoewel we dus nog niet zeker weten wat de volgende stap zal zijn, is het wel duidelijk dat met de ART de theorie van de zwaartekrachtverschijnselen niet af is. Met name in het regime van de quantumfysica moeten er nog nieuwe paden worden gebaand. Maar ook in het raamwerk van de ART zelf is er veel dat experimenteel niet getoetst is. Het direkt waarnemen van zwaartekrachtgolven, het bestuderen van de koppeling van gravitatie aan de spin en impulsmoment van massa's, zoals in de zgn. gravitomagnetische verschijnselen, en mogelijke afwijkingen van pure Newtonse gravitatie in de kracht tussen stationaire massa's op submillimeterafstanden zijn onderzoeksrichtingen waarin de komende jaren veel nieuwe ontwikkelingen tegemoet gezien kunnen worden.

[9] R. Brustein, M. Gasperini en G. Veneziano, Phys. Rev. D55 (1997), 3882.
Zie ook: http://www.to.infn.it/∼ gasperini/

Samenvatting van hoofdstuk 8

De ART is conceptueel een grote sprong vooruit in vergelijking met Newtons zwaartekrachtwet gebaseerd op direkte werking op afstand. Hoewel er nog veel experimenteel onderzoek te doen valt, is het duidelijk dat de relativistische grondslagen van de ART onontbeerlijke zijn voor een goede beschrijving van alle bekende zwaartekrachtverschijnselen. Toch is de ART zeker niet het laatste woord op het gebied van de gravitatiefysica. Op kosmische schaal dan wel in het submillimetergebied zouden nieuwe krachtvelden een rol kunnen spelen, i.h.b. krachten die beschreven worden door scalaire velden. In het regime van de quantumverschijnselen speelt gravitatie op afstandsschalen die experimenteel toegankelijk zijn geen rol. Toch is er voldoende reden te denken, dat gravitatie bij de Planckschaal allesoverheersend wordt. De huidige beschrijving in termen van de klassieke ART voldoet dan niet meer. Mogelijke nieuwe ingredienten in de quantumtheorie zijn supersymmetrie en extra dimensies. Een raamwerk dat uitzicht biedt op een volledige oplossing van alle wiskundige problemen wordt geboden door de superstringtheorie. Tot heden heeft deze theorie nog weinig contact gemaakt met de wereld van waarneembare verschijnselen, maar de verwachtingen zijn niettemin hoog.