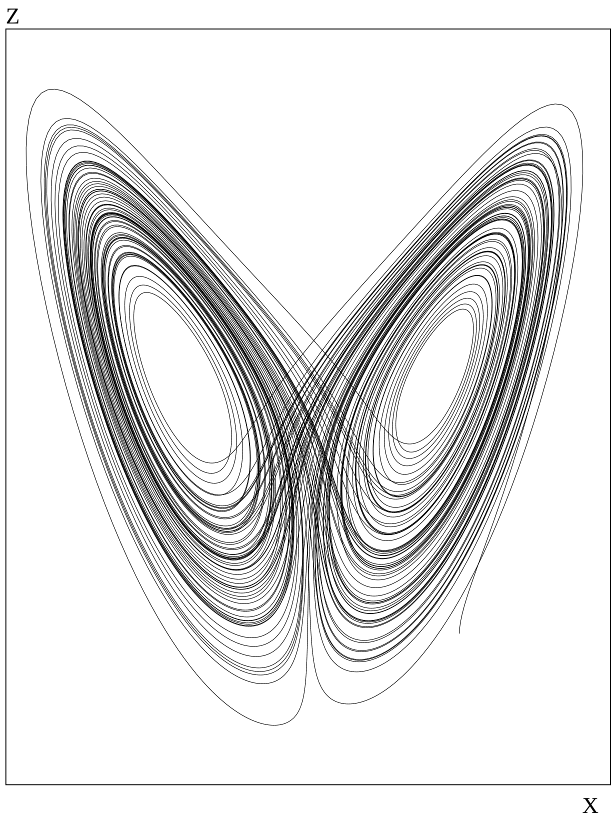


Universiteit Utrecht

Department of Mathematics

Wiskundige Modellen, deel 2

door

F. Beukers en G. Sleijpen

Najaar 1998

Inhoudsopgave

5 Differentiaalvergelijkingen van orde 1, 2

5.1 Eerste orde differentiaalvergelijkingen

In de modellen voor populatiegroei hebben we ons tot nu toe beperkt tot *discrete modellen*. Ten eerste zijn de tijdvakken waarover we populaties bekeken discreet van aard en ten tweede zijn de grootheden waarvan we de groei bekeken, de populatie aantallen, in principe geheel en dus ook discreet. Er zijn veel situaties waarin deze discreetheid er eigenlijk niet toe doet. Beschouw bijvoorbeeld een radioactief preparaat van Co^{60}. In werkelijkheid bestaat een dergelijk preparaat uit vele quadriljoenen Co^{60} atomen, die elk per tijdseenheid een zekere kans hebben om te vervallen tot Ni^{60}. In de praktijk werken we echter met gewichtshoeveelheden Co^{60}, grammen of milligrammen, en in onze ogen zijn het percentages van deze gewichtshoeveelheid die in Ni^{60} overgaan ten gevolge van radioaktief verval. De hoeveelheid Co^{60} wordt dus gemeten door de continue variabele 'gewicht' en niet door het werkelijke aantal atomen. Voor de praktijk is dit meer dan voldoende. Verder is radioaktief verval een continu doorgaand proces zonder duidelijke cyclus. Ten behoeve van de modelvorming bestaat er dus geen voorkeur voor een bepaald tijdvak, zoals bij een aantal modellen met recursies. In een dergelijk geval laten we het tijdvak naar nul gaan en gebruiken de afgeleide van het gewicht naar de tijd in ons model. We spreken over **continue modellen**.

Stel, dat ons preparaat ten tijde t een hoeveelheid $q(t)$ aan Co^{60} bevat. De waarde $q(0)$ is gegeven. De vraag is om $q(t)$ te bepalen. Het model dat we gebruiken bestaat slechts uit de volgende aanname:

Aannamen 5.1 *de afname van de hoeveelheid radioactief materiaal per tijdseenheid is evenredig met de totale hoeveelheid materiaal.*

Dit kunnen we explicieter met symbolen formuleren,

$$\frac{dq}{dt}(t) = -\alpha q(t) \tag{5.1}$$

waarin α de genoemde evenredigheidsconstante is. Deze differentiaalvergelijking lossen we als volgt op. Deel aan beide zijden door $q(t)$ en integreer van 0 tot t,

$$\int_0^t \frac{1}{q}\frac{dq}{dt}dt = -\int_0^t \alpha dt. \tag{5.2}$$

We vinden $\log q(t) - \log q(0) = -\alpha t$, en dus $q(t) = q(0)e^{-\alpha t}$. De hoeveelheid radioaktief materiaal neemt dus exponentieel af. Voor de volledigheid geven we het verband tussen α en de halveringstijd T. Per definitie (van T dus) geldt namelijk, $q(0)e^{-\alpha T} = q(T) = q(0)/2$. Gevolg, $\alpha T = \log 2$, waaruit volgt $\alpha = \log 2/T$. Onze funktie q kan dus herschreven worden als $q(t) = q(0)2^{-t/T}$.

Kalium-Argon methode. Een fraaie toepassing van radioactief verval is de ouderdomsbepaling van gesteenten door de K-Ar methode. In de natuur bestaat een klein deel ($\approx\ 0.012\%$) van het element kalium uit de radioactieve isotoop K^{40} met een zeer lange vervaltijd. Dit verval kan ofwel door uitzending van een elektron gebeuren, waarbij het stabiele Ca^{40} ontstaat, of door elektronenvangst waaruit het stabiele edelgas Ar^{40} gevormd wordt. De kansen per jaar voor een K^{40}-atoom op deze vervalmogelijkheden zijn $c = 4.72 \times 10^{-10}$ respektievelijk $a = 0.585 \times 10^{-10}$. De hoeveelheid K^{40} in een steenmonster voldoet daarmee aan

$$N'(t) = -(a + c)N(t)$$

met als oplossing $N(t) = N(0)e^{-(a+c)t}$. De totale ontstane hoeveelheid Ca^{40} en Ar^{40} is $N(0) - N(t)$. Helaas is de gevormde calcium niet meer te achterhalen omdat Ca^{40} ook op natuurlijke wijze veel in gesteenten voorkomt. De argon heeft zich sinds de stolling van het gesteente opgehoopt in holten en poriën en kan door smelting worden teruggewonnen en gemeten. De totale hoeveelheid argon is gelijk aan

$$\frac{a}{a+c}(N(0) - N(t))$$

en dus vinden we

$$Q = \frac{\text{hoeveelheid Ar}^{40}\text{ op tijdstip } t}{\text{hoeveelheid K}^{40}\text{ op tijdstip } t} = \frac{a}{a+c}\frac{1-e^{-(a+c)t}}{e^{-(a+c)t}}.$$

In een gebied in midden-Zweden werd $Q = 0.043$ gevonden. Omrekening met bovenstaande formule geeft een ouderdom van 622 miljoen jaar. Soortgelijke methoden (U-Pb, Th-Pb, Rb-Sr), indien ze kunnen worden toegepast, geven over het algemeen resultaten die redelijke overeenstemming vertonen.

Radioaktief afval. Stel dat we in een afvalvat een hoeveelheid $q(t)$ radiokatief materiaal hebben. Per tijdseenheid wordt er een constante hoeveelheid b bijgestort. Nu bestaat dq/dt uit twee bijdragen, het verval $-\alpha q$ en het gedumpte materiaal b. Dus,

$$\frac{dq}{dt} = -\alpha q + b.$$

Deze d.v. kan op dezelfde manier als boven worden opgelost,

$$\int_0^t \frac{1}{-\alpha q + b}\frac{dq}{dt} = \int_0^t dt$$

en dus $-[\log(-\alpha q(t) + b) - \log(-\alpha q(0) + b)]/\alpha = t$. Na een kleine berekening volgt,

$$q(t) = \frac{b}{\alpha} + (q(0) - \frac{b}{\alpha})e^{-\alpha t}.$$

Merk op, dat $\lim_{t\to\infty} q(t) = b/\alpha$, de hoeveelheid afval stabiliseert zich op de waarde $b/\alpha = bT/\log 2$. Dit is met name van belang voor afvalmateriaal met een korte halveringstijd. Stel nu, dat in plaats van een constante afvalstroom b er een periodiek varierende afvalstroom is door bijvoorbeeld dagelijks of wekelijks terugkerende aktiviteiten. Als voorbeeld nemen we aan dat er per tijdseenheid een hoeveelheid $b + a\sin\omega t$ gedumpt wordt. De waarde $2\pi/\omega$ is de em periode van de dumping, $\omega/2\pi$ is de *frequentie* en ω zelf wordt de *hoekfrequentie* genoemd. De d.v. wordt,

$$\frac{dq}{dt} = -\alpha q + b + a\sin\omega t.$$

De oplossingsmethode gaat nu anders. Breng de term αq naar links en vermenigvuldig met $e^{\alpha t}$,

$$e^{\alpha t}\frac{dq}{dt} + e^{\alpha t}\alpha q = (b + a\sin\omega t)e^{\alpha t}.$$

Merk nu op dat door deze truc links precies de afgeleide van $qe^{\alpha t}$ staat. Integreer aan beide zijden van 0 tot t en los $q(t)$ uit de gevonden uitdrukking op. We vinden,

$$q(t) = Ae^{-\alpha t} + \frac{b}{\alpha} + a\frac{\alpha\sin\omega t - \omega\cos\omega t}{\alpha^2 + \omega^2}$$

waarin A een constante is, gegeven door

$$A = \frac{a\omega}{\alpha^2 + \omega^2} + q(0) - \frac{b}{\alpha}.$$

Dit antwoord kan iets sprekender gemaakt worden door de hoek δ in te voeren, vastgelegd door $\cos\delta = \alpha/\sqrt{\alpha^2+\omega^2}$ en $\sin\delta = \omega/\sqrt{\alpha^2+\omega^2}$. Dan geldt, $\alpha \sin\omega t - \omega\cos\omega t = \sqrt{\alpha^2+\omega^2}\sin(\omega t - \delta)$ en dus,

$$q(t) = Ae^{-\alpha t} + \frac{b}{\alpha} + \frac{a\sin(\omega t - \delta)}{\sqrt{\alpha^2+\omega^2}}.$$

De term met $e^{-\alpha t}$ gaat exponentieel naar nul als $t \to \infty$. Als t dus groot is ten opzichte van de halveringstijd T, dan draagt deze term vrijwel niets bij en kan verwaarloosd worden. Dit is met name van belang voor afvalmateriaal met een korte halveringstijd. Verder herkennen we weer de constante term b/α en tenslotte heeft de periodieke dumping een respons in $q(t)$ die een *fase* δ achterloopt bij de dumping zelf.

De differentiaalvergelijkingen (afgekort met d.v.) die we gezien hebben zijn van de **eerste orde**, hooguit de eerste afgeleide van de onbekende funktie komt in de vergelijking voor. De algemene gedaante van een d.v. van eerste orde luidt,

$$\frac{dy}{dt} = f(t, y(t)) \tag{5.3}$$

waarin $y(t)$ de te bepalen funktie is. Als f niet expliciet van t afhangt noemen we de vergelijking **autonoom**, in andere gevallen **niet-autonoom**. Onze eerste twee voorbeelden zijn autonoom, het derde is niet-autonoom. In de praktijk zijn de meeste d.v.'s niet zo eenvoudig op te lossen als hierboven. Bijvoorbeeld, $y' = -y^2 + ty + 2$ is een d.v. waarvan de oplossingen totaal nieuwe funkties zijn, die geen enkel verband hebben met de ons bekende 'klassieke' funkties zoals $\log$, $\cos$ of $\sin$. In dergelijke gevallen moeten we onze toevlucht nemen tot **numerieke berekeningen** op de computer of **kwalitatieve methoden** die alleen maar een globale indruk van het gedrag van de oplossingen geven. Beide methoden zullen we in latere paragrafen behandelen. Soms hebben we geluk en kunnen we de d.v. expliciet oplossen, d.w.z. de oplossingen uitdrukken in 'klassieke' funkties en hun integralen. In deze paragraaf behandelen we twee van dergelijke typen.

De eerste methode die men kan proberen is die van de **scheiding van variabelen**. Stel dat de d.v. de speciale gedaante

$$\frac{dy}{dt} = \frac{f(t)}{g(y)} \tag{5.4}$$

heeft. Stel dat F een primitieve van f is en G een primitieve van g. Vermenigvuldig aan beide zijden met dt en $g(y)$. We krijgen $g(y)dy = f(t)dt$.Wiskundig gezien is dit onzin (althans met de wiskunde die we hier bekend veronderstellen), maar als formele tussenstap erg nuttig. Neem aan beide zijden een primitieve en we vinden $G(y) = F(t) + C$,waarin C een willekeurige constante is. Zoals gezegd is bovenstaande oplossingsmethode zeer slordig, maar het antwoord is wel juist (vul maar in!) en dus bruikbaar. Hoe het op de nette manier zou moeten gaan, hebben we in de eerste twee voorbeelden over radioaktief verval gezien. Bij deze voorbeelden was de constante C gelijk aan $q(0)$. Met andere woorden, C kan worden vastgelegd door nadere eisen aan de oplossing $y(t)$ op te leggen, zoals een gegeven waarde in 0. Dergelijke eisen noemen we **randvoorwaarden** of **beginvoorwaarden**.

Voorbeeld 5.1 Bepaal de funktie $y(t)$ die voldoet aan

$$\begin{cases} y' = t(1-y^2), \\ \text{en } y(0) = 0. \end{cases}$$

Scheiding van variabelen levert $dy/(1-y^2) = tdt$ en na integratie, $\log|\frac{1+y}{1-y}| = t^2 + C$. Oplossing van y geeft

$$y(t) = \frac{C_1 e^{t^2} - 1}{C_1 e^{t^2} + 1}$$

waarin $C_1 = e^C$. De beginvoorwaarde $y(0) = 0$ impliceert dat $C_1 = 1$.

Een tweede klasse van oplosbare vergelijkingen is die van de **lineaire differentiaalvergelijkingen**, d.w.z. vergelijkingen van de vorm

$$\frac{dy}{dt} + p(t)y = r(t). \tag{5.5}$$

Lineariteit slaat op het feit dat zowel y als y' als term voorkomen met een macht die hooguit 1 is. Merk op dat al onze voorbeelden over het radioaktief verval, met name het derde, van dit type zijn. Als $r(t) \equiv 0$ dan noemen we de vergelijking **homogeen**, anders **inhomogeen**. De oplossingsmethode bestaat uit vermenigvuldiging met een slim gekozen **integrerende faktor** namelijk $\exp(\int pdt)$,

$$e^{\int pdt}\frac{dy}{dt} + p(t)e^{\int pdt}y = r(t)e^{\int pdt}.$$

Links staat precies de afgeleide van $\exp(\int pdt)y$. Primitiveer aan beide zijden en we vinden als algemene oplossing

$$y(t) = Ce^{-\int pdt} + e^{-\int pdt}\int e^{\int pdt}r(t)dt$$

waarin C weer een willekeurige integratieconstante is die door beginvoorwaarden kan worden vastgelegd.

Voorbeeld 5.2 Bepaal de algemene oplossing van $(t^2+1)y' + ty = t^3$.Breng eerst deze vergelijking in de standaardvorm $y' + ty/(t^2+1) = t^3/(t^2+1)$. De integrerende factor is $\exp(\int tdt/(t^2+1)) = \sqrt{t^2+1}$. Vermenigvuldiging hiermee en integratie levert $y\sqrt{t^2+1} = \int t^3 dt/(t^2+1) = C + \frac{1}{3}(t^2-2)\sqrt{t^2+1}$. Dus,

$$y(t) = \frac{C}{\sqrt{t^2+1}} + \frac{1}{3}(t^2-2).$$

5.2 Oefeningen

Oefening 5.1 Bepaal de algemene oplossingen van

a. $\dfrac{dy}{dx} = \dfrac{1+2y}{4-x}$

b. $\dfrac{dy}{dx} = -\dfrac{\cos x}{\sin y}$

c. $\dfrac{dy}{dx} = -(1+y^2)\cdot\ln x$

d. $\dfrac{dy}{dx} = \dfrac{y}{x^2 - 3x + 2}$

e. $\dfrac{dy}{dx} = \dfrac{yx}{x^3 - 3x + 2}$

Oefening 5.2 Bepaal de algemene oplossing van

a. $\dot{x} = x + e^t$

b. $\dot{x} = \sqrt{t}\, x + t^2$

c. $\dot{x} = 2x + e^{-t}$

d. $\dot{x} = x + \cos t$

e. $\dot{x} = tx + 2t$

f. $\dot{x} = \dfrac{1}{1+t}\, x + \sin t$

Oefening 5.3 *De levensduur van een gloeidraad*. We beschouwen een gloeidraad in een lamp als een dunne rechte cylinder-vormige draad. Door de draad loopt een elektrische stroom waardoor de draad een constante hoge temperatuur heeft. We nemen verder nog aan dat de lengte ℓ van de draad constant is. Ten gevolge van de hoge temperatuur verdampt er voortdurend materiaal van de draad. Per tijdseenheid verdampt er een hoeveelheid materiaal evenredig met het oppervlak van de draad. Op een tijdstip $t = 0$ is de dikte van de draad d_0.

a. Geef een differentiaalvergelijking voor de dikte d van de draad.

b. Los deze differentiaalvergelijking op.

c. Bepaal nu de levensduur van de draad als gegeven is dat op $t = 0$ de dikte 0.1mm is en op $t = 100$ de dikte 0.09mm is.

Oefening 5.4 Een bekende methode om archeologische vondsten te dateren is de C^{14}-methode ontdekt door Willard Libby rond 1949. Het idee is verbluffend eenvoudig maar zeer effektief. De aardatmosfeer vangt continu kosmische straling op die neutronen produceert en die op hun beurt weer met stikstof combineren tot de radioaktieve koolstofisotoop C^{14}. De halfwaardetijd van deze isotoop is 5568 jaar. Hierdoor bestaat een klein percentage van de kooldioxide in de atmosfeer uit radioactieve CO_2. Een belangrijke aanname is dat dit percentage constant is over perioden van duizenden jaren. Zolang een organisme leeft zal, door voortdurende koolstofuitwisseling met de omgeving, zijn C^{12}/C^{14} verhouding gelijk zijn aan die van de atmosfeer. Nadat een organisme gestorven is stopt de uitwisseling en vervalt de aanwezige radioaktieve koolstof. Hierdoor zal in bijvoorbeeld hout, dat enkele duizenden jaren oud is, veel minder C^{14} voorkomen dan in nog levend hout.

a. Houtskool gevonden in de woonruimten van de beroemde grotten van Lascaux, Frankrijk, vertoonde 0.97 C^{14} desintegraties per minuut per gram. Levend hout vertoont 6.68 desintegraties. Bereken hoelang geleden de Lascaux grotten bewoond werden.

b. Bij opgravingen in Nippur, Babylonië, gaf houtskool afkomstig uit een dakbalk 4.09 desintegraties per minuut per gram te zien. Levend hout vertoont 6.68 desintegraties. Aannemende dat de houtskool werd gevormd in Hammurabi's tijd, vind een schatting voor de tijd van Hammurabi's opvolging.

Er moet aan toegevoegd worden dat er in werkelijkheid nog fluktuaties optreden van de C^{14}-fraktie in de atmosfeer. De C^{14}-methode moet daarom voortdurend geijkt worden tegen andere tijdsschalen. Vooral na de atoomproeven in de vijftiger jaren is de C^{14}-fraktie toegenomen.

Oefening 5.5 Een mottebal is een bolletje kamfer dat snel verdampt. Door dit verdampen nemen omvang en gewicht van het bolletje af. Het gewichtsverlies is evenredig met de oppervlakte van het bolletje. Het gewicht G (in gram) van het bolletje is een functie van de tijd t (in weken).

a. Laat zien dat er een positieve constante c bestaat zodat geldt:

$$\frac{dG}{dt} = -cG^{2/3}.$$

b. Stel dat het bolletje aanvankelijk 10 gram woog en na tien weken nog maar 1 gram. Na hoeveel weken is het bolletje geheel verdampt?

Oefening 5.6 We beschouwen een meer met constante inhoud V in km^3 waarin zich een hoeveelheid chemisch afval bevindt. Het afval is voortdurend gelijk verdeeld over het hele meer. Op tijdstip t is de concentratie $c(t)$. Een vervuilde rivier komt uit in het meer, de concentratie van chemisch afval in de rivier is constant k en per tijdseenheid komt een hoeveelheid s rivierwater in het meer, terwijl eenzelfde hoeveelheid water per tijdseenheid het meer uitstroomt.

a. Geef een differentiaalvergelijking voor $c(t)$ en los deze op.
Bereken $\lim_{t\to\infty} c(t)$.

b. Laat er nu ook een gewichtshoeveelheid chemisch afval g per tijdseenheid illegaal in het meer geloosd worden. Los de nieuwe differentiaalvergelijking op en bepaal $\lim_{t\to\infty} c(t)$ indien deze bestaat.

c. Neem als tijdseenheid 1 jaar. Stel de vervuiling van de rivier wordt beëindigd en de illegale lozingen worden gestaakt op het tijdstip t_1 zodanig dat $c(t_1) = 100c(0)$, uiteraard $c(0) < \frac{1}{100}$. We noemen nu $\frac{s}{v} = \frac{1}{10}$ snelle doorstroming en $\frac{s}{v} = \frac{1}{100}$ langzame doorstroming. Bepaal voor beide gevallen hoe lang het duurt voordat de concentratie $c(t)$ weer de oorspronkelijke waarde $c(0)$ heeft.

Oefening 5.7 [Tentamensom september 1993]

In een meer met inhoud V m^3 bevindt zich een hoeveelheid chemisch afval. Het afval is voortdurend gelijk verdeeld over het hele meer. Op tijdstip t is de concentratie $c(t)$. Een vervuilde rivier komt uit in het meer. De instroomsnelheid van het vervuild rivier water is konstant σ m^3 per etmaal. Een zelfde hoeveelheid water stroomt uit het meer, ook met konstante snelheid. De concentratie van chemisch afval in de rivier bij het instroompunt varieert in de loop van de dag volgens $\kappa(\cos(\pi t) + 1)$.

a. Laat zien dat de differentiaalvergelijking

$$c'(t) = \frac{\sigma\kappa}{V}[\cos(\pi t) + 1] - \frac{\sigma}{V}c(t)$$

een goed mathematisch model is voor de beschreven situatie.

b. Neem voor $\sigma = 10^{-2}V$ en los de differentiaalvergelijking op: geef de oplossing in termen van t, $c(0)$ en κ.

c. Beschrijf wat op den lange duur (t zeer groot) de concentratie van het afval in het meer zal zijn (varieert de concentratie nog in de loop van de dag; zo ja, hoe? Wat is de gemiddelde concentratie per etmaal?)

Oefening 5.8 De aanwezigheid van een bepaald gif doodt een bacteriesoort met een snelheid die evenredig is met de hoeveelheid aanwezige bacteriën en de hoeveelheid gif. Noem de evenredigheidsfactor a. Als er geen gif zou zijn zou de bacteriepopulatie groeien met een snelheid evenredig aan het aantal aanwezige bacteriën. Noem deze evenredigheidsfaktor b. Stel dat de hoeveelheid gif T in constant tempo toeneemt, d.w.z. $dT/dt = c$ en dat de gifproduktie begint op tijdstip $t = 0$. Zij $y(t)$ het aantal levende bacteriën op tijdstip t. Stel een differentiaalvergelijking op voor $y(t)$ en onderzoek wat er gebeurt als $t \to \infty$.

Oefening 5.9 Een regendruppel valt in een wolk onder invloed van de zwaartekracht naar beneden. Door condensatie groeit de waterdruppel: de massatoename per tijdseenheid is evenredig met de oppervlakte. Neem aan dat de druppel steeds bolvormig is. Stel differentiaalvergelijkingen op voor de massa en voor de snelheid van de regendruppel. (De wet van Newton luidt: $\frac{d}{dt}(mv) = mg$.) Los deze vergelijkingen op.

In oefening 5.10 kan gebruik worden gemaakt van de volgende definities en eigenschappen. De definities en eigenschappen komen voor 2-dimensionale modellen uitgebreidt aan de orde in §6.4.

Voor een zekere kontinue funktie f, en zekere skalair q_0, zijn we geïnteresseerd in de oplossing q van het volgende probleem

$$q'(t) = f(q(t)) \text{ voor } t \geq t_0 \quad \text{met } q(0) = q_0. \tag{5.6}$$

De oplossing is in **evenwicht** als $q(t) = q_0$ voor iedere t, of equivalent hiermee, als $q'(t) = 0$ voor iedere t.

Er geldt:

De oplossing $q_0 = \alpha$ is evenwicht precies dan als $f(\alpha) = 0$;

Een evenwichts oplossing $q(t) = q_0 = \alpha$ is stabiel als $f'(\alpha) < 0$ en instabiel als $f'(\alpha) > 0$.

Oefening 5.10 [Tentamensom maart 1996]

Zie inleidende tekst hierboven.

Voor de groei van de hoeveelheid gras in een wei die door schapen begraasd wordt gebruiken we het volgende kontinue model:

$$G' = \kappa\, G \left(1 - \frac{G}{G_0}\right) - \gamma\, S \frac{G^2}{G_1^2 + G^2} \tag{5.7}$$

waarbij $G(t)$ de totale hoeveelheid gras is in de wei op tijdstip t. Verder is S het aantal schapen in de wei. De overige grootheden zijn konstanten: γ en κ zijn evenredigheids konstanten en G_0 representeerd de “maximale” hoeveelheid gras.

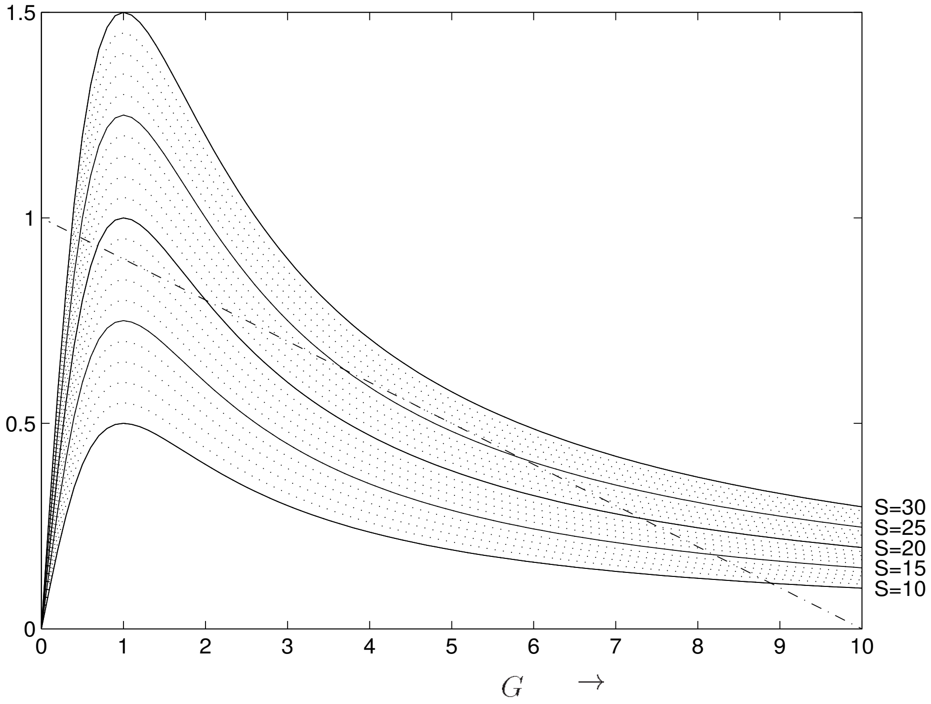
a. Interpreteer de term $\gamma\, S\, G^2/(G_1^2 + G^2)$. Wat voor invloed heeft de konstant G_1?

b. Hoe heet de groei in dit model als er geen schapen in de wei zijn ($S = 0$) (dezelfde naam als in geval van een diskreet model)?

Neem verder $G_0 = 10$, $G_1 = 1$ en $\kappa = 1$, $\gamma = 0.1$.

In FIG. 1 staan de grafieken van $1 - G/10$ (-.-.), $0.1\, S\, G/(1 + G^2)$ voor alle S tussen 10 en 30. Voor $S = 10, 15, 20, 25, 30$ is de grafiek aangegeven met een doorlopende lijn (—), voor de overige S met stippels (· · ·). Bij het beantwoorden van de volgende vragen kan je gebruik maken van de informatie die je uit de grafieken kan halen. Het snijpunt van $1 - G/10$ en $\frac{1}{10} S\, G/(1 + G^2)$ tussen 0 en 1 heet α_S (als het bestaat), dat tussen 1 en ± 5 heet β_S (ook weer alleen als het bestaat) en tussen ± 5 en 10 heet γ_S (als het bestaat). Je hoeft de numerieke waarde van deze snijpunten hier niet expliciet te berekenen.

c. Beschrijf de evenwichten (in termen van bijvoorbeeld α_S, β_S en γ_S; zie plaatje) en hun stabiliteit voor $S = 10$, $S = 20$ en $S = 30$.



FIGUUR 1. *Grafieken ten behoeve van de stabiliteitsanalyse in Oefening* 5.10.

(Hint: maak een plaatje langt de "G-as" waarin je de evenwichten aangeeft en geef middels pijltjes aan hoe de oplossing in andere punten op die G-as zal verlopen).

d. De boer wacht telkens met veranderingen tot de hoeveelheid gras in de wei min of meer in evenwicht is.
Wat gebeurt er met zijn grasopbrengst als hij telkens een schaap meer in zijn wei laat grazen en zo zijn kudde geleidelijk uitbreidt van 10 naar 30 schapen? Wat is het effekt van een daarop volgende geleidelijke reduktie van de kudde?

5.3 Lineaire d.v.'s met constante coëfficienten

Onder een d.v. van **orde** k verstaan we een d.v. van de vorm

$$\frac{d^k y}{dt^k} = F\left(t, y, \frac{dy}{dt}, \ldots, \frac{d^{k-1}y}{dt^{k-1}}\right). \tag{5.8}$$

Onder een *lineaire d.v.* van orde k verstaan we een vergelijking van de vorm

$$\frac{d^k y}{dt^k} + p_1(t)\frac{d^{k-1}y}{dt^{k-1}} + \cdots p_{k-1}(t)\frac{dy}{dt} + p_k(t)y = r(t). \tag{5.9}$$

Als $r(t) \equiv 0$ noemen we de vergelijking **homogeen** anders **inhomogeen**. Voor onze doeleinden is zelfs de klasse van lineaire d.v.'s te uitgebreid en we beperken ons tot tweede orde lineaire d.v.'s waarvan de coëfficienten p_1, p_2 constante funkties zijn. Dus,

$$y'' + py' + qy = r(t) \tag{5.10}$$

waarin $p, q \in \mathbb{R}$. Eerst bekijken we het homogene geval $r(t) \equiv 0$.

Stelling 5.2 *Beschouw de d.v.* (5.10) *en stel* $r(t) \equiv 0$. *Dan vormt de oplossingsverzameling een lineaire vektorruimte over* $\mathbb{R}$ *van dimensie* 2. *Met andere woorden, er bestaan twee oplossingen* $y_1(t)$ *en* $y_2(t)$ *zó dat elke oplossing op unieke wijze in de vorm* $Ay_1(t)+By_2(t),\ A, B \in \mathbb{R}$ *geschreven kan worden.*

Uit de lineaire algebra weten we ook dat we voor de basis y_1, y_2 ieder tweetal lineair onafhankelijke oplossingen kunnen nemen. Anders gezegd, ieder tweetal niet-triviale oplossingen $y_1(t)$, $y_2(t)$ zó dat y_1/y_2 niet-constant is voldoet als basis.

Opmerking. Bovengenoemde stelling gaat ook op voor willekeurige homogene lineaire d.v.'s van orde k. De oplossingsverzameling is dan een vektorruimte van dimensie k. Omdat een precieze formulering nog wat extra notatie vergt en we er verder niet naar kijken, laten we deze algemenere stelling achterwege.

We geven voor d.v. (5.10) met $r(t) \equiv 0$ een basis y_1, y_2.
Stel eerst de **karakteristieke vergelijking** op, $\lambda^2 + p\lambda + q = 0$. De nulpunten geven we aan met λ_1, λ_2. We onderscheiden drie gevallen.

i) λ_1, λ_2 reëel en verschillend. We kunnen dan $y_1 = e^{\lambda_1 t}$ en $y_2 = e^{\lambda_2 t}$ nemen.

ii) $\lambda_1 = \lambda_2$. We kunnen $y_1 = e^{\lambda_1 t}$ en $y_2 = te^{\lambda_1 t}$ nemen.

iii) λ_1, λ_2 verschillend en complex. Dan geldt $\lambda_{1,2} = \rho \pm \sigma i$. We kunnen nu $y_1 = e^{\rho t}\cos(\sigma t)$ en $y_2 = e^{\rho t}\sin(\sigma t)$ nemen.

We beschouwen nu de inhomogene vergelijking

$$y'' + py' + qy = r(t). \tag{5.11}$$

Oplossing hiervan gaat als volgt. Bepaal op de één of andere manier een oplossing die we de **particuliere** oplossing $y_{\text{part}}(t)$ noemen. (Vergelijk het Engelse woord *particular*.) Er geldt dus $y''_{\text{part}} + py'_{\text{part}} + qy_{\text{part}} = r(t)$. Trek dit af van onze vergelijking. We vinden,

$$(y - y_{\text{part}})'' + p(y - y_{\text{part}})' + q(y - y_{\text{part}}) = 0.$$

Het verschil $y - y_{\text{part}}$ voldoet dus aan de homogeen gemaakte d.v. $y'' + py' + qy = 0$, waarvoor we reeds de algemene oplossing hebben. Noem deze oplossing $y_{\text{hom}}(t)$. De algemene oplossing van onze inhomogene d.v. luidt dus $y_{\text{hom}}(t) + y_{\text{part}}(t)$.

Voorbeeld 5.3 We lossen de vergelijking $y'' + 3y' + 2y = r(t)$ op voor diverse $r(t)$. De algemene oplossing van de homogeen gemaakte vergelijking is eenvoudig. De karakteristieke vergelijking luidt $\lambda^2 + 3\lambda + 2 = 0$ en dus $y_{\text{hom}} = Ae^{-t} + Be^{-2t}$.

- $r(t) = t$. Probeer als particuliere oplossing $y = at + b$ voor nader te bepalen a, b. Dan geldt

$$y'' + 3y' + 2y = 2at + 2b + 3a.$$

Opdat dit gelijk is aan t moet gelden $2a = 1$ en $2b + 3a = 0$ en dus $a = 1/2$, $b = -3/4$.

- $r(t) = e^t$. Probeer als particuliere oplossing $y = ae^t$. Dan geldt

$$y'' + 3y' + 3y = 6ae^t$$

en dus moeten we $a = 1/6$ nemen.

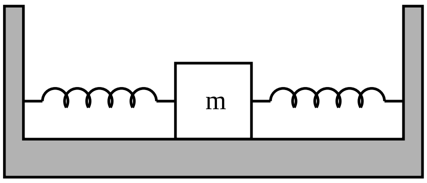
- $r(t) = \cos t$. Probeer als particuliere oplossing $a\cos t + b\sin t$. Dan geldt

$$y'' + 3y' + 2y = (-a + 3b + 2a)\cos t + (-b - 3a + 2b)\sin t.$$

Dus $a + 3b = 1$ en $b - 3a = 0$, waaruit volgt $a = 1/10$, $b = 3/10$.

5.4 Harmonische oscillator, resonantie

We beschouwen het mechanische systeem bestaande uit een massa m die in één richting over een grondvlak kan bewegen en aan weerszijden door een veer aan een muur is bevestigd



De oorsprong van de x-as, $x = 0$, zetten we op de evenwichtstoestand van het systeem. Vraag is nu hoe de positie $x(t)$ van de massa zich als funktie van de tijd t gedraagt.

Stel eerst dat de massa zich wrijvingsloos over het grondvlak beweegt. De kracht die m ondervindt is gelijk aan $F_{\text{veer}} = -kx$, waarin $k > 0$ de zogenaamde *veerconstante* is. Voor niet al te grote afwijkingen is dit een redelijke aanname die bekend staat als de *wet van Hooke* (Robert Hooke, Engels fysicus en uitvinder van het draagbare uurwerk, 1635-1702). De massa m beweegt zich onder invloed van deze kracht volgens de *tweede wet van Newton*, $F = mx''$. We vinden dus de d.v.

$$m\frac{d^2x}{dt^2} = -kx.$$

De algemene oplossing luidt $a\cos\sqrt{k/m}t + b\sin\sqrt{k/m}t$, met a en b willekeurige integratieconstanten. We hebben dus een periodieke beweging met periode $2\pi\sqrt{m/k}$. Kies δ zó dat $\sin\delta = a/\sqrt{a^2+b^2}$, $\cos\delta = -b/\sqrt{a^2+b^2}$ en stel $A = \sqrt{a^2+b^2}$. Onze oplossing kunnen we herschrijven als $A\sin(\omega t - \delta)$, $\omega = \sqrt{k/m}$. Dit is de algemene gedaante van een *harmonische trilling* met *amplitude* A, *hoekfrequentie* ω en *faseverschuiving* δ. Het mechanische systeem zelf noemen we een *harmonische oscillator*.

We nemen nu aan dat er wel wrijving optreedt bij de beweging over het grondvlak. Als modellering van de wrijvingskracht kiezen we $F_{\text{wrijving}} = -cx'$ waarin $c > 0$ de *wrijvingscoëficient* is. We merken hier wel nadrukkelijk op dat er vele andere mogelijkheden zijn voor de wrijvingsterm, afhankelijk van de situatie. Zo is er bijvoorbeeld de *statische wrijving*, de kracht die overwonnen moet worden om een voorwerp van zijn plaats te trekken. In onze situatie nemen we aan dat de statische wrijving verwaarloosbaar is en dat we alleen *dynamische wrijving* hebben. Onder de dynamische wrijvingskrachten zijn er ook diverse mogelijkheden. De luchtremming die een parachutespringer ondervindt, of de spaceshuttle bij terugkeer in de dampkring, is evenredig met een macht van x' die ongeveer 2 is. Zou dit niet het geval zijn, dan was parachutespringen een zeer hachelijke onderneming en zouden er bij terugkeer van de spaceshuttle geen problemen met hitteontwikkeling zijn. Hoe interessant ook, we gaan hier verder niet op in en keren terug naar onze situatie. De d.v. luidt

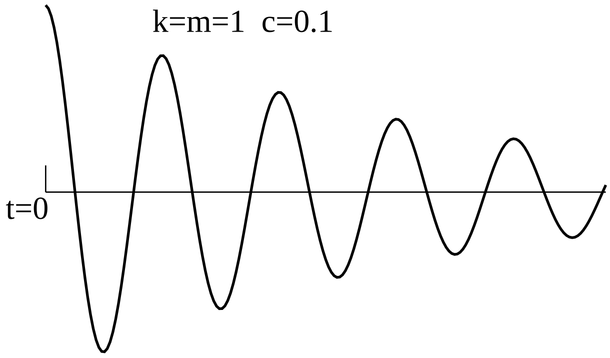
$$m\frac{d^2x}{dt^2} = -c\frac{dx}{dt} - kx.$$

Dit is een lineaire, homogene d.v. met constante coëfficienten. De karakteristieke vergelijking luidt $m\lambda^2 + c\lambda + k =$ met de oplossingen $\lambda_{1,2} = -(c \pm \sqrt{c^2-4km})/2m$. We onderscheiden drie gevallen naar toenemende c.

a) $c^2 < 4km$. Dan zijn $\lambda_{1,2}$ complex en de algemene oplossing luidt

$$x(t) = e^{-ct/2m}(a\cos\omega t + b\sin\omega t), \qquad \omega = \frac{1}{2m}\sqrt{c^2-4km}.$$

De grafiek ziet er globaal uit als

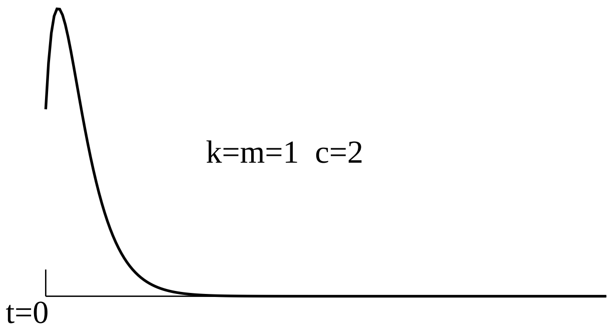


en we noemen dit geval de *gedempte trilling*.

b) $c^2 = 4km$. Dan geldt $\lambda_1 = \lambda_2$ en de algemene oplossing luidt

$$x(t) = (at + b)e^{-ct/2m}.$$

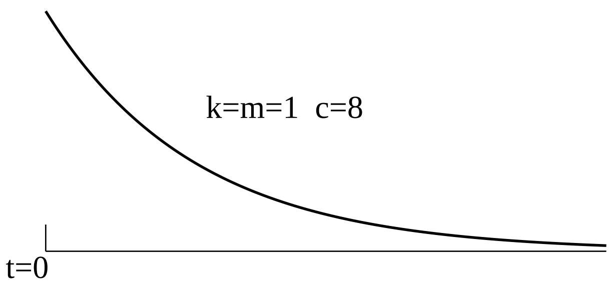
Dit geval staat bekend onder de naam *kritische demping*. Een typische grafiek,



c) $c^2 > 4km$. Dan zijn $\lambda_{1,2}$ reëel en negatief. De algemene oplossing luidt

$$x(t) = ae^{\lambda_1 t} + be^{\lambda_2 t}.$$

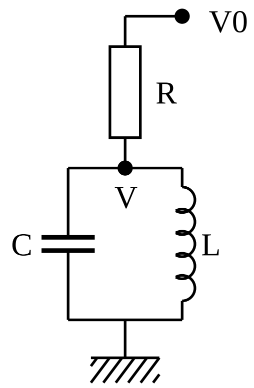
Dit is het geval van *overdemping*. Een grafiek,



In alle drie gevallen geldt voor iedere oplossing dat $\lim_{t\to\infty} x(t) = 0$. Stel dat we k, m vasthouden en c variëren. Dan blijkt dat de convergentie naar nul van de algemene oplossing het snelst is bij $c^2 = 4km$ (Ga na!). Dit verklaart de benaming kritische demping.

Het verschijnsel (gedempte) harmonische trilling is een universeel begrip in de natuurwetenschap en komt in vele gedaanten voor. Dat deze gedaante niet mechanisch hoeft te zijn blijkt uit het volgende voorbeeld.

Beschouw een elektrische schakeling met een weerstand R, een condensator C en een spoel L,



De spanning V_0 is extern en kan al of niet ingeschakeld zijn. De spanning $V(t)$ voldoet aan de volgende d.v.

$$\frac{d^2V}{dt^2} + \frac{1}{RC}\frac{dV}{dt} + \frac{1}{LC}V = \frac{1}{RC}\frac{dV_0}{dt}.$$

Als V_0 constant wordt gehouden hebben we weer te maken met een gedempte harmonische oscillator waarvan de demping door $1/RC$ bepaald wordt. Het gedrag van dergelijke schakelingen wordt interessant als we V_0 periodiek laten variëren.

Laten we daarom eens kijken naar een harmonische oscillator die door een uitwendige kracht wordt aangedreven. In het bijzonder speelt het begrip *resonantie* hierbij een grote rol. Voor de modellering van het begrip resonantie voeren we aan onze mechanische harmonische oscillator een periodieke uitwendige kracht van de vorm $F_{\text{uit}} = f \sin \nu t$ toe. De totale kracht die op m werkt is nu $F_{\text{uit}} + F_{\text{veer}} + F_{\text{wrijving}} = -cx' - kx - f \sin \nu t$. De d.v. wordt,

$$m\frac{d^2x}{dt^2} + c\frac{dx}{dt} + kx = f \sin \nu t.$$

Dit is een inhomogene lineaire d.v. en we moeten dus een particuliere oplossing zien te vinden. Probeer hiervoor $x_{\text{part}} = \alpha \cos \nu t + \beta \sin \nu t$ met nader te bepalen α, β. Enig rekenwerk levert

$$\alpha = \frac{f}{m}\frac{-c\nu}{c^2\nu^2 + (k - m\nu^2)^2} \qquad \beta = \frac{f}{m}\frac{k - m\nu^2}{c^2\nu^2 + (k - m\nu^2)^2}.$$

Kies nu δ zó dat

$$\sin \delta = \frac{c\nu}{\sqrt{c^2\nu^2 + (k - m\nu^2)^2}} \qquad \cos \delta = \frac{k - m\nu^2}{\sqrt{c^2\nu^2 + (k - m\nu^2)^2}}.$$

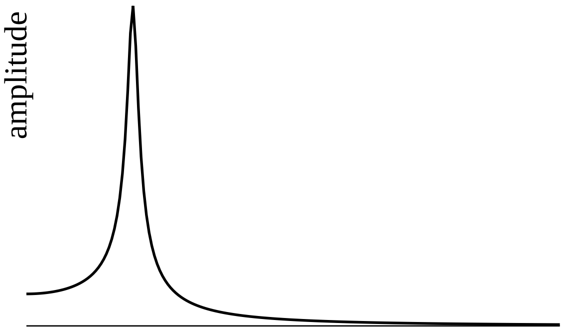
De algemene oplossing van de inhomogene d.v. wordt

$$x(t) = x_{\text{homogeen}}(t) + \frac{f}{m}\frac{\sin(\nu t - \delta)}{\sqrt{c^2\nu^2 + (k - m\nu^2)^2}}$$

waarin x_{homogeen} de algemene oplossing van de homogeen gemaakte d.v. is. We hebben al gezien dat, zodra $c > 0$, de term x_{homogeen} op den duur naar 0 toe gaat. We noemen dit de *uitdemping* van de eigentrillingen van de oscillator. Op den duur krijgen we alleen met de tweede term van onze oplossing te maken. Deze laatste term, de respons van de oscillator op $F_{\rm uit}$, beschrijft een harmonische trilling die een fase δ bij $F_{\rm uit}$ achterloopt en amplitude

$$\frac{f}{m}\frac{1}{\sqrt{c^2\nu^2+(k-m\nu^2)^2}}$$

heeft. Zetten we deze amplitude uit tegen ν, de aandrijffrequentie, dan krijgen we een grafiek die grofweg de volgende gedaante heeft,



Deze funktie heeft een maximum bij $\nu_0 = (2m)^{-1}\sqrt{4mk-2c^2}$. De eigenfrequentie van de oscillator is $\nu_{\rm eig} = (2m)^{-1}\sqrt{4mk-c^2}$. Neem aan dat c klein is. Verwaarlozen we termen van orde c^2 dan zien we dat de piek optreedt bij $\nu_0 \approx \sqrt{k/m} \approx \nu_{\rm eig}$. De respons van de oscillator is dus het hevigst als de aandrijffrequentie in de buurt van de eigenfrequentie van de oscillator ligt. Dit verschijnsel staat bekend als resonantie. De piek heeft bij benadering de hoogte $f/mc\nu$. Als $c \downarrow 0$ dan zien we dat deze piek zelfs willekeurig hoog gemaakt kan worden. Een interessant detail is dat bij resonantie, d.w.z. $\nu^2 \approx k/m$, de faseverschuiving δ ongeveer $\pi/2$ bedraagt, de respons loopt vrijwel $\pi/2$ achter bij de aandrijving.

Het is interessant te zien dat er situaties zijn waarin resonanties vermeden dienen te worden, en situaties waarin er juist gebruik van wordt gemaakt. Van de eerste soort is de ramp met de *Tacoma-brug*, een hangbrug, de bekendste. Op 1 juli 1940 werd de Tacoma Narrows Bridge by Puget Sound in de staat Washington voor het verkeer geopend. De daarop volgende dagen begon de brug een merkwaardig gedrag te vertonen door in verticale oscillatie te geraken. Deze beweging bleef enkele maanden aanhouden en de brug werd al gauw een toeristische attractie. Tot 7 november 1940. In de ochtend, 7 uur, begon de brug te golven met uitwijkingen die tot een meter op konden lopen. Om ongeveer 10 uur knapte er iets en de oscillaties werden heviger. Op het ene moment was één eind van de weg acht meter hoger dan het andere, het volgende moment acht meter lager. Om 10u30 traden de eerste scheuringen op en om 11u10 stortte de brug in. De verklaring die uiteindelijk werd gevonden is de volgende. Als een object in een luchtstroom wordt geplaatst treden er onder zekere condities wervelstromen achter het object op. Deze wervels raken vervolgens los van het object en worden met de luchtstroom meegevoerd. Als de frequentie waarmee deze wervels loslaten dicht bij de eigenfrequentie van de brug ligt, en de brug ook nog slecht gestroomlijnd is, treedt er resonantie op die in het geval van de Tacoma-brug fataal was.

Een voorbeeld waarbij juist gebruikt wordt gemaakt van resonantie is het elektrische circuit dat we hierboven zagen. Stel dat $V_0(t) = V_0 \sin \nu t$. Een eenvoudige berekening leert dat, na uitdemping van de eigentrillingen, geldt

$$V(t) = \frac{V_0}{\sqrt{1+(RC)^2(1-\frac{1}{LC\nu^2})^2}}\sin(\nu t-\delta)$$

voor zekere δ. Bij geschikte keuze van R, C, L vertoont de amplitude van deze trilling een scherpe piek ter hoogte V_0 bij $\nu = 1/\sqrt{LC}$. Blijkbaar worden dus trillingen met hoekfrequentie $1/\sqrt{LC}$ door deze schakeling uitgefilterd, de rest wordt als het ware kortgesloten. Van dit principe wordt gebruik gemaakt bij afstemcircuits voor radios, waarin we een bepaalde frequentie uit het gehele radiospektrum willen selecteren.

5.5 Oefeningen

Oefening 5.11 Bepaal de algemene oplossing van

a. $\ddot{x} - 3\dot{x} - 4x = 0$

b. $\ddot{x} - 2\dot{x} + x = 0$

c. $\ddot{x} + 4\dot{x} + 5x = 0$

d. $\ddot{x} + x = \sin 2t$

e. $\ddot{x} + 3\dot{x} + 2x = t$

f. $\ddot{x} + 3\dot{x} + 2x = e^{\alpha t}$ 1) $\alpha = -1$ 2) $\alpha = 1$

Oefening 5.12 Een bekende oliemaatschappij heeft het volgende probleem. Bij het boren naar olie raakt de boorstaaf in horizontale richting in trilling. Aangezien de boorkop breder is dan de boorstaaf is dit aanvankelijk niet zo erg, maar als de amplitude van de trilling te groot wordt, gaat de boorstaaf tegen de wand van het boorgat slaan. De boorstaaf raakt hierdoor beschadigd en er is zelfs eens een boorstaaf gebroken. We beschouwen het volgende vereenvoudigde model:

$$\ddot{x} + 2\mu\dot{x} + \omega^2 x = a\sin(\nu t).$$

Waarin: x de horizontale uitwijking van de boorstaaf is; μ een dempingsconstante; $\omega = \frac{1}{\sqrt{L}}$ is de eigenfrequentie van de boorstaaf met lengte L; a en ν zijn respectievelijk de sterkte en de frequentie van de aandrijving, onder andere bepaald door de omwentelingssnelheid van de boorstaaf.

Van de oliemaatschappij hebben we de volgende parameterwaarden opgekregen:

$$\mu = \frac{1}{4}, \quad a = 1, \quad \text{en} \quad \nu^2 = \frac{8}{9}.$$

Als eenheid van lengte in verticale richting hebben we de kilometer gekozen. De lengte L (in kilometers) van de boorstaaf geeft tevens de diepte waarop de boorkop zich bevindt. Bij horizontale uitwijking 2 (in inches) slaat de boorstaaf tegen de wand van het boorgat. We nemen aan dat het uitdempende deel van de uitwijking $x(t)$ verwaarloosd mag worden. Op welke boordiepte verwacht je de eerste problemen?

Oefening 5.13 Een parachutespringer die samen met zijn parachute een massa m heeft springt van grote hoogte naar beneden en opent meteen de parachute. De valsnelheid van de parachutist nadert naar een constante waarde. We bekijken twee modellen, één waarin de remkracht van de parachute gelijk is aan cV en één waarin de remkracht gelijk is aan cV^2. Hierin is V de valsnelheid en c een zekere constante, die experimenteel bepaald moet worden. Bereken voor elk van de twee modellen de limietwaarde van de valsnelheid. Let in het bijzonder op de afhankelijkheid van m.

Oefening 5.14 Een projectiel wordt vanaf de aarde verticaal omhoog geschoten met een snelheid V_0. We verwaarlozen de luchtweerstand, maar nemen wel de afhankelijkheid van de zwaartekracht als funktie van de afstand van het projectiel tot de aarde in beschouwing. De aantrekkingskracht van de aarde bedraagt mgR^2/r^2 waarin m de massa van het projectiel is, g de versnelling van de zwaartekracht, R de straal van de aarde en r de afstand van het projectiel tot het centrum van de aarde. We willen de minimale V_0 berekenen opdat het projectiel niet meer terugvalt naar de aarde. M.a.w. we willen de verticale ontsnappingssnelheid van de aarde weten.

a. Laat zien dat de bewegingsvergelijking luidt:

$$mr'' = -mg\frac{R^2}{r^2}.$$

b. Vermenigvuldig deze vergelijking aan weerszijden met r' en primitiveer. (We krijgen een integratieconstante E en de relatie die ontstaat tussen r' en r kan gezien worden als de wet van energiebehoud).

c. Maak nu gebruik van $(r')^2 \geq 0$ en vind, afhankelijk van E een bovengrens voor r. Voor welke E kan r onbegrensd toenemen?

d. Bereken nu de ontsnappingssnelheid als funktie van g en R. Zij tenslotte gegeven dat $g = 10m/s^2$ en $R = 6400km$. Bereken de expliciete waarde van de ontsnappingssnelheid.

Er is ook een ontsnappingsnelheid, met een kleinere waarde, in het geval we het projectiel in horizontale richting wegschieten. Deze wordt in een latere opgave bepaald.

6 Stelsels differentiaalvergelijkingen

6.1 Algemeen

De meest algemene gedaante van een stelsel eerste orde d.v.'s is als volgt.
Zij $f_1, f_2, \ldots, f_m$ een m-tal funkties van $m+1$ variabelen $t, x_1, x_2, \ldots, x_m$. We nemen aan dat het domein D van de f_i van de vorm $\{(t, x_1, \ldots, x_m) \mid t_1 < t < t_2, (x_1, \ldots, x_m) \in U\}$ is, waarin U een open deelverzameling in $\mathbb{R}^m$ is. Verder nemen we aan dat de f_i continue partiële afgeleiden hebben in D. Beschouw het stelsel d.v.'s,

$$\left\{ \begin{array}{rcl} x_1'(t) & = & f_1(t, x_1(t), \ldots, x_m(t)) \\ x_2'(t) & = & f_2(t, x_1(t), \ldots, x_m(t)) \\ & & \cdots \\ x_m'(t) & = & f_m(t, x_1(t), \ldots, x_m(t)) \end{array} \right. \tag{6.1}$$

waarin de accenten differentiatie naar t voorstellen. Noteren we

$$\mathbf{x}(t) = (x_1(t), \ldots, x_m(t))^T \quad \text{en} \quad \mathbf{f} = (f_1, \ldots, f_m),$$

dan kunnen we het stelsel (6.1) verkort schrijven als

$$\mathbf{x}'(t) = \mathbf{f}(t, \mathbf{x}(t)). \tag{6.2}$$

Het belang van deze stelsels blijkt onder andere uit de volgende opmerking. Beschouw een m-de orde d.v.

$$y^{(m)} = F(t, y, y', \ldots, y^{(m-1)})$$

waarin de accenten differentiatie naar t voorstellen. Schrijf nu

$$x_1(t) = y(t),\ x_2(t) = y'(t), \ldots, x_m(t) = y^{m-1}(t).$$

Dan geldt

$$\left\{ \begin{array}{rcl} x_1'(t) & = & x_2(t) \\ x_2'(t) & = & x_3(t) \\ \cdots & & \cdots \\ x_{m-1}'(t) & = & x_m(t) \\ x_m'(t) & = & F(t, x_1(t), x_2(t), \ldots, x_m(t)) \end{array} \right.$$

met andere woorden, een m-de orde d.v. kan overgevoerd worden in een stelsel van m eerste orde d.v's.
We willen niet te lang stil blijven staan bij deze algemene systemen, maar vermelden wel de volgende *existentie- en uniciteitsstelling* voor de oplossingen.

Stelling 6.1 *Beschouw het stelsel* (6.1) *en kies een punt* $(t_0, \xi_1, \ldots, \xi_m)$ *in het domein van de* f_i. *Dan zijn er* m *funkties* $x_1(t), \ldots, x_m(t)$ *in een omgeving* I *van* t_0 *zó dat* $x_i(t_0) = \xi_i$ *voor* $i = 1, \ldots, m$ *en die voldoen aan het stelsel* (6.1)*. Bovendien is deze oplossing uniek op het interval* I.

Een bijzonder, maar zeer veel voorkomend, geval is wanneer de funkties f_i niet expliciet van t afhangen. Ons stelsel ziet er dan uit als

$$\begin{cases} x_1'(t) &= f_1(x_1(t), \ldots, x_m(t)) \\ x_2'(t) &= f_2(x_1(t), \ldots, x_m(t)) \\ & \ldots \\ x_m'(t) &= f_m(x_1(t), \ldots, x_m(t)) \end{cases} \tag{6.3}$$

of afgekort,

$$\mathbf{x}'(t) = \mathbf{f}(\mathbf{x}(t)).$$

Een stelsel van de vorm (6.3) noemen we een **autonoom** stelsel. Dergelijke stelsels hebben een mooie meetkundige interpretatie. De vektor, of beter vektorwaardige funktie, $\mathbf{x}(t)$ kunnen we interpreteren als een punt in $\mathbb{R}^m$ dat met de tijd beweegt. Het beeld van deze funktie is dan niets anders dan de baan van dit punt. De afgeleide van $\mathbf{x}(t)$ naar t kan gezien worden als de *snelheid* van deze beweging. De funktie $\mathbf{f}(\mathbf{x})$ kan geïnterpreteerd worden als een *vektorveld*. In ieder punt $\mathbf{x}$ van het domein van $\mathbf{f}$ is de vektor $\mathbf{f}(\mathbf{x})$ gegeven. Een bekend en zeer aanschouwelijk voorbeeld van een vektorveld is een stationaire (=onafhankelijk van de tijd) vloeistofstroming. In elk punt waar zich vloeistof bevindt is een snelheidsvektor van de stroming gegeven. Het stelsel d.v.'s (6.3) kan nu geïnterpreteerd worden als de eis dat in ieder punt van de baan $\mathbf{x}(t)$ de snelheid van het punt overeenstemt met de vektor uit het vektorveld ter plekke. In het voorbeeld van de vloeistofstroming betekent dit dat de oplossingen van het stelsel precies de beweging beschrijven van bijvoorbeeld een zeer licht deeltje dat met de vloeistofstroom meebeweegt. Zeer toepasselijk noemen we de baan die door een oplossing $\mathbf{x}(t)$ beschreven wordt een **stroomlijn** van het vektorveld $\mathbf{f}$. Als we de nadruk meer op het stelsel d.v.'s willen leggen spreken we ook van **integraalkromme** of **oplossingskromme**. Voordat we overgaan naar concrete gevallen vermelden we een gevolg van de uniciteitsstelling voor oplossingen.

Gevolg uniciteit oplossingen (autonoom geval).
Door elk punt $\mathbf{x}_0$ in het domein van $\mathbf{f}$, met $\mathbf{f}(\mathbf{x}_0) \neq \mathbf{0}$, gaat precies één stroomlijn.

6.2 Autonome stelsels in $\mathbb{R}^2$, lineaire geval

In deze paragraaf bestuderen we het stelsel

$$\begin{cases} x_1'(t) = f_1(x_1(t), x_2(t)) \\ x_2'(t) = f_2(x_1(t), x_2(t)) \end{cases} \tag{6.4}$$

d.w.z. autonome stelsels in $\mathbb{R}^2$. We nemen aan dat het domein G van de funkties f_1, f_2 een open gebied is en dat ze continue partiële afgeleiden hebben in G. Om te beginnen merken we op dat iedere eerste orde d.v. $y' = F(t, y)$ door een kunstgreep in een dergelijk stelsel kan worden omgevormd. Stel namelijk $x_1 = t$ en $x_2 = y$. Dan geldt trivialerwijs dat $x_1' = 1$ en dus,

$$\begin{cases} x_1'(t) = 1 \\ x_2'(t) = F(x_1(t), x_2(t)) \end{cases}$$

Verder kunnen we tweede orde homogene lineaire d.v.'s $y'' + py' + qy = 0$ met constante coëfficienten omvormen door $x_1(t) = y(t),\ x_2(t) = y'(t)$ te stellen. We vinden,

$$\begin{cases} x_1'(t) = x_2(t) \\ x_2'(t) = -px_2(t) - qx_1(t) \end{cases} \tag{6.5}$$

We beginnen met het eenvoudigste geval, dat van een **lineair vektorveld**,

$$\begin{cases} x_1'(t) = ax_1(t) + bx_2(t) \\ x_2'(t) = cx_1(t) + dx_2(t) \end{cases} \tag{6.6}$$

of korter,

$$\mathbf{x}' = A\mathbf{x}, \quad \text{waarin} \quad A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

een matrix is met constante coëfficienten. De gedaante van het vektorveld $A\mathbf{x}$, en dus van de bijbehorende stroomlijnen, hangt sterk af van de eigenwaarden van A. Neem als eerste voorbeelden van vektorvelden, met x, y als afkortingen voor x_1 resp. x_2,

$$\text{a)} \begin{bmatrix} 2x \\ y \end{bmatrix} \qquad \text{b)} \begin{bmatrix} -2x \\ -y \end{bmatrix} \qquad \text{c)} \begin{bmatrix} x \\ -y \end{bmatrix}.$$

De oplossingen van de bijbehorende stelsels (6.6) zijn,

$$\text{a)} \begin{bmatrix} C_1e^{2t} \\ C_2e^{t} \end{bmatrix} \qquad \text{b)} \begin{bmatrix} C_1e^{-2t} \\ C_2e^{-t} \end{bmatrix} \qquad \text{c)} \begin{bmatrix} C_1e^{t} \\ C_2e^{-t} \end{bmatrix}$$

waarin C_1, C_2 willekeurige integratieconstanten zijn. Controleer dit en maak zelf een schets van deze stelsels krommen.

Een ander drietal voorbeelden is,

$$\text{d)} \begin{bmatrix} x - y \\ y + x \end{bmatrix} \qquad \text{e)} \begin{bmatrix} -x - y \\ -y + x \end{bmatrix} \qquad \text{f)} \begin{bmatrix} -y \\ x \end{bmatrix}$$

De oplossingen van de bijbehorende stelsels (6.6) zijn,

$$\text{d)}\ e^{t}\mathbf{f}(t) \qquad \text{e)}\ e^{-t}\mathbf{f}(t) \qquad \text{f)}\ \mathbf{f}(t) = \begin{bmatrix} C_1\cos t + C_2\sin t \\ C_1\sin t - C_2\cos t \end{bmatrix}$$

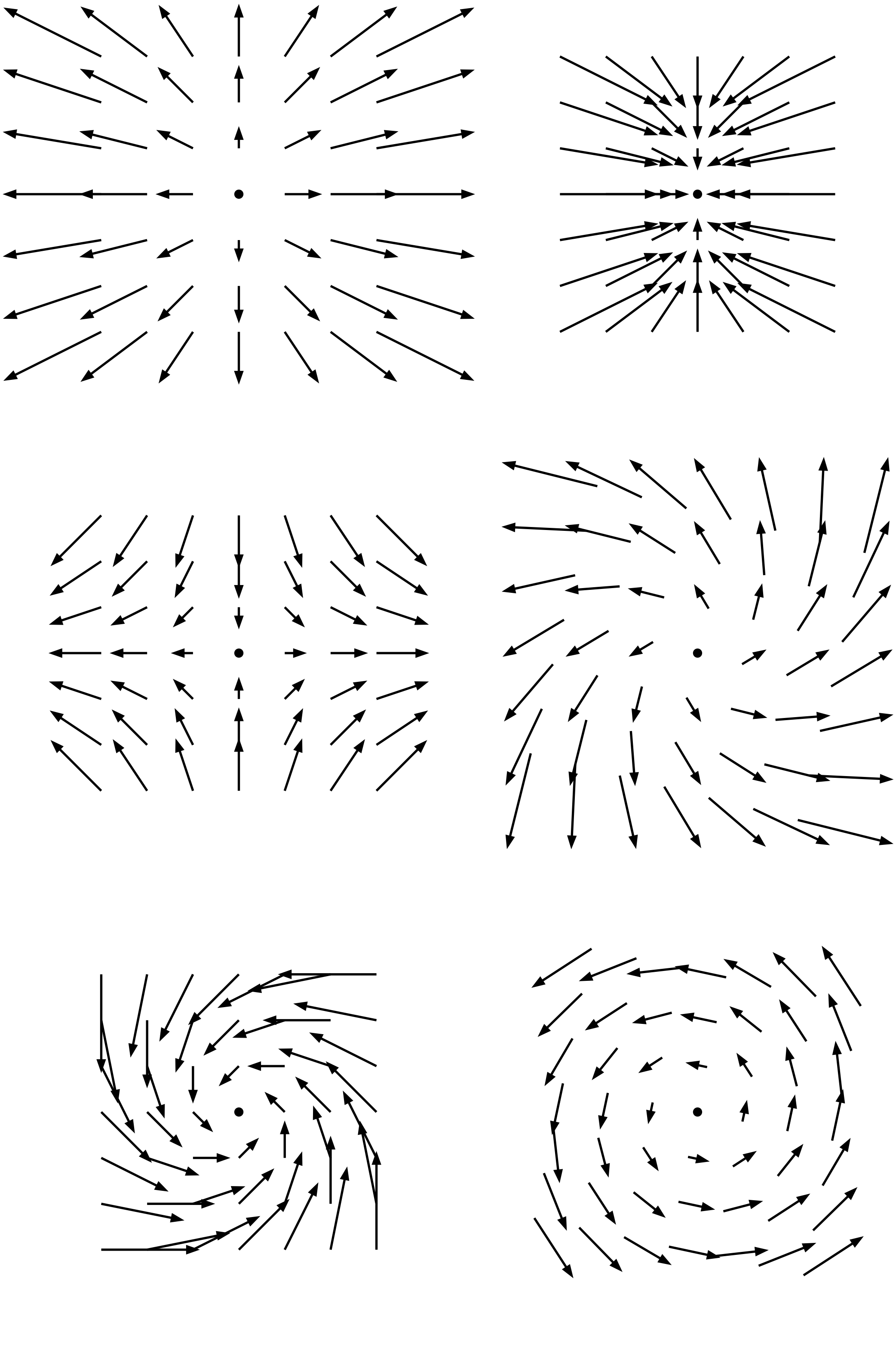
waarin C_1, C_2 willekeurige integratieconstanten zijn.

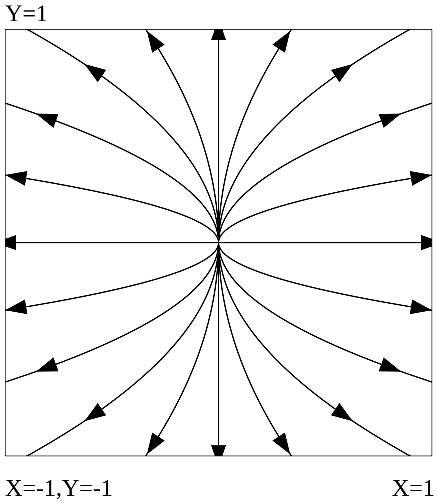
Op de volgende twee bladzijden geven we een schets van deze zes vektorvelden en van de bijbehorende integraalkrommen. Bij de schetsen van de vektorvelden hebben we de lengte van de vektoren voor de overzichtelijkheid een factor drie kleiner getekend hebben dan ze in werkelijkheid zijn.

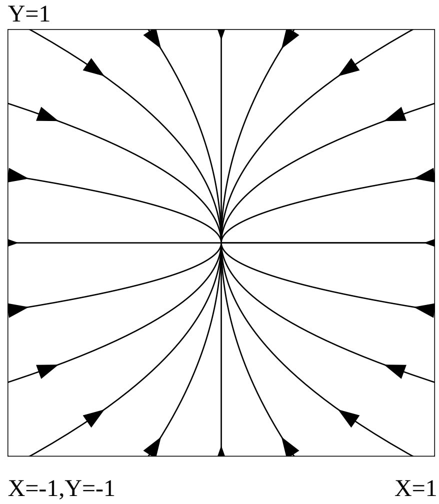
Bovenstaande situaties zijn typerend voor de oplossingen die we voor een willekeurig stelsel (6.6) kunnen hebben. Zij $\lambda_{1,2}$ het tweetal eigenwaarden van A en stel dat ze reëel en verschillend zijn. Dan is er een basis van eigenvektoren en A is diagonaliseerbaar, d.w.z. er is een matrix S zó dat $SAS^{-1} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$. Voer de nieuwe variabelen u, v in d.m.v. $(u, v)^T = S(x, y)^T$. Uit (6.6) volgt nu

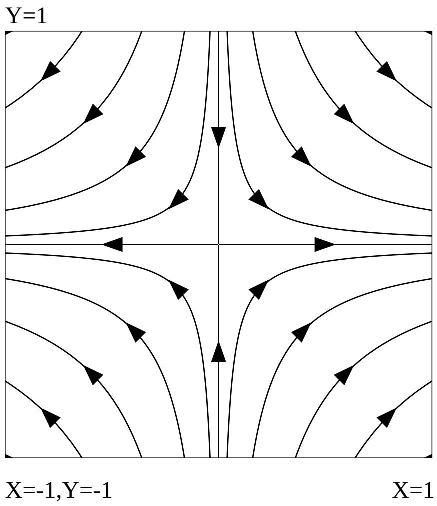
$$\begin{cases} u'(t) = \lambda_1 u(t) \\ v'(t) = \lambda_2 v(t) \end{cases} \tag{6.7}$$

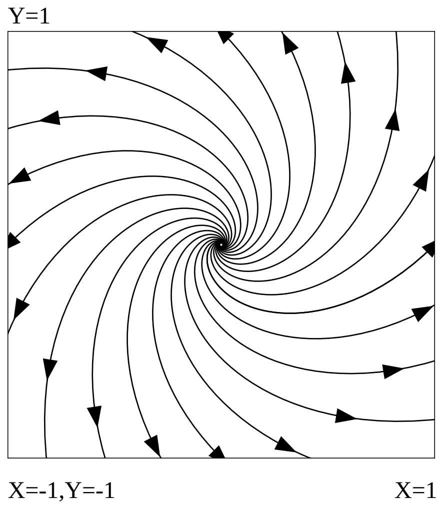
welke eenvoudig kunnen worden opgelost, zie voorbeelden a), b) en c).

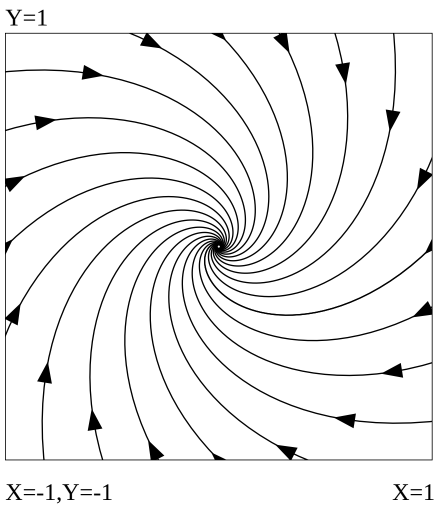


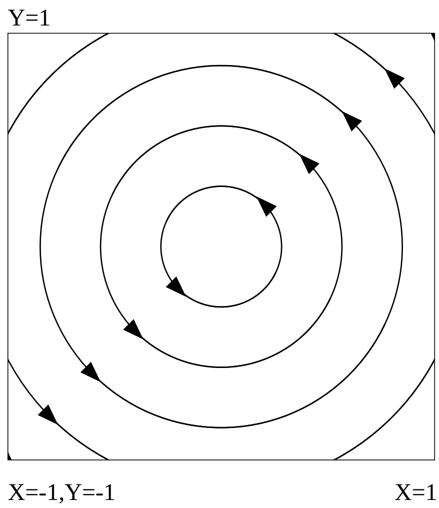

Y=1
X=-1,Y=-1
X=1


Y=1
X=-1,Y=-1
X=1


Y=1
X=-1,Y=-1
X=1


Y=1
X=-1,Y=-1
X=1


Y=1
X=-1,Y=-1
X=1


Y=1
X=-1,Y=-1
X=1

Het is duidelijk dat het teken van de λ_i belangrijk is. Als A complexe eigenwaarden heeft dan is er een S zó dat SAS^{-1} van de vorm

$$R\begin{bmatrix}\cos\phi & -\sin\phi\\ \sin\phi & \cos\phi\end{bmatrix}$$

is, waarin $R = |\lambda_1|$, $\phi = \arg(\lambda_1)$. Uit de voorbeelden blijkt dat nu het teken van $\cos\phi$ belangrijk is. Globaal gezien hebben we de volgende situaties,

a) $\lambda_1 > \lambda_2 > 0$, het **instabiel knooppunt**.

b) $\lambda_1 < \lambda_2 < 0$, het **stabiele knooppunt**.

c) $\lambda_1 > 0 > \lambda_2$, het **zadelpunt**.

d) $\lambda_{1,2}$ complex en Re$\lambda_1 > 0$, het instabiele **spiraalpunt**.

e) $\lambda_{1,2}$ complex en Re$\lambda_1 < 0$, het **stabiele spiraalpunt**.

f) $\lambda_{1,2}$ complex en Re$\lambda_1 = 0$, het **centrumpunt**.

Als voorbeeld kunnen we het stelsel (6.5) nemen, dat van de vergelijking $y'' + py' + qy = 0$ afkomt. De eigenwaardevergelijking luidt $\lambda^2 + p\lambda + q = 0$ en dit is precies de karakteristieke vergelijking van onze tweede orde d.v. We hebben reeds eerder gezien hoe het karakter van de λ_i bepalend is voor het gedrag van de oplossingen van de d.v.

6.3 Oefeningen

Oefening 6.1 [Diskreet versus kontinu]

Model met discrete tijd (zie ook Oefening 2.1):

$$\begin{bmatrix}x_{n+1}\\ y_{n+1}\end{bmatrix} = \begin{bmatrix}x_n\\ y_n\end{bmatrix} \qquad \text{met} \quad \mathbf{A} = \begin{bmatrix}0.4 & 2\\ 0.2 & 0\end{bmatrix}$$

Model met continue tijd:

$$\begin{bmatrix}x'\\ y'\end{bmatrix} = \mathbf{A}\begin{bmatrix}x\\ y\end{bmatrix} \qquad \text{ook met} \quad \mathbf{A} = \begin{bmatrix}0.4 & 2\\ 0.2 & 0\end{bmatrix} \tag{6.8}$$

a. Bereken de eigenvectoren en de eigenwaarden van $\mathbf{A}$ (zie Oefening 2.1).
b. Schets de vectoren $\mathbf{A}^n\,\mathbf{x}$, $\quad n = 0, 1, 2, ...$, waarbij $\mathbf{x}$ een eigenvector is (voor beide eigenvectoren; zie Oefening 2.1). Schets ook de kromme $\begin{bmatrix}x(t)\\ y(t)\end{bmatrix}$ voor $t \geq 0$ waarbij $\begin{bmatrix}x(t)\\ y(t)\end{bmatrix}$ de oplossing is van (6.8) met $\begin{bmatrix}x(0)\\ y(0)\end{bmatrix}$ een eigenvektor van $\mathbf{A}$.

c. Dezelfde vraag als in b., maar nu met, respektievelijk, $\mathbf{x} = e_1$, en $\begin{bmatrix}x(0)\\ y(0)\end{bmatrix} = e_1$.

d. Bepaal voor beide systemen de evenwichtspunten en maak voor beide systemen het faseportret, waarmee je in onderdeel c. begonnen bent, af.
Hoe zit het met de stabiliteit van de evenewichten?

Oefening 6.2
Beschouw het volgende model met continue tijd:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 1 & 4 \\ -2 & -3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

a. Bereken de eigenvectoren met de bijbehorende eigenwaarden.

b. Bepaal de evenwichtspunt(en). Wat is de aard van de evenwicht(en)?

c. Schets het faseportret, d.w.z, schets een aantal oplossingskrommen zodat je een goed beeld krijgt over de oplossing bij iedere mogelijke beginwaarde.

6.4 Autonome stelsels in $\mathbb{R}^2$, algemene geval

We bekijken nu algemene autonome stelsels (6.4) in $\mathbb{R}^2$. Bij iteraties zochten we naar evenwichtspunten, hier zoeken we het analogon daarvan, de zogenaamde **stationaire oplossingen** of **evenwichtsoplossingen**. Dit zijn oplossingen (a, b), die constant zijn in de tijd, hetgeen uitsluitend mogelijk is als $f_1(a, b) = f_2(a, b) = 0$. Stel dat (a, b) een stationaire oplossing is. Net als in §4.3 lineariseren we ons stelsel rond het punt (a, b). Voer daartoe de nieuwe variabelen $u = x - a$, $v = y - b$ in (met x, y afkorting voor x_1, resp. x_2.). Ons stelsel gaat over in

$$\begin{cases} u' = f_1(u + a, v + b) \\ v' = f_2(u + a, v + b) \end{cases}$$

Stel dat $D(\mathbf{f})(\xi)$ de Jacobi-matrix van $\mathbf{f} = (f_1, f_2)^T$ in het punt $\mathbf{a} = (a, b)^T$ is (zie (4.17)). Ons stelsel kan herschreven worden in de vorm

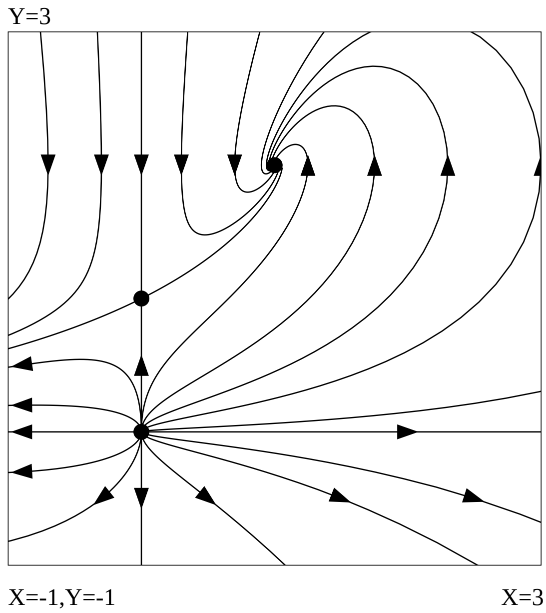
$$\begin{bmatrix} u' \\ v' \end{bmatrix} = D(\mathbf{f})(\mathbf{a}) \begin{bmatrix} u \\ v \end{bmatrix} + \text{hogere orde termen in } u, v.$$

en de gelineariseerde vorm krijg je door de hogere orde termen weg te gooien. De constante term in de ontwikkeling van $\mathbf{f}$ vervalt omdat $\mathbf{f}(a, b) = (0, 0)$. Het idee is, dat als we in de buurt van de stationaire oplossing blijven, d.w.z. u, v klein, de hogere orde termen nauwelijks van invloed zijn en we, lokaal gezien, te maken hebben met een lineair autonoom stelsel van het type dat we hierboven bestudeerd hebben. Het blijkt dat de kwalitatieve vorm van de oplossingen rond een stationaire oplossing tot één van de boven genoemde gevallen a) tot en met f) behoort. Voor de matrix A moeten we nu de matrix $D(\mathbf{f})(\xi)$ nemen. Voorzichtigheid is geboden in het geval f). In het lineaire geval zien we gesloten banen rond het punt $(0, 0)$, maar in het niet-lineaire geval is de geslotenheid van deze banen meestal verstoord.

Voorbeeld 6.1 Beschouw het stelsel

$$\begin{cases} x' = 2x - xy \\ y' = y + xy - y^2 \end{cases}$$

Evenwichtsoplossingen volgen uit de vergelijkingen $2x - xy = y + xy - y^2 = 0$ en dus $(0, 0)$, $(0, 1)$, $(1, 2)$. Rond het punt $(0, 0)$ staat onze vergelijking reeds in de voor lineariseren geschikte gedaante. We lezen direkt de Jacobi-matrix $\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$ af en we hebben te maken met een instabiele knoop. In het punt $(0, 1)$ luidt de Jacobi-matrix $\begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix}$. De eigenwaarden



FIGUUR 2. *Integraalkrommen voor het stelsel d.v. in Voorbeeld* 6.1.

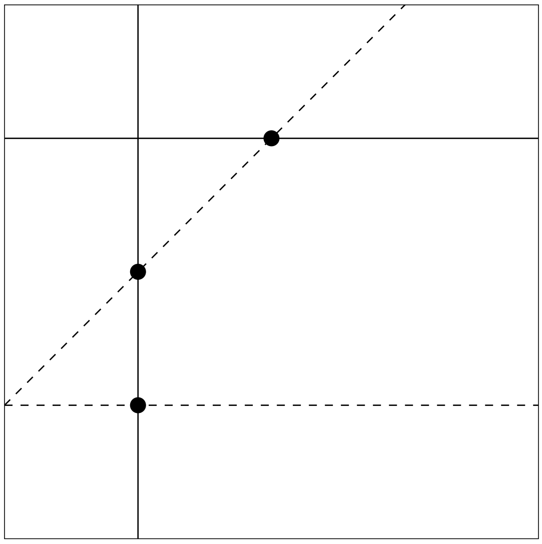
zijn ± 1 en we hebben een zadelpunt. In $(1, 2)$ is de Jacobi-matrix $\begin{bmatrix} 0 & -1 \\ 2 & -2 \end{bmatrix}$ met eigenwaarden $-1 \pm i$, we hebben een stabiel spiraalpunt. Controleer deze berekeningen! Het plaatje met de banen van de oplossingen ziet er als in FIG. 2.

Toegegeven, bovenstaand plaatje is met de computer gegenereerd en uit de ligging van de evenwichtspuntspunten en hun aard is het niet meteen duidelijk dat de banen er als boven uit zullen zien. Het is echter mogelijk om met behulp van wat schetsen een inzicht te krijgen in het verloop van de banen zonder daarbij de computer te gebruiken. We geven in het vierkant $-1 \leq x, y \leq 3$ de punten waarvoor geldt dat $x' = 0$ aan met een gesloten lijn, de punten waarvoor $y' = 0$ met een onderbroken lijn. De verzameling punten met $x' = 0$ bestaat uit de rechten $x = 0$ en $y = 2$. De verzameling punten met $y' = 0$ bestaat uit de rechten $y = 0$ en $y = x + 1$. De evenwichtspunten zijn natuurlijk precies de snijpunten van deze twee verzamelingen (zie FIG. 3.

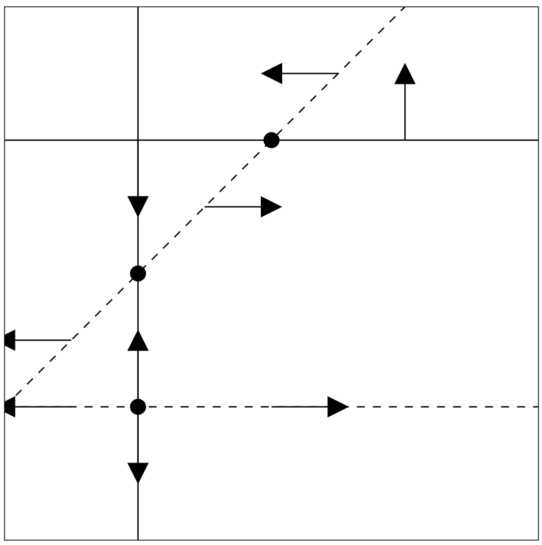
Op de stukken van de doorlopende lijnen zijn de vektoren van het vektorveld verticaal, x' is daar immers 0. We moeten alleen nagaan of de richting naar boven of naar beneden is. Op de gearceerde verzameling zijn de vektoren horizontaal want daar is $y' = 0$. We moeten alleen nog nagaan of de richting links of rechts is. We vinden het plaatje in FIG. 4 (ga na!),

Tenslotte merken we op dat de verzamelingen $x' = 0$ en $y' = 0$ het platte vlak in een aantal gebieden opdelen. In elk van deze gebieden geldt dat de vektoren uit het vektorveld allemaal in hetzelfde kwadrant wijzen. Men zou dat ook nog kunnen aangeven door bijvoorbeeld in het rechts-midden gebied een noordoost pijl te tekenen. Je moet dat dan wel losser interpreteren: Terwijl de getekende horizontale en vertikale pijlen aangeven dat het vektorveld precies horizontaal/vertikaal loopt, symboliseert de noordoost pijl slechts een kwadrant. Zo krijgen we dus ook voor deze gebieden een beeld hoe het vektorveld verloopt.

Hopelijk wordt het nu wat duidelijker waarom de integraalkrommen lopen zoals we aan het begin getoond hebben. De aard van de punten $(0, 0)$ en $(0, 1)$ is reeds duidelijk te zien. Ook zien we dat $(1, 2)$ een spiraalpunt gaat worden, maar het is de vraag of deze afstotend of aantrekkend is. Dit moeten we beslissen met behulp van de eigenwaarden van de lokale



FIGUUR 3. *De snijpunten van de "krommen"* $x' = 0$ *(doorgetrokken lijn —) en "*$y' = 0$ *(gestreept lijn − − −) zijn precies evenwichtsoplossingen.*



FIGUUR 4. *Richting van verabdering in de buurt van de "krommen"* $x' = 0$ *en "*$y' = 0$.

Jacobimatrix van het vektorveld.

Volterra-vergelijking. In de 20er jaren ontdekte de Italiaanse bioloog Umberto d'Ancona dat gedurende de eerste wereldoorlog het percentage roofvissen (haaien en roggen) in de Middellandse zee sterk toenam. Hier volgt een tabel van het percentage roofvissen in de totale visvangst rond de haven van Fiume.

1914	1915	1916	1917	1918	1919	1920	1921	1922	1923
11.9%	21.4%	22.1%	21.2%	36.4%	27.3%	16.0%	15.9%	14.8%	10.7%

De reden voor deze procentuele toename moet gezocht worden in het feit dat er gedurende de oorlog minder gevist werd. Het mechanisme waarmee dit gebeurde was echter een volkomen raadsel voor d'Ancona. Een verklaring werd tenslotte ontdekt door de bekende Italiaanse wiskundige Vito Volterra (1860-1940). Het model is een voorbeeld van een continue beschrijving van populatiedynamica.
Stel dat de populatie prooivissen gegeven wordt door $x(t)$ en de populatie roofvissen door $y(t)$. In principe ondergaan de prooivissen een exponentiële groei gegeven door $x' = \alpha x$ voor zekere $\alpha > 0$, maar door de aanwezigheid van roofvissen wordt deze groei geremd en we kiezen $x'(t) = \alpha x(t) - \beta x(t)y(t)$ voor zekere $\beta > 0$. Voor de roofvissen geldt dat $y'(t)$ bestaat uit een bijdrage $-\gamma y(t)$, met $\gamma > 0$, door sterfte en een bijdrage $\delta x(t)y(t)$, met $\delta > 0$, door geboorte, die natuurlijk sterk afhangt van de hoeveelheid aanwezige prooivissen. We vinden het stelsel

$$\begin{cases} x'(t) = \alpha x(t) - \beta x(t)y(t) \\ y'(t) = -\gamma y(t) + \delta x(t)y(t) \end{cases} \tag{6.9}$$

Merk op, dat $(e^{\alpha t}, 0)$ en $(0, e^{-\gamma t})$ oplossingen zijn, d.w.z. de positieve X en Y-as zijn integraalkrommen. In het bijzonder betekent dit dat een integraalkromme die in het eerste kwadrant van $\mathbb{R}^2$ start, dit kwadrant niet verlaat. Banen van oplossingen snijden elkaar immers niet. De evenwichtspunten blijken $(0,0)$ en $(\gamma/\delta, \alpha/\beta)$ te zijn. We zien direkt dat $(0,0)$ een zadelpunt is. De eigenwaarden van de gelineariseerde d.v. bij $(\gamma/\delta, \alpha/\beta)$ zijn $\pm i\sqrt{\alpha\gamma}$. Dit wijst op een centrumpunt. Evenals in het vorige voorbeeld kunnen we hier ook de verzameling punten aangeven waar $x' = 0$ (gesloten lijn) en $y' = 0$ (onderbroken lijn) samen met enkele vektoren uit het vektorveld: zie FIG. 5. Voor een schets van de banen van de oplossingen bij $\alpha = \beta = \gamma = \delta = 1$, zie FIG. 6.

Merk op dat de banen gesloten zijn. Dit is een zeer bijzondere situatie en een gevolg van het feit dat het systeem **integreerbaar** is. Er bestaat namelijk een funktie $F(x, y)$ zó dat voor elke oplossing $\overline{x(t), y(t)}$ geldt $F(x(t), y(t)) =$ constant. Uit het stelsel d.v.'s volgt namelijk dat voor elke oplossing geldt,

$$x'(\frac{\gamma}{x} - \delta) + y'(\frac{\alpha}{y} - \beta) = 0.$$

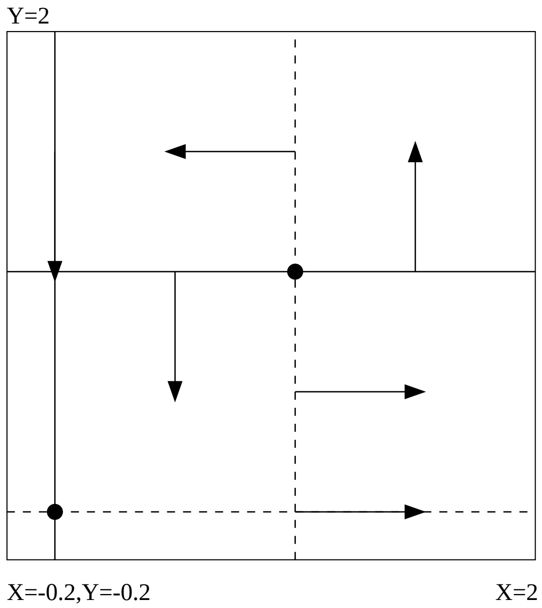
Primitiveren aan beide zijden levert

$$\gamma \log x(t) - \delta x(t) + \alpha \log y(t) - \beta y(t) = \text{constant}.$$

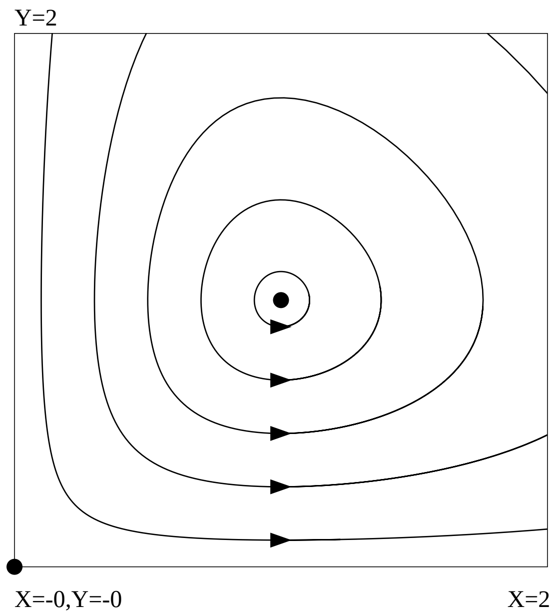
De funktie $F(x, y) = \gamma \log x - \delta x + \alpha \log y - \beta y$ heeft in het eerste kwadrant een maximum in $(\gamma/\delta, \alpha/\beta)$. Verder gaat $F(x, y)$ naar $-\infty$ als x of y naar 0 of ∞ gaan. De integraalkrommen zijn nu gewoon de niveaulijnen (hoogtelijnen) van $F(x, y)$ en ze zijn gesloten omdat de grafiek van $F(x, y)$ in het eerste kwadrant een 'berg' met een enkele top is.

Volterra nam aan dat de *gemiddelde* verhouding van de populaties over een jaar, welke we uiteindelijk meten, door de getallen $\gamma/\delta : \alpha/\beta$ (prooivissen : roofvissen) wordt weergegeven. Introduceer nu een vangstterm in de vergelijkingen door aan $x'(t)$ en $y'(t)$ de term $-\varepsilon x(t)$ respectievelijk $-\varepsilon y(t)$ toe te voegen. We vinden het nieuwe stelsel

$$\begin{cases} x'(t) = (\alpha - \varepsilon)x(t) - \beta x(t)y(t), \\ y'(t) = -(\gamma + \varepsilon)y(t) + \delta x(t)y(t). \end{cases}$$



FIGUUR 5. *De snijpunten De "krommen" $x' = 0$ (doorgetrokken lijn —) en "$y' = 0$ (gestreept lijn – – –) met hun snijpunten en sommige richtingsvectoren.*



FIGUUR 6. *Een schets van de banen van de oplossingen bij $\alpha = \beta = \gamma = \delta = 1$ van het model* (eq:LotVol).

Het nieuwe centrumpunt is

$$\left(\frac{\gamma + \varepsilon}{\delta}, \frac{\alpha - \varepsilon}{\beta} \right) .$$

We zien nu wat er gebeurt als we de visvangst-coëfficient ε laten toenemen. De populatie prooivissen neemt toe (!) in verhouding tot de populatie roofvissen. Dit staat bekend als het

Volterra effekt in de populatiedynamica. Omgekeerd, als de visvangst afneemt, zal het percentage roofvissen toenemen. Dit is een mogelijke verklaring voor de waargenomen toename in de eerste wereldoorlog.

Het zal duidelijk zijn dat er veel kritiek kan worden uitgeoefend op het Volterra model, dat tegenwoordig niet meer als erg realistisch beschouwd wordt. Historisch is het echter een belangrijk model. Het was één van de eersten in zijn soort en de mogelijkheid tot verklaring van oscillaties binnen populatie-aantallen kwam erdoor ook binnen bereik. Later zijn er vele varianten en verbeteringen voorgesteld, waarvan we er één tijdens het praktikum zullen bestuderen. Vaak voegt men ook de naam **Lotka** toe aan het **Volterra-model**. Alfred James Lotka (1880-1949) was een Amerikaan die de wiskundige aspekten van de populatiedynamica uitgebreid bestudeerde.

Bij niet-lineaire autonome stelsels in $\mathbb{R}^2$ kan het gebeuren dat er een geïsoleerde periodieke oplossing optreedt. Een dergelijke oplossing herkennen we aan het feit dat de integraalkromme gesloten is.

Limiet cykel. Beschouw het stelsel

$$\begin{cases} x' = x - y - x(x^2+y^2) \\ y' = y + x - y(x^2+y^2) \end{cases}$$

Zonder al te veel te rekenen kunnen we zien dat hier iets merkwaardigs gebeurt. Allereerst de stationaire punten. Er blijkt er slechts één te zijn, $(0,0)$ en dit is een afstotend spiraalpunt (ga na!). Bekijk nu het vektorveld beperkt tot de cirkel $x^2+y^2=2$. Dit is $(-x-y, x-y)$. Het inwendig produkt met de normaalvector op de cirkel $x^2+y^2=2$ is $((x,y),(-x-y,x-y)) = -x^2-y^2 = -2$. De hoek tussen de (naar buiten gerichte) normaal en het vektorveld is dus overal groter dan $\pi/2$. De integraalkrommen kunnen dus niet uit de cirkelschijf $x^2+y^2<2$ "ontsnappen". Verder is er ook geen stationaire oplossing waar ze naar toe kunnen convergeren, het enige stationaire punt is immers $(0,0)$ en dat is afstotend. Bovendien mag een integraalkromme zichzelf niet doorsnijden. Door al deze eisen wordt een integraalkromme als het ware gedwongen naar een periodieke oplossing te convergeren. Dat zien we inderdaad terug in FIG. 7 als we naar het plaatje van de integraalkrommen kijken.

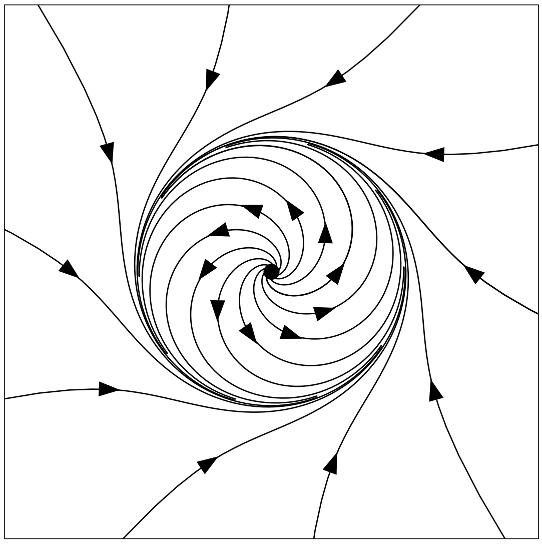
Terzijde merken we op dat in dit voorbeeld de vergelijking zó is gemaakt dat we in poolcoördinaten r en ϕ een eenvoudig stelsel krijgen. Merk namelijk op dat $x = r\cos\phi$, $y = r\sin\phi$. Differentiatie van beide gelijkheden geeft $x' = r'\cos\phi - r\phi'\sin\phi$ en $y' = r'\sin\phi + r\phi'\cos\phi$. Uit deze vergelijkingen volgt $xx'+yy' = rr'$ en $xy'-yx' = (x^2+y^2)\phi'$. Met behulp van het stelsel d.v.'s substitueren we x' en y' in deze gelijkheden en na overgang op poolcoördinaten vinden we,

$$\begin{cases} r' = r(1-r^2) \\ \phi' = 1 \end{cases}$$

Merk op, als $r<1$ dan geldt $r'>0$, dus r neemt toe, en als $r>1$ dan geldt $r'<0$, dus r neemt af. Verder is $r=1, \phi=t$ een periodieke oplossing.

Met hetzelfde soort argumenten als in bovenstaand voorbeeld, maar dan wat exacter! kan men de volgende stelling bewijzen,

Stelling 6.2 (Poincaré-Bendixon) *Stel dat een oplossing $x_1(t), x_2(t)$ van* (6.4) *voor $t \geq 0$ beperkt blijft tot een begrensd gebied G dat geen stationaire oplossingen bevat. Dan nadert de baan van $x_1(t), x_2(t)$ op den duur naar een gesloten kromme, die zelf de baan is van een periodieke oplossing van* (6.4)



FIGUUR 7. *De integraalkrommen worden als het ware gedwongen naar een periodieke oplossing te convergeren.*

Een ander gebied waaruit autonome systemen in $\mathbb{R}^2$ voortkomen is dat van de klassieke mechanica in één dimensie. Stel dat een massapunt m zich over (een deel van) de X-as beweegt onder invloed van een kracht die alleen van de plaats en eventueel de snelheid van het massapunt afhangt, $F(x(t), x'(t))$. Volgens de tweede wet van Newton geldt $mx''(t) = F(x(t), x'(t))$. Stel $x_1 = x$ en $x_2 = x'$ dan kan de bewegingsvergelijking omgeschreven worden tot het stelsel

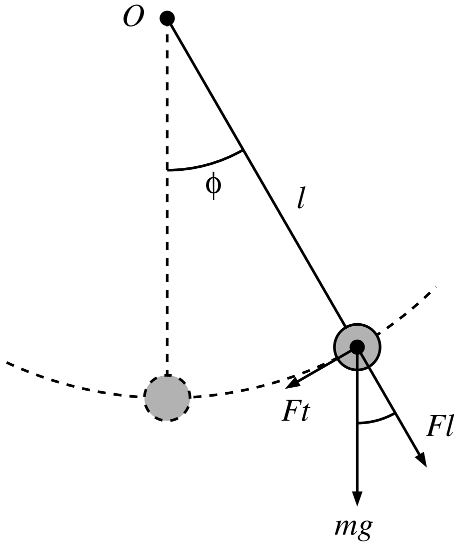
$$\begin{cases} x_1'(t) = x_2(t) \\ x_2'(t) = \frac{1}{m}F(x_1(t), x_2(t)). \end{cases}$$

De verzameling van mogelijke toestanden (x, x') waarin een massapunt zich kan bevinden noemen we de **faseruimte** van het mechanisch systeem. Dit is meestal het x_1x_2-vlak.

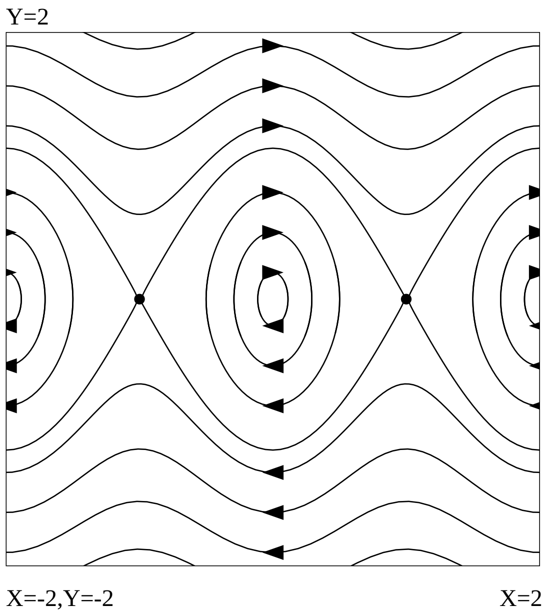
De mathematische slinger. Stel dat de massa m aan een starre gewichtsloze staaf met lengte l is opgehangen aan het punt O. Onder invloed van de zwaartekracht g beschrijft m een slingerbeweging. De positie van m wordt gegeven door de hoek ϕ: zie FIG. 8. We zullen eerst aannemen dat er geen wrijving is met de lucht of met het ophangpunt. De versnelling van m is gelijk aan $l\phi''$. De kracht die op m werkt is precies de component van de zwaartekracht die loodrecht op de staaf staat, $-mg\sin\phi$. De bewegingsvergelijking wordt $ml\phi'' = -mg\sin\phi$ en dus,

$$l\phi'' + g\sin\phi = 0.$$

Als we weten dat de beweging een kleine uitwijking heeft kunnen we de benadering $\sin\phi \approx \phi$ gebruiken. Dit is een voorbeeld van de boven besproken techniek van lineaire benadering. De gelineariseerde vergelijking wordt $l\phi'' + g\phi = 0$ en dit is precies de vergelijking van de harmonische oscillator met periode $2\pi\sqrt{l/g}$. De slingertijd voor kleine uitwijkingen is dus $2\pi\sqrt{l/g}$. We maken nu een studie van alle bewegingen, dus ook grote uitwijkingen, maar doen dit kwalitatief. D.w.z. het gaat meer om de vorm van de integraalkrommen dan om de precieze vergelijkingen. De posities ϕ van de slinger kunnen worden voorgesteld door de cirkel



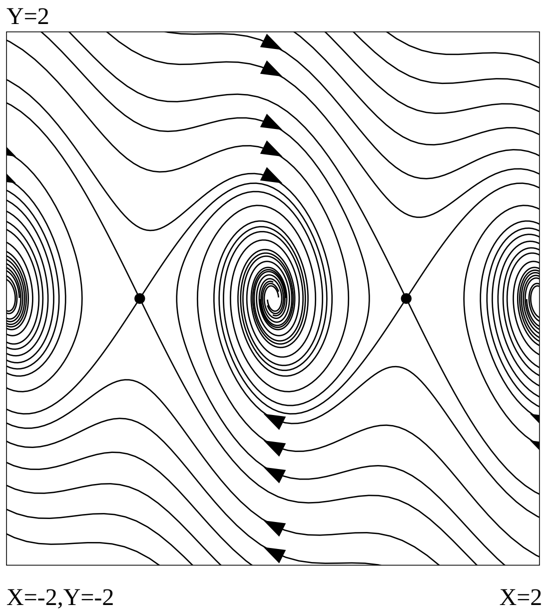
FIGUUR 8. *Onder invloed van de zwaartekracht g beschrijft m een slingerbeweging.*



FIGUUR 9. *Faseplaatje van de bewegingen van een slinger.*

in het platte vlak met straal l en middelpunt O. Bij iedere positie is het in principe mogelijk de slinger een willekeurige snelheid te geven. Deze kunnen we loodrecht op het platte vlak uitzetten. Doen we dit voor iedere positie dan zien we dat onze faseruimte $\{(\phi, \phi') \mid \phi \in \mathrm{cirkel}, \phi' \in \mathbb{R}\}$ een cylinder is. Deze kunnen we in het platte vlak weergeven door hem langs de lijn $\phi = \pi \equiv -\pi$ open te knippen en uit te vouwen. Overtuig je ervan dat de mogelijke bewegingen van de slinger door het faseplaatje in FIG. 9 worden weergegeven,

Merk op dat dit systeem een centrumpunt heeft corresponderend met de stabiele bewegingsloze ophanging en een zadelpunt corresponderend met het balanceren van m recht boven het ophangpunt. Dat dit laatste punt instabiel is weten we ook uit ervaring. Tevens zijn de integraalkrommen gesloten en dit zit hem in het feit dat ons systeem integreerbaar is. Verme-



FIGUUR 10. *Faseplaatje voor de beweging van een slinger met wrijving.*

nigvuldig de vergelijking namelijk met ϕ' en primitiveer. We vinden,

$$\frac{l}{2}\phi'(t)^2 - g\cos\phi(t) = \text{constant}. \tag{6.10}$$

Dit is een voorbeeld van een behoudswet. Vermenigvuldigen we even in gedachten deze gelijkheid met m, dan is $ml/2\phi'(t)^2$ de kinetische energie en $-mg\cos\phi(t)$ de potentiële energie ten tijde t. De som van deze twee is constant in de tijd en we hebben te maken met energiebehoud. Hoe dan ook, onze integraalkrommen zijn precies de niveaulijnen van $(l/2)(\phi')^2 - g\cos\phi$. Verifieer zelf of dit met het faseplaatje overeenstemt.

Neem nu aan dat de slinger een wrijvingskracht $-c\phi'$ ondervindt. De nieuwe bewegingsvergelijking luidt,

$$l\phi''(t) + \frac{c}{m}\phi'(t) + g\sin\phi(t) = 0.$$

De integraalkrommen in het faseplaatje zien er nu als in FIG. 10.

Merk op dat het centrumpunt nu een stabiel spiraalpunt is geworden en dat we nog steeds het zadelpunt hebben.

6.5 Oefeningen

Oefening 6.3 Bereken de stationaire punten. Bereken in elk stationair punt de eigenwaarden van het in dat punt gelineariseerde stelsel.

a.

$$\begin{cases} \dot{x} = y - xy \\ \dot{y} = -x + xy \end{cases}$$

b.

$$\begin{cases} \dot{x} = x^2 - 3x + 2 \\ \dot{y} = xy^2 - x \end{cases}$$

c.

$$\begin{cases} \dot{x} = \cos x + \sin y \\ \dot{y} = \sin x + \cos y \end{cases}$$

Oefening 6.4 Beschouw:

$$\begin{cases} \dot{x} = 5\left(1 - kx - \dfrac{2y}{1+x}\right) x \\ \dot{y} = \left(-1 + \frac{2x}{1+x}\right) y. \end{cases}$$

a. Bepaal de stationaire punten.

b. Bereken bij elk vast punt voor welke waarden van k dit punt stabiel is.

Oefening 6.5 Stel we laten een slinger zonder demping, zoals in de voorgaande paragraaf beschreven, vanuit een positie met hoek ϕ_0 los vanuit stilstand. Geef, m.b.v. (6.10), een (intgraal)-uitdrukking voor de slingertijd van deze slingerbeweging en laat zien dat deze toeneemt als we ϕ_0 laten toenemen.

Oefening 6.6 Een hond rent met snelheid h in de richting van zijn baas die met een constante snelheid w in een rechte lijn loopt. Om de gedachten te bepalen zullen we aannemen dat de baas op tijdstip $t = 0$ in $(0,0)$ vertrekt in de positieve y-richting. De hond start dan in het punt (x_0, y_0). Geef het stelsel differentiaalvergelijkingen voor de baan $(x(t), y(t))$ die de hond aflegt. Merk op dat we door de substitutie $y(t) = u(t) + wt$ er een autonoom stelsel van kunnen maken. Kun je een interpretatie geven van deze substitutie?

Oefening 6.7 Het volgende stelsel modelleert twee diersoorten die elkaar om dezelfde voedselvoorraad beconcurreren,

$$\begin{cases} N_1' = (a_1 - d_1(bN_1 + cN_2))N_1 \\ N_2' = (a_2 - d_2(bN_1 + cN_2))N_2 \end{cases}$$

De constanten zijn allen positief. Bepaal de evenwichtspunten van dit stelsel. Laat zien dat als $a_1 d_2 > a_2 d_1$ de populatie N_2 uitsterft en N_1 een evenwicht bereikt (Volterra's exclusie principe).

Oefening 6.8 We hebben een meer dat als uitstekend viswater bekend staat. Uiteraard is dit een trekpleister voor veel sportvissers. Om de fluktuaties in de visstand te bestuderen maken we een simpel model voor de vis-visser interaktie.

Aannamen voor de vissen:

1. Vissen groeien logistisch bij afwezigheid van vissers

2. De aanwezigheid van vissers remt de populatiegroei in een grootte evenredig met het aantal vissers en het aantal vissen.

Aannamen voor de vissers:

1. Vissers worden tot het meer aangetrokken in een mate evenredig met de hoeveelheid vis.
2. Vissers worden ontmoedigd van een bezoek aan het meer in een mate die evenredig is met het aantal aanwezige vissers.

Stel een model op voor deze interaktie en analyseer het. Stel dat een visclub besluit regelmatig vis uit te zetten in het meer. Hoe ziet het model er dan uit? Analyseer deze situatie ook. Wat is het effekt van het uitzetten van vis op de vispopulatie?

Oefening 6.9 Beddington en May (1982) stelden het volgende model op voor de interaktie tussen baleinwalvissen en hun voedselbron, krill (een soort plankton), in de poolzee.

$$\begin{cases} \text{krill: } \dfrac{dx}{dt} = rx\dfrac{1-x}{K} - axy \\ \text{walvissen: } \frac{dy}{dt} = sy\frac{1-y}{bx} \end{cases}$$

Analyseer dit model door de evenwichtspunten te bestuderen.

Oefening 6.10 Stel we hebben een zware massa M in de oorsprong van een rechthoekig coördinatenstelsel en een massa m op positie $\mathbf{r}$. De kracht die de zware massa op m via zwaartekracht uitoefent is volgens Newton gelijk aan

$$\mathbf{F} = -\frac{GMm}{|\mathbf{r}|^3}\mathbf{r}.$$

Hierin is G de gravitatieconstante waarvan de waarde pas lang na Newton's tijd voor het eerst gemeten is (Cavendish, 1798). Toch kon Newton uitstekend uit de voeten zonder de waarde van G. Stel dat m zeer klein is t.o.v. M, we kunnen dan veronderstellen dat M vast in de oorsprong blijft en m, onder invloed van de gravitatie, beweegt. Stel dat de beweging in het x, y-vlak plaatsvindt. Uit Newton's bewegingsvergelijkingen $\mathbf{F} = m\frac{d^2\mathbf{r}}{dt^2}$ volgt:

$$\frac{d^2x}{dt^2} = -GM\frac{x}{|\mathbf{r}|^3}, \quad \frac{d^2y}{dt^2} = -GM\frac{y}{|\mathbf{r}|^3}.$$

De waarde van m is gewoon weggevallen, hetgeen impliceert dat de beweging onder invloed van een zwaartekrachtveld onafhankelijk van de massa van het bewegende object is. We zullen de vergelijkingen niet volledig oplossen, maar alleen kijken naar oplossingen in de vorm van cirkelvormige banen die met constante snelheid doorlopen worden.

a. Stel dat we zo'n cirkelvormige baan als oplossing hebben met straal R en omloopstijd T. Geef een verband tussen R en T. (Hint: vul de oplossing $x = R\cos\omega t, y = R\sin\omega t$ in de vergelijking in)

b. Bij benadering geldt dat de aarde een cirkelvormige baan rond de zon beschrijft en de maan een cirkelvormige baan rond de aarde. De gemiddelde afstanden zijn hierbij 1.5×10^8 respectievelijk 3.84×10^5 kilometer, de omloopstijden zijn 365 respectievelijk 28 dagen. Bereken de verhouding van de massa's van de aarde en de zon. Hoe ver is een geostationaire satelliet van de aarde verwijderd? De aardstraal bedraagt 6370 km. Voeg daar voor de spaceshuttle nog eens 300 km bij. Wat is de omlooptijd van de spaceshuttle?

c. Om de zonnemassa zelf te bepalen moeten we uiteraard G kennen. Uit (subtiele) metingen volgt: $G = 6.67 \times 10^{-11}$ m^{3}s^{-2}kg^{-1}. Bereken de zonnemassa.

Oefening 6.11 We schieten met snelheid v_0 een kogel af onder een hoek α met het aardoppervlak. We brengen een coördinatenstelsel aan met een x-as in horizontale richting en de y-as verticaal. Onze kogel wordt vanuit $(0,0)$ afgeschoten. We nemen aan dat de baan in het eerste kwadrant van ons x,y-vlak ligt. De versnelling van de zwaartekracht bedraagt g en de massa van de kogel m. We nemen aan dat de kogel op tijdstip $t=0$ wordt afgeschoten en we verwaarlozen de wrijving met de lucht. De baan van de kogel wordt gegeven door de funkties $x(t)$ en $y(t)$.

a. Geef een tweede orde differentiaalvergelijking voor x en een vergelijking voor y.

b. Los de vergelijkingen op en geef de baan van de kogel in x,y-coördinaten aan.

c. Gegeven v_0 en α, op welke afstand komt de kogel op de grond terecht?

Oefening 6.12 We hebben hetzelfde coördinatensysteem als in de vorige opgave. We schieten nu een kogel vanaf een (kleine) hoogte h vanuit $(0,h)$ in (horizontale) positieve x-richting af. Geef een vergelijking voor de baan van de kogel in x,y-coördinaten.

We bekijken nu het probleem van de horizontale ontsnappingssnelheid. Vraag: wat is de minimale snelheid waarmee een kogel horizontaal kan worden afgevuurd opdat hij niet op de grond terugvalt. Het principe dat we zullen gebruiken is dat de kromming van de kogelbaan kleiner moet zijn dan de kromming van de aarde. Een goede maat voor de kromming van de grafiek van een funktie f in een punt met horizontale raaklijn is de waarde van $|f''(x)|$ in dat punt. Bereken de kromming van de kogelbaan bij afvuren. Bereken ook de kromming van het aardoppervlak dat gegeven wordt door $y=\sqrt{R^2-x^2}-R$ in het punt $x=0$. Hierin is R de straal van de aarde. Bereken nu de horizontale ontsnappingssnelheid als funktie van g en R. Gegeven dat $g=10m/s^2$ en $R=6400km$ bereken de expliciete waarde van de horizontale ontsnappingssnelheid.

Oefening 6.13 [Tentamensom september 1993]

Gemiddeld over een zeker gebied is, op tijdstip t, $x(t)$ de dichtheid van een gewas X en $y(t)$ de dichtheid van een ander gewas Y. De ontwikkeling van de dichtheden kan beschreven worden met het volgende model:

$$\begin{cases} x' &= \left(9-x^2-y^2\right)x \\ y' &= (4-x-y)\,y. \end{cases} \tag{6.11}$$

a. Hoe heet de ontwikkeling van de dichtheid van gewas Y indien gewas X afwezig is? Wat voor invloed heeft gewas X op de individuele groeisnelheid van gewas Y? Wat is de invloed van Y op X? Hoe zou je de interaktie biologisch interpreteren?

b. Geef in het deel van het x-y–vlak dat biologisch relevant is aan waar $x'=0$, respektievelijk $y'=0$. Geef het tekenverloop van x' en y' aan (middels een pijltje) in het diagram dat zo ontstaat.

c. Het punt $(x_0,y_0)=\left(2-\sqrt{\frac{1}{2}},2+\sqrt{\frac{1}{2}}\right)$ is een evenwichtspunt. Bepaal de overige evenwichtspunten van dit groeimodel. Voor welke evenwichtspunten kan je op grond van de schets in (b) de stabiliteit of de instabiliteit vaststellen? Beschrijf van welk type de evenwichtspunten zijn ("zadelpunt", of iets dergelijks). (Voer dit onderzoek grafisch uit, maar beargumenteer wel je antwoorden)

d. Vier oplossingskrommen "verdwijnen" of "ontstaan" in het evenwichtspunt (x_0,y_0). De eigenvektoren van het stelsel dat ontstaat door te lineairiseren rond (x_0,y_0) geven ons zekere informatie over deze vier speciale oplossingskrommen. Welke informatie geven ze ons?

Lineairiseer rond (x_0,y_0), bereken de eigenvektoren, teken ze op de relevante plaats in het x-y–vlak en schets de vier speciale krommen (in de buurt van (x_0,y_0)).

e. Twee van de vier oplossingskrommen uit het onderdeel (d) spelen mathematisch en biologisch een belangrijke rol. Beschrijf de wiskundige rol en geef een biologische interpretatie. Schets in het relevante deel van het x-y–vlak een paar karakteristieke krommen waar duidelijk de funktie van die twee belangrijke krommen uit blijkt.

Oefening 6.14 [Tentamensom maart 1995]

We bestuderen de verspreiding van een besmettelijke ziekte en kiezen voor een kontinu model. $x(t)$ is het deel van de bevolking dat op tijdstip t dat vatbaar is voor de ziekte maar niet ziek is. Mensen die ziek zijn zijn ook (vrijwel) onmiddellijk besmettelijk. Op tijdstip t is het deel dat ziek is $y(t)$. Mensen die genezen zijn immuun en niet meer besmettelijk. Het aantal mensen dat per tijdseenheid geneest of sterft is een vast deel van het aantal zieke mensen. Verder nemen we aan dat het aantal mensen in de bestudeerde periode niet toeneemt en maar alleen afneemt doordat mensen bezwijken aan de ziekte.

Voor zekere positieve konstanten ρ en γ kan het verspreiding van de ziekte beschreven worden door

$$\begin{cases} x' = -\rho xy, \\ y' = \rho xy - \gamma y. \end{cases} \tag{6.12}$$

a. Beschrijf van welke veronderstellingen uitgegaan is bij het opstellen van beide termen "ρxy". Hoe interpreteer je de term γy.

b. Geef in het biologisch relevante deel van het x-y-vlak aan waar $x' = 0$, respektievelijk $y' = 0$. Geef het tekenverloop van x' en y' aan (middels pijltjes) in het diagram dat zo ontstaat. Bepaal de evenwichtspunten. Zijn die stabiel? Alleen voor $x(0) > \gamma/\rho$ en $y(0) > 0$ ontstaat er een epidemie. Waarom?

c. Bepaal een funktie $F(x, y)$ waarvoor geldt

$$\frac{\partial F}{\partial y} = 1 \quad \text{en} \quad \frac{\partial F}{\partial x} = 1 - \frac{\gamma}{\rho}\frac{1}{x}.$$

Hoe kan je F gebruiken om oplossingskrommen te vinden?
Wat voor effect heeft de term ρxy op de oplossing?

d. Stel nu $\gamma/\rho = 0.75$. Gebruik de funktie F uit (c) om, in het biologisch relevante deel van het x-y-vlak, de kromme te schetsen van de oplossing met $x(0) = 0.9999$ en $y(0) = 0.0001$. Geef duidelijk aan hoe de kromme doorlopen wordt. Hoe groot is het deel van dat bevolking dat op het hoogtepunt van de epidemie ziek is ($\ln 0.75 \approx -.288$)?

Oefening 6.15 Een tweetal chemische stoffen A en B wordt in een oplossing op vaste concentratie a respectievelijk b gehouden. Stof A vervalt met zekere snelheid in een nieuwe stof X. Deze reageert samen met B en heeft als reaktieprodukten de stoffen Y en D. Stof D is stabiel, maar X werkt als katalysator op Y die tot X vervalt. Tenslotte gaat X zelf over in de stabiele stof E. De reaktieschemas zien er als volgt uit,

$$\begin{aligned} A &\rightarrow X \\ B + X &\rightarrow Y + D \\ 2X + Y &\rightarrow 3X \\ X &\rightarrow E \end{aligned}$$

Alle reaktiecoefficienten worden 1 verondersteld.

a. Toon aan dat voor de concentraties x, y van X resp. Y geldt:

$$\begin{cases} x' = a - b\,x + x^2\,y - x \\ y' = b\,x - x^2\,y \end{cases} \tag{6.13}$$

b. Bepaal een evenwichtsoplossing $(\bar{x}, \bar{y})$ met $\bar{x}, \bar{y} > 0$.

c. Stel $a = 1$. Het evenwichtspunt wordt nu door $(1, b)$ gegeven. Ga de stabiliteit van dit evenwicht na voor $0 < b < 4$.

Oefening 6.16 Een voorwerp beweegt onder invloed van de zwaartekracht en zonder wrijving langs de kromme K gegeven door de vergelijking $h(x) = x^3 - 3x$. Hierbij is $h(x)$ de hoogte en x de horizontale verplaatsing. De snelheid van het voorwerp noemen we v. (Met het zelfde teken als $\frac{dx}{dt}$). De versnelling van de zwaartekracht is $g = 10$ en de massa van het voorwerp is m. De zwaartekracht is natuurlijk verticaal gericht en is van de grootte mg, maar we zijn nu ook geïnteresseerd in de component F_{langs} van de zwaartekracht langs de bewegingsrichting. (Er is nog een andere kracht, dwars op de bewegingsrichting, die het voorwerp op de kromme houdt, maar deze is voor ons niet van belang.)

a. Maak een schets van de kromme K, bepaal in het bijzonder de lokale maxima en minima.

b. Zijn er periodieke oplossingen van de bewegingsvergelijking? Zijn er ook niet-periodieke? (Nog niet rekenen s.v.p.) Geef toelichting.

c. Als ”fasevlak”nemen we het x, v vlak. Maak een schets die een goed beeld van de integraalkrommen geeft en geef aan waar de stationaire punten liggen. (Nog steeds nauwelijks rekenen s.v.p.)

d. Bewijs dat het deeltje zich in het fasevlak beweegt volgens de vergelijkingen

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} x \\ v \end{bmatrix} = \frac{1}{\sqrt{1 + h'(x)^2}} \begin{bmatrix} v \\ -gh'(x) \end{bmatrix}.$$

(Hint: Zij α de hoek die de raaklijn aan K boven het punt x met de X-as maakt. We weten dat $\tan\alpha = h'(x)$. Interpreteer $1/\sqrt{1 + (h')^2}$ en $h'/\sqrt{1 + (h')^2}$)

e. Bereken de stationaire punten van bovenstaand stelsel en hun aard. Controleer of dit met de schets uit onderdeel (c) overeenstemt.

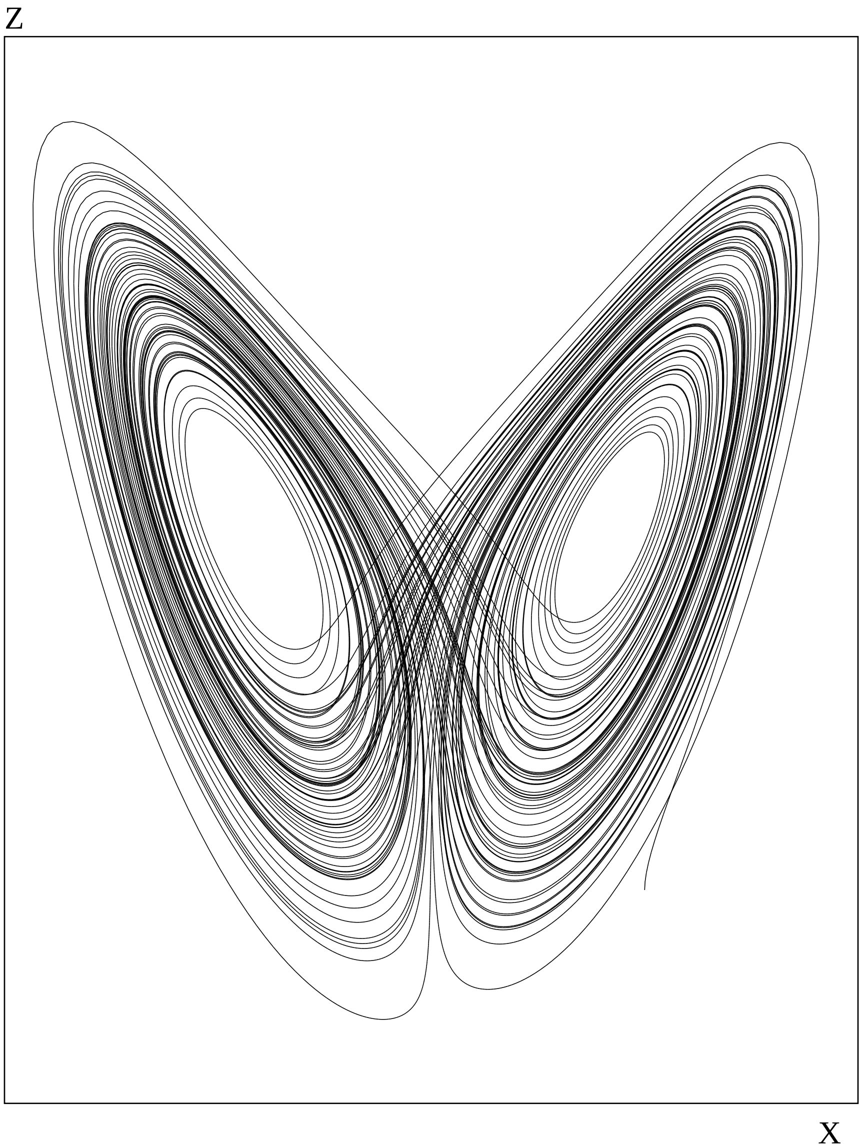
We verwachten een wet van behoud van energie die zegt dat

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{2}mv^2 + mgh(x)\right) = 0.$$

f. Leid die wet af uit de bewegingsvergelijkingen.

6.6 Niet-autonome stelsels en stelsels in $\mathbb{R}^3$

Zoals eerder opgemerkt kunnen de integraalkrommen van een autonoom stelsel elkaar niet snijden. In dimensie twee heeft dit tot gevolg dat de familie van integraalkrommen van (6.4) redelijk geordende plaatjes geeft. We hebben gezien dat in $\mathbb{R}^2$ alleen stationaire punten en limiet cykels als uiteindelijke limietbanen kunnen optreden. Gaan we echter over naar dimensie drie en hoger dan is dit geenszins het geval meer. Er treden nieuwe typen limietbanen (attractoren) op, waarvan de bekendste de *quasi-periodieke banen* zijn. Deze ontstaan als er bij een proces meerdere periodieke bewegingen in het spel zijn waarvan de frequenties een irrationale verhouding hebben. Een ander type, dat pas in de zestiger jaren werd ontdekt, is dat van de *vreemde attractor* of *chaotische attractor* welke veelal een fractalachtige struktuur hebben.



FIGUUR 11. *Een projektie van een baan van het model van Lorenz.*

De vlinder van Lorenz. Beschouw het volgende stelsel d.v.'s

$$\begin{cases} x' = 10(y - x) \\ y' = x(28 - z) - y \\ z' = xy - (8/3)z \end{cases}$$

Dit stelsel werd rond 1962 bestudeerd door Lorenz naar aanleiding van een model dat een beschrijving gaf van opstijgende lucht ten gevolge van de opwarming van de aarde. Nemen we als startpunt bijvoorbeeld $(10, 10, 10)$ dan krijgen we een baan te zien waarvan een projektie er als in FIG. 11 ziet. De baan beschrijft een deel van de ruimte dat de vorm heeft van een vlinder met vleugels van positieve dikte. Het punt beweegt zich enkele malen om één vleugel en wipt vervolgens over op de andere, en zo voort. We krijgen zo een rij getallen die het aantal malen op de linkervleugel, op de rechtervleugel, op de linkervleugel, en zo voort aangeeft. Het blijkt dat deze rij getallen zeer sterk van het beginpunt afhangt. Nemen we een startpunt dat dicht bij $(10, 10, 10)$ ligt, dan zal de rij verblijfstijden in de vleugels in het begin lijken op dat bij startpunt $(10, 10, 10)$. We zien na vrij korte tijd echter al dat de rijen van elkaar af gaan wijken en wel zo sterk, dat men helemaal niet meer vermoedt dat de banen ooit nabij gelegen startpunten hebben gehad. Om het gedrag van dit systeem dus op langere termijn te kunnen voorspellen moet men het startpunt tot op zeer hoge precisie kennen. In de praktijk is dit lang niet altijd mogelijk. De toestand van het weer over een groot gebied is slechts met een beperkte nauwkeurigheid bekend, we moeten het doen met metingen door een netwerk van meetstations, en wat daartussen gebeurt kunnen we alleen maar gokken. In weersmodellen doet deze onzekerheid zich al snel gelden en weersvoorspellingen voor langere termijnen dan een week lijken, zelfs indien de modellen zelf correct zijn, op deze manier fundamenteel uitgesloten.

Wat wel opvalt bij het plaatje is, dat hoewel de baan sterk van het beginpunt afhangt en dus snel onbetrouwbaar wordt, ze toch netjes binnen het gebied van de twee vleugels blijft. Dit wordt veroorzaakt door het feit dat de baan wordt aangetrokken door een limietbaan die aan het systeem ten grondslag ligt. Deze limietbaan is een voorbeeld van een chaotische attractor met een zeer rijke fractalachtige struktuur. Het plaatje is te vinden in alle populaire boekjes over chaostheorie. De term 'vlinder van Lorenz' refereert naar de wijsheid dat het gedrag van een vlinder in China gevolgen heeft voor gebeurtenissen in de rest van de wereld.

Het blijkt dat niet-autonome stelsels in dimensie twee ook vaak een chaotisch gedrag kunnen vertonen. De reden hiervoor is dat dergelijke stelsels omgeschreven kunnen worden als autonoom stelsel in $\mathbb{R}^3$. Beschouw het stelsel

$$\begin{cases} x' = f(t, x, y) \\ y' = g(t, x, y). \end{cases}$$

Door $z = t$ te stellen kan dit stelsel omgevormd worden tot

$$\begin{cases} x' = f(z, x, y) \\ y' = g(z, x, y) \\ z' = 1, \end{cases}$$

een autonoom stelsel.

6.7 Computer-opdrachten

Aanwijzingen voor het gebruik van de computer

Start het programma `des` in het 'Natuurwetenschappen' menu. Je komt in het hoofdmenu van `des`. Opties in dit en alle volgende menu's kunnen geselecteerd worden door met de pijltjestoetsen ($\leftarrow$ en $\rightarrow$) door het menu te lopen en op de gewenste optie `enter` te geven. Met de optie `edit problem` kun je een stelsel vergelijkingen invoeren en tevens het startpunt en de stapgrootte. De vergelijkingen mogen parameters (van één letter) bevatten die in de optie `edit macros` waarden kunnen worden gegeven. In de macros kunnen ook functies worden ingevoerd. Doe dit zo veel mogelijk. Bijvoorbeeld als een functie $f(x)$ meerdere keren voorkomt, definieer f dan in de macro en gebruik die verder in het programma. Assen kunnen worden gespecificeerd onder `edit axes`. Je kunt willekeurige functies en/of variabelen tegen elkaar uitzetten. Om vergelijkingen, parameters, etc. te veranderen moeten deze eerst geselecteerd worden. Veranderingen moeten worden afgesloten met `enter`. De berekeningen worden gestart door in het hoofdmenu `solve problem` te selecteren en kunnen worden gestopt met `enter`.

Computer Opgave 6.I [Een roofdier-prooidier model]

We bekijken populaties van twee soorten dieren P en R die in een gesloten systeem naast elkaar voorkomen en waarbij soort R jaagt op soort P. We geven met $P(t)$ en $R(t)$ de populatiegrootte van de beide soorten op tijdstip t aan.

We veronderstellen dat de groei van de prooidieren bij afwezigheid van de roofdieren logistisch verloopt, d.w.z. volgens

$$P'(t) = r\left[1 - \frac{P(t)}{a}\right]P(t).$$

Het aantal prooidieren dat geconsumeerd wordt door één roofdier noemen we C. Het zal groter zijn als de populatie prooidieren groter is, maar het zal niet onbeperkt toenemen. Er zijn verschillende formules om een dergelijk eetgedrag te beschrijven. Wij kiezen hier $C = \frac{vP}{c+P}$ met c en v positieve constantes. De vergelijking voor $P'(t)$ wordt nu

$$P'(t) = r\left[1 - \frac{P(t)}{a}\right]P(t) - \frac{vP(t)}{c + P(t)}R(t)$$

met r, c, v en a positieve constantes.

We veronderstellen verder dat de populatie van de roofdieren bij afwezigheid van de prooidieren exponentieel afneemt, d.w.z.

$$R'(t) = -bR(t).$$

De consumptie van prooidieren wordt echter (met efficiëntie e) gebruikt voor de groei van de populatie volgens $R'(t) = (-b + eC)R(t)$. De vergelijking voor $R'(t)$ wordt dus

$$R'(t) = \left[-b + \frac{evP(t)}{c + P(t)}\right]R(t)$$

met b en e positieve constantes.

Ter vereenvoudiging kiezen we nu $c = 1$, $b = 1$, $r = 5$, $v = 10$ en $e = \frac{1}{5}$ zodat we krijgen

$$\begin{cases} P'(t) = 5\left[1 - \dfrac{P(t)}{a} - \dfrac{2R(t)}{1 + P(t)}\right]P(t) \\ R'(t) = \left[-1 + \frac{2P(t)}{1+P(t)}\right]R(t). \end{cases} \tag{6.14}$$

a. Bepaal de evenwichtspunten van dit systeem.

De ligging van en het aantal evenwichtspunten is afhankelijk van de waarde van a. Als we ook op de types van de evenwichtspunten letten, treden er 4 verschillende situaties op, namelijk voor

$$0 < a < 1, \quad 1 < a < \tfrac{5}{3}, \quad \tfrac{5}{3} < a < 3 \quad \text{en voor} \quad 3 < a.$$

b. Geef hier een wiskunde verklaring voor. Beschrijf ook de aard van de verschillende evenwichten. (Hint: Zoek het antwoord in termen van eigenschappen van de eigenwaarden. Zijn λ's reëel of complex? Is Re$(\lambda) < 0$ Etc.) (*)

We onderzoeken de verschillen met behulp van het programma `des`. Na het starten van dit programma voeren we het stelsel d.v.'s in en kunnen a kiezen onder `'edit macros'`. Stel om te beginnen de assen in op $0 < P < 2$ en $0 < R < 1$.

c. Laat zien dat de positieve X-as en Y-as integraalkrommen zijn (d.w.z. er zijn oplossingen die over deze assen lopen). (*)

d. Kies verschillende waarden van a, minstens één in elk van de bovengenoemde intervallen. Maak voor elk van deze waarden een schets van de integraalkrommen door verschillende startpunten te kiezen en het stelsel op te lossen. Geef ook een schatting van de positie van de stationaire punten en hun aard. Als de aard niet duidelijk is kun je desnoods de assen zó instellen dat alleen het kleine gebied rond het stationaire punt zichtbaar is.

e. Kan je aangeven voor welke van de vier boven aangegeven gebieden voor a er sprake is van een stabiel spiraalpunt?

f. Bepaal in alle bovengenoemde gevallen het gedrag van de oplossing voor $t \to \infty$.

Computer Opgave 6.II [Een model voor een immuun reactie]

Antigenen ("ziektemakers") worden in ons lichaam gebonden door een antilichaam en vervolgens onschadelijk gemaakt. We bestuderen hieronder een model voor de verhouding tussen deze twee.

$x(t)$ is de koncentratie (gemiddeld aantal) antigenen in het lichaam op tijdstip t, inklusief de gebonden antigenen waarvan de koncentratie $\bar{x}(t)$ is. De toename ("sterfte & geboorte") (per tijdseenheid) van de koncentratie antigenen is evenredig met (die tijdseenheid en) deze koncentratie ($x' = \lambda x$). Per tijdseenheid wordt een deel van de gebonden antigenen onschadelijk gemaakt. De verandering in de koncentratie antigenen kan dus, voor zekere evenredigheidskonstante λ en α, beschreven worden door de vergelijking

$$x' = \lambda\, x - \alpha\, \bar{x}. \tag{6.15}$$

We nemen hier verder $\lambda = 1$ en $\alpha = 2$.

$y(t)$ is de koncentratie antilichamen op tijdstip t, weer inklusief gebonden materiaal. $\bar{y}(t)$ is de koncentratie van het gebonden deel. De verandering in $y(t)$ wordt bepaald door twee faktoren: "natuurlijke sterfte"($y' = -\mu\, y$) en produktie. De produktie wordt geprikkeld door de aanwezigheid van gebonden antilichamen: de produktie is evenredig met $\bar{y}(t)$. De evenredigheidsfaktor neemt af bij hogere koncentratie antilichamen. De verandering in $y(t)$ wordt dus beschreven door

$$y' = -\mu\, y + \beta\,(1 - \vartheta\, y)\,\bar{y}. \tag{6.16}$$

Voor de evenredigheidskonstante μ en β nemen we verder $\mu = 1$ en $\beta = 4$.

We nemen aan dat de hoeveelheid gebonden materiaal evenredig is met de "kans" dat ongebonden antigeen ongebonden antilichamen ontmoeten: het evenwicht tussen gebonden en ongebonden materiaal wordt beschreven door

$$\bar{x} = \kappa\,(x - \bar{x})\,(y - \bar{y}) \tag{6.17}$$

We nemen tenslotte aan (niet zo realistisch) dat één antilichaam precies één antigeen bindt:

$$\bar{y} = \bar{x}. \tag{6.18}$$

Omdat $\bar{x}^2 \ll \bar{x}$ mogen we $\bar{x}^2$ verwaarlozen. Dan impliceren (6.17) en (6.18) dat

$$\bar{x} = \frac{\kappa\, x\, y}{1 + \kappa\,(x + y)}. \tag{6.19}$$

Invullen van (6.18) en (6.19) in (6.15)-(6.16) levert

$$\begin{cases} x' = \dfrac{1 + \kappa x - \kappa\, y}{1 + \kappa\,(x + y)} x \\[2ex] y' = \dfrac{-1 + 3\,\kappa x - \kappa\, y - 4\,\vartheta\,\kappa\, x\, y}{1 + \kappa\,(x + y)} y \end{cases} \tag{6.20}$$

Beschouw ook het stelsel

$$\begin{cases} x' = (1 + \kappa x - \kappa\, y)\, x \\ y' = (-1 + 3\,\kappa\, x - \kappa y - 4\,\vartheta\,\kappa\, x\, y)\, y \end{cases} \tag{6.21}$$

a. Wat is het verschil tussen de fase-banen van stelsel (6.20) en (6.21)?

We bestuderen verder (6.21).

b. Teken de lijnen $x' = 0$ en $y' = 0$.
(We zouden graag de krommen $1/\kappa + x$ en $(-1 + 3\,\kappa\, x)/(\kappa + 4\,\vartheta\,\kappa\, x)$ willen tekenen. Maar de oplossing x van de differentiaalvergelijking loopt niet van 0 tot ???. Het programma `des` is niet gemaakt om grafieken van funkties te tekenen, maar met een truukje en de `edit macros` optie kan dat toch.
We tekenen de $P = 1/\kappa + t$ en $Q = (-1 + 3\,\kappa\, t)/(\kappa + 4\,\vartheta\,\kappa\, t)$ en zetten t botweg langs de horizontale as, de x-as. Voer deze krommen in in “`edit macros`”; definieer de assen: horizontaal t, verticaal p; `solve problem`; daarna vertikaal q; `solve problem`.

Neem $\kappa = 1.3$ en $\vartheta = 0.06$.

c. Spoor de evenwichtspunten op en klassificeer ze. Bereken een paar oplossingen.

d. Wat is het effect van $\vartheta = 0.06$ vervangen door $\vartheta = 0$?

e. Een van de separatrices voor $\vartheta = 0.06$ is van (levens)belang. Welke is dat? Teken die separatrix.
(Hint: Start in de buurt van het zadelpunt en laat de tijd teruglopen)

f. Geef een biologische interpretatie van de resultaten.

6 Numeriek oplossen van differentiaal vergelijkingen

6.1 Inleiding

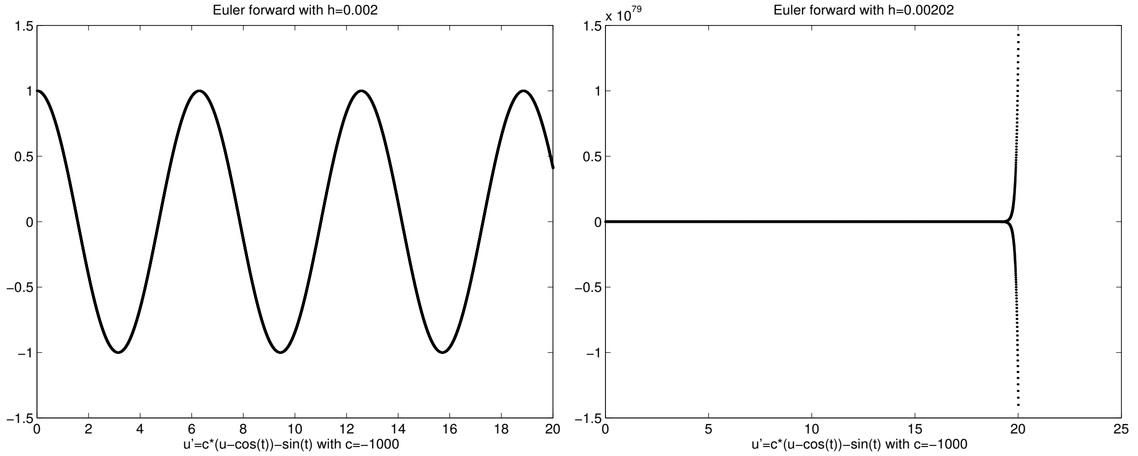
De meeste differentiaalvergelijkingen zijn niet *analytisch* oplosbaar, d.w.z. voor de oplossingen kan geen funktievoorschrift gegeven worden. Toch kunnen we ook voor die differentiaalvergelijkingen vaak veel te weten komen over de oplossing. Een aantal wiskundige technieken die ons daarbij helpen hebben we in de voorgaande hoofdstukken besproken. De informatie die we krijgen is globaal: we weten bijvoorbeeld of de oplossing zich op den duur stabilizeert, en of dat afhangt van de begintoestand, of we weten dat de oplossing periodiek varieert, etc.. Voor een groot aantal toepassingen zijn we echter geïnteresseerd in een *numerieke oplossing*, d.w.z. in preciese numerieke waarden van de oplossing in specifieke tijdstippen (niet alleen voor $t \to \infty$). Het is leuk om te weten dat het aantal vluchten op Schiphol jaarlijks toeneemt, exponentieel wellicht, maar worden het nu 30 of 40 miljoen in het jaar 2000? Rekentechnieken die ons daarbij kunnen helpen, en bedoeld zijn voor computer gebruik, bespreken we in dit hoofdstuk.

Analytische technieken hebben hun beperkingen: ze geven slechts globale informatie en soms is die zelfs niet volledig. Numerieke technieken geven preciese waarde voor specifieke parameters en tijdstippen maar geven daarentegen weer niet het globale beeld. In de praktijk zal je beide type technieken moeten gebruiken zelfs als je alleen in getalwaarden geïnteresseerd bent: numerieke technieken vertellen je bijvoorbeeld niet hoe gevoelig de uitkomsten afhangen van fouten in de ingangsgegevens, fouten in het model, etc.. Bovendien introduceren numerieke technieken zelf fouten. De computer kan bijvoorbeeld een afgeleide slechts benaderend uitrekenen (wat is “voor h oneindig klein”?) en werkt met een groot (16) maar toch beperkt aantal cijfers. Wat het effect is van met name deze fouten op de nauwkeurigheid van de numerieke uitkomsten, staat centraal in onze diskussie in dit hoofdstuk.

De differentiaalvergelijking zelf kan gevoelig reageren op verstoringen: dat zal bijvoorbeeld het geval zijn voor een oplossing die gestart is vlakbij of op een separatrix, of als de parameterwaarden in ’n chaotisch gebied liggen. Zulk soort differentiaalvergelijkingsproblemen zijn *instabiel* en ieder foutje gemaakt met wat voor numerieke methode dan ook zal een dramatisch effect hebben op de uitkomst. Maar als we niet zorgvuldig omspringen met onze numerieke oplosmethode zal zo’n ongewenste foutgroei ook kunnen optreden bij stabiele problemen: *de methode is dan instabiel*. FIG. 12 illustreert wat we met “zorgvuldig” bedoelen.

Er is helaas niet een numerieke methode die stabiel is (die geen of zeer beperkte foutgroei met zich meebrengt) voor ieder stabiel probleem. We kunnen dus niet zomaar een standaard methode uit de kast halen en blindelings toepassen op ons probleem. We moeten weten waar we op moeten letten. We moeten weten welke numerieke methoden geschikt zijn voor wat voor soort problemen. En als we niet zeker zijn over de betrouwbaarheid van de gekozen numeriek methode voor ons specifiek probleem, dan zouden we toch graag willen weten hoe we aan de numerieke uitkomsten kunnen zien hoe betrouwbaar (hoe nauwkeurig) ze zijn. Bij geschiktheid spelen ook “economische” motieven een rol: naarmate een methode meer problemen aankan is hij ook duurder, d.w.z. heeft hij meer rekentijd nodig. We zouden graag willen weten hoe we tot een weloverwogen keuze kunnen komen rekening houdend met kosten en nauwkeurigheid.

In dit hoofdstuk bespreken we geen “sofisticated methods”; we beperken ons tot een paar populaire methoden. We besteden meer aandacht aan de zojuist geschetste vragen.



FIGUUR 12. *De numerieke uitkomst van* $u' = c\,(u - \cos(t)) - \sin(t)$*,* $u(0) = 1$*, met* $c = -1000$ *wordt met de voorwaartse Euler methode berekend in de tijdstippen* $0, h, 2h, 3h, \ldots$ *De exacte, analytische, uitkomst is* $u = \cos(x)$*. Het linkse figuur laat de resultaten zien voor* $h = 0.002$*, het rechtse voor een ietsje grotere* $h = 0.00202$*. Blijkbaar moet h zorgvuldig gekozen worden. Merk op dat met* $h = 0.002$ *en met* $h = 0.00202$ *op het getoonde traject ongeveer 10000 funktiewaarden berekend zijn.*

6.2 De voorwaartse Euler regel

Hoe berekenen we numeriek de oplossing van

$$u'(t) = F(t, u(t)), \qquad u(t_0) = a \tag{6.1}$$

of van

$$\begin{cases} x' = f(x, y) \\ y' = g(x, y) \end{cases} \quad \text{met} \quad \begin{cases} x(t_0) = x_0 \\ y(t_0) = y_0 ? \end{cases} \tag{6.2}$$

Hierbij zijn x_0 en y_0 bekende waarden op het begintijdstip t_0 en is a een bekende konstante (parameter).

De computer kan niet differentiëren. Het ligt voor de hand om, voor een geschikte kleine h, de afgeleide te vervangen door een benadering:

$$u'(t) \approx \frac{u(t+h) - u(t)}{h}. \tag{6.3}$$

Voor $h \to 0$ is dit immers precies de definitie van $u'(t)$. Negeren we eventjes de fout die deze benadering met zich meebrengt dan gaat (6.1), na vermenigvuldiging met h en herrangschikking van de termen, over in

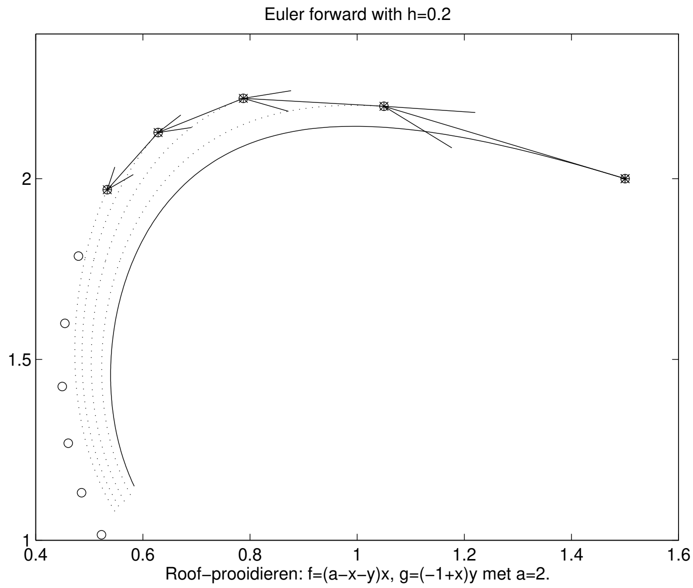
$$u(t+h) \approx u(t) + hF(t, u(t)).$$

Deze formule ziet er heel bruikbaar uit. We kennen immers u op tijdstip t_0. We kunnen dus gemakkelijk u in $t_1 = t_0 + h$ benaderend berekenen door

$$u(t_1) \approx u_1 = u(t_0) + hF(t_0, u(t_0)) = u_0 + hF(t_0, u_0). \tag{6.4}$$

Bedenk hierbij dat F een bekend funktievoorschrift is: in (6.1) is $u(t)$ niet bekend en kunnen we daarom $F(t, u(t))$ niet uitrekenen, maar in (6.4) kennen we u_0. Dit biedt weer de mogelijkheid om u in het tijdstip $t_2 = t_1 + h$ benaderend te berekenen: $u(t_2) \approx u(t_1) + hF(t_1, u(t_1)) \approx u_1 + hF(t_1, u_1)$ en $F(t_1, u_1)$ kan inmiddels berekend worden. Dit kunnen we herhalen:

$$t_{n+1} = t_n + h, \quad u(t_{n+1}) \approx u_{n+1} = u_n + hF(t_n, u_n) \quad \text{voor} \quad n = 0, 1, \ldots : \tag{6.5}$$



FIGUUR 13. *Een meetkundige interpretatie van voorwaarts Euler. De ononderbroken lijn is de exacte oplossing van het roof-prooidieren model $x' = (a-x-y)x$, $y' = (-1+x)y$ met $a = 2$, voor t tussen $t_0 = 0$ en $T = 2$, met als beginwaarde $x(0) = 1.5$ en $y(0) = 2$. De circels $\circ$ geven de voorwaartse Euler oplossing aan verkregen met stapgrootte $h = 0.2$. De pijlen zijn de vectoren $h(f_n, g_n)$ getekend vanuit de punten (x_n, y_n) voor $n = 0, 1, 2, 3$. Duidelijk is te zien dat voorwaarts Euler in iedere stap "overstapt" naar een oplossing met andere beginwaarde, nl. het gevonden punt (x_n, y_n): de gestippelde (...) krommes zijn de exacte oplossing met beginwaarde (x_n, y_n).*

u_n is dus de *berekende*, numerieke waarde voor u in het tijdstip t_n; $u(t_n)$ is de exacte waarde voor u in datzelfde tijdstip. Omdat er fouten gemaakt worden zullen die waarden verschillend zijn. Voor het gemak schrijft men gewoonlijk

$$F_n \quad \text{i.p.v.} \quad F(t_n, u_n).$$

Deze notatie houdt de formules overzichtelijk maar verbergt ook belangrijke informatie: F_n, maar ook t_n en u_n, hangt van h af. Hou in de gaten dat bij een andere **stapgrootte** h deze grootheden een andere waarde hebben!

Exact dezelfde aanpak kunnen we volgen voor andere meer-dimensionale problemen. Voor bijvoorbeeld (6.2) krijgen we

$$\begin{cases} x(t_{n+1}) \approx x_{n+1} = x_n + h\, f(x_n, y_n) = x_n + h\, f_n \\ y(t_{n+1}) \approx y_{n+1} = y_n + h\, g(x_n, y_n) = y_n + h\, g_n. \end{cases} \tag{6.6}$$

Deze methode om, zeg, u in de tijdstippen $t_n = t_0 + nh$ te benaderen door u_n die berekend is door de differentievergelijking (6.5) recursief te doorlopen, heet de **voorwaartse Euler methode** ("forward Euler"). Voor, bijvboorbeeld, het 2-dimensionale geval heeft de methode de volgende meetkundige interpretatie: (x_{n+1}, y_{n+1}) ontstaat door vanuit (x_n, y_n) in de richting (f_n, g_n) te lopen over een afstand $h\,(f_n, g_n)$ (zie ook FIG. 13).

Voorbeeld. We voeren de berekening uit voor het roof-prooidieren model

$$\begin{cases} x' = f(x, y) = (a - x - y)x, \\ y' = g(x, y) = (-1 + x)y \end{cases} \qquad \text{voor} \quad t \in [0, 2]$$

met $a = 2$ en startwaarden $x(0) = 1.5$ en $y(0) = 2$. We nemen $h = 0.2$ als stapgrootte. We vinden dus een numerieke oplossingen in de tijdstippen $t_0 = 0$, $t_1 = 0.2$, $t_2 = 0.4$,..., $t_{10} = 2$.

Uiteraard is $x_0 = 1.5$ en $y_0 = 2$.

$$x_1 = x_0 + hf(x_0, y_0) = 1.5 + 0.2\,(2 - 1.5 - 2)\,1.5 = 1.05,$$
$$y_1 = y_0 + h\,g(x_0, y_0) = 2.0 + 0.2\,(-1 + 1.5)\,2 = 2.2,$$
$$x_2 = x_1 + hf(x_1, y_1) = 1.05 + 0.2\,(2 - 1.05 - 2.2)\,1.5 = 0.7875,$$
$$y_2 = y_1 + h\,g(x_1, y_1) = 2.2 + 0.2\,(-1 + 1.05)\,2 = 2.222, \text{ etc..}$$

In onze verdere uitleg zullen we ons voornamelijk koncentreren op het 1-dimensionale probleem (6.1), maar vrijwel alle uitspraken gelden ook voor meer dimensionale differentiaal-vergelijkingen.

6.3 De nauwkeurigheid van voorwaarts Euler

Voorwaarts Euler biedt ons een efficiënte methode om een differentiaalvergelijking numeriek op te lossen. De aanpak introduceert echter fouten. Het is van belang te weten hoe groot die fouten ongeveer zijn. Ze zullen afhangen van de stapgrootte h. Als h klein genoeg zullen we er vast geen last van hebben. Maar hoe groot is “klein genoeg”? Voor de bereking zullen we een concrete waarde moeten kiezen. Bovendien is een al te kleine h ongewenst. Immers het aantal F-waarden dat berekend moet worden om de oplossing over een tijdsinterval ter lengte T te kunnen volgen is ongeveer T/h. De kosten zijn dus omgekeerd evenredig met h.

Op ieder van de tijdstippen t_n introduceren we een fout door $u'(t_n)$ te vervangen door een differentie quotiënt (zie (6.3)). Bovendien vervangen we de exacte functiewaarden $u(t_n)$ door de berekende u_n. Deze laatste substitutie introduceert ook weer de oude fouten in de nieuwe benadering u_{n+1}. We maken in iedere stap nieuwe fouten en we baseren onze berekening op waarden die al fout zijn. Wat is het effect hiervan? FIG. 13 geeft een andere manier om tegen het opstapelen van fouten aan te kijken. In iedere voorwaartse Euler stap, stap je in feite over op een nieuwe naburige oplossing.

6.3.1 Lokale fouten

Over de nieuwe fout in elke stap, de zgn. lokale fout, kunnen we gemakkelijk iets zeggen. Omdat, volgens de 2-de orde Taylor reeks

$$u(t + h) \approx u(t) + hu'(t) + \tfrac{1}{2}\,h^2 u''(t),$$

is[1]

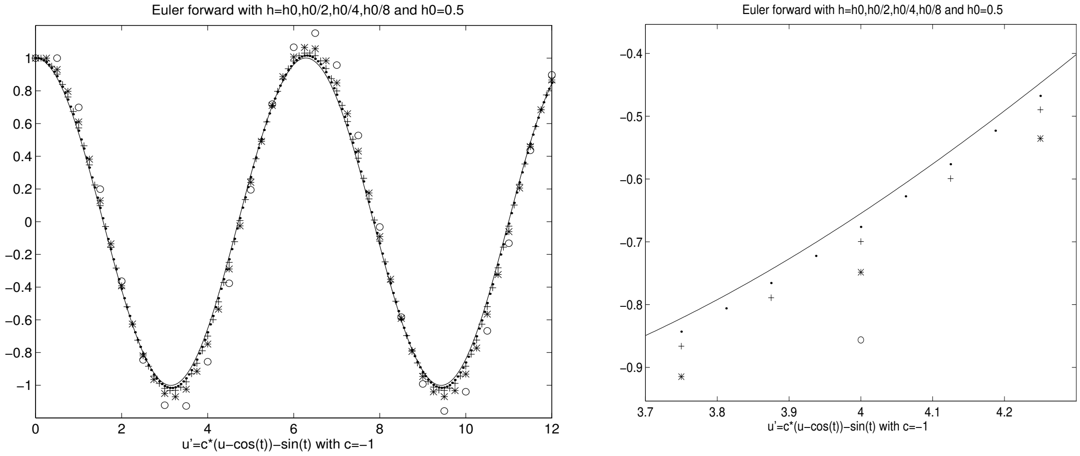
$$\frac{u(t + h) - u(t)}{h} = u'(t) + \delta(t) \quad \text{waarbij} \quad \delta(t) \approx \tfrac{1}{2}\,hu''(t). \tag{6.7}$$

$\delta(t)$ is de **lokale diskretisatie fout** op tijdstip t in de Euler stap. δ is de fout die onstaat door het *diskretiseren* van de afgeleide. Dus

$$u(t_{n+1}) = u(t_n) + h\,F(t_n, u(t_n)) + h\delta(t_n). \tag{6.8}$$

Het numeriek rekenschema (voorwaarts Euler) ontstaat door het wegpoetsen van $h\delta(t_n)$. $h\delta(t_n)$ is dus nieuwe fout in het rekenschema, de **lokale fout**. Deze is dus $h\times$(de lokale diskretisatie fout) (met excuses voor het misleidende woordgebruik en het subtiele verschil tussen “lokale fout in het rekenschema” en “lokale diskretisatie fout”).

[1] Formule (6.7) bevat, in de uitdrukking voor δ, nog steeds een $\approx$ teken, maar de benadering in deze formule is natuurlijk veel beter dan die in (6.3).



FIGUUR 14. *De exacte en numerieke oplossingen van probleem* $u' = c\,(u - \cos(t)) - \sin(t)$, $u(0) = 1$, *met* $c = -1$. *De numerieke oplossing is verkregen met voorwaarts Euler met achtereenvolgens stapgrootte* $h = 0.5$ *(de '*∘*' geven de gevonden waarden),* $h = 0.25$ *(*$* * *$*),* $h = 0.125$ *(*$+ + +$*),* $h = 0.0625$ *(*$\cdots$*). De exacte oplossing* $u(t) = \cos(t)$ *is de ononderbroken lijn. Het linker figuur laat de oplossing zien van* $t = 0$ *tot* $t = T = 12$. *In het rechter figuur is de oplossing in de buurt van* $t = 4$ *uitvergroot. Duidelijk is de halvering van de fout te zien.*

6.3.2 Globale fouten

De belangrijke vraag is natuurlijk hoe nauwkeurig het numerieke resultaat is, m.a.w. hoe groot het verschil $u_n - u(t_n)$ is tussen de met voorwaarts Euler berekende waarde u_n en de beoogde exacte waarde $u(t_n)$. Dit verschil is ook een fout en noemen we de **globale fout** en noteren we met e_n. Dus

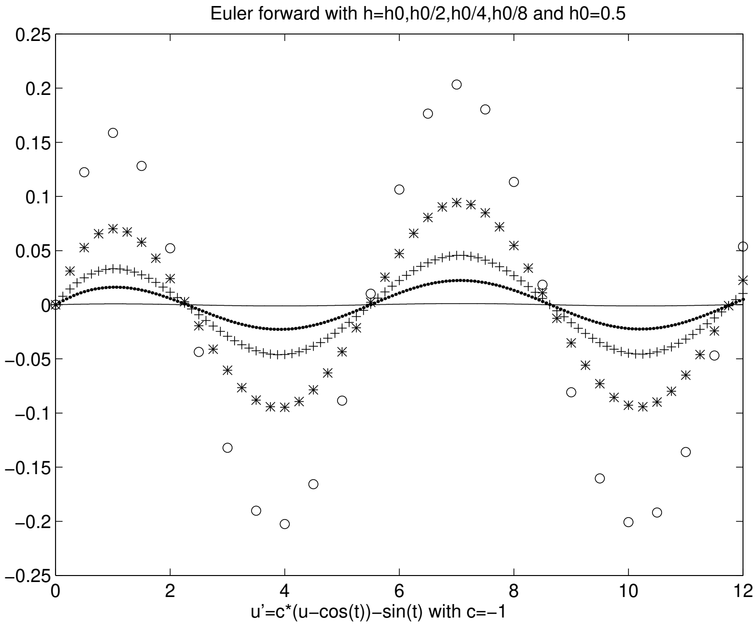
$$e_n = u_n - u(t_n).$$

De globale fout is het totaal effect van alle eerder gemaakte lokale fouten.

Kijken we naar u_1 dan verwachten we dat de globale fout ongeveer evenredig is met h^2, immers $u_1 - u(t_1) = -h\,\delta(t_0) \approx -\frac{1}{2}\,h^2 u''(t_0)$, terwijl (6.7) wellicht suggereert dat de globale fout ongeveer evenredig is met h. We doen een experiment om na te gaan wat er echt gebeurt. Als de globale fout ongeveer evenredig is met h moet die fout ongeveer halveren als we de stapgrootte halveren, terwijl die fout ongeveer 4× zo klein wordt in geval hij ongeveer evenredig is met h^2. FIG. 14 laat duidelijk een halvering van de fout zien bij halvering van h.

In iedere stap introduceren we een nieuwe fout $h\delta(t)$ die evenredig is met h^2. De nieuwe lokale fouten zijn 4× zo klein als we met $h = \frac{1}{2}h_0$ werken in plaats van met $h = h_0$. Toch halveert de globale fout. Hoe valt dat te verklaren? De verklaring is eenvoudig: de lokale fouten $h\delta(t)$ ($h\times$ de lokale diskretisatie fout) worden inderdaad 4× zo klein bij halvering van h_0, echter het *aantal* lokale fouten dat we langs het hele traject maken verdubbelt. We hebben immers eens zo veel stappen nodig om hetzelfde tijdstraject te volgen. Eenmaal gemaakte fouten neem je mee in de volgende stappen. Het gunstigste wat je mag verwachten is een opeenhoping (som) van eerder gemaakte lokale fouten. Met wat meer pech blaast de differentiaalvergelijking lokale fouten op of doet de numerieke methode dat. Blijkbaar was dat in FIG. 14 niet het geval (in FIG. 12 wel!). Over een klein tijdstrajectje, bijvoorbeeld van t tot $t + h_0$, zal de differentiaalvergelijking fouten niet al te zeer opblazen en kunnen we de halvering van de fout precieser verklaren.

Met stapgrootte h_0 maken we over dat tijdstrajectje één nieuwe lokale fout van $\approx \frac{1}{2}h_0^2 u''(t)$, terwijl we met stapgrootte $\frac{1}{2}h_0$ twee nieuwe lokale fouten maken ieder ter grootte van $\approx \frac{1}{2}(\frac{1}{2}h_0)^2 u''(t) = \frac{1}{4}\left(\frac{1}{2}h_0^2 u''(t)\right)$, samen dus $\approx \frac{1}{2}\left(\frac{1}{2}h_0^2 u''(t)\right)$.



FIGUUR 15. *Het figuur laat, voor verschillende h, de funktie hw zien voor het probleem uit* FIG. 14*. Voor h zijn dezelfde waarden gekozen als in* FIG. 14 *(met dezelfde symbolen). Geplot is de numerieke oplossing minus de exacte oplossing.*

De evenredigheid van de globale fout met h kan men precisiëren. De volgende stelling geldt als je (6.1) oplost net voorwaarts Euler.

Stelling 6.1 *Als F glad is ($2\times$ continu differentieerbaar) dan is er een gladde functie w (continu differentieerbaar) zodat voor iedere h, mits voldoende klein, geldt*

$$e_n \approx h\, w(t_n) \quad \textit{voor iedere} \quad n = 0, 1, 2, \ldots.$$

De funktie w hangt niet van h af. In feite is w de oplossing van

$$\begin{cases} w' = Jw + \frac{1}{2}\, u'' \quad \text{met} \quad J(t) = \dfrac{\partial F}{\partial u}(t, u(t)) \\ w(t_0) = 0. \end{cases}$$

(Voorwaarts Euler toepassen op $(hw)' = J(hw) - \frac{1}{2}\, h\, u''$ levert voor de oplossing hw_n precies dezelfde relatie als voor de e_n in (6.22). Bovendien is $e_0 = 0$ en $hw_0 = 0$. Hieruit volgt eenvoudig dat $e_n = hw_n \approx hw(t_n)$ voor alle n.)

"Mits h voldoende klein" is een essentiële voorwaarde: ook het volledige bewijs van de stelling bevat $\approx$-tekens en die veronderstellen dat h voldoende klein is. In FIG. 12 hebben we al gezien dat dat wel eens kan betekenen dat h heel erg klein moet zijn. We komen hier uitgebreid op terug in §6.5.

6.3.3 Foutschatten

Globaal foutschatten. Het feit dat de globale fout $\approx h\times$ een gladde funktie is (zie Stelling 6.1), kan je als volgt gebruiken om die fout te schatten ook al ken je de gladde funktie zelf niet. Het feit dat de funktie bestaat is al voldoende informatie.

Als je u in het tijdstip t benaderd hebt met het voorwaarts Euler resultaat met stapgrootte $h = h_0$ en met het voorwaartse Euler resultaat met stapgrootte $h = 2h_0$ dan geldt (pas Stelling 6.1 toe):

$$u_{h_0}(t) - u(t) \approx \tfrac{1}{2} 2h_0 w(t) \approx \tfrac{1}{2}(u_{2h_0}(t) - u(t)),$$

hierbij stelt $u_h(t)$ het voorwaarts Euler resultaat voor bij stapgrootte h (dus $u_h(t_n) = u_n$) en wordt gekeken in een vast tijdstip t van de vorm $t = t_0 + n\,(2\,h_0)$. Dus

$$\begin{aligned} u_{h_0}(t) - u(t) &\approx \tfrac{1}{2}(u_{2h_0}(t) - u(t)) \\ &= \tfrac{1}{2}(u_{2h_0}(t) - u_{h_0}(t)) + \tfrac{1}{2}(u_{h_0}(t) - u(t)) \end{aligned}$$

en

$$u_{h_0}(t) - u(t) \approx u_{2h_0}(t) - u_{h_0}(t) : \tag{6.9}$$

de globale fout bij stapgrootte h_0 (dat is $u_{h_0}(t) - u(t)$) is blijkbaar ongeveer het verschil tussen de oplossing bij stapgrootte $2h_0$ en de oplossing bij stapgrootte h_0. Dat verschil is berekenbaar.

Bij de foutschatting (6.9) hebben we impliciet aangenomen dat h_0 (of nauwkeuriger $2h_0$) klein genoeg is (zie Stelling 6.1). In §6.5 zullen we zien dat we niet lichtvaardig met die aanname mogen omspringen (zie ook FIG. 12). Door ook nog de voorwaartse Euler oplossing bij stapgrootte $4h_0$ te bepalen kan je jezelf enige zekerheid verschaffen in de juistheid van de veronderstelling dat h_0 voldoende klein is. Immers als de fout min of meer $hw(t)$ is dan is (pas (6.9) twee maal toe, eenmaal met $2h_0$ i.p.v. h_0):

$$\begin{aligned} u_{2h_0}(t) - u_{h_0}(t) &\approx u_{h_0}(t) - u(t) \approx \tfrac{1}{2}\,(u_{2h_0}(t) - u(t)) \\ &\approx \tfrac{1}{2}\,(u_{4h_0}(t) - u_{2h_0}(t)) \,. \end{aligned}$$

Met andere woorden, ook het verschil tussen twee opeenvolgende oplossingen zal je, bij halverende h, zien halveren als de fout inderdaad evenredig is met h.

Hierboven zijn we er impliciet van uitgegaan dat, in het tijdstip t, de funktie w niet nul is. Als $w(t) = 0$ zullen de hogere orde Taylor reeks termen van belang zijn en is $u_h(t) - u(t) \not\approx 0$: het $\approx$ moet telkens in een relatieve zin opgevat worden. De funktie w zal niet al te vaak 0 zijn. De kans dat $w(t)$ precies 0 is zal dan ook klein zijn. Je kunt wat meer zekerheid krijgen dat $w(t) \neq 0$ door te kijken naar $u_{2h_0}(t) - u_{h_0}(t)$ ($\approx hw(t)$) voor een aantal tijdstippen t: de nulpunten van w kun je dan wel ongeveer zien zitten.

Uiteraard hangt aan (deze manier van) foutschatten een prijskaartje: foutschatting (6.9) verlangt dat je niet alleen een oplossing bepaalt met stapgrootte $h = h_0$ maar ook een met stapgrootte $h = 2h_0$, terwijl je voor het opbouwen van enig vertrouwen in de juistheid van de schatting ook nog de oplossing met stapgrootte $h = 4h_0$ moet bepalen. In de praktijk blijft vaak het opbouwen van enig vertrouwen achterwegen en neemt men aan dat $h = h_0$ voldoende klein is als de numerieke oplossing er niet al te gek uit ziet (FIG. 12 illustreert wat een “gekke” numerieke oplossing is).

Lokaal foutschatten. Vaak beperkt men zich tot het schatten van de lokale nieuwe fout $\delta(t_j)$ en neemt men aan dat de globale fout in het tijstip t niet veel groter dan $(t - t_0)\varepsilon$ zal zijn als

$$|\delta(t_j)| \leq \varepsilon \quad \text{voor ieder tijdstip} \quad t_j : \tag{6.10}$$

men neemt aan dat

$$|e_n| \leq C(t - t_0)\varepsilon \quad \text{in} \quad t = t_n,$$

met C niet al te groot (denk aan 'n $C \leq 5$). Het idee hierachter is dat de globale fout e_n in $t = t_n$ niet veel groter zal zijn dan de som van $h\times$ de lokale fouten:

$$\sum_{j<n} |h\delta(t_j)| \leq \sum_{j<n} h\varepsilon = nh\varepsilon = (t_n - t_0)\varepsilon = (t - t_0)\varepsilon.$$

Men negeert het feit dat de differentiaalvergelijking fouten kan opblazen en men neemt aan dat de numerieke methode de lokale fouten net zo behandeld als de differentiaalvergelijking zelf. Deze laatste aanname is inderdaad correct voor de voorwaartse Euler methode (en andere methoden die verderop aan de orde komen) mits h voldoende klein is. Hier heeft "h voldoende klein" exact dezelfde betekenis als in Stelling 6.1.

• Het lokaal foutschatten kan heel goedkoop. Met bijvoorbeeld

$$\widetilde{u}_{n+1} = u_n + h\,F(t_{n+1}, u_{n+1})$$

is

$$\widetilde{u}_{n+1} - u_{n+1} \approx 2h\delta(t_n) \quad \text{en dus} \quad \delta(t_n) \approx \frac{\widetilde{u}_{n+1} - u_{n+1}}{2h}. \tag{6.11}$$

Het uitrekenen van $\widetilde{u}_{n+1}$ kost vrijwel niets: de waarde $h\,F(t_{n+1}, u_{n+1})$ moeten we immers sowieso kennen voor de volgende Euler stap, voor de berekening van u_{n+2}.

• Het lokaal foutschatten biedt de mogelijkheid om *de stapgrootte te besturen*: als bijvoorbeeld, met stapgrootte $h = h_0$ de schatting van $|\delta(t_n)|$ vertelt dat deze grootheid $\varepsilon/5$ is kunnen we in de volgende stap ook wel met $h = 5h_0$ werken. Dat schiet harder op. We moeten daarentegen h verkleinen als (de schatting voor) $|\delta(t_n)|$ significant veel groter is dan ε.

Om te voorkomen dat in iedere stap met een nieuwe stapgrootte gewerkt wordt verandert men h in de praktijk alleen als de schatting voor $|\delta(t_n)|$ een waarde oplevert die $\gamma\times$ groter of $1/\gamma\times$ kleiner is dan ε, met bijvoorbeeld $\gamma = 3/2$.

Dat (6.11) correct is kun je als volgt inzien

$$\begin{aligned}
\widetilde{u}_{n+1} - u_{n+1} &= h\,[F(t_{n+1}, u_{n+1}) - F(t_n, u_n)] \\
&\approx h\,[F(t_{n+1}, u(t_{n+1})) - F(t_n, u(t_n))] \\
&\approx h\,[u'(t_{n+1}) - u'(t_n)] \approx h\,h\,u''(t_n) \approx 2\,h\,\delta(t_n).
\end{aligned}$$

6.4 Lokaal nauwkeurigere methoden

6.4.1 Trapezium regel, Adams-Bashford

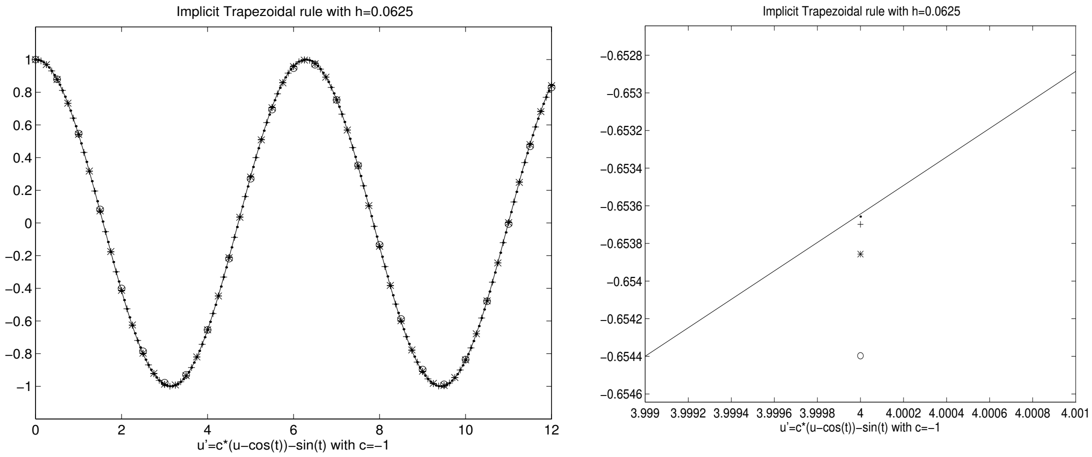
Analoog aan (6.7) geldt

$$\frac{u(t+h) - u(t)}{h} = u'(t+h) - \delta(t) \quad \text{waarbij} \quad \delta(t) \approx +\tfrac{1}{2}\,hu''(t). \tag{6.12}$$

Op het teken na, is de lokale fout min of meer gelijk aan die in (6.7). Bij middeling van de twee benaderingen verdwijnt de fout vrijwel geheel. Een preciesere analyse (die we hier niet geven), die gebaseerd op meer termen van de Taylor reeks, levert een lokale diskretisatie fout van orde h^2:

$$\frac{u(t+h) - u(t)}{h} = \tfrac{1}{2}\Big(u'(t+h) + u'(t)\Big) + \delta_{\text{trap}}(t) \tag{6.13}$$

$$\text{waarbij} \quad \delta_{\text{trap}}(t) \approx -\tfrac{1}{12}h^2 u'''(t).$$



FIGUUR 16. *De exacte en numerieke oplossingen van probleem* $u' = c\,(u - \cos(t)) - \sin(t)$, $u(0) = 1$, *met* $c = -1$. *De numerieke oplossing is verkregen met de impliciete trapezium regel met achtereenvolgens stapgrootte* $h = 0.5$ *(de '*$\circ$*' geven de gevonden waarden),* $h = 0.25$ *(*$* * *$*),* $h = 0.125$ *(*$+ + +$*),* $h = 0.0625$ *(*$\cdot\cdot\cdot$*). De exacte oplossing* $u(t) = \cos(t)$ *is de ononderbroken lijn. Het linker figuur laat de oplossing zien van* $t = 0$ *tot* $t = T = 12$. *In het rechter figuur is de oplossing in de buurt van* $t = 4$ *uitvergroot. Duidelijk is te zien dat de methode veel nauwkeuriger is dan voorwaarts Euler (zie* FIG. 14*) en dat halveren van* h *resulteert in een reduktie van de fout met een faktor* 4.

Vervangen we de u' door F, immers $u'(t) = F(t, u(t))$, dan krijgen we een bruikbare relatie. Vermenigvuldiging met h en schuiven met termen leidt tot

$$u(t_{n+1}) = u(t_n) + \tfrac{1}{2}\,h\,\Big[F(t_{n+1}, u(t_{n+1})) + F(t_n, u(t_n))\Big] + h\,\delta_{\text{trap}}(t_n) \tag{6.14}$$

Poetsen we de $h\,\delta_{\text{trap}}(t_n)$ weg dan krijgen we een bruikbare relatie: de **trapezium regel**

$$u_{n+1} = u_n + \tfrac{1}{2}\,h\,\Big[F(t_{n+1}, u_{n+1}) + F(t_n, u_n)\Big]\,. \tag{6.15}$$

Het nadeel van deze aanpak is dat u_{n+1} nu ook in de rechterkant, in een F-waarde voorkomt: we moeten u_{n+1} *oplossen*. We zeggen daarom de trapezium regel **impliciet** is (u_{n+1} moet opgelost worden), terwijl de voorwaarts Euler methode **expliciet** is (u_{n+1} volgt door *rechtstreekse* berekening). Vooral als F niet linear is kan dat een reden zijn om de trapezium regel maar niet te kiezen.

Door de u_{n+1} in de F waarde te vervangen door een goede gok voor u_{n+1} kunnen we de regel weer expliciet maken. We zouden de voorwaartse Euler methode kunnen gebruiken om u_{n+1} te gokken:

$$\begin{cases} \widetilde{u}_{n+1} = u_n + h\,F(t_n, u_n) \\ u_{n+1} = u_n + \tfrac{1}{2}\,h\,\Big[F(t_{n+1}, \widetilde{u}_{n+1}) + F(t_n, u_n)\Big]\,. \end{cases} \tag{6.16}$$

Deze **expliciete** trapezium regel wordt ook wel trapezium regel met **voorzegger** genoemd. We gebruiken immers een voorwaartse Euler stap om $\widetilde{u}_{n+1}$ te berekenen. Het verschil tussen $\widetilde{u}_{n+1}$ en u_{n+1} is van orde h^2 (waarom?). Hieruit volgt dat het verschil tussen $\frac{1}{2}\,h\,F(t_{n+1}, \widetilde{u}_{n+1})$ en $\frac{1}{2}\,h\,F(t_{n+1}, u_{n+1})$ van orde h^3 is (de h voor de F geeft een extra h), dit is van dezelfde orde van grootte als $h\,\delta_{\text{trap}}$.

In §6.5.5 zullen we zien dat, met name voor grotere stapgrootte h, de "duurdere" impliciete trapezium regel ook voordelen heeft ten opzichte van de expliciete trapezium regel.

Andere manier om aan hogere orde expliciete methoden te komen is door formules analoog aan (6.13) te ontwerpen die alleen u'-waarden hebben in tijdstippen t en eerder. Bijvoorbeeld

$$\frac{u(t+h)-u(t)}{h} = \tfrac{1}{2}\big(3\,u'(t) - u'(t-h)\big) + \delta_{\text{A-B}}(t) \tag{6.17}$$

$$\text{waarbij} \quad \delta_{\text{A-B}}(t) \approx \tfrac{5}{12}h^2 u'''(t).$$

De uitdrukking voor de lokale diskretisatie fout fout $\delta_{\text{A-B}}$ kan men weer afleiden met Taylor reeksen. De inmiddels gebruikelijke manipulaties leiden tot

$$u_{n+1} = u_n + \tfrac{1}{2}\,h\left[3F(t_n,u_n) - F(t_{n-1},u_{n-1})\right].$$

Dit is de zgn. **Adams-Bashford** methode. De methode is expliciet en de lokale fouten zijn van dezelfde orde van grootte als in de trapeziumregel. De methode is echter niet *zelfstartend*. Immers uit u_0 en u_1 kunnen we u_2 berekenen, u_1 en u_2 leidt to u_3, etc., maar waar halen we u_1 vandaan? Ook hier kunnen we weer een Euler stap doen: neem voor $u_1 = u_0 + hF(t_0,u_0)$. Je kunt je afvragen of het wel zin heeft een nauwkeurigere methode te gebruiken als je in de eerste stap toch een relatief onnauwkeurige Euler stap doet. De Euler stap doe je echter maar één keer terwijl je de nauwkeurigere stappen keer op keer uitvoert.

Globale fout. Voor methoden als de trapezium regel (de impliciete en de expliciete) en Adams-Bashford geldt dat, voor voldoende kleine stapgrootte h, de globale fout $\approx h^2\times$ een gladde funktie is:

$$u_n - u(t_n) \approx h^2 w(t_n) \quad \text{voor iedere} \quad n = 0,1,2,$$

De funktie w hangt af van de differentiaalvergelijking en van de methode, maar niet van de stapgrootte h. Het feit dat h^2 voor de w staat hangt samen met de h^2 in de uitdrukking voor de lokale diskretiesatie fout in de methoden (zie (6.13) en (6.17)). Voor de Adams-Bashford methode moet de startwaarde u_1 wel voldoende nauwkeurig zijn: in feite moet gelden $u_1 - u(t_1) \approx h^2\alpha$ voor zekere constante α. Met $u_1 = u_0 + hF(t_0,u_0)$ (Euler stap) geldt dat inderdaad: $u_1 - u(t_1) = -\frac{1}{2}h^2\,u''(t_0)$.

Globaal foutschatten. In dit soort methoden "van orde h^2" zal dus bij halvering van de stapgrootte de globale fout met een factor 4 kleiner worden. Dit inzicht kan weer gebruikt worden om de globale fout te schatten:

$$u_{h_0}(t) - u(t) \approx \tfrac{1}{3}[u_{2h_0}(t) - u_{h_0}(t)] \tag{6.18}$$

hierbij is $u_h(t)$ het numerieke resultaat met stapgrootte h. Dat de foutschattingsformule correct is volgt eenvoudig uit het bovenste diagram in FIG. 17. Hierin zijn langs de getallenrechte de verhoudingen aangegeven tussen de verschillen van diverse grootheden.

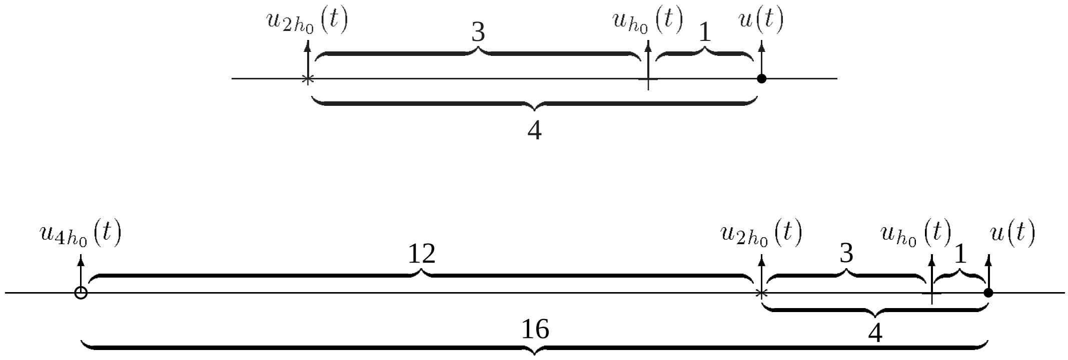
De evenredigheid van de globale fout met h^2 veronderstelt weer dat h voldoende klein is. Is h voldoende klein dan vertelt het onderste diagram in FIG. 17 ons dat

$$u_{2h_0}(t) - u_{h_0}(t) \approx \tfrac{1}{4}[u_{4h_0}(t) - u_{2h_0}(t)]. \tag{6.19}$$

Door te checken of de opeenvolgende verschillen tussen numerieke waarden in een tijdstip met een factor 4 afnemen bij halvering van de stapgrootte kan men enig vertrouwen krijgen in de aanname dat de stapgrootte klein genoeg is.

Lokaal foutschatten. Ook hier kan men de lokale fout schatten en hopen dat dat informatie geeft over de globale fout:

$$|\delta(t_j)| \leq \varepsilon \quad \text{alle} \quad j \quad \Rightarrow \quad |e_n| \leq C(t_n - t_0)\varepsilon$$



FIGUUR 17. *De numerieke waarden $u_h(t)$ met $h = h_0, 2h_0, 4h_0$ zijn berekend met een orde h^2 methode (als de trapezium regel of Adams-Bashford). De getallen geven de verhoudingen aan tussen de verschillen. De verhoudingsgetallen 1, 4 en 16 volgen onmiddelijk uit het feit dat de globale fout ongeveer evenredig is met h^2.*

voor een niet al te grote C. Omdat nu $\delta(t) = h^2 v(t)$ voor zekere $v(t)$ $(v(t) = \frac{1}{12}u'''(t)$ voor de trapezium regel) zal men bijvoorbeeld h moeten halveren als de schatting voor $|\delta(t_j)|$ aangeeft dat de stapgrootte $4\times$ te groot is.

Het schatten van de lokale fout kan weer goedkoop. De lokale fout in bijvoorbeeld de impliete trapeziumregel kan gemakkelijk geschat worden middels een Adams-Bashford stap: als u_n trapezium regel resultaten zijn bij stapgrootte h en

$$\widetilde{u}_{n+1} = u_n + \tfrac{1}{2}\,h[3\,F(t_n, u_n) - F(t_{n-1}, u_{n-1})].$$

dan

$$\delta_{\text{trap}}(t_n) \approx \frac{\widetilde{u}_{n+1} - u_{n+1}}{6h}. \qquad (6.20)$$

Merk op dat, als je toch een Adams-Bashford stap doet om de lokale fout te schatten, je het resultaat $\widetilde{u}_{n+1}$ ook kunt gebruiken als voorzegger voor een "expliciete" trapezium regel: Adams-Bashford is immers lokaal nauwkeuriger dan Euler. In de praktijk worden dit soort paren van methoden (van gelijke orde) vaak gebruikt. De voorzegger heet **predictor**, de tweede methode, die in feite de numerieke oplossing produceert, heet **corrector**. Predictor heeft hier een dubbele betekenis: hij dient als voorzegger voor de impliciete methode en hij wordt gebruikt om de lokale fout te schatten. Het gekoppeld gebruikt van de methoden wordt een **predictor-corrector methode** genoemd.

6.4.2 Runge Kutta en multistep methoden

Hierboven hebben we gezien hoe we methoden kunnen construeren die lokaal (en globaal) nauwkeuriger zijn dan Euler. Globaal gesproken zijn er twee typen manieren omdat de doen. In de eerste manier maak je gebruik van allerlei tussenstappen om nauwkeurige F-waarden te vinden. Dit soort methoden worden **Runge-Kutta methoden** genoemd. De expliciete trapezium regel is hier een voorbeeld van. De **klassieke Runge-Kutta** methode van Runge en Kutta zelf is een ander voorbeeld:

$$\begin{cases} k_0 = F(t_n, u_n) \\ v_{\frac{1}{2}} = u_n + \frac{1}{2}\, h\, k_0, \qquad k_{\frac{1}{2}} = F(t_{n+\frac{1}{2}}, v_{\frac{1}{2}}) \\ v'_{\frac{1}{2}} = u_n + \frac{1}{2}\, h\, k_{\frac{1}{2}}, \qquad k'_{\frac{1}{2}} = F(t_{n+\frac{1}{2}}, v'_{\frac{1}{2}}) \\ v_1 = u_n + h\, k'_{\frac{1}{2}}, \qquad k_1 = F(t_{n+1}, v_1) \\ u_{n+1} = u_n + h\, \left[\frac{1}{6}\, k_0 + \frac{1}{3}\, k_{\frac{1}{2}} + \frac{1}{3}\, k'_{\frac{1}{2}} + \frac{1}{6}\, k_1\right] \end{cases}$$

De k-waarden k_0, $k_{\frac{1}{2}}, \ldots k_1$ en de v-waarden $v_{\frac{1}{2}}, \ldots$ worden in iedere stap door nieuwe overschreven.
De klassieke Runge-Kutta methode is een methode van orde h^4.

In de tweede aanpak maak je om u_{n+1} te berekenen niet alleen gebruik van u_n maar ook van u_j-waarden uit het verleden ($j = n-1, n-2, \ldots$). Per stap bepaal je wel maar één nieuwe F-waarde. Dit soort methoden worden **multistep methoden** genoemd. Adams-Bashford is hier een voorbeeld van. BDF (backward difference formula) formules zijn andere veel gebruikte multistep methoden. De BDF van orde h^2 ziet er als volgt uit

$$u_{n+1} = \tfrac{4}{3}u_n - \tfrac{1}{3}u_{n-1} + \tfrac{2}{3}\, h\, F(t_{n+1}, u_{n+1}), \quad \text{met} \quad \delta(t) \approx -\tfrac{2}{9}h^2 u'''(t).$$

De multistep methoden zijn per stap goedkoper dan de Runge-Kutta methoden: in een multistep methode hoeft er, per stap, maar een nieuwe F-waarden uitgerekend te worden. Het is overigens niet zomaar duidelijk of dat in het totaal voordeliger is: dat hangt immers ook af van de grootte van de stapgrootte waarmee gewerkt kan worden. De Runge-Kutta methoden zijn een-staps methoden (er zijn geen u-waarden of F-waarden uit het verleden nodig). Daardoor is het tussentijds veranderen van de stapgrootte (stapgrootte besturing) wat gemakkelijker.

Foutschatten, etc., gaat voor deze hogere orde methoden op dezelfde manier als voor Euler, trapezium regel en Adams-Bashford. Uiteraard na voor de hand liggende aanpassingen, aanpassingen die te maken hebben met de orde van de methode.

Opmerking. Niet ieder differentie qoutiënt voor $u'(t)$ leidt tot een bruikbare methode.
Voor de liefhebbers het volgende illustratieve voorbeeld.

Er geldt $([u(t_{n+1} - u(t_n)] + 2[u(t_n - u(t_{n-1})])\,/(3h) = u'(t_n) + \delta(t)$ met $\delta(t)$ van orde h. De resulterende nmerieke methode $u_{n+1} = -u_n + 2u_{n-1} + h\, F(t_n, u_n)$ ziet er bruikbaar uit, maar is toch waardeloos. Pas deze methode maar eens toe op het flauwe probleem $u' = 0$, $u(0) = 0$ en neem $u_0 = 0$ (gegeven) en $u_1 = \varepsilon$ (bij het bepalen van u_1 maken we een klein foutje). De oplossing u_n zou nul moeten zijn, of, omdat fouten onvermijdelijk zijn, ≈ 0. Echter $|u_n| \geq 2^n|\varepsilon|$. Dus $|u_{20}| \geq 10^6\,|\varepsilon|$.

6.5 Numerieke stabiliteit

De resultaten in het vorige hoofdstuk gaan er vanuit dat h voldoende klein is. Als h niet voldoende klein is zijn hoger orde Taylor reeks termen niet verwaarloosbaar en zal de globale fout daardoor in bijvoorbeeld voorwaarts Euler niet evenredig zijn met h. Het voorbeeld in FIG. 12 laat zien dat h ook om een andere reden niet voldoende klein kan zijn. In dat voorbeeld is $u(t) = \cos(t)$ en $h \approx 0.002$ en de verwaarloosde Taylor term, in zeg $t = \pi/4$, is $\frac{1}{6}h^3 u'''(t) \approx 0.9\,10^{-9}$ en is veel kleiner dan de term $\frac{1}{2}h^2 u''(t) \approx 1.4\,10^{-6}$ die wel meegenomen is. In dit hoofdstuk proberen we die andere reden te begrijpen.

We analyseren zorgvuldig het voorbeeld in FIG. 12.

$$\begin{cases} u'(t) = c\,(u - \cos(t)) - \sin(t), & t \in [0, T] \\ u(0) = 1 \end{cases} \tag{6.21}$$

6.5.1 De stabiliteit van het beginwaarde probleem

De slechte resultaten van voorwaarts Euler in FIG. 12 zijn niet te wijten aan opblazen van foutjes door de differentiaalvergelijking zelf. Immers als u de oplossing is van (6.21) en $\widetilde{u}$ is ook de oplossing van (6.21) maar met een licht verstoorde beginwaarde $\widetilde{u}(0) = 1 + \varepsilon$ dan is de fout $\widetilde{u}(t) - u(t)$ ten gevolge van deze verstoring van de beginwaarde gelijk aan

$$\widetilde{u}(t) - u(t) = \varepsilon\, e^{ct}$$

($v = \widetilde{u} - u$ voldoet aan de vergelijking $v'(t) = c\,v$ en $v(0) = \varepsilon$.Dit volgt gemakkelijk door de differentiaalvergelijkingen voor $\widetilde{u}$ en u van elkaar af te trekken.)
Voor positieve c groeit de fout op de beginwaarde, voor negatieve c — het geval waarin we hier geïnteresseerd zijn — krimpt de fout. Voor $c = -1000$ reduceert de differentiaalvergelijking zelf ieder foutje bijzonder krachtig: in bv. $t = 0.1$ is de reduktie al

$$e^{-1000*0.1} = 3.7\,10^{-44}.$$

Of, in het algemeen, een differentiaalvergelijking zelf fouten opblaast weten we vaak al uit het soort globale analyse uit het vorige hoofdstuk. Van het fase portret weten we of oplosssingen van de differentiaalvergelijking die in naburige punten starten min of meer bij elkaar in de buurt blijven.

6.5.2 De stabiliteit van voorwaarts Euler

We passen voorwaarts Euler toe op (6.21).
Het verschil tussen (6.5),

$$u_{n+1} = u_n + hF(t_n, u_n),$$

en (6.8),

$$u(t_{n+1}) = u(t_n) + hF(t_n, u(t_n)) + h\,\delta_{\text{Euler}}(t),$$

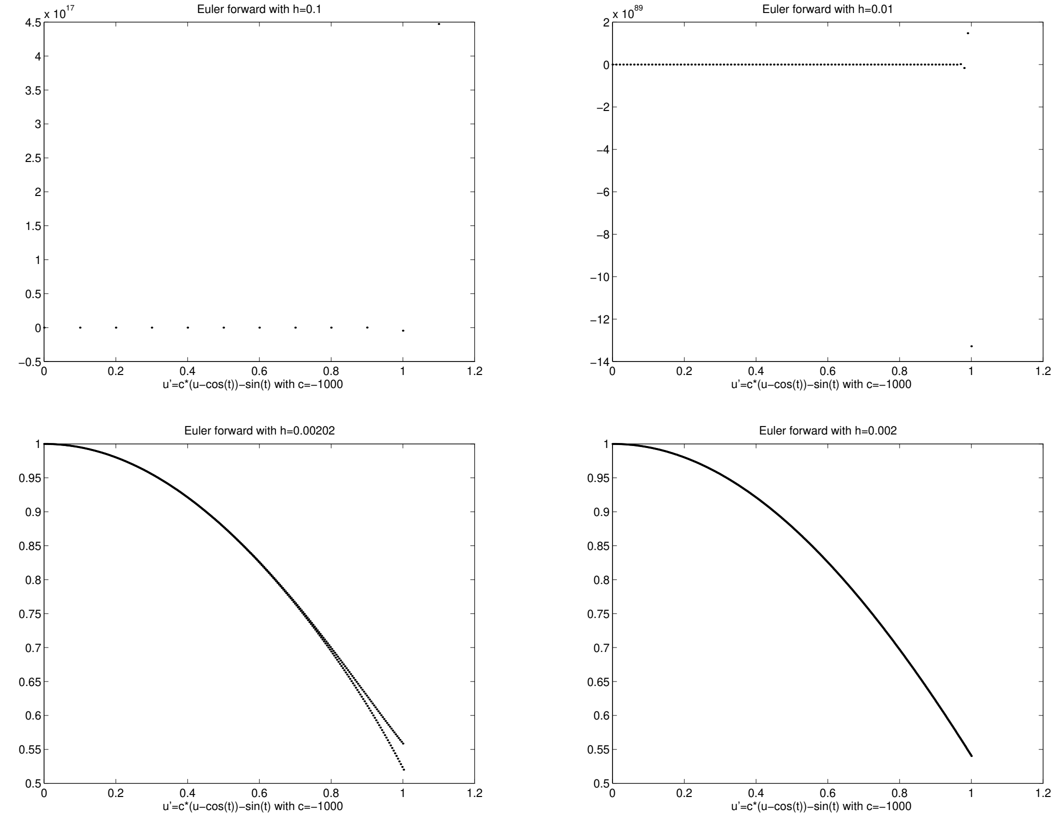
levert de volgende relatie op voor de globale fouten e_n.

$$\begin{aligned} e_{n+1} &= e_n + h\left[F(t_n, u_n) - F(t_n, u(t_n))\right] - h\,\delta_{\text{Euler}}(t_n) \\ &= e_n + h\,c\,e_n - h\,\delta_{\text{Euler}}(t_n). \end{aligned} \tag{6.22}$$

In de laatste stap hebben we gebruikt dat $F(t, u) = c(u - \cos(t)) - \sin(t)$. De laatste relatie voor de globale fouten leert ons dat oude fouten (gerepresenteerd door e_n) worden vermenigvuldigd met $1 + h\,c$ voordat de nieuwe fouten erbij komen:

$$e_{n+1} = (1 + h\,c)\,e_n - h\,\delta_{\text{Euler}}(t_n). \tag{6.23}$$

We bekijken eens wat dat betekent voor $c = -1000$.
A. Uit FIG. 18 (rechtsboven) blijkt dat, voor $h = 0.01$, de globale fouten zeer groot zijn. Dit ligt niet aan de nieuwe lokale fouten. Die zijn, omdat $u = \cos(t)$ voor (6.21), gelijk aan



FIGUUR 18. *De figuren tonen de numerieke resultaten met voorwaarts Euler voor* (6.21) *met* $c = -1000$ *voor* $h = 0.01$ *(linksboven),* $h = 0.01$ *(rechtsboven),* $h = 0.00202$ *(linksonder) en* $h = 0.002$ *(rechtsonder). Let op de schaling langs de vertikale as.*

$|h\delta_{\text{Euler}}(t_n)| = |\frac{1}{2}\,h^2\,u''(t_n)| = 0.5\,10^{-4}$. Dus behoorlijk klein. De vermenigvuldigings factor voor oude fouten is echter gelijk aan

$$1 + h\,c = 1 - 0.01 * 1000 = -9.$$

Blijkbaar worden oude fouten in iedere stap met een factor -9 vergroot. Na 11 stappen is dat met een factor $(-9)^{11} = -3.1\,10^{10}$. Dit verklaard duidelijk waarom stapgrootte $h = 0.01$ voor dit probleem ((6.21) met $c = -1000$) een slechte keus is.
B. Met $h = 0.00202$ hebben we ook weer grote globale fouten (zie FIG. 12 en FIG. 18, linksonder). Nu ligt dat zeker niet aan de nieuwe lokale fouten: die zijn nu nog kleiner. De vermenigvuldigings factor voor oude fouten is nu echter ook niet zo beroerd:

$$1 + h\,c = 1 - 0.00202 * 1000 = -1.02\ :$$

per stap worden de oude fouten, afgezien van het teken, met 2% vergroot. De vergroting per stap valt dus wel mee. De misère zit echter in het groot aantal stappen dat we moeten doen. Om in bijvoorbeeld $t = 1$ te komen zijn, met $h = 0.00202$, ongeveer 500 stappen nodig. Dat betekent dat een fout gemaakt op tijdstip $t = 0$ door voorwaarts Euler in tijdstip $t = 1$ opgeblazen wordt met een factor $(1.02)^{500} \approx e^{0.02*500} = e^{10} \approx 2.2\,10^4$.
C. Met $h = 0.002$ is er geen probleem (zie FIG. 18 rechtsonder): oude fouten worden per stap vermenigvuldigd met $1 + h\,c = -1$: geen groei. Als $h < 0.002$ worden oude fouten zelfs gereduceerd.

D. Voor $c = +1$ en $h = 0.02$ worden oude fouten ook telkens vergroot met 2% ($1 + h\,c = 1 + 0.02 * 1 = 1.02$). Nu zijn er echter maar 50 stappen nodig om van $t = 0$ tot $t = 1$ te komen en de fout wordt opgeblazen met slechts een factor $(1.02)^{50} \approx e \approx 2.7$ en dit is precies de factor waarmee de differentiaalvergelijking zelf de fouten opblaast (zie §6.5.1).

Uit bovenstaande beschouwingen mogen we het volgende konkluderen.

Konklusie voor voorwaarts Euler.
• Als $|h\,c|$ klein is — denk aan $|h\,c| \leq 0.3$ — dan levert de voorwaartse Euler methode geen problemen (kleine foutgroei in een beperkt aantal stappen). De probleem die eventueel optreden komen van de differentiaalvergelijking zelf (zie D).
• Als $|h\,c|$ niet klein is — denk aan $|h\,c| > 0.3$ — dan moet h zo zijn dat

$$|1 + h\,c| \leq 1. \tag{6.24}$$

Deze numerieke stabiliteits eis kan niet verzwakt worden. Als h maar een beetje te groot is (zie B) hebben we al last van *numerieke instabiliteiten*: kleine fouten worden gigantisch opgeblazen en de effekten daarvan vertroebelen het numeriek resultaat volledig.

Bovenstaande konklusie blijkt ook in zekere zin te gelden voor algemenere niet-lineaire en meer-dimensionale problemen.
We zullen met voorwaarts Euler voor (6.1) geen last hebben van numerieke instabiliteiten als, voor iedere t met $c = \frac{\partial F}{\partial u}(t, u(t))$ geldt

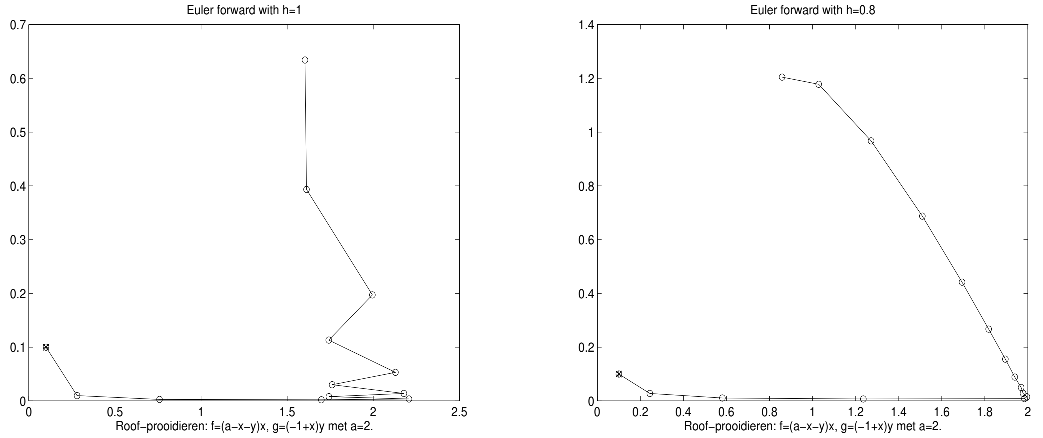
$$|h\,c| \text{ klein} \quad \text{of} \quad |1 + h\,c| \leq 1. \tag{6.25}$$

Voor bijvoorbeeld $u' = -u + u^3 - \cos(t)$, $u(0) = 1$ is $F(t, u) = -u + u^3 - \cos(t)$ en $c = -1 + 3(u(t))^2$. Hier is $u(t)$ de oplossing is van de differentiaalvergelijking in het tijdstip t.
Voor (6.2) moet (6.25) gelden met $c = \lambda_i$ voor iedere t en voor iedere eigenwaarde λ_i van de Jacobiaan

$$\begin{bmatrix} \dfrac{\partial f}{\partial x}(x(t), y(t)) & \dfrac{\partial f}{\partial y}(x(t), y(t)) \\[2ex] \dfrac{\partial g}{\partial x}(x(t), y(t)) & \dfrac{\partial g}{\partial y}(x(t), y(t)) \end{bmatrix}.$$

Bij niet-lineaire problemen kan er gezondigd worden tegen de numerieke stabiliteits eis (6.25) gedurende een klein tijdsinterval, terwijl voor er voor andere tijdstippen t prima aan de eis voldaan is. In dat tijdstrajektje van numerieke instabiliteit zal de numerieke instabiliteit zich manifesteren in de vorm van onnatuurlijke pieken ("wiggels") die in het stabiele tijdstraject weer kunnen verdwijnen (zie FIG. 19).

Het lijkt een beetje dwaas om de numerieke stabiliteit te analyseren met behulp van eigenwaarden van een Jacobi matrix in een oplossing die we niet kennen. Soms, in gevallen waar de gebruikelijke numerieke oplosmethoden merkwaardige resultaten produceren, neemt men inderdaad de moeite en bepaalt men de eigenwaarden (in de numerieke oplossing) in een aantal tijdstippen (niet in alle). Uit de ligging van $h\times$ de eigenwaarden probeert men het merkwaardige gedrag van de numerieke oplosmethoden te begrijpen en tot een geschikte alternatief te komen. Gewoonlijk is het moeizaam bepalen van eigenwaarden niet nodig. Uit de analyse van het simpel voorbeeldje (6.21) weten we dat als h maar iets te groot is de gevolgen desastreus zijn. Numerieke instabiliteiten produceren gelukkig geen oplossingen die er natuurlijk uitzien. Aan de oplossing zien we al of h te groot is (of groot genoeg) voor numerieke stabiliteit.



FIGUUR 19. *Voor het roof-prooidieren model vertoont voorwaarts Euler instabiliteiten met $h = 1$ (linkse figuur) in de vorm van "wiggels". Voor $h = 0.8$ zijn die instabiliteiten verdwenen. De oplossing is nog steeds niet erg nauwkeurig, maar dit komt hier omdat de lokale diskretisatie fouten nog te groot zijn.*

6.5.3 Wanneer is h klein genoeg?

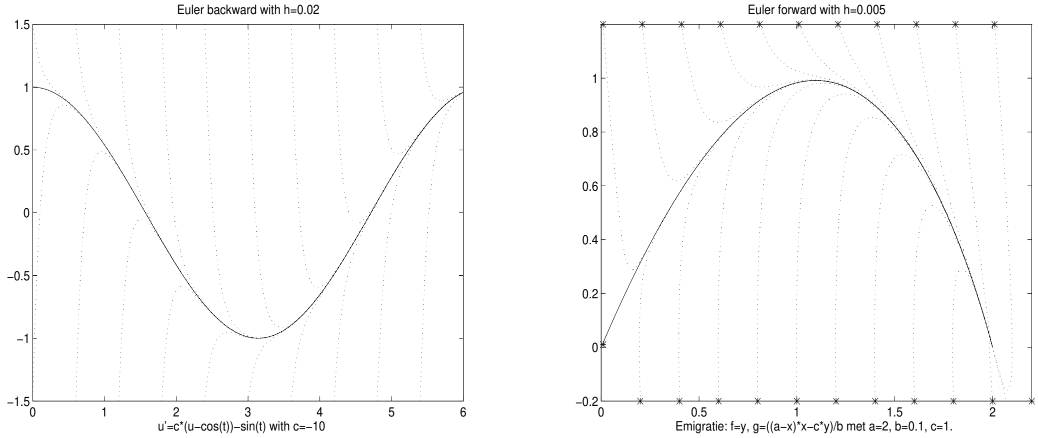
Er zijn een drietal redenen om de grootte van h te willen beperken.

Informatief. Voor beleid of voor een biologische analyse heeft men gewoonlijk geen behoefte aan een grote reeks getallen. Als men bijvoorbeeld de ontwikkeling van een roofdierenpopulatie wil volgen zal men geen behoefte hebben aan aantallen per dag, maar is het al interessant om te weten wat er per seizoen (of wellicht per jaar) gebeurt. Te veel informatie kan zelfs vermoeiend en afleidend werken. Het is voldoende en genoeg om de oplossing van een roof-prooidieren model te kennen in tijdstippen t_n met stapgrootte h is 3 maanden.

Iedere specialisme heeft overigens zijn eigen tijdschaal: de interessante fenomenen in bijvoorbeeld de micro-biologie spelen zich af op een kleinere tijdsschaal.

Nauwkeurigheid. De numerieke resultaten van een model moeten uiteraard ook voldoende nauwkeurig zijn, bijvoorbeeld 2 cijfers achter de komma als het om biologische dichtheden gaat, 20 cijfers achter de komma als de baan van een raket naar Mars berekend moet worden. De lokale fouten in de numerieke methode moeten daartoe voldoende klein zijn en die lokale fouten moeten niet vergroot worden door de numerieke methode (numerieke stabiliteit). Als we geen last hebben van numerieke instabiliteit kunnen we checken of h klein genoeg is door stapgrootte halvering. Dit levert ook een schatting voor de globale fout (zie §6.3.3). De h die hierbij uit de bus komt rollen is vaak veel kleiner dan de h die we kozen om "informatieve" redenen. In geval van het roof-prooidier model zou dat best een h van 1 dag kunnen zijn.

Numerieke stabiliteit. Vaak is de h die een lokaal nauwkeurig resultaat oplevert ook klein genoeg voor numerieke stabiliteit. Helaas is dat niet altijd het geval. De stabiliteitseis $|1 + h\,c| \leq 1$ kan een h opleggen die veel kleiner is dan vereist is voor (lokale) nauwkeurigheid ($|\delta_{\text{Euler}}(t)| \leq \varepsilon$).
Voor (6.21) met $c = -10\,000$ is $|\delta_{\text{Euler}}(t)| \approx |\frac{1}{2}\,h\,u''(t)| \leq 0.005$ als $h = 0.01$ (2 cijfers achter de komma goed) terwijl $|1 + h\,c| \leq 1$ verlangt dat $h \leq 2\,10^{-4}$. Deze "stabiele" h is 200 (!) keer kleiner dan de "nauwkeurige" h. Omdat, met $c = -10\,000$, de differentiaalvergelijking zelf fouten krachtig reduceert (zie §6.5.1), zou je met $h = 0.01$ zelfs een nauwkeurigheid verwachten van $h|\delta_{\text{Euler}}| \leq 0.5\,10^{-4}$. Uiteraard krijg je met de "stabiele" $h = 2\,10^{-4}$ en veel nauwkeuriger resultaat dan je verlangt: de fout is nu kleiner dan ($h|\delta_{\text{Euler}}| \leq$) $2\,10^{-8}$. Het is



FIGUUR 20. *Voorbeelden van stijve problemen. De gestippelde oplossingen lopen relatief snel naar de langzaam variërende oplossing (ononderbroken lijn).*

niet anders: met voorwaarts Euler krijg je, voor dit soort zogenaamde "stijve" problemen (zie §6.5.4) òf een onzinnig òf een heel nauwkeurig resultaat.

Als om reden van numerieke nauwkeurigheid of numerieke stabiliteit h kleiner gekozen moet worden dan om informatieve redenen wenselijk is, zit er niets anders op om h maar klein te kiezen en bij weergave van de resultaten een aantal tussenresultaten weg te laten (zie §6.5.5 voor een alternatief).

6.5.4 Stijve problemen

De eis voor numerieke stabiliteit is met name voor de zogenaamde stijve differentiaalvergelijkings problemen veel zwaarder dan de lokale nauwkeurigheids eis. Een probleem is **stijf** als de oplossing waarin we geïnteresseerd zijn relatief langzaam varieert (dus als u'' niet al te groot is), terwijl naburige oplossingen snel variëren, d.w.z., snel naar de langzaam variërende oplossing "gezogen" worden (zie FIG. 20). De vergelijking (6.21) met c negatief maar groot in absolute waarde (denk aan $c = -1000$) is daar een voorbeeld van (oplossingen van deze vergelijkingen met $c = -10$ staan in het linker figuur van FIG. 20). De naburige snel variërende oplossingen bepalen voor voorwaarts Euler de stapgrootte, terwijl de nauwkeurigheid af zou moeten hangen van de langzaam variërende oplossing.

Stijve problemen komen in de praktijk vaak voor als een model "verfijnd" wordt. We denken aan het volgende:
Stel dat $u(t) - \cos(t) = 0$ een model is voor een of ander fenomeen. Stel dat bij nadere beschouwing blijkt dat de afgeleide toch zwak gekoppeld is aan u. Dan zou $-\frac{1}{1000}(u'(t) + \sin(t)) = u(t) - \cos(t)$ een beter model kunnen zijn. Dit is precies het probleem wat we zonet in §6.5.2 besproken hebben.
Het modelleren van kleine effekten in hogere orde afgeleiden leidt tot stijve problemen. Voor het oorspronkelijke model werkt voorwaarts Euler met redelijke stapgrootte prima, terwijl voor het verfijnde model voorwaarts Euler het laat afweten (d.w.z., een piep kleine h vereist).

Uiteraard heb je Euler niet nodig om $u(t) - \cos(t) = 0$ op te lossen. Hieronder *voor de geïnteresseerde* een voorbeeld waar je dat wel nodig hebt en waarbij modelverfijning tot numerieke problemen leidt.
Voorbeeld. Voorwaarts Euler kan toegepast worden op logistische model $u' = (a - u)u$ voor bevolkingsgroei. Als een klein deel van de populatie wegtrekt (emi– en immigratie) dan leidt

een verfijning van het model waarin met dit klein effect rekening wordt gehouden (onder zekere aanname) tot $u' = (a-u)u - bu''$ met b klein, bv. $b = 0.01$. Omdat $u'' = \frac{1}{b}\,[(a-u)u - u']$ met $v = u'$ geschreven kan worden als

$$\begin{cases} u' = v \\ v' = \frac{1}{b}\,[(a-u)u - v] \end{cases}$$

kan voorwaarts Euler ook toegepast worden op het verfijnde model. Het rechter figuur in FIG. 20 (met $b = 0.1$) laat zien waarom dit probleem stijf is.

6.5.5 Als h "klein genoeg" te klein is

De stabiliteits eis $|1+h\,c| \leq 1$ voor voorwaarts Euler kan, met name voor stijve problemen, wel erg beperkend zijn. Een kleine stapgrootte betekent immers veel stappen om in een specifiek tijdstip te komen, dus veel rekenwerk. Met name voor hoog-dimensionale problemen kan dat wel eens betekenen dat het het produceren van een nauwkeurige numerieke oplossing te lang duurt (als het berekenen van de weersverwachting voor morgen drie dagen duurt heeft de berekening geen zin). Er zijn gelukkig andere methoden die grotere stappen toelaten, die minder last hebben van numerieke instabiliteiten. Een zo'n methode ontstaat uit (6.12).

Uit het differentie quotiënt (6.12),

$$\frac{u(t+h) - u(t)}{h} = u'(t+h) + \delta_{\text{EB}}(t) \quad \text{waarbij} \quad \delta_{\text{EB}} \approx -\tfrac{1}{2}\,hu''(t), \tag{6.26}$$

volgt de numerieke methode

$$u_{n+1} = u_n + h\,F(t_{n+1}, u_{n+1}). \tag{6.27}$$

Dit is de **terugwaartse Euler** methode ("backward Euler"): "terugwaarts" omdat de methode gebaseerd is op een terugwaartse differentie (6.12).

Terugwaarts Euler toepassen op (6.21) levert globale fouten die voldoen aan de relatie (ga te werk als in §6.5.2)

$$e_{n+1} = e_n + h\,c\,e_{n+1} + h\delta_{\text{EB}}(t_n).$$

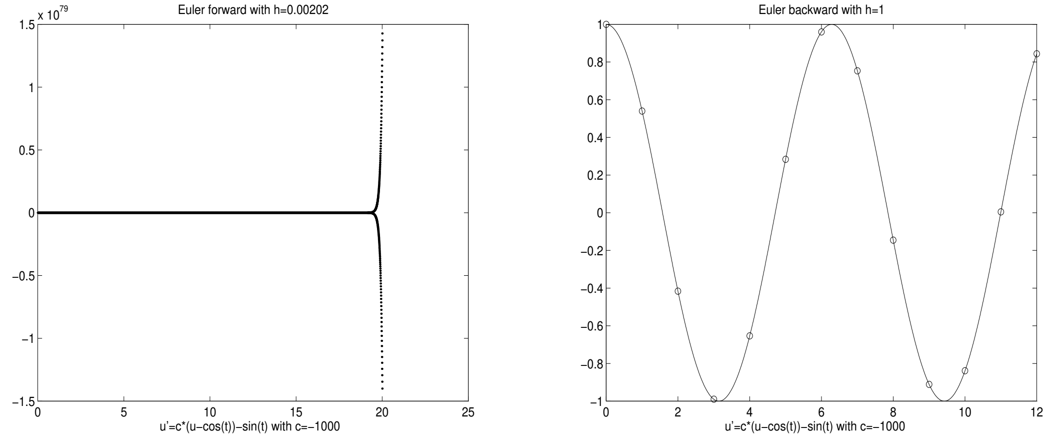
Blijkbaar geldt

$$e_{n+1} = \frac{1}{1-h\,c}\,e_n + \frac{h\,\delta_{\text{EB}}(t_n)}{1-h\,c}. \tag{6.28}$$

De lokale fouten zijn, op teken na, ongeveer gelijk aan de lokale fouten in voorwaarts Euler met dezelfde stapgrootte. De foutvoortplanting is echter totaal anders. Voor bijvoorbeeld $c = -1000$ worden oude fouten vermenigvuldigd met

$$\frac{1}{1-h\,c} = \frac{1}{1+1000\,h}.$$

Deze factor is ≤ 1 voor iedere stapgrootte h ($h > 0$): voor iedere h worden oude fouten altijd gedempt. Terugwaarts Euler is (voor probleem (6.21)) numeriek stabiel voor iedere h en h kan volledig bepaald worden door lokale nauwkeurigheid: zie FIG. 21. Men zegt dat terugwaarts Euler **A-stabiel** is. Voor niet-lineaire stijve differentiaalvergelijkingen kunnen er bij terugwaarts Euler ook weer numerieke stabiliteitsproblemen optreden, maar de h waarbij dit gebeurt zal toch veel en veel groter zijn dan in geval van voorwaarts Euler.



FIGUUR 21. *De figuren laten zien dat terugwaarts Euler een veel grotere stapgrootte toelaat dan voorwaarts Euler. In beide gevallen is* (6.21) *met* $c = -1000$ *opgelost. In het linker figure produceert voorwaarts Euler met* $h = 0.00202$ *een onzinnige oplossing (let op de schaling langs de vertikale as), terwijl in het rechter figuur terugwaarts Euler een heel nauwkeurige oplossing produceert (met* $\circ$*) al voor grote* h*, nl.* $h = 1$*. De ononderbroken lijn is de exacte oplossing.*

Terugwaarts Euler is even nauwkeurig als voorwaarts Euler, maar numeriek veel stabieler. Bij deze methode kan de stapgrootte wel bijna volledig bepaald worden door de lokale nauwkeurigheid. Het nadeel is echter dat de methode impliciet is: in iedere stap moet u_{n+1} opgelost worden. Er zijn relatief grote stappen mogelijk, maar de stappen zijn duur. Voor stijve problemen is het alternatief goedkope maar kleine en dus vele stappen met voorwaarts Euler (of met een andere methode zoals Adams-Bashford of de expliciete trapezium regel). Het is niet duidelijk welke aanpak de beste is, maar in de praktijk blijkt terugwaarts Euler het met grote voorsprong te winnen van voorwaarts Euler (en methoden als Adams-Bashford). Voor niet-stijve problemen is er geen reden om terugwaarts Euler te gebruiken en wint voorwaarts Euler (of een andere expliciete methode).

De impliciete trapezium regel heeft eenzelfde prettige foutvoortplantingsgedrag als terugwaarts Euler. Deze regel vermenigvuldigt oude fouten voor probleem (6.21) met

$$\frac{1 + \frac{1}{2}\,h\,c}{1 - \frac{1}{2}\,h\,c}.$$

en ook hier worden alle fouten gedempt voor iedere h als $c < 0$. De impliciete trapezium regel is ook A-stabiel. De methode is nauwkeuriger (orde h^2) dan terugwaarts Euler (orde h).

De andere *expliciete* methoden, de expliciete trapezium regel, Adams-Bashford, klassieke Runge–Kutta hebben wel numerieke stabiliteitsproblemen voor stijve problemen. Het analyseren van de foutvoortplanting door deze methode is lastig en laten we hier achterwege. Wel melden we dat voor (6.21) en reële c de volgende numerieke stabiliteitsrestrikties op h gelden:

$h\,c \geq -2$ (voorwaarts Euler),

$h\,c \geq -2$ (expliciete trapezium regel),

$h\,c \geq -1$ (Adams-Bashford),

$h\,c \geq -2.7$ (klassiek RK).

Als vuistregel geldt dat de impliciete methoden veel stabieler zijn dan de expliciete methoden. Iedere expliciete methode stelt strikte numerieke stabiliteitseisen aan h. Voor stijve problemen verlangen expliciete methoden daarom zeer kleine stapgrootte. Voor dit soort problemen is het vaak efficiënter een impliciete methode met grotere stappen te gebruiken ook al zijn de stappen duurder.

6.6 Programma pakketten

— Bij de meeste standaardprogramma's kan men een zekere tolerantie aangeven, maar die heeft doorgaans betrekking op de lokale diskretisatie fout (dus op $\delta(t_n)$). Hoeveel van die fouten gemaakt worden en hoe ze voortgeplant worden krijgt men niet te horen. Men tast omtrent de nauwkeurigheid van het resultaat dus volkomen in het duister. Er zit dan weinig anders op dan het programma nog eens te laten lopen met een $10\times$ zo kleine tolerantie en maar aan te nemen dat de fouten dan beduidend ($10\times$?) kleiner zijn dan het verschil tussen beide uitkomsten.

— Voor differentiaalvergelijkingen zijn *adaptieve methoden* ontwikkeld (die dus zelf een gunstige stapgrootte bepalen in afhankelijkheid van het gedrag van de oplossing). Deze methoden zijn gebaseerd op predictor-corrector technieken en werken alleen als er geen serieuze numerieke stabiliteits restrikties op de stapgrootte h liggen (ze werken niet of slecht voor stijve problemen).

— Er zijn voor verschillende klassen van differentiaalvergelijkingen aparte programma's ontwikkeld, niet alleen voor stijf en niet-stijf, maar ook voor vergelijkingen waarin bijvoorbeeld een of meer afgeleiden ontbreken of die snel-oscillerende oplossingen hebben. Bij Hamilton systemen is het bijvoorbeeld beter om methoden als voorwaarts Euler en terugwaarts Euler te mengen: $x_{n+1} = x_n + h\, f(x_n, y_n)$, $y_{n+1} = y_n + h\, g(x_{n+1}, y_n)$.

6.7 Oefeningen

Oefening 6.1 [Tentamensom november 1997]

Men lost een zekere differentiaalvergelijking, gedefinieerd op het tijdsinterval $[0, 1]$, op met behulp van de voorwaarts Euler methode en met de trapeziumregel (een zekere hogere orde methode; zie eventueel onderdeel (e)). Werkend met verschillende stapgrootten h vindt men onderstaande benaderingen voor $u(1)$, de exacte oplossing u in het tijdstip $t = 1$.

h	I	II
0.04	4.6673	3.6029
0.02	3.9569	3.4446
0.01	3.6593	3.4105
0.005	3.5249	3.4026
0.0025	3.4612	3.4007

a. Helaas is ten gevolge van een administratieve fout niet na te gaan of tabel I bij Euler hoort en tabel II bij de trapeziumregel of andersom. Kan je met een redelijke zekerheid vaststellen welke kolom bij welke methode hoort?

Beschouw verder, voor $\varepsilon = 10^{-3}$, de differentiaalvergelijking

$$\begin{cases} u'(t) = f(t, u(t)) = \frac{1}{\varepsilon}(-u(t) + \cos(t)) \quad \text{voor } t \geq 0 \\ u(0) = u_0 = 1. \end{cases}$$

Voor $h > 0$ is $t_n = nh$ en is u_n de met voorwaarts Euler berekende benadering voor $u(t_n)$.

b. Bepaal de voorwaarts Euler oplossing met stapgrootte $h = 0.01$ in $t = 0.02$.

c. In $t = 1$ is de voorwaarts Euler oplossing met $h = 0.01$ gelijk aan $-1.0625\,10^{90}$ terwijl de exacte oplossing ongeveer gelijk is aan $\cos(1)$. Verklaar hoe deze grote fout kon ontstaan en geef aan hoe je met behulp van de voorwaarts Euler methode zo'n grote fout had kunnen vermijden (motiveer met name je keuze voor de stapgrootte h zorgvuldig).

d. De trapeziumregel benadering $\widetilde{u}_n$ voor $u(t_n)$ berekent men als volgt[2]

$$\begin{cases} \widetilde{u}_{n+1} = \widetilde{u}_n + \frac{1}{2}h[f(t_n, \widetilde{u}_n) + f(t_{n+1}, \widetilde{u}_{n+1})] \\ \widetilde{u}_0 = u_0. \end{cases}$$

Verwacht je dat, met $h = 0.01$, de fout in de trapeziumregel benadering $\widetilde{u}_{100}$ voor $u(1)$ ook zo'n grote fout zal vertonen als die in de voorwaarts Euler benadering? Motiveer je antwoord.

6.8 Computer-opdrachten

Computer Opgave 6.I [Numerieke oplosmethoden]

a. De maker van het programma `des` weet zeker dat hij een van de volgende methoden gebruikt heeft:

Voorwaarts Euler, Adams-Bashford, impliciete trapezium regel.

Hij is echter vergeten welke van die drie.
Zoek experimenteel uit welke methode hij waarschijnlijk gebruikt heeft.
Hint. Los

$$\begin{cases} u' = c(u - \cos(t)) - \sin(t), \\ u(0) = 1 \end{cases} \tag{6.29}$$

op met `des` voor geschikte h en c en kijk naar oplossing in bv $t = 3$.

Vertoont de numerieke oplossing van (6.29) voor $c = -1$ in $t = 4$ ook het verwachte fout gedrag? Verklaring?

b. Op het college hebben we gezien dat we met voorwaarts Euler voor (6.29) alleen nauwkeurige resultaten mogen verwachten als $|1 + h\,c| \leq 1$. Voor negatieve reële c betekent dit dat $-2 \leq h\,c$.
Voor de `des` methode (de methode die gebruikt is in `des`) is er ook zo'n scherpe grens $-\beta \leq h\,c$. Bepaal β experimenteel (in 2 signifikante cijfers). Ga na of de kombinatie c, h hetzelfde stabiliteits gedrag vertoont als de kombinatie $c/10$, $10h$. Hoe zit het met de nauwkeurigheid? Is $c = 1$, $h = 0.2$ even nauwkeurig als $c = 0.1$ en $h = 2$? Neem nu $c = -100$ bereken op 2 cijfers nauwkeurig de oplossing voor tijdstippen tussen 0 en 20 (globale fout kleiner dan 10^{-2}). Geef een betrouwbare schatting van de fout. (Hoe onderbouw je de betrouwbaarheid van je schatting?)

Bekijk ook eens de lineaire differentiaalvergelijking

$$\begin{cases} x' = d(-x + y) \\ y' = d(-x - y) \end{cases} \quad \text{of, in matrix vorm,} \quad \begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} -d & d \\ -d & -d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

---

[2] Voor de exacte oplossing geldt

$$\begin{cases} u(t_{n+1}) = u(t_n) + \frac{1}{2}h[f(t_n, u(t_n)) + f(t_{n+1}, u(t_{n+1}))] + h\delta_{\text{trap}}(t_n) \\ \text{met} \quad \delta_{\text{trap}}(t_n) \approx -\frac{1}{12}h^2 u'''(t_n). \end{cases}$$

Voor de eigenwaarden $c = -d \pm d\,i$ en stapgrootte h moet weer gelden dat $|hc| \leq \widetilde{\beta}$. Bepaal $\widetilde{\beta}$ (voor voorwaarts Euler is $\widetilde{\beta} = 1$).

c. We onderzoeken verder het gedrag van de `des` methode voor het immuun-reaktie model

$$\begin{cases} x' = (1 + \kappa x - \kappa\, y)\, x \\ y' = (-1 + 3\,\kappa\, x - \kappa y - 4\,\vartheta\,\kappa\, x\, y)\, y \end{cases} \tag{6.30}$$

Neem $\kappa = 1.3$ en $\vartheta = 0$ (en start in $x_0 = 2$, $y_0 = 3$).

Voor welke stapgrootte h lijken de numerieke instabiliteiten geen rol meer te spelen? Voor welke h is dat voor geval $\vartheta = 0.06$? Kan je verklaren waarom de numerieke stabiliteit verbetert? (Hint: Wat gebeurt er met de eigenwaarden van de Jacobiaan in het evenwicht?)

Plot de oplossing voor $\vartheta = 0.06$ in het fasevlak voor $h = h_0$, $h = h_0/2$, $h = h_0/4$ en $h = h_0/8$ met $h_0 = 0.2$ ($T = 10$). Kan je uit de resultaten de globale fout schatten? Kan je dat wel als je de koncentratie antigenen uitzet tegen de tijd? Schat de fout in $t = 4$ voor $h = h_0/8$. Is de schatting betrouwbaar? Maak het verschil of je naar $t = 9.2$ kijkt of naar $t = 2$? Kijk ook eens naar de oplossing voor t tussen 36 en 40 ($T = 40$). Wat valt je op? Is het geobserveerde fenomeen toe te schrijven aan numeriek instabiliteit? Wat verwacht je voor de stapgrootte h die je een voldoen nauwkeurig resultaat (2 cijfers) oplevert voor $T = 100$? Kan je dit “vervelend” gedrag ook zien in het fasevlak? Kan je dit verklaren?