# Onderzoek naar Monotoniciteit en Stabiliteit van numerieke methoden voor de Convectie Vergelijking

Marco Baro
*CWI*
*Postbus 94079, 1090 GB Amsterdam*

Amsterdam, November 1996

# Voorwoord

Hierbij wil ik het CWI heel hartelijk bedanken voor het bieden van de gelegenheid voor het maken van mijn afstudeerwerk.

Verder wil ik dr.ir. B. Koren heel hartelijk bedanken voor de uitstekende begeleiding tijdens mijn afstudeerwerk en prof.dr.ir. P. Wesseling voor het begeleiden van mijn afstudeerstage vanuit de universiteit en voor het bestuderen en corrigeren van de verschillende versies van deze scriptie.

Daarnaast wil ik graag drs. P. de Zeeuw heel hartelijk bedanken voor de vele stimulerende gesprekken, ir. W. J. H. Stortelder en ir. J. Noordmans voor de lekkere koffie, drs. R. M. Elkenbracht-Huizing voor de vele nuttige sollicitatie tips en verder alle medewerkers van het CWI, bij wie ik altijd terecht kon voor hulp of gewoon een gezellig gesprek.

M. Baro

# Hoofdstuk 1

# Samenvatting

In dit verslag worden numerieke methoden voor de één-dimensionale tijdsafhankelijke convectievergelijking geanalyseerd. Ter voorkoming van "wiggles" wordt gebruik gemaakt van een hogere-orde TVD schema met niet-lineaire flux-limiters voor de plaatsdiscretisatie. Door Zijlema zijn voorwaarden afgeleid voor de flux-limiter, die leiden tot $\mathcal{O}(h^2)$ nauwkeurigheid in gladde extrema. In dit verslag worden voorwaarden afgeleid voor $\mathcal{O}(h^3)$ nauwkeurigheid in gladde extrema. Aan de hand van die voorwaarden worden drie nieuwe limiters afgeleid. Deze drie limiters worden getest met behulp van een aantal testproblemen.

Bij de tijdsdiscretisatie zijn we op zoek gegaan naar een nieuw RK4 schema, welks stabiliteitsgebied het spectrum, behorende bij onze plaatsdiscretisatie, zo goed mogelijk omsluit en dat een zo ruim mogelijke CFL-conditie oplevert. Een probleem hierbij is dat het spectrum nogal moeilijk is te bepalen doordat de plaatsdiscretisatie niet-lineair is. Er wordt daarom op drie manieren een benadering gemaakt van het spectrum. Met behulp van een testprobleem dat wordt opgelost met een aantal RK3 schema's wordt getest welk spectrum met bijbehorende CFL-conditie de beste voorspelling oplevert van hetgeen in de praktijk wordt waargenomen. Het blijkt dat twee spectra bevredigende resultaten opleveren. Deze spectra worden gebruikt om een RK4 schema te vinden dat, net als de plaatsdiscretisatie, derde orde nauwkeurig is voor zowel lineaire als niet-lineaire problemen, en dat een stabiliteitsgebied heeft dat een zo ruim mogelijke CFL-conditie oplevert voor elk van de beide spectra. Dit schema wordt getest om te onderzoeken welke van de twee CFL-condities het beste overeenkomt met de praktijk.

Tenslotte wordt gekeken wat de resultaten zijn van deze CFL-condities als ze worden toegepast op een twee-dimensionaal testprobleem (De Molenkamp test).

# Hoofdstuk 2

# Introductie

Convectieve processen spelen een belangrijke rol in de stromingsleer. Een belangrijke toepassing van convectie is bijvoorbeeld het transport van chemische stoffen in de atmosfeer. Grootschalige, twee- en drie-dimensionale transport- en scheikundige modellen vormen hierbij een belangrijk gereedschap om het gedrag van chemische stoffen in de atmosfeer te kunnen begrijpen en voorspellingen te kunnen doen naar de invloed op bijvoorbeeld het milieu of de gezondheid. Het is in deze modellen dan ook van het grootste belang om de convectie nauwkeurig te modelleren. Het probleem is echter dat het numeriek modelleren van convectie vaak gepaard gaat met veel moeilijkheden.

Zo kan het gebeuren dat een numeriek schema een oplossing oplevert met negatieve waarden van bijvoorbeeld concentraties van chemische bestanddelen; hetgeen natuurlijk fysisch niet mogelijk is. Het verhelpen van dit probleem kan echter het gevolg hebben dat we met een ander probleem te maken krijgen. Zo is het bijvoorbeeld mogelijk om een numeriek schema te construeren dat deze negatieve waarden voorkomt. Het gevolg hiervan kan zijn dat dit 'verbeterde' schema gaat zorgen voor te grote numerieke diffusie.

We beginnen in dit verslag voor de volledigheid met de afleiding van de convectie-diffusie vergelijking voor een willekeurige dimensie. Voordat deze vergelijking zal worden gediscretiseerd in de plaats, zal deze worden omgeschreven naar een integraalvorm. De reden hiertoe is dat we op deze manier een grotere klasse van oplossingen toelaten.

Vervolgens zal deze integraalvorm in de plaats gediscretiseerd worden met behulp van de cel-gecentreerde eindige-volumemethode. Het zal blijken dat met deze methode automatisch wordt voldaan aan een belangrijke eis: behoud van massa. Bij de toepassing van de eindige volume methode zal een vereenvoudiging plaats vinden naar de één-dimensionale lineaire convectie vergelijking, waarbij de convectiesnelheid constant gehouden wordt.

De reden voor deze vereenvoudiging is dat ten eerste de exacte oplossing bekend is van deze vergelijking en dus de numerieke oplossing hiermee goed kan worden vergeleken. Ten tweede bemoeilijken variërende convectiesnelheden en meerdere dimensies het maken van een numeriek schema en als het numerieke schema al niet goed werkt in het simpele geval, dan werkt het zeker niet in het moeilijkere, meer-dimensionale geval.

Het doel van dit verslag is om de voorwaarden af te leiden voor een nauwkeurig numeriek schema voor de convectie vergelijking, dat bovendien monotoon en stabiel is. Om dit doel te bereiken zal een niet-lineaire functie worden geïntroduceerd, welke de flux-limiter zal worden genoemd. Het zal dan blijken dat deze flux-limiter de eigenschap heeft om in gebieden van de numerieke oplossing, waar valse niet-monotoniciteit dreigt (met als gevolg misschien negatieve waarden), dit te voorkomen.

Verder zullen er voorwaarden voor de flux-limiter worden afgeleid, welke zullen leiden tot een gebied, waarbinnen de flux-limiter zich moet bevinden, wil het numerieke schema monotoon zijn.

Vervolgens zullen er enige veel gebruikte limiters worden bekeken en worden hun eigenschappen toegelicht. Daarna zullen nieuwe voorwaarden worden afgeleid, waarvan we graag willen dat een flux-limiter daaraan voldoet. Met deze nieuwe voorwaarden zullen drie limiters worden geconstrueerd. Vervolgens zullen deze limiters worden getest en aan de hand daarvan worden uitspraken gedaan naar de beste limiter.

Als eenmaal de plaatsdiscretisatie vastligt gaan we proberen een RK4 schema te construeren waarbij een zo groot mogelijke CFL-conditie hoort. De werkwijze hierbij is dat we allereerst het spectrum behorende bij plaatsdiscretisatie gaan proberen af te leiden. Daarna zullen we een nieuw RK4 schema construeren, welks stabiliteitgebied ons benaderde spectrum zo goed mogelijk omsluit en een zo groot mogelijke CFL-conditie garandeert.

Tenslotte gaan we eens kijken hoe onze resultaten het doen als ze worden toegepast op een tweedimensionaal testprobleem.

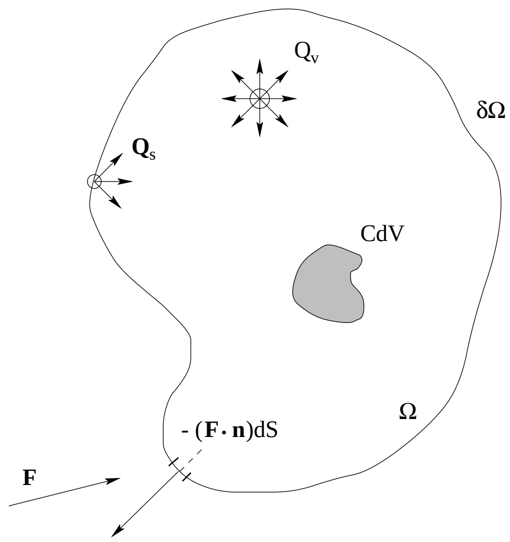
# Hoofdstuk 3

# De Convectie-Diffusie vergelijking

## 3.1 Afleiding

In deze paragraaf zal de convectie-diffusie vergelijking worden afgeleid. We zullen dit doen aan de hand van de algemene behoudswet, die als eerste zal worden afgeleid. We volgen hier de tekst zoals deze staat in [5].

Laten we een vast (dus niet met de stroming meebewegend) volume $\Omega$ beschouwen, dat wordt omsloten door een oppervlak $\partial\Omega$, waarop $\mathbf{n}$ de buitennormaal is. Verder nemen we aan dat binnen het beschouwde volume $\Omega$ zich een scalaire grootheid $C$ per eenheid van volume bevindt. Eén en ander is nog eens weergegeven in figuur 3.1.




Figuur 3.1: Volume $\Omega$.

Het begrip behoud houdt in dat de verandering van een bepaalde grootheid, binnen een bepaald volume, het gevolg is van de netto bijdragen van bronnen binnen het volume en de hoeveelheid, die door de rand van het volume naar binnen en naar buiten stroomt. Deze hoeveelheid duiden we aan met *flux.*

Met het oog op hetgeen hierboven staat kunnen we nu de algemene vorm van de behoudswet als volgt opschrijven.
Er geldt dat de verandering van de grootheid $C$ per eenheid van tijd *binnen het volume* $\Omega$, gegeven wordt door de term

$$\frac{d}{dt}\int_{\Omega} C dV. \tag{3.1.1}$$

Deze term moet gelijk zijn aan de netto bijdrage van de fluxen, over de rand $\partial\Omega$

$$-\oint_{\partial\Omega}(\mathbf{F}\cdot\mathbf{n})dS, \tag{3.1.2}$$

waarbij $\mathbf{F}$ de flux-vector voorstelt. Bij deze term moeten nog de bijdragen van de bronnen worden opgeteld. Deze zijn te verdelen in twee soorten, namelijk de volumebronnen, $Q_v$, en de oppervlaktebronnen, $\mathbf{Q}_s$. Dus de totale bijdrage van de bronnen wordt dan

$$\int_\Omega Q_v dV + \oint_{\partial\Omega}(\mathbf{Q}_s\cdot\mathbf{n})dS. \tag{3.1.3}$$

Als we alles bij elkaar nemen dan krijgen we de volgende algemene vorm van de behoudswet met betrekking tot de grootheid $C$

$$\frac{d}{dt}\int_\Omega C dV + \oint_{\partial\Omega}(\mathbf{F}\cdot\mathbf{n})dS = \int_\Omega Q_v dV + \oint_{\partial\Omega}(\mathbf{Q}_s\cdot\mathbf{n})dS. \tag{3.1.4}$$

Of, na toepassing van de divergentie-stelling van Gauss (voor continue fluxen en oppervlaktebronnen)

$$\frac{d}{dt}\int_\Omega C dV + \int_\Omega(\nabla\cdot\mathbf{F})dV = \int_\Omega Q_v dV + \int_\Omega(\nabla\cdot\mathbf{Q}_s)dV. \tag{3.1.5}$$

Omdat formule (3.1.5) nu geldt voor een willekeurig volume $\Omega$, kunnen we de volgende behoudswet in differentiaalvorm opschrijven

$$\frac{\partial C}{\partial t} + \nabla\cdot\mathbf{F} = Q_v + \nabla\cdot\mathbf{Q}_s. \tag{3.1.6}$$

We gaan vervolgens de fluxvector $\mathbf{F}$ splitsen in twee componenten, namelijk een convectief deel, $\mathbf{F}_c$, en een diffusief deel, $\mathbf{F}_d$. We kunnen nu schrijven voor $\mathbf{F}_c$:

$$\mathbf{F}_c = C\mathbf{u}. \tag{3.1.7}$$

Hierbij is $\mathbf{u}$ het snelheidsveld van de stroming. Formule (3.1.7) geeft de hoeveelheid $C$ aan, die wordt getransporteerd door de stroming met snelheid $\mathbf{u}$.
Voor $\mathbf{F}_d$ kunnen we de volgende formule opschrijven met behulp van de gradiëntwet van Fick:

$$\mathbf{F}_d = -k\rho\nabla c. \tag{3.1.8}$$

Formule (3.1.8) stelt de fluxbijdrage voor ten gevolge van de moleculaire effecten, als het medium in rust is. Hierbij stelt $c$ de grootheid $C$ per eenheid van massa voor. Verder stelt $\rho$ de dichtheid van het medium voor, waarin $\Omega$ zich bevindt (dus er geldt $C = \rho c$) en is $k$ de diffusiecoëfficiënt.
Als we dit allemaal weer invullen in formule (3.1.6) dan krijgen we

$$\frac{\partial(\rho c)}{\partial t} + \nabla\cdot(\rho c\mathbf{u}) = \nabla\cdot(k\rho\nabla c) + Q_v + \nabla\cdot\mathbf{Q}_s. \tag{3.1.9}$$

Formule (3.1.9) staat bekend als de algemene vorm van de transportvergelijking voor de grootheid $C = \rho c$.

Vervolgens gaan we de formule voor het behoud van massa afleiden omdat we met behulp hiervan formule (3.1.9) in een alternatieve vorm kunnen schrijven. In feite hebben we al het werk al gedaan. We kunnen formule (3.1.9) gaan toepassen, waarbij we voor de grootheid $C$ de dichtheid $\rho$ gebruiken. We nemen hierbij het volgende aan:

Ten eerste nemen we aan dat er geen bronnen zijn, dus de termen $Q_v$ en $\mathbf{Q}_s$ doen niet mee. (Dit zou bijvoorbeeld het geval kunnen zijn als we meerfasen stromingen zouden beschouwen waar, ten gevolge van chemische reacties, massabronnen aanwezig zijn. We beschouwen hier alleen eenfase-stromingen.)

Ten tweede bestaat de diffusieve flux niet als we massatransport beschouwen. (Dit houdt in dat er alleen massa kan worden getransporteerd door middel van convectie.) Als we dit allemaal in acht nemen dan vinden we de volgende formule voor het behoud van massa:

$$\frac{d}{dt}\int_{\Omega}\rho dV + \oint_{\partial\Omega}\rho(\mathbf{u}\cdot\mathbf{n})dS = 0. \tag{3.1.10}$$

Toepassen van de divergentie-stelling van Gauss geeft vervolgens

$$\frac{\partial\rho}{\partial t} + \nabla\cdot(\rho\mathbf{u}) = 0. \tag{3.1.11}$$

De term binnen de divergentie-operator kan nog verder worden uitgewerkt. Als namelijk de *totale afgeleide* wordt geïntroduceerd:

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{u}\cdot\nabla, \tag{3.1.12}$$

dan leidt dit, na toepassen op formule (3.1.11), tot de volgende vorm voor de wet van behoud van massa:

$$\frac{D\rho}{Dt} + \rho(\nabla\cdot\mathbf{u}) = 0. \tag{3.1.13}$$

Indien $\frac{D\rho}{Dt} = 0$, dan verandert $\rho$ niet voor een partikel dat met de stroming meebeweegt. De stroming heet in dit geval *incompressibel.* Uit (3.1.13) volgt dan

$$\nabla\cdot\mathbf{u} = 0. \tag{3.1.14}$$

Het stromingsveld is dan *divergentievrij.*

Formule (3.1.9) kan nu met behulp van formule (3.1.11) op een alternatieve wijze worden herschreven. Als namelijk formule (3.1.11) vermenigvuldigd wordt met $c$ en vervolgens wordt afgetrokken van formule (3.1.9), dan krijgen we

$$\rho\frac{\partial c}{\partial t} + \rho(\mathbf{u}\cdot\nabla c) = \nabla\cdot(k\rho\nabla c) + Q_v + \nabla\cdot\mathbf{Q}_s, \tag{3.1.15}$$

of met behulp van de totale afgeleide, formule (3.1.12)

$$\rho\frac{Dc}{Dt} = \nabla\cdot(k\rho\nabla c) + Q_v + \nabla\cdot\mathbf{Q}_s. \tag{3.1.16}$$

We gaan nu de veronderstelling maken, dat de dichtheid in het medium onafhankelijk is van de tijd én de plaats. In dat geval gaat formule (3.1.16) over in:

$$\frac{DC}{Dt} = \nabla\cdot(k\nabla C) + Q. \tag{3.1.17}$$

Waarbij voor $Q$ geldt: $Q = Q_v + \nabla\cdot\mathbf{Q}_s$. Formule (3.1.17) staat bekend als de inhomogene convectie-diffusie vergelijking, voor een willekeurige grootheid $C = \rho c$ (per eenheid van volume) en bronterm $Q$.

Als er geen bronnen zijn in de stroming (dus $Q = 0$), dan houden we een iets eenvoudiger formule over. In dat geval gaat formule (3.1.17) namelijk over in

$$\frac{DC}{Dt} = \nabla\cdot(k\nabla C). \tag{3.1.18}$$

Formule (3.1.18) staat bekend als de homogene convectie-diffusie vergelijking.

# Hoofdstuk 4

# Plaatsdiscretisatie

Als nu formule (3.1.17) of (3.1.18) direct zou worden gediscretiseerd, dan zou dit veelal niet leiden tot de correcte numerieke oplossing. De reden hiervoor is dat bij het direct toepassen van een numeriek schema, het exacte behoud van $C$ verloren kan gaan. (In het geval van bijvoorbeeld een stationair probleem, $\frac{\partial C}{\partial t} = 0$, zullen de massa-fluxen op de randen van aangrenzende cellen niet precies tegen elkaar wegvallen, met als gevolg dat in het numerieke schema de totale massa niet exact constant blijft; er zal dus massa verloren gaan of bijkomen!)

Om het bovenstaande probleem tegen te gaan, gaan we formule (3.1.17) omschrijven in de *conservatieve vorm.* We nemen als het ware weer een stapje terug:

$$\frac{d}{dt}\int_{\Omega} C dV + \oint_{\partial\Omega}\left[C(\mathbf{u}\cdot\mathbf{n}) - k\frac{\partial C}{\partial n}\right] dS = \int_{\Omega} Q dV. \tag{4.0.1}$$

Een ander voordeel van deze schrijfwijze is dat we op deze manier een grotere klasse van oplossingen toelaten. Het is namelijk zo dat differentiaalvergelijkingen niet geldig zijn op discontinuïteiten. Door nu in plaats van formule (3.1.17), formule (4.0.1) toe te passen, laten we nu wel oplossingen toe, die discontinu kunnen zijn. Het belang van deze conservatieve formulering is reeds door Lax in [12] aangestipt.

Formule (4.0.1) vormt het uitgangspunt voor de eindige-volumemethode (EVM). Bij de EVM worden behouds-formules in integraalvorm, zoals formule (4.0.1), als uitgangspunt genomen voor ieder volumetje rondom een roosterpunt. Dit zal in de volgende paragraaf verder uiteen worden gezet.

## 4.1 De Eindige-volumemethode

Zoals reeds in de inleiding is uitgelegd, gaan we voor de plaatsdiscretisatie EVM toepassen.

Bij het toepassen van de EVM wordt het gebied $\Omega$ opgedeeld in kleine volumina $\Omega_i$ met rand $\partial\Omega_i$, met $i = 1, \ldots, N$. Meestal wordt een regelmatige verdeling genomen. Zo worden bijvoorbeeld in twee dimensies rechthoeken of driehoeken gebruikt. Bij de keuze van de volumina, die ook wel *controle-volumina* worden genoemd, moet wel voldaan zijn aan een aantal eisen:

1. De vereniging van alle volumina $\Omega_i$ moet het gehele domein $\Omega$ bedekken.

2. Aangrenzende volumina $\Omega_i$ mogen elkaar niet overlappen.

3. De flux door een rand $\partial\Omega_i$ is identiek voor de twee volumina die aan $\partial\Omega_i$ grenzen.

Deze drie eisen zorgen ervoor dat voor het gehele rekengebied $\Omega$ aan behoud van massa wordt voldaan. Hier zal later op worden teruggekomen. De hoeveelheid onbekenden, die bij elk volume wordt betrokken, hangt af van het beschouwde probleem. In het geval van een scalair probleem, bijvoorbeeld, associëren we bij elk controle-volume één onbekende. In dat geval denken we ons de onbekenden gelegen te zijn in de zwaartepunten van de volumina (cel-gecentreerde EVM).

Het handige bij de EVM, naast de grote flexibiliteit waarmee een willekeurig domein kan worden opgedeeld, is dat vooral natuurlijke randvoorwaarden heel gemakkelijk kunnen worden verwerkt, dit in tegenstelling tot de eindige differentie methode. Men moet er dan wel voor zorgen dat de rand $\partial\Omega$ van het rekengebied $\Omega$ opgebouwd is uit randen $\partial\Omega_i$ van de controle-volumina.

Voor een willekeurig controle-volume $\Omega_i$ met rand $\partial\Omega_i$, kunnen we nu schrijven

$$\frac{d}{dt}\int_{\Omega_i} C dV + \oint_{\partial\Omega_i}\left[C(\mathbf{u}\cdot\mathbf{n}) - k\frac{\partial C}{\partial n}\right] dS = \int_{\Omega_i} Q dV. \tag{4.1.2}$$

De bovenstaande derde eis is van groot belang voor de eigenschap dat de massa ook behouden blijft in discrete vorm. Dit kan als volgt worden ingezien. Elk lijnstuk $\partial\Omega_i$ van de rand van een controle-volume komt in het binnengebied twee keer voor, omdat het de rand vormt van twee naast elkaar liggende volumina. De uitwendige normaal op dit lijnstuk wijst dus voor die twee volumina precies in tegengestelde richting. Dit heeft tot gevolg dat de *bijdragen* van dat lijnstuk aan de twee vergelijkingen (geldend voor die twee volumina) gelijk van grootte zijn, maar van tegengesteld teken.

Als dus alle vergelijkingen voor de afzonderlijke volumina bij elkaar worden opgeteld, dan krijgen we formule (4.0.1) weer terug, die geldt voor het hele domein $\Omega$. Dus de totale flux blijft bij het toepassen van de EVM ook in de gediscretiseerde vorm behouden. Discretisatiemethoden, die hieraan voldoen, worden ook wel *conservatieve* schema's genoemd.

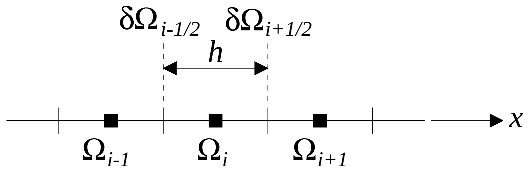
## 4.2 De Convectie vergelijking

In veel toepassingen in de stromingsleer domineert de convectieterm over de diffusieterm. Dit is het geval als het Reynolds getal $Re \gg 1$ is. In dat geval zijn de diffusietermen klein in grote delen van het domein.

Als toepassing van de EVM, bij de plaatsdiscretisatie, beschouwen we daarom de pure convectie vergelijking en wel in één dimensie, omdat uitgaande van deze situatie al veel aspecten kunnen worden belicht. Verder stellen we $Q = 0$ en de convectiesnelheid $\mathbf{u} = u$ constant en positief. We beschouwen dus

$$\frac{\partial C}{\partial t} + u\frac{\partial C}{\partial x} = 0. \tag{4.2.3}$$

We verdelen het domein $\Omega$ (hier de $x$-as) in kleine volumina (hier lijnstukken) ter lengte $h$. De controle-volumina zijn zo gekozen dat in het midden van elk volume zich één onbekende $c_i = c(ih)$, bevindt. Eén en ander wordt nog eens toegelicht in figuur 4.1.




Figuur 4.1: Cel-gecentreerde eindige volumes in één dimensie.

Toepassing van de EVM op (4.2.3) betekent integreren over $\Omega_i$, hetgeen resulteert in

$$\frac{d}{dt}\int_{\Omega_i} C dx + \int_{\partial\Omega_{i+\frac{1}{2}}} C(\mathbf{u}\cdot\mathbf{n}) ds + \int_{\partial\Omega_{i-\frac{1}{2}}} C(\mathbf{u}\cdot\mathbf{n}) ds = 0. \tag{4.2.4}$$

De randintegralen in formule (4.2.4) worden nu benaderd met behulp van de middenpuntsregel en de volume-integraal wordt bepaald met behulp van eenpunts-integratie. Zo krijgen we de volgende semi-gediscretiseerde vergelijking

$$\frac{dC_i}{dt} + \frac{u}{h}(C_{i+\frac{1}{2}} - C_{i-\frac{1}{2}}) = 0. \tag{4.2.5}$$

De nauwkeurigheid van EVM discretisatie wordt nu verder bepaald door de manier waarop de flux $uC_{i+\frac{1}{2}}$ op de celwand worden berekend. Omdat $u$ constant is komt het aan op de benadering van $C_{i+\frac{1}{2}}$, welke grootheid we gemakshalve ook de flux zullen noemen. Als we nu aannemen dat de stroming zich beweegt in de positieve richting van de $x$-as in figuur 4.1, dan is een mogelijke benadering voor $C_{i+\frac{1}{2}}$

$$C_{i+\frac{1}{2}} = C_i. \tag{4.2.6}$$

Invullen in formule (4.2.5) geeft dan het eerste orde upwind schema:

$$\frac{dC_i}{dt} + \frac{u}{h}(C_i - C_{i-1}) = 0. \tag{4.2.7}$$

Om de nauwkeurigheid van deze discretisatie te bepalen, kan de term $C_{i-1}$ worden benaderd door een Taylor-reeks. We krijgen dan

$$\left(\frac{\partial C}{\partial t} + u\frac{\partial C}{\partial x} - \frac{h}{2}u\frac{\partial^2 C}{\partial x^2}\right)_i = \mathcal{O}(h^2). \tag{4.2.8}$$

Dit wordt de *vervangende vergelijking* genoemd (modified equation). Deze formule laat zien dat de beschouwde benadering voor $C_{i+\frac{1}{2}}$ een $\mathcal{O}(h)$ schema oplevert. In (4.2.8) is nog iets interessants te zien. De derde term in de vervangende vergelijking is een diffusie term. Dit geeft aan dat deze discretisatie diffusief is.

Om een hogere orde nauwkeurigheid te verkrijgen voor de convectie-term nemen we de volgende lineaire combinatie:

$$C_{i+\frac{1}{2}} = C_i + \left(\frac{1+\kappa}{4}\right)(C_{i+1} - C_i) + \left(\frac{1-\kappa}{4}\right)(C_i - C_{i-1}), \quad \kappa \in [-1, 1]. \tag{4.2.9}$$

De formule voor $C_{i-\frac{1}{2}}$ wordt verkregen door in formule (4.2.9) de index $i$ te vervangen door index $i-1$. Deze formule wordt het *$\kappa$-schema* genoemd. Het schema is door Van Leer geïntroduceerd in [13]. Door de waarde van $\kappa$ te variëren kunnen verschillende discretisatie schema's worden verkregen. Zo krijgen we voor $\kappa = 0$ het schema van Fromm, $\kappa = 1$ geeft het standaard $2^e$ orde centrale schema, bij $\kappa = -1$ vinden we het $\mathcal{O}(h^2)$ fully upwind schema van Beam en Warming en voor $\kappa = \frac{1}{3}$ krijgen we het $\mathcal{O}(h^3)$ upwind-biased schema.

Hieronder zal nu worden aangetoond dat voor $\kappa = \frac{1}{3}$ de hoogste nauwkeurigheid wordt bereikt. Hiertoe gaan we weer de vervangende vergelijking afleiden. Er geldt

$$C_{i+\frac{1}{2}} - C_{i-\frac{1}{2}} = C_i - C_{i-1} + \left(\frac{1+\kappa}{4}\right)(C_{i+1} - 2C_i + C_{i-1}) + \left(\frac{1-\kappa}{4}\right)(C_i - 2C_{i-1} + C_{i-2}). \tag{4.2.10}$$

We gaan weer op dezelfde wijze te werk als bij het bepalen van de nauwkeurigheid van het $\mathcal{O}(h)$ upwind schema. We ontwikkelen de termen $C_{i-1}$, $C_{i+1}$ en $C_{i-2}$ weer in een Taylor-reeks. Als dit allemaal gesubstitueerd wordt in formule (4.2.10) en de termen bij elkaar worden genomen, dan vinden we tenslotte voor de vervangende vergelijking

$$\left(\frac{\partial C}{\partial t} + u\frac{\partial C}{\partial x} + \left(\frac{3\kappa - 1}{12}\right)h^2\frac{\partial^3 C}{\partial x^3}\right)_i = \mathcal{O}(h^3). \tag{4.2.11}$$

Het is nu gemakkelijk om in te zien dat als $\kappa \neq \frac{1}{3}$, dit resulteert in $\mathcal{O}(h^2)$ nauwkeurigheid. Als $\kappa = \frac{1}{3}$ dan krijgen we het meest nauwkeurige schema, namelijk $\mathcal{O}(h^3)$.

Als we (4.2.9) in een Taylorreeks gaan schrijven rond $i+\frac{1}{2}$, dan kunnen we nog iets interessants zien. We vinden dan namelijk

$$C_{i+\frac{1}{2}} = \left(C + \left(\frac{2\kappa - 1}{8}\right)h^2\frac{\partial^2 C}{\partial x^2} + \mathcal{O}(h^3)\right)_{i+\frac{1}{2}}. \tag{4.2.12}$$

We zien nu dat voor $\kappa = \frac{1}{2}$ de fluxberekening $\mathcal{O}(h^3)$ is. Met andere woorden voor $\kappa = \frac{1}{2}$ (dit wordt ook wel het QUICK-schema genoemd) is de berekening van de flux op een celwandje het meest nauwkeurig. Echter (4.2.11) laat zien dat voor die waarde van $\kappa$, de discrete vergelijking $\mathcal{O}(h^2)$ nauwkeurig is.

## 4.3 Monotoniciteit

Het is aantrekkelijk om bij de plaatsdiscretisatie van de convectie term het $\kappa$-schema te gaan gebruiken. Het liefst zouden we het ($\kappa = \frac{1}{3}$)-schema willen gebruiken omdat, zoals in paragraaf 4.2 is aangetoond, dan de hoogste nauwkeurigheid wordt bereikt.

Er is echter één probleem. Er is reeds in de inleiding verteld over het belang van monotoniciteit van de numerieke oplossing. We zouden nu graag ook bij het toepassen van bijvoorbeeld het ($\kappa = \frac{1}{3}$)-schema behoud van monotoniciteit willen hebben. Helaas is dit niet zomaar gewaarborgd zoals in het vervolg van deze paragraaf zal blijken.

Allereerst gaan we eens kijken wat we precies bedoelen met een monotoon schema. Laten we de semi-gediscretiseerde vergelijking (4.2.7) eens schrijven in de volgende vorm

$$\frac{dC_i}{dt} = H\left(C_{i-k}, C_{i-k+1}, \ldots, C_{i+k}\right). \tag{4.3.13}$$

De volgende definitie voor monotoniciteit wordt veel gebruikt (zie [6], blz. 525).

**Definitie 4.1** *Schema (4.3.13) is een monotoon schema als*

$$\frac{\partial H}{\partial C_j} \geq 0, \quad \forall j \in [i-k, i+k]. \tag{4.3.14}$$

We zien dus dat (4.3.13) monotoon is als $H$ een monotoon stijgende functie is in ieder van zijn argumenten. De volgende stelling geeft aan waarom een hoog orde nauwkeurig schema (hoger dan $\mathcal{O}(h)$), zoals bijv. het ($\kappa = \frac{1}{3}$)-schema, geen monotoniciteit garandeert.

**Stelling 4.1** *Er bestaat geen lineair convectie schema dat een hogere nauwkeurigheid heeft dan $\mathcal{O}(h)$ én dat tevens monotoon is.*

(Zie voor het bewijs hiervan [1].) Dus het ($\kappa = \frac{1}{3}$)-schema mag dan wel heel nauwkeurig zijn, toch kunnen er bij deze discretisatie van de convectie term wiggles ontstaan.

Om nu toch hogere orde (dan $\mathcal{O}(h)$) schema's te willen toepassen, dienen we niet-lineaire schema's te gebruiken. Voorts kunnen we ook acceptabele oplossingen verkrijgen als de eis van monotoniciteit wordt vervangen door de zwakkere eis van *total variation diminishing* (TVD). De total variation van een numerieke oplossing $\{C_i\}$ is gedefinieerd door

$$TV(C) = \sum_i \mid C_{i+1} - C_i \mid. \tag{4.3.15}$$

Een schema is nu TVD als de totale variatie niet toeneemt bij het nemen van een tijdstap. Er moet dan dus gelden

$$TV(C^{n+1}) \leq TV(C^n). \tag{4.3.16}$$

Als nu aan de volgende monotoniciteits eigenschappen voldaan is tijdens het nemen van een tijdstap:

1. Er ontstaan geen nieuwe extrema.
2. De waarde van een lokaal minimum neemt niet af en de waarde van een lokaal maximum neemt niet toe.

dan is het schema *monotoniciteit behoudend*. Dit wil dus zeggen dat als $C^n$ monotoon is, $C^{n+1}$ monotoon blijft. De volgende stelling is nu heel belangrijk (zie [3]).

**Stelling 4.2**
*1. Een monotoon schema is TVD.*
*2. Een TVD schema is monotoniciteit behoudend.*

Er kan nu worden aangetoond, dat voor lineaire schema's, behoud van monotoniciteit hetzelfde is als monotoniciteit. Dus ieder lineair TVD-schema is monotoon en dus $\mathcal{O}(h)$ nauwkeurig. De belemmering in stelling 3.1 gaat echter niet op voor niet-lineaire schema's, zodat het mogelijk is voor niet-lineaire hogere orde schema's om toch monotoniciteit behoudend te zijn.

We zullen dus, om hogere orde nauwkeurigheid én monotoniciteit af te dwingen, niet-lineaire schema's moeten gaan toepassen. Dit is op zich een verrassende stap in het geval van een lineaire differentiaal vergelijking, die evenwel bij de numerieke behandeling van hyperbolische differentiaalvergelijkingen geheel is ingeburgerd.

In de volgende paragraaf zal een niet-lineair schema geïntroduceerd worden met behulp van *flux-limiting*.

## 4.4 De Flux-limiter

Een manier om een niet-lineair hogere-orde TVD schema te formuleren is met behulp van *flux-limiting*. Beschouw de volgende algemene uitdrukking voor de flux (voor het geval dat $u > 0$)

$$C_{i+\frac{1}{2}} = C_i + \frac{1}{2}\phi(r_{i+\frac{1}{2}})(C_i - C_{i-1}), \tag{4.4.17}$$

waarbij geldt

$$r_{i+\frac{1}{2}} = \frac{C_{i+1} - C_i}{C_i - C_{i-1}}. \tag{4.4.18}$$

Formule (4.4.17) kan worden beschouwd als de som van een $\mathcal{O}(h)$ upwind term (waarvan bekend is dat deze zeer diffusief is) en een term die *anti-diffusief* is. De hogere orde anti-diffusieve term wordt vermenigvuldigd met een functie $\phi$, die de *flux-limiter* wordt genoemd.

Deze flux-limiter $\phi$ is een niet-lineaire functie van naburige roosterpunt waarden van $C_i$ en kan worden beschouwd als soort schakelaar tussen een hoog-orde schema en een laag-orde schema. Dat wil zeggen, in gebieden waar de oplossing glad (monotoon) is, zal de flux-limiter zorgen voor een hoog-orde schema; terwijl in gebieden waar de oplossing niet monotoon dreigt te worden, de limiter ter voorkoming van valse niet-monotoniciteiten, voor een laag-orde schema zal zorgen.

Een mogelijkheid hiertoe zou zijn om in gebieden waar niet-monotoniciteit dreigt, de flux-limiter naar 0 te laten gaan. Gemakkelijk is in te zien dat we dan weer het $\mathcal{O}(h)$ upwind schema krijgen, waarvan we reeds weten dat dit monotoon is. We komen hier later op terug.

Laten we eens de vervangende vergelijking gaan opstellen voor het niet-lineaire schema met de flux-limiter. Het doel is om hiermee enige eisen af te leiden voor de nauwkeurigheid waaraan de flux-limiter moet voldoen. Net zoals in paragraaf 4.2, beschouwen we weer de pure convectie vergelijking. Allereerst gaan we het quotiënt in formule (4.4.18) anders opschrijven:

$$r_{i+\frac{1}{2}} = \frac{C_{i+1} - C_i}{C_i - C_{i-1}} = \frac{C_i - C_{i-1}}{C_i - C_{i-1}} + \frac{C_{i+1} - 2C_i + C_{i-1}}{C_i - C_{i-1}} = 1 + \frac{C_{i+1} - 2C_i + C_{i-1}}{C_i - C_{i-1}}. \tag{4.4.19}$$

Vervolgens gaan we $\phi(r_{i+\frac{1}{2}})$ in een Taylorreeks schrijven rond het punt 1 met behulp van formule (4.4.19). Dit geeft

$$\phi(r_{i+\frac{1}{2}}) = \phi(1) + \left(\frac{C_{i+1} - 2C_i + C_{i-1}}{C_i - C_{i-1}}\right)\phi'(1) + \frac{1}{2}\left(\frac{C_{i+1} - 2C_i + C_{i-1}}{C_i - C_{i-1}}\right)^2\phi''(1) + \dots \tag{4.4.20}$$

Vervolgens gaan we in formule (4.4.20) bij de kwadraatterm de termen $C_{i+1}$ en $C_{i-1}$, in zowel de teller als de noemer, schrijven in een Taylorreeks rond het punt $x_i$. We krijgen dan

$$\left(\frac{C_{i+1} - 2C_i + C_{i-1}}{C_i - C_{i-1}}\right)^2 = h^2\left(\frac{C''_i}{C'_i}\right)^2 + \mathcal{O}(h^3). \tag{4.4.21}$$

Als we nu al deze resultaten substitueren in formule (4.4.17) dan krijgen we na enig rekenen

$$
\begin{aligned}
C_{i+\frac{1}{2}} &= C_i + \frac{1}{2}\phi'(1)(C_{i+1} - C_i) + \frac{1}{2}(\phi(1) - \phi'(1))(C_i - C_{i-1}) \\
&+ \frac{1}{4}\phi''(1)\left(h^2\left(\frac{C_i''}{C_i'}\right)^2 + \mathcal{O}(h^3)\right)(C_i - C_{i-1})\,. 
\end{aligned} \tag{4.4.22}
$$

Vervolgens gaan we nu hetzelfde doen voor $C_{i-\frac{1}{2}}$. We krijgen dan als we deze twee termen van elkaar aftrekken

$$
\begin{aligned}
C_{i+\frac{1}{2}} - C_{i-\frac{1}{2}} &= (C_i - C_{i-1}) + \frac{1}{2}\phi'(1)(C_{i+1} - 2C_i + C_{i-1}) \\
&+ \frac{1}{2}(\phi(1) - \phi'(1))(C_i - 2C_{i-1} + C_{i-2}) \\
&+ \frac{1}{4}\phi''(1)\left(h^2\left(\frac{C_i''}{C_i'}\right)^2 + \mathcal{O}(h^3)\right)(C_i - 2C_{i-1} + C_{i-2}).
\end{aligned} \tag{4.4.23}
$$

Tayloriseren van de termen $C_{i+1}$, $C_{i-1}$ en $C_{i-2}$ in formule (4.4.23) geeft vervolgens

$$
\begin{aligned}
C_{i+\frac{1}{2}} - C_{i-\frac{1}{2}} &= h\frac{\partial C_i}{\partial x} - \frac{h^2}{2}\frac{\partial^2 C_i}{\partial x^2} + \frac{h^3}{6}\frac{\partial^3 C_i}{\partial x^3} + \mathcal{O}(h^4) + \frac{1}{2}\phi'(1)\left(h^2\frac{\partial^2 C_i}{\partial x^2} + \mathcal{O}(h^4)\right) \\
&+ \frac{1}{2}(\phi(1) - \phi'(1))\left(h^2\frac{\partial^2 C_i}{\partial x^2} - h^3\frac{\partial^3 C_i}{\partial x^3} + \mathcal{O}(h^4)\right) \\
&+ \frac{1}{4}\phi''(1)\left(h^2\left(\frac{C_i''}{C_i'}\right)^2 + \mathcal{O}(h^3)\right)\left(h^2\frac{\partial^2 C_i}{\partial x^2} + \mathcal{O}(h^3)\right).
\end{aligned} \tag{4.4.24}
$$

Als we nu de laatste term in formule (4.4.24) beschouwen als een $\mathcal{O}(h^4)$ term en vervolgens de betreffende termen bij elkaar nemen, dan vinden we de volgende vervangende vergelijking

$$
\left(\frac{\partial C}{\partial t} + u\frac{\partial C}{\partial x} + (\phi(1) - 1)\frac{h}{2}u\frac{\partial^2 C}{\partial x^2} + (1 - 3\phi(1) + 3\phi'(1))\frac{h^2}{6}u\frac{\partial^3 C}{\partial x^3} + \mathcal{O}(h^3)\right)_i = 0. \tag{4.4.25}
$$

Het is nu gemakkelijk om in te zien dat als $\phi(1) = 1$ er sprake is van $\mathcal{O}(h^2)$ nauwkeurigheid. Verder valt op te merken dat voor $\phi(1) = 1$ én $\phi'(1) = \frac{2}{3}$, we $\mathcal{O}(h^3)$ nauwkeurigheid hebben.

Als we kijken naar (4.4.25), dan kunnen we nog iets interessants zien. Als $\phi(1)$ in de derde term van (4.4.25) groter wordt, dan zal het schema meer anti-diffusief worden. Evenzo zal, als $\phi(1)$ kleiner wordt, het schema diffusiever worden. Beide situaties zouden we graag willen voorkomen. We zullen hier later nog op terugkomen.

Het is ook mogelijk om (4.4.22) direct te vergelijken met (4.2.9), om dan direct voorwaarden te vinden voor de flux-limiter. Vergelijking van de twee formules geeft dat ze zoveel mogelijk overeenstemmen indien

$$
\phi'(1) = \frac{1+\kappa}{2}. \tag{4.4.26}
$$

We zien dus dat de waarde van $\phi'(1)$ bepaalt op welk $\kappa$-schema een limiter wordt toegepast, namelijk: $\kappa = 2\phi'(1) - 1$. Voor $\kappa = \frac{1}{3}$ vinden we $\phi'(1) = \frac{2}{3}$, hetgeen we reeds hebben afgeleid. Met behulp van (4.4.26), (4.4.22) en (4.2.9) vinden we dat moet gelden

$$
\phi(1) - \frac{1+\kappa}{2} = \frac{1-\kappa}{2} \quad \Rightarrow \quad \phi(1) = 1. \tag{4.4.27}
$$

Dit komt ook overeen met onze eerdere resultaten.

In de volgende paragraaf gaan we de voorwaarden afleiden voor de flux-limiter, zodanig dat het numerieke schema TVD wordt.

## 4.5 Het TVD-gebied van de flux-limiter

In deze paragraaf gaan we de TVD voorwaarden voor de flux-limiter formuleren.

We beschouwen weer het semi-gediscretiseerde schema in formule (4.2.5). Hierbij geldt weer voor $C_{i+\frac{1}{2}}$ formule (4.4.17). We kunnen met (4.4.17) nu het verschil $C_{i+\frac{1}{2}} - C_{i-\frac{1}{2}}$ op de volgende manier opschrijven

$$C_{i+\frac{1}{2}} - C_{i-\frac{1}{2}} = \left(1 + \frac{1}{2}\phi(r_{i+\frac{1}{2}}) - \frac{\phi(r_{i-\frac{1}{2}})}{2r_{i-\frac{1}{2}}}\right)(C_i - C_{i-1}). \tag{4.5.28}$$

Formule (4.2.5) gaat hiermee over in

$$\frac{dC_i}{dt} + \frac{u}{h}\left(1 + \frac{1}{2}\phi(r_{i+\frac{1}{2}}) - \frac{\phi(r_{i-\frac{1}{2}})}{2r_{i-\frac{1}{2}}}\right)(C_i - C_{i-1}) = 0. \tag{4.5.29}$$

In [8] wordt nu aangetoond dat als geldt

$$\left(1 + \frac{1}{2}\phi(r_{i+\frac{1}{2}}) - \frac{\phi(r_{i-\frac{1}{2}})}{2r_{i-\frac{1}{2}}}\right) \geq 0, \tag{4.5.30}$$

het beschouwde numerieke schema TVD is.

Vergelijking (4.5.30) zal vervolgens worden omgeschreven naar de volgende eis voor de flux-limiter. Toewerkend naar Spekreijse gaan we eisen dat geldt

$$0 \leq 1 + \frac{1}{2}\phi(r) - \frac{\phi(s)}{2s} \leq \mathcal{M}, \tag{4.5.31}$$

waarin $\mathcal{M}$ een nog vrij te kiezen bovengrens is. Hierbij geldt dat $r = r_{i+\frac{1}{2}}$ en $s = r_{i-\frac{1}{2}}$.

Met behulp van (4.5.31) zijn we nu in staat om het TVD-gebied volgens Spekreijse te tekenen. Voeren we in $M = \mathcal{M} - 1$, dan is in te zien dat aan (4.5.31) wordt voldaan voor

$$\alpha \leq \phi(r) \leq M, \tag{4.5.32a}$$

$$-M \leq \frac{\phi(s)}{s} \leq 2 + \alpha. \tag{4.5.32b}$$

Met behulp van formule (4.5.32a) komen we terug op hetgeen in de vorige paragraaf is besproken. Zoals daar reeds werd aangestipt willen we graag dat het numerieke schema niet te diffusief of te anti-diffusief wordt. De onder- en bovenbegrenzing van $\phi$ in (4.5.32a) zorgt er juist voor dat dit niet gebeurt.

In figuur 4.2 is te zien hoe het monotoniciteitsgebied van Spekreijse eruit komt te zien met behulp van (4.5.32a) en (4.5.32b). Hierbij geldt voor $\alpha$ en $M$: $\alpha \in [-2, 0]$ en $M > 0$.

Het TVD-gebied van Sweby in [15] is te zien in figuur 4.3. Zoals te zien is dit gebied, voor $M = 2$ en $\alpha = 0$, geheel bevat in het monotoniciteitsgebied van Spekreijse. Verder kan worden opgemerkt dat Sweby, in tegenstelling tot Spekreijse, een strengere eis heeft opgelegd, namelijk

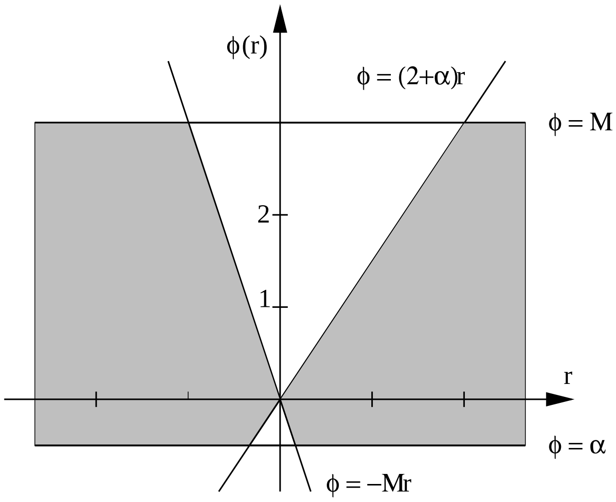
$$\phi(r) = 0, \quad \text{als } r \leq 0. \tag{4.5.33}$$

In gebieden waar valse niet-monotoniciteit dreigt, reduceert het schema dus tot het (monotone) eerste orde upwind schema.

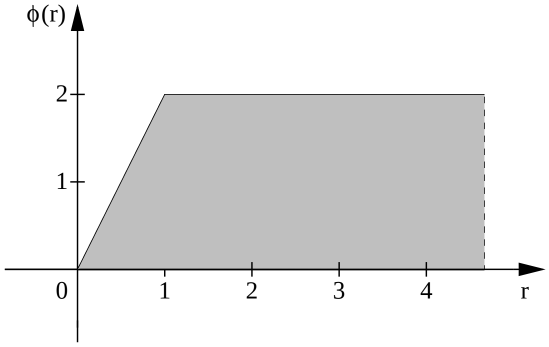
Sweby geeft in [15] ook de voorwaarden aan, waaraan de flux-limiter moet voldoen, zodat een $\mathcal{O}(h^2)$-schema wordt verkregen. We kunnen dit op de volgende manier inzien. Als we namelijk het $\kappa$-schema in formule (4.2.9) noteren in termen van $r = r_{i+\frac{1}{2}}$ dan krijgen we

$$\phi(r) = \frac{1-\kappa}{2} + \frac{1+\kappa}{2}r; \quad -1 \leq \kappa \leq 1. \tag{4.5.34}$$

De doorsnijding van de familie $\kappa$-schema's in formule (4.5.34) en het TVD-gebied van Sweby in figuur 4.3 geeft dan figuur 4.4.




Figuur 4.2: Spekreijse's monotoniciteitsgebied.




Figuur 4.3: Het TVD gebied van Sweby.

In figuur 4.4 is duidelijk terug te zien dat voor $\mathcal{O}(h^2)$ nauwkeurigheid het noodzakelijk is dat $\phi(1) = 1$. Dit is dus geheel in overeenstemming met onze eerdere resultaten.

In de volgende paragraaf zullen we eens een aantal limiters gaan bekijken, die bekend zijn in de literatuur.

## 4.6 Voorbeelden van enige limiters

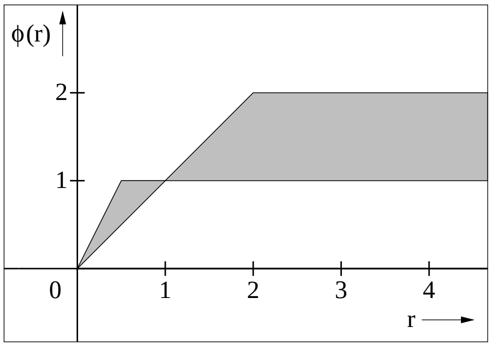
Er zijn in de literatuur vele limiters bekend. In feite is het zo dat er voor elk apart probleem wel een geschikte limiter aanwezig is. Hierdoor is het moeilijk om te spreken over dé beste limiter omdat een zeer geschikte limiter voor één bepaald probleem helemaal niet geschikt hoeft te zijn voor een ander probleem.
Laten we eens een aantal limiters bekijken:

$$\phi(r) = \frac{r + \mid r \mid}{1 + r} \quad \text{Van Leer limiter} \tag{4.6.35a}$$

$$\phi(r) = \max[0, \min(r, \theta)] \quad \text{Chakravarthy-Osher limiter} \tag{4.6.35b}$$

$$\phi(r) = \max[0, \min(\Phi r, 1), \min(r, \Phi)] \quad \text{Sweby } \Phi \text{ limiter} \tag{4.6.35c}$$

$$\phi(r) = \max(0, \min(2r, \min(\frac{1}{3} + \frac{2}{3}r, 2))) \quad \kappa = \tfrac{1}{3} \text{ limiter} \tag{4.6.35d}$$




Figuur 4.4: Het $\mathcal{O}(h^2)$ TVD gebied van Sweby.

$$\phi(r) = \begin{cases} 0, & r < 0 \\ \frac{r}{(1+r)}\{\kappa r + (2-\kappa)\}, & 0 \le r \le 1, \quad 0 \le \kappa \le 1 \\ \frac{2r}{(1+r)^2}\{\kappa r^2 + (1-\kappa)r + 1\}, & 0 \le r \le 1, \quad -1 \le \kappa < 0 \\ \frac{2}{(1+r)^2}\{r^2 + (1+\kappa)r - \kappa\}, & r > 1. \end{cases} \tag{4.6.35e}$$

Hierbij geldt $1 \le \theta \le 2$ en $1 \le \Phi \le 2$. Door deze parameters te variëren kan een familie van limiters worden gegenereerd. Zo geeft $\Phi = 1$ of $\theta = 1$ de Minmod limiter, welke de meest diffusieve limiter is. Verder geeft $\Phi = 2$ de Superbee limiter, welke minder diffusief is dan de Van Leer limiter en de Minmod limiter.

Bij de ($\kappa = \frac{1}{3}$)-limiter (4.6.35d), voorgesteld in [9], is duidelijk te zien dat het $\kappa = \frac{1}{3}$ schema erin gebakken is. Voor $\frac{1}{4} \le r \le \frac{5}{2}$ komt de limiter overeen met het betreffende $\kappa = \frac{1}{3}$ schema. Voor $r < \frac{1}{4}$ en $r > \frac{5}{2}$, daar waar de flux-limiter buiten het TVD gebied van Sweby raakt, volgt de flux-limiter de randen van het TVD gebied.

Merk op dat al deze limiters door het punt $\phi(1) = 1$ gaan. Dus al deze limiters geven een $\mathcal{O}(h^2)$ TVD schema.

De flux-limiter in (4.6.35e), voorgesteld in [18], is op een andere manier geconstrueerd dan de limiters in (4.6.35a) – (4.6.35d). Bij de afleiding van deze limiter is gebruik gemaakt van de *Variable Normalisation* methode. De essentie van deze methode, die door Leonard in [14] is uiteengezet, is dat men alles beschouwd in de NV-ruimte met behulp van NV-variabelen. Een NV-variable $\hat{\phi}$ in het punt $i + k$ is gedefinieerd als

$$\hat{\phi}_{i+k} = \frac{\phi_{i+k} - \phi_{i-1}}{\phi_{i+1} - \phi_{i-1}}, \quad k = -1, -\frac{1}{2}, 0, \frac{1}{2}, 1. \tag{4.6.36}$$

Er geldt dus $\hat{\phi}_{i-1} = 0$ en $\hat{\phi}_{i+1} = 1$. De achterliggende gedachte is nu dat als $\hat{\phi}_{i+k}$ waarden heeft tussen 0 en 1, $\phi_{i+k}$ lokaal monotoon is. Terwijl waarden hierbuiten aangeven dat we met een extremum hebben te maken. Met behulp van deze NV-variabelen in de NV-ruimte, kunnen nu op gemakkelijke wijze lineaire en niet-lineaire schema's worden geconstrueerd en de condities voor monotoniciteit en nauwkeurigheid gemakkelijk worden afgeleid. Vervolgens kunnen met simpele transformaties deze NV-schema's weer terug worden getransformeerd naar flux-limited schema's. Op deze manier is limiter (4.6.35e) zo afgeleid, dat in monotone gebieden van de oplossing de limiter $\mathcal{O}(h^2)$ is voor willekeurige $\kappa$. Verder wordt $\mathcal{O}(h^3)$ nauwkeurigheid bereikt voor $\kappa = \frac{1}{3}$.

## 4.7 $\mathcal{O}(h^2)$ nauwkeurigheid in gladde extrema

Het is opvallend dat alle in de vorige paragraaf beschouwde limiters, lokaal naar $\mathcal{O}(h)$ nauwkeurigheid reduceren bij extrema, ondanks de orde van nauwkeurigheid in monotone gebieden. In [22] wordt door Zijlema afgeleid dat er flux-limiters geconstrueerd kunnen worden, zodanig dat een hogere orde nauwkeurigheid dan $\mathcal{O}(h)$ verkregen kan worden in de gladde extrema. Deze afleiding zal hieronder worden samengevat. We beschouwen de algemene formule voor de flux-limiter (4.4.17). We kunnen nu schrijven

$$C_{i+\frac{1}{2}} - C_{i-\frac{1}{2}} = \left[1 + \frac{1}{2}\phi(r_{i+\frac{1}{2}})\right](C_i - C_{i-1}) - \frac{1}{2}\phi(r_{i-\frac{1}{2}})(C_{i-1} - C_{i-2}). \tag{4.7.37}$$

Vervolgens gaan we de termen $C_{i-1}$ en $C_{i-2}$ Tayloriseren. We krijgen dan als we dit weer substitueren in (4.7.37) en de betreffende termen bij elkaar brengen

$$\begin{aligned}\frac{C_{i+\frac{1}{2}} - C_{i-\frac{1}{2}}}{h} &= \left[1 + \frac{1}{2}\phi(r_{i+\frac{1}{2}}) - \frac{1}{2}\phi(r_{i-\frac{1}{2}})\right]\frac{\partial C_i}{\partial x} \\ &+ \left[-1 - \frac{1}{2}\phi(r_{i+\frac{1}{2}}) + \frac{3}{2}\phi(r_{i-\frac{1}{2}})\right]\frac{h}{2}\frac{\partial^2 C_i}{\partial x^2} \\ &+ \left[1 + \frac{1}{2}\phi(r_{i+\frac{1}{2}}) - \frac{7}{2}\phi(r_{i-\frac{1}{2}})\right]\frac{h^2}{6}\frac{\partial^3 C_i}{\partial x^3} + \mathcal{O}(h^3).\end{aligned} \tag{4.7.38}$$

Het is nu niet moeilijk om in te zien dat als $x_i$ geen extremum is, met behulp van Taylorisatie in de differentieverhoudingen $r_{i+\frac{1}{2}}$ en $r_{i-\frac{1}{2}}$ volgt

$$r_{i+\frac{1}{2}} = r_{i-\frac{1}{2}} = 1 + \mathcal{O}(h). \tag{4.7.39}$$

Substitutie van (4.7.39) in (4.7.38) geeft dan onze $\mathcal{O}(h^2)$ voorwaarde weer terug, namelijk $\phi(1) = 1$. Als $x_i$ wel een extremum is, dan geldt $C_i' = 0$ maar $C_i'' \neq 0$. Er geldt dan

$$r_{i+\frac{1}{2}} = -1 + \mathcal{O}(h), \quad r_{i-\frac{1}{2}} = \frac{1}{3} + \mathcal{O}(h). \tag{4.7.40}$$

Als we dit substitueren in (4.7.38), dan krijgen we de volgende voorwaarde voor $\mathcal{O}(h^2)$ in de *gladde extrema*

$$3\phi(\frac{1}{3}) - \phi(-1) = 2. \tag{4.7.41}$$

Het is gemakkelijk om in te zien dat alle $\kappa$-schema's uit (4.5.34) aan (4.7.41) voldoen, zoals was te verwachten.

Als we kiezen $\phi(-1) = 0$ dan volgt uit (4.7.41): $\phi(\frac{1}{3}) = \frac{2}{3}$. We vinden dus twee nieuwe punten, naast het punt $\phi(1) = 1$, waar we graag willen dat de limiter daar doorheen gaat. Met behulp van deze twee punten kunnen we het volgende voorschrift opstellen voor een *Nieuw Gelimite $\kappa = \frac{1}{3}$ schema*:

$$\phi(r) = \begin{cases} 0, & r < 0 \\ 2r, & 0 \leq r < \frac{1}{3} \\ \frac{2}{3}, & \frac{1}{3} \leq r < \frac{1}{2} \\ \frac{1}{3} + \frac{2}{3}r, & \frac{1}{2} \leq r < \frac{5}{2} \\ 2, & r \geq \frac{5}{2}. \end{cases} \tag{4.7.42}$$

Of in een iets compactere vorm

$$\phi(r) = \max\left(0, \min\left(2r, \frac{2}{3}\right), \min\left(\min\left(\frac{1}{3} + \frac{2}{3}r, 2r\right), 2\right)\right). \tag{4.7.43}$$

## 4.8 $\mathcal{O}(h^3)$ nauwkeurigheid in gladde extrema

Het is ook mogelijk om voorwaarden te vinden voor $\mathcal{O}(h^3)$ nauwkeurigheid in gladde extrema.

We schrijven hiertoe $\phi(r_{i+\frac{1}{2}})$ en $\phi(r_{i-\frac{1}{2}})$ in de volgende Taylorreeksen rond $\phi(r_{i+\frac{1}{2}}) = -1$, resp. $\phi(r_{i-\frac{1}{2}}) = \frac{1}{3}$

$$\phi(r_{i+\frac{1}{2}}) = \phi(-1) + (r_{i+\frac{1}{2}}+1)\phi'(-1) + \frac{1}{2}(r_{i+\frac{1}{2}}+1)^2\phi''(-1) + \mathcal{O}(r_{i+\frac{1}{2}}+1)^3, \tag{4.8.44}$$

$$\phi(r_{i-\frac{1}{2}}) = \phi(\frac{1}{3}) + (r_{i-\frac{1}{2}}-\frac{1}{3})\phi'(\frac{1}{3}) + \frac{1}{2}(r_{i-\frac{1}{2}}-\frac{1}{3})^2\phi''(\frac{1}{3}) + \mathcal{O}(r_{i-\frac{1}{2}}-\frac{1}{3})^3. \tag{4.8.45}$$

Vervolgens gaan we apart de termen $r_{i+\frac{1}{2}}+1$ en $r_{i-\frac{1}{2}}-\frac{1}{3}$ bekijken

$$r_{i+\frac{1}{2}}+1 = \frac{C_{i+1}-C_i}{C_i-C_{i-1}}+1 = \frac{C_{i+1}-C_{i-1}}{C_i-C_{i-1}} = \frac{\frac{h^3}{3}C_i''' + \mathcal{O}(h^4)}{-\frac{h^2}{2}C_i'' + \frac{h^3}{6}C_i''' + \mathcal{O}(h^4)} = -\frac{2}{3}h\frac{C_i'''}{C_i''} + \mathcal{O}(h^2), \tag{4.8.46}$$

$$r_{i-\frac{1}{2}}-\frac{1}{3} = \frac{C_i-C_{i-1}}{C_{i-1}-C_{i-2}}-\frac{1}{3} = \frac{C_i-\frac{4}{3}C_{i-1}+\frac{1}{3}C_{i-2}}{C_{i-1}-C_{i-2}} = \frac{-\frac{4}{18}h^3C_i''' + \mathcal{O}(h^4)}{-\frac{3}{2}h^2C_i'' + \mathcal{O}(h^3)} = \frac{4}{27}h\frac{C_i'''}{C_i''} + \mathcal{O}(h^2). \tag{4.8.47}$$

Als we de resultaten (4.8.46) en (4.8.47) substitueren in (4.4.17) dan krijgen we na enig rekenen

$$\begin{aligned} C_{i+\frac{1}{2}} &= C_i + \frac{1}{2}\phi(-1)(C_i-C_{i-1}) + \frac{1}{2}\phi'(-1)(C_{i+1}-C_{i-1}) \\ &+ \frac{1}{9}h^2\left(\frac{C_i'''}{C_i''}\right)^2\phi''(-1)(C_i-C_{i-1}) + \mathcal{O}(h^3)(C_i-C_{i-1}). \end{aligned} \tag{4.8.48}$$

Net zoals gedaan is in (4.8.46) en (4.8.47) gaan we weer de termen $C_{i+1}$ en $C_{i-1}$ schrijven in een taylorreeks rond i. Bij elkaar brengen van betreffende termen geeft

$$C_{i+\frac{1}{2}} = C_i - \frac{1}{4}\phi(-1)h^2C_i'' + \frac{1}{6}\left(\frac{1}{2}\phi(-1)+\phi'(-1)\right)h^3C_i''' + \mathcal{O}(h^4). \tag{4.8.49}$$

Als we dezelfde werkwijze herhalen voor $C_{i-\frac{1}{2}}$ dan krijgen we

$$C_{i-\frac{1}{2}} = C_i + \left(\frac{1}{2}-\frac{3}{4}\phi(\frac{1}{3})\right)h^2C_i'' + \left(-\frac{1}{6}+\frac{7}{12}\phi(\frac{1}{3})-\frac{1}{9}\phi'(\frac{1}{3})\right)h^3C_i''' + \mathcal{O}(h^4). \tag{4.8.50}$$

Aftrekken van de termen $C_{i+\frac{1}{2}}$ en $C_{i-\frac{1}{2}}$ geeft tenslotte

$$\begin{aligned} C_{i+\frac{1}{2}} - C_{i-\frac{1}{2}} &= \left(-\frac{1}{4}\phi(-1)-\frac{1}{2}+\frac{3}{4}\phi(\frac{1}{3})\right)h^2C_i'' \\ &+ \frac{1}{3}\left(\frac{1}{4}\phi(-1)+\frac{1}{2}\phi'(-1)+\frac{1}{2}-\frac{7}{4}\phi(\frac{1}{3})+\frac{1}{3}\phi'(\frac{1}{3})\right)h^3C_i''' + \mathcal{O}(h^4). \end{aligned} \tag{4.8.51}$$

We kunnen nu met behulp van (4.8.51) de volgende voorwaarden vinden voor $\mathcal{O}(h^3)$ nauwkeurigheid in gladde extrema

$$3\phi(\frac{1}{3}) - \phi(-1) = 2, \tag{4.8.52a}$$

$$21\phi(\frac{1}{3}) - 4\phi'(\frac{1}{3}) - 3\phi(-1) - 6\phi'(-1) = 6. \tag{4.8.52b}$$

Met behulp van de nieuwe voorwaarde (4.8.52b) zijn we in staat nieuwe limiters te construeren. In deze paragraaf zullen twee limiters worden geconstrueerd. Allereerst zal een limiter worden geconstrueerd waarvoor geldt, net zoals in de vorige paragraaf, $\phi(r)=0$ als $r\le 0$ (maar die nu wel $\mathcal{O}(h^3)$ nauwkeurig is in gladde extrema). Daarna wordt een limiter geconstrueerd die waarden ongelijk aan 0 aanneemt als $r \le 0$ is.

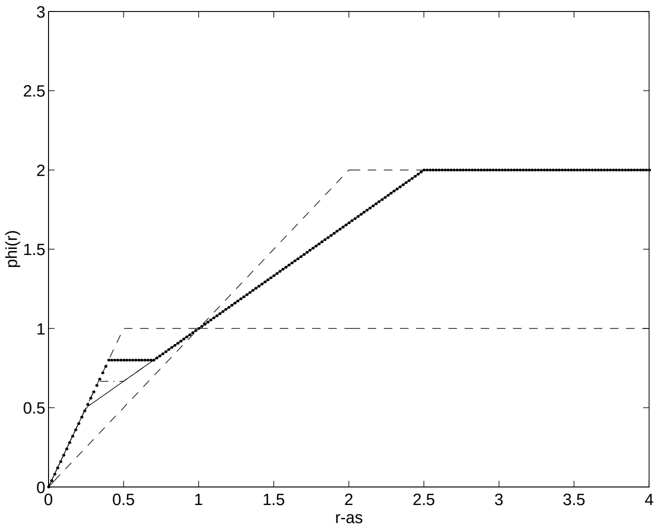
Er geldt $\phi(-1) = 0$. Als we dit invullen in (4.8.52a), dan vinden we $\phi(\frac{1}{3}) = \frac{2}{3}$. Vervolgens beschouwen we (4.8.52b). Substitutie daarin van $\phi(-1) = 0$, $\phi'(-1) = 0$ en $\phi(\frac{1}{3}) = \frac{2}{3}$ levert $\phi'(\frac{1}{3}) = 2$. Als we (4.7.43) beschouwen, dan zien we dat aan deze voorwaarde niet wordt voldaan. Immers in het punt $r = \frac{1}{3}$ heeft (4.7.43) een knik en is daar dus niet differentieerbaar. Bij het construeren van de nieuwe limiter wordt daarom de $2r$-lijn iets langer gevolgd; we volgen tot het (willekeurig) gekozen punt $r = \frac{2}{5}$, waarna we $\phi = \frac{4}{5}$ nemen tot aan de $\phi = \frac{1}{3} + \frac{2}{3}r$ – lijn. Op deze manier wordt wel voldaan aan de voorwaarde $\phi'(\frac{1}{3}) = 2$. De nieuwe limiter komt er dan als volgt uit te zien

$$\phi(r) = \begin{cases} 0, & r < 0 \\ 2r, & 0 \leq r < \frac{2}{5} \\ \frac{4}{5}, & \frac{2}{5} \leq r < \frac{7}{10} \\ \frac{1}{3} + \frac{2}{3}r, & \frac{7}{10} \leq r < \frac{5}{2} \\ 2, & r \geq \frac{5}{2}. \end{cases} \tag{4.8.53}$$

Of in een iets compactere vorm

$$\phi(r) = \max\left(0, \min\left(2r, \frac{4}{5}\right), \min\left(\min\left(\frac{1}{3} + \frac{2}{3}r, 2r\right), 2\right)\right). \tag{4.8.54}$$

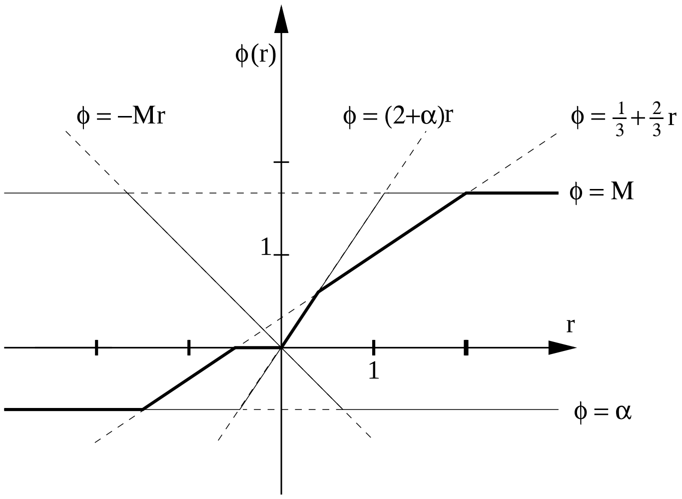
In figuur 4.5 is deze nieuwe limiter geplot tezamen met (4.6.35d) en (4.7.43).




Figuur 4.5: De drie limiters (4.6.35d) (– lijn), (4.7.43) (–· lijn) en (4.8.54) (· · · lijn) in het $\mathcal{O}(h^2)$ TVD gebied van Sweby.

Bij het construeren van de tweede limiter wordt gepoogd de $\kappa = \frac{1}{3}$ – lijn zolang mogelijk te volgen, ook daar waar $r$ negatief wordt. Op de $\kappa = \frac{1}{3}$ – lijn (i.e. $\phi(r) = \frac{1}{3} + \frac{2}{3}r$) geldt $\phi(\frac{1}{3}) = \frac{5}{9}$. Invullen in (4.8.52a) geeft $\phi(-1) = -\frac{1}{3}$. Gemakkelijk valt in te zien dat dit punt ook op de $\kappa = \frac{1}{3}$ – lijn ligt. Vervolgens beschouwen we (4.8.52b), waar achtereenvolgens in wordt gesubstitueerd: $\phi(\frac{1}{3}) = \frac{5}{9}$, $\phi(-1) = -\frac{1}{3}$ en $\phi'(\frac{1}{3}) = \frac{2}{3}$. Hieruit volgt, zoals te verwachten was, $\phi'(-1) = \frac{2}{3}$. De vorm van de te construeren limiter is te zien in figuur 4.6.

De waarde van $\alpha$ is nu bepalend voor de limiter om zowel in Spekreijse's monotoniciteitsgebied te blijven als te voldoen aan de extra nauwkeurigheidseisen. We zouden graag willen dat $\alpha$ een waarde aanneemt zodat het snijpunt van de $\kappa = \frac{1}{3}$ – lijn met de lijn $(2 + \alpha)r$ zover mogelijk naar links ligt, zodat de limiter in het punt $r = \frac{1}{3}$ differentieerbaar is. Tegelijkertijd willen we dat de ondergrens $\alpha$ van Spekreijse's monotoniciteitsgebied een zodanige waarde heeft dat de limiter ook in het punt $r = -1$ differentieerbaar is. We willen dus bij het bepalen van $\alpha$ dat geldt




Figuur 4.6: De vorm van de te construeren nieuwe limiter.

$$\frac{1}{3}+\frac{2}{3}r=(2+\alpha)r \quad \wedge \quad r<\frac{1}{3}. \tag{4.8.55}$$

Oplossen voor $\alpha$ geeft: $\alpha > -\frac{1}{3}$. Dit strijdt met de nauwkeurigheidseis $\phi(-1) = -\frac{1}{3}$; er kan dus niet volledig worden voldaan aan de extra nauwkeurigheidseisen en de TVD-voorwaarden. Toch wordt er zo goed mogelijk aan de nauwkeurigheidseisen voldaan voor $\alpha = -\frac{1}{3}$. We kiezen daarom voor deze waarde van $\alpha$ en verder kiezen we $M = 2$. De limiter komt er dan als volgt uit te zien

$$\phi(r)=\begin{cases} -\frac{1}{3}, & r<-1 \\ \frac{1}{3}+\frac{2}{3}r, & -1\le r<-\frac{1}{2} \\ 0, & -\frac{1}{2}\le r<0 \\ \frac{5}{3}r, & 0\le r<\frac{1}{3} \\ \frac{1}{3}+\frac{2}{3}r, & \frac{1}{3}\le r<\frac{5}{2} \\ 2, & r\ge\frac{5}{2}. \end{cases} \tag{4.8.56}$$

In figuur 4.7 is de limiter te zien in Spekreijse's monotoniciteitsgebied met $\alpha = -\frac{1}{3}$ en $M = 2$.

## 4.9 Numerieke experimenten I

Om nu te kijken of de nieuwe limiters (4.7.43), (4.8.54) en (4.8.56) inderdaad nauwkeurigere resultaten opleveren gaan we deze tezamen met de oude flux-limiter (4.6.35d) en (4.6.35e), voor $\kappa = \frac{1}{3}$, toepassen op het volgende testprobleem:

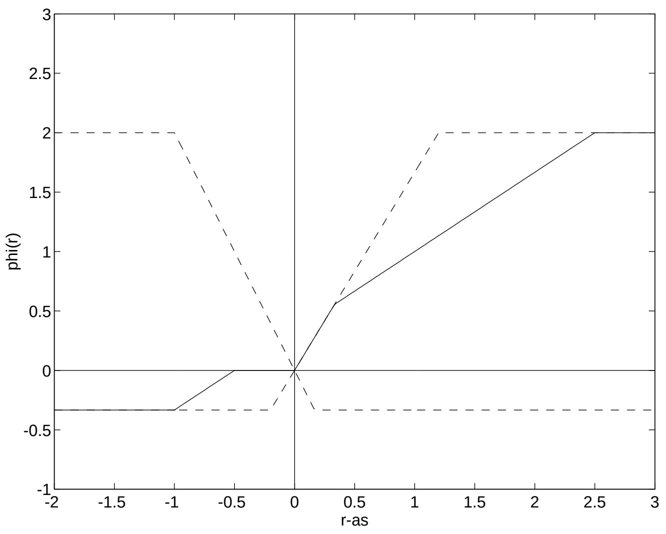
**Vergelijking:**

$$\frac{\partial C}{\partial t}+u\frac{\partial C}{\partial x}=k\frac{\partial^2 C}{\partial x^2}, \quad 0\le x\le 1. \tag{4.9.57}$$

**Beginwaarden:**

$$C(x,0)=\frac{1}{4}\left(1-\cos 2\pi\frac{x-a}{b-a}\right)^2, \ a\le x\le b \quad \text{en} \quad C(x,0)=0 \text{ daarbuiten}, \tag{4.9.58a}$$

$$C(x,0)=\sin\pi\frac{x-a}{b-a}, \ a\le x\le b \quad \text{en} \quad C(x,0)=0 \text{ daarbuiten}, \tag{4.9.58b}$$




Figuur 4.7: De limiter (4.8.56) in Spekreijse's monotoniciteitsgebied ($\alpha = -\frac{1}{3}$, $M = 2$).

$$C(x,0) = 1,\ a \leq x \leq b \quad \text{en} \quad C(x,0) = 0 \text{ daarbuiten.} \tag{4.9.58c}$$

**Randwaarden:**

$$C(0,t) = 0, \quad C(1,t) = 0. \tag{4.9.59}$$

Dit probleem modelleert bijvoorbeeld de stroming en uitsmering van een stofconcentratie of een temperatuurverdeling, in een in zee uitmondende rivier. (De rivier stroomt met snelheid $u$, het rivierwater heeft een moleculaire diffusie- resp. warmtegeleidingscoëfficiënt $k$, $C$ is de stofconcentratie resp. watertemperatuur, en $x = 1$ is de zeemonding.) Het probleem komt uit [17]. Net als daar kiezen we: $u = 1$, $a = 0.2$, $b = 0.4$. Verder nemen we aan $k = 0$. We beschouwen dus een puur advectie probleem.

Bij de discretisatie hebben we gebruik gemaakt van de eindige volume methode, zoals beschreven is in paragraaf 4.1. Verder hebben we bij de discretisatie van de randvoorwaarden de formules gebruikt in [9]. Om de nauwkeurigheid te onderzoeken, hebben we de numerieke oplossingen op tijdstip $t = 0.3$ berekend op een reeks van roosters, waarbij het aantal cellen steeds werd verdubbeld. We beschouwden roosters met $20, 40, 80, \ldots, 640$ cellen. Bij de bepaling van de numerieke fout werd er vergeleken met de exacte oplossing. Voor de tijdsdiscretisatie werd de klassieke 4-stage, $4^e$ orde nauwkeurige Runge-Kutta methode gebruikt. Hierbij werd de tijdstap lineair met de maaswijdte $h$ genomen en klein genoeg gekozen om ervoor te zorgen dat de tijdsdiscretisatie fouten te verwaarlozen waren met de plaatsdiscretisatie fouten.

In de figuren 4.8 t/m 4.13 zijn de grafieken te zien van de exacte en numerieke oplossingen voor het probleem met (4.9.58a) als beginoplossing. (Ik heb alleen deze beginoplossing beschouwd omdat in dat geval het verschil tussen de verschillende limiters het meest zichtbaar is. Ik heb ook de grafieken van de andere twee beginoplossingen bekeken, maar niet erbij gezet omdat er geen nieuwe informatie aan ontleend wordt.) In deze figuren is duidelijk te zien dat het ongelimiteerde ($\kappa = \frac{1}{3}$)-schema het meest nauwkeurig is in de pieken van de oplossing. Het nadeel is dat de oplossing op bepaalde plaatsen negatief wordt, hetgeen we juist willen voorkomen. De grafieken 4.9 t/m 4.13 laten allemaal een numerieke oplossing zien, die nergens foutieve maxima of minima heeft en overal positief is. Het nadeel is dat de oplossingen meer diffusie vertonen dan het ongelimiteerde schema. Dat is de prijs die betaald moet worden voor monotoniciteit en positiviteit. Als we de grafieken 4.9 t/m 4.13 met elkaar vergelijken dan zien we dat (4.6.35e), met $\kappa = \frac{1}{3}$, het meest diffusief is van alle vijf limiters. Verder geeft (4.8.54) de beste resultaten van de vijf limiters. Van de nieuwe geconstrueerde limiters (4.8.54) en (4.8.56) geeft (4.8.56) in figuur 4.13 de minst goede resultaten. De reden hiervoor is waarschijnlijk dat bij de constructie van deze limiter niet goed kon worden voldaan aan de TVD-voorwaarden en de nieuwe nauwkeurigheids eisen.

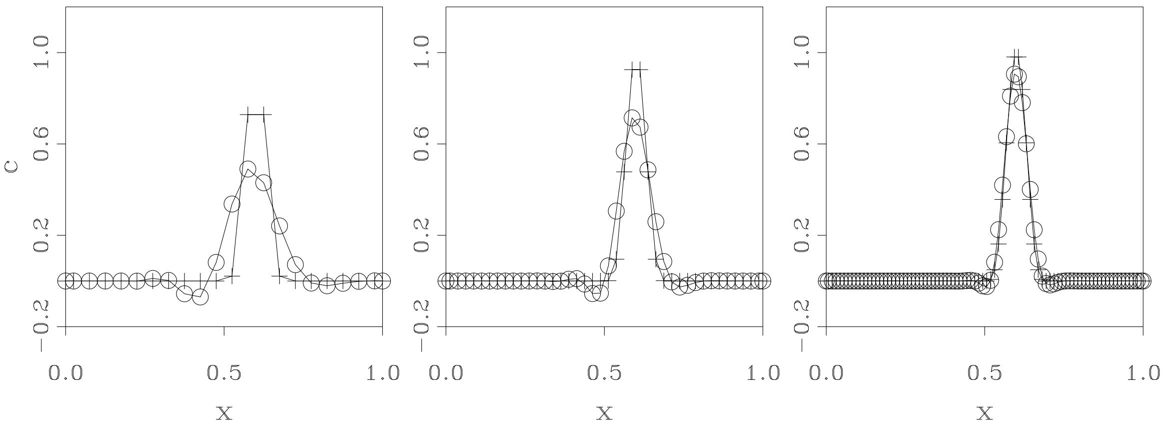
De resultaten voor elk van de drie beginvoorwaarden zijn terug te vinden in de tabellen 4.1, 4.2 en 4.3. Hierbij is voor elk van de zes grids, de $L_1$- en de $L_\infty$-norm van de fout gegeven. Dit is in de tabellen aangegeven met resp. $\|\Delta C\|_1$ en $\|\Delta C\|_\infty$. Verder is in deze tabellen ook het gedrag van $\|\Delta C\|_1$ en $\|\Delta C\|_\infty$ te zien in termen van $\mathcal{O}(h^p)$; de waarden van resp. $p_{\|\Delta C\|_1}$ en $p_{\|\Delta C\|_\infty}$ zijn opgenomen in de tabel.

Als we in deze tabellen louter naar de waarden van $\|\Delta C\|_1$ en $\|\Delta C\|_\infty$ kijken, dan kunnen we weer de conclusie trekken dat (4.8.54) de beste resultaten geeft en dat (4.6.35e) met $\kappa = \frac{1}{3}$ de minst goede resultaten geeft.

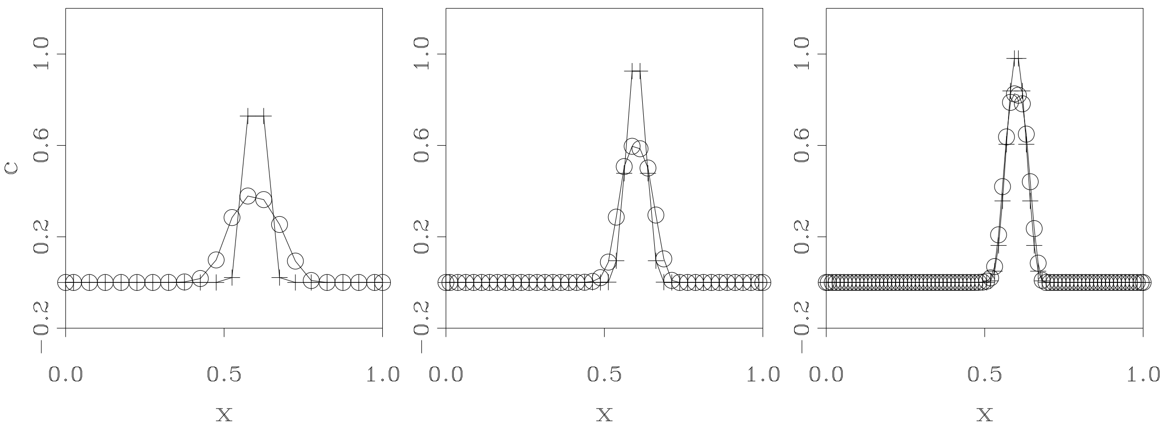
Als we echter gaan kijken hoe $\|\Delta C\|_1$ en $\|\Delta C\|_\infty$ zich gedragen in orde van $h$, dan zal blijken dat hier nieuwe informatie aan ontleend kan worden. Bij de beginoplossing die overal glad is en geen knikken of discontinuïteiten vertoont (formule (4.9.58a)), geeft (4.8.54) weer de beste resultaten. Echter wanneer de beginoplossing knikken of discontinuïteiten gaat vertonen (4.9.58b of 4.9.58c), zal het orde gedrag van (4.8.54) slechter worden. Zie hiervoor tabel 4.2 en 4.3.

Iets anders dat opvalt is dat we geen $\mathcal{O}(h^2)$ of zelfs $\mathcal{O}(h^3)$ gedrag in de gladde extrema terugvinden in onze resultaten. Dit zou te zien moeten zijn in de kolom van $p_{\|\Delta C\|_\infty}$ in tabel 4.1. We zouden daar waarden in de buurt van 2 of 3 hebben verwacht. De reden hiervoor is vermoedelijk terug te vinden in het feit dat de vierde afgeleide van beginoplossing (4.9.58a) discontinu is in $x = a$ en $x = b$. De invloed van deze afgeleide is te voelen wanneer de maaswijdte fijner wordt.

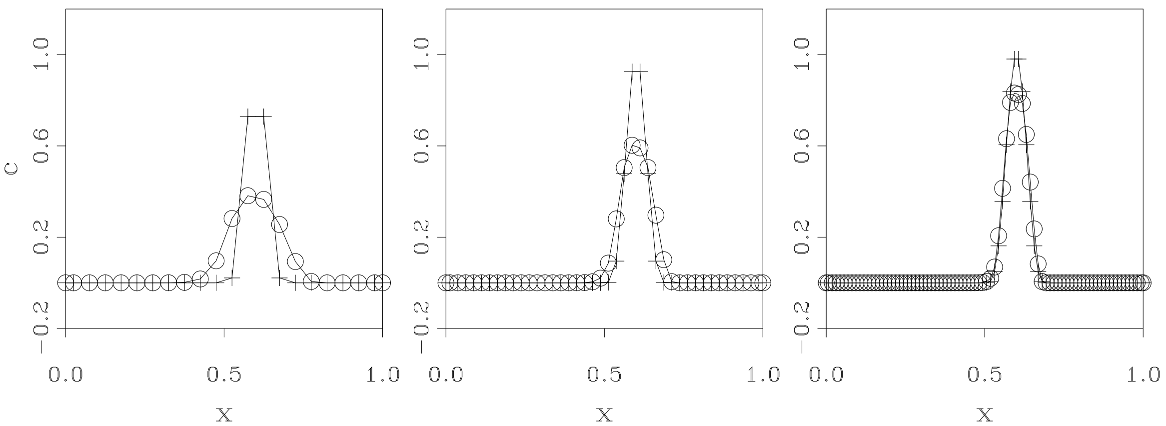
We kunnen nu de volgende conclusies trekken. In het geval van een gladde beginoplossing geven (4.6.35d), (4.7.43) en (4.8.54) de beste resultaten. Hierbij geeft (4.8.54) zelfs nog iets betere resultaten dan de overige twee limiters. We constateerden niet het verwachte $\mathcal{O}(h^2)$ of zelfs $\mathcal{O}(h^3)$ gedrag in gladde extrema, maar deze zullen waarschijnlijk wel naar voren komen als een beginoplossing beschouwd wordt die overal oneindig vaak differentieerbaar is. Dit zal gaan gebeuren in de volgende paragraaf. Al met al kunnen we constateren dat de nieuwe limiters, (4.7.43) en (4.8.54), toch een verbetering opleveren ten opzichte van de oude limiters.



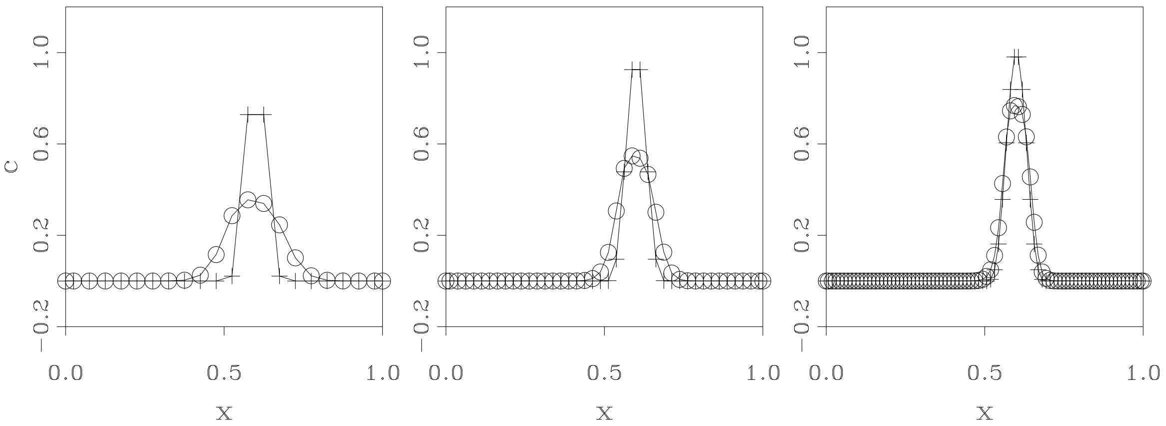

Figuur 4.8: Exacte(+) en numerieke(∘) oplossing m.b.v. ongelimiteerde $\kappa = \frac{1}{3}$ limiter voor 20, 40 en 80 cellen.



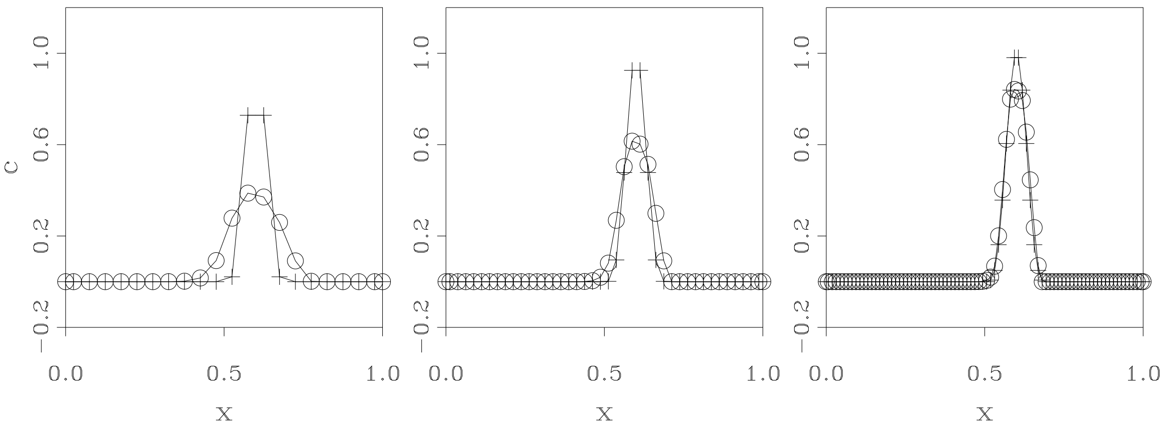

Figuur 4.9: Exacte(+) en numerieke(∘) oplossing m.b.v. (4.6.35d) voor 20, 40 en 80 cellen.



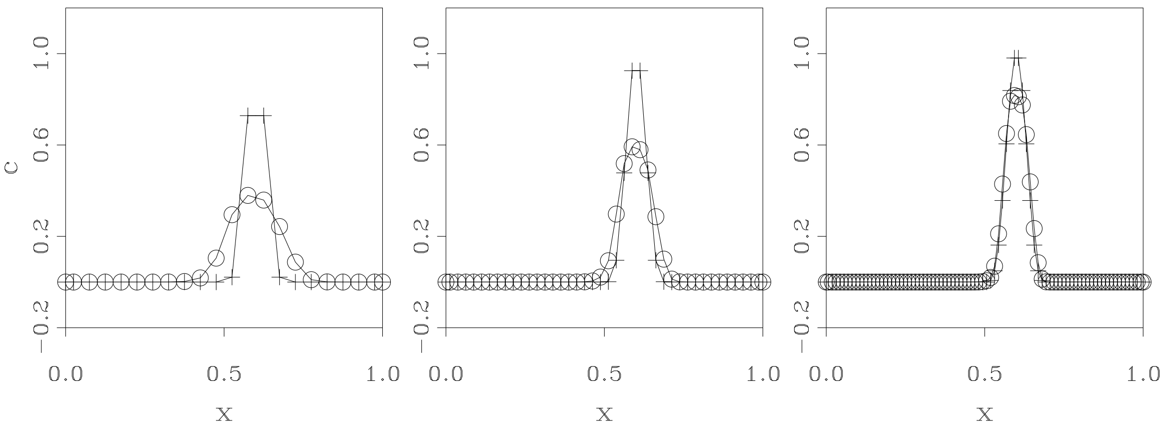

Figuur 4.10: Exacte(+) en numerieke(∘) oplossing m.b.v. (4.7.43) voor 20, 40 en 80 cellen.




Figuur 4.11: Exacte(+) en numerieke($\circ$) oplossing m.b.v. (4.6.35e) (met $\kappa = \frac{1}{3}$) voor 20, 40 en 80 cellen.




Figuur 4.12: Exacte(+) en numerieke($\circ$) oplossing m.b.v. (4.8.54) voor 20, 40 en 80 cellen.




Figuur 4.13: Exacte(+) en numerieke($\circ$) oplossing m.b.v. (4.8.56) voor 20, 40 en 80 cellen.

Tabel 4.1: Vergelijking van de limiters met (4.9.58a) als beginoplossing.

| Aantal cellen | Flux-limiter | $\|\Delta C\|_1$ | $p_{\|\Delta C\|_1}$ | $\|\Delta C\|_\infty$ | $p_{\|\Delta C\|_\infty}$ |
|---|---|---|---|---|---|
| 20 | (4.6.35d) | 0.716E-01 | * | 0.366E+00 | * |
| | (4.6.35e) met $\kappa = \frac{1}{3}$ | 0.762E-01 | * | 0.389E+00 | * |
| | (4.7.43) | 0.710E-01 | * | 0.363E+00 | * |
| | (4.8.54) | 0.699E-01 | * | 0.358E+00 | * |
| | (4.8.56) | 0.719E-01 | * | 0.370E+00 | * |
| 40 | (4.6.35d) | 0.334E-01 | 1.098 | 0.340E+00 | 0.106 |
| | (4.6.35e) met $\kappa = \frac{1}{3}$ | 0.388E-01 | 0.972 | 0.388E+00 | 0.005 |
| | (4.7.43) | 0.328E-01 | 1.113 | 0.334E+00 | 0.120 |
| | (4.8.54) | 0.317E-01 | 1.142 | 0.323E+00 | 0.147 |
| | (4.8.56) | 0.340E-01 | 1.082 | 0.346E+00 | 0.097 |
| 80 | (4.6.35d) | 0.106E-01 | 1.662 | 0.161E+00 | 1.078 |
| | (4.6.35e) met $\kappa = \frac{1}{3}$ | 0.158E-01 | 1.295 | 0.218E+00 | 0.834 |
| | (4.7.43) | 0.102E-01 | 1.691 | 0.155E+00 | 1.104 |
| | (4.8.54) | 0.943E-02 | 1.748 | 0.146E+00 | 1.146 |
| | (4.8.56) | 0.111E-01 | 1.613 | 0.169E+00 | 1.029 |
| 160 | (4.6.35d) | 0.227E-02 | 2.221 | 0.572E-01 | 1.493 |
| | (4.6.35e) met $\kappa = \frac{1}{3}$ | 0.414E-02 | 1.936 | 0.906E-01 | 1.264 |
| | (4.7.43) | 0.216E-02 | 2.237 | 0.540E-01 | 1.524 |
| | (4.8.54) | 0.198E-02 | 2.249 | 0.491E-01 | 1.572 |
| | (4.8.56) | 0.254E-02 | 2.129 | 0.638E-01 | 1.409 |
| 320 | (4.6.35d) | 0.435E-03 | 2.382 | 0.188E-01 | 1.605 |
| | (4.6.35e) met $\kappa = \frac{1}{3}$ | 0.106E-02 | 1.969 | 0.357E-01 | 1.344 |
| | (4.7.43) | 0.402E-03 | 2.421 | 0.175E-01 | 1.623 |
| | (4.8.54) | 0.360E-03 | 2.460 | 0.156E-01 | 1.659 |
| | (4.8.56) | 0.518E-03 | 2.291 | 0.219E-01 | 1.541 |
| 640 | (4.6.35d) | 0.849E-04 | 2.356 | 0.620E-02 | 1.600 |
| | (4.6.35e) met $\kappa = \frac{1}{3}$ | 0.235E-03 | 2.171 | 0.135E-01 | 1.400 |
| | (4.7.43) | 0.774E-04 | 2.378 | 0.577E-02 | 1.606 |
| | (4.8.54) | 0.652E-04 | 2.467 | 0.503E-02 | 1.629 |
| | (4.8.56) | 0.105E-03 | 2.300 | 0.744E-02 | 1.560 |

Tabel 4.2: Vergelijking van de limiters met (4.9.58b) als beginoplossing.

| Aantal cellen | Flux-limiter | $\|\Delta C\|_1$ | $p_{\|\Delta C\|_1}$ | $\|\Delta C\|_\infty$ | $p_{\|\Delta C\|_\infty}$ |
|---|---|---|---|---|---|
| 20 | (4.6.35d) | 0.612E-01 | * | 0.315E+00 | * |
| | (4.6.35e) met $\kappa = \frac{1}{3}$ | 0.688E-01 | * | 0.354E+00 | * |
| | (4.7.43) | 0.603E-01 | * | 0.311E+00 | * |
| | (4.8.54) | 0.584E-01 | * | 0.303E+00 | * |
| | (4.8.56) | 0.617E-01 | * | 0.320E+00 | * |
| 40 | (4.6.35d) | 0.184E-01 | 1.732 | 0.132E+00 | 1.257 |
| | (4.6.35e) met $\kappa = \frac{1}{3}$ | 0.267E-01 | 1.366 | 0.182E+00 | 0.959 |
| | (4.7.43) | 0.180E-01 | 1.745 | 0.126E+00 | 1.298 |
| | (4.8.54) | 0.169E-01 | 1.790 | 0.117E+00 | 1.370 |
| | (4.8.56) | 0.186E-01 | 1.732 | 0.139E+00 | 1.200 |
| 80 | (4.6.35d) | 0.523E-02 | 1.816 | 0.440E-01 | 1.584 |
| | (4.6.35e) met $\kappa = \frac{1}{3}$ | 0.811E-02 | 1.719 | 0.688E-01 | 1.403 |
| | (4.7.43) | 0.507E-02 | 1.825 | 0.441E-01 | 1.519 |
| | (4.8.54) | 0.484E-02 | 1.805 | 0.444E-01 | 1.399 |
| | (4.8.56) | 0.560E-02 | 1.728 | 0.440E-01 | 1.663 |
| 160 | (4.6.35d) | 0.185E-02 | 1.498 | 0.264E-01 | 0.735 |
| | (4.6.35e) met $\kappa = \frac{1}{3}$ | 0.361E-02 | 1.167 | 0.443E-01 | 0.635 |
| | (4.7.43) | 0.178E-02 | 1.514 | 0.264E-01 | 0.738 |
| | (4.8.54) | 0.158E-02 | 1.610 | 0.265E-01 | 0.747 |
| | (4.8.56) | 0.198E-02 | 1.502 | 0.264E-01 | 0.734 |
| 320 | (4.6.35d) | 0.572E-03 | 1.694 | 0.164E-01 | 0.685 |
| | (4.6.35e) met $\kappa = \frac{1}{3}$ | 0.118E-02 | 1.619 | 0.279E-01 | 0.667 |
| | (4.7.43) | 0.555E-03 | 1.678 | 0.164E-01 | 0.685 |
| | (4.8.54) | 0.505E-03 | 1.651 | 0.164E-01 | 0.686 |
| | (4.8.56) | 0.602E-03 | 1.718 | 0.164E-01 | 0.685 |
| 640 | (4.6.35d) | 0.186E-03 | 1.618 | 0.104E-01 | 0.665 |
| | (4.6.35e) met $\kappa = \frac{1}{3}$ | 0.390E-03 | 1.590 | 0.174E-01 | 0.684 |
| | (4.7.43) | 0.183E-03 | 1.604 | 0.104E-01 | 0.665 |
| | (4.8.54) | 0.169E-03 | 1.577 | 0.104E-01 | 0.665 |
| | (4.8.56) | 0.193E-03 | 1.639 | 0.104E-01 | 0.665 |

Tabel 4.3: Vergelijking van de limiters met (4.9.58c) als beginoplossing.

| Aantal cellen | Flux-limiter | $\|\Delta C\|_1$ | $p_{\|\Delta C\|_1}$ | $\|\Delta C\|_\infty$ | $p_{\|\Delta C\|_\infty}$ |
|---|---|---|---|---|---|
| 20 | (4.6.35d) | 0.878E-01 | * | 0.357E+00 | * |
| | (4.6.35e) met $\kappa = \frac{1}{3}$ | 0.996E-01 | * | 0.363E+00 | * |
| | (4.7.43) | 0.867E-01 | * | 0.359E+00 | * |
| | (4.8.54) | 0.842E-01 | * | 0.362E+00 | * |
| | (4.8.56) | 0.887E-01 | * | 0.351E+00 | * |
| 40 | (4.6.35d) | 0.500E-01 | 0.813 | 0.363E+00 | -2.072E-02 |
| | (4.6.35e) met $\kappa = \frac{1}{3}$ | 0.578E-01 | 0.786 | 0.374E+00 | -4.131E-02 |
| | (4.7.43) | 0.495E-01 | 0.808 | 0.364E+00 | -2.194E-02 |
| | (4.8.54) | 0.482E-01 | 0.806 | 0.369E+00 | -2.924E-02 |
| | (4.8.56) | 0.506E-01 | 0.809 | 0.361E+00 | -4.230E-02 |
| 80 | (4.6.35d) | 0.297E-01 | 0.751 | 0.384E+00 | -8.172E-02 |
| | (4.6.35e) met $\kappa = \frac{1}{3}$ | 0.352E-01 | 0.714 | 0.399E+00 | -9.354E-02 |
| | (4.7.43) | 0.294E-01 | 0.750 | 0.385E+00 | -8.082E-02 |
| | (4.8.54) | 0.286E-01 | 0.752 | 0.390E+00 | -8.105E-02 |
| | (4.8.56) | 0.300E-01 | 0.754 | 0.382E+00 | -8.108E-02 |
| 160 | (4.6.35d) | 0.176E-01 | 0.759 | 0.401E+00 | -6.258E-02 |
| | (4.6.35e) met $\kappa = \frac{1}{3}$ | 0.213E-01 | 0.724 | 0.421E+00 | -7.837E-02 |
| | (4.7.43) | 0.174E-01 | 0.757 | 0.402E+00 | -6.136E-02 |
| | (4.8.54) | 0.169E-01 | 0.758 | 0.407E+00 | -6.047E-02 |
| | (4.8.56) | 0.177E-01 | 0.762 | 0.400E+00 | -6.457E-02 |
| 320 | (4.6.35d) | 0.104E-01 | 0.762 | 0.415E+00 | -5.002E-02 |
| | (4.6.35e) met $\kappa = \frac{1}{3}$ | 0.128E-01 | 0.731 | 0.440E+00 | -6.182E-02 |
| | (4.7.43) | 0.103E-01 | 0.760 | 0.416E+00 | -4.897E-02 |
| | (4.8.54) | 0.100E-01 | 0.758 | 0.421E+00 | -4.747E-02 |
| | (4.8.56) | 0.104E-01 | 0.764 | 0.414E+00 | -5.136E-02 |
| 640 | (4.6.35d) | 0.611E-02 | 0.761 | 0.427E+00 | -4.157E-02 |
| | (4.6.35e) met $\kappa = \frac{1}{3}$ | 0.771E-02 | 0.736 | 0.455E+00 | -4.938E-02 |
| | (4.7.43) | 0.608E-02 | 0.760 | 0.428E+00 | -4.070E-02 |
| | (4.8.54) | 0.592E-02 | 0.757 | 0.432E+00 | -3.883E-02 |
| | (4.8.56) | 0.614E-02 | 0.764 | 0.426E+00 | -4.247E-02 |

## 4.10 Numerieke experimenten II

In de vorige paragraaf is bij het numerieke experiment het $\mathcal{O}(h^2)$ en $\mathcal{O}(h^3)$ gedrag in de gladde extrema uitgebleven. Theorie en praktijk kwamen niet overeen met elkaar. Als oorzaak werd gegeven dat de (gladde) beginoplossing niet oneindig vaak differentieerbaar was. De vierde afgeleide van (4.9.58a) bleek een knik te bevatten. Om voor dit probleem een oplossing te vinden, gaan we weer het testprobleem toepassen maar nu met een beginoplossing die overal oneindig vaak differentieerbaar is. We nemen de volgende beginoplossing

$$C(x,0) = \frac{1}{2}\left(\sin(2\pi x) + 1\right), \quad 0 \leq x \leq 1. \tag{4.10.60}$$

Voordat het testprobleem met de nieuwe beginoplossing wordt toegepast op de vijf limiters, wordt nog een verandering doorgevoerd. In [9] wordt opgemerkt dat het flux-gelimite schema niet goed toepasbaar is op de randen van het rekengebied. Om ervoor te zorgen dat de discretisatie aan de randen niet van enige invloed is op het $\mathcal{O}(h^2)$ of $\mathcal{O}(h^3)$ gedrag in de gladde extrema, gaan we het probleem periodiek maken. Dit houdt in dat de randen van het rekengebied aan elkaar worden geknoopt. Dus bijvoorbeeld celwand $\partial\Omega_{N+\frac{1}{2}}$ correspondeert dan met $\partial\Omega_{\frac{1}{2}}$. Op deze manier kan steeds zonder problemen het flux-gelimite schema worden toegepast.

Een ander voordeel van het periodiek maken van het probleem is meer van praktische aard. Bij onderzoek naar stabiliteit van tijdstapschema's is het vaak noodzakelijk om over lange tijden te integreren. Door het periodiek maken kun je in principe oneindig lang gaan integreren in de tijd door steeds maar rondjes te blijven lopen. Dit komt nog aan de orde bij tijdsintegratie.

In tabel 4.4 zijn de resultaten te zien van dit numerieke experiment. Het eerste dat opvalt als we de resultaten vergelijken met de tabellen 4.1 t/m 4.3 (en met name met tabel 4.1) is dat de numerieke fouten gemeten in de $L_1$- en de $L_\infty$-norm in tabel 4.4 minstens een factor 10 kleiner zijn. Aangezien we hier te maken hebben met een enorm gladde beginoplossing, is dit niet zo verwonderlijk. We zien dat elke limiter zijn werk goed doet met als kleine uitschieter de limiter in (4.8.54). Als we naar het orde gedrag kijken van $\|\Delta C\|_1$ en $\|\Delta C\|_\infty$, dan geeft (4.8.54) weer de beste resultaten. Echter het verwachte $\mathcal{O}(h^2)$ of $\mathcal{O}(h^3)$ gedrag in de gladde extrema blijft weer uit. We zien zelfs als we de resultaten vergelijken met tabel 4.1 dat het orde gedrag zelfs verminderd is. Omdat ik dit zeer verwonderlijk vond heb ik ter controle het numerieke experiment ook uitgevoerd voor het ongelimiteerde ($\kappa = \frac{1}{3}$)-schema. Deze resultaten zijn ook opgenomen in tabel 4.4. We zien dat dit schema de beste resultaten geeft van alle schema's. Dit konden we verwachten aangezien de beginoplossing overal enorm glad is. Bij het toepassen van het ($\kappa = \frac{1}{3}$)-schema treden dan ook nergens valse niet-monotoniciteiten op. Zoals reeds is aangetoond is dit schema $\mathcal{O}(h^3)$ nauwkeurig, hetgeen direct is terug te vinden in de tabel.

De reden voor het grote verschil in orde gedrag is waarschijnlijk terug te vinden in het feit dat bij de constructie van (4.8.54) en (4.8.56) uitgegaan is van de situatie dat het gladde extremum precies in een cel-centrum ligt, de eveneens mogelijke situatie van een glad extremum op een celwandje is (nog) niet beschouwd.

In tabel 4.4 zijn ook de resultaten terug te vinden voor het ($\kappa = \frac{1}{2}$)-schema (QUICK), zowel voor het ongelimiteerde als het gelimiteerde geval. In het gelimiteerde geval is de limiter in (4.8.54) gebruikt, maar dan omgeschreven voor het $\kappa = \frac{1}{2}$-schema. We krijgen dan de volgende formule

$$\phi(r) = \max\left(0, \min\left(2r, \frac{4}{5}\right), \min\left(\min\left(\frac{1}{4} + \frac{3}{4}r, 2r\right), 2\right)\right). \tag{4.10.61}$$

Zoals duidelijk uit de tabel blijkt, is het ($\kappa = \frac{1}{2}$)-schema, zowel met als zonder limiter $\mathcal{O}(h^2)$ nauwkeurig. Dit komt precies overeen met de theorie zoals uiteen is gezet in paragraaf 4.2. We zien dat ondanks het feit dat QUICK $\mathcal{O}(h^3)$ nauwkeurig is voor de berekening van de flux op een celwandje, de plaatsdiscretisatie, hetgeen waar wij in geïnteresseerd zijn, toch $\mathcal{O}(h^2)$ nauwkeurig is.

Zoals reeds is opgemerkt hebben we bij de numerieke experimenten, de tijdstapgrootte $\tau$ op empirische wijze zo klein gemaakt, dat de tijdsdiscretisatie fout kleiner was dan de plaatsdiscretisatie fout. Verder hebben we bij de keuze van de tijdstapgrootte, op emperische wijze monotoniciteit en stabiliteit gerealiseerd. Het probleem echter is, dat we graag een waarde $\tau_0$ zouden willen hebben, zodanig dat we op mathematische wijze stabiliteit kunnen *garanderen* voor $\tau \leq \tau_0$. Het vervolg van het onderzoek zal zijn om deze waarde te gaan bepalen voor een aantal tijdstapschema's, waaronder een door ons nieuw geconstrueerd RK4 schema. Bij het zoeken van deze waarde zal bij de plaatsdiscretisatie steeds het flux-gelimite schema met als flux-limiter (4.8.54) gebruikt worden, omdat deze limiter bij al onze numerieke experimenten steeds de kleinste numerieke fout gaf.

Tabel 4.4: Vergelijking van de limiters met (4.10.60) als beginoplossing.

| Aantal cellen | Flux-limiter | $\|\Delta C\|_1$ | $p_{\|\Delta C\|_1}$ | $\|\Delta C\|_\infty$ | $p_{\|\Delta C\|_\infty}$ |
|---|---|---|---|---|---|
| 20 | (4.6.35d) | 0.732E-02 | * | 0.230E-01 | * |
| | (4.6.35e) met $\kappa = \frac{1}{3}$ | 0.139E-01 | * | 0.332E-01 | * |
| | (4.7.43) | 0.688E-02 | * | 0.219E-01 | * |
| | (4.8.54) | 0.644E-02 | * | 0.203E-01 | * |
| | (4.8.56) | 0.815E-02 | * | 0.247E-01 | * |
| | $\kappa = \frac{1}{3}$ schema | 0.153E-02 | * | 0.241E-02 | * |
| | $\kappa = \frac{1}{2}$ schema | 0.283E-02 | * | 0.445E-02 | * |
| | (4.10.61) | 0.709E-02 | * | 0.202E-01 | * |
| 40 | (4.6.35d) | 0.173E-02 | 2.085 | 0.836E-02 | 1.462 |
| | (4.6.35e) met $\kappa = \frac{1}{3}$ | 0.344E-02 | 2.012 | 0.133E-01 | 1.324 |
| | (4.7.43) | 0.158E-02 | 2.119 | 0.786E-02 | 1.481 |
| | (4.8.54) | 0.147E-02 | 2.131 | 0.714E-02 | 1.508 |
| | (4.8.56) | 0.201E-02 | 2.022 | 0.926E-02 | 1.413 |
| | $\kappa = \frac{1}{3}$ schema | 0.193E-03 | 2.984 | 0.304E-03 | 2.987 |
| | $\kappa = \frac{1}{2}$ schema | 0.640E-03 | 2.144 | 0.101E-02 | 2.142 |
| | (4.10.61) | 0.192E-02 | 1.884 | 0.715E-02 | 1.501 |
| 80 | (4.6.35d) | 0.375E-03 | 2.204 | 0.298E-02 | 1.488 |
| | (4.6.35e) met $\kappa = \frac{1}{3}$ | 0.814E-03 | 2.077 | 0.514E-02 | 1.365 |
| | (4.7.43) | 0.334E-03 | 2.246 | 0.278E-02 | 1.498 |
| | (4.8.54) | 0.299E-03 | 2.298 | 0.249E-02 | 1.517 |
| | (4.8.56) | 0.451E-03 | 2.154 | 0.340E-02 | 1.446 |
| | $\kappa = \frac{1}{3}$ schema | 0.242E-04 | 2.996 | 0.380E-04 | 2.997 |
| | $\kappa = \frac{1}{2}$ schema | 0.156E-03 | 2.040 | 0.245E-03 | 2.042 |
| | (4.10.61) | 0.476E-03 | 2.014 | 0.253E-02 | 1.499 |
| 160 | (4.6.35d) | 0.769E-04 | 2.284 | 0.104E-02 | 1.513 |
| | (4.6.35e) met $\kappa = \frac{1}{3}$ | 0.184E-03 | 2.147 | 0.196E-02 | 1.392 |
| | (4.7.43) | 0.689E-04 | 2.276 | 0.970E-03 | 1.519 |
| | (4.8.54) | 0.576E-04 | 2.376 | 0.861E-03 | 1.535 |
| | (4.8.56) | 0.947E-04 | 2.253 | 0.122E-02 | 1.480 |
| | $\kappa = \frac{1}{3}$ schema | 0.303E-05 | 2.999 | 0.477E-05 | 2.996 |
| | $\kappa = \frac{1}{2}$ schema | 0.386E-04 | 2.010 | 0.607E-04 | 2.011 |
| | (4.10.61) | 0.116E-03 | 2.042 | 0.107E-02 | 1.245 |
| 320 | (4.6.35d) | 0.148E-04 | 2.378 | 0.359E-03 | 1.539 |
| | (4.6.35e) met $\kappa = \frac{1}{3}$ | 0.420E-04 | 2.130 | 0.736E-03 | 1.412 |
| | (4.7.43) | 0.133E-04 | 2.371 | 0.333E-03 | 1.541 |
| | (4.8.54) | 0.112E-04 | 2.356 | 0.293E-03 | 1.557 |
| | (4.8.56) | 0.185E-04 | 2.356 | 0.427E-03 | 1.512 |
| | $\kappa = \frac{1}{3}$ schema | 0.378E-06 | 3.000 | 0.598E-06 | 2.995 |
| | $\kappa = \frac{1}{2}$ schema | 0.964E-05 | 2.003 | 0.151E-04 | 2.003 |
| | (4.10.61) | 0.275E-04 | 2.073 | 0.388E-03 | 1.458 |
| 640 | (4.6.35d) | 0.286E-05 | 2.372 | 0.122E-03 | 1.562 |
| | (4.6.35e) met $\kappa = \frac{1}{3}$ | 0.931E-05 | 2.174 | 0.274E-03 | 1.428 |
| | (4.7.43) | 0.255E-05 | 2.387 | 0.113E-03 | 1.563 |
| | (4.8.54) | 0.207E-05 | 2.443 | 0.981E-04 | 1.577 |
| | (4.8.56) | 0.371E-05 | 2.319 | 0.147E-03 | 1.539 |
| | $\kappa = \frac{1}{3}$ schema | 0.473E-07 | 3.000 | 0.751E-07 | 2.994 |
| | $\kappa = \frac{1}{2}$ schema | 0.241E-05 | 2.001 | 0.379E-05 | 2.001 |
| | (4.10.61) | 0.653E-05 | 2.073 | 0.159E-03 | 1.289 |

# Hoofdstuk 5

# Tijdsdiscretisatie

In het voorafgaande hebben we ons alleen nog maar bezig gehouden met de plaatsdiscretisatie. Op deze manier hebben we het volgende semi-gediscretiseerde stelsel differentiaalvergelijkingen verkregen

$$\frac{dC_i}{dt} = -\frac{u}{h}\left(1 + \frac{1}{2}\phi(r_{i+\frac{1}{2}}) - \frac{\phi(r_{i-\frac{1}{2}})}{2r_{i-\frac{1}{2}}}\right)(C_i - C_{i-1}). \tag{5.0.1}$$

De volgende stap die nu zal worden genomen is een discretisatie van het linkerlid in de tijd, waarbij tevens besloten moet worden op welke tijdstap het rechterlid geëvalueerd wordt. Deze methode bij het oplossen van tijdsafhankelijke problemen, waarbij eerst een discretisatie naar plaats wordt gemaakt en daarna naar de tijd, wordt de *methode der lijnen* genoemd.

Laten we eens opnieuw naar het $\mathcal{O}(h^2)$ centrale schema kijken

$$\frac{dC_i}{dt} = -\frac{u}{2h}\left(C_{i+1} - C_{i-1}\right). \tag{5.0.2}$$

De convectie term is hier gediscretiseerd met behulp van centrale differenties. We zouden het linkerlid van (5.0.2) kunnen gaan discretiseren met voorwaartse differenties. Verder wordt het rechterlid geëvalueerd op tijdstap $n$. We krijgen dan

$$\frac{C_i^{n+1} - C_i^n}{\tau} = -\frac{u}{2h}\left(C_{i+1}^n - C_{i-1}^n\right). \tag{5.0.3}$$

Dit wordt de methode van Euler genoemd. Als we verder naar (5.0.3) kijken, dan zien we dat voor iedere $i$ er één onbekende is op tijdstap $n+1$. Een schema met zo'n eigenschap wordt ook wel een *expliciet* schema genoemd. In dit geval wordt het schema in (5.0.2) ook wel *Euler expliciet* genoemd.

Als het rechterlid in (5.0.2) op tijdstip $n+1$ wordt geëvalueerd, dan wordt het schema in (5.0.2) *Euler impliciet* genoemd

$$\frac{C_i^{n+1} - C_i^n}{\tau} = -\frac{u}{2h}\left(C_{i+1}^{n+1} - C_{i-1}^{n+1}\right). \tag{5.0.4}$$

We zien dat voor iedere $i$ er drie onbekenden zijn op tijdstip $n+1$. Vergelijking (5.0.4) leidt tot een stelsel vergelijkingen met een tridiagonale matrix, waarvoor in de literatuur voldoende efficiente oplosmethoden bekend zijn (zie bijvoorbeeld [5], hoofdstuk 12). Impliciete tijdstapschema's zullen verder niet in dit verslag aan de orde komen.

Laten we eens verder kijken naar (5.0.3). We schrijven (5.0.3) om naar

$$C_i^{n+1} = C_i^n - \frac{\nu}{2}\left(C_{i+1}^n - C_{i-1}^n\right), \quad \text{met} \quad \nu = \frac{u\tau}{h}. \tag{5.0.5}$$

De parameter $\nu$ wordt ook wel het *Courant getal* genoemd. Als we nu een experiment zouden gaan doen met dit schema, met bijbehorende begin- en randvoorwaarden, dan zou bij de keuze van $\nu = 1$ de numerieke oplossing na een aantal tijdstappen in toenemende mate gaan verschillen van de exacte oplossing. Het blijkt dat dit schema *onstabiel* is, zelfs voor willekeurige waarden van $\nu$. Zo'n schema noemen we *onvoorwaardelijk onstabiel.* Hoe dit komt zal ter sprake komen als we het zullen gaan hebben over *stabiliteit.*

## 5.1 Enige definities aangaande stabiliteit

Voordat we het gaan hebben over stabiliteit en enige methodes gaan behandelen om stabiliteit te onderzoeken, zullen we eerst enige definities aangaande stabiliteit beschouwen. Een mogelijke definitie voor stabiliteit is dat fouten in het numerieke schema niet onbegrensd mogen groeien, zoals gebeurt bij het schema in (5.0.5), als opeenvolgende tijdstappen worden genomen. Als we de numerieke fout $\varepsilon$ definieren als het verschil tussen de numerieke en de exacte oplossing van het gediscretiseerde probleem

$$\varepsilon_i^n = C_i^n - \overline{C}_i^n, \tag{5.1.6}$$

dan is een mogelijke stabiliteitsconditie

$$\lim_{n\to\infty} |\varepsilon_i^n| \leq K \quad \text{bij vaste } \tau. \tag{5.1.7}$$

Hierbij is $K$ onafhankelijk van $n$. Een nadeel van deze stabiliteitsconditie is dat deze niet kan garanderen dat de fout niet te groot wordt op vaste tussengelegen tijdstappen. Door nu als stabiliteitsconditie *absolute stabiliteit* te gebruiken zijn we in staat dit probleem te omzeilen (zie [20], paragraaf 4.6).

**Definitie 5.1** *We noemen een schema* absoluut stabiel *als er een constante $K > 0$ bestaat (onafhankelijk van $n$ en $e^0$) zodanig dat*

$$\|e^n\| \leq K\|e^0\|, \quad \forall n, \quad \forall e^0, \tag{5.1.8}$$

bij een geschikt gekozen norm. In deze definitie stelt $e^n$ de fout, $\varepsilon_i^n$, voor geschreven in een kolomvector

$$e^n = \begin{pmatrix} \varepsilon_1^n \\ \varepsilon_2^n \\ \vdots \\ \varepsilon_i^n \\ \vdots \end{pmatrix}. \tag{5.1.9}$$

Waar het nu in essentie om gaat bij absolute stabiliteit is dat, bij willekeurige, maar vaste maaswijdten en tijdstap, beginstoringen niet onbegrensd kunnen toenemen in de loop van de berekening (ook als $n \to \infty$).

Een andere stabiliteitsconditie in [20], paragraaf 4.6, heeft als essentie dat bij een berekening tot een vast tijdstip $T$, initieel aanwezige storingen niet onbegrensd kunnen toenemen als maasverfijning en/of tijdstapverkleining wordt toegepast. Dit wordt *nulstabiliteit* genoemd. In dit verslag zal alleen absolute stabiliteit worden beschouwd en als er gesproken wordt over stabiliteit zal daar altijd absolute stabiliteit mee worden bedoeld.

In de volgende paragrafen zullen drie methoden worden behandeld om stabiliteit te onderzoeken. Het zal blijken dat één hiervan de meeste voordelen bevat en deze zal dan ook worden gebruikt bij het stabiliteitsonderzoek van ons TVD-schema met de flux-limiter.

## 5.2 Stabiliteitsanalyse met behulp van de Vervangende Vergelijking

In paragraaf 4.2 is de *vervangende vergelijking* geïntroduceerd om de nauwkeurigheid van de plaatsdiscretisatie te onderzoeken. Het is ook mogelijk om met behulp van de vervangende vergelijking de stabiliteit te onderzoeken bij de gebruikte discretisatie in de plaats en tijd.

We beschouwen weer de pure convectie vergelijking, waarbij voor de tijdsdiscretisatie Forward Euler is gebruikt en voor de plaatsdiscretisatie $\mathcal{O}(h^2)$ centraal (zie (5.0.5)). Vervolgens gaan we zowel in de plaats als in de tijd tayloriseren. Er geldt

$$\begin{aligned} C_i^{n+1} &= C_i^n + \tau (C_t)_i^n + \frac{\tau^2}{2}(C_{tt})_i^n + \frac{\tau^3}{6}(C_{ttt})_i^n + \cdots \\ C_{i+1}^n &= C_i^n + h(C_x)_i^n + \frac{h^2}{2}(C_{xx})_i^n + \frac{h^3}{6}(C_{xxx})_i^n + \cdots \\ C_{i-1}^n &= C_i^n - h(C_x)_i^n + \frac{h^2}{2}(C_{xx})_i^n - \frac{h^3}{6}(C_{xxx})_i^n + \cdots \end{aligned} \tag{5.2.10}$$

Alles substitueren in (5.0.5) geeft

$$(C_t + uC_x)_i^n = -\frac{\tau}{2}(C_{tt})_i^n - \frac{uh^2}{6}(C_{xxx})_i^n + \mathcal{O}(\tau^2, h^4). \tag{5.2.11}$$

Vervolgens gaan we proberen om de afgeleiden naar de tijd weg te werken. We substitueren daartoe $C_t = -uC_x$ in het rechterlid. Op die manier kunnen we schrijven

$$(C_t + uC_x)_i^n = -\frac{\nu^2 h^2}{2}(C_{xx})_i^n + \mathcal{O}(h^2) \quad \text{met} \quad \nu = \frac{u\tau}{h}. \tag{5.2.12}$$

We kunnen nu zien waarom het schema in (5.0.5) instabiel is. De eerste term in het rechterlid kan worden beschouwd als numerieke viscositeit, veroorzaakt door de discretisatiefout van het schema. De min voor deze term geeft negatieve viscositeit aan en deze term heeft de eigenschap om oscillaties te versterken in tegenstelling tot positieve viscositeit. Dit is de reden waarom het schema in (5.0.5) onstabiel is voor willekeurige waarden van $\nu$.

Laten we eens hetzelfde gaan doen maar nu voor een Euler expliciet schema waar voor de plaatsdiscretisatie het $\mathcal{O}(h)$ upwind schema is gebruikt

$$C_i^{n+1} = C_i^n - \nu\left(C_i^n - C_{i-1}^n\right). \tag{5.2.13}$$

Als we dezelfde bovenstaande procedure voor dit schema afwerken, dan krijgen we

$$(C_t + uC_x)_i^n = \frac{uh}{2}(1-\nu)(C_{xx})_i^n + \frac{uh^2}{6}(\nu+1)(\nu-1)(C_{xxx})_i^n + \frac{uh^3}{24}(1-\nu)(1+\nu+\nu^2)(C_{xxxx})_i^n + \mathcal{O}(h^4). \tag{5.2.14}$$

De eerste term in het rechterlid geeft de numerieke viscositeit weer. De tweede term in het rechterlid geeft numerieke dispersie weer en de derde term geeft numerieke dissipatie weer. Voor stabiliteit moet gelden dat de coëfficiënt in de eerste term in het rechterlid positief is. Hieraan wordt voldaan als $\nu < 1$ is. Aan de tweede term in het rechter lid kan geen informatie ontleent worden aangaande stabiliteit, echter voor de derde term in het rechterlid moet gelden dat de coëfficiënt negatief is. Hieraan wordt voldaan als $\nu > 1$ is. We komen dus op een tegenspraak. En dit geeft voor mij de reden aan waarom deze manier van stabiliteitsanalyse niet handig is. Immers hoeveel termen moeten er nu worden meegenomen bij de stabiliteitsanalyse?

## 5.3 Stabiliteitsanalyse met behulp van de von Neumann methode

Bij de von Neumann methode wordt Fouriertransformatie toegepast om de stabiliteit van het numerieke schema te onderzoeken. De toepasbaarheid van deze methode is beperkt tot schema's met constante coëfficiënten (zoals in (5.0.5)), waarbij het rooster uniform is en tot beginwaarde-problemen of begin-randwaarde problemen, waarbij *periodieke* randvoorwaarden gegeven zijn.

Laten we eens deze methode gaan toepassen op ons lineaire schema in (5.0.5), waarbij we aannemen dat de gegeven randvoorwaarden periodiek zijn. Zij $\overline{C}_i^n$ de exacte oplossing, $C_i^n$ de numerieke benadering en $\varepsilon_i^n$ de numerieke fout in roosterpunt $i$ op tijdstap $n$. We kunnen dan schrijven

$$C_i^n = \overline{C}_i^n + \varepsilon_i^n. \tag{5.3.15}$$

Als we dit substitueren in het lineaire schema in (5.0.5), dan moet gelden voor de fout $\varepsilon_i^n$ (immers $\overline{C}_i^n$ voldoet exact aan de gediscretiseerde vergelijking)

$$\varepsilon_i^{n+1} = \varepsilon_i^n - \frac{\nu}{2}(\varepsilon_{i+1}^n - \varepsilon_{i-1}^n). \tag{5.3.16}$$

We zien dus dat de numerieke fout $\varepsilon_i^n$ zich op dezelfde wijze ontwikkelt in de tijd als de numerieke oplossing $C_i^n$. Voor vaste $n$ is $\varepsilon_i^n$ een roosterfunctie in ons rekenrooster. Het is nu mogelijk om roosterfuncties in discrete Fouriersommen te ontwikkelen. We stellen daarom

$$\varepsilon_i^n = \sum_{j=1}^{N} E_j^n e^{I\theta_j}, \quad \text{met} \quad \theta_j = \frac{2\pi j}{N}, \tag{5.3.17}$$

met $I = \sqrt{-1}$ en $E_j^n$ stelt de amplitude voor van de $j$-de harmonische golf. Verder stelt $N$ het aantal controle-volumina voor waar ons domein is in verdeeld. De coëfficiënten $E_j^n$ moeten zo gekozen worden, dat (5.3.17) een oplossing wordt van (5.3.16).

Omdat we te maken hebben met een lineair schema, waaraan de fout $\varepsilon_i^n$ voldoet, is het niet moeilijk om in te zien dat ook iedere harmonische functie afzonderlijk hieraan moet voldoen. Laten we daarom één harmonische functie, $E_j^n e^{I\theta_j}$, beschouwen en deze substitueren in (5.3.16). We krijgen dan, als we voor het gemak de subscript $j$ even weglaten

$$E^{n+1}e^{Ii\theta} = E^n e^{Ii\theta} - \frac{\nu}{2}\left(E^n e^{I(i+1)\theta} - E^n e^{I(i-1)\theta}\right), \tag{5.3.18}$$

of nadat we gedeeld hebben door $e^{Ii\theta}$

$$E^{n+1} = E^n - \frac{\nu}{2}E^n\left(e^{I\theta} - e^{-I\theta}\right). \tag{5.3.19}$$

Als we nu (5.3.19) vervolgens omschrijven naar

$$G = 1 - I\nu\sin\theta, \tag{5.3.20}$$

waarbij $G$ gedefinieerd is als

$$G = \frac{E^{n+1}}{E^n}, \tag{5.3.21}$$

dan wordt er aan de stabiliteitsconditie in (5.1.7) voldaan als de amplitude $E^n$ niet toeneemt in de tijd. Dat wil zeggen als

$$|G| = \left|\frac{E^{n+1}}{E^n}\right| \leq 1, \quad \text{voor alle} \quad \theta. \tag{5.3.22}$$

De grootheid G wordt de *amplificatie factor* genoemd. Als er dus geldt

$$|G|^2 = 1 + \nu^2\sin^2\theta \leq 1, \tag{5.3.23}$$

dan is het schema stabiel. Gemakkelijk is in te zien dat hieraan nooit kan worden voldaan. Het schema is dus onvoorwaardelijk onstabiel. En dit komt overeen met hetgeen we ook reeds hadden gevonden in de vorige paragraaf.

Vervolgens gaan we de methode toepassen op het schema in (5.2.13). Als we weer één harmonische functie, $E^n e^{Ii\theta}$, beschouwen, dan krijgen we als we dit substitueren in de vergelijking voor de fout $\varepsilon_i^n$

$$E^{n+1}e^{Ii\theta} = E^n e^{Ii\theta} + \nu E^n\left(e^{Ii\theta} - e^{I(i-1)\theta}\right), \tag{5.3.24}$$

of na deling van $E^n e^{I i\theta}$,

$$G = 1 - \nu + \nu e^{-I\theta} = 1 - 2\nu \sin^2 \frac{\theta}{2} - I\nu \sin\theta. \tag{5.3.25}$$

Na enig rekenen vinden we dan voor de stabiliteitsvoorwaarde

$$0 < \nu \leq 1. \tag{5.3.26}$$

We zien dus dat dit schema voorwaardelijk stabiel is. De voorwaarde in (5.3.26) staat ook wel bekend als de *Courant-Friedrichs-Lewy* of de *CFL*-conditie. In tegenstelling tot de stabiliteitsmethode m.b.v. de vervangende vergelijking, zien we dat met behulp van de von Neumann methode wel op vrij eenvoudige wijze de stabiliteitscondities zijn te vinden.

## 5.4 Stabiliteitsanalyse met behulp van de Matrixmethode

De laatste stabiliteitsmethode die ik kort wil gaan behandelen is de matrixmethode. De matrixmethode is flexibeler dan de von Neumann methode, in het bijzonder met betrekking tot de randvoorwaarden en niet-uniforme roosters. Een nadeel van deze methode, zoals nog zal blijken, is dat deze moeilijk is toe te passen.

Stel we hebben het volgende semi-gediscretiseerde stelsel vergelijkingen, geschreven in matrix-vector notatie

$$\frac{dC}{dt} = SC + P. \tag{5.4.27}$$

Hierbij stelt $S$ de discretisatie matrix voor en $P$ de bijdragen van de randvoorwaarden.

Laten we eens bijvoorbeeld als tijdstapmethode Forward Euler gaan toepassen. Hierdoor gaat (5.4.27) over in

$$C^{n+1} = AC^n + \bar{P}, \tag{5.4.28}$$

met $A = I + \tau S$.

Als we nu aannemen dat er zich op de randen geen storingen bevinden in de randvoorwaarden, dan is het niet moeilijk om in te zien dat de foutvector $e^n$ voldoet aan de homogene vergelijking

$$e^{n+1} = Ae^n = (A)^n e^0. \tag{5.4.29}$$

Het schema is (absoluut) stabiel als er een constante $K > 0$ bestaat, zodanig dat

$$\|A^n\| \leq K, \quad \forall n. \tag{5.4.30}$$

Om te onderzoeken of hieraan wordt voldaan maken we gebruik van stelling 4.8.1 in [20], blz. 140

**Stelling 5.4.1** *Het schema is stabiel d.e.s.d.a.*

$$\rho(A) < 1 \quad \vee \quad [\rho(A) = 1 \quad \wedge \quad |\lambda| = 1 \quad \textit{niet defect}]\,. \tag{5.4.31}$$

Hierbij geeft $\rho(A)$ de spectraalradius aan van de matrix $A$. En hier zien we het nadeel van het gebruik van deze methode. Want het berekenen van de in absolute zin grootste eigenwaarde van een matrix kan soms erg lastig zijn.

Zoals ik reeds heb opgemerkt, heeft de matrixmethode als voordeel boven de methode van von Neumann, dat de invloed van randvoorwaarden en niet-uniforme roosters op eenvoudigere wijze dan bij de von Neumann methode kan worden meegenomen. Immers al deze informatie wordt verwerkt in de bij het schema behorende matrix. Toch schijnt dit geen beperking te zijn voor von Neumann volgens [5], blz 403. Daar wordt namelijk verteld dat in vele omstandigheden de aanwezigheid van niet-periodieke randvoorwaarden weinig effect heeft op de uitkomst van een von Neumann analyse, uitgevoerd op het inwendige van het rooster, in vergelijking met de matrixmethode. Verder wordt verteld dat in geval van discrepanties, de resultaten van de von Neumann methode veelal het meest nauwkeurig blijken te zijn.

Wij gebruiken als reden voor het gebruik van von Neumann boven de matrixmethode dat von Neumann op eenvoudigere wijze kan worden toegepast. Zoals al reeds is opgemerkt heeft de matrixmethode als nadeel dat het bepalen van de grootste eigenwaarde soms erg lastig kan zijn. Omdat ook al de tekortkomingen van de stabiliteitsanalyse met behulp van de Vervangende Vergelijking zijn aangetoond, zijn dit voor ons voldoende redenen om in het vervolg de von Neumann analyse te gaan toepassen. Hierbij gaan we op een iets andere manier te werk als in paragraaf 5.3. De tijdsintegratie zal namelijk nog niet meteen worden geschematiseerd. Allereerst zal geprobeerd worden om van de plaatsdiscretisatie het spectrum te bepalen. Daarna zal een zo geschikt mogelijk tijdstapschema genomen worden zodanig dat het stabiliteitsgebied, behorend bij dit tijdstapschema, het spectrum zo goed mogelijk omsluit. Immers alleen wanneer het spectrum zich volledig binnen het stabiliteitsgebied bevindt kunnen we, bij een niet te groot gekozen tijdstap, stabiliteit garanderen.

We zullen eerst het spectrum bepalen behorende bij onze plaatsdiscretisatie, inclusief de flux-limiter.

## 5.5 Bepaling van het spectrum

Zoals reeds bekend is hebben we bij de plaatsdiscretisatie gebruik gemaakt van een niet-lineair TVD-schema. Dit maakt het niet gemakkelijk om het spectrum te gaan bepalen. Immers met welk schema hebben we te maken? Als er in de oplossing valse niet-monotoniciteiten zijn, schakelt de flux-limiter naar het $\mathcal{O}(h)$-Upwind schema. In gebieden waar de oplossing glad is hebben we te maken met het ($\kappa = \frac{1}{3}$)-schema. Dit alles maakt het nogal lastig om een spectrum te construeren, dat bij onze plaatsdiscretisatie hoort.

Om toch een idee te krijgen hoe dit spectrum eruit ziet gaan we op drie manieren tegen de plaatsdiscretisatie aan kijken. Verder onderzoek moet dan aantonen welk van de drie manieren het spectrum zo goed mogelijk representeert. Bij elk van de drie manieren zal gebruik gemaakt worden van de *methode der bevroren coëfficiënten.* Dit wil zeggen dat de term in (5.0.1) tussen de grote haken in het rechterlid constant gehouden wordt, als het ware bevroren. Op deze manier zijn we in staat om enige uitspraken te doen over de stabiliteit van niet-lineaire schema's.

### Methode I

De eerste methode die we gaan beschouwen wordt veel gebruikt (zie [15]). Bij het bepalen van het spectrum op deze methode gaan we het rechterlid in (5.0.1) afschatten. Als we naar de flux-limiter (4.8.54) kijken, dan kunnen we de volgende afschatting maken

$$1 + \frac{1}{2}\phi(r_{i+\frac{1}{2}}) - \frac{\phi(r_{i-\frac{1}{2}})}{2r_{i-\frac{1}{2}}} \leq 1 + \frac{1}{2} \cdot 2 - 0 = 2. \tag{5.5.32}$$

Op deze manier gaat (5.0.1) over in

$$\frac{dC_i}{dt} = -\frac{2u}{h}\left(C_i - C_{i-1}\right). \tag{5.5.33}$$

We gaan hierop de von Neumann methode toepassen. Omdat de numerieke fout zich op dezelfde manier ontwikkelt als de numerieke oplossing (ze voldoen immers allebei aan hetzelfde lineaire schema), volstaan we door voor $C_i$ direct één harmonische golf,

$$C_i = Ee^{Ii\theta}, \tag{5.5.34}$$

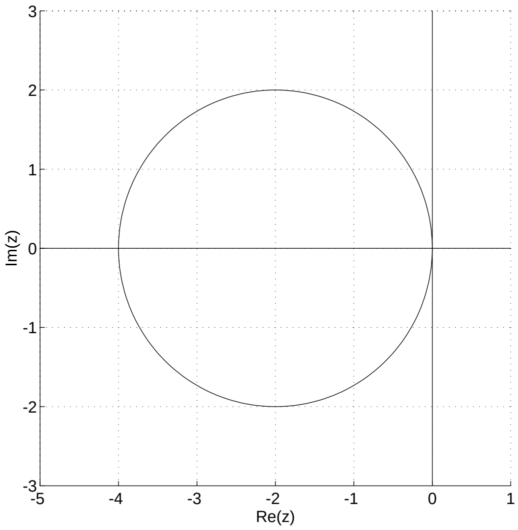
te substitueren in (5.5.32). Uitwerken geeft dan

$$\frac{dC_i}{dt} = -\frac{2u}{h}\left[(\cos(\theta) - 1) - I\sin(\theta)\right]C_i. \tag{5.5.35}$$

Het spectrum volgens Methode I wordt dan gegeven door de volgende formule

$$\lambda_1(\theta) = 2[(\cos(\theta) - 1) - I\sin(\theta)], \quad \text{met} \quad \theta \in [-\pi, \pi]. \tag{5.5.36}$$

Zoals te zien is in figuur 5.1, wordt het spectrum van de plaatsdiscretisatie met behulp van methode I voorgesteld door een cirkel met middelpunt $(-2, 0)$ en straal 2.




Figuur 5.1: Het spectrum volgens Methode I.

## Methode II

Bij Methode II gaan we eens kijken naar de sinus-functie in (4.10.60). We zien dat deze glad is en één maximum heeft. Dit heeft als gevolg dat de flux-limiter vooral zijn waarden zal aannemen in de omgeving van 1 (in het monotone gedeelte van de oplossing) of in de omgeving van 0 (in de buurt van het maximum). Ons eerste idee is dan ook om de flux-limiter opgebouwd te zien uit twee lijnen, namelijk de lijn $\phi = \frac{1}{3} + \frac{2}{3}r$ en de lijn $\phi = 0$. In dat geval zal het spectrum bestaan uit de omhullende van het spectrum behorende bij het ($\kappa = \frac{1}{3}$)-schema en het spectrum behorende bij het $\mathcal{O}(h)$ upwind schema.

Allereerst gaan we het spectrum behorende bij het $\mathcal{O}(h)$ upwind schema bepalen. We beschouwen dus het volgende semi-gediscretiseerde schema

$$\frac{dC_i}{dt} = -\frac{u}{h}\left(C_i - C_{i-1}\right). \tag{5.5.37}$$

We gaan hierop net zoals bij Methode I de von Neumann methode toepassen. Omdat (5.5.37) vrijwel identiek is aan (5.5.32), volstaan we door in één keer de formule voor het $\mathcal{O}(h)$ upwind spectrum te geven. Deze luidt

$$\lambda_2(\theta) = (\cos(\theta) - 1) - I\sin(\theta), \quad \text{met} \quad \theta \in [-\pi, \pi]. \tag{5.5.38}$$

Bij het $\kappa = \frac{1}{3}$ schema hebben we het volgende semi-gediscretiseerde stelsel

$$\frac{dC_i}{dt} = -\frac{u}{h}\left(C_i - C_{i-1} + \frac{1}{3}\left(C_{i+1} - 2C_i + C_{i-1}\right) + \frac{1}{6}\left(C_i - 2C_{i-1} + C_{i-2}\right)\right). \tag{5.5.39}$$

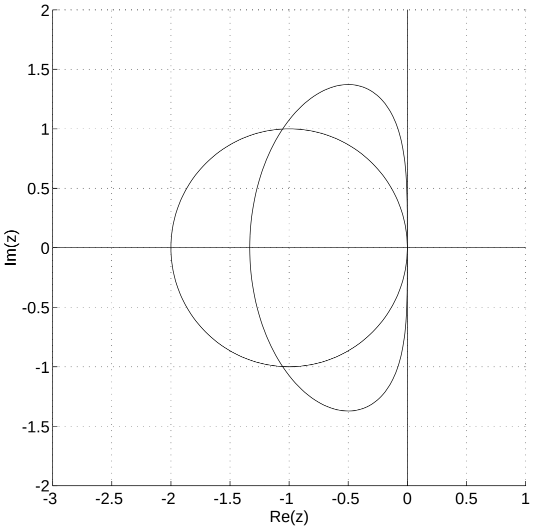
We substitueren weer (5.5.34) en uitwerken geeft dan

$$\frac{dC_i}{dt} = -\frac{u}{h}\left[\left(\frac{2}{3}\cos(\theta) - \frac{1}{6}\cos(2\theta) - \frac{1}{2}\right) + I\left(\frac{1}{6}\sin(2\theta) - \frac{4}{3}\sin(\theta)\right)\right]C_i. \tag{5.5.40}$$

Het spectrum behorende bij het $\kappa = \frac{1}{3}$ schema wordt gegeven door de volgende formule

$$\lambda_3(\theta) = \left(\frac{2}{3}\cos(\theta) - \frac{1}{6}\cos(2\theta) - \frac{1}{2}\right) + I\left(\frac{1}{6}\sin(2\theta) - \frac{4}{3}\sin(\theta)\right), \quad \text{met} \quad \theta \in [-\pi, \pi]. \tag{5.5.41}$$

Het spectrum van de plaatsdiscretisatie verkregen met behulp van deze methode is de omhullende van $\lambda_1(\theta)$ (de cirkel met middelpunt $(-1, 0)$ en straal 1) en $\lambda_2(\theta)$ (de ovaal-vormige figuur) in figuur 5.2.




Figuur 5.2: Het spectrum volgens Methode II.

## Methode III

De derde methode, [10], is eigenlijk het omgekeerde van de von Neumann methode. Bij die methode stopten we er steeds een hele verzameling frequenties in en hadden we steeds één discretisatie. Bij deze methode stoppen we er één frequentie in en krijgen daarvoor een hele verzameling discretisaties terug.

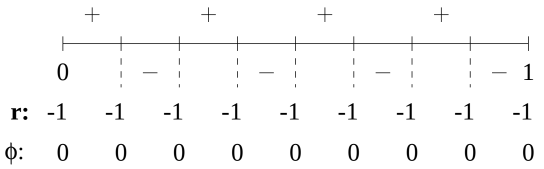
We beschouwen steeds één sinus-vormige golf, dus met één bepaalde frequentie $\theta$, en bekijken dan hoe deze golf in één tijdstap wordt vervormd. Bij elke $\theta$ krijgen we een wolk van eigenwaarden. De in absolute waarde grootste eigenwaarden zullen we het spectrum laten vormen.

Alvorens we met deze methode gaan beginnen, zouden we ons eerst af kunnen vragen in hoeverre deze aanpak juist is. Immers voor afzonderlijke frequenties is deze methode correct. Maar als we een willekeurige oplossing beschouwen, welke is opgebouwd uit meerdere frequenties, dan kunnen er vraagtekens bij deze methode worden geplaatst. Immers het superpositie principe is hier niet geldig. In hoeverre deze aanpak werkt zal wel blijken als we dit gaan testen met behulp van een testprobleem. Wat we alvast wél kunnen opmerken is dat deze analyse in ieder geval zal leiden tot enig inzicht aangaande het voor ons tot nu toe onbekende spectrum behorende bij onze plaatsdiscretisatie en dát is in ieder geval beter dan niets.

We gaan als volgt te werk. Allereerst beschouwen we een golf met frequentie $\theta = 0$ (dit is een constante) op ons domein, dat bijvoorbeeld is opgedeeld in acht cellen. Vervolgens gaan we op elk celwandje de $r$-waarde bepalen. Omdat we hier een constante lijn hebben, zal de breuk $r$ overal op elk celwandje de waarde 1 hebben. Daarna substitueren we deze waarde in de flux-limiter. Omdat we bij $r = 1$ ons bevinden op de $(\kappa = \frac{1}{3})$-lijn, zien we dat we te maken hebben met het $(\kappa = \frac{1}{3})$-schema. Op deze manier kunnen we in elk celletje de $C_{i+\frac{1}{2}}$ en de $C_{i-\frac{1}{2}}$ bepalen met behulp van (4.4.17). Omdat dit, bij deze frequentie, in ieder celletje dezelfde waarde oplevert (immers we hebben hier een constante oplossing), geldt er voor ieder celletje (5.5.39). En hiervan hebben we reeds het spectrum bepaald. We substitueren daarom $\theta = 0$ in (5.5.41) en krijgen dan als enige eigenwaarde het punt 0.

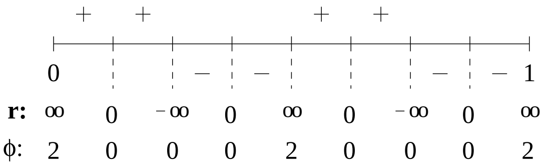
Vervolgens beschouwen we de hoogst mogelijk frequentie $\theta = \pi$. We gaan weer op dezelfde wijze te werk als bij $\theta = 0$. Gemakkelijk is in te zien dat $r$ overal de waarde $-1$ aanneemt. Op dit punt heeft de flux-limiter de waarde 0. Eén en ander is nogmaals op schematische wijze weergegeven in figuur 5.3 tezamen met de $r$- en de flux-limiter-waarden op elk celwandje.
We zien dus dat voor ieder celletje (5.5.37) geldt en ook hiervoor hebben we reeds het spectrum bepaald. We substitueren daarom $\theta = \pi$ in (5.5.38) en krijgen als enige eigenwaarde het punt $-2$.

Ik heb bij deze analyse nog drie andere frequenties bekeken, namelijk $\theta = \frac{\pi}{2}$, $\theta = \frac{\pi}{4}$ en $\theta = \frac{3\pi}{4}$. Ik zal alleen $\theta = \frac{\pi}{2}$ behandelen, bij de overige twee frequenties worden de eigenwaarden op dezelfde wijze verkregen. Bij $\theta = \frac{\pi}{2}$ ziet de golf op ons domein met acht cellen er op schematische wijze als volgt uit.




Figuur 5.3: Schematische voorstelling van een golf met frequentie $\pi$.

Zie figuur 5.4. Net als in figuur 5.3 zijn de waarden van $r$ en de flux-limiter op de celwandjes in de figuur weergegeven.




Figuur 5.4: Schematische voorstelling van een golf met frequentie $\frac{\pi}{2}$.

We zien dat de $\phi$-waarden periodiek herhaald worden en daarom beschouwen we één zo'n periode: $(2,0,0,0,2)$. Met behulp hiervan kunnen vier discrete vergelijkingen worden afgeleid voor vier opeenvolgende celletjes op, zeg, de roosterpunten $i$, $i+1$, $i+2$ en $i+3$:

$$\begin{array}{rclcl}
(\phi_{i-\frac{1}{2}},\phi_{i+\frac{1}{2}}) & = & (2,0): \frac{dC_i}{dt} & = & \frac{u}{h}(C_{i-\frac{1}{2}}-C_{i+\frac{1}{2}}) = \frac{u}{h}(-C_{i-2}+2C_{i-1}-C_i), \\
(\phi_{i-\frac{1}{2}},\phi_{i+\frac{1}{2}}) & = & (0,0): \frac{dC_{i+1}}{dt} & = & \frac{u}{h}(C_{i+\frac{1}{2}}-C_{i+\frac{3}{2}}) = \frac{u}{h}(C_i-C_{i+1}), \\
(\phi_{i-\frac{1}{2}},\phi_{i+\frac{1}{2}}) & = & (0,0): \frac{dC_{i+2}}{dt} & = & \frac{u}{h}(C_{i+\frac{3}{2}}-C_{i+\frac{5}{2}}) = \frac{u}{h}(C_{i+1}-C_{i+2}), \\
(\phi_{i-\frac{1}{2}},\phi_{i+\frac{1}{2}}) & = & (0,2): \frac{dC_{i+3}}{dt} & = & \frac{u}{h}(C_{i+\frac{5}{2}}-C_{i+\frac{7}{2}}) = \frac{u}{h}(2C_{i+2}-2C_{i+3}),
\end{array} \tag{5.5.42}$$

of in matrixvorm:

$$\frac{d}{dt}\begin{pmatrix} C_i \\ C_{i+1} \\ C_{i+2} \\ C_{i+3} \end{pmatrix} = \frac{u}{h}\begin{pmatrix} -1 & 2 & -1 & & & \\ & & 1 & -1 & & \\ & & & 1 & -1 & \\ & & & & 2 & -2 \end{pmatrix}\begin{pmatrix} C_{i-2} \\ C_{i-1} \\ C_i \\ C_{i+1} \\ C_{i+2} \\ C_{i+3} \end{pmatrix}. \tag{5.5.43}$$

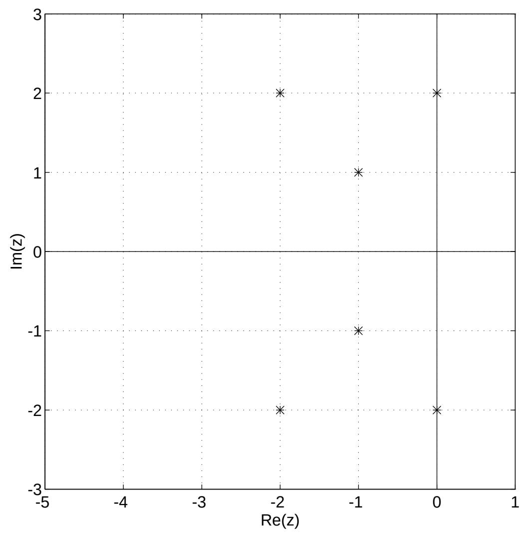
De volgende stap is het elimineren van $C_{i-2}$ en $C_{i-1}$ en het diagonaliseren van de matrix middels de substitutie $C_i = Ee^{Ii\theta}$. We krijgen dan

$$\frac{d}{dt}\begin{pmatrix} C_i \\ C_{i+1} \\ C_{i+2} \\ C_{i+3} \end{pmatrix} = \frac{u}{h}\mathrm{diag}\begin{pmatrix} e^{-2I\theta}+2e^{-I\theta}-1, \\ e^{-I\theta}-1, \\ e^{-I\theta}-1, \\ 2e^{-I\theta}-2 \end{pmatrix}\begin{pmatrix} C_i \\ C_{i+1} \\ C_{i+2} \\ C_{i+3} \end{pmatrix}. \tag{5.5.44}$$

Vervolgens substitueren we de beschouwde frequentie $\theta = \frac{\pi}{2}$ (en $\theta = -\frac{\pi}{2}$ voor de complex geconjugeerde eigenwaarden) en krijgen dan

$$\frac{d}{dt}\begin{pmatrix} C_i \\ C_{i+1} \\ C_{i+2} \\ C_{i+3} \end{pmatrix} = \frac{u}{h}\mathrm{diag}\begin{pmatrix} \pm 2I, \\ -1\pm I, \\ -1\pm I, \\ -2\pm 2I \end{pmatrix}\begin{pmatrix} C_i \\ C_{i+1} \\ C_{i+2} \\ C_{i+3} \end{pmatrix}. \tag{5.5.45}$$

De eigenwaarden behorende bij deze frequentie worden dus gegeven door de punten $\pm 2I$, $-1\pm I$ en $-2\pm 2I$. In figuur 5.5 is de ligging van deze eigenwaarden in het complexe vlak te zien.




Figuur 5.5: De wolk van eigenwaarden behorende bij $\theta = \frac{\pi}{2}$.

Bij het bepalen van de wolk van eigenwaarden bij de frequentie $\theta = \frac{\pi}{4}$ moet even opgelet worden. In tegenstelling tot figuur 5.3 en 5.4 hebben in dit geval de $+$ en de $-$ niet allemaal dezelfde waarden. In dit geval moeten de waarden in de celcentra allemaal expliciet worden bepaald. Bij het bepalen van deze waarden maken we gebruik van de functie $\sin(2\pi x)$. (Bij $\theta = \frac{\pi}{4}$ hebben we te maken met een golf over $8h$. De functie $\sin(2\pi x)$ geeft dit precies goed weer.).

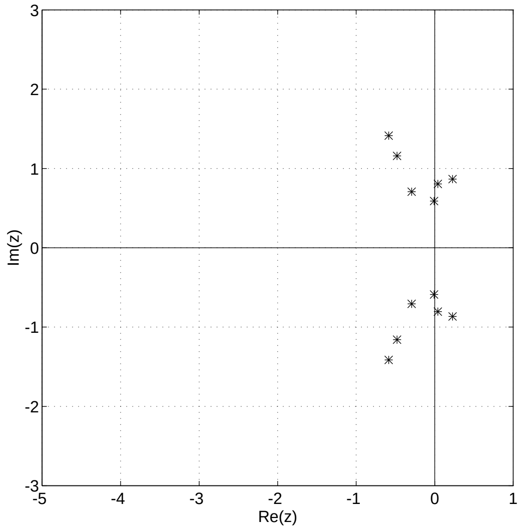
We passen weer de bovenstaande werkwijze toe bij het bepalen van de eigenwaarden en komen na enig rekenen op het volgende semi-gediscretiseerde stelsel vergelijkingen, waarvan de matrix met behulp van Fourier is gediagonaliseerd

$$\frac{d}{dt}\begin{pmatrix} C_i \\ C_{i+1} \\ C_{i+2} \\ C_{i+3} \\ C_{i+4} \\ C_{i+5} \\ C_{i+6} \\ C_{i+7} \end{pmatrix} = \frac{u}{h}\text{diag}\begin{pmatrix} (\frac{1}{2}\sqrt{2}-\frac{2}{3}) \pm (\frac{1}{3}\sqrt{2}+\frac{1}{3})I, \\ (\frac{7}{12}\sqrt{2}-\frac{5}{6}) \pm \frac{5}{12}\sqrt{2}I, \\ (\frac{1}{2}\sqrt{2}-1) \pm \frac{1}{2}\sqrt{2}I, \\ (\frac{1}{4}\sqrt{2}-\frac{5}{6}) \pm (\frac{7}{12}\sqrt{2}+\frac{1}{3})I, \\ (\frac{1}{2}\sqrt{2}-\frac{2}{3}) \pm (\frac{1}{3}\sqrt{2}+\frac{1}{3})I, \\ (\frac{7}{12}\sqrt{2}-\frac{5}{6}) \pm \frac{5}{12}\sqrt{2}I, \\ (\sqrt{2}-2) - \sqrt{2}I, \\ (\frac{3}{4}\sqrt{2}-\frac{5}{6}) \pm (\frac{12}{13}\sqrt{2}-\frac{2}{3})I \end{pmatrix}\begin{pmatrix} C_i \\ C_{i+1} \\ C_{i+2} \\ C_{i+3} \\ C_{i+4} \\ C_{i+5} \\ C_{i+6} \\ C_{i+7} \end{pmatrix}. \qquad (5.5.46)$$

In figuur 5.6 zijn de eigenwaarden in het complexe vlak te zien.

In figuur 5.6 zien we dat bij deze frequentie eigenwaarden voorkomen met een positief reel deel. Dit zijn de volgende eigenwaarden: $(\frac{1}{2}\sqrt{2}-\frac{2}{3}) \pm (\frac{1}{3}\sqrt{2}+\frac{1}{3})I$ en $(\frac{3}{4}\sqrt{2}-\frac{5}{6}) \pm (\frac{12}{13}\sqrt{2}-\frac{2}{3})I$. Als we nu als tijdstapschema het klassieke 4-stage Runge-Kutta schema zouden willen toepassen, dan zouden we in de problemen kunnen raken. Immers het stabiliteitsgebied van het RK4 schema laat geen eigenwaarden toe met een positief reëel deel. We moeten echter bedenken dat we slechts gekeken hebben naar één tijdstap. Bij een volgend tijdstap is het goed mogelijk dat alle eigenwaarden wel weer binnen het stabiliteitsgebied komen, dankzij de werking van de flux-limiter. We zullen daarom niet al te veel aandacht schenken aan deze eigenwaarden als we het spectrum gaan construeren.

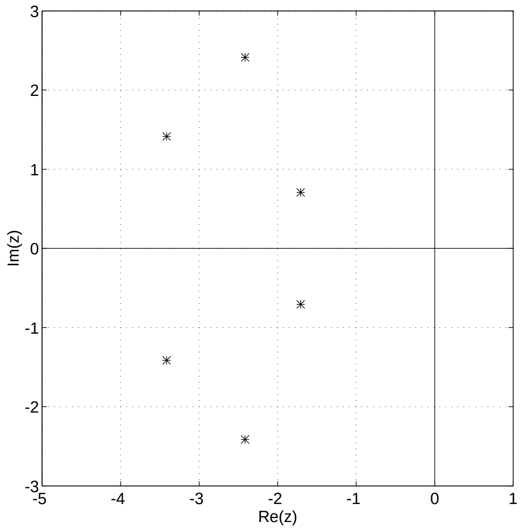
Voor de frequentie $\theta = \frac{3\pi}{4}$ vinden we na enig rekenen




Figuur 5.6: De wolk van eigenwaarden behorende bij $\theta = \frac{\pi}{4}$.

$$\frac{d}{dt}\begin{pmatrix} C_i \\ C_{i+1} \\ C_{i+2} \\ C_{i+3} \\ C_{i+4} \\ C_{i+5} \\ C_{i+6} \\ C_{i+7} \end{pmatrix} = \frac{u}{h}\mathrm{diag}\begin{pmatrix} -(\frac{1}{2}\sqrt{2}+1) \pm \frac{1}{2}\sqrt{2}I \quad (6\times), \\ -(2+\sqrt{2}) \pm \sqrt{2}I, \\ -(1+\sqrt{2}) \pm (1+\sqrt{2})I \end{pmatrix}\begin{pmatrix} C_i \\ C_{i+1} \\ C_{i+2} \\ C_{i+3} \\ C_{i+4} \\ C_{i+5} \\ C_{i+6} \\ C_{i+7} \end{pmatrix}. \tag{5.5.47}$$

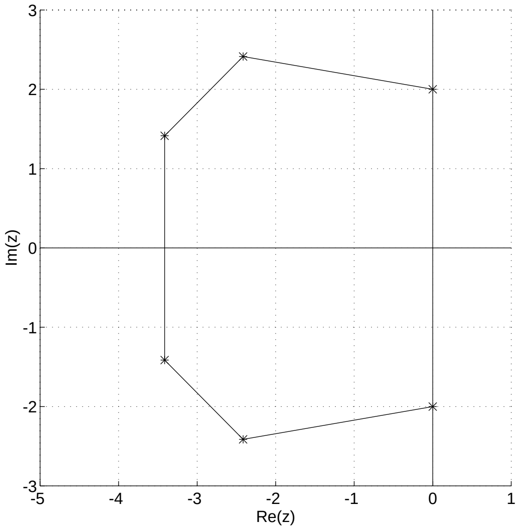
In figuur 5.7 is deze wolk van eigenwaarden weer afgedrukt in het complexe vlak. In tegenstelling tot de vorige frequentie, hebben alle eigenwaarden nu wel een negatief reëel deel.




Figuur 5.7: De wolk van eigenwaarden behorende bij $\theta = \frac{3\pi}{4}$.

We hebben nu alle eigenwaarden bepaald behorende bij deze vijf frequenties en we willen nu een omhullende kiezen waarbinnen alle eigenwaarden met een reëel deel kleiner of gelijk aan nul zich bevinden. De eigenwaarden die dit gebied vormen zullen we de kritieke eigenwaarden noemen. Het gebied gevormd door deze eigenwaarden is het spectrum behorende bij methode III.

Als eigenwaarden, behorende bij de beschouwde frequenties, die het spectrum vormen, vinden we $\pm 2I$, $-(1+\sqrt{2}) \pm (1+\sqrt{2})I$ en $-(2+\sqrt{2}) \pm \sqrt{2}I$. Het het spectrum is te zien in figuur 5.8.




Figuur 5.8: Het spectrum volgens Methode III.

We hebben nu op drie manieren een spectrum bepaald behorende bij onze plaatsdiscretisatie en geplot in figuur 5.1, 5.2 en 5.8. De vraag die ons nu gaat bezighouden is welk spectrum het meest realistisch is. We hopen dat dit duidelijker gaat worden als we enige Runge-Kutta schema's gaan bekijken. Hierbij is ons uiteindelijke doel om een stabiliteitsgebied, behorende bij een Runge-Kutta schema, te construeren zodanig dat ons meest nauwkeurige spectrum zo goed mogelijk wordt omsloten.

## 5.6 De Runge-Kutta methoden

Veel toegepaste tijdsintegratiemethoden zijn de Runge-Kutta schema's. Het idee bij de Runge-Kutta methoden is om bij het systeem van differentiaalvergelijkingen van de vorm

$$\frac{dC}{dt} = f(C,t), \tag{5.6.48}$$

het rechterlid in verschillende waarden van $C$ te evalueren op het tijdsinterval tussen $n\tau$ en $(n+1)\tau$, en de evaulaties op zo'n manier met elkaar te combineren, dat een zeer nauwkeurige benadering wordt verkregen voor $C^{n+1}$. Een zeer gedetailleerde beschrijving van de Runge-Kutta methoden is te vinden in [7]. Omdat het rechterlid van (5.6.48) in ons geval onafhankelijk van de tijd is, wordt de algemene vorm van een $K$-stage Runge-Kutta methode in ons geval

$$\begin{array}{rcl} C^{(i)} & = & C^n + \tau \sum_{j=1}^{i-1} a_{ij} f(C^{(j)}), \;\; i = 1, \ldots, K \\ C^{n+1} & = & C^n + \tau \sum_{i=1}^{K} b_i f(C^{(i)}). \end{array} \tag{5.6.49}$$

We zien dat deze methode volledig is bepaald door de reële coëfficiënten $a_{ij}$ en $b_i$ en het aantal stages $K$. De methode kan ook op compacte wijze worden weergegeven door het volgende schema

$$\begin{array}{c|c} & A \\ \hline & b^T \end{array}$$

waarbij $A$ een benedendriehoeks matrix is en $b = (b_i)$ een vector ter lengte $K$. Zo'n schema wordt ook wel een *Butcher array* genoemd.

Als we eens het standaard RK4 schema gaan beschouwen, dan ziet deze er als volgt uit in formule vorm

$$\begin{aligned}
C^{(1)} &= C^n \\
C^{(2)} &= C^n + \tfrac{1}{2}\tau f(C^{(1)}) \\
C^{(3)} &= C^n + \tfrac{1}{2}\tau f(C^{(2)}) \\
C^{(4)} &= C^n + \tau f(C^{(3)}) \\
\\
C^{n+1} &= C^n + \tfrac{1}{6}\tau \left(f(C^{(1)}) + 2f(C^{(2)}) + 2f(C^{(3)}) + f(C^{(4)})\right).
\end{aligned} \tag{5.6.50}$$

Op compacte wijze weergegeven in een Butcher array ziet het RK4 schema er als volgt uit, zie onderstaand schema.

$$\begin{array}{c|cccc}
 & \frac{1}{2} & & & \\
 & 0 & \frac{1}{2} & & \\
 & 0 & 0 & 1 & \\
\hline
 & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6}
\end{array}$$

Het RK4 schema

Laten we eens gaan kijken naar het stabiliteitsgebied van RK4. We passen hierbij lineaire stabiliteitsanalyse toe, d.w.z we nemen aan dat geldt: $f(C) = \lambda C$. Als we dit invullen in (5.6.50) en de formules in elkaar substitueren, dan krijgen we

$$\begin{aligned}
C^{(1)} &= C^n \\
C^{(2)} &= \left(1 + \tfrac{1}{2}\tau\lambda\right) C^n \\
C^{(3)} &= \left(1 + \tfrac{1}{2}\tau\lambda + \tfrac{1}{4}(\tau\lambda)^2\right) C^n \\
C^{(4)} &= \left(1 + \tau\lambda + \tfrac{1}{2}(\tau\lambda)^2 + \tfrac{1}{4}(\tau\lambda)^3\right) C^n \\
\\
C^{n+1} &= \left(1 + \tau\lambda + \tfrac{1}{2}(\tau\lambda)^2 + \tfrac{1}{6}(\tau\lambda)^3 + \tfrac{1}{24}(\tau\lambda)^4\right) C^n \equiv k(\tau\lambda) C^n.
\end{aligned} \tag{5.6.51}$$

Hierbij wordt $k(\tau\lambda)$ de versterkingsfactor genoemd en er geldt dat het schema stabiel is d.e.s.d.a. geldt

$$|k(\tau\lambda)| \leq 1. \tag{5.6.52}$$

Als we nu stellen $\lambda \in \mathbb{C}$, vast en gegeven, en we beschouwen $\tau$ variabel, dan is de vraag voor welke waarden van $\tau$ nu aan (5.6.52) wordt voldaan.

Het stabiliteitsgebied is als volgt gedefinieerd. Als we stellen $z = \tau\lambda$, met $z \in \mathbb{C}$, dan is het stabiliteitsgebied $S$ gedefinieerd door

$$S = \{z \in \mathbb{C} : |k(z)| \leq 1\}. \tag{5.6.53}$$

Als $\tau\lambda \in S$ dan is de methode stabiel. Met behulp van deze definities kunnen we het stabiliteitsgebied van RK4 als volgt opschrijven

$$S_4 = \left\{ z \in \mathbb{C} : \left| 1 + z + \frac{1}{2}z^2 + \frac{1}{6}z^3 + \frac{1}{24}z^4 \right| \leq 1 \right\} \tag{5.6.54}$$

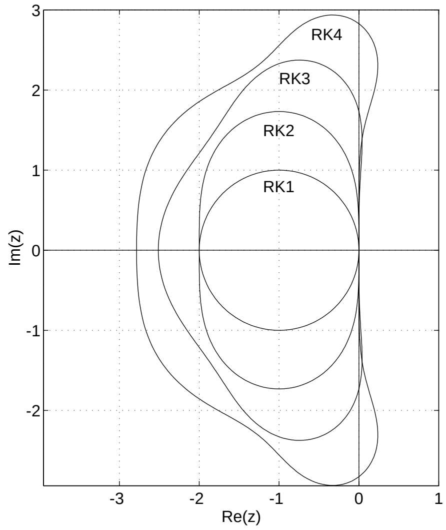
Op geheel dezelfde wijze kunnen de stabiliteitsgebieden van andere RK schema's worden afgeleid. Hierbij geldt dat alle RK schema's met $K$ stages hetzelfde lineaire stabiliteitsgebied hebben. We vinden dan bijvoorbeeld voor de stabiliteitsgebieden $S_1$, $S_2$ en $S_3$ behorende bij RK1 (Forward Euler), RK2 (Heun) en RK3

$$S_1 = \{z \in \mathbb{C} : |1+z| \leq 1\}, \tag{5.6.55}$$

$$S_2 = \left\{z \in \mathbb{C} : \left|1+z+\frac{1}{2}z^2\right| \leq 1\right\}, \tag{5.6.56}$$

$$S_3 = \left\{z \in \mathbb{C} : \left|1+z+\frac{1}{2}z^2+\frac{1}{6}z^3\right| \leq 1\right\}. \tag{5.6.57}$$

In figuur 5.9 zijn de stabiliteitsgebieden van RK1, RK2, RK3 en RK4 in één figuur afgedrukt.




Figuur 5.9: Het stabiliteitsgebied van RK1 t/m RK4.

Over de orden van nauwkeurigheid van de RK schema's kunnen we het volgende zeggen. Een RK methode heeft een *globale afbreekfout* van de *orde p* als de lokale afbreekfout van de orde $p+1$ is. Voor RK4 geldt dat

$$e^{\tau\lambda} - k(\tau\lambda) = \mathcal{O}(\tau^5). \tag{5.6.58}$$

Daarmee is de lokale afbreekfout van de orde 5 en de globale afbreekfout van de orde 4. We hebben hier alleen lineaire functies $f(C)$ beschouwd. Ook voor willekeurige lineaire functies geldt dit maar het bewijs hiervan zal achterwege worden gelaten.

Zoals reeds in de vorige paragraaf was aangeduid, willen we graag weten welke van de drie spectra, bepaald door de drie methoden, het meest realistisch is. Omdat de fout in onze plaatsdiscretisatie van de orde 3 is gaan we daarom bij het testen van de drie spectra, RK3 schema's gebruiken. Van de RK3 schema's die we gaan gebruiken eisen we dat ze ook in het niet-lineaire geval van de orde 3 zijn (we weten reeds dat ze in het lineaire geval wel van de orde 3 zijn).

We beschouwen in ons onderzoek de volgende vier RK3 schema's, op compacte wijze weergegeven in tabel 5.1.

Allereerst gaan we onderzoeken hoe het zit met de orde van nauwkeurigheid in de tijd van deze schema's. Beschouw hiertoe de volgende algemene Taylorreeks:

$$C^{n+1} = C^n + \tau\frac{dC^n}{dt} + \frac{1}{2}\tau^2\frac{d^2C^n}{dt^2} + \frac{1}{6}\tau^3\frac{d^3C^n}{dt^3} + \mathcal{O}(\tau^4). \tag{5.6.59}$$

Nu kunnen we schrijven met behulp van (5.6.48), waarbij het rechterlid onafhankelijk van de tijd is verondersteld

Tabel 5.1: De vier RK3 schema's in Butcher arrays.

$$\begin{array}{c|ccc} & \frac{1}{3} & & \\ & 0 & \frac{2}{3} & \\ \hline & \frac{1}{4} & 0 & \frac{3}{4} \end{array}$$

a. Heun [4]

$$\begin{array}{c|ccc} & \frac{1}{2} & & \\ & -1 & 2 & \\ \hline & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \end{array}$$

b. Kutta [11]

$$\begin{array}{c|ccc} & 1 & & \\ & \frac{1}{4} & \frac{1}{4} & \\ \hline & \frac{1}{6} & \frac{1}{6} & \frac{2}{3} \end{array}$$

c. RK3 (Fehlberg) [2]

$$\begin{array}{c|ccc} & \frac{8}{15} & & \\ & \frac{1}{4} & \frac{5}{12} & \\ \hline & \frac{1}{4} & 0 & \frac{3}{4} \end{array}$$

d. Wray [21]

$$\frac{dC^n}{dt} = f(C^n), \tag{5.6.60a}$$

$$\frac{d^2C^n}{dt^2} = f(C^n)f'(C^n), \tag{5.6.60b}$$

$$\frac{d^3C^n}{dt^3} = f(C^n)f'^2(C^n) + f^2(C^n)f''(C^n). \tag{5.6.60c}$$

Als we (5.6.60a), (5.6.60b) en (5.6.60c) substitueren in (5.6.59), dan geeft dit

$$C^{n+1} = C^n + \tau f(C^n) + \frac{1}{2}\tau^2 f(C^n)f'(C^n) + \frac{1}{6}\tau^3 \left(f(C^n)f'^2(C^n) + f^2(C^n)f''(C^n)\right) + \mathcal{O}(\tau^4). \tag{5.6.61}$$

Bij het nauwkeurigheids onderzoek van elk der vier RK3 schema's hopen we dat we op (5.6.61) uitkomen. In dat geval weten we zeker dat de RK3 schema's ook derde orde nauwkeurig zijn voor niet-lineaire problemen.

Laten we eens het Heun [4] schema in tabel 5.1a gaan onderzoeken. In formulevorm ziet dit schema er als volgt uit

$$
\begin{aligned}
C^{(1)} &= C^n \\
C^{(2)} &= C^n + \tfrac{1}{3}\tau f(C^{(1)}) \\
C^{(3)} &= C^n + \tfrac{2}{3}\tau f(C^{(2)}) \\
\\
C^{n+1} &= C^n + \tfrac{1}{4}\tau f(C^{(1)}) + \tfrac{3}{4}\tau f(C^{(3)}).
\end{aligned} \tag{5.6.62}
$$

Beschouw allereerst de term $C^{n+1}$ in (5.6.62). Hierin substitueren we $C^{(3)}$. Dit geeft

$$
C^{n+1} = C^n + \frac{1}{4}\tau f(C^n) + \frac{3}{4}\tau f(C^n + \frac{2}{3}\tau f(C^{(2)})). \tag{5.6.63}
$$

Allereerst gaan we de term $\frac{2}{3}\tau f(C^{(2)})$ omschrijven. We kunnen namelijk schrijven $\frac{2}{3}\tau f(C^{(2)}) = C^{(3)} - C^n$. Vervolgens schrijven we $f$ in een Taylorreeks rond $C^n$. Dit geeft

$$
\begin{aligned}
C^{n+1} &= C^n + \frac{1}{4}\tau f(C^n) + \frac{3}{4}\tau \left[ f(C^n) + \left(C^{(3)} - C^n\right) f'(C^n) \right. \\
&\quad \left. + \frac{1}{2}\left(C^{(3)} - C^n\right)^2 f''(C^n) + \mathcal{O}\left(C^{(3)} - C^n\right)^3 \right].
\end{aligned} \tag{5.6.64}
$$

Om(5.6.64) uit te werken, gaan we apart de term $C^{(3)} - C^n$ beschouwen. We kunnen dan gaan schrijven

$$
\begin{aligned}
C^{(3)} - C^n &= \frac{2}{3}\tau f(C^{(2)}) = \frac{2}{3}\tau f(C^n + \frac{1}{3}\tau f(C^n)) \\
&= \frac{2}{3}\tau f(C^n) + \frac{2}{9}\tau^2 f(C^n) f'(C^n) + \frac{1}{27}\tau^3 f^2(C^n) f''(C^n) + \mathcal{O}(\tau^4).
\end{aligned} \tag{5.6.65}
$$

Als we (5.6.65) substitueren in (5.6.64) en de haakjes uitwerken, dan levert dit

$$
C^{n+1} = C^n + \tau f(C^n) + \frac{1}{2}\tau^2 f(C^n) f'(C^n) + \frac{1}{6}\tau^3 f'^2(C^n) f(C^n) + \frac{1}{6}\tau^3 f^2(C^n) f''(C^n) + \mathcal{O}(\tau^4). \tag{5.6.66}
$$

We zien dat (5.6.66) overeenkomt met (5.6.61). Dus Heun [4] heeft ook voor niet-lineaire problemen een globale afbreekfout van de derde orde.

Op geheel dezelfde wijze kan de bovenstaande werkwijze ook worden toegepast op de overige drie RK3 schema's. We vinden dan in elk van de drie gevallen dezelfde formule (5.6.61) terug. Dus elk van de vier RK3 schema's heeft zowel voor lineaire als niet-lineaire problemen een globale afbreekfout van de orde drie.

De volgende stap in ons onderzoek is om voor elk van onze drie spectra, de CFL-conditie te gaan bepalen. Deze CFL-conditie zal tevoorschijn komen als we van elk van de drie spectra het snijpunt gaan proberen te vinden met het lineaire stabiliteitsgebied van RK3. Met deze CFL waarde zullen we in staat zijn om het beste spectrum te gaan bepalen.

## 5.7 Bepaling van de CFL-conditie voor de drie spectra

We hebben reeds afgeleid dat de stabiliteitseis, geldend voor ieder RK3 schema, als volgt luidt

$$
\left| 1 + \lambda\tau + \frac{1}{2}(\lambda\tau)^2 + \frac{1}{6}(\lambda\tau)^3 \right| \leq 1. \tag{5.7.67}
$$

Bij het bepalen van het snijpunt van de verschillende spectra met het stabiliteitsgebied, gaan we als volgt te werk. De $\lambda$'s in (5.7.67) stellen de spectra van eigenwaarden voor. Omdat het te lastig is de formule van ieder spectrum direct in te vullen en vervolgens het snijpunt te bepalen, substitueren we de punten waarvan we vermoeden dat deze het eerst zullen gaan snijden met het stabiliteitsgebied. Stel dat dit betreffende kritieke punt, $\lambda_{kr}$, er als volgt uitziet

$$\lambda_{kr} = \frac{u}{h}(a + bI), \tag{5.7.68}$$

waarbij $a$ en $b$ bekend verondersteld mogen worden. Als we dit substitueren in (5.7.67), dan krijgen we

$$\left|1 + \frac{u\tau}{h}(a + bI) + \frac{1}{2}\left(\frac{u\tau}{h}\right)^2 (a + bI)^2 + \frac{1}{6}\left(\frac{u\tau}{h}\right)^3 (a + bI)^3\right| \leq 1. \tag{5.7.69}$$

Als we dit uitwerken en het Courant-getal, $\nu = \frac{u\tau}{h}$, substitueren, dan krijgen we de volgende voorwaarde

$$\left\{1 + \nu a + \frac{1}{2}\nu^2\left(a^2 - b^2\right) + \frac{1}{6}\nu^3 a\left(a^2 - 3b^2\right)\right\}^2 + b^2\left\{\nu + \nu^2 a + \frac{1}{6}\nu^3\left(3a^2 - b^2\right)\right\}^2 \leq 1. \tag{5.7.70}$$

We zien nu hoe de CFL-conditie tot stand komt bij deze werkwijze. Als namelijk het kritieke punt wordt ingevuld, dan houden we een vergelijking over als functie van $\nu$. De oplossing van deze vergelijking geeft precies die waarde aan waarmee we het spectrum kunnen opblazen, totdat het stabiliteitsgebied wordt geraakt. Dit is precies de CFL-conditie die we zoeken. We weten dan ook meteen de tijdstapgrootte $\tau$, welke stabiliteit garandeert.

## Spectrum I

Allereerst beginnen we met het spectrum verkregen door de afschatting van de term in het rechterlid van (5.0.1), zoals uiteen was gezet in methode I. Als we goed kijken naar het spectrum in figuur 5.1 dan kunnen we opmerken dat het snijpunt met de negatieve reële as (het punt $-4$) het eerste punt zal zijn dat het RK3 stabiliteitsgebied zal raken. We substitueren dan ook in (5.7.70) $a = -4$ en $b = 0$. Dit geeft dan de volgende formule voor $\nu$

$$\frac{1024}{9}\nu^6 - \frac{512}{9}\nu^5 + \frac{448}{3}\nu^4 - \frac{256}{3}\nu^3 + 32\nu^2 - 8\nu = 0. \tag{5.7.71}$$

Met behulp van MAPLE vinden we voor $\nu$ de waarde $\nu_I = 0.628$ en de CFL-conditie luidt daarmee

$$\nu_I \leq 0.628. \tag{5.7.72}$$

Als we het spectrum vermenigvuldigen met deze waarde $\nu_I$ dan zien we in figuur 5.10 dat de twee gebieden elkaar raken.

## Spectrum II

Vervolgens beschouwen we het spectrum dat verkregen was met behulp van methode II. Net zoals bij de bepaling van het snijpunt van Spectrum I met het RK3 stabiliteitsgebied, verwachten we ook hier dat het kritieke punt het snijpunt met de negatieve reële as is (zie figuur 5.2). We substitueren nu in (5.7.70) de waarden $a = -2$ en $b = 0$. Dit geeft dan de volgende formule voor $\nu$

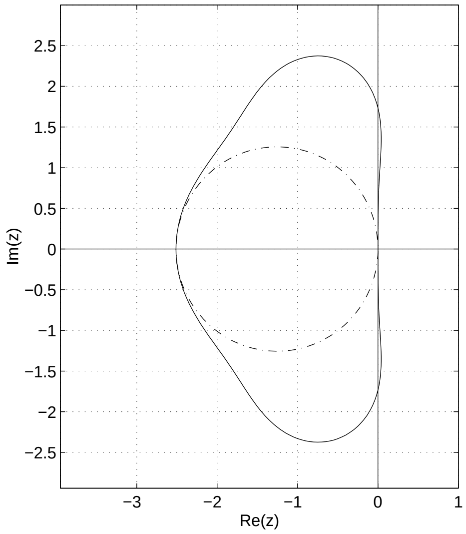
$$\frac{16}{9}\nu^6 - \frac{16}{3}\nu^5 + \frac{28}{3}\nu^4 - \frac{32}{3}\nu^3 + 8\nu^2 - 4\nu = 0. \tag{5.7.73}$$

Oplossen van deze vergelijking op numerieke wijze met behulp van MAPLE voor $\nu$ geeft dan $\nu_{II} = 1.256$. De CFL-conditie luidt dus

$$\nu_{II} \leq 1.256. \tag{5.7.74}$$

Als we Spectrum II vermenigvuldigen met $\nu_{II}$ en dit plotten in een figuur tezamen met het RK3 stabiliteitsgebied, dan is het resultaat te zien in figuur 5.11.

Zoals te zien is, raken de twee gebieden elkaar bij deze waarde van $\nu_{II}$.




Figuur 5.10: Het RK3 stabiliteitsgebied en Spectrum I vermenigvuldigd met $\nu_I = 0.628$.

### Spectrum III

Bij het spectrum verkregen door methode III, is het lastiger om de kritieke punten direct eruit te lichten. Er is daarom besloten om de drie eigenwaarden die het gebied opspannen (zie figuur 5.8), alle drie te beschouwen als kritieke punten. Vervolgens zal voor elk van de drie punten de bijbehorende $\nu$-waarde worden berekend. De *kleinste* $\nu$-waarde leidt dan tot de bij Spectrum III behorende CFL-conditie. Hieronder zijn de resultaten te zien van de $\nu$-waarden $\nu_1$, $\nu_2$ en $\nu_3$, behorende bij elk van de drie eigenwaarden $\lambda_1$, $\lambda_2$ en $\lambda_3$, berekend op dezelfde wijze als hiervoor

$$\lambda_1 = 2I \quad \text{en} \quad \nu_1 = \frac{1}{2}\sqrt{3}, \tag{5.7.75}$$

$$\lambda_2 = -(2+\sqrt{2}) + \sqrt{2}I \quad \text{en} \quad \nu_2 = 0.646, \tag{5.7.76}$$

$$\lambda_3 = -(1+\sqrt{2}) + (1+\sqrt{2})I \quad \text{en} \quad \nu_3 = 0.695. \tag{5.7.77}$$

We zien dat de kleinste $\nu$-waarde $\nu_2$ is, en dit leidt tot de volgende CFL-conditie

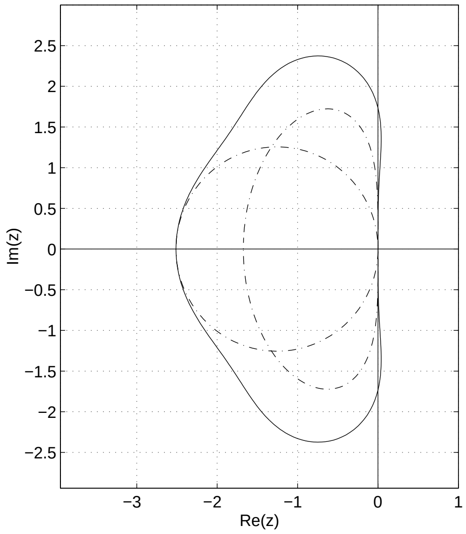
$$\nu_{III} \leq 0.646. \tag{5.7.78}$$

In figuur 5.12 is het RK3 stabiliteitsgebied te zien, tezamen met Spectrum III, welke is vermenigvuldigd met de zojuist bepaalde $\nu_{III}$.

## 5.8 Onderzoek naar het bepalende spectrum

Nu we bij elk spectrum de CFL-conditie hebben gevonden, zijn we klaar om te gaan onderzoeken welke van de drie spectra als beste het spectrum, behorende bij onze plaatsdiscretisatie, benadert. Onze hoop gaat uit naar Spectrum II. Immers bij dit spectrum hadden we het grootste CFL-getal gevonden. En dat geeft aan dat onze tijdstapgrootte groter gekozen kan worden.

Bij het bepalen van het beste spectrum, worden de verschillende $\nu$-waarden getest met behulp van het periodieke testprobleem, zoals beschreven is in paragraaf 4.10. We houden hierbij het rooster vast en zetten het aantal cellen op 80. Om de stabiliteit te onderzoeken worden de verschillende $\nu$-waarden gevarieerd. Zo beschouwen we achtereenvolgens $\frac{1}{2}\nu$, $0.95\nu$, $\nu$, $1.05\nu$ en $2\nu$. Op deze wijze behoren we volgens de theorie achtereenvolgens ruim in het stabiliteitsgebied, vlak voor de rand van het stabiliteitsgebied te zitten, precies op de rand, er net overheen en ruim er overheen. Verder laten we




Figuur 5.11: Het RK3 stabiliteitsgebied en Spectrum II vermenigvuldigd met $\nu_{II} = 1.256$.

de beginoplossing een flink aantal rondjes rondlopen. De manier van testen is in de onderstaande tabel verder uitgewerkt.

Tabel 5.2: Manier van testen van de verschillende $\nu$-waarden.

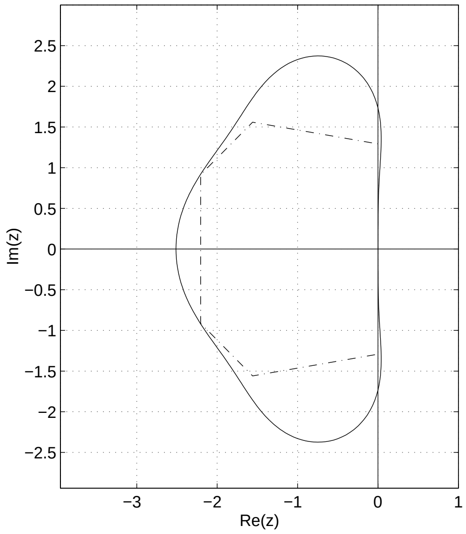
| Spectrum | I | II | III |
|---|---|---|---|
| $(\nu, n)$ | $(\frac{1}{2}\nu_I,\, 4000)$<br>$(0.95\nu_I,\, 2105)$<br>$(\nu_I,\, 2000)$<br>$(1.05\nu_I,\, 1905)$<br>$(2\nu_I,\, 1000)$ | $(\frac{1}{2}\nu_{II},\, 4000)$<br>$(0.95\nu_{II},\, 2105)$<br>$(\nu_{II},\, 2000)$<br>$(1.05\nu_{II},\, 1905)$<br>$(2\nu_{II},\, 1000)$ | $(\frac{1}{2}\nu_{III},\, 4000)$<br>$(0.95\nu_{III},\, 2105)$<br>$(\nu_{III},\, 2000)$<br>$(1.05\nu_{III},\, 1905)$<br>$(2\nu_{III},\, 1000)$ |
| $t_{max}$ | 15.7 | 31.4 | 16.2 |

In tabel 5.2 stelt $n$ het aantal tijdstappen voor. Voor elk spectrum laten we de beginoplossing hetzelfde aantal rondjes lopen. Omdat we te maken hebben met drie verschillende $\nu$-waarden, houdt dit natuurlijk in dat we per spectrum te maken hebben met een andere eindtijd $t_{max}$.

Om nog het verschil tussen de vier verschillende RK3 schema's te bekijken, wordt tabel 5.2 voor ieder RK3 schema afgewerkt. Op deze manier zien we in één klap welk spectrum het meest bepalend is en welk RK3 schema het meest stabiel is. De resultaten van de testen zijn te zien in de onderstaande drie tabellen.
In deze tabellen geven de getallen de relatieve rekentijden weer. Hierbij is de rekentijd die nodig was om het testprobleem op te lossen met behulp van Heun [4] en het CFL-getal $\nu = \frac{1}{2}\nu_{II}$, als referentietijd genomen. Deze referentietijd bedroeg 3.47 sec. De $\times$-en geven instabiliteit weer. Bij bepaalde waarden van $\nu$ en bij bepaalde RK3 schema's werd de oplossing, die steeds na een aantal tijdstappen werd weggeschreven, namelijk onrealistisch groot.

We kunnen aan de hand van deze tabellen de volgende conclusies trekken. Allereerst zien we in de tabellen behorende bij Spectrum I en II dat Heun [4] en Wray [21] stabieler zijn dan Kutta [11] en RK3 (Fehlberg) [2]. Bij toepassing van Heun [4] en Wray [21] blijft het schema bij toename van de $\nu$-waarden, langer stabiel. Qua rekensnelheid zijn alle vier de RK3 schema's ongeveer even snel, zoals de getallen in de tabellen duidelijk aangeven.

Als we verder naar de tabellen kijken, dan zien we dat de CFL-conditie behorende bij Spectrum I te conservatief is. We zien dat als het CFL-getal wordt verdubbeld, we dan instabiliteit krijgen bij bepaalde RK3 schema's. Bij Heun [4] en Wray [21] is er dan net wel stabiliteit, bij Kutta [11] en RK3




Figuur 5.12: Het RK3 stabiliteitsgebied en Spectrum III vermenigvuldigd met $\nu_{III} = 0.646$.

(Fehlberg) [2] is er net instabiliteit. Met andere woorden, een kleine verstoring kan in dit geval net wel of net niet voor instabiliteit zorgen.

Bij Spectrum II zien we dat er met deze CFL-conditie beter aan de theorie wordt voldaan. We zien dat $\nu_{II}$ een goede begrenzing aangeeft. Vlak voor deze begrenzing is er stabiliteit, eroverheen is er instabiliteit en precies op deze waarde zelf kan een kleine verstoring net wel of net geen stabiliteit geven.

Tenslotte zien we in de tabel van Spectrum III dat net als bij Spectrum I het CFL-getal te conservatief blijkt te zijn, echter in mindere mate dan bij Spectrum I.

In tabel 5.4 kunnen we nog iets interessants zien. Ik heb daar opnieuw Spectrum I en II getest met hetzelfde periodieke testprobleem. Er is echter nu een RK2 methode gebruikt voor de tijdsdiscretisatie; namelijk Improved Euler. Net zoals bij de vorige testen heb ik allereerst de $\nu$-waarden bepaald, zodanig dat het spectrum, vermenigvuldigd met deze waarde, precies het RK2-stabiliteitsgebied zou raken. Ik vond na enig rekenen

$$\begin{array}{ll} \text{Spectrum I}: & \nu_{Ia} = 0.500, \\ \text{Spectrum II}: & \nu_{IIa} = 0.873. \end{array} \tag{5.8.79}$$

Nu is het bekend dat het stabiliteitsgebied van een RK2 methode een ovaal is (zie figuur 5.9). Met andere woorden, als een spectrum eigenwaarden heeft op de imaginaire as, dan zal een RK2 methode altijd instabiel zijn. De testresultaten in tabel 5.4 laten duidelijk zien dat er bij dit testprobleem zich geen eigenwaarden op de imaginaire as bevinden.

We zouden op grond van de resultaten uit de tabellen 5.3 en 5.4 kunnen concluderen dat Spectrum II het echte spectrum, behorende bij onze plaatsdiscretisatie, het beste benadert. Echter deze conclusie is nog te voorbarig. Immers we hebben bij het testen slechts één testprobleem beschouwd. Het zou best wel eens kunnen dat bij een ander testprobleem het spectrum wel eigenwaarden heeft op de imaginaire as.

Ik wil daarom bij het vervolg van het onderzoek zowel Spectrum II als Spectrum III meenemen. Mischien dat verder onderzoek (zoals bijvoorbeeld het beschouwen van meerdere frequenties bij de constructie van Spectrum III) en meerdere testproblemen een beter beeld kunnen geven over welk spectrum beter is.

De volgende stap die we nu gaan nemen, is het vinden van RK4 schema's, waarvan de stabiliteitsgebieden resp. Spectrum II als Spectrum III zo goed mogelijk omsluiten. Mogelijk zijn we dan in staat om $\nu$-waarden te vinden, die ruimschoots (meer dan $\frac{4}{3}$ keer) groter zijn dan de $\nu$-waarden $\nu_{II}$ en $\nu_{III}$, die we eerder hadden gevonden.

Tabel 5.3: De resultaten van het stabiliteitsonderzoek voor elk van de drie spectra.

| Spectrum I | $\nu = \frac{1}{2}\nu_I$ | $\nu = 0.95\nu_I$ | $\nu = \nu_I$ | $\nu = 1.05\nu_I$ | $\nu = 2\nu_I$ |
|---|---|---|---|---|---|
| Heun [4] | 0.99 | 0.54 | 0.53 | 0.50 | 0.22 |
| Kutta [11] | 1.02 | 0.56 | 0.54 | 0.52 | × |
| RK3 (Fehlberg) [2] | 1.02 | 0.56 | 0.54 | 0.51 | × |
| Wray [21] | 1.00 | 0.56 | 0.53 | 0.50 | 0.22 |

| Spectrum II | $\nu = \frac{1}{2}\nu_{II}$ | $\nu = 0.95\nu_{II}$ | $\nu = \nu_{II}$ | $\nu = 1.05\nu_{II}$ | $\nu = 2\nu_{II}$ |
|---|---|---|---|---|---|
| Heun [4] | 1.00 | 0.46 | 0.38 | × | × |
| Kutta [11] | 1.02 | 0.46 | × | × | × |
| RK3 (Fehlberg) [2] | 1.03 | 0.42 | × | × | × |
| Wray [21] | 1.00 | 0.45 | 0.39 | × | × |

| Spectrum III | $\nu = \frac{1}{2}\nu_{III}$ | $\nu = 0.95\nu_{III}$ | $\nu = \nu_{III}$ | $\nu = 1.05\nu_{III}$ | $\nu = 2\nu_{III}$ |
|---|---|---|---|---|---|
| Heun [4] | 0.99 | 0.55 | 0.52 | 0.49 | × |
| Kutta [11] | 1.02 | 0.56 | 0.54 | 0.51 | × |
| RK3 (Fehlberg) [2] | 1.03 | 0.56 | 0.54 | 0.51 | × |
| Wray [21] | 1.00 | 0.56 | 0.53 | 0.51 | × |

Tabel 5.4: Verder stabiliteitsonderzoek bij Spectrum I en II.

| Spectrum I | $\nu = \frac{1}{2}\nu_{Ia}$ | $\nu = 0.95\nu_{Ia}$ | $\nu = \nu_{Ia}$ | $\nu = 1.05\nu_{Ia}$ | $\nu = 2\nu_{Ia}$ |
|---|---|---|---|---|---|
| Improved Euler | 0.68 | 0.38 | 0.37 | 0.35 | 0.16 |

| Spectrum II | $\nu = \frac{1}{2}\nu_{IIa}$ | $\nu = 0.95\nu_{IIa}$ | $\nu = \nu_{IIa}$ | $\nu = 1.05\nu_{IIa}$ | $\nu = 2\nu_{IIa}$ |
|---|---|---|---|---|---|
| Improved Euler | 0.68 | 0.38 | 0.30 | 0.28 | × |

## 5.9 Optimalisering van het RK4 schema

We beschouwen het volgende RK4 schema, op compacte wijze weergeven in het onderstaande Butcher array.

$$
\begin{array}{c|cccc}
 & a_{21} & & & \\
 & 0 & a_{32} & & \\
 & 0 & 0 & a_{43} & \\
\hline
 & b_1 & b_2 & b_3 & b_4
\end{array}
$$

Vorm van het te optimaliseren RK4 schema.

Zoals te zien is, staan in dit Butcher array op dezelfde plaatsen de niet-nullen als bij het standaard RK4 schema. Een schema van deze vorm willen we nu gaan optimaliseren. In formule vorm ziet het schema er als volgt uit

$$\begin{aligned}
C^{(1)} &= C^n \\
C^{(2)} &= C^n + a_{21}\tau f(C^{(1)}) \\
C^{(3)} &= C^n + a_{32}\tau f(C^{(2)}) \\
C^{(4)} &= C^n + a_{43}\tau f(C^{(3)}) \\
\\
C^{n+1} &= C^n + \tau\left(b_1 f(C^{(1)}) + b_2 f(C^{(2)}) + b_3 f(C^{(3)}) + b_4 f(C^{(4)})\right).
\end{aligned} \tag{5.9.80}$$

Laten we eens de formule voor het lineaire stabiliteitsgebied gaan opstellen. Allereerst gaan we de formule voor de versterkingsfactor afleiden. Op geheel dezelfde wijze als gedaan is in paragraaf 5.6 vinden we

$$\begin{aligned}
k(\tau\lambda) &= 1 + (b_1+b_2+b_3+b_4)(\tau\lambda) + (b_2a_{21}+b_3a_{32}+b_4a_{43})(\tau\lambda)^2 \\
&+ (b_3a_{32}a_{21}+b_4a_{43}a_{32})(\tau\lambda)^3 + (a_{21}a_{32}a_{43}b_4)(\tau\lambda)^4.
\end{aligned} \tag{5.9.81}$$

In paragraaf 5.6 is verteld dat als (5.6.58) geldt, de globale afbreekfout van de orde vier is. Dan moet dus gelden

$$\begin{aligned}
b_1+b_2+b_3+b_4 &= 1 \\
b_2a_{21}+b_3a_{32}+b_4a_{43} &= \tfrac{1}{2} \\
b_3a_{32}a_{21}+b_4a_{43}a_{32} &= \tfrac{1}{6} \\
a_{21}a_{32}a_{43}b_4 &= \tfrac{1}{24}
\end{aligned} \tag{5.9.82}$$

Omdat de plaatsdiscretisatie van de orde drie is, één orde lager dan bij ons RK4 schema, kunnen we een orde van ons RK4 schema laten vallen. We laten daarom de coëfficiënt van $(\tau\lambda)^4$ variabel. Op deze wijze kunnen we het volgende stabiliteitsgebied opstellen, behorende bij ons RK4 schema, dat nu derde orde nauwkeurig is voor lineaire problemen[1]

$$S_{4\beta} = \left\{ z \in \mathbb{C} : \left| 1 + z + \frac{1}{2}z^2 + \frac{1}{6}z^3 + \beta z^4 \right| \le 1 \right\}, \quad \text{met} \quad \beta = a_{21}a_{32}a_{43}b_4. \tag{5.9.83}$$

De volgende stap is om de variabele $\beta$ zodanig te gaan bepalen, dat ten eerste het spectrum zo goed mogelijk wordt omsloten en ten tweede het CFL-getal zo groot mogelijk wordt. We passen hierbij dezelfde werkwijze toe als bij paragraaf 5.7.

Allereerst beschouwen we de stabiliteitseis, geldend voor ons RK4$\beta$ schema. Deze luidt

$$\left| 1 + \lambda\tau + \frac{1}{2}(\lambda\tau)^2 + \frac{1}{6}(\lambda\tau)^3 + \beta(\lambda\tau)^4 \right| \le 1. \tag{5.9.84}$$

Vervolgens substitueren we (5.7.68) in (5.9.84). Als we alles uitwerken dan resulteert dit in een soortgelijke formule als (5.7.70)

$$\begin{aligned}
&\left\{1 + \nu a + \tfrac{1}{2}\nu^2\left(a^2-b^2\right) + \tfrac{1}{6}\nu^3 a\left(a^2-3b^2\right) + \beta\nu^4(a^4-6a^2b^2+b^4)\right\}^2 + \\
&b^2\left\{\nu + \nu^2 a + \tfrac{1}{6}\nu^3\left(3a^2-b^2\right) + 4\beta\nu^4(a^3-ab^2)\right\}^2 \le 1.
\end{aligned} \tag{5.9.85}$$

We kunnen met behulp van (5.9.85) bij gegeven $\beta$, $a$ en $b$ de waarde $\nu$ bepalen, zodanig dat bij deze waarde het spectrum precies raakt aan het stabiliteitsgebied. Hieruit volgt dan meteen de CFL-conditie. Om nu de $\beta$ te vinden waarbij de optimale CFL-conditie wordt gevonden ben ik als volgt te werk gegaan bij de beide spectra.

Ik heb als basis dezelfde methode toegepast zoals is uitgelegd in paragraaf 5.7. We kunnen in dit geval echter niet zomaar op voorhand een kritiek punt uit het spectrum kiezen. De reden hiervoor is dat het mogelijk is dat bij een bepaalde $\beta$ het stabiliteitsgebied zo zou kunnen veranderen, dat een ander punt van het spectrum een kritiek punt wordt. Ik ben daarom bij een gegeven $\beta$ voor een flink

[1] De situatie voor niet-lineaire problemen komt nog later aan de orde.

aantal punten van het spectrum de $\nu$-waarde gaan bepalen. Vervolgens ben ik op zoek gegaan naar de *minimale* $\nu$-waarde behorende bij al die punten. Bij deze waarde van $\nu$ (bij een nog steeds vaste en gegeven waarde van $\beta$) wist ik namelijk dat er tenminste één punt het stabiliteitsgebied raakte en de overige punten van het spectrum er binnenin zaten. Vervolgens ben ik dit voor een aantal waarden van $\beta$ gaan doen. Bij al deze $\beta$'s kreeg ik een $\nu$-waarde en de maximale waarde van al deze minimale $\nu$'s resulteerde in een optimale CFL-conditie en een stabiliteitsgebied dat het betreffende spectrum zo goed mogelijk omsloot.

Voor deze hele bovenstaande procedure heb ik een MAPLE programma geschreven. Zo verdeelde ik bij Spectrum III elk van de drie stukken, waaruit de bovenste helft van het spectrum bestaat, in 100 stukken. Verder beschouwde ik $\beta$'s uit het interval $[0, 0.05]$, dat ik eveneens in 100 stukken had verdeeld. Door enig onderzoek wist ik namelijk dat in dit interval de optimale $\beta$ zou liggen. Met behulp van dit MAPLE programma vond ik op numerieke wijze als optimale $\beta$

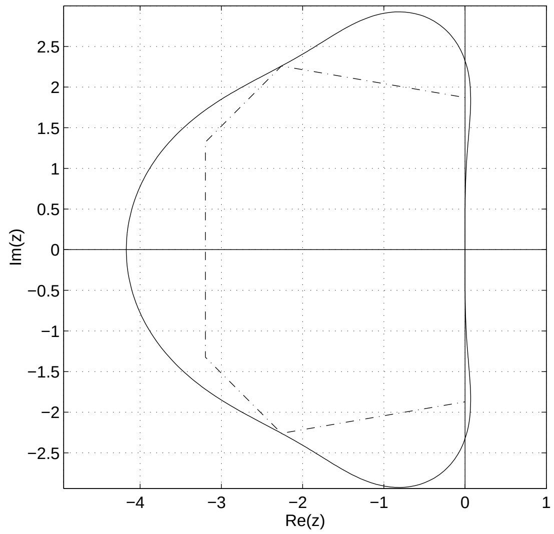
$$\beta_{opt} = \frac{1}{40}. \tag{5.9.86}$$

Als optimale CFL-conditie werd gevonden

$$\nu_{opt_{III}} \leq 0.936. \tag{5.9.87}$$

(Later heb ik uit het interval $[0, 0.05]$ nog een deelinterval beschouwd, namelijk $[0.023, 0.026]$. Dit deelinterval verdeelde ik ook weer in 100 stukken en met behulp van het MAPLE programma vond ik toen $\beta_{opt} = 0.0248$ en als CFL-conditie vond ik een waarde die pas in de vijfde decimaal achter de komma verschilde van hetgeen ik de eerste keer vond. Ik vond dit zo een marginaal verschil dat ik de resultaten in (5.9.86) en (5.9.87) onveranderd heb gelaten.)

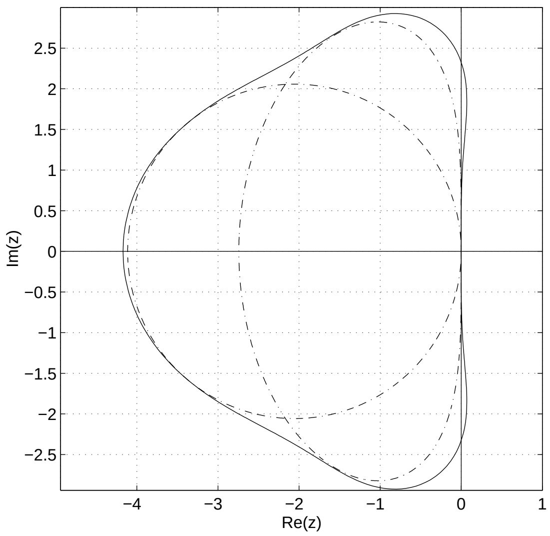
Het stabiliteitsgebied behorende bij $\beta_{opt}$ en het spectrum vermenigvuldigd met $\nu_{opt_{III}}$ is in figuur 5.13 te zien.




Figuur 5.13: Het stabiliteitsgebied bij $\beta_{opt} = \frac{1}{40}$ en het spectrum vermenigvuldigd met $\nu_{optIII} = 0.936$.

Bij Spectrum II ben ik op dezelfde wijze te werk gegaan als hierboven. Hierbij ben ik in eerste instantie de twee spectra, waaruit Spectrum II bestaat, apart gaan beschouwen. Allereerst bekeek ik het ($\kappa = \frac{1}{3}$)-spectrum en op dezelfde werkwijze vond ik met behulp van het MAPLE programma $\beta = 0.0245$ en $\nu = 2.064$. Bij deze waarden bevond zich ook het $\mathcal{O}(h)$ upwind schema geheel binnen het stabiliteitsgebied. Daarna onderzocht ik apart het $\mathcal{O}(h)$ upwind schema. Ik kreeg toen de waarden $\beta = 0.0238$ en $\nu = 2.072$. Weliswaar vond ik toen een grotere CFL-waarde, echter bij deze waarde bleek dat het ($\kappa = \frac{1}{3}$)-spectrum voor een gedeelte buiten het stabiliteitsgebied kwam te liggen. We kunnen nu concluderen dat een optimale $\beta$- en $\nu$-waarde tussen de beschouwde waarden moet inliggen.

Ik heb ook nog gekeken in hoeverre de $\nu$-waarde zou verschillen van de gevonden waarden indien Spectrum II zou worden opgeblazen, totdat het stabiliteitsgebied bij $\beta = \frac{1}{40}$ zou worden geraakt. Het bleek dat dit gebeurde bij $\nu = 2.057$. Omdat deze waarde zo weinig verschilt van $\nu = 2.064$ of $\nu = 2.072$, kan in feite voor zowel Spectrum II als Spectrum III hetzelfde RK4$\beta$ schema worden gebruikt. Ik zal deze $\nu$-waarde ($\nu = 2.057$) behorende bij $\beta = \frac{1}{40}$ beschouwen als een optimale $\nu$-waarde voor Spectrum II en deze aanduiden met $\nu_{optII}$. In figuur 5.14 is Spectrum II, vermenigvuldigd met deze $\nu$-waarde, te zien in het RK4$\beta$ stabiliteitsgebied.




Figuur 5.14: Het stabiliteitsgebied bij $\beta_{opt} = \frac{1}{40}$ en het spectrum vermenigvuldigd met $\nu_{optII} = 2.057$.

Nu we weten hoe het stabiliteitsgebied van het RK4$\beta$ schema eruit ziet, willen we graag weten hoe het schema er uit ziet in een Butcher array. We hebben al de volgende vier vergelijkingen

$$\begin{aligned} b_1 + b_2 + b_3 + b_4 &= 1 \\ b_2 a_{21} + b_3 a_{32} + b_4 a_{43} &= \tfrac{1}{2} \\ b_3 a_{32} a_{21} + b_4 a_{43} a_{32} &= \tfrac{1}{6} \\ a_{21} a_{32} a_{43} b_4 &= \tfrac{1}{40}. \end{aligned} \tag{5.9.88}$$

Echter deze vergelijkingen waren afgeleid voor een RK4$\beta$ schema dat derde orde nauwkeurig was voor *lineaire* problemen. We willen graag dat het schema ook derde orde nauwkeurig is voor *niet-lineaire* problemen. We gaan daarom dezelfde nauwkeurigheidsanalyse doen als in paragraaf 5.6. Op geheel dezelfde wijze vinden we dan dat moet gelden, wil dit schema derde orde nauwkeurig zijn voor niet-lineaire problemen

$$\begin{aligned} C^{n+1} &= (b_1 + b_2 + b_3 + b_4)\tau f(C^n) \\ &+ (b_2 a_{21} + b_3 a_{32} + b_4 a_{43})\tau^2 f(C^n) f'(C^n) \\ &+ (b_3 a_{32} a_{21} + b_4 a_{43} a_{32})\tau^3 f(C^n) f'^2(C^n) \\ &+ \tfrac{1}{2}(b_2 a_{21}^2 + b_3 a_{32}^2 + b_4 a_{43}^2)\tau^3 f^2(C^n) f''(C^n) + \mathcal{O}(\tau^4). \end{aligned} \tag{5.9.89}$$

Voor derde orde nauwkeurigheid voor niet-lineaire problemen moet dan gelden, als we (5.9.89) vergelijken met (5.6.61)

$$\begin{aligned} b_1 + b_2 + b_3 + b_4 &= 1 \\ b_2 a_{21} + b_3 a_{32} + b_4 a_{43} &= \tfrac{1}{2} \\ b_3 a_{32} a_{21} + b_4 a_{43} a_{32} &= \tfrac{1}{6} \\ b_2 a_{21}^2 + b_3 a_{32}^2 + b_4 a_{43}^2 &= \tfrac{1}{3}. \end{aligned} \tag{5.9.90}$$

Als we vervolgens (5.9.88) combineren met (5.9.90), dan krijgen we precies de vergelijkingen waarmee het Butcher array behorende bij dit RK4$\beta$ schema blijkt te kunnen worden opgesteld. De vergelijkingen luiden

$$\begin{aligned}
b_1 + b_2 + b_3 + b_4 &= 1 \\
b_2 a_{21} + b_3 a_{32} + b_4 a_{43} &= \tfrac{1}{2} \\
b_3 a_{32} a_{21} + b_4 a_{43} a_{32} &= \tfrac{1}{6} \\
b_2 a_{21}^2 + b_3 a_{32}^2 + b_4 a_{43}^2 &= \tfrac{1}{3} \\
a_{21} a_{32} a_{43} b_4 &= \tfrac{1}{40}.
\end{aligned} \tag{5.9.91}$$

We zien dat we vijf vergelijkingen hebben met zeven onbekenden. We kunnen dus twee onbekenden vrij kiezen. Om het schema zo eenvoudig mogelijk te houden, stellen we die twee nog vrij te kiezen onbekenden gelijk aan nul. Gemakkelijk is het om in te zien dat als er twee onbekenden uit $a_{21}$, $a_{32}$, $a_{43}$ of $b_4$ nul worden gekozen, dit leidt tot inconsistentie. Immers er wordt dan niet meer voldaan aan de vijfde vergelijking in (5.9.91). De twee onbekenden moeten dus worden gekozen uit $b_1$, $b_2$ of $b_3$. Als we bijvoorbeeld kiezen $b_2 = b_3 = 0$, dan resulteert dit in

$$\begin{aligned}
b_1 + b_4 &= 1 \\
b_4 a_{43} &= \tfrac{1}{2} \\
b_4 a_{43} a_{32} &= \tfrac{1}{6} \\
b_4 a_{43}^2 &= \tfrac{1}{3} \\
a_{21} a_{32} a_{43} b_4 &= \tfrac{1}{40}.
\end{aligned} \tag{5.9.92}$$

Na enig rekenen, kunnen we het volgende Butcher array construeren, behorende bij ons nieuwe schema RK4$\beta_1$.

$$\begin{array}{c|cccc}
 & \frac{3}{20} & & & \\
 & 0 & \frac{1}{3} & & \\
 & 0 & 0 & \frac{2}{3} & \\
\hline
 & \frac{1}{4} & 0 & 0 & \frac{3}{4}
\end{array}$$

Het RK4$\beta_1$ schema.

Als we kiezen $b_1 = b_3 = 0$ en dit substitueren in (5.9.91) dan krijgen we het volgende stelsel vergelijkingen

$$\begin{aligned}
b_2 + b_4 &= 1 \\
b_2 a_{21} + b_4 a_{43} &= \tfrac{1}{2} \\
b_4 a_{43} a_{32} &= \tfrac{1}{6} \\
b_2 a_{21}^2 + b_4 a_{43}^2 &= \tfrac{1}{3} \\
a_{21} a_{32} a_{43} b_4 &= \tfrac{1}{40}.
\end{aligned} \tag{5.9.93}$$

Oplossen van dit stelsel vergelijkingen leidt tot het volgende Butcher array behorende bij ons nieuwe RK4$\beta_2$ schema.

$$\begin{array}{c|cccc}
 & \frac{3}{20} & & & \\
 & 0 & \frac{247}{651} & & \\
 & 0 & 0 & \frac{31}{42} & \\
\hline
 & 0 & \frac{100}{247} & 0 & \frac{147}{247}
\end{array}$$

Het RK4$\beta_2$ schema.

Tenslotte kiezen we $b_1 = b_2 = 0$. Wederom substitueren in (5.9.91) resulteert in het volgende stelsel vergelijkingen

$$\begin{aligned} b_3 + b_4 &= 1 \\ b_3 a_{32} + b_4 a_{43} &= \tfrac{1}{2} \\ b_3 a_{32} a_{21} + b_4 a_{43} a_{32} &= \tfrac{1}{6} \\ b_3 a_{32}^2 + b_4 a_{43}^2 &= \tfrac{1}{3} \\ a_{21} a_{32} a_{43} b_4 &= \tfrac{1}{40}. \end{aligned} \tag{5.9.94}$$

Oplossen van dit stelsel vergelijkingen geeft tenslotte het Butcher array behorende bij ons nieuwe RK4$\beta_3$ schema.

$$\begin{array}{c|cccc} & 0.319 & & & \\ & 0 & 0.351 & & \\ & 0 & 0 & 1.058 & \\ \hline & 0 & 0 & 0.789 & 0.211 \end{array}$$

Het RK4$\beta_3$ schema.

## 5.10 Numerieke testresultaten voor het RK4$\beta_1$ schema

In de vorige paragraaf zijn drie RK4$\beta$ butcher arrays afgeleid en we gaan in deze paragraaf met één vam deze butcher arrays Spectrum II en Spectrum III testen. Vanwege de eenvoud van het RK4$\beta_1$ schema, heb ik dit schema gekozen bij het testen van de beide spectra.

Netzoals in paragraaf 5.8 beschouwen we weer het periodieke testprobleem en passen we weer dezelfde wijze van testen toe zoals te zien is in de onderstaande tabel 5.5.

Tabel 5.5: Manier van testen van de verschillende $\nu$-waarden.

| Spectrum | II | III |
|---|---|---|
| $(\nu, n)$ | $(\frac{1}{2}\nu_{II}, 4000)$<br>$(0.9\nu_{II}, 2222)$<br>$(\nu_{II}, 2000)$<br>$(1.05\nu_{II}, 1818)$<br>$(2\nu_{II}, 1000)$ | $(\frac{1}{2}\nu_{III}, 4000)$<br>$(0.9\nu_{III}, 2222)$<br>$(\nu_{III}, 2000)$<br>$(1.05\nu_{III}, 1818)$<br>$(2\nu_{III}, 1000)$ |
| $t_{max}$ | 51.425 | 23.400 |

De resultaten van deze testen zijn te zien in de onderstaande tabellen.

Tabel 5.6: Stabiliteitsonderzoek van het RK4$\beta_1$ schema bij Spectrum II en III.

| Spectrum II | $\nu = \frac{1}{2}\nu_{optII}$ | $\nu = 0.9\nu_{optII}$ | $\nu = \nu_{optII}$ | $\nu = 1.1\nu_{optII}$ | $\nu = 2\nu_{optII}$ |
|---|---|---|---|---|---|
| RK4$\beta_1$ | 1.30 | 0.62 | $\times$ | $\times$ | $\times$ |

| Spectrum III | $\nu = \frac{1}{2}\nu_{optIII}$ | $\nu = 0.9\nu_{optIII}$ | $\nu = \nu_{optIII}$ | $\nu = 1.1\nu_{optIII}$ | $\nu = 2\nu_{optIII}$ |
|---|---|---|---|---|---|
| RK4$\beta_1$ | 1.30 | 0.69 | 0.68 | 0.61 | 0.30 |

De kruisen geven weer instabiliteit aan en de getallen zijn net zoals in paragraaf 5.8 bepaald aan de hand van dezelfde referentietijd 3.47 sec.

We zien dat bij Spectrum III het CFL-getal te conservatief is, zoals we eigenlijk al op grond van de resultaten in paragraaf 5.8 hadden kunnen concluderen. We zien dat als we de $\nu$-waarde verdubbelen, er nog steeds geen instabiliteit optreedt. Het blijkt uit verder onderzoek dat pas bij drie keer de $\nu$-waarde, de numerieke oplossing pas echt instabiel wordt.

In de tabel met Spectrum II zien we opnieuw dat bij dit testprobleem, de CFL-waarde het beste voldoet aan de theorie. Precies bij de waarde $\nu = \nu_{optII}$ is er instabiliteit, even ervoor bij $\nu = 0.9\nu_{optII}$ is de numerieke oplossing wel stabiel.

## 5.11 Vereenvoudiging van het RK4$\beta_1$ schema

In deze paragraaf wil ik in het kort ingaan op een methode, waarmee op eenvoudige wijze von Neumann stabiliteitscondities kunnen worden afgeleid voor een groot aantal tijdsdicretisatie schema's voor de convectie-diffusie vergelijking. Deze methode is uiteengezet in [19] en in dit artikel is voor de plaatsdiscretisatie zowel het $\kappa$-schema als het vierde orde centrale schema beschouwd.

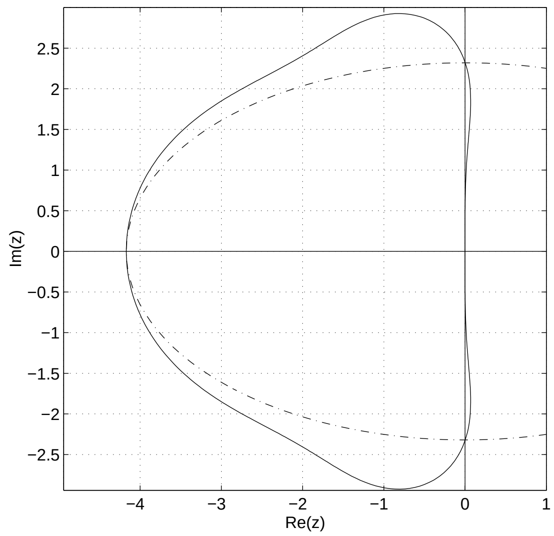
Waar het nu in essentie om gaat in [19] is dat een stabiliteitsgebied, behorende bij een bepaalde tijdstapsmethode, wordt opgevuld met behulp van een mal. Zo'n mal kan bijvoorbeeld een ellips, een ovaal of een parabool zijn, maar ook bijvoorbeeld een combinatie van die figuren. Om nu een bepaalde mal in het stabiliteitsgebied te kunnen passen, moeten er eerst aan een aantal voorwaarden worden voldaan. En deze voorwaarden zijn equivalent aan een tijdstaprestrictie, die garandeert dat de numerieke oplossing niet instabiel wordt.

Ik wil nu gaan onderzoeken in hoeverre de methode in dit artikel, toegepast op het RK4$\beta_1$ schema, al dan niet zal gaan leiden tot een te conservatieve CFL-conditie, vergeleken met de numerieke testresultaten in de vorige paragraaf. Voor de plaatsdiscretisatie zullen we het ($\kappa = \frac{1}{3}$)-schema nemen, alhoewel dit een sterke vereenvoudiging is op de door ons gebruikte plaatsdiscretisatie.

Omdat we een pure convectie vergelijking hebben in één dimensie, kunnen we alleen voldoen aan de voorwaarden in stelling 3.3 van [19]. Toepassen van deze stelling geeft dat we in het RK4$\beta_1$ stabiliteitsgebied de volgende ellips kunnen passen

$$\left(\frac{v}{4.17}\right)^2 + \left(\frac{w}{2.32}\right)^2 = 1, \quad \text{met} \quad v + iw = z. \qquad (5.11.95)$$

Hierbij zijn de lengte van de korte en lange as van de ellips zo bepaald dat deze samenvallen met de snijpunten van het stabiliteitsgebied met de reële en de imaginaire as. In figuur 5.15 is deze ellips te zien in het RK4$\beta_1$ stabiliteitsgebied. Zoals te zien is vult de ellips het gebied niet optimaal op.




Figuur 5.15: Het stabiliteitsgebied bij $\beta_{opt} = \frac{1}{40}$ (–) en de ellips (-·).

De tijdstaprestrictie volgens stelling 3.3 van [19], vereenvoudigd voor ons één-dimensionale pure convectie probleem, luidt dan

$$\tau \leq \min\left[\frac{\frac{1}{2}ah}{(1-\kappa)|u|}, \frac{bh}{(2-\kappa)|u|}\right]. \tag{5.11.96}$$

Als we vervolgens al onze gegevens substitueren; $a = 4.17$, $b = 2.32$, $h = \frac{1}{80}$, $\kappa = \frac{1}{3}$ en $|u| = 1$, dan geeft dit

$$\tau \leq \min[0.0391, 0.0174] = 0.0174, \tag{5.11.97}$$

hetgeen resulteert in de volgende CFL-conditie

$$\nu \leq 1.392. \tag{5.11.98}$$

Vergeleken met de numerieke resultaten in de vorige paragraaf, zien we dat deze methode tot een enigszins te conservatieve CFL-conditie leidt in vergelijking tot de CFL-conditie behorende bij Spectrum II.

Ik heb reeds opgemerkt dat de ellips in figuur 5.15 het stabiliteitsgebied niet geheel opvult. Ik heb gezocht naar een mal die dit beter doet. Het blijkt dat als de formule van een verschoven ellips wordt gebruikt er betere resultaten kunnen worden verkregen. De formule voor een ellips met middelpunt $(-\alpha, 0)$ luidt

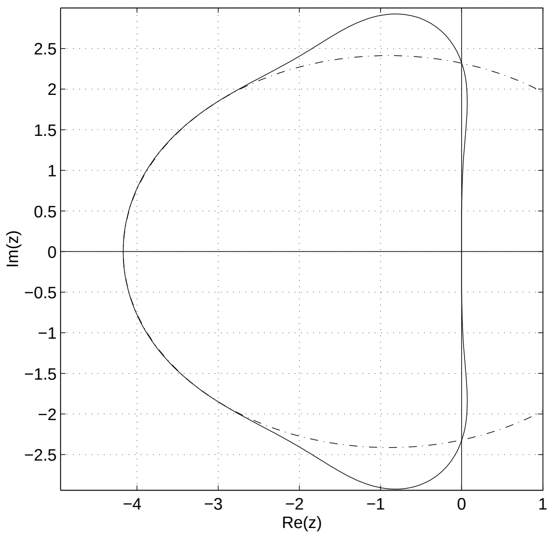
$$\frac{(v+\alpha)^2}{a^2} + \frac{w^2}{b^2} = 1 \quad \text{met} \quad v + iw = z. \tag{5.11.99}$$

Om ervoor te zorgen dat de ellips dezelfde raakpunten als het stabiliteitsgebied heeft met de twee assen, moet er gelden

$$a = \alpha - 4.17, \tag{5.11.100}$$

$$b = \sqrt{\frac{(2.32)^2}{1 - \left(\frac{\alpha}{\alpha-4.17}\right)^2}}. \tag{5.11.101}$$

Als we kiezen $\alpha = 0.9$, dan zien we het resultaat in figuur 5.16.




Figuur 5.16: Het stabiliteitsgebied bij $\beta_{opt} = \frac{1}{40}$ (–) en de verschoven ellips (-·).

Zoals te zien is wordt met deze ellips het RK4$\beta_1$ stabiliteitsgebied beter opgevuld. Het probleem echter is dat het mij niet gelukt is om bij deze formule voor de verschoven ellips een formule te vinden voor de tijdstaprestrictie. Mischien dat verder onderzoek hiernaar betere resultaten zullen opleveren.

## 5.12 De Molenkamp test

In deze paragraaf wil ik de twee optimale CFL-condities, die bepaald waren in paragraaf 5.9 voor de twee spectra, gaan testen op een twee-dimensionaal testprobleem. Het doel hiervan is om te kijken in hoeverre deze CFL-condities al dan niet in het twee-dimensionale geval te ruim gekozen zijn of juist te klein.

Het twee-dimensionale testprobleem dat ik ga beschouwen staat bekend als *De Molenkamp test.* Dit probleem modelleert de transport van een 'wolk' materiaal (bijvoorbeeld een giftige stof in de atmosfeer) zonder dat de geometrie van deze wolk verandert. Dit probleem wordt beschreven door de volgende formules:

$$\frac{\partial C}{\partial t} + u\frac{\partial C}{\partial x} + v\frac{\partial C}{\partial y} = 0 \quad -1 \leq x \leq 1 \qquad -1 \leq y \leq 1, \tag{5.12.102}$$

waarbij het snelheidsveld $(u, v)$ een pure rotatie beschrijft van een star lichaam (in dit geval de wolk, waarvan we eisen dat de geometrie niet verandert)

$$u = -\omega y, \quad v = \omega x, \quad \text{met} \quad \omega = 2\pi. \tag{5.12.103}$$

De beginoplossing is een Gaussische verdeling, welke enigszins buiten het domein treedt:

$$C(x, y, 0) = 0.01^{4r^2}, \quad \text{met} \quad r = \sqrt{\left(x + \frac{1}{2}\right)^2 + y^2}. \tag{5.12.104}$$

Op de instroomrand, waar de exacte oplossing,

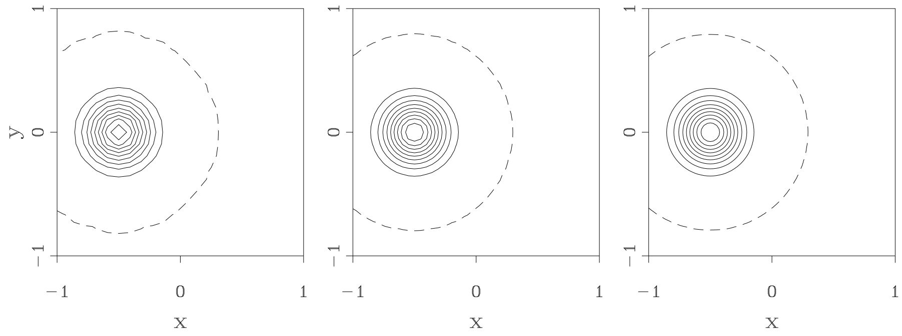
$$C(x, y, t) = 0.01^{4r^2}, \quad \text{met} \quad r = \sqrt{\left(x + \frac{1}{2}\cos(\omega t)\right)^2 + \left(y + \frac{1}{2}\sin(\omega t)\right)^2} \tag{5.12.105}$$

is opgelegd, hebben we de volgende Dirichlet randvoorwaarden

$$\{(x, y) \mid x = -1, -1 \leq y \leq 0\}, \tag{5.12.106a}$$
$$\{(x, y) \mid x = +1, 0 \leq y \leq 1\}, \tag{5.12.106b}$$
$$\{(x, y) \mid y = -1, 0 \leq x \leq 1\}, \tag{5.12.106c}$$
$$\{(x, y) \mid y = +1, -1 \leq x \leq 0\}. \tag{5.12.106d}$$

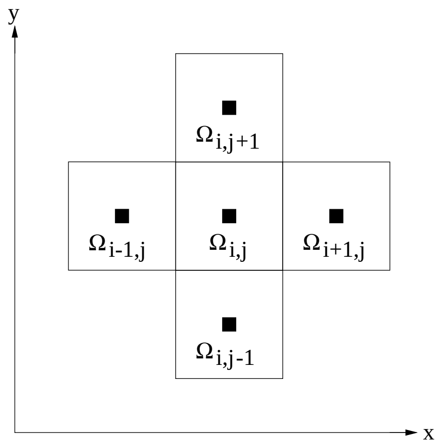
We willen nu graag de numerieke oplossing hebben na 1 omwenteling; d.w.z. op $t = 1$. De exacte oplossing is dan precies gelijk aan de beginoplossing. In figuur 5.17 is de exacte discrete oplossing te zien op drie verschillende roosters. Tevens zijn er contour-lijnen te zien voor $C = 10^{-5} + 0.1n$, met $n = 0, 1, 2, \ldots, 9$. De contour-lijn $C = 10^{-5}$ is gearceerd.

Bij het op numerieke wijze oplossen van dit probleem gaan we opnieuw de methode der lijnen toepassen. Allereerst gaan we ons bezig houden met de plaatsdiscretisatie, welke zal worden uitgevoerd met behulp van de EVM. Hiertoe gaan we allereerst het domein, wat een vierkant is, opdelen in celgecentreerde eindige volumina $\Omega_{i,j}$, zoals is weergegeven in figuur 5.18.
Net zoals in het één dimensionale geval zijn de volumina zo gekozen, dat in het midden van elk volume zich precies één onbekende, $C_{i,j} = C(ih, jh)$, bevindt. Hierbij komen de $i$- en $j$-richtingen overeen met de richting van de $x$- en $y$-as. Toepassen van de EVM op (5.12.102) betekent weer integreren over $\Omega_{i,j}$, hetgeen resulteert in




21 × 21-grid 41 × 41-grid 81 × 81-grid

Figuur 5.17: De exacte discrete oplossingen van de Molenkamp test op drie verschillende roosters.




Figuur 5.18: Celgecentreerde eindige volumes in twee dimensies.

$$\frac{d}{dt}\int\int_{\Omega_{i,j}} C dx dy + \int_{\partial\Omega_{i+\frac{1}{2},j}} C(\mathbf{u}\cdot\mathbf{n})ds + \int_{\partial\Omega_{i-\frac{1}{2},j}} C(\mathbf{u}\cdot\mathbf{n})ds \int_{\partial\Omega_{i,j+\frac{1}{2}}} C(\mathbf{u}\cdot\mathbf{n})ds + \int_{\partial\Omega_{i,j-\frac{1}{2}}} C(\mathbf{u}\cdot\mathbf{n})ds = 0. \tag{5.12.107}$$

Uitwerken van de bovenstaande volume-integraal en de randintegralen, geeft dan de volgende semi-gediscretiseerde vergelijking

$$\frac{dC_{i,j}}{dt} + \frac{1}{h}\left\{(uC)_{i+\frac{1}{2},j} - (uC)_{i-\frac{1}{2},j}\right\} + \frac{1}{h}\left\{(vC)_{i,j+\frac{1}{2}} - (vC)_{i,j-\frac{1}{2}}\right\} = 0. \tag{5.12.108}$$

De formules waarmee de fluxen, $(uC)_{i+\frac{1}{2},j}$ en $(vC)_{i,j+\frac{1}{2}}$, op de horizontale en verticale celwanden worden berekend, zijn te zien in [9], blz 124. In deze formules wordt voor de flux-limiter, $\phi = \phi(r)$, (4.8.54) gebruikt. De tijdsdiscretisatie wordt gerealiseerd met behulp van het RK4$\beta_1$ schema.

Met behulp van een door Barry Koren geschreven Fortran programma, heb ik getest in hoeverre de CFL-condities behorende bij Spectrum II en III in het twee-dimensionale probleem bevredigende resultaten zouden geven. Als CFL-getal definieerden we

$$\nu = \max\left\{\frac{\tau u}{h}, \frac{\tau v}{h}\right\}. \tag{5.12.109}$$

Het bleek echter dat beide CFL-condities te zwak waren. Pas bij $0.125\nu_{optIII}$ hadden we convergentie.

De reden van dit grote verschil tussen de CFL-condities in ons één- en twee-dimensionale probleem heeft te maken met het feit dat ten eerste ons spectrum er in het twee dimensionale probleem er een stuk anders uit ziet. Ten tweede hebben we in het één-dimensionale geval de beide CFL-condities bepaald met behulp van von Neumann. Dit was gerechtvaardigd omdat we in het één-dimensionale probleem periodieke randwaarden hadden. De grote discrepantie in het twee-dimensionale probleem heeft er dus alles mee te maken met het feit dat we hier geen periodieke randwaarden hebben. Dus in dat geval gaat de von Neumann analyse *niet* op. En dit is dan ook volgens ons de verklaring voor het grote verschil tussen de CFL-condities bij het één- en twee-dimensionale probleem.

# Hoofdstuk 6

# Conclusies

Aan het einde van dit verslag, alle resultaten in beschouwing nemend, kunnen we tot de volgende conclusies komen.

- Omdat we graag bij de plaatsdiscretisatie behoud van massa willen hebben is het omschrijven naar de conservatieve vorm van de convectie-diffusie vergelijking een goede stap. Deze conservatieve vorm is een goede uitgangspositie om de eindige volume methode (EVM) toe te passen voor de plaatsdiscretisatie.
- Ondanks het feit dat QUICK $\mathcal{O}(h^3)$ nauwkeurigheid geeft voor de flux op een celwandje, laten de testresultaten zien dat alleen het $\kappa = \frac{1}{3}$-schema echt $\mathcal{O}(h^3)$ nauwkeurigheid geeft in de plaatsdiscretisatie.
- Het is niet mogelijk om met een ($\kappa = \frac{1}{3}$)-schema, een lineair monotoon schema te construeren. Het $\mathcal{O}(h)$ upwind schema is het enige lineaire schema dat monotoon is.
- Om hoge orde schema's te willen toepassen en behoud van monotoniciteit te willen handhaven, moeten niet-lineaire schema's worden gebruikt. Tevens dient de strenge eis van monotoniciteit vervangen te worden door de zwakere eis van *total variation diminishing* (TVD).
- Een manier om een niet-lineair hoger orde TVD-schema te formuleren is met behulp van flux-limiting. Het schema wat wij hebben gebruikt is een mix tussen het (diffusieve) $\mathcal{O}(h)$ upwind schema en een term die anti-diffusief is. In dit schema speelt de flux-limiter de rol van schakelaar; in gebieden waar de oplossing glad (monotoon) is zal de flux-limiter overschakelen naar een hoog-orde schema terwijl in gebieden waar de oplossing niet monotoon dreigt te worden, de limiter ter voorkoming van valse niet-monotoniciteiten zal overgaan op het $\mathcal{O}(h)$ upwind schema.
- Het is niet nodig voor een limiter om bij extrema te reduceren naar $\mathcal{O}(h)$ nauwkeurigheid in gladde extrema. Zijlema heeft voorwaarden afgeleid voor $\mathcal{O}(h^2)$ nauwkeurigheid in gladde extrema. Het is ons gelukt om voorwaarden voor $\mathcal{O}(h^3)$ nauwkeurigheid te vinden.
- Van de drie nieuwe flux-limiters die zijn geconstrueerd, geeft (4.8.54) de meest bevredigende resultaten. Het verwachte $\mathcal{O}(h^3)$-gedrag in gladde extrema blijkt echter uit. Reden hiervoor is mischien dat bij het testprobleem de gladde extrema op een celwandje lagen, terwijl wij voorwaarden hebben afgeleid voor een glad extremum in het centrum van een celletje.
- Bij de bepaling van een zo goed mogelijke CFL-conditie voor een RK4 schema, zijn we eerst het spectrum van onze plaatsdiscretisatie op drie manieren gaan bepalen. Het direct afleiden van het spectrum is te lastig bij onze plaatsdiscretisatie. Met behulp van RK3 schema's hebben we de drie spectra met bijbehorende CFL-condities getest. Hierbij kwam Spectrum II als beste uit de bus. Het bleek dat Spectrum III als enige eigenwaarden had op de imaginaire as. Omdat we slechts één testprobleem hebben beschouwd kan het goed mogelijk zijn dat bij een ander testprobleem er mischien wel eigenwaarden zijn op de imaginaire as. We hebben daarom zowel Spectrum II als III mee in ons verdere onderzoek meegenomen.

- Bij de bepaling van een RK4 schema waarvan het stabiliteitsgebied onze beide spectra zo goed mogelijk omsluit, zijn we uitgegaan van RK4 schema's die derde orde nauwkeurig zijn voor zowel lineaire als niet-lineaire problemen. Immers onze plaatsdiscretisatie was dit ook. We krijgen dan een vrijheidsgraad in ons stabiliteits polynoom wat we kunnen gebruiken om een stabiliteitsgebied te construeren dat ten eerste onze spectra zo goed mogelijk omsluit en dat ten tweede een zo zwak mogelijke stabiliteitsconditie oplevert. Onderzoek hiernaar leverde voor beide spectra hetzelfde stabiliteitsgebied op en nagenoeg voor beide spectra een zo optimaal mogelijke CFL-conditie.

- Aan de hand van dit stabiliteitsgebied zijn voorwaarden afgeleid waaruit we drie RK4$\beta$ schema's kunnen construeren. Eén hiervan hebben we gebruikt om de twee CFL-condities behorende bij de twee spectra te testen met behulp van ons periodieke testprobleem. Opnieuw bleek dat Spectrum II bij dit testprobleem de beste resultaten gaf.

- Met behulp van een stabiliteitsmethode in [19] hebben we voor het beschouwde RK4$\beta_1$ stabiliteitsgebied opnieuw een CFL-conditie afgeleid. Vergeleken met de testresultaten in paragraaf 5.10 gaf deze methode vergeleken met de CFL-conditie behorende bij Spectrum II een iets te conservatieve CFL-conditie.

- Tenslotte hebben we de CFL-condities uit paragraaf 5.10 getest voor een twee dimensionaal testprobleem. Hierbij gebruikten we onze beste limiter en ons RK4$\beta_1$ schema. Uit de testen bleken deze condities te zwak.

# Referenties

[1] S.K. GODUNOV, Finite Difference method for numerical computation of discontinuous solutions of the equations of fluid dynamics, *Math. Sbornik*, **47**, 271-306 (1959).

[2] E. HAIRER, S. P. NØRSETT, G. WANNER, Solving ordinary differential equations I–Nonstiff Problems, *Comp. Math.*, Vol. 8, Section II.4 (Springer-Verlag, Berlin, 1987).

[3] A. HARTEN, High resolution schemes for hyperbolic conservation laws, *J. Comput. Phys.*, **49**, 357-393, (1983).

[4] K. HEUN, Neue Methode zur approximativen Integration der Differential-gleichungen einer unabhängigen Veräderlichen, *Z. Math. Phys.*, **45**, 23-28, (1900).

[5] CH. HIRSCH, *Numerical Computation of Internal and External Flows*, Vol. 1 (Wiley, Chichester, 1988).

[6] CH. HIRSCH, *Numerical Computation of Internal and External Flows*, Vol. 2 (Wiley, Chichester, 1990).

[7] P.J. VAN DER HOUWEN, *Construction of integration formulas for initial value problems*, (Amsterdam: North Holland, 1977).

[8] W. HUNDSDORFER, B. KOREN, M. VAN LOON EN J.G. VERWER, A positive finite-difference advection scheme, *J. Comput. Phys.*, **117**, 35-46 (1995).

[9] B. KOREN, Chapter 5: A robust upwind discretization method for advection, diffusion and source terms. In B. Koren en C.B. Vreugdenhill (editors): *Notes on Numerical Fluid Mechanics*, **45**, 117-138 (Vieweg, Braunschweig, 1993).

[10] B. KOREN: Private communication (1996).

[11] W. KUTTA, Beirag zur näherungsweisen Integration totaler Differentialgleichungen, *Z. Math. Phys.*, **46**, 435-453 (1901).

[12] P.D. LAX: Hyperbolic systems of conservation laws II, *Comm. Pure Appl. Math.*, **10**, 537-566 (1957).

[13] B. VAN LEER: Upwind-difference methods for aerodynamic problems governed by the Euler equations, *Lectures in Applied Mathematics*, **22** – **Part 2**, 327-336 (American Mathematical Society, Providence, RI, 1985).

[14] B.P. LEONARD Simple high-accuracy resolution program for convective modelling of discontinuities, *Int. J. Numer. Meth. Fluids*, **8**, 1291-1318 (1988).

[15] P.K. SWEBY: High resolution schemes using flux limiters for hyperbolic conservation laws, *SIAM J. Numer. Anal.*, **21**, 995-1011 (1984).

[16] S.P. SPEKREIJSE: Multigrid solution of the steady Euler equations, Ph.D. thesis, Delft University of Technology (1987).

[17] C.B. VREUGDENHIL, Chapter 1: Introduction. In B. Koren en C.B. Vreugdenhill (editors): *Notes on Numerical Fluid Mechanics*, **45**, 1-25 (Vieweg, Braunschweig, 1993).

[18] P. WESSELING: Private communication (1996).

[19] P. WESSELING: von Neumann stability conditions for the convection-diffusion equation, *IMA J. Numer. Anal.*, **16**, 583-598 (1996).

[20] P. WILDERS EN P. WESSELING, *Dictaat Numerieke Stromingsleer A (a110A)*, (Technische Universiteit Delft, Faculteit der Technische Wiskunde en Informatica, 1989).

[21] P.R. SPALART, R.D. MOSER EN M.M. ROGERS, Spectral methods for the Navier Stokes equations with one infinite and two periodic directions, *J. Comput. Phys.*, **96**, 297-324 (1991).

[22] M. ZIJLEMA: Computational modeling of turbulent flow in general domains, Ph.D. thesis, Delft University of Technology (1996).

# Inhoud