“Basisboek Analyse”

Not another ε-book

Dutch one-page intro on $y = f(x)$ on second page

Between table of contents and the first chapter:

another Dutch page on the whole book;

yet another few Dutch pages on recent stuff:

Chapters 7, 8 and some rude material behind the index

Also new is a cross-referencing Chapter 1 on fundamentals in real context:

Archimedes, Babylon, Newton, Cardano, zero’s and roots

Old fundamenten hoofdstuk moved to the end: Hoofdstuk 26

JH&FF, Oegstgeest, Amsterdam (2018)

"Basisboek Analyse"

Not another ε-book about

$$y = f(x)$$

Een eerste cursus analyse omvat de differentiaal- en integraalrekening voor zeg maar wat op school $y = f(x)$ betreft. Die rekening kan niet zonder de middelwaardestelling (Sectie 4.1), die weer niet kan zonder de uitspraak dat een continue functie $f : [a, b] \to \mathbb{R}$ een maximum en een minimum heeft (Opgave 2.35). Daartoe volstaan Stelling 2.30 als *definitie* van continuiteit met behulp van convergente rijen en Stelling 3.1 over het bestaan van convergente deelrijen van begrensde rijen als *axioma* voor de verzameling $\mathbb{R}$ van de reële getallen.

De elementaire opgaven in Sectie 3.2 zijn vervolgens meer dan voldoende om Hoofdstuk 3 de verhaallijn van integraalrekening in Hoofdstuk 2 naar differentiaalrekening in Hoofdstuk 4 niet te veel te laten onderbreken. Via het bestaan van monotone deelrijen in Stelling 3.14 kan dit wat ongebruikelijke axioma gerelateerd worden aan de discussie over kleinste bovengrenzen in Sectie 26.2 aan de hand van de voorbeelden in Hoofdstuk 26 voor het speelkwartier begint in Sectie 26.1. Zo'n discussie zou iedere cursus Analyse vooraf moeten gaan.

Cauchy zal het misschien niet leuk vinden maar continuiteit en convergentie zijn zo hulpmiddelen van maar geen doel op zich in een cursus differentiaal- en integraalrekening voor exacte wetenschappers in spe, en passen mijns inziens als onderwerp veel beter in een module voor wiskundestudenten over metrische ruimten waarin $\mathbb{R}$ als prominent voorbeeld kan figureren.

Sectie 12.3, een oud Leids dictaatje, is zo'n module, waarin aan het eind ook het voor de daar maximale doelstelling nog niet benodigde begrip eindige deeloverdekkingen van open overdekkingen wordt behandeld. Dat begrip is pas nodig[1] voor de partities met *fading*[2] functions in de analyse op manifolds, zie Sectie 8.15 in het betreffende nieuwe Chapter 8 dat voorafgaand aan de inhoudsopgave van dit boek apart wordt besproken. I also added a historical chapter Hoofdstuk 1 in English to discuss some historical context[3] which may replace what is now Hoofdstuk 26 over wat sommigen nu als fundamenten menen te moeten zien. Meer beschouwingen na de inhoudsopgave.

---

[1] Zie ook het ruwe materiaal in Hoofdstuk 15 dat aan maattheorie vooraf kan gaan.

[2] https://www.youtube.com/watch?v=y8OrKLH8GBY

[3] Predating the story line I chose in Hoofdstuk 4 before Section 4.1.

The Netherlands

© 2016 tekst Joost Hulshof
© 2016 illustratie Ruud Hulshof

Fotografie omslag: nog onbekend
Vormgeving omslag: nog onbekend

ISBN
NUR

All rights reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written consent of the publisher.

De auteursinkomsten[4] uit dit boek komen ten goede aan de “geluksmachine”:

`http://orchestra18c.com`

[4] Als het ervan komt...

Inhoudsopgave

1 Archimedes, algebra and all that really **16**
- 1.1 Area in the Cartesian plane . . . 18
  - 1.1.1 Really? . . . 20
  - 1.1.2 Counting it? . . . 21
  - 1.1.3 The square root of two? . . . 22
  - 1.1.4 Elementary school representations of real numbers . . . 23
  - 1.1.5 More bounded increasing sequences . . . 23
- 1.2 Integrals now! . . . 24
- 1.3 Back to Babylonian times: YBC7289 . . . 27
- 1.4 Differential calculus founded in algebra . . . 29
- 1.5 The mystery of the tangent . . . 30
- 1.6 So what about continuity? . . . 30
  - 1.6.1 Quadratics . . . 31
  - 1.6.2 Cubics . . . 32
  - 1.6.3 Imagine what? . . . 32
  - 1.6.4 Minimise it! . . . 33
  - 1.6.5 Generalise it! . . . 35
  - 1.6.6 Estimates and order in the complex plane . . . 35
  - 1.6.7 In search of minima and minimisers . . . 36
  - 1.6.8 Conclusion . . . 38

2 Integraalrekening **39**
- 2.1 Integralen via onder- en bovensommen . . . 44
- 2.2 Toch maar wat meer over continuïteit en limieten . . . 51
- 2.3 Integreerbaar: definities, regels, voorbeelden . . . 54
- 2.4 Terug- en vooruitblik . . . 58
- 2.5 Verdiepingsmoment: tussensommen . . . 60

3 Convergente (deel)rijen en zo **62**
- 3.1 Spelen met kwantoren . . . 62
- 3.2 Elementaire opgaven over convergente rijen . . . 63
- 3.3 Limietpunten en convergente deelrijen . . . 64
- 3.4 Begrensde rijen continue functies . . . 66
- 3.5 Overaftelbaarheid van de reële getallen . . . 66
- 3.6 Directer bewijs voor de convergente deelrijstelling . . . 67
- 3.7 Onvoorwaardelijke convergentie . . . 69

4 Van integraal- naar differentiaalrekening **72**
4.1 De middelwaardestelling . . . 74
4.2 Integralen via primitieve functies . . . 76
4.3 Inverse functies . . . 79
4.4 Impliciete functies . . . 80
4.5 Differentieerbare impliciete functies . . . 86
4.6 Totale afgeleide, ketting- en Leibnizregel . . . 90
4.7 Stationair onder randvoorwaarde . . . 93
4.8 Zijstapje: convergentie van Newton's methode . . . 94
4.9 Integralen met parameters . . . 96
4.10 Partieel integreren en Taylorpolynomen . . . 98
4.11 Meer calculus: substitutieregels voor integralen . . . 102
4.12 Conclusie . . . 102
4.13 Wat uitwerkingen . . . 104

5 Rekenregels en voorbeelden **106**
5.1 Rekenregels voor de afgeleide . . . 107
5.2 Mooie voorbeelden: machtreeksen . . . 110
5.3 De natuurlijke logaritme . . . 115
5.4 Zijstapje: tetratie . . . 119
5.5 Meer differentiaalvergelijkingen . . . 120
5.6 Meer inverse functies . . . 123
5.7 Rare voorbeelden . . . 124
5.8 Stirling's formule via schalen en limieten . . . 125
5.9 Makkelijke convergentiestellingen . . . 129

6 Analyse met meer variabelen **131**
6.1 Zijstapje: de hoofdstelling van de algebra . . . 132
6.2 Differentiaalrekening, ook met complexe getallen . . . 134
6.3 Randvoorwaarden en de methode van Lagrange . . . 140
6.4 Toepassing: de ongelijkheid van Hölder . . . 144
6.5 Aftelbare sommen van vierkante matrices . . . 146
6.6 Impliciete functiestelling in Euclidische ruimten . . . 148
6.7 Flashback . . . 151
6.8 Karakterisaties van deelvarieteiten . . . 153
6.8.1 Niet per se lineaire deelvarieteiten . . . 154
6.8.2 Beelden (van randen) van bollen . . . 157
6.8.3 Overgang op andere coördinaten . . . 158
6.8.4 Meer dan continu differentieerbaar . . . 158
6.9 Integralen over rechthoeken . . . 159
6.9.1 Notationele kwesties met die d-tjes . . . 159

6.9.2 Formele d-algebra zonder betekenis? 161
6.9.3 Transformaties en parametrisaties 163
6.9.4 Een transformatiestelling 164

7 Standing at the crossroads **167**

8 Multilinear algebra and integration **171**
8.1 Local integrals with the normal at the boundary 173
8.2 The length of curve 176
8.3 Line integrals of vector fields along curves 177
8.4 Surface area 179
8.5 The transpose, quadratic forms and operator norms 180
8.6 Eigenvalues of compact linear symmetric operators 182
8.7 Singular values and measures of parallelotopes 184
8.8 Surface integrals 187
8.9 Integrals of functions over compact manifolds 188
8.10 More integration of differential forms 189
8.11 From Green's Theorem to Stokes' curl theorem 193
8.12 Pullbacks and the action of d 195
8.13 From Gauss' to Stokes' Theorem 199
8.14 More exercises 202
8.15 Cut-off functions and partitions of unity 209
8.16 Partitions of compact manifolds 210
8.17 Changing partitions 211
8.18 Again: local descriptions of a manifold 213
8.19 Coordinate transformations 215

9 Toepassingen met een zuiver toetje **217**
9.1 Integraalrekening in poolcoördinaten 217
9.2 Integraalrekening en kansverdelingen 219
9.3 Gradient, kettingregel, coördinatentransformaties 221
9.3.1 Gradient, divergentie en Laplaciaan 221
9.3.2 Kettingregel uitgeschreven voor transformaties 224
9.3.3 Kettingregel met Jacobimatrices 226
9.3.4 Omschrijven van differentiaaloperatoren 227
9.4 Harmonische polynomen 229
9.5 Intermezzo: het waterstofatoom 233
9.6 Lijnintegralen over polygonen en Coursat 234
9.7 Machtreeksen via een Cauchy integraalformule 240
9.8 De Cauchy Integraal Transformatie 244
9.9 Kromme lijnintegralen 245

9.10 Still to fit in: stable polynomials 248

10 Analyse in genormeerde ruimten **251**
10.1 Calculus voor X-waardige functies 253
10.2 Differentiaalrekening voor functies op X 256
10.3 Partieel differentieerbaar $\Longrightarrow$? 259
10.4 Differentiaal- en integraalvergelijkingen 262
10.5 Impliciete functiestelling: integraalvergelijkingen 267
10.6 Calculus in Banachalgebra's van operatoren 268

11 Extremisme **277**
11.1 Tweede afgeleide, Hessiaan, Lemma van Morse 277
11.2 Nog een keer de methode van Lagrange 285

12 Appendix: al of niet metrische topologie **287**
12.1 Metrische ruimten 287
12.2 Omgevingen, open en gesloten verzamelingen 289
12.3 Metrische ruimten; continue afbeeldingen 293

13 Terug naar het platte vlak **315**
13.1 Punten en vectoren in het platte vlak 315
13.2 Kortste afstanden 318
13.3 Vlakke meetkunde met het inprodukt 320
13.4 Projecteren op convexe verzamelingen 322
13.5 Andere inprodukten en bilineaire vormen 324
13.6 Om te onthouden 327
13.7 Poolcoördinaten in het (complexe) vlak 328

14 De Hilbertruimte in **330**
14.1 Standaardassenkruizen 331
14.2 Symmetrische matrices 333
14.3 Reële Hilbertruimten 334
14.4 De standaard Hilbertruimte 339
14.5 Riesz of geen Riesz? 342
14.6 Convergente deelrijen 347
14.7 Positieve eigenwaarden van wat? 348
14.8 The French Connection 352

15 Lebesgue's Differentiatie Stelling **354**

16 Functies op de cirkel **360**
16.1 Een Hilbert ruimte voor (periodieke) functies? 361
16.2 Standaard Hilbertruimten voor 'functies' 365
16.3 Fourierreeksen 367
16.4 Convergentie van Fourierreeksen 372
16.5 Dat andere inprodukt met afgeleiden 379
16.6 Blipfuncties 382
16.7 Intermezzo: out of Hilbertspace 384

17 A very partial reader for Olver's PDE book **386**
17.1 First order equations 388
17.1.1 The method of characteristics 389
17.1.2 Shocks, mass conservation, Rankine-Hugoniot condition 391
17.1.3 Appearance of the shock 392
17.1.4 ODE-system for the shock 393
17.2 The one-dimensional wave equation 394
17.3 Fourier series 395
17.4 The integral Fourier transform 398
17.5 The fast Fourier transform 400

18 Sobolev spaces **403**
18.1 Sobolev spaces 403
18.2 Estimates and embeddings for $W_0^{1,p}(U)$ 404
18.3 Compactness for $W_0^{1,p}(U)$ 405
18.4 Statements for $W^{1,p}(U)$ via extension 409
18.5 Density of smooth functions in $W^{1,p}(U)$ 410
18.6 The extension and trace operators 412
18.7 Additional exercises 414
18.8 Hölder continuity via Morrey's estimate 415
18.9 GNS-estimates 416

19 Evans' Chapter 6 and Navier-Stokes **419**
19.1 Existence of weak solutions via Lax-Milgramm 419
19.1.1 Weak solutions 419
19.1.2 The Lax-Milgramm Theorem 420
19.1.3 Lax-Milgramm; boundedness condition 422
19.1.4 Lax-Milgramm; coercivity 422
19.1.5 The general case with first order terms 423
19.2 The selfadjoint case 424
19.2.1 Second hand in homework set 424
19.2.2 Maximum principles 424

19.3 The Navier-Stokes equations 424
19.4 Navier-Stokes related exercises 426

20 Geostuff **430**
20.1 Submanifolds of $\mathbb{R}^d$ are Riemannian 430
20.2 Covariant differentiation 432
20.3 Tangent vectors as derivatives 433
20.4 Commutators of tangent vector fields 435
20.5 Covariant differentiation of tangent vectors 436
20.6 Second fundamental form 437
20.7 Curvature 437
20.8 Geodesic curves 439
20.9 The Jacobi equations 442

21 Newton's method the hard way **443**
21.1 Newton's method: a convergence proof 443
21.2 The optimal result 445
21.3 A suboptimal result 445
21.4 Alternative proof of convergence 446
21.5 The optimal alternative result 446
21.6 A suboptimal alternative result 447
21.7 A lousy alternative result 448
21.8 A much better suboptimal alternative result 448

22 Nash' modification of Newton's method **449**
22.1 The modified scheme 450
22.2 The new error term 451
22.3 The system of inequalities 452
22.4 Estimating the increments 453
22.5 Estimating the error terms 454
22.6 Sufficient conditions for a convergence result 456
22.7 Sufficient convergence condition on initial value 457
22.8 The optimal choice of parameters 458
22.9 Continuity 460

23 The Nash embedding theorem **461**

24 Airy scary **462**

25 Hartman-Grobman stelling **476**

26 Welke fundamenten? 481
26.1 Academisch speelkwartier: kolomcijferen 482
26.1.1 Optellen 487
26.1.2 Vermenigvuldigen? 491
26.1.3 Andere aftelbare sommen? 494
26.1.4 Een cijfer keer een kommagetal 496
26.1.5 Produkten van kommagetallen 497
26.2 Kleinste bovengrenzen 500
26.3 Absoluut convergente reeksen 502
26.4 Verzamelingen in de praktijk 503
26.5 Equivalentierelaties 506
26.6 Analyse in en van wat? 508

27 Welke horizon? 513

'I like fonctions of one variable'

Xavier Cabré adressing Abel prize winner Louis Nirenberg and a small analysis group at Tor Vergata in June 2015.

Denk aan $y = f(x)$ maar dan iets algemener. Dit vak gaat over integraal- en differentiaalrekening voor $f : \mathbb{R} \to \mathbb{R}$, met generalisaties naar $f : X \to \mathbb{R}$, $f : \mathbb{R} \to X$, $f : X \to Y$, bijvoorbeeld met $X = Y = \mathbb{R}^2$.

Na een inleidend overzicht in Hoofdstuk 26 of Chapter 1 beginnen we met integraalrekening in Hoofdstuk 2, waarin convergente deelrijen nodig zijn om integreerbaarheid van continue functies te bewijzen. Na Hoofdstuk 3 over die deelrijen en wat je zeker wil weten over absoluut convergente reeksen, volgt dan de link met (en) differentiaalrekening in Hoofdstuk 4 voor functies $f : \mathbb{R} \to \mathbb{R}$, waaronder ook impliciete functies, met de bijbehorende stelling in Hoofdstuk 4, *voor* de kettingregel. Het bewijs van de impliciete functiestelling doen we in detail en is gebaseerd op een contractieargument dat nog vaker zal terugkomen. Daarna zijn er wat inkoppertjes, bijvoorbeeld de multiplicator(en) van Lagrange, waarvan de resultaten later twee keer overgeschreven[5] kunnen worden. Rekenregels, voorbeelden en toepassingen in Hoofdstuk 5 sluiten wat je Analyse 1 zou kunnen noemen af.

Integraalrekening leidt in Hoofdstuk 2 al tot een eerste Banachruimte ruimte X, $X = C([a, b])$ de verzameling van continue functies $f : [a, b] \to \mathbb{R}$ op een begrensd gesloten interval $[a, b]$. Met de goede versie van de kettingregel in de bagage kan de lezer in principe Hoofdstuk 5 en Hoofdstuk 6 overslaan en zich in Hoofdstuk 10 de benodigde analyse met X i.p.v. $\mathbb{R}$ eigen maken om de toepassingen op differentiaal- en integraalvergelijkingen met parameters deels zelf in te vullen. Waar je een college Reële Analyse mee zou kunnen beginnen, met in Hoofdstuk 11 een (hele) kleine aanzet tot een vak Variatierekening.

Daarvoor is er nog een aanzet tot wat meestal Analyse 2 heet. Aan het begin van Hoofdstuk 6 een kort bewijsje van de hoofdstelling van de algebra, en dan verder met vooral differentiaal- en een beetje integraalrekening in meer variabelen, waarbij stellingen en bewijzen vooral overgeschreven worden uit de eerste hoofdstukken en de betekenis duidelijk wordt bij het uitpakken van de notatie. Eerst voor functies van $\mathbb{C}$ naar $\mathbb{C}$, en dan voor afbeeldingen van $\mathbb{R}^n$ naar $\mathbb{R}^m$. Verschillende perspectieven en toepassingen komen aan de orde. Als afsluiting wordt de complexe structuur direct en zonder kromme poespas gebruikt om met lijnintegralen over lijnstukjes te laten zien dat complex differentieerbare functies (zonder enige extra aannames) geweldige mooie eigenschappen hebben, denk ook aan $f(A)$ met $A \notin \mathbb{C}$.

[5] Uitgangspunt: alles voor $f : \mathbb{R} \to \mathbb{R}$ op zo'n manier dat het naadloos generaliseert.

Nog een woord na het lezen van een geschiedenisboek

Ik verwijs af en toe naar het boekje Wiskunde in je Vingers met Ronald Meester [HM]. Het verwerken van reacties van collega's (opdrachtgever Bas Edixhoven, Jan Wiegerinck, Bob Rink, René Swarttouw) heb ik in de zomer van 2017 afgewisseld met het lezen in [Eves], de zesde editie van An Introduction to the History of Mathematics van (vader) Howard Eves (en zoon Jamie voor de cultural connections). Ik verwijs daarom af en toe naar [Eves]. Verder is ook commentaar van studenten van de Mastermath zij-instroomcursus Analyse voor leraren zoveel mogelijk verwerkt. Ik verwijs ook naar [BE], het Leidse dictaat:

`https://drive.google.com/file/d/0B_AYW-l1PhK-TVo0X2FVNWpYdGM/view`

Hoofdstuk 26 van het huidige boek is bedoeld als enerzijds een alternatieve beschrijving voor de fundamenten waar we op bouwen in de analyse. Dat betreft met name de verzameling van de reële getallen, de eigenschappen van deze verzameling, en de bewerkingen met elementen ervan[6]. Anderzijds beschrijft Hoofdstuk 26 ook al wat van de inhoud die komen gaat, in samenhang met die fundamenten van de wiskunde. Chapter 1 offers the alternative approach I would prefer now.

Hoofdstuk 2 begint de facto bij Archimedes in een nu veel uitgebreidere inleiding die aan Sectie 2.1 vooraf gaat. De besproken formules voor sommen van machten zouden voorkennis voor dit vak kunnen zijn. Voor bewijzen (met volledige inductie) van die formules zie eventueel ook Hoofdstuk 6 van [HM].

Hoofdstuk 3 was een tussenhoofdstuk met twee opties voor het behandelen van de stelling over convergente deelrijen van begrensde rijen die nodig was in Hoofdstuk 2 om de Riemann integreerbaarheid van continue functies te bewijzen, maar is nu uitgebreid tot een algemene behandeling van het convergentiebegrip voor rijen en reeksen, met elementaire opgaven in een nieuwe Sectie 3.2.

In Hoofdstuk 4 vat Stelling 4.12 nu samen hoe differentiaal- en integraalrekening samenhangen. Sectie 4.1 over de middelwaardestelling is uitgebreid met opgaven die hiertoe dienend zijn. Sectie 4.4, de moeilijkste sectie van het hoofdstuk over de impliciete functiestelling, heeft nu een inleiding die het verband legt met de inverse functiestelling die als speciaal geval besproken wordt (en wellicht beter is om eerst te doen), en een nodeloos technische tussenformulering is geschrapt.

Nieuw[7]: het outlook Chapter 7 en de behandeling van integralen over manifolds die daarna volgt. Nummering schuift op. Achter de nog niet bijwerkte index ruw materiaal voor diverse doeleinden.

[6] Meningen zijn soms verdeeld of een cursus Fundamenten Analyse vooraf moet gaan.

[7] Ad hoc geschreven tijdens het geven van een college aan 2-dejaars.

About Chapter 8

Hoofdstuk 8 over multilineaire algebra en lijn- en oppervlakteintegralen is nieuw, lees het stukje van formule (8.1) en de eerste sectie van Hoofdstuk 8 met de lokale vorm van de analysestellingen waar het om gaat, gevonden na de vraagstelling met betrekking tot (8.7) over wat er komt uit de integraal van een partiele afgeleide over een begrensd gebied in (eerst maar) $\mathbb{R}^2$ en het antwoord in (8.9).

In dat antwoord (8.9) verschijnt voor het eerst de normaalvector op de rand van het gebied, en de manier waarop je die over de rand integreert met de technieken in Sectie 8.8: eerst in het speciale[8] geval dat $\mathrm{N} = 2$ en de parameterisatie van de vorm (8.36) is met $n = 1$.

De stap van lokaal naar globaal kan via de introductie van de fading[9] functions al informeel gemaakt worden om (8.10) in ieder geval al te noemen met als speciaal geval de eerste integration by parts formule (8.11). Voor de speciale fading function in (8.11) is de stap van lokaal naar globaal niet eens nodig.

Na Sectie 8.1 komen Secties 8.2,8.3 over lijnintegralen. Sectie 8.8 over integralen van functies over beelden van blokken onder parametriserende afbeeldingen $\Phi : \mathbb{R}^n \to \mathbb{R}^{\mathrm{N}}$ kan pas komen na Stelling 8.4 over maten van paralellelotopen en de daarmee verband houdende voorkennis uit de (multi)-lineaire algebra.

Secties 8.5, 8.6 (met deze vorm van de spectraalstelling[10]) en 8.7 passen goed in een college Lineaire Algebra, in een module over de algemene stelling van Pythagoras, te beginnen met de vraag naar de oppervlakte van een parallellelogram in $\mathbb{R}^3$.

Een alternatieve versie van Sectie 8.8 begint gewoon met Sectie 8.4, en heeft eventueel Sectie 8.2 als deelsectie over lijnintegralen, waarvoor een rechtstreeks aanpak met Riemannsommen[11] en reparametrisatie naar booglengte[12] mogelijk is.

De stap van lokaal naar globaal is met Secties 8.9, 8.15, 8.16, 8.17, 8.18, 8.19 nog niet helemaal compleet omdat we ook over manifolds met rand willen integreren, teneinde in Sectie 8.13 de algemene stelling van Stokes over integralen over manifolds van niet alleen dimensie $n = \mathrm{N} - 1$ maar ook

[8] Zie Sectie 9.6 en verder voor een polygonale aanpak in complexe context.
[9] Cut-off functions.
[10] In inprodruktruimten!
[11] Zoals nu alleen genoemd aan het begin van Sectie 8.2.
[12] Nu alleen genoemd aan het eind van Sectie 8.2.

$n \leq \mathrm{N} - 2$ in $\mathbb{R}^{\mathrm{N}}$ te formuleren. Alleen Sectie 8.9 is onderdeel van de flow in het hoofdstuk. De andere secties zijn naar achteren geschoven.

Het ligt voor de hand om daarvan eerst Sectie 8.15 over partities van 1 te doen, dat sluit aan bij de appendix van Sectie 12.3. over kompaktheid met open overdekkingen.

Ik kan me goed voorstellen dat de stap van lokaal naar globaal pas helemaal achteraf gedaan wordt en dat eerst de differentiaalvormen worden geintroduceerd, en voorlopig heb ik daar nu voor gekozen.

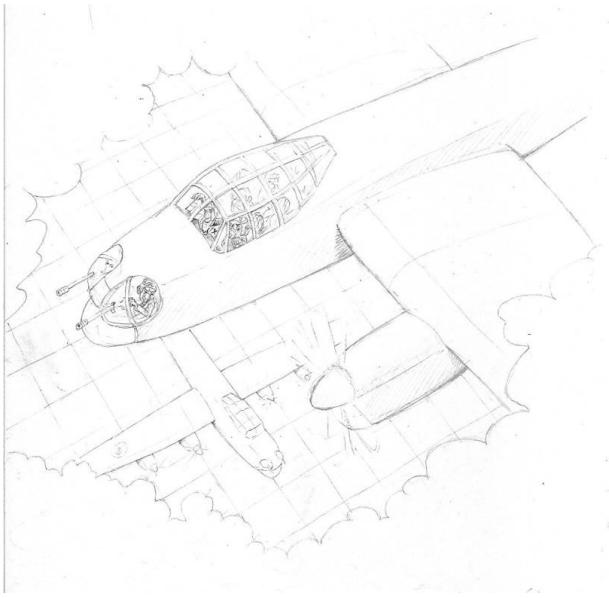
De stap van integralen met dS-notatie naar integralen met produkten van dx_i-tjes in differentiaalvormen begint met Sectie 8.10, met aan het eind de lijnintegralen uit Sectie 8.3, waarin ik voor het eerst[13] differentiaalvormen benoem. De subsecties van Hoofdstuk 6.9 zijn hier relevant.

In Sectie 8.14 (met leerzame en besproken sommen nu) verwijs ik naar (8.34) en (8.32). Er moet nog een stukje bij over stuksgewijs gladde randen[14], en wel naar aanleiding van de failure besproken na Opgave 8.11, maar wat er daarna komt vond ik voor nu leuker: het folium van Descartes voor algemene dimensie n en exponent p, via eerst de impliciete functiestelling in Opgave 8.13 en dan Opgave 8.13, resulterend in formules voor oppervlakte en inhoud in (8.89) en Opgave 8.17.

Tenslotte heb ik nauwelijks iets gezegd over de werking van differentiaalvormen op vectorvelden. Dat kan volgens mij het best gebeuren in samenhang met de fysische interpretaties die tot nu toe in dit college nooit aan de orde kwamen denk ik. Maar zie ook Hoofdstuk 20 en het plan in Hoofdstuk 23.

[13] In dit Hoofdstuk 8 althans.

[14] Met wat overwegingen over hoe het zou gaan met $\partial\Omega$ alleen maar Lipschitz continu.



De tekening hierboven is van mijn broertje, die ook de 11 tekeningen in het groene boekje met Ronald Meester maakte. Dit is nummer 12. Met 11 hoofdstukken bleef er eentje over die illustreert wat er op 3 maart 1945 in het Bezuidenhout in Den Haag gebeurde. De bommen kwamen niet op de Duitse lanceerinstellingen in het Haagse Bos terecht maar op woonwijken. De school van mijn moeder, vlak bij de hoge toren tussen de A12 en die laan waar NWO tegenwoordig zit, bleef net gespaard.

De Nederlandse wikipedia pagina over het bombardement op Den Haag is de moeite waard, en vermeldt het verwisselen van coördinaten. De tekening illustreert dit met een x en een y in de cockpit, en met vliegtuigen die in richtingen loodrecht op elkaar vliegen. Het is altijd leuk om in woorden een idee te beschrijven en te kijken wat de tekenaar er van bakt. Helaas paste deze tekening niet in het format van het groene boekje.

De keuze van letters als namen voor wiskundige objecten is verder een bron van leerzaam vermaak. Bij het introduceren van nieuwe stof is het een goede gewoonte om systematisch te zijn in de keuze van de letters, maar vroeg of laat moet je de stap maken om flexibel te zijn en letters en notaties te veranderen, al is het alleen maar om te leren de notaties van anderen te begrijpen. Sterker, andere vakgebieden gebruiken soms notaties die je vanuit een star en strict wiskundig standpunt als fout zou moeten duiden. Het is soms wat contraproductief om dat dan vervolgens ook altijd maar te doen.

Dit boek is opgedragen aan het thuisfront dat voor de klas en in de strijd voor beter onderwijs haar mannetje staat.

1 Archimedes, algebra and all that really

Forget about the math you had if you can. Some 23 centuries ago Archimedes knew that

$$1+4+9+16+25+36+49+64+81+100 = \frac{1000}{3}+\frac{100}{2}+\frac{10}{6}.$$

On the left you see a calculation with positive integers, how's that for a start? On the right you see a sum of positive fractions, that you should think of as rationals, because they involve ratio's, the ratio being that dividing positive integers we encounter noninteger positive numbers that we often think of as ratio's.

The positive integers are the numbers 1, 2, 3, 4, etc, aka the natural numbers, which together make up what mathematicians call the set $\mathbb{N}$ of all natural numbers. We say that 1 belongs to $\mathbb{N}$, notation $1 \in \mathbb{N}$, and with $1 \in \mathbb{N}$ also $2 = 1+1 \in \mathbb{N}$, $3 = 2+1 \in \mathbb{N}$, $4 = 3+1 \in \mathbb{N}$, and so on. This is not a course about the undeniable joys of set theory, so we simply write

$$\mathbb{N} = \{1, 2, 3, 4, \dots\},$$

a notation that is of course dubious to some[1] of us mathematicians.

The sum of the first 10 squared natural numbers is another natural number, 385, the product of the primes 5, 7 and 11. I take it for granted that every natural number is a product of primes. This is a course about analysis and not about number theory.

A modern notation for the curious sum above is

$$(C_{10}) \qquad \sum_{k=1}^{10} k^2 = \frac{10^3}{3}+\frac{10^2}{2}+\frac{10}{6},$$

in which k is an index that runs from 1 to 10. It's a dummy index, you can replace it by any other letter in any alphabet you like. I've labelled this sum by (C_{10}), the C is for curious, and the 10 is obvious.

[1] https://www.youtube.com/watch?v=4bez05AT1u4

On the right you recognise 10 again, one of your favorite natural numbers: you should learn to count to 10 as soon you can:

$$1, 2, 3, 4, 5, 6, 7, 8, 9, 10!$$

Be careful with ! because 10-factorial is defined by and written as

$$10! = 1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8 \times 9 \times 10 = \prod_{k=1}^{10} k = 3628800,$$

the *product* of the first 10 natural numbers, which happens to be 6 weeks in seconds, whereas 6 weeks in days is 42, the answer to life, the universe and everything really, for which no nice algebraic formula exists.

The *sum* of the first 10 natural numbers is of course

$$\sum_{k=1}^{10} k = \frac{10^2}{2} + \frac{10}{2} = \frac{1}{2} \times 11 \times 10,$$

the young Gauss knew that, but Archimedes was (perhaps not) the first, replacing 10 by any other natural number 11, 12, 42, 10^{10}, $10^{10^{10}}$ or whatever[2], to consider such sums for arbitrarily large n. Long before Gauss was young Archimedes knew that[3]

$$\sum_{k=1}^{n} k = \frac{n^2}{2} + \frac{n}{2}, \quad \sum_{k=1}^{n} k^2 = \frac{n^3}{3} + \frac{n^2}{2} + \frac{n}{6}, \quad \sum_{k=1}^{n} k^3 = \frac{n^4}{4} + \frac{n^3}{2} + \frac{n^2}{4},$$

lovely sums from which a pattern emerges.

The one in the middle is

$$(C_n) \qquad \sum_{k=1}^{n} k^2 = \frac{n^3}{3} + \frac{n^2}{2} + \frac{n}{6}.$$

You may prove (C_n) for every natural number by first guessing[4] it and then showing that starting with $n = 1$ and (C_1) being a statement that is trivially true, it holds that

$$(C_n) \implies (C_{n+1}),$$

by evaluating

$$\sum_{k=1}^{n+1} k^2 = \frac{n^3}{3} + \frac{n^2}{2} + \frac{n}{6} + (n+1)^2$$

[2] Have a look at Section 5.4 when you have some time.

[3] I'm using search and replace to substitute $10 = n$ in our modern notation.

[4] Guess what, you should learn how to guess.

and verifiying you then get what you have guessed[5], since it happens be the case that

$$\frac{n^3}{3}+\frac{n^2}{2}+\frac{n}{6}+(n+1)^2=\frac{(n+1)^3}{3}+\frac{(n+1)^2}{2}+\frac{n+1}{6}.$$

This proves a merry little but supernatural theorem[6]: For every natural number (C_n) holds, as a rearrangement of larger and larger sums of squared natural numbers into simple 3-term sums of supernatural numbers.

Note that k is still the dummy, but n is a variable natural number that you can take as large as you want, because a number bounding all the natural numbers is simply not real. In particular the sums get larger with n, not as fast as $n!$ of course, but still.

However, dividing the sum of the first n squares by n-cubed, you get

$$\frac{1}{n^3}\sum_{k=1}^{n}k^2=\frac{1}{3}+\frac{1}{2n}+\frac{1}{6n^2}=\frac{1}{n}\sum_{k=1}^{n}\left(\frac{k}{n}\right)^2,$$

in which $2n=2\times n$, $6n^2=6\times n^2$, as we do away with the symbol $\times$ in multiplications with variable numbers[7] as soon as we leave elementary school. Now you see that for n larger and larger, you get a sum that is closer and closer to $\frac{1}{3}$, a number well known to the Egyptians in the days of the pyramids that relates to the context presented next[8]. Pyramids later found themselves an application as learning gadgets, see

`https://www.aob.nl/nieuws/het-is-gewoon-crap-bullshit/`

.

1.1 Area in the Cartesian plane

Draw yourself the nice Cartesian xy-plane[9] that Archimedes never had, in which you make a big sketch by hand of the parabola consisting of all points with equal distances to the point $(0,\frac{1}{4})$ and to the line defined by $y=-\frac{1}{4}$. This parabola is in fact the curve defined by $y=x^2$, and you may enjoy verifying that in particular $(1,1)$ fits its geometric description[10], before you go on.

[5] Or discovered via tables of squares, cubes and differences, see Opgave 2.5.

[6] This theorem predates set theory by at least two millennia.

[7] Denoted by letters.

[8] I actually don't know if Archimedes ever considered this particular context.

[9] You were likely not able to forget about that plane anyhow, let alone about x and y.

[10] Since $3^2+4^2=5^2$.

Now evaluate $y = x^2$ with

$$0 = \frac{0}{10}, \frac{1}{10}, \frac{2}{10}, \frac{3}{10}, \frac{4}{10}, \frac{5}{10}, \frac{6}{10}, \frac{7}{10}, \frac{8}{10}, \frac{9}{10}, \frac{10}{10} = 1.$$

These particular x-values give you points in the unit square S, the square consisting of all points (x, y) with $0 \leq x \leq 1$ and $0 \leq y \leq 1$. Look at the region A_2 in S bounded by $y = 0$, $x = 1$ and $y = x^2$, and let us agree that the area of S is equal to 1. Can you convince yourself that consequently the area of A_2 is less than

$$\frac{1}{1000} + \frac{4}{1000} + \frac{9}{1000} + \frac{16}{1000} + \frac{25}{1000} + \frac{36}{1000} + \frac{49}{1000} + \frac{64}{1000} + \frac{81}{1000} + \frac{100}{1000},$$

but more than

$$\frac{1}{1000} + \frac{4}{1000} + \frac{9}{1000} + \frac{16}{1000} + \frac{25}{1000} + \frac{36}{1000} + \frac{49}{1000} + \frac{64}{1000} + \frac{81}{1000}?$$

To do so you have to relate every term in the sums above to the area of a rectangle[11] with 4 corner points, one of which is the point on the curve. There are two obvious ways to do so, providing you with the above (upper and lower) sums[12] for the area of A_2.

If this worked, you can also convince yourself, denoting the area of A_2 by $|A_2|$ from now on, that

$$|A_2| < \frac{1}{n^3} \sum_{k=1}^{n} k^2 = \frac{1}{3} + \frac{1}{2n} + \frac{1}{6n^2} < \frac{1}{3} + \frac{1}{n}$$

for every natural number n, and likewise that

$$|A_2| > \frac{1}{n^3} \sum_{k=1}^{n-1} k^2 = \frac{1}{3} + \frac{1}{2n} + \frac{1}{6n^2} - \frac{1}{n} = \frac{1}{3} - \frac{1}{2n} + \frac{1}{6n^2} > \frac{1}{3} - \frac{1}{n}.$$

Combining it follows that

$$\frac{1}{3} - \frac{1}{n} < |A_2| < \frac{1}{3} + \frac{1}{n},$$

whence

$$-\frac{1}{n} < |A_2| - \frac{1}{3} < \frac{1}{n}$$

[11] Parallel to the axes.

[12] Riemann sums in fact.

for every natural number n, and that will imply that

$$|A_2| - \frac{1}{3} = 0 \quad \text{whence} \quad |A_2| = \frac{1}{3},$$

won't it?

Think about it, but think[13] about pyramids first, and see how the layered structures of the voluminous worn off pyramids of Cheops and Chefren consist of squares that were bound to lead to sums Riemann would have liked.

1.1.1 Really?

A number a with

$$-\frac{1}{n} < a < \frac{1}{n}$$

for every natural number simply can't be different from 0, can it? Otherwise it is either positive, in which case

$$0 < a < \frac{1}{n},$$

or negative, in which case

$$0 < -a < \frac{1}{n},$$

and positive numbers with $a < \frac{1}{n}$ for every natural number n cannot be real, because that would imply that

$$n < M = \frac{1}{a}$$

for all natural numbers n, something you thought unreal probably[14]. No real number is an upperbound for all the positive numbers.

How's that? Nowadays mathematicians[15] think of the set

$$\mathbb{N} = \{1, 2, 3, 4, \dots\}$$

as a subset of the set $\mathbb{R}^+$, whose elements are all the positive real numbers, whatever real may be. The set $\mathbb{R}^+$ in turn is a subset of the set $\mathbb{R}$ of all real numbers. This set can be introduced in many ways[16], but it most certainly contains $\mathbb{N}$ and all the numbers obtained from $\mathbb{N}$ via subtraction

[13] Don't think of activities.
[14] https://www.youtube.com/watch?v=V8l7ErLb-qM
[15] But who are they?
[16] Read all about it in Dutch: Chapter 26.

and division, such as $0 = 1 - 1$, $-1 = 1 - 2$, the sometimes forbidden[17] number $\frac{1}{3}$ as 1 divided by 3, and so on.

These numbers derived from the natural numbers by algebra are called the rational numbers, and the set whose elements are all the rational numbers is denoted by $\mathbb{Q}$, which contains the set

$$\mathbb{Z} = \{\ldots, -4, -3, -2, -1, 0, 1, 2, 3, 4, \ldots\}$$

whose elements are all the integers, positive, negative, and the zero number that separates them. You may think of $\mathbb{Z}$ as a bi-infinite sequence of marked points on a line with no endpoints, the other numbers in $\mathbb{Q}$ lying in the intervals between.

1.1.2 Counting it?

It is easy to enumerate $\mathbb{Q}$. Just take a path in your Cartesian xy-plane that starts in $(0, 1)$ and then passes through every point (x, y) with $y \in \mathbb{N}$ and $0 \neq x \in \mathbb{Z}$. Along this path $\frac{x}{y}$ runs through the whole of $\mathbb{Q}$, with multiple hits of course[18], but skipping the fractions $\frac{x}{y}$ with common factors in numerator and denominator, every element of $\mathbb{Q}$ occurs precisely once in an enumeration of fractions thus obtained. It is thereby clear that the set $\mathbb{N}$ is in a one to one correspondence with the set $\mathbb{Q}$.

You really should convince yourself that such a one to one correspondence between $\mathbb{N}$ and the set of all points on the number line via an enumeration

$$x_1, x_2, x_3, \ldots$$

is unreal. Such an enumeration would allow you to completely cover the number line with intervals[19]

$$(x_1 - \frac{1}{4}, x_1 + \frac{1}{4}),$$

$$(x_2 - \frac{1}{8}, x_1 + \frac{1}{8}),$$

$$(x_3 - \frac{1}{16}, x_3 + \frac{1}{16}),$$

$$(x_4 - \frac{1}{32}, x_4 + \frac{1}{32}),$$

[17] In Dutch educational law.

[18] In view of $\frac{1}{3} = \frac{2}{6} = \frac{3}{9} = \cdots$ for instance.

[19] For numbers $a < b$ we denote by (a, b) the set of all numbers x with $a < x < b$.

$$(x_4 - \frac{1}{64}, x_4 + \frac{1}{64}),$$
$$(x_4 - \frac{1}{128}, x_4 + \frac{1}{3128}),$$
$$(x_4 - \frac{1}{256}, x_4 + \frac{1}{256}),$$
$$(x_4 - \frac{1}{512}, x_4 + \frac{1}{512}),$$

et cetera, the total length of these intervals being

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \frac{1}{32} + \frac{1}{64} + \frac{1}{128} + \frac{1}{256} + \cdots \leq 1,$$

an absurdity that we are not quite willing to accept[20].

1.1.3 The square root of two?

In $\mathbb{Q}$ we have the usual operations $+, -\times, /$ with the usual rules. You should employ your knowledge of those natural numbers to prove that the magic number $\sqrt{2}$ cannot belong to $\mathbb{Q}$. That is to say, $\mathbb{Q}$ does not contain a positive number r with $r^2 = 2$. Every $q \in \mathbb{Q}$ either has $q^2 > 2$ or $q^2 < 2$, and it is not so difficult to generate the increasing sequence of numbers

$$q_1 = 1 < q_2 = 1.4 < q_3 = 1.41 < q_4 = 1.414 < q_5 = 1.4142 < q_6 = 1.41421$$
$$< q_7 = 1.414213 < q_8 = 1.4142135 < q_9 = 1.41421356 < \cdots < 2,$$

2 being an upper bound because of the 1 in front of every decimal point of every single number in this sequence.

Every new digit in every next number in this sequence is chosen as large as possible to have the square below 2. Of course the resulting sequence of digits can be only be generated by a machine that runs a suitable programme till the very end of time, but still we can agree to say that the number r is defined by putting all those digits in an never-ending expansion

$$r = 1.41421356\ldots,$$

even if it is impossible to write all these damned digits down.

Clearly the number r is then larger than all *finitely represented* expansions in the increasing sequence $q_1 < q_2 < q_3 < \ldots$, and r is thus an upper bound below the crude upper bound 2. You should try to convince yourself that there is actually no lower upper bound than r, making r the lowest upper bound for the sequence.

[20] A strict inequality $<$ in this last estimate may be something you thought unreal.

1.1.4 Elementary school representations of real numbers

A proper definition of the multiplication of such numbers represented by never-ending decimal expansions then must yield that

$$r^2 = 2,$$

a possible convention being that 2 is represented as

$$2 = 1.9999999\cdots = 1 + \frac{9}{10} + \frac{9}{10^2} + \frac{9}{10^3} + \cdots,$$

excluding expansions which only have finitely many nonzero digits. In particular $1 \in \mathbb{N}$ should be considered as represented in $\mathbb{R}$ by

$$1 = 0.9999999\ldots,$$

thereby having *all* real numbers represented by never-ending sequences[21] of digits, with a decimal point that separates the integer part of the number from its remainder[22].

All this can be done, but is somewhat cumbersome[23] in base 10. Binary representations with base $2 = 10$ do a much beter job here. With

$$1 = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \cdots$$

represented by the bits $1, 1, 1, 1, 1, 1\ldots$, and a zero before the binary point in

$$1 = 0.11111111\ldots,$$

you may enjoy the details of the resulting construction of the real numbers and their arithmetic[24].

1.1.5 More bounded increasing sequences

Anyhow, you can now consider bounded sequences

$$r_1 < r_2 < r_3 < \cdots$$

of such numbers.

With the proper definition of $<$ it is not hard to then construct the lowest possible upper bound of such a sequence as such a number. The existence

[21] Why can such representations not be enumerated? In Dutch: Section 3.5.
[22] If you like after division by 1.
[23] But quite amusing, see Section 26.1 on what might be called kolomcijferen in Dutch.
[24] https://www.youtube.com/watch?v=x9DejWMJoOA

of such lowest upper bounds is a crucial property of the set $\mathbb{R}$, either via the construction above, or by any other construction: the real numbers[25] really have the property that every strictly increasing bounded sequence has a lowest upper bound.

It is this property of the real numbers really then that proves[26] to mathematicians that $\mathbb{N}$ is not bounded by any real number, for that would imply the existence of a smallest $S \in \mathbb{R}$ with $n \leq S$ for all $n \in \mathbb{N}$. But then $S - \frac{1}{2}$ is not an upper bound whence $n \in \mathbb{N}$ exists with $n > S - \frac{1}{2}$ and thus the number $n + 1 \in \mathbb{N}$ satisfies $n + 1 > S + \frac{1}{2} > S$ contradicting S being an upper bound for $\mathbb{N}$, as you must have thought of as being unreal from the very beginning[27], since you've used the smallest upper bound property ever since elementary school: did you not take it for granted that

$$0.333333333\ldots$$

defines a real number, which happens to be a rational equal to $\frac{1}{3}$ really, the area $|A_2|$ of the set A_2 introduced above?

1.2 Integrals now!

Of course you've already been told that $|A_2| = \frac{1}{3}$, via

$$|A_2| = \int_0^1 x^2\,dx = \Big[\frac{x^3}{3}\Big]_0^1 = \frac{1}{3},$$

but that's based on differential calculus, which did not exist in the days of Archimedes. Summing up, based on the properties of the real numbers, including the Archimedian property that below every positive number there exits a $\frac{1}{n}$ with $n \in \mathbb{N}$, we have shown that

$$\int_{x\in[0,1]} x^2\,dx = \int_0^1 x^2\,dx = \frac{1}{3},$$

if we define and compute the integral of x^2 from 0 to 1 via lower and upper sums as above, with n arbitrarily large.

As for notation, using the values of x^2 for

$$x \in [0,1]_n = \big\{\frac{k}{n} : k = 1, 2, \ldots, n\big\} \quad \text{and setting} \quad \Delta_n x = \frac{1}{n},$$

[25] You may wish to reflect on how to state that there is only one set of real numbers.
[26] More in Dutch here: Section 26.2.
[27] Some progress report, is this the beginning or the end?

the upper sums rewrite as

$$\frac{1}{n}\sum_{k=1}^{n}\left(\frac{k}{n}\right)^2 = \sum_{x\in[0,1]_n} x^2\,\Delta_n x,$$

not so different from the notation for the integral. It's really just

$$\int_{x\in[0,1]} x^2\,dx \quad \text{versus} \quad \sum_{x\in[0,1]_n} x^2\,\Delta_n x,$$

the latter being a finite sum, the former having all subscripts n dropped, Σ replaced by the curly $\int$, and $\Delta_n x$ by dx.

An almost immediate consequence now is that it follows that

$$\int_0^b x^2\,dx = \frac{b^3}{3},$$

by similar definition and computation, with all the sketches you may have drawn copied with b replacing 1 on the x-axis and b^2 replacing 1 on the y-axis, multiplying every term in the lower and upper sums by b^3. Moreover, the pattern you saw emerging replacing the exponent 2 by $3, 4, 5, \ldots$, can and should be proved outside the domain of analysis.

It thus follows that

$$\int_0^b x^n\,dx = \frac{b^{n+1}}{n+1}$$

for all $b \in \mathbb{R}^+$ and $n \in \mathbb{N}$. Likewise for any $0 \le a < b$, a similar definition leads[28] to

$$\int_a^b x^n\,dx = \frac{b^{n+1}}{n+1} - \frac{a^{n+1}}{n+1},$$

and such identities and the inequalities that prove them combine. Zum Beispiel, with

$$p_n(x) = 1 + x + x^2 + \cdots + x^n \quad \text{and} \quad P_n(x) = x + \frac{x^2}{2} + \cdots + \frac{x^{n+1}}{n+1},$$

it holds that

$$\int_0^1 p_n(x)\,dx = P_n(1) - P_n(0),$$

i.e.

$$\int_0^1 \left(1 + x + x^2 + \cdots + x^n\right)\,dx = 1 + \frac{1}{2} + \cdots + \frac{1}{n+1}.$$

[28] Subtraction may also convince you.

It really all adds up, also if we first multiply by positive scalars. With

$$E_n(x) = 1 + x + \frac{x^2}{2} + \frac{x^3}{2 \times 3} + \cdots + \frac{x^n}{n!}$$

for instance we have[29]

$$\int_0^1 E_n(x)\, dx = E_{n+1}(1) - E_{n+1}(0),$$

i.e.

$$\int_0^1 \left(1 + x + \frac{x^2}{2} + \frac{x^3}{2 \times 3} + \cdots + \frac{x^n}{n!}\right) dx = 1 + \frac{1}{2!} + \cdots + \frac{1}{(n+1)!},$$

and the question about what will hold for

$$p(x) = 1+x+x^2+\cdots+, \quad P_n(x) = x+\frac{x^2}{2}+\cdots, \quad E(x) = 1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\cdots$$

is immediate. Does

$$1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \cdots$$

define a positive number for instance?

Never-ending summings and integrals, we can define and compute them guided by what Archimedes knew long before differential calculus was invented, with a complete theory for expressions of the form

$$f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 + \cdots,$$

in which x is a positive real variable and a_0, a_1, a_2, a_2 real nonnegative coefficients[30]. Of course we also consider negative x and negative $a_0, a_1, a_2, a_2, \ldots$, not much of a worry, but calling for some care[31]. All this can be done before even worrying about the natural domain and properties of the map

$$f \to \int_a^b f(x)dx$$

as defined using the sums of Riemann. The worries are minor, starting in Section 2.1 with first Stelling 2.20 and then Stelling 2.23, based on the natural concept of uniform continuity in Defintie 2.22.

[29] Pay attention to the constant 1 in $E_{n+1}(1)$ and $E_{n+1}(0)$.

[30] See Section 5.9.

[31] Read Section 26.3 in Dutch to get a feel.

1.3 Back to Babylonian times: YBC7289

So what about differential calculus? You may have heard of Newton and Newton's method for a calculation of algebraic numbers like $\sqrt{2}$ which is more efficient then the generation of digits by trial and error. The first recorded attempt to find a positive r with $r^2 = 2$ that I know of can be found on YBC7289, dating back to some 40 centuries ago, using calculations carried out in base 60, long before the distinction between $\mathbb{Q}$ and $\mathbb{R}$ as we just discussed it.

These calculations are now believed to employ the approximation

$$\sqrt{1+x} \approx 1 + \frac{x}{2}$$

for x close to 0. The right hand side is an expression which you may relate to the line defined by $y = 1 + \frac{x}{2}$ in your Cartesian plane, and $y = \sqrt{1+x}$ with $x + 1 \geq 0$ defines half of a parabola via $1 + x = y^2$. You should relate this to your knowledge[32] of tangent lines and the differential calculus in highschool, $\frac{1}{2}$ being the value of $F'(0)$ if

$$F(x) = \sqrt{1+x}.$$

Replacing x by $y - 1$ the approximation is equivalent to

$$g(y) = \sqrt{y} \approx 1 + \frac{1}{2}(y-1),$$

which relates to $g'(1) = \frac{1}{2}$.

Apparently this idea of linear approximations was used in the Babylonian calculations to find $\sqrt{2}$, which in our modern language would look like

$$\sqrt{2} = \sqrt{r^2 + 2 - r^2} = r\sqrt{1 + \frac{2-r^2}{r^2}} \approx r\left(1 + \frac{2-r^2}{2r^2}\right) = \frac{r}{2} + \frac{1}{r},$$

in which the above approximation of $F(x) = \sqrt{1+x}$ is used as an intermediate step, with

$$x = \frac{2-r^2}{r^2}$$

to be precise. In terms of g the approximation is strikingly equivalent to

$$\sqrt{2} = g(2) \approx g(r^2) + \frac{1}{2r}(2 - r^2),$$

[32] You did not forget all of your differential calculus, did you?

in which the coefficient $\frac{1}{2r}$ is the value of the derivative $g'(y)$ in $y = r^2$. Thus the line tangent in $y = r^2$ to the graph $x = g(y)$ is used to approximate $\sqrt{2}$.

In terms of $f(x) = x^2$ this is equivalent to intersecting the line defined by

$$y = r^2 + 2r(x - r)$$

with the line defined by $y = 2$, and solving voor x to find a new and better approximation for $\sqrt{2}$, starting from a previous $x = r$. You may recognise this as Newton's method. Starting with $r = 1$ the approximation $\frac{3}{2}$ voor $\sqrt{2}$ is obtained (which is not that bad really). Redoing the approximation with $r = \frac{3}{2}$ then gives $\frac{17}{12}$, much better, and $r = \frac{17}{12}$ in turn gives

$$\frac{17}{24} + \frac{12}{17} = \frac{289 + 288}{24 \times 17} = \frac{577}{408} = 1 + \frac{24}{60} + \frac{51}{60^2} + \frac{10}{60^3} + \cdots,$$

the Babylonian approximation on YBC7289, in which the 10 will return if you continue the expansion in base 60.

Iterating the input-output relation[33]

$$r \to \frac{r}{2} + \frac{1}{r},$$

as was also done by Heron in Alexandria some 22 centuries later, one obtains

$$\frac{3}{2} > \frac{17}{12} > \frac{577}{408} > \frac{665857}{470832} > \frac{886731088897}{627013566048} > \cdots > 1$$

and thus an enumerated set of decreasing numbers which is bounded from below by 1. The largest such lower bound is of course $\sqrt{2}$, as you may want[34] to know for sure. Again the existence of such a largest lower bound L is a property of the set $\mathbb{R}$, either by assumption, or by construction, and a proof that $L^2 = 2$ is easily concocted. See e.g. my Dutch book with Ronald Meester on fingers and math.

The limits of difference quotients you have seen in highschool will be avoided in this material, as everything you have seen in the calculus of derivatives is to be based on linear approximations if we want to develop an underlying theory that will allow a generalisation to multivariate calculus. All this may start from elementary examples like

$$x^2 = 1 + x^2 - 1 = 1 + (x + 1)(x - 1) = 1 + 2(x - 1) + (x - 1)^2,$$

[33] You don't actually need the considerations above to come up with this method...

[34] I would.

to obtain $y = 1 + 2(x - 1) = 2x - 1$ as the formula for the line tangent in $(1, 1)$ to the parabola given by $y = x^2$. Babylonian math really. And again, if you know how to deal with x^2 you know how to deal with

$$f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 + \cdots ,$$

really, it's not the big deal you may have thought it was, as a little algebra will help to clarify.

1.4 Differential calculus founded in algebra

The formula

$$(BL_7) \qquad x^7 = a^7 + 7a^6(x-a) + (x^5 + 2ax^4 + 3a^2x^3 + 4a^3x^2 + 5a^4x + 6a^5)(x-a)^2$$

is a nice piece of easy but meaningful algebra that tells you a lot about hard analyis. You can find it[35] using long division to first write the *difference quotient*

$$\frac{x^7 - a^7}{x - a}$$

as

$$\frac{x^7 - a^7}{x - a} = x^6 + ax^5 + a^2x^4 + a^3x^3 + a^4x^2 + a^5x + a^6 = 7a^6 + (\cdots)(x - a),$$

and what's on the dots you find from another long division[36]. It gives you the long term in the algebra formula which I labeled (BL_7), BL for Babylonian linear approximation, the first two terms on the right hand side being a linear approximation of x^7 around $x = a$, just like we saw in the context of YBC7289 above.

To fix ideas restrict the attention to the fixed a and variable x lying in the interval $[-1, 1]$ which consists of all numbers between -1 and 1, including -1 and 1. The error in the linear approximation is then bounded by

$$(1 + 2 + 3 + 4 + 5 + 6)(x - a)^2,$$

that is to say, writing

$$x^7 = a^7 + 7a^6(x - a) + R_7(x, a),$$

we have

$$|R_7(x, a)| \leq \frac{7 \times 6}{2} (x - a)^2$$

[35] By looking at this page.
[36] Which one?

as young Gauss would write is. No other linear approximation comes with such an estimate, so therefore

$$a^7 + 7a^6(x - a)$$

is the *only linear approximation* to be called THE linear approximation of x^7 around $x = a$. This then is the algebraic foundation of differential calculus, and again it all adds up in Section 5.2. There are no limits it would seem[37].

1.5 The mystery of the tangent

So, if we believe the historians, the only thing about

$$f(x) = \sqrt{1 + x} \approx 1 + \frac{x}{2}$$

not going back to Babylonian times is the notation really. Neither does the concept of a tangent line I guess. We say that

$$y = 1 + \frac{x}{2}$$

defines a line tangent tangent in $(x, y) = (0, 1)$ to the graph of f, which should be the same as saying that

$$x = 2y - 2$$

defines a line tangent to the graph of g in $(y, x) = (1, 0)$ as in

$$x = g(y) = y^2 - 1.$$

Are the statements about tangent lines symmetric in x and y? Section 6.7 adresses this question in the context of one-to-one order preserving correspondences

$$y = f(x) \quad \text{and} \quad x = g(y)$$

between x- and y-intervals, first in the case that $f(0) = 0$ and $x = y$ defines a line tangent to the graph of f in $(0, 0)$. Enter the epsilons again.

1.6 So what about continuity?

Good question. It badly messes up analysis, that's for sure, as you inevatibly are about to find out, the sad fact being that most algebraic equations[38] simply don't have solutions we can find by algebraic methods. As a consequence the main theorem of algebra just cannot do without continuity. Read on.

[37] Indeed: Sectie 5.2, though under Opgave 5.14 the first ϵ and δ appear.

[38] See also Section 9.10.

1.6.1 Quadratics

You may be inclined to think of the quadratic equation

$$(I) \qquad 1 + x^2 = 0$$

as an example. But solutions of (I) are so unreal, imagine a real number that squares to -1! Square roots of negative numbers don't exist in $\mathbb{R}$. Thus your general quadratic equation

$$x^2 + px + q = 0$$

with real coefficients p, q is not always solvable by completing the square via

$$\left(x + \frac{p}{2}\right)^2 = \frac{p^2}{4} - q,$$

because you're really done if the above right hand side is negative.

Of course the solution formula

$$x = -\frac{p}{2} + \sqrt{\frac{p^2}{4} - q}$$

make good sense if

$$\frac{p^2}{4} > q,$$

although you should not forget that

$$x = -\frac{p}{2} - \sqrt{\frac{p^2}{4} - q}$$

defines another solution then, making for two solutions[39].

The borderline case is the equation

$$x^2 + px + \frac{p^2}{4} = 0,$$

which only has $x = -\frac{p}{2}$ as a solution, counted with multiplicity 2 as mathematicians rightly prefer to do.

[39] And no more, a little algebra theorem.

1.6.2 Cubics

What a releaf, no continuity arguments were needed for the quadratic equation. The story is the same for not quite your general cubic equation

$$x^3 + px + q = 0,$$

which was solved in the days of Cardano by putting

$$x = u + v \quad \text{with} \quad 3uv + p = 0,$$

leading to a quadratic equation for u^3 solvable by completing the square. The study of this method and the general solution of the cubic offers a true feast[40] in itself that we here sadly leave for some other party to crash. It does however unavoidably lead to the consideration of squares of square roots of negative numbers. Thus it really must be done: imagine a square root of -1.

1.6.3 Imagine what?

Since no square root of -1 exists as a real number we imagine one square root of minus one, call it not just anyone, but i will do[41]. The square of this imaginairy i then equals -1, and so then does the square of $-i$ if we stick by the usual rules for $+, -, \times, /$. Combining with the real numbers whose collective[42] fundamental properties Archimedes already understood we now find many complex numbers.

Here they all are:

$$z = x + iy = x + yi$$

with x and y real. Every point in the Cartesian xy-plane defines a complex number, numbers we can add, substract and multiply as we like, simplifying complicated expressions using the old rules combined with the new rule that $i^2 = -1$, as in a cute little trick for division in which $u^2 + v^2$ appears:

$$\frac{x+iy}{u+iv} = \frac{x+iy}{u+iv}\,\frac{u-iv}{u-iv} = \frac{ux+yv+i(uy-vx)}{u^2+v^2} = \frac{ux+yv}{u^2+v^2} + i\,\frac{uy-vx}{u^2+v^2}.$$

The Pythagorean sum u^2+v^2 is nowadays called the square of the absolute value of

$$w = u + iv,$$

[40] The quartic is another feast, but then the party's over, as Galois tragically found out.
[41] https://www.youtube.com/watch?v=WOHPuY88Ry4
[42] Sometimes collection is another word for set.

the complex number defined by the reals u and v, the absolute value of w being defined by

$$|w| = |u + iv| = \sqrt{u^2 + v^2}, \quad \text{and likewise} \quad |z| = |x + iy| = \sqrt{x^2 + y^2},$$

and you should note that

$$|z| = 0 \iff z = 0$$

generalises what you know for $|x|$ and x. In fact, for any complex expression $p(z)$ with z an unknown complex number, the equation

$$p(z) = 0$$

is equivalent to solving

$$|p(z)| = 0,$$

just like solving $p(x) = 0$ is equivalent to solving $|p(x)| = 0$ for real expressions, with x a real unknown.

Now if $p(z)$ is just $p(x)$ with x replaced by z it gets really interesting. Par exemple, the function

$$x \to |x^2 + 1|$$

is positive for every real x, whereas

$$z \to |z^2+1| = |(x^2-y^2+1+2ixy| = \sqrt{x^2 + (y+1)^2}\,\sqrt{x^2 + (y-1)^2} = f(x, y)$$

is a realvalued function of $z \in \mathbb{C}$ with two global zero minima (in $z = i$ and $z = -i$) as you see immediately from the formula above, and these global mimimisers are the roots of the equation

$$z^2 + 1 = 0,$$

which solved by merely *finding* the roots as $z = i$ and $z = -i$.

1.6.4 Minimise it!

Let's imagine[43] that you could not find these roots[44]. Then you may look at the formula for $f(x, y)$ to see that for (x, y) on large circles centered in the origin, $f(x, y)$ is bounded from below by a positive number. The larger the circle, the larger the lower bound. This should guarantee the existence of a smallest value m of $f(x, y)$ in some (x_0, y_0), i.e.

$$f(x_0, y_0) \leq f(x, y)$$

[43] I know that's unreal really, unless we lost it.

[44] Or the solution of any other polynomial equation! N.B. not every zero is a root.

for alle (x, y) with $x, y \in \mathbb{R}$, equivalent to $|z^2 + 1|$ being minimal and equal to m in $z_0 = x_0 + iy_0$.

Now substitute $z = z_0 + w$. Then

$$|(z_0 + w)^2 + 1| = |1 + z_0^2 + 2z_0w + w^2|$$

is minimal in $w = 0$. Can it happen that $1 + z_0^2 \neq 0$? If so then

$$m = |1 + z_0^2| > 0,$$

and thus the complex number $1 + z_0^2$ corresponds to a point in the xy-plane which, among all points corresponding to values of $1+z_0^2+2z_0w+w^2$ varying w, has minimal distance m to the origin.

Suppose then that the coefficient of w is nonzero. In this example it would mean that $z_0 \neq 0$, and all complex numbers w with

$$|w| = \frac{\epsilon}{2|z_0|}$$

for some $\epsilon > 0$ then produce values of $2z_0w$. It's a little exercise to show that these values have $|2z_0w| = \epsilon$, and that all complex numbers with absolute value equal to $\epsilon > 0$ appear as such a product $2z_0w$.

It follows that for some such w the complex number $1 + z_0^2 + 2z_0w$ corresponds to a point with distance $m - \epsilon$ to the origin, while

$$|w^2| = |w|^2 = \frac{\epsilon^2}{4|z_0|^2}$$

then guarantees that the point corresponding to $1 + z_0^2 + 2z_0w + w^2$ has distance to the origin at most

$$m - \epsilon + \frac{\epsilon^2}{4|z_0|^2} < m$$

if $\epsilon > 0$ is chosen small, a contradiction.

Thus $|z_0| > 0$ cannot be if $m > 0$, so m is achieved by $z_0 = 0$: $|1 + z^2|$ is minimal in $z = 0$, and $m = 1$. But this too cannot be: examine the values of $|z^2|$ for $|z| = \epsilon > 0$ and derive another contradiction. If follows that the assumption $m > 0$ was false and therefore $|z_0^2 + 1| = m = 0$. Therefore z_0 is a solution[45] of $z^2 + 1 = 0$.

[45] We found it again.

1.6.5 Generalise it!

Not a big deal, but the key idea applies to every equation of the form

$$z^2 + \alpha z + \beta = 0$$

with $\alpha, \beta \in \mathbb{C}$, and then to every equation of the form

$$z^3 + \alpha z^2 + \beta z + \gamma = 0,$$

as you may enjoy figuring out, and likewise for every polynomial equation

$$z^n + \alpha_{n-1} z^{n-1} + \cdots + \alpha_1 z + \alpha_0 = 0.$$

The existence of a global minimiser $z = z_0$ for

$$z \to |z^n + \alpha_{n-1} z^{n-1} + \cdots + \alpha_1 z + \alpha_0|$$

leads to the conclusion that

$$z_0^n + \alpha_{n-1} z_0^{n-1} + \cdots + \alpha_1 z_0 + \alpha_0 = 0,$$

and existence of a global minimiser in turn is guaranteed by the continuity of

$$(x, y) \to x + iy = z \to |z^n + \alpha_{n-1} z^{n-1} + \cdots + \alpha_1 z + \alpha_0| = f(x, y) \geq 0,$$

and the largeness of f on large circles centered in the origin. The latter involves a simple geometric estimate explained in Section 1.6.6 next, the former a proper introduction and definition of the concept of continuity that allows a simple existence proof, based on properties of real number sequences only, in search of minima and minimisers, see Section 1.6.7 below.

1.6.6 Estimates and order in the complex plane

For $\alpha, \beta, \gamma \in \mathbb{R}$[46] and z a variable complex number[47] the third order polynomial

$$C(z) = z^3 + \alpha z^2 + \beta z + \gamma$$

defines a varying complex number which corresponds to a varying point in the Cartesian plane. Note that $C(z)$ is the sum of z^3 and the quadratic[48] polynomial $Q(z) = \alpha z^2 + \beta z + \gamma$.

[46] No, that's not an exponent. To be read as: for real numbers α, β, γ.
[47] Notation: for variable $z \in \mathbb{C}$.
[48] Quadratic, unless $\alpha = 0$.

For fixed $R > 0$ the complex numbers z with $|z| = R$ correspond to a circle with radius R centered in the origin. Since $|z^3| = |z|^3$, the varying complex number z^3 corresponds to a point on the circle with the same center and (larger) radius R^3 (if we take $R > 1$).

For the absolute value of the smaller sum $Q(z)$ we have

$$|Q(z)| = |\alpha z^2 + \beta z + \gamma| \leq |\alpha z^2| + |\beta z| + |\gamma|,$$

which says that the distance to the origin of the point corresponding to $Q(z)$ is at most the sum of the distances of the points corresponding to the 3 indivual terms, which have respective distances

$$|\alpha z^2| = |\alpha|\,|z^2| = |\alpha|\,|z|^2 = |\alpha|\,R^2, \quad |\beta z| = |\beta|\,|z| = |\beta|\,R, \quad |\gamma|$$

to the origin if $|z| = R$. Thus the distance to the origin of the point corresponding to $C(z)$ is at least

$$R^3 - |\alpha|\,R^2 - |\beta|\,R - |\gamma| \geq R^3 - (|\alpha| + |\beta| + |\gamma|)R^2$$

$$= R^3(1 - \frac{|\alpha| + |\beta| + |\gamma|}{R}) \geq \frac{1}{2}\,R^3$$

if we choose $R \geq 2(|\alpha| + |\beta| + |\gamma|)$.

Without words this argument reads

$$|C(z)| = |z^3 + \alpha z^2 + \beta z + \gamma| \geq |z^3| - |\alpha z^2 + \beta z + \gamma| \geq |z^3| - |\alpha z^2| - |\beta z| - |\gamma|$$

$$\geq |z|^3(1 - \frac{|\alpha| + |\beta| + |\gamma|}{|z|}) \geq \frac{1}{2}\,|z|^3 = \frac{1}{2}\,R^3$$

if $|z| = R \geq 1$ and $|z| = R \geq 2(|\alpha| + |\beta| + |\gamma|)$.

Now fix such an R. Clearly then

$$|C(0)| = |\gamma| \leq \frac{R}{2} \quad \text{while} \quad |C(z)| > \frac{1}{2}R^3 \geq \frac{1}{2}R \quad \text{for} \quad |z| > R,$$

whence a smallest value of $|C(z)|$, *if it exists*, can only be attained by a $z \in \mathbb{C}$ with $|z| \leq R$. If it exists.

1.6.7 In search of minima and minimisers

Now define

$$f(x, y) = |C(x + iy)|$$

with $C(z) = C(x + iy)$ as just above, and let R be as chosen. Then

$$f(x, y) \geq 0$$

for every real number pair (x, y), and in addition it holds that

$$f(x, y) > f(0, 0) \quad \text{if} \quad x^2 + y^2 > R^2.$$

We look for a globally minimal value m of $f(x, y)$ and a corresponding minimiser (x_0, y_0) with $f(x_0, y_0) = m$.

Now if $f(x_0, y_0) = 0$ for some (x_0, y_0) we're done[49]. So suppose that $f(x, y) > 0$ for every real x and y. In search of the minimum we can forget about the points (x, y) outside the circle defined by $x^2 + y^2 = R^2$, as we know that we can only have $m \leq f(0, 0)$. If there are no values $f(x, y) < f(0, 0)$ we're also done[50].

So assume that $f(0, 0)$ is not a global minimal value of $f(x, y)$. Without loss of any generality we now assume that $f(0, 0) = 1$ and it may happen that there are values $f(x, y) < \frac{1}{2}$. If so let $I_1 = [0, \frac{1}{2})$.

If not it may happen that there are values $f(x, y) < \frac{3}{4}$, if so let $I_1 = [\frac{1}{2}, \frac{3}{4})$, if not, then consider $[\frac{3}{4}, \frac{7}{8})$, and so on. This proces must terminate otherwise Archimedes happily tells us there are no values $f(x, y)$ below 1 at all. It follows that in the sequence of intervals

$$[0, \frac{1}{2}), [\frac{1}{2}, \frac{3}{4}), [\frac{3}{4}, \frac{7}{8}), [\frac{7}{8}, \frac{15}{16}), \ldots$$

there is a *first* interval I_1 that contains a value $f(x, y)$, say $p_1 = f(x_1, y_1)$. In that interval we repeat the argument above and obtain a first interval I_2 in a similar sequence with a $p_2 = f(x_2, y_2) \in I_2$, and so one.

It may happen that $p_2 = p_1$ and or $a_1 = a_2$ and likewise in every next step, but no matter what, we get

$$a_n \leq p_n = f(x_n, y_n) < b_n$$

with both a_n and b_n fractions whose denominators are pure powers of 2, and

$$0 < b_n - a_n \leq \frac{1}{2^n}.$$

By construction the minimum we look for must lie in every $[a_n, b_n)$.

The sequence a_n has

$$0 \leq a_1 \leq a_2 \leq a_3 \leq \cdots < 1$$

and its smallest lower bound defines a real number that we call m. It's really Archimedes again who now tells us this m is in fact *the only possible value a global minimum* of $f(x, y)$ can have, if it exists.

[49] By the way, this is what we wanted in this example, remember?

[50] This is what we wanted to force a contradiction.

1.6.8 Conclusion

The by construction bounded sequences x_n and y_n do not necessarily have the property that

$$x_1 \leq x_2 \leq x_3 \leq x_4 \leq \cdots \quad \text{or} \quad x_1 \geq x_2 \geq x_3 \geq x_4 \geq \cdots$$

and likewise for y_n, but a nice little argument[51] provides us with subsequences x_n and y_n which do have this property and thereby define real numbers x_0 and y_0, either as lowest upperbound of the particular subsequence, or as largest lower bound. What remains is to come up with a definition of continuity that applies to the example from Section 1.6.6 under consideration and implies that $f(x_0, y_0) = m$. A quick way to do so is via the concept of convergent sequences and to say that continuity amounts to $x_n \to x_0$ and $y_n \to y_0$ implying that $f(x_n, y_n) \to f(x_0, y_0)$. In Sectie 2.2 this characterisation is discussed in the context of functions of one variable and related to the concept of uniform continuity in Stelling 2.32.

[51] See Sectie 3.6. BTW, Thomas Rot worked out Opgave 3.16.

2 Integraalrekening

Voor functies $f : \mathbb{R} \to \mathbb{R}$ begint integraalrekening met de vraag naar de oppervlakte van verzamelingen in het x, y-vlak begrensd door de x-as en de *grafiek* van f. Die grafiek wordt gegeven door $y = f(x)$. Om niet meteen over onbegrensde gebieden te praten kijken we eerst naar verzamelingen van de vorm

$$A_+ = \{(x, y) : x, y \in \mathbb{R},\ 0 < y < f(x),\ a < x < b\}$$

en

$$A_- = \{(x, y) : x, y \in \mathbb{R},\ f(x) < y < 0,\ a < x < b\},$$

met $a, b \in \mathbb{R}$ en $a < b$, waarbij we opmerken dat de "grotere" verzamelingen die ontstaan als we hierin elke $<$ door $\leq$ vervangen dezelfde oppervlakte zouden moeten hebben. Althans bij een zinvolle definitie van het begrip oppervlakte, een definitie die we hier echter achterwege laten!

Dat lijnstukken oppervlakte nul hebben, en grafieken van niet al te lelijke functies ook, lijkt vanzelfsprekend, maar komt hier nog niet aan de orde. In wat hieronder volgt gaat het alleen over welomschreven benaderingen van oppervlakten van verzamelingen waarvan we kunnen aannemen dat die op de een of andere manier meetbaar zijn, bijvoorbeeld met een gereedschapskist vol rechthoekjes waarvan de oppervlakte als lengte en breedte buiten kijf staat. De wiskunde hieronder is correct ook zonder dat aan het begrip oppervlakte wordt gedacht. Uitgangspunt is wel de consensus dat de oppervlakte tussen de lijnen $x = 0$, $x = 1$, $y = 0$, $y = 1$ in het xy-vlak gelijk is aan $1 \times 1 = 1 \cdot 1 = 1$, let wel, zonder fysische eenheden erbij.

We beginnen met het geval dat $A_- = \emptyset$: we nemen aan dat $f(x) \geq 0$ voor $x \in [a, b]$. Het geval $[a, b] = [0, 1]$ en $f(x) = x^n$ met $n \in \mathbb{N}$ is een leerzaam eerste voorbeeld. We rekenen wat we de oppervlakte van de verzameling

$$A_n = \{(x, y) : 0 \leq y \leq x^n \leq 1\}$$

zouden willen noemen expliciet uit. Als notatie voor de oppervlakte[1] van een verzameling $A \subset \mathbb{R}^2$ gebruiken we $|A|$. Als uitgangspunt nemen we dat over de oppervlakte van een rechthoek geen twijfel bestaat, waarbij we zonder eenheden (dus geen vierkante centimeters of zo) werken. De oppervlakte van een rechthoek die parallel ligt aan de assen wordt eenvoudig bepaald door de schaalverdeling op de assen.

Voor $n = 1$ gaat het om het gebied in het xy-vlak begrensd door

$$x = 1, \quad y = 0 \quad \text{en} \quad y = x,$$

[1] Of elke $A \subset \mathbb{R}^2$ een oppervlakte $|A|$ heeft is natuurlijk wel de vraag.

een driehoek met oppervlakte $\frac{1}{2}$, precies de helft van de oppervlakte van het vierkant met hoekpunten $(0,0)$, $(1,0)$, $(1,1)$ en $(0,1)$. Kortom $|A_1| = \frac{1}{2}$.

Hoe zit het met A_2? Omdat $A_2 \subset A_1$ is het duidelijk dat moet gelden dat $|A_2| \leq |A_1|$, nog voor je de vraag stelt hoe en of de oppervlakte eigenlijk precies gedefinieerd is. Het is ook duidelijk[2] dat $|A_2|$ niet kleiner kan zijn zijn dan de oppervlakte van een rechthoek die geheel in A_2 ligt, bijvoorbeeld de rechthoek met hoekpunten $(\frac{1}{2},0)$, $(1,0)$, $(1,\frac{1}{4})$ en $(\frac{1}{2},\frac{1}{4})$. We concluderen hieruit dat $|A_2|$ minimaal $\frac{1}{2} \times \frac{1}{4}$ moet zijn.

Opgave 2.1. Laat met rechthoekjes zien dat ook moet gelden dat

$$\frac{1}{3}\cdot\frac{1}{9} + \frac{1}{3}\cdot\frac{4}{9} < |A_2| < \frac{1}{3}\cdot\frac{1}{9} + \frac{1}{3}\cdot\frac{4}{9} + \frac{1}{3}\cdot\frac{9}{9}.$$

Opgave 2.2. Laat zien dat voor elke $N \in \mathbb{N}$ moet gelden dat

$$\frac{1^2 + 2^2 + \cdots + (N-1)^2}{N^3} < |A_2| < \frac{1^2 + 2^2 + \cdots + N^2}{N^3}.$$

Als je even nadenkt over hoe je de grenzen in Opgave 2.2 hebt gevonden dan is duidelijk dat het verschil precies $\frac{1}{N}$ is, en dat kunnen we zo klein maken als we willen door N groot te nemen. Bovendien kunnen we deze onder- en bovengrenzen gewoon uitrekenen:

Opgave 2.3. Laat zien dat voor elke $N \in \mathbb{N}$ geldt dat

$$\sum_{n=1}^{N} n^2 = 1^2 + 2^2 + \cdots + N^2 = \frac{N^3}{3} + \frac{N^2}{2} + \frac{N}{6}.$$

We gebruiken hier de somnotatie waarin de index n aangeeft dat we de som nemen van de kwadraten van n over alle n in de verzameling

$$\{n \in \mathbb{N} : 1 \leq n \leq N\}.$$

[2] Maak een plaatje.

In [Eves] lezen we op pagina 47 dat de formule in Opgave 2.3 al door Archimedes gevonden is. Het rechterlid factoriseert als

$$\frac{N(N+1)(2N+1)}{6}$$

maar het gaat ons in het vervolg alleen om de coëfficiënt van N^3 in het rechterlid. De bovengrens voor $|A_2|$ in Opgave 2.2 is nu immers gelijk aan

$$\frac{1}{3}+\frac{1}{2N}+\frac{1}{6N^2}$$

en wordt kleiner als N groter genomen wordt. De ondergrens is (niet opnieuw uitrekenen maar gewoon $\frac{1}{N}$ eraf)

$$\frac{1}{3}-\frac{1}{2N}+\frac{1}{6N^2},$$

en dat wordt juist groter als N groter genomen wordt (al zie je dat misschien niet meteen). De onvermijdelijke conclusie is dat $|A_2|=\frac{1}{3}$. In termen van de integralen[3] die je natuurlijk al eerder gezien hebt is hiermee de uitspraak dat

$$\int_0^1 x^2\,dx=\frac{1}{3}$$

gedaan zonder dat er een functie geprimitiveerd is!

Wat aardig en nuttig is om op te merken is dat we van het derdegraadspolynoom in Opgave 2.3 alleen de kopcoëfficiënt in het antwoord terugzien. Het is een leerzame oefening dat het voor bijvoorbeeld[4] A_{42} precies hetzelfde gaat, je bewijst en gebruikt

$$1^{42}+2^{42}+\cdots+N^{42}=\frac{N^{43}}{43}+\cdots$$

om via oppervlakten van rechthoekjes op $|A_{42}|=\frac{1}{43}$ uit te komen.

Opgave 2.4. Doe $|A_1|$ ook op deze manier. Hint: $1+2+\cdots+N$ is[5] een som die Gauss ooit als strafwerk zou hebben gekregen.

[3] Maar denk ook aan volumes van piramiden, zie Hoofdstuk 2 van [Eves].
[4] Amusant in [Eves, §3.16]: bewijs meetkundig dat $1^3+\cdots+N^3=(1+\cdots+N)^2$.
[5] Met $N=100$.

Opgave 2.5. Schrijf de differenties $(n+1)^2 - n^2$, $(n+1)^3 - n^3$, $(n+1)^4 - n^4$, enzovoorts uit.

Opgave 2.6. De ad hoc notaties

$$N_1 = \sum_{n=1}^{N} n, \quad N_2 = \sum_{n=1}^{N} n^2, \quad N_3 = \sum_{n=1}^{N} n^3, \ldots$$

definiëren sommen die voor elke index $1, 2, 3, \ldots$ van $N \in \mathbb{N}$ afhangen[6]. Gebruik nu Opgave 2.5: laat met de differenties van n^2 zien dat

$$(N+1)^2 = 1 + N + 2N_1,$$

met de differenties van n^3 dat

$$(N+1)^3 = 1 + N + 3N_1 + 3N_2,$$

met de differenties van n^4 dat

$$(N+1)^4 = 1 + N + 4N_1 + 6N_2 + 4N_3,$$

enzovoorts.

Opgave 2.7. Laat voor $k = 1, 2, 3, \ldots$ zien[7] dat N_k een polynoom is van graad $k+1$ met kopcoëfficiënt $\frac{1}{k+1}$. Hint: los uit de in Opgave 2.6 gevonden indentiteiten achtereenvolgens $N_1, N_2, N_3, \ldots$ op; voor elke volgende k alle vorige resultaten gebruiken en goed kijken wat de graad van elke term is.

Nog voor dat we aan precieze definities zijn toegekomen is de onvermijdelijke conclusie dat wel moet gelden dat

$$|A_n| = \int_0^1 x^n \, dx = \frac{1}{n+1},$$

en een kleine variant hier op is:

[6] Met eventueel

$$N_0 = \sum_{n=1}^{N} 1.$$

[7] Index k omdat de index n al in gebruik is, het geval $k = 42$ is al genoemd.

Opgave 2.8. Voor $b > 0$ en $n \in \mathbb{N}$ is de oppervlakte van

$$\{(x, y) : 0 \leq y \leq x^n \leq b^n\}$$

gelijk aan het quotiënt van b^{n+1} en $n + 1$. Laat dit zien door met schaling van de eenheden op de assen de vraag te herleiden tot het geval $b = 1$.

Lezend in [Eves] is het duidelijk dat Archimedes deze conclusie al had kunnen trekken en dat wellicht ook gedaan heeft voor $n = 2$ en $n = 3$. In moderne notatie:

$$\int_0^b x^n \, dx = \frac{b^{n+1}}{n+1}.$$

Opgave 2.9. Voor $0 \leq a < b$ en $n \in \mathbb{N}$ is de oppervlakte van

$$\{(x, y) : a \leq x \leq b, 0 \leq y \leq x^n\}$$

(in moderne notatie) gelijk aan

$$\int_a^b x^n \, dx = [\frac{x^n}{n+1}]_a^b = \frac{b^{n+1}}{n+1} - \frac{a^{n+1}}{n+1}.$$

De functie

$$x \to \frac{x^{n+1}}{n+1}$$

noemen we tegenwoordig de primitieve van de functie

$$x \to x^n$$

omdat de afgeleide van de eerste functie gelijk is aan de tweede. Het is niet waarschijnlijk dat de lezer[8] de begrippen primitieve en afgeleide functies nog niet in de een of andere vorm gezien heeft maar het advies is om een en ander bij verder lezen in het achterhoofd te houden. In (4.6) komt de uitspraak hierboven terug in algemene vorm.

Opgave 2.10. Tijdens de vakantiecursus voor wiskundeleraren op 25-8-2017 in Eindhoven maakte een van de toehoorders de terechte opmerking dat uit $\int_0^1 x^2 \, dx = \frac{1}{3}$ ook de integraal $\int_0^1 \sqrt{x} \, dx$ volgt. Kun je $\int_0^1 \sqrt[n]{x} \, dx$ berekenen voor iedere $n \in \mathbb{N}$?

[8] Ik zal de lezer doorgaans tutoyeren, sorry.

Opgave 2.11. De dag na die cursus mailde Frits Beukers me een manier om op $\int_0^1 x^k\,dx = \frac{1}{k+1}$ uit te komen zonder een formule voor $1^k + 2^k + \cdots + n^k$, gebruikmakend van de differenties in Opgave 2.5, maar dan uitgeschreven via het merkwaardige product $a^{k+1} - b^{k+1} = (a-b)(a^k + \cdots + b^k)$. Vul de puntjes in en schrijf uit wat je krijgt met $a = n+1$ en $b = n$. Je krijgt $k+1$ termen, die elk uit factoren $n+1$ en n bestaan. Als je al zulke factoren n vervangt door $n+1$ wordt de som groter en als je al zulke $n+1$ vervangt door n wordt de som kleiner. Sommeer de ongelijkheden die je krijgt van $n = 0$ tot $n = N$ en deel door $k+1$. Laat zien dat

$$1^k + 2^k + \cdots + N^k < \frac{(N+1)^{k+1}}{k+1} < 1^k + 2^k + \cdots + N^k + (N+1)^k,$$

en gebruik dit om via onder- en bovensommen $\int_0^1 x^k\,dx$ te bepalen.

2.1 Integralen via onder- en bovensommen

De integraal van $x = 0$ tot $x = 1$ van x^2 in de inleiding hierboven is bij afspraak de nog niet wiskundig gedefinieerde oppervlakte van de verzameling[9]

$$A = \{(x, y) : x, y \in \mathbb{R},\ 0 \le y \le f(x),\ a \le x \le b\}, \tag{2.1}$$

met in dit voorbeeld $a = 0$, $b = 1$ en $f(x) = x^2$. Voor $a, b \in \mathbb{R}$ met $a < b$ en een functie $f : [a, b] \to \mathbb{R}$ met $f(x) \ge 0$ voor $a \le x \le b$, beoogt de wiskundige definitie van

$$\int_a^b f(x)\,dx \tag{2.2}$$

wiskundig betekenis te geven aan de oppervlakte $|A|$ van A in (2.1).

Daartoe worden zogenaamde ondersommen en bovensommen gedefinieerd, waarvan het evident is dat ze bij de beoogde definitie van de oppervlakte van A kleiner respectievelijk groter[10] dan $|A|$ zijn. Als vervolgens blijkt dat er precies één reëel getal is dat groter is dan alle ondersommen en kleiner dan alle bovensommen, dan is dit getal de enige kandidaat voor $|A|$, en vervolgens wordt dit getal de integraal van f over $[a, b]$ genoemd, i.e. (2.2). Het is natuurlijk niet verplicht om dan vervolgens te zeggen dat deze integraal gelijk is aan de oppervlakte $|A|$ van A.

We bekijken deze aanpak eerst voor het geval dat $f : [a, b] \to \mathbb{R}$ een niet-dalende niet-negatieve functie is.

[9] Nu maar met $\le$ i.p.v. $<$.

[10] Precies gezegd: groter dan of gelijk aan.

Definitie 2.12. *Als $a, b \in \mathbb{R}$ met $a < b$ en $f : [a, b] \to \mathbb{R}$, dan heet f niet-negatief als $f(x) \geq 0$ voor alle $x \in [a, b]$; f heet niet-dalend als de implicatie*

$$x_1 \leq x_2 \implies f(x_1) \leq f(x_2)$$

geldt voor alle $x_1, x_2 \in [a, b]$.

Voor zo'n niet-dalend niet-negatieve functie ligt het voor de hand om de oppervlakte $|A|$ van A in (2.1) van onder te benaderen met in dit geval *ondersommen* van de vorm

Teken plaatje!

$$\underline{S}_N = \sum_{k=1}^{N} f(x_{k-1})(x_k - x_{k-1}) \tag{2.3}$$

$$= f(x_0)(x_1 - x_0) + \cdots + f(x_{N-1})(x_N - x_{N-1}),$$

waarin

$$a = x_0 \leq x_1 \leq \cdots \leq x_N = b \quad \text{met} \quad N \in \mathbb{N}. \tag{2.4}$$

Zo'n keuze van $x_1, \ldots, x_{N-1}$ noemen we een *partitie* P van $[a, b]$ en $\underline{S}_N$ is niets anders dan de som van de oppervlakten van de rechthoekjes[11]

$$R_k = [x_{k-1}, x_k] \times [0, f(x_{k-1})] \quad (k = 1, \ldots, N).$$

Hoe $|A|$ verder ook gedefinieerd wordt, je kunt je zelf er makkelijk van overtuigen dat bij iedere zinvolle definitie van $|A|$ wel moet gelden dat $\underline{S}_N \leq |A|$. Teken maar een plaatje met grafiek en rechthoekjes.

Evenzo maken de oppervlakten van de rechthoekjes

$$R_k = [x_{k-1}, x_k] \times [0, f(x_k)] \quad (k = 1, \ldots, N)$$

de *bovensom*

$$\bar{S}_N = \sum_{k=1}^{N} f(x_k)(x_k - x_{k-1}). \tag{2.5}$$

Voor een zinvolle definitie van de oppervlakte van A moet dus wel gelden dat

$$\underline{S}_N \leq |A| \leq \bar{S}_N. \tag{2.6}$$

Slordig[12] gezegd, de gezochte waarde $|A|$ zitten *tussen*[13] de ondersom $\underline{S}_N$ en de bovensom $\bar{S}_N$, en het verschil van die twee is

$$\bar{S}_N - \underline{S}_N = \sum_{k=1}^{N} (f(x_k) - f(x_{k-1}))(x_k - x_{k-1}).$$

[11] Zo'n rechthoekje kan ook een lijnstuk of een punt zijn.

[12] Het kan gebeuren dat $\underline{S}_N = |A|$ en/of $|A| = \bar{S}_N$.

[13] Als $\underline{S}_N < |A| < \bar{S}_N$ spreken we van strict tussen.

Als we nu de partitie P in (2.4) equi-distant kiezen, i.e.

$$x_k - x_{k-1} = \frac{b-a}{N} \quad (k = 1, \ldots, N),$$

dan volgt

$$\bar{S}_N - \underline{S}_N = \sum_{k=1}^{N} (f(x_k) - f(x_{k-1})) \frac{b-a}{N} = (f(b) - f(a)) \frac{b-a}{N}$$

en dat kunnen we willekeurig klein maken door N groot te kiezen[14]. De onvermijdelijke conclusie is dan dat er maar één getal tussen alle ondersommen en bovensommen kan liggen.

Opgave 2.13. Met $N = 1, 2, 4, 8, 16 \ldots$ vormen de ondersommen $\underline{S}_1, \underline{S}_2, \underline{S}_4, \underline{S}_8, \ldots$ een niet-dalende rij getallen. Laat dit zien.

Evenzo vormen de bovensommen $\bar{S}_1, \bar{S}_2, \bar{S}_4, \bar{S}_8, \ldots$ een niet-stijgende rij. Omdat

$$\underline{S}_1 \leq \underline{S}_2 \leq \underline{S}_4 \leq \underline{S}_8 \leq \cdots \leq \bar{S}_8 \leq \bar{S}_4 \leq \bar{S}_2 \leq \bar{S}_1,$$

zijn beide rijen begrensd.

Definitie 2.14. *Als $a \leq M \in \mathbb{R}$ respectievelijk $a \geq M$ voor alle $a \in A \subset \mathbb{R}$ dan heet M een bovengrens respectievelijk een ondergrens voor A, en A naar boven respectievelijk beneden begrensd.*

De kleinste bovengrens[15] $\underline{J}$ van deze rij ondersommen kan niet groter zijn dan een bovensom en is daarmee een ondergrens voor alle bovensommen in deze rij bovensommen. Voor de grootste ondergrens[16] $\bar{J}$ van alle bovensommen geldt dus $\bar{J} \geq \underline{J}$. Omdat

$$\bar{J} - \underline{J} \leq \bar{S}_{2^n} - \underline{S}_{2^n} = (f(b) - f(a)) \frac{b-a}{2^n}$$

en 2^n willekeurig groot gekozen kan worden volgt $\bar{J} - \underline{J} = 0$. Zo definiëren we een uniek getal $J = \bar{J} = \underline{J}$ dat vervolgens genoteerd wordt als

$$J = \int_{[a,b]} f = \int_a^b f = \int_a^b f(x)\, dx,$$

[14] Anders zou $\mathbb{N}$ naar boven begrensd zijn in $\mathbb{R}$.

[15] Ander woord voor kleinste bovengrens: supremum. Waarom bestaat die/dat?

[16] Ander woord voor grootste ondergrens: infimum. Waarom bestaat die/dat?

spreek naar keuze uit als: de *integraal* van f over $[a,b]$, respectievelijk de integraal van f of $f(x)$ van a of $x=a$ tot b of $x=b$. In de notatie is x een *dummy* variabele, die door elke andere letter vervangen kan worden, maar bij voorkeur niet door a of b, en liever ook niet door d of f.

Opgave 2.15. Voor niet-stijgende niet-negatieve functies is het verhaal hierboven op details na hetzelfde. Schrijf dat uit.

Opgave 2.16. Niet-stijgende functies en niet-dalende functies worden monotone functies genoemd. Ook als deze functies zowel positieve als negatieve waarden aannemen, definiëren eindige sommen als in (2.3) en (2.5) via $N=2^n$ een uniek getal J dat de integraal van f over $[a,b]$ genoemd wordt. Wat is de interpretatie van deze integraal in termen van de gebieden in het x,y-vlak begrensd door $x=a$, $x=b$, $y=0$ en $y=f(x)$? Leg uit waarom.

De integraal van een monotone functie f is nu gedefinieerd aan de hand van equi-distante partities van het interval $[a,b]$. Hoe zit het met andere partities en de ordening van onder- en bovensommen? Een vraag om algemeen te stellen maar nu eerst nog even het voor specifieke geval van monotone functies.

Opgave 2.17. Neem weer het geval dat f niet-dalend is. Neem een partitie P gedefinieerd door (2.4) en een partitie Q gedefinieerd door

$$a=y_0\le y_1\le\cdots\le y_M=b\quad\text{met}\quad M\in\mathbb{N}.$$

Neem[17] bij P de bovensom $\bar{S}$ in (2.5) en bij Q de ondersom

$$\underline{S}=\sum_{l=1}^{M} f(y_{l-1})(y_l-y_{l-1})$$

Laat zien dat $\underline{S}\le\bar{S}$. Hint: eerst zelf proberen, maar kijk eventueel al naar (2.11).

Definitie 2.18. *Laat A en B twee deelverzamelingen zijn van $\mathbb{R}$. We zeggen dat $A\le B$ als $a\le b$ voor alle $a\in A$ en alle $b\in B$. In dat geval zeggen we dat $J\in\mathbb{R}$ tussen A en B ligt als $A\le\{J\}\le B$.*

[17] Even zonder index in de notatie.

Al onze onder- en bovensommen maken twee zulke verzamelingen A en B. De volgende opgave is nu van belang.

Opgave 2.19. Laat A en B twee niet-lege deelverzamelingen zijn van $\mathbb{R}$ met de eigenschap dat $A \leq B$. Bewijs dat er een getal $J \in \mathbb{R}$ is dat tussen A en B ligt. Hint: zie de redenatie onder Definitie 2.14.

Stelling 2.20. *Laat $f : [a, b] \to \mathbb{R}$ een monotone functie zijn. Dan bestaat er precies één getal J dat tussen de verzameling van alle sommen van de vorm (2.3) en de verzameling van alle sommen van de vorm (2.5) ligt. Dit getal J is per definitie de integraal van $f(x)$ van $x = a$ tot $x = b$, notatie*

$$J = \int_a^b f(x)\,dx.$$

Bewijs. Het werk is hierboven gedaan, tenzij Opgave 2.17 niet gelukt is, maar lees dan toch maar verder. Zonder beperking der algemeenheid[18] nemen we f niet-dalend. De ondersommen vormen een niet-lege verzameling A en de bovensommen een niet-lege verzameling B. Er geldt dat $A \leq B$, dus er bestaat een J tussen A en B. Met equi-distante partities zien we dat er maar één zo'n J kan zijn. Klaar.

Monotone functies zijn dus integreerbaar. Merk op dat gegeven een partitie P zoals in (2.4), en iedere keuze van zogenaamde *tussenpunten*

$$x_{k-1} \leq \xi_k \leq x_k \quad (k = 1, \ldots, N),$$

ook

$$S = \sum_{k=1}^{N} f(\xi_k)(x_k - x_{k-1}) \tag{2.7}$$

een voor de hand liggende benadering van $\int_a^b f(x)\,dx$ zou moeten zijn, ook als $f : [a, b] \to \mathbb{R}$ niet monotoon is, maar we kunnen dan niet meer op een natuurlijke manier van boven- of ondersommen spreken.

Als de waardenverzameling[19]

$$\{f(x) :\ a \leq x \leq b\} = \cup_{x\in[a,b]}\{f(x)\}$$

echter begrensd is, i.e. als er $m, M \in \mathbb{R}$ bestaan zodat $m \leq f(x) \leq M$ voor alle $x \in [a, b]$, dan is op elke deelinterval $[x_{k-1}, x_k]$ de waardenverzameling

[18] Afkorting: zbda.

[19] Het bereik B_f van $f : [a, b] \to \mathbb{R}$, door sommigen ook het codomein van f genoemd.

ook begrensd en bestaan er bijbehorende ondergrenzen m_k en bovengrenzen M_k met $m \leq m_k \leq M_k \leq M$, waarmee we

$$\underline{S} = \sum_{k=1}^{N} m_k(x_k - x_{k-1}) \quad \text{en} \quad \bar{S} = \sum_{k=1}^{N} M_k(x_k - x_{k-1}) \tag{2.8}$$

weer als onder- en bovensommen voor zowel (2.7) als de te definiëren integraal $\int_a^b f(x)\,dx$ kunnen introduceren. Bij dezelfde partitie P geldt dan onmiddellijk dat $\underline{S} \leq \bar{S}$.

Wat hebben we nu nodig om het bewijs van Stelling 2.20 over te schrijven? Twee dingen goed begrijpen nu. Ten eerste, de verzamelingen van ondersommen A en bovensommen B moeten weer voldoen aan $A \leq B$. Ten tweede, we moeten de uniciteit van het getal J tussen A en B aantonen, en daartoe nemen we voor m_k en M_k natuurlijk de grootst respectievelijk kleinst mogelijk waarde, teneinde het verschil

$$\bar{S} - \underline{S} = \sum_{k=1}^{N} (M_k - m_k)(x_k - x_{k-1})$$

zo klein mogelijk[20] te krijgen.

Kies dus

$$M_k = \sup_{[x_{k-1},x_k]} f = \sup\{f(x) : x_{k-1} \leq x \leq x_k\} \quad \text{en} \quad m_k = \inf_{[x_{k-1},x_k]} f.$$

Hierin staat sup voor *supremum* van de verzameling in kwestie. Iedere niet-lege verzameling $A \subset \mathbb{R}$ die begrensd is naar boven, middels een *bovengrens* $M \geq a$ voor alle $a \in A$, heeft een kleinste bovengrens die het supremum van A genoemd wordt. In Sectie 26.2 heb je hier al over gelezen natuurlijk. Dit is één van de vanzelfsprekende eigenschappen die $\mathbb{R}$ moet[21] hebben. Evenzo staat inf voor *infimum* of grootste ondergrens.

Opgave 2.21. Laat zien dat bij het *verfijnen* van de partitie, i.e. extra punten toevoegen in (2.4), ondersommen niet kleiner en bovensommen niet groter kunnen worden.

Als nu voor alle $k = 1, \ldots, N$ geldt dat

$$M_k - m_k \leq \varepsilon \tag{2.9}$$

[20] Hopelijk willekeurig klein.

[21] Anders $\mathbb{R}$ terugsturen naar de verzamelingsgeleerden!

(voor een $\varepsilon > 0$), dan volgt dat

$$\bar{S} - \underline{S} = \sum_{k=1}^{N} (M_k - m_k)(x_k - x_{k-1}) \leq \varepsilon(b-a). \tag{2.10}$$

Wat we dus nodig hebben voor uniciteit van J is dat elke $\varepsilon > 0$ realiseerbaar is in (2.9), middels een geschikte keuze van (2.4). In dat geval geldt voor iedere andere $\tilde{J}$ tussen A en B dat

$$|\tilde{J} - J| \leq \bar{S} - \underline{S} \leq \varepsilon(b-a),$$

en omdat $\varepsilon > 0$ dan zo klein gekozen mag worden als we maar willen volgt[22] $\tilde{J} = J$.

De definitie van (meteen maar) *uniforme continuïteit* brengt nu uitkomst. Dit is de eerste keer dat we een uitspraak formuleren waarin *voor alle* $\varepsilon > 0$ iets moet gelden dat met die ε te maken heeft, en zo'n uitspraak meteen gebruiken om een fundamentele stelling te bewijzen. Merk op dat je zo'n uitspraak alleen maar *echt gebruikt* als je meerdere (de facto oneindig veel) keren een ε kiest[23].

Definitie 2.22. *Als $a, b \in \mathbb{R}$ met $a < b$ en $f : [a,b] \to \mathbb{R}$, dan heet f uniform continu als er voor alle $\varepsilon > 0$ een $\delta > 0$ bestaat zodanig dat voor alle $x, y \in [a,b]$ geldt dat*

$$|x - y| \leq \delta \implies |f(x) - f(y)| \leq \varepsilon.$$

Stelling 2.23. *Laat $f : [a,b] \to \mathbb{R}$ een uniform continue functie zijn. Dan bestaat er precies één getal J dat tussen de verzameling van alle ondersommen en de verzameling van alle bovensommen ligt, waarbij we onder de onder- en bovensommen de sommen in (2.8) verstaan. Dit getal J is per definitie de integraal van $f(x)$ van $x = a$ tot $x = b$, notatie*

$$J = \int_a^b f(x)\,dx.$$

Bewijs. Wederom is vrijwel al het werk al gedaan. De ondersommen vormen een niet-lege verzameling A en de bovensommen een niet-lege verzameling B. Met $\delta > 0$ bij $\varepsilon > 0$ zoals in Definitie 2.23 volgt[24] (2.9) als de partitie P in (2.4) zo gekozen wordt dat $x_k - x_{k-1} \leq \delta$ voor alle $k = 1, \ldots, N$.

[22] Ook dit gebruikt weer dat $\mathbb{N}$ niet begrensd is in $\mathbb{R}$ via het kiezen van $\varepsilon = \frac{1}{n}$.
[23] Houd het met $\varepsilon = \frac{1}{n}$ bij voorkeur simpel, en je slaagt met vlag en wimpel.
[24] Ga dit na, al volgt aanstonds dat m_k en M_k aangenomen worden in $[x_{k-1}, x_k]$.

Als we nu ook nog laten zien dat $A \leq B$ zijn we klaar. Neem daartoe P als in (2.4) en Q net als in Opgave 2.17, en vorm de *gemeenschappelijke* verfijning

$$a = z_0 \leq z_1 \leq \cdots \leq z_K = b, \tag{2.11}$$

door $x_1 \leq \cdots \leq x_{N-1}$ en $y_1 \leq \cdots \leq y_{M-1}$ op één gemeenschappelijke volgorde te zetten. Dus $K - 1 = M - 1 + N - 1$ en iedere z_i is een x_k of een y_l, waarmee met

$$m_l = \inf_{[y_{l-1},y_l]} f, \quad \tilde{m}_i = \inf_{[z_{i-1},z_i]} f \leq \sup_{[z_{i-1},z_i]} f = \tilde{M}_i, \quad M_k = \sup_{[x_{k-1},x_k]} f$$

volgt dat

$$\sum_{l=1}^{M} m_l(y_l - y_{l-1}) \leq \sum_{i=1}^{K} \tilde{m}_l(z_i - z_{i-1}) \leq \sum_{i=1}^{K} \tilde{M}_l(z_i - z_{i-1}) \leq \sum_{k=1}^{N} M_k(x_k - x_{k-1}).$$

Waarom? Omdat de eerste ongelijkheid de ondersommen betreft van een partitie en een verfijning[25] daarvan, en de derde ongelijkheid de bovensommen van een andere partitie naar dezelfde verfijning. De tweede ongelijkheid spreekt vanzelf en Stelling 2.23 is nu voor eens en altijd inzichtelijk met onder- en bovensommen bewezen.

Opgave 2.24. Bewijs dat je antwoord in Opgave 2.10 juist is.

2.2 Toch maar wat meer over continuïteit en limieten

Uniform continue functies $f : [a, b] \to \mathbb{R}$ zijn dus integreerbaar met dezelfde aanpak als monotone functies. Maar hoe weet je of een functie uniform continu is? Wel, bijvoorbeeld als je een schatting van de vorm

$$|f(x) - f(y)| \leq C|x - y|^\alpha \tag{2.12}$$

hebt voor alle $x, y \in [a, b]$ met een vaste $\alpha > 0$ en een vaste $C > 0$. In het geval dat $\alpha \in (0, 1)$ spreekt men van uniform *Hölder continu* met exponent α, en als $\alpha = 1$ van uniform *Lipschitz continu*. Het geval $\alpha > 1$ is niet heel relevant[26].

Uniforme continuïteit betreft functies op hun hele domein, hierboven steeds $[a, b]$. *Gewone continuïteit* betreft het gedrag van een functie in de buurt van een gegeven punt.

[25] Zie Opgave 2.21.

[26] Waarom niet? Leuke opgave: bewijs dat zulke functies constant zijn.

Definitie 2.25. *Als we in Definitie 2.22 de y vast nemen en gegeven iedere $\varepsilon > 0$, altijd een $\delta > 0$ kunnen vinden waarvoor de implicatie geldt voor alle x met $x \in [a, b]$, dan heet f continu in y. We zeggen dan ook dat*

$$f(y) = \lim_{x \to y} f(x),$$

of ook wel

$$f(x) \to f(y)$$

als $x \to y$ (steeds met $x \in [a, b]$). Als deze uitspraak geldt met uitsluiting van $x = y$, en de functiewaarde $f(y)$ vervangen door een $L \in \mathbb{R}$, dan zeggen we dat L de limiet is van $f(x)$ voor $x \to y$.

In Definitie 2.25 kan $[a, b]$ vervangen worden door een willekeurige $I \subset \mathbb{R}$ en in de uitspraak $f(x) \to L$ hoeft y dan niet per se in I te liggen.

De voor de hand liggende vraag is natuurlijk of een $f : [a, b] \to \mathbb{R}$ die continu is in elke $y \in [a, b]$ ook uniform continu is op heel $[a, b]$. Daartoe herformuleren we nu eerst de definitie van $f(x) \to L$ als $x \to y$ in termen van *convergente rijen*.

Definitie 2.26. *Een door $n \in \mathbb{N}$ genummerde*[27] *rij reële getallen x_n heet convergent als er een (limiet) $\bar{x} \in \mathbb{R}$ bestaat zodanig dat er voor alle $\varepsilon > 0$ een $N \in \mathbb{N}$ bestaat waarvoor de implicatie*

$$n \geq N \implies |x_n - \bar{x}| \leq \varepsilon$$

geldt voor alle $n \in \mathbb{N}$. Zonder woorden luidt deze uitspraak

$$\exists \bar{x} \in \mathbb{R} \, \forall \varepsilon > 0 \, \exists N \in \mathbb{N} \, \forall n \in \mathbb{N} : n \geq N \implies |x_n - \bar{x}| \leq \varepsilon.$$

We zeggen dat $x_n \to \bar{x}$ (als $n \to \infty$).

Opgave 2.27. Laat zien dat een convergente rij reële getallen x_n maar één limiet heeft. Hint: stel niet dan zijn er minstens twee limieten, zeg $\bar{x}$ en $\hat{x}$; gebruik

$$|\bar{x} - \hat{x}| = |\bar{x} - x_n + x_n - \hat{x}| \leq |\bar{x} - x_n| + |x_n - \hat{x}|$$

en kies één geschikte ε om een tegenspraak af te leiden met Definitie 2.26.

[27] Of door $n \in k + \mathbb{N} = \{k + j : j \in \mathbb{N}\}$, met $k \in \mathbb{Z}$.

Opgave 2.28. Laat zien dat iedere convergente rij reële getallen x_n begrensd is, i.e. er is een $M \geq 0$ waarvoor geldt dat $|x_n| \leq M$ voor alle n in $\mathbb{N}$. Hint: pas Definitie 2.26 toe met één $\varepsilon > 0$ naar vrije keuze en kies M zo dat $M \geq |x_n|$ voor alle n kleiner dan de bijbehorende N en $M \geq |x_N| + \varepsilon$. Geef een voorbeeld van een begrensde rij die niet convergent en maak die rij convergent door voldoende veel termen van de rij weg[28] te laten.

Opgave 2.29. Laat $f : I \to \mathbb{R}$ en $y, L \in \mathbb{R}$. De uitspraak $f(x) \to L$ als $x \to y$ is equivalent met de uitspraak dat voor elke rij x_n in I de implicatie

$$x_n \to y \implies f(x_n) \to L$$

geldt. Bewijs dit. Hint: rijgen van definities, noem de ε in de uitspraak dat $x_n \to y$ daartoe maar δ.

NB. Het kan gebeuren in Opgave 2.29 dat $y \notin I$ en dat er geen rijen x_n in I zijn met $x_n \to y$. In dat geval zijn beide uitspraken zinloos[29] voor elke $L \in \mathbb{R}$. Zo'n punt $y \notin I$ heet *geen limietpunt* van I. Opgave 2.29 bewijst de volgende stelling[30]:

Stelling 2.30. *Laat $I \subset \mathbb{R}$, $f : I \to \mathbb{R}$ en $x_0 \in I$. Dan is f continu in x_0 dan en slechts dan als voor elke rij x_n in I met $x_n \to x_0$ geldt dat $f(x_n) \to f(x_0)$.*

Opgave 2.31. De enige manier waarop $f : [a, b] \to \mathbb{R}$ niet uniform continu kan zijn is als er een $\varepsilon > 0$ bestaat waarvoor geen bijbehorende $\delta > 0$ te vinden is waarmee de uitspraak in Definitie 2.22 geldt. Kies nu een willekeurige rij $0 < \delta_n \to 0$. Voor iedere $n \in \mathbb{N}$ zijn er dan $x_n, y_n \in [a, b]$ met wel $|x_n - y_n| \leq \delta_n$ maar niet $|f(x_n) - f(y_n)| \leq \varepsilon$. Neem nu aan[31] dat $y_n \to \eta \in [a, b]$ en leid een tegenspraak af met de continuïteit van f in η. Hint: laat zien dat ook $x_n \to \eta$; wat geldt er dus voor $f(x_n)$ en $f(y_n)$?

Stelling 2.32. *Laat $f : [a, b] \to \mathbb{R}$ continu zijn in elk punt van $[a, b]$. Dan is $f : [a, b] \to \mathbb{R}$ uniform continu.*

Bewijs. Stel dat $f : [a, b] \to \mathbb{R}$ niet uniform continu is. Redeneer dan als in Opgave 2.31: als de rij y_n niet convergent is dan kan de rij convergent

[28] Grapje: denk aan ongewenste data.
[29] Maar wel waar.
[30] Waarin n in $\mathbb{N}$ en 0 beschikbaar is voor x_0 als naam voor de limiet van de rij x_n.
[31] Absurde aanname?

gemaakt worden door voldoende elementen van de rij weg te laten (met andere woorden, neem een deelrij) zodanig dat de rij convergent wordt (en de andere rij dus ook). Gebruik hiertoe Stelling 3.1 die zegt dat *iedere begrensde rij in* $\mathbb{R}$ een convergente deelrij heeft. Noem de limiet van de deelrij η, laat zien dat $\eta \in [a, b]$ en het werk is gedaan.

De in het bewijs gebruikte Stelling 3.1 is belangrijk genoeg om een eigen hoofdstuk te krijgen. Het is de eerste stelling die we formuleren over het bestaan van een limiet van een (in dit geval deel-)rij zonder die limiet al op de een of andere manier bepaald te hebben. Een poging om het convergent zijn van een rij x_n te formuleren *zonder* de limiet in de definitie te gebruiken gaat terug tot Cauchy en correspondeert met we wat tegenwoordig een Cauchyrij noemen. Vergelijk de volgende definitie daarom nauwkeurig met de eerste zin in Definitie 2.26.

Definitie 2.33. *Een door* $n \in \mathbb{N}$ *genummerde rij reële getallen* x_n *heet een Cauchyrij als er voor alle* $\varepsilon > 0$ *een* $N \in \mathbb{N}$ *bestaat waarvoor de implicatie*

$$m, n \geq N \implies |x_n - x_m| \leq \varepsilon$$

geldt voor alle $m, n \in \mathbb{N}$.

In Definitie 2.33 is $m, n \geq N$ te lezen als $m \geq N$ en ook $n \geq N$. Evenzo is $m, n \in \mathbb{N}$ te lezen als $m \in \mathbb{N}$ en ook $n \in \mathbb{N}$. In het volgende hoofdstuk volgt het convergent zijn van Cauchyrijen via Stelling 3.1 uit het begrensd zijn van Cauchyrijen. Voor nu volstaat een eenmalig beroep op Stelling 3.1 zelf.

Opgave 2.34. Laat $f : [a, b] \to \mathbb{R}$ uniform continu zijn. Bewijs dat f begrensd is op $[a, b]$.

Opgave 2.35. Laat $f : [a, b] \to \mathbb{R}$ continu zijn in elk punt van $[a, b]$. Bewijs dat f een maximum M en een minimum m op $[a, b]$ heeft. Hint: omdat f uniform continu is op $[a, b]$, is f begrensd op $[a, b]$. Laat $M = \sup\{f(x) : a \leq x \leq b\}$. Dan bestaat er een rij x_n in $[a, b]$ met $f(x_n) \to M$. Neem daarvan een convergente deelrij.

2.3 Integreerbaar: definities, regels, voorbeelden

We weten tot nu toe alleen dat voor monotone en voor continue $f : [a, b] \to \mathbb{R}$ met behulp van onder- en bovensommen een ondubbelzinnige betekenis kan

worden gegeven aan de *integreerbaarheid* van f en de waarde van de integraal

$$\int_a^b f(x)\,dx.$$

Voor iedere begrensde functie kunnen we echter bij een partitie (2.4)

$$M_k = \sup_{[x_{k-1},x_k]} f \quad \text{en} \quad m_k = \inf_{[x_{k-1},x_k]} f$$

definiëren en de bijbehorende onder- en bovensom. Net als in het bewijs van Stelling 2.23 volgt nu dat de verzameling van alle ondersommen A en de verzameling van alle bovensommen B voldoen aan $A \leq B$. Op grond van Opgave 2.19 is er dus tenminste één getal J dat tussen A en B ligt.

Definitie 2.36. *Een begrensde functie $f : [a, b] \to \mathbb{R}$ heet Riemann integreerbaar als er precies één getal J tussen de verzameling A van alle ondersommen en de verzameling B van alle bovensommen ligt. In dat geval is per definitie*

$$J = \int_a^b f(x)\,dx.$$

Het is wel handig om een paar variaties op het thema te hebben bij de vraag welke functies nog meer integreerbaar zijn, en welke regeltjes er gelden die wel moeten gelden. De volgende opgaven zijn daartoe nuttig. Zorg dat je begrijpt wat de uitspraken zijn en waarom ze waar (moeten) zijn. Het begrip oneigenlijk integreerbaar komt ook langs en daar doen we vooral niet moeilijk over.

Opgave 2.37. Bewijs dat met

$$M_k = \sup\{f(x) : x_{k-1} < x < x_k\} \quad \text{en} \quad m_k = \inf\{f(x) : x_{k-1} < x < x_k\}$$

dezelfde begrensde functies integreerbaar worden als in Definitie 2.36, met dezelfde waarde van J.

Opgave 2.38. Bewijs dat de functie f gedefinieerd door $f(x) = 1$ voor alle $x \in \mathbb{Q}$ en $f(x) = 0$ voor alle $x \notin \mathbb{Q}$ niet integreerbaar is op $[0, 1]$.

Opgave 2.39. Bewijs dat de functie f gedefinieerd door $f(0) = 0$ en $f(x) = \sin \frac{1}{x}$ voor $x \neq 0$ integreerbaar is op $[0, 1]$. Hint: gebruik van $\sin$ alleen dat $x \to \sin x$ begrensd, periodiek en continu is. In welke x is $x \to \sin \frac{1}{x}$ dan continu? Waarom niet en waarom wel?

Opgave 2.40. Iedere begrensde integreerbare functie $f : [a,b] \to \mathbb{R}$ die op de punten van een partitie

$$a = x_0 < x_1 < \cdots < x_N = b \quad \text{met} \quad N \in \mathbb{N}$$

wordt veranderd door $g(x_k) = c_k$ te stellen voor $k = 0, \ldots, N$ voor gegeven waarden $c_k \in \mathbb{R}$, en $g(x) = f(x)$ voor alle $x \in [a,b]$ die niet to P behoren, definieert een begrensde integreerbare functie $g : [a,b] \to \mathbb{R}$ met $\int_a^b f(x)\,dx = \int_a^b g(x)\,dx$. Bewijs dit.

Opgave 2.41. Iedere begrensde functie $f : [a,b] \to \mathbb{R}$ waarvoor er een partitie

$$a = x_0 < x_1 < \cdots < x_N = b \quad \text{met} \quad N \in \mathbb{N}$$

bestaat met f monotoon of continu op elk open interval (x_{k-1}, x_k) is integreerbaar. Bewijs dit[32].

Definitie 2.42. *Als $a < b$ dan is*

$$\int_a^b f(x)\,dx = -\int_b^a f(x)\,dx,$$

vermits $f : [b,a] \to \mathbb{R}$ begrensd en integreerbaar is[33].

Opgave 2.43. Laat $f : [a,b] \to \mathbb{R}$ begrensd zijn en $c \in (a,b)$. Bewijs dat f integreerbaar is op $[a,b]$ dan en slechts dan als f integreerbaar is op zowel $[a,c]$ als $[c,b]$, en dat in dat geval

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx.$$

Definitie 2.44. *Laat $-\infty \le a < b \le \infty$ en $f : (a,b) \to \mathbb{R}$ integreerbaar zijn op elke interval $[A,B] \subset (a,b)$. Dan is per definitie de (eventueel) oneigenlijke*[34] *integraal van a tot b gelijk aan*

$$\int_a^b f(x)\,dx = \lim_{A \downarrow a} \lim_{B \uparrow b} \int_A^B f(x)\,dx,$$

als deze limiet van een limiet bestaat.

[32] Hiermee ben je van een hoop gezeur af.

[33] Consistent met de intuïtie dat dx in $\int_a^b f(x)\,dx$ negatief is als $a > b$.

[34] Denk aan a en of b niet eindig, of f niet gewoon integreerbaar op $[a,b]$.

Limieten van limieten zijn verwarrend. Daarom deze opgave die via splitsen tot (twee) gewone limieten leidt, zoals dat bij calculus doorgaans aangeleerd wordt.

Opgave 2.45. Neem in Definitie 2.44 een $c \in (a,b)$. Bewijs dat

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx$$

als beide integralen, eventueel als

$$\int_a^c f(x)\,dx = \lim_{A\downarrow a} \int_A^c f(x)\,dx \quad \text{en} \quad \int_c^b f(x)\,dx = \lim_{B\uparrow b} \int_c^B f(x)\,dx,$$

bestaan[35].

Terug naar de gewone integralen merken we nu op dat de Riemann integreerbare functies een (vector)ruimte[36] over $\mathbb{R}$ vormen. As je nog niet weet wat een vectorruimte is lees dan de voetnoot, doe even alsof

$$f = \begin{pmatrix} f_1 \\ f_2 \end{pmatrix} \quad \text{en} \quad g = \begin{pmatrix} g_1 \\ g_2 \end{pmatrix}$$

2-vectoren in $\mathbb{R}^2$ zijn, waarvoor de gebruikelijk vectoroptelling en vermenigvuldiging met scalairen λ, μ uit het lichaam $\mathbb{R}$ wel bekend zullen zijn, en je begrijpt waarom $\mathbb{R}^2$ een vectorruimte genoemd wordt. Iedere $f \in \mathbb{R}^2$ heeft coördinaten f_i genummerd door $i \in \{1,2\}$ en met functies $f : [a,b] \to \mathbb{R}$ is het net zo. De "coördinaten" van f worden alleen niet genoteerd met f_x maar met $f(x)$, genummerd door $x \in [a,b]$ (en niet in een kolommetje gezet). De verzameling van alle functies $f : [a,b] \to \mathbb{R}$ is zo op een vanzelfsprekende manier net zo'n vectorruimte als $\mathbb{R}^2$.

Meer over vectorruimten hoef je hier voorlopig niet te weten, maar is het wel goed om op te merken dat dat deze vectorruimte van alle functies $f : [a,b] \to \mathbb{R}$ een heleboel deelvectorruimten heeft, bijvoorbeeld $\mathrm{RI}([a,b])$, de verzameling van alle Riemann integreerbare functies $f : [a,b] \to \mathbb{R}$.

Opgave 2.46. De verzameling van begrensde integreerbare functies $f : [a,b] \to \mathbb{R}$ is een vectorruimte[37] over $\mathbb{R}$, hier $\mathrm{RI}([a,b])$ genaamd. Bewijs dit, en ook dat

$$f \to \int_a^b f(x)\,dx$$

[35] Vaak schrijft men $\lim_{A\downarrow a} = \lim_{A\to a^+}$ en $\lim_{B\uparrow b} = \lim_{B\to b^-}$.

[36] Met f, g Riemann integreerbaar is $\lambda f + \mu g$ het ook voor alle $\lambda, \mu \in \mathbb{R}$.

[37] Is met met $f, g \in \mathrm{RI}([a,b])$ ook $fg \in \mathrm{RI}([a,b])$? Hint: neem f, g eerst positief.

een lineaire[38] afbeelding is van deze vectorruimte naar $\mathbb{R}$. Hint: laat met name zien dat met $f, g \in \mathrm{RI}([a,b])$ ook $f+g \in \mathrm{RI}([a,b])$ en

$$\int_a^b (f(x)+g(x))\,dx = \int_a^b f(x)\,dx + \int_a^b g(x)\,dx.$$

Opgave 2.47. Als $f \in \mathrm{RI}([a,b])$ dan is ook $x \to |f(x)|$ begrensd en integreerbaar op $[a,b]$ en geldt de "3-hoeksongelijkheid"

$$|\int_a^b f(x)\,dx| \leq \int_a^b |f(x)|\,dx,$$

te vergelijken met (25). Bewijs dit rechtstreeks uit de definities en geef een voorbeeld van een begrensde functie $f : [a,b] \to \mathbb{R}$ die niet integreerbaar is waarvoor $x \to |f(x)|$ het wel is.

Opmerking 2.1. *Een functie $f : [a,b] \to \mathbb{R}$ kan alleen maar Riemann integreerbaar zijn als f begrensd is. Maar welke begrensde functies integreerbaar zijn is een vraag die we nog niet gesteld hebben. Zonder bewijs: het antwoord is dat daartoe de verzameling $\mathcal{N}$ van x-waarden waarin f niet continu is een zogenaamde nulverzameling moet zijn. Dat is een verzameling die voor elke $\varepsilon > 0$ overdekt kan worden met intervallen $[a_n, b_n]$ genummerd door $n \in \mathbb{N}$, waarbij voor elke $N \in \mathbb{N}$ geldt dat*

$$\sum_{n=1}^{N} (b_n - a_n) \leq \varepsilon.$$

Overdekken betekent hier dat

$$\mathcal{N} \subset \cup_{n \in \mathbb{N}} [a_n, b_n],$$

en de overdekking hangt af van ε. Intuïtief: met een willekeurig klein intervalletje opgegeknipt in aftelbaar veel stukjes is $\mathcal{N}$ te bedekken.

2.4 Terug- en vooruitblik

Begrensde functies $f : [a,b] \to \mathbb{R}$ kunnen via onder- en bovensommen die willekeurig dicht bij elkaar liggen integreerbaar zijn met integraal $\int_a^b f(x)\,dx$.

[38] Wat lineair betekent staat op de volgende bladzijde.

Zulke integreerbare functies vormen een lineaire vectorruimte over $\mathbb{R}$, hier genoteerd met $\mathrm{RI}([a,b])$. Deze vectorruimte bevat alle (uniform) continue functies $f : [a,b] \to \mathbb{R}$, die zelf ook een lineaire vectorruimte over $\mathbb{R}$ vormen, doorgaans met $C([a,b])$ aangeduid. Ook bevat $\mathrm{RI}([a,b])$ alle monotone functies $f : [a,b] \to \mathbb{R}$ maar die vormen geen vectorruimte[39].

De vectorruimte $\mathrm{RI}([a,b])$ is zelf ook weer een deelvectorruimte van alle functies $f : [a,b] \to \mathbb{R}$, die met de bewerkingen gedefinieerd door

$$(f+g)(x) = f(x) + g(x), \quad (\lambda f)(x) = \lambda f(x)$$

voor alle $f, g : [a,b] \to \mathbb{R}$, alle $\lambda \in \mathbb{R}$ en alle $x \in [a,b]$, een lineaire vectorruimte over $\mathbb{R}$ vormen, waarbinnen de integraaltheorie zoals hierboven behandeld zich afspeelt.

De vectorruimte $C([a,b])$ kan voorzien worden van diverse normen[40], bijvoorbeeld de maximum-norm

$$f \to \max_{a \le x \le b} |f(x)| = |f|_\infty ,$$

of de zogenaamde 1-norm

$$f \to \int_a^b |f(x)|\, dx = |f|_1 .$$

Met een norm kunnen we praten over convergentie, waarbij $f_n \to f$ betekent dat $|f_n - f| \to 0$ als $n \to \infty$. Alles hierboven dat met convergente rijen is gekarakteriseerd kan worden overschreven, maar begrensde rijen hebben i.h.a. geen convergente deelrijen.

De maximum-norm maakt van $C([a,b])$ een genormeerde ruimte waarin Cauchyrijtjes convergent zijn. Zulke ruimten heten *Banachruimten*, abstract vaak genoteerd met hoofdletters achterin het alphabet, zie Hoofdstuk 10. Met deze norm is de lineaire afbeelding in Opgave 2.46 continu in de zin dat $f_n \to f$ in de maximum norm (uniforme convergentie) impliceert dat

$$\int_a^b f_n \to \int_a^b f,$$

een uitspraak die onder vele andere zwakkere voorwaarden ook geldt. Merk op dat $C([a,b])$ ook een algebra[41] is, i.e. met $f, g \in C([a,b])$ zit ook $x \to f(x)g(x)$ in $C([a,b])$ en aan een voor de hand liggend lijstje[42] axioma's is voldaan.

In Hoofdstuk 4 beginnen we met de differentiaalrekening[43] voor functies

[39] Waarom niet?

[40] Wikipidia: normed vector space: Hier: notatie met enkele strepen, $|x|$ i.p.v. $||x||$.

[41] Een commutatieve Banachalgebra zelfs.

[42] See in wikipedia: algebra over a field.

[43] Af en toe terugbladerend verder lezen vanaf (9.37) is nu een optie.

van $\mathbb{R}$ naar $\mathbb{R}$, om uiteindelijk ook differentiaalrekening te doen voor functies[44] van X naar Y. Daartoe zullen we ook integraalrekening voor functies van $\mathbb{R}$ naar X nodig hebben. De beschouwingen over integralen van functies van $\mathbb{R}$ naar $\mathbb{R}$ sluiten we nu af met een reflecterende opgave, en in $\mathbb{R}$ is er dan eerst nog een hoofdstukje over die convergente deelrijen en zo.

Teken plaatje!

Opgave 2.48. Teken een plaatje met een x-as en een y-as zonder pijltjes, markeer een $x = a$ en een $x = b$ op de x-as en teken een grafiek van zo maar een continue $f(x)$ zonder nulpunten of tekenwisselingen gedefinieerd voor x tussen a en b. Kies pas na dat je dit gedaan hebt de pijltjes op de assen die aangeven of x naar links of naar rechts loopt en y naar beneden of naar boven. Er zijn vier mogelijkheden om dat te doen. Relateer in alle vier de gevallen de integraal

$$\int_a^b f(x)\,dx$$

aan de (bij algemene afspraak) positieve oppervlakte van het gebied ingesloten door $x = a$, $x = b$, $y = 0$ en $y = f(x)$. In welk geval is zowel f als $\int_a^b f(x)\,dx$ positief? Verifieer dat dan $b > a$ en dat je de positieve x-as 3 uren tegen de klok in om de oorsprong moet draaien om de positieve y-as te krijgen. Hint: als je dit een stomme opgave vindt lees dan alvast Sectie 6.9 en kijk naar Opgave 6.21.

2.5 Verdiepingsmoment: tussensommen

We zijn begonnen met onder- en bovensommen om integralen van monotone functies $f : [a, b] \to \mathbb{R}$ zo eenvoudig mogelijk te behandelen. Voor continue functies kunnen we echter net zo goed kijken naar sommen van de vorm

$$S_N = \sum_{k=1}^{N} f(\xi_k)(x_k - x_{k-1}), \quad \text{met} \quad \xi_k \in [x_{k-1}, x_k] \tag{2.13}$$

gegeven de partitie (2.4) vrij te kiezen voor elke $k = 1, \ldots, N$. Vanzelfsprekend geldt

$$\underline{S}_N \leq S_N \leq \bar{S}_N,$$

met $\underline{S}_N$ en $\bar{S}_N$ als in (2.8) en het bewijs van Stelling 2.23 herlezend is via (2.10) duidelijk dat ook $|S_N - J| \leq \varepsilon(b - a)$ als $x_k - x_{k-1} \leq \delta$ voor alle $k = 1, \ldots, N$.

Het is nu een aardige en nuttige oefening om wederom via een rij equidistante partities met $N = 1, 2, 4, 8, 16, \ldots$, met voor ξ_k steeds $\xi_k = x_k$

[44] Of, ander woord, afbeeldingen.

gekozen, direct en *zonder de ordening van boven- en ondersommen* te gebruiken te laten zien dat er een limietwaarde J bestaat, en daarna voor deze J op dezelfde conclusie uit te komen.

Stelling 2.49. *Als $f : [a, b] \to \mathbb{R}$ uniform continu is dan is er een unieke $J \in \mathbb{R}$ waarvoor voor iedere S_N als in (2.13) geldt dat $|S_N - J| \leq \varepsilon(b - a)$, mits $x_k - x_{k-1} \leq \delta$ voor alle $k = 1, \ldots, N$, waarbij $\delta > 0$ bij $\varepsilon > 0$ gekozen is zoals in de definitie van uniforme continuïteit van $f : [a, b] \to \mathbb{R}$.*

Opgave 2.50. Bewijs Stelling 2.49 zonder de ordening van boven- en ondersommen te gebruiken. Hint: nodig voor zo'n bewijs is de convergentie van Cauchyrijen.

3 Convergente (deel)rijen en zo

In het bewijs van Stelling 2.32 hebben we voor het eerst gebruik gemaakt van de volgende stelling, waarvan het belang in de analyse moeilijk is te onderschatten, ook als je van mening bent dat'ie eigenlijk triviaal is en verder geen betoog behoeft.

Stelling 3.1. *Iedere begrensde rij x_n in $\mathbb{R}$ heeft een convergente deelrij.*

De lezer die de verhaallijn over integraal- en differentiaalrekening niet te veel wil onderbreken kan eventueel de volgende drie secties overslaan en doorlezen in Sectie 3.6 voor een direct bewijs van Stelling 3.1, maar mist dan leerzame oefeningen met kwantoren en een belangrijke techniek uit de analyse: de truc van de diagonaalrij, waarmee meerdere belangrijke stellingen worden bewezen, waaronder een stelling over begrensde rijen in $C([a,b])$.

3.1 Spelen met kwantoren

Voor we laten zien waarom deze stelling eigenlijk triviaal is spelen we eerst nog wat met de kwantoren $\forall$ en $\exists$.

De kale definitie in Definitie 2.26 met kwantoren van het convergent zijn van de rij x_n wordt ook wel getekstzet als

$$\exists_{\bar{x}\in\mathbb{R}}\,\forall_{\varepsilon>0}\,\exists_{N\in\mathbb{N}}\,\forall_{n\in\mathbb{N}}\ :\ n\geq N\implies |x_n-\bar{x}|\leq\varepsilon,$$

hetgeen misschien net iets prettiger oogt dan alles op dezelfde regelhoogte. We kunnen ook zeggen dat $\bar{x}\in\mathbb{R}$ een limiet is van de rij $x_n\in\mathbb{R}$ als geldt dat

$$\forall_{\varepsilon>0}\,\exists_{N\in\mathbb{N}}\,\forall_{n\in\mathbb{N}}\ :\ n\geq N\implies |x_n-\bar{x}|\leq\varepsilon,$$

en vergelijken deze definitie met de kale versie van $f:[a,b]\to\mathbb{R}$ uniform continu:

$$\forall_{\varepsilon>0}\,\exists_{\delta>0}\,\forall_{x,y\in[a,b]}\ :\ |x-y|\leq\delta\implies |f(x)-f(y)|\leq\varepsilon. \tag{3.1}$$

Beide definities eindigen met een implicatie, maar $\bar{x}\in\mathbb{R}$ limiet van de rij $x_n\in\mathbb{R}$ schrijven we net zo lief als

$$\forall_{\varepsilon>0}\,\exists_{N\in\mathbb{N}}\,\forall_{n\geq N}\ :\ |x_n-\bar{x}|\leq\varepsilon,$$

waarin we stilzwijgend n altijd in $\mathbb{N}$ nemen. De variant op deze uitspraak die Cauchy bedacht, zie Definitie 2.33, is

$$\forall_{\varepsilon>0}\,\exists_{N\in\mathbb{N}}\,\forall_{m,n\geq N}\ :\ |x_n-x_m|\leq\varepsilon,$$

een rij die hieraan voldoet noemen we een Cauchyrij.

De (logische) ontkenning van $f : [a, b] \to \mathbb{R}$ uniform continu is gebruikt in Opgave 2.31. Die ontkenning krijg[1] je door in een keten van zulke kwantoren de verwisseling

$$\forall \leftrightarrow \exists$$

toe te passen en de laatste uitspraak te ontkennen. Dit geeft

$$\exists_{\varepsilon>0}\, \forall_{\delta>0}\, \exists_{x,y\in[a,b]}\ :\ |x-y| \leq \delta \wedge |f(x)-f(y)| > \varepsilon$$

als ontkenning van de uniforme continuiteit van $f : [a, b] \to \mathbb{R}$, waarbij het de implicatie is die in de laatste stap ontkend is: x en y voldoen wel $|x - y| \leq \delta$ maar het is niet zo dat $|f(x) - f(y)| \leq \varepsilon$.

Evenzo is de ontkenning van het convergent zijn van de rij $x_n \in \mathbb{R}$ te formuleren met als laatste stap de ontkenning van de implicatie, maar de formulering dat $\bar{x} \in \mathbb{R}$ geen limiet is van de rij $x_n \in \mathbb{R}$ is natuurlijk ook gewoon de ontkenning van de uitspraak zonder implicatie, i.e.

$$\exists_{\varepsilon>0}\, \forall_{N\in\mathbb{N}}\, \exists_{n\geq N}\ :\ |x_n - \bar{x}| > \varepsilon.$$

Dat en-tekentje $\wedge$ voor de ontkenning van een implicatie is dan niet nodig.

Opgave 3.2. Formuleer de logische ontkenningen van alle quantorenuitspraken hierboven.

3.2 Elementaire opgaven over convergente rijen

Een sectie om over te slaan als je de verhaallijn integraal- en differentiaalrekening niet te veel wil onderbreken, maar deze opgaven zijn een must voor de beheersing van de stof.

Opgave 3.3. Iedere convergente rij in $\mathbb{R}$ is begrensd. Bewijs dit.

Opgave 3.4. Iedere Cauchyrij in $\mathbb{R}$ is begrensd. Bewijs dit.

[1] Overtuig zelf hiervan zonder waarheidstabellen.

Opgave 3.5. Iedere convergente rij in $\mathbb{R}$ is een Cauchyrij. Bewijs dit.

Opgave 3.6. Als x_n en y_n convergente rijen zijn in $\mathbb{R}$ dan is de rij $x_n + y_n$ ook convergent in $\mathbb{R}$ met de limiet die je verwacht. Bewijs dit.

Opgave 3.7. Als x_n en y_n convergente rijen zijn in $\mathbb{R}$ dan is de rij $x_n y_n$ ook convergent in $\mathbb{R}$ met de limiet die je verwacht. Bewijs dit.

Opgave 3.8. Als x_n een convergente rij is in $\mathbb{R}$ met $x_n \neq 0$ voor alle n dan is de rij $\frac{1}{x_n}$ ook convergent in $\mathbb{R}$ met de limiet die je verwacht, tenzij de limiet van de rij x_n gelijk is aan 0. Bewijs dit.

Opgave 3.9. Als x_n een convergente rij is in $\mathbb{R}$ dan is de rij $|x_n|$ ook convergent in $\mathbb{R}$ met de limiet die je verwacht. Bewijs dit.

3.3 Limietpunten en convergente deelrijen

Het grappige is dat een wellicht per abuis opgeschreven mix van de uitspraak dat $\bar{x} \in \mathbb{R}$ de limiet van de rij $x_n \in \mathbb{R}$ en de uitspraak dat $\bar{x} \in \mathbb{R}$ juist geen limiet is van de rij $x_n \in \mathbb{R}$ precies de definitie is van een nieuw belangrijk begrip: we zeggen dat $\bar{x} \in \mathbb{R}$ *limietpunt* van de rij $x_n \in \mathbb{R}$ is als

$$\forall_{\varepsilon>0}\, \forall_{N\in\mathbb{N}}\, \exists_{n\geq N} \;:\; |x_n - \bar{x}| \leq \varepsilon, \tag{3.2}$$

hetgeen iets anders is dan $\bar{x} \in \mathbb{R}$ is limiet van de rij $x_n \in \mathbb{R}$, want dat was

$$\forall_{\varepsilon>0}\, \exists_{N\in\mathbb{N}}\, \forall_{n\geq N} \;:\; |x_n - \bar{x}| \leq \varepsilon. \tag{3.3}$$

De uitspraak in (3.2) is er een om goed te lezen. *Je ziet namenlijk vast niet meteen deze uitspraak equivalent is met het bestaan van een convergente deelrij van x_n met limiet $\bar{x}$. Om te onthouden nu is dat een rij met een limiet begrensd is en maar één limiet heeft, terwijl een begrensde rij altijd tenminste één limietpunt (de limiet van een convergente deelrij) heeft*[2]. Merk

[2] En wellicht een heleboel limietpunten, je kunt het zo gek niet verzinnen.

op dat iedere strict stijgende rij $n_k \in \mathbb{N}$ een deelrij van de rij x_n in $\mathbb{R}$ definieert, genummerd door k in x_{n_k}. Als we het hebben over een deelrij van x_n is dit precies wat we bedoelen: we maken de deelrij door uit de door n genummerde rij x_n alle x_n weg te laten waarvoor n niet als n_k in de gegeven strict stijgende rij $n_k \in \mathbb{N}$ voorkomt. De laatste cursief gedrukte uitspraak is een herformulering van Stelling 3.1 waar dit hoofdstuk in eerste instantie over gaat.

Opgave 3.10. Bewijs dat een limietpunt van de rij x_n altijd de limiet is van een convergente deelrij x_{n_k}.

Opgave 3.11. Bewijs dat iedere convergente rij begrensd is en geef een voorbeeld van een rij $x_n \in [0,1]$ die als limietpunten alle punten in $[0,1]$ heeft.

Bewijs van Stelling 3.1. Neem nu een rij $x_n \in \mathbb{R}$ die begrensd is, zeg zonder beperking der algemeenheid geheel bevat in $[0,1]$. Dan moet tenminste één van de twee intervallen $[\frac{0}{2}, \frac{1}{2}]$ of $[\frac{1}{2}, \frac{2}{2}]$ voor oneindig veel n het element x_n bevatten. Noem dat interval $I_1 = [\frac{m_1}{2}, \frac{m_1+1}{2}]$. Dus $m_1 = 0$ of $m_1 = 1$, al naar gelang het interval waarin die oneindig veel x_n zitten. Nummer de bijbehorend n met een index j als een strict stijgende rij $n_{1j} \in \mathbb{N}$. De eerste index 1 geeft aan dat dit de eerste deelrij is die we kiezen.

Herhaal vervolgens het argument voor deze deelrij maar nu met de intervallen $[\frac{m_1}{2} + \frac{0}{4}, \frac{m_1}{2} + \frac{1}{4}]$ en $[\frac{m_1}{2} + \frac{1}{4}, \frac{m_1}{2} + \frac{2}{4}]$. Er is dus een verdere deelrij die geheel in één van de twee intervallen zit. Noem het betreffende interval $I_2 = [\frac{m_1}{2} + \frac{m_2}{4}, \frac{m_1}{2} + \frac{m_2+1}{4}]$ en de indices van die verdere deelrij $n_{2j} \in \mathbb{N}$. Enzovoorts. Dit geeft

$$x_{n_{kj}} \in I_k = [\sum_{l=1}^{k} \frac{m_l}{2^l}, \sum_{l=1}^{k} \frac{m_l}{2^l} + \frac{1}{2^{k+1}}]$$

voor alle $j \in \mathbb{N}$. Omdat we steeds verdere deelrijen nemen zijn de n_{kj} strict stijgend in j en niet-dalend in k. We zien dat $x_{n_{kk}}$ een deelrij van de (door $n \in \mathbb{N}$ genummerde) rij x_n is omdat $n_{kk} \in \mathbb{N}$ een strict stijgende rij is.

Opgave 3.12. Maak het bewijs af door de laten zien dat de door k genummerde "diagonaalrij" $x_{n_{kk}}$ een convergente deelrij is. Bewijs dat de limiet $\bar{x}$ van deze deelrij een limietpunt is van de rij, dus dat voldaan is aan (3.2).

3.4 Begrensde rijen continue functies

We merken we op dat in onze mooie Banachruimte[3] $C([a,b])$ de uitspraak over convergente deelrijen van begrensde rijen niet waar is. Niet elke begrensde rij $f_n \in C([a,b])$ heeft een deelrij die in de maximumnorm convergent is[4]. Maar wel als de rij aan deze extra aanname voldoet:

$$\forall_{\varepsilon>0}\,\exists_{\delta>0}\,\forall_{n\in\mathbb{N}}\,\forall_{x,y\in[a,b]}\,:\,|x-y|\leq\delta \implies |f_n(x)-f_n(y)|\leq\varepsilon. \tag{3.4}$$

De rij f_n heet dan equicontinu op $[a,b]$.

De stelling van Ascoli-Arzéla zegt nu dat iedere begrensde en op $[a,b]$ equicontinue rij $f_n : [a,b] \to \mathbb{R}$ een uniform convergente[5] deelrij heeft met limiet $\bar{f}$ in $C([a,b])$. Deze stelling is iets minder[6] triviaal maar je bewijst hem met Stelling 3.1, en weer een "diagonaalargument".

Opgave 3.13. Moeilijk: formuleer en bewijs de Stelling van Ascoli-Arzéla zoals hierboven verwoord. Hint: neem zonder beperking der algemeenheid $[a,b]=[0,1]$ en gebruik diagonaalargumenten om een deelrij f_{n_k} te maken die in elk binaire getal

$$\frac{k}{2^l}$$

convergeert. Met de equicontinuiteit van de rij f_n kun je daarna bewijzen dat f_{n_k} een Cauchyrij is in de maximumnorm.

3.5 Overaftelbaarheid van de reële getallen

De diagonaalargumenten hierboven zijn varianten op het argument in het bewijs dat ($\mathbb{Q}$ wel maar) $\mathbb{R}$ geen aftelbare verzameling is. Een bekend bewijs uit het ongerijmde gaat uit van een aftelling van alle getallen tussen 0 en 1 met

$$r_k = \sum_{j=1}^{\infty} \frac{n_{kj}}{10^j},$$

waarbij de decimalen n_{kj} in een blok

[3] Met de maximumnorm.

[4] Verzin een voorbeeld met functies f_n die alleen waarden tussen 0 en 1 aannemen.

[5] Moeilijke woorden voor convergentie in de maximumnorm.

[6] Lees: gewoon moeilijk.

$$\begin{array}{cccccccccc}
n_{11} & n_{12} & n_{13} & n_{14} & n_{15} & n_{16} & n_{17} & n_{18} & \cdots \\
n_{21} & n_{22} & n_{23} & n_{24} & n_{25} & n_{26} & n_{27} & n_{28} & \cdots \\
n_{31} & n_{32} & n_{33} & n_{34} & n_{35} & n_{36} & n_{37} & n_{38} & \cdots \\
n_{41} & n_{42} & n_{43} & n_{44} & n_{45} & n_{46} & n_{47} & n_{48} & \cdots \\
n_{51} & n_{52} & n_{53} & n_{54} & n_{55} & n_{56} & n_{57} & n_{58} & \cdots \\
n_{61} & n_{62} & n_{63} & n_{64} & n_{65} & n_{66} & n_{67} & n_{68} & \cdots \\
n_{71} & n_{72} & n_{73} & n_{74} & n_{75} & n_{76} & n_{77} & n_{78} & \cdots \\
n_{81} & n_{82} & n_{83} & n_{84} & n_{85} & n_{86} & n_{87} & n_{88} & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots &
\end{array}$$

gezet worden dat naar rechts en naar onder doorloopt. Gegeven de doorlopende diagonaalrij n_{kk} kies je dan decimalen d_k met $|d_k - n_{kk}| = 2$ of zo. Het getal

$$\sum_{j=1}^{\infty} \frac{d_j}{10^j}$$

kan dan niet in de rij r_k voorkomen.

3.6 Directer bewijs voor de convergente deelrijstelling

Wat nu volgt is de uitwerking van een opgave die Han Peters me ooit gaf: bewijs dat een rij x_n altijd een monotone deeltij heeft.

Als een rij $x_n \in \mathbb{R}$ begrensd is dan heeft'ie een kleinste bovengrens $M_1 \in \mathbb{R}$. Kan het zijn dat M_1 geen rijelement is? Vast wel, en als dat het geval is dan moeten er dus rijelementen dicht van onder bij M_1 liggen. Bijvoorbeeld boven de $M_1 - 1$, dus zeg dat x_{n_1} het eerste rijelement is waarvoor

$$M_1 - 1 < x_{n_1} < M_1$$

geldt. Omdat x_{n_1} zelf geen bovengrens kan zijn, is er dus ook een eerste n_2 waarvoor

$$M_1 - 1 < x_{n_1} < x_{n_2} < M_1,$$

en dan geldt $n_2 > n_1$ want anders was x_{n_1} niet de eerste in de rij die boven[7] $M_1 - 1$ ligt. En zo gaat dat verder. De conclusie is dat een kleinste bovengrens M_1 van een rij $x_n \in \mathbb{R}$ die zelf niet in de rij voorkomt, de limiet is van een strict stijgende rij x_{n_k} in $\mathbb{R}$, met n_k een strict stijgende rij n_k in $\mathbb{N}$.

[7] Voor de beeldvorming zien we $\mathbb{R}$ hier even als verticale getallenlijn.

Wat nu als we aannemen dat de rij x_n zulke strict stijgende deelrijen helemaal niet heeft? Dan is M_1 dus een rijelement, zeg $M_1 = x_{n_1}$ (met een totaal andere n_1 dan hierboven[8]). Maar de deelrij x_n die begint bij $n = n_1+1$ heeft dan een kleinste[9] bovengrens M_2 die op zijn hoogst gelijk is aan M_1, en die M_2 moet weer aangenomen worden door een x_{n_2} met $n_2 > n_1$. Dat gaat zo weer door en dus is er dan een strict stijgende rij n_k in $\mathbb{N}$ waarvoor x_{n_k} een niet-stijgende rij in $\mathbb{R}$ is.

We concluderen dat iedere begrensde rij x_n in $\mathbb{R}$ een (niet per se strict) monotone deelrij x_{n_k} heeft, en als je er even over nadenkt dan geldt de uitspraak voor elke rij in $\mathbb{R}$. Deze stelling, die Han weer kende uit het boek Elementary Analysis van Ross, trivialiseert[10] Stelling 3.1, wederom vanwege het bestaan van grootste ondergrenzen en kleinste bovengrenzen van begrensde verzamelingen in $\mathbb{R}$, en dus onthouden we hem!

Stelling 3.14. *Iedere rij x_n in $\mathbb{R}$ heeft een monotone deelrij.*

Voor wie de twee secties hierboven heeft overgeslagen, de stelling in Opgave 3.13 is wel erg relevant voor later, maar merk nu in de context van rijen en reeksen in $\mathbb{R}$ vooral op hoe de conclusies aan het eind van Sectie 26.3 met het bestaan van kleinste bovengrenzen samenhangen. Daarom na de volgende opgaven nog een belangrijke subsectie over (onvoorwaardelijke) convergentie van reeksen.

Opgave 3.15. Vul de details in: iedere Cauchyrij x_n in $\mathbb{R}$ is begrensd, heeft een monotone en dus convergente deelrij, en is daarmee convergent.

Opgave 3.16. Achteraf merken we nu op dat Stelling 3.14 ook geldt voor rijen in $\mathbb{Q}$, want $\mathbb{Q} \subset \mathbb{R}$. Dus dat moet te bewijzen zijn zonder die kleinste bovengrenzen die alleen maar in $\mathbb{R}$ altijd bestaan. Denk daar maar eens over na.

---

[8] Nog afgezien van het feit dat die n_1 hierboven nu niet meer hoeft te bestaan.

[9] Als $s_k = \sup_{n\geq k} x_n$ dan is s_k een niet-stijgende rij met limiet

$$\lim_{k\to\infty} s_k = \inf_{k\in\mathbb{N}} s_k = \inf_{k\in\mathbb{N}} \sup_{n\geq k} x_n = \limsup_{n\to\infty} x_n$$

[10] Maakt'ie ook het diagonaalargument voor de overaftelbaarheid van $\mathbb{R}$ overbodig?

3.7 Onvoorwaardelijke convergentie

Dat gezeur over die convergentie van reeksen willen we zo veel mogelijk vermijden. Maar toch. We willen nu eenmaal met aftelbare sommen als

$$S = \sum_{n=0}^{\infty} a_n = a_0 + a_1 + a_2 + \cdots \tag{3.5}$$

rekenen alsof het getallen zijn en alles mag[11]. En wat moet mogen is

$$a_0 + a_1 = a_1 + a_0,$$

de volgorde van optellen moet niet uitmaken. Voor sommen als in (3.5) betekent dit dat we willen dat

$$S = \sum_{n=0}^{\infty} a_{\phi(n)} \tag{3.6}$$

voor elke bijectie[12] $\phi : \mathbb{N}_0 \to \mathbb{N}_0$, maar waarom zou dit zo moeten zijn in de alledaagse wiskundige werkelijkheid?

De basale aanname die we steeds maken is (23), i.e.

$$\sum_{n=0}^{\infty} |a_n| < \infty,$$

te lezen als dat de rij partiële sommen

$$\sum_{n=0}^{N} |a_n|$$

begrensd is, met dus een kleinste bovengrens die we T noemen. Bij iedere $\varepsilon > 0$ is er dan een $N = N_\varepsilon \in \mathbb{N}_0$ waarvoor met $m = N_\varepsilon$ geldt dat

$$T - \varepsilon < \sum_{n=0}^{N_\varepsilon} |a_n| \leq T, \tag{3.7}$$

want anders was T geen *kleinste* bovengrens. Bijgevolg is

$$\sum_{n=N_\varepsilon+1}^{m} |a_n| < \varepsilon \tag{3.8}$$

[11] Lees: goed is.

[12] Ongetwijfeld kent de lezer deze tot het algemene cultuurgoed behorende term.

voor alle $m > N_\varepsilon$, want anders was T geen *bovengrens*.

Hopelijk betekent dit voor

$$S_N = \sum_{n=0}^{N} a_n \quad \text{en} \quad S_N^\phi = \sum_{n=0}^{N} a_{\phi(n)}$$

en

$$T_N = \sum_{n=0}^{N} |a_n| \quad \text{en} \quad T_N^\phi = \sum_{n=0}^{N} |a_{\phi(n)}|$$

dat

$$S_N \to S, \quad S_N^\phi \to S, \quad T_N \to T, \quad T_N^\phi \to T \tag{3.9}$$

als $N \to \infty$, en

$$|S| \le T, \tag{3.10}$$

maar voorlopig weten we behalve $|S_N| \le T$ en $|S_N^\phi| \le T$ alleen dat $T_N \to T$.

Het gaat dus nog om convergentie van de andere drie in (3.9). Hoe zit het met T_N^ϕ? De bijectie $\phi : \mathbb{N}_0 \to \mathbb{N}_0$ is een permutatie van $\mathbb{N}_0$. Als we

$$\mathbb{N}_0 = \{n = \phi(m) : \, m \in \mathbb{N}_0\}$$

aftellen met $m \in \mathbb{N}_0$, dan hebben we vroeg of laat, zeg met $m = 0, 1 \ldots, M_\varepsilon$, alle $n \in \mathbb{N}_0$ tot en met N_ε gehad. Bijgevolg is

$$T - \varepsilon < T_{N_\varepsilon} \le T_{M_\varepsilon}^\phi \le T.$$

Voor de gewone partiële sommen hebben we

$$|S_{M_\varepsilon}^\phi - S_{N_\varepsilon}| \le \sum_{n=N_\varepsilon+1}^{\infty} |a_n| \le \varepsilon$$

want $S_{M_\varepsilon}^\phi - S_{N_\varepsilon}$ is een eindige som van termen a_n met alleen maar $n > N_\varepsilon$, en evenzo is voor alle $m > N_\varepsilon$

$$|S_m - S_{N_\varepsilon}| \le \sum_{n=N_\varepsilon+1}^{\infty} |a_n| \le \varepsilon.$$

Dan staat het er eigenlijk al wat we willen.

Opgave 3.17. Gebruik (3.7) en (3.8) en met $\varepsilon = \frac{1}{k}$, $k \in \mathbb{N}$, om te laten zien dat ook $T_N^\phi \to T$, en dat S_N en S_N^ϕ convergeren, naar één en dezelfde $S \in \mathbb{R}$, eerst langs (verschillende) deelrijen N_k.

Hiermee zijn (3.9) en (3.10) bewezen, en we vatten het resultaat samen in een nog iets algemenere stelling die toereikend is voor wat we later nog nodig hebben, en die ook geldt voor aftelbare sommen van complexe getallen.

Stelling 3.18. *Laat a_n genummerd door $n \in \mathbb{N}_0$ een rij in $\mathbb{R}$ zijn. Als de partiële sommen*

$$T_N = \sum_{n=0}^{N} |a_n|$$

een door $N \in \mathbb{N}_0$ genummerde begrensde rij vormen, dan zijn voor elke bijectie $\phi : \mathbb{N}_0 \to \mathbb{N}_0$ de rijen

$$S_N^\phi = \sum_{n=0}^{N} a_{\phi(n)} \quad en \quad T_N^\phi = \sum_{n=0}^{N} |a_{\phi(n)}|$$

convergent, met limieten respectievelijk S en T die niet van ϕ afhangen, en er geldt dat $|S| \leq T$. We schrijven

$$S = \sum_{n=0}^{\infty} a_n = \sum_{n=0}^{\infty} a_{\phi(n)} \quad en \quad |\sum_{n=0}^{\infty} a_n| \leq \sum_{n=0}^{\infty} |a_n| = T.$$

De uitspraak over de rijen S_N^ϕ en T_N^ϕ geldt ook voor (iedere) bijectie $\phi : \mathbb{N}_0 \to A$ als A een (dan per definitie) aftelbare verzameling is, en voor de getallen $a_\alpha \in \mathbb{R}$ met $\alpha \in A$ geldt dat

$$\sum_{n=0}^{N} |a_{\phi(n)}|$$

een begrensde rij is. In dat geval schrijven we

$$S = \sum_{\alpha \in A} a_\alpha, \quad T = \sum_{\alpha \in A} |a_\alpha|, \quad met\ wederom \quad |S| \leq T.$$

4 Van integraal- naar differentiaalrekening

Wat voor eigenschappen heeft de functie $F : [a, b] \to \mathbb{R}$ gedefinieerd door

$$F(x) = \int_a^x f(s)\, ds \tag{4.1}$$

als $f \in \mathrm{RI}([a, b])$? Om deze vraag te beantwoorden schrijven we, voor een vaste $x_0 \in [a, b]$ en een variable h, de integraal met bovengrens $x_0 + h$ als

$$F(x_0 + h) = \int_a^{x_0+h} f(s)\, ds = \underbrace{\int_a^{x_0} f(s)\, ds}_{F(x_0)} + \int_{x_0}^{x_0+h} f(s)\, ds.$$

Als M een bovengrens is voor de waarden die $|f(x)|$ aanneemt voor $x \in [a, b]$ dan geldt voor de tweede term dat

$$\left| \int_{x_0}^{x_0+h} f(s)\, ds \right| \leq M|h|,$$

en dus is F uniform Lipschitz continu op $[a, b]$, want met $x = x_0 + h$ en $y = x_0$ volgt

$$|F(x) - F(y)| \leq M|x - y|$$

voor alle $x, y \in [a, b]$.

Dit scherpen we aan door

$$F(x_0 + h) = F(x_0) + \int_{x_0}^{x_0+h} (f(x_0) + f(s) - f(x_0))\, ds$$

$$= F(x_0) + \int_{x_0}^{x_0+h} f(x_0)\, ds + \int_{x_0}^{x_0+h} (f(s) - f(x_0))\, ds$$

te schrijven. Met

$$R(h; x_0) = \int_{x_0}^{x_0+h} (f(s) - f(x_0))\, ds, \tag{4.2}$$

volgt nu dat

$$F(x_0 + h) = F(x_0) + f(x_0)h + R(h; x_0), \tag{4.3}$$

waarin we de eerste twee termen zien als *lineaire*[1] *benadering* van $F(x_0 + h)$ als h klein is.

[1] Beter: affien. De afbeelding $h \to f(x_0)h$ is lineair, $h \to F(x_0) + f(x_0)h$ niet.

Hoe groot kan de *restterm* (4.2) zijn? Als f continu is in x_0, dan is er gegeven iedere $\varepsilon > 0$ een $\delta > 0$ zodanig dat

$$|f(s) - f(x_0)| \leq \varepsilon \quad \text{als} \quad |s - x_0| \leq \delta.$$

Voor de s in de integrand is dit laatste zeker het geval als $|h| \leq \delta$ en met de 3-hoeksongelijkheid voor integralen volgt dan dat

$$|R(h; x_0)| \leq \varepsilon |h| \quad \text{als} \quad |h| \leq \delta \tag{4.4}$$

en $x_0 + h \in [a, b]$. Deze uitspraak wordt genoteerd als

$$R(h; x_0) = o(h) \quad \text{als} \quad h \to 0,$$

spreek uit: $R(h; x_0)$ is *kleine* o van h voor h gaat naar nul[2].

Definitie 4.1. *Een functie $F : [a, b] \to \mathbb{R}$ heet differentieerbaar in $x_0 \in [a, b]$ als er een getal $f(x_0) \in \mathbb{R}$ bestaat zodanig dat (4.3) geldt voor alle $x = x_0 + h \in [a, b]$, met een restterm $R(h; x_0)$ waarvoor geldt dat $R(h; x_0) = o(h)$ als $h \to 0$.*

In deze definitie kan $[a, b]$ vervangen worden door een willekeurig interval $I \subset \mathbb{R}$. Als $x_0 = a$ dan wel $x_0 = b$ spreken we van rechts- dan wel linksdifferentieerbaar. Alleen $h \geq 0$ dan wel $h \leq 0$ zijn dan toegestaan. De waarde $f(x_0)$ heet de *afgeleide* van F in x_0, notatie $F'(x_0) = f(x_0)$.

In Hoofdstuk 5 komen we uitgebreid te spreken over de basale rekenregels voor en voorbeelden van differentieerbare functies. De enige rekenregel die we hier zometeen al nodig hebben is:

Opgave 4.2. Als $F : [a, b] \to \mathbb{R}$ en $G : [a, b] \to \mathbb{R}$ differentieerbaar zijn in $x_0 \in [a, b]$, dan is ook de verschilfunctie $F - G$ differentieerbaar in $x_0 \in [a, b]$ met $(F - G)'(x_0) = F'(x_0) - G'(x_0)$. Bewijs dit.

De andere rekenregels voor differentieerbare functies zullen we in dit hoofdstuk al tegenkomen als bijprodukten in de behandeling van een eerste vorm van de kettingregel in Sectie 4.6 na de impliciete functies in Sectie 4.4 die hier eerder aan de orde komen dan doorgaans gebruikelijk.

De afgeleide kan voor zo'n vaste x_0 gezien worden als een lineaire afbeelding[3] van $\mathbb{R}$ naar $\mathbb{R}$, gegeven door

$$h \to F'(x_0)h. \tag{4.5}$$

[2] Niet te verwarren met de grote O van h die nog komt (en je misschien al gezien hebt).
[3] Voor $F : X \to \mathbb{R}$ wordt dit $F'(x_0) : X \to \mathbb{R}$ in Sectie 10.2.

Als F in elke $x_0 \in [a,b]$ *differentieerbaar* is, met afgeleide $F'(x_0)$, dan is daarmee een functie $f : [a,b] \to \mathbb{R}$ gedefinieerd die de afgeleide functie van F wordt genoemd. De functie F heet dan een *primitieve functie* voor f (ook wel: anti-afgeleide van). Daarvan hebben we met (4.1) hierboven een belangrijk voorbeeld al gezien, en dat is meteen een eerste *hoofdstelling* van de integraalrekening.

Stelling 4.3. *Definieer voor $f \in \mathrm{RI}([a,b])$ de functie $F \in C([a,b])$ door*

$$F(x) = \int_a^x f(s)\,ds.$$

Dan is F differentieerbaar in elke $x_0 \in [a,b]$ waar f continu is, met afgeleide $F'(x_0) = f(x_0)$.

Stelling 4.3 zegt dus dat iedere continue functie $f : [a,b] \to \mathbb{R}$ een primitieve functie heeft. Met de functie F uit de stelling geldt[4] dat

$$\int_a^b f(x)\,dx = F(b) - F(a), \tag{4.6}$$

omdat $F(a) = 0$. Tellen we bij deze F een constante op dan blijft (4.6) natuurlijk waar omdat we in het rechterlid een verschil van F-waarden hebben staan. Maar geldt (4.6) nu ook voor *alle* primitieve functies van f? Het antwoord is ja, maar daar is een nieuwe stelling voor nodig die nog veel meer toepassingen heeft. Een stelling over *differentiequotiënten*, differentiequotiënten waarvan we de limiet[5] pas nemen in het volgende hoofdstuk, zie Opgave 5.2. Deze subsectie besluiten we met een opgave waarin we de boel omdraaien:

Opgave 4.4. Gebruik Opgave 2.9 en de resultaten hierboven om te laten zien dat de afgeleide van $x \to x^n$ gelijk is aan $x \to nx^{n-1}$. Hint: niet de definitie van differentieerbaarheid zelf gebruiken nu. Beperk je argumenten tot het geval $x \geq 0$.

4.1 De middelwaardestelling

Zijn er functies die differentieerbaar zijn op een interval met afgeleide overal nul, maar zelf niet constant? De stelling die dit uitsluit is de middelwaardestelling:

[4] In je achterhoofd: vergelijk dit met Opgave 2.9.

[5] De limiet van een differentiequotiënt krijgt pas daar de naam differentiaalquotiënt.

Stelling 4.5. *Laat $F \in C([a,b])$ differentieerbaar zijn op het open interval (a,b). Dan is er een ξ in (a,b) zodanig dat*

$$\frac{F(b)-F(a)}{b-a} = F'(\xi)$$

Onthoud de uitspraak in deze stelling bijvoorbeeld als dat een differentiequotiënt (alleen van $\mathbb{R}$-waardige functies!) altijd een afgeleide is in een tussenpunt, en teken een plaatje van de grafiek met (niet zo maar) twee lijnen die allebei dezelfde richtingscoëfficiënt hebben. *Teken plaatje!*

Deze middelwaardestelling is een gevolg van de uitspraak in Opgave 2.35 over maxima en minima van functies $F \in C([a,b])$, en een uitspraak over de afgeleide in het geval dat zo'n extreme waarde wordt aangenomen in een $x_0 \in (a,b)$ waar F differentieerbaar is. De volgende opgaven laten dat zien.

Stelling 4.6. *Laat $F : (a,b) \to \mathbb{R}$ een functie zijn. Als er een punt $x_0 \in (a,b)$ is waarin F differentieerbaar is, en ook geldt dat $F(x) \leq F(x_0)$ voor alle $x \in (a,b)$, dan is $F'(x_0) = 0$.*

Definitie 4.7. *Een $x_0 \in (a,b)$ waarin $F : (a,b) \to \mathbb{R}$ differentieerbaar is met $F'(x_0) = 0$ heet een stationair punt van F en F heet dan stationair in x_0.*

Opgave 4.8. Bewijs Stelling 4.6 rechtstreeks vanuit Definitie 4.1.

Opgave 4.9. Bewijs Stelling 4.5 in het geval dat $F(a) = F(b) = 0$ (dit geval staat bekend als de Stelling van Rolle[6]). Hint: gebruik Stelling 4.6 en Opgave 2.35. Onderscheid de gevallen $m < 0 \leq M$, $m \leq 0 < M$ en $m = 0 = M$.

Opgave 4.10. Herleid het algemene geval in Stelling 4.5 tot het resultaat in Opgave 4.9 door van F een lineaire[7] functie G met $G(x) = \alpha x + \beta$ af te trekken, met $\alpha, \beta \in \mathbb{R}$ zo gekozen dat $F(a) = G(a)$ en $F(b) = G(b)$.

Opgave 4.11. Als $(a,b) \subset \mathbb{R}$ een interval is en $F : (a,b) \to \mathbb{R}$ een differentieerbare functie met $F'(x) = 0$ voor alle $x \in (a,b)$ dan is er een $C \in \mathbb{R}$ zo dat $F(x) = C$ voor

[6] Read about Rolle in wikipidia.

[7] Beter gezegd: een functie waarvan de grafiek een rechte lijn is.

alle $x \in (a,b)$. Bewijs deze niet-triviale uitspraak! Wat betekent deze uitspraak in het licht van Opgave 4.2 voor twee differentieerbare functies F en G van (a,b) naar $\mathbb{R}$ waarvoor geldt dat $F'(x) = G'(x)$ voor alle $x \in (a,b)$?

4.2 Integralen via primitieve functies

Met $I = [a,b]$ verwijst Definitie 4.1 naar het voorbeeld gedefinieerd door (4.1), waarin F differentieerbaar in elk punt $x_0 \in [a,b]$ waar f continu is. In het geval dat een functie $F : [a,b] \to \mathbb{R}$ de eigenschap heeft dat $F' \in C([a,b])$ concluderen we nu dat

$$F(x) = F(a) + \int_a^x F'(s)\,ds. \tag{4.7}$$

Links en rechts in (4.7) staan immers twee functies die als afgeleide $F' = f$ hebben: de integraal in het rechterlid is met $F'(x) = f(x)$ precies de formule (4.1) waarmee we begonnen zijn en tot Stelling 4.3 zijn gekomen, waarna ook (4.6) volgde.

De functie

$$x \to F(x) - \int_a^x F'(s)\,ds$$

heeft dus als afgeleide de nulfunctie en is daarmee vanwege Stelling 4.5 toegepast op iedere $[\tilde{a},\tilde{b}] \subset [a,b]$ constant op $[a,b]$. In Opgave 4.11 is dit argument uitgespeld. Door $x = 0$ in te vullen zien we dat de constante $F(a)$ is. Zo komen we tot een *tweede hoofdstelling* van de integraalrekening:

Stelling 4.12. *Voor $f \in C([a,b])$ geldt dat*

$$\int_a^b f(x)\,dx = F(b) - F(a),$$

als F een primitieve functie is van f. Als G een andere primitieve is dan de primitieve F gedefinieerd door (4.1), dan is $F - G$ constant op $[a,b]$.

Bewijs. Wie goed leest ziet dat het bewijs hierboven al gegeven is: de functie F gedefinieerd door (4.1) is op grond van Stelling 4.3 een primitieve waarvoor vanwege (4.6) de uitspraak in Stelling 4.12 geldt; als G een andere primitieve[8] is dan geldt vanwege Stelling 4.5 (nu zonder Opgave 4.11 maar met Opgave 4.2) toegepast op de verschilfunctie dat $F(b) - G(b) = F(a) - G(a)$ en dus

[8] Bij afspraak boven Stelling 4.3 een functie die differenteerbaar is op heel $[a,b]$.

$F(b) - F(a) = G(b) - G(a)$. Opgave 4.11 geeft tenslotte weer de uitpraak dat $F - G$ constant is.

Kijk nog eens naar Opgave 2.9. De formule in Stelling 4.12 wordt ook wel geschreven als

$$\int_{[a,b]} dF = F(x)\Big|_a^b \quad \text{met} \quad dF = F'(x)dx = f(x)dx \tag{4.8}$$

en

$$F(x)\Big|_a^b = \big[F(x)\big]_a^b = F(b) - F(a),$$

een formele notatie met de d van F die later in de vectorcalculus nog (algemenere) betekenis zal krijgen. De uitdrukking $f(x)dx$ heet een 1-vorm, $F = F(x)$ heet een 0-vorm, en een 1-vorm kan de d van een 0-vorm zijn. Een en ander is hier nog van latere zorg: te weten als we bij integraal- en differentiaalrekening met twee of meer variabelen onvermijdelijk uitdrukkingen met dx en dy voor onze kiezen krijgen, waarbij de d van $F(x, y)$ iets met dx en dy moet zijn, dus $dF = f(x, y)dx + g(x, y)dy$. Met expressies $f(x, y)dx$ en $g(x, y)dy$ die niet alleen bij elkaar opgeteld worden maar ook met elkaar vermenigvuldigd.

Zoals gezegd, van later zorg, maar merk nu alvast op dat in Stelling 4.12 de uitdrukking links kan worden gezien als

$$\int_a^b \quad \text{werkend op} \quad f(x)dx,$$

en de uitdrukking rechts als

$$\Big|_a^b \quad \text{werkend op} \quad F(x),$$

een wisselwerking tussen "integralen" en differentiaalvormen, waarin verwisselen van a en b een minteken geeft. Zie in dit verband ook Definitie 2.42 en Opgave 2.48 daaronder.

Opgave 4.13. Als I een interval is en $f : I \to \mathbb{R}$ de eigenschap heeft dat $f \in \mathrm{RI}([a, b])$ voor elk gesloten interval $[a, b] \subset I$, dan varieert $\int_a^b f$ continu met a en b. Leg nog een keer uit wat dat betekent en waarom.

Opgave 4.14. Stelling 4.12 kan ook geformuleerd worden voor $F : [a, b] \to \mathbb{R}$ continu differentieerbaar, i.e. $F : [a, b] \to \mathbb{R}$ is differentieerbaar en $x \to F'(x)$ definieert een continue functie op $[a, b]$. Herschrijf de uitspraak

$$F(b) - F(a) = \int_a^b F'(x)\,dx$$

via de substitutie

$$x = (1-t)a + tb = a + t(b-a)$$

als

$$F(b)-F(a) = \int_0^1 F'((1-t)a+tb)(b-a)\,dt = \int_0^1 F'((1-t)a+tb)\,dt\,(b-a), \quad (4.9)$$

en bewijs deze uitspraak rechtstreeks vanuit de definities, dus zonder het regeltje

$$dx = (b-a)dt$$

te gebruiken.

De formule (4.9) zal nog terugkomen[9], ook met b en a variabel, niet alleen met a en b in $\mathbb{R}$, en niet alleen met $\mathbb{R}$-waardige functies. Zo'n *differentie* $F(b) - F(a)$ kan altijd als een integraal van $t = 0$ tot $t = 1$ geschreven worden, als de integrand maar continu is. Maar alleen voor $\mathbb{R}$-waardige functies kunnen we vervolgens zeggen dat de integraal in ligt tussen het minimum en het maximum van de integrand, en dus gelijk moet zijn aan een waarde van $F'(x)$ op het interval $[a, b]$: een $F'(\xi)$ met $\xi \in [a, b]$. Een net iets zwakkere uitspraak dan in Stelling 4.5, onder een veel sterkere aanname dan in Stelling 4.5 , nog steeds exclusief voor $\mathbb{R}$-waardige functies.

Merk op dat als $x \to F'(x)$ zelf Lipschitz continu is op $[a, b]$, met Lipschitz constante L, de eerste integraal in (4.9) met $b = x$ herschrijft als

$$\int_0^1 F'(a)(x-a)\,dt + \int_0^1 (F'((1-t)a+tx) - F'(a))(x-a)\,dt,$$

waarmee

$$F(x) = F(a) + F'(a)(x-a) + R(x;a) \quad (4.10)$$

met

$$R(x;a) = \int_0^1 (F'((1-t)a+tx) - F'(a))(x-a)\,dt,$$

en dus

$$|R(x;a)| \le \int_0^1 Lt|x-a|^2\,dt = \frac{L}{2}|x-a|^2. \quad (4.11)$$

In (4.10) zien we een lineaire benadering met een restterm, die in (4.11) expliciet wordt afgeschat op een constante keer $|x-a|^2$. We schrijven dit als,

$$R(x;a) = O(|x-a|^2),$$

[9] In principe is vanaf hier de sprong naar Sectie 9.6 mogelijk.

spreek uit *grote O van* $x-a$ in het kwadraat. Zo'n schatting is voldoende om convergentie van de methode van Newton te bewijzen onder milde voorwaarden, zie Sectie 4.8. Merk op dat $O(|x-a|^2) = o(|x-a|)$ als $x \to a$ maar i.h.a. niet[10] $o(|x-a|) = O(|x-a|^2)$.

4.3 Inverse functies

Als I een open interval is met $F : I \to \mathbb{R}$ differentieerbaar en $F'(x) > 0$ voor alle $x \in I$, dan volgt uit de middelwaardestelling (Stelling 4.5) dat F strict monotoon stijgend is op I. Voor iedere $[a,b] \subset I$ met $a < b$ is het beeld onder F van $[a,b]$ dus bevat in $(F(a), F(b))$, het beeld van (a,b) onder F in $(F(a), F(b))$, en kan elke waarde $c \in (F(a), F(b))$ maar één keer[11] worden aangenomen.

Wordt elke waarde $c \in (F(a), F(b))$ ook daadwerkelijk aangenomen? Het antwoord is ja[12] en daarvoor is alleen maar nodig dat $F \in C([a,b])$. Zie Hoofdstuk 10 van [HM] of [BE] over die fundamenten van de wiskunde. Of maak de volgende opgave.

Opgave 4.15. Laat $F \in C([a,b])$ en neem aan dat $F(a) < c < F(b)$. Beschouw de kleinste bovengrens ξ van de niet-lege verzameling x-waarden in $[a,b]$ waarvoor $F(x) < c$, dus

$$\xi = \sup\{x \in [a,b] : F(x) < c\}.$$

Bewijs dat $F(\xi) = c$.

Stelling 4.16. *Laat $F \in C([a,b])$ en neem aan dat F differentieerbaar is met $F'(x_0) > 0$ in elke $x_0 \in (a,b)$. Dan bestaat er een functie $G \in C([F(a), F(b)])$ met $G(F(x)) = x$ voor alle $x \in [a,b]$, $F(G(y)) = y$ voor alle $y \in [F(a), F(b)]$ en G is differentieerbaar in elke $y_0 \in [F(a), F(b)]$ met afgeleide gedefinieerd door $F'(x_0)G'(y_0) = 1$ als $y_0 = F(x_0)$ (en $x_0 = G(y_0)$).*

Opgave 4.17. Het bestaan van G volgt onmiddellijk[13] uit de overwegingen boven de stelling en de opgave. We hoeven alleen nog te laten zien dat G differentieerbaar is. Neem daartoe zonder beperking der algemeenheid het geval dat $x_0 = 0$, $F(0) = 0 = y_0$ (dus $G(0) = 0$), en $F'(x_0) = 1$. Maak[14] het bewijs af door $F'(0) = 1$ te relateren aan het liggen van de grafiek $y = F(x)$ tussen $y = (1+\varepsilon)x$ en $x = (1+\varepsilon)y$ in de *Teken plaatje!*

[10] Het is gelijkteken wordt hier op een niet-symmetrische manier gebruikt!

[11] Men zegt dan wel eens dat F injectief is.

[12] Men zegt dan wel eens dat F surjectief is.

[13] F en G zijn allebei zowel injectief als surjectief en heten daarom *bijecties*.

[14] Na discussie met Thomas Rot is Sectie 6.7 toegevoegd, zou te lezen moeten zijn.

buurt van $(x, y) = (0, 0)$ voor gegeven willekeurige $\epsilon > 0$. Zorg dat je herformulering symmetrisch is in x en y en concludeer dat $G'(0)$ bestaat en gelijk is aan 1.

Opgave 4.18. Reduceer het algemene geval via geschikte hulpfuncties tot het speciale geval $F(0) = 0$ en $F'(0) = 1$ in Opgave 4.17.

Opgave 4.19. De functies $x \to x^n$ op $\mathbb{R}^+ = \{x \in \mathbb{R} : x > 0\}$ met $n \in \mathbb{N}$ hebben inverse functies. Gebruik Opgave 4.4 en Stelling 4.17 om de afgeleiden van de inverse functies $y \to \sqrt[n]{y}$ te bepalen.

Op dit moment hebben we nog niet zoveel andere voorbeelden. Maar zie later en verderop Sectie 5.6, en leg Stelling 4.16 daar dan nog eens naast. Wel is het aardig om hier op te merken dat lineaire benaderingen[15] van wortels al 300 jaar voor Christus voorkwamen, zie [Eves, §2.17].

4.4 Impliciete functies

Als voor een functie van twee reële variabelen, zeg

$$\mathbb{R}^2 = \{(x, y) : x, y \in \mathbb{R}\} \xrightarrow{F} \mathbb{R},$$

geldt dat $F(0, 0) = 0$, dan heeft de vergelijking

$$F(x, y) = 0$$

in het algemeen nog meer oplossingen in de *buurt*[16] van $(x, y) = (0, 0)$. In deze sectie gaat het over het vinden van die andere oplossingen. De aanpak die we presenteren is in wezen hetzelfde als[17] die voor $F : X \times Y \to Y$.

Een bijzonder geval van de vraagstelling is het geval waarin

$$F(x, y) = g(y) - x,$$

en we dus eigenlijk vragen naar de inverse functie f van een gegeven functie g, maar anders dan in de vorige sectie gaat het nu om de inverse in de buurt van een gegeven punt (het punt $(0, 0)$ als $g(0) = 0$) op de grafiek van g, en *niet* om een globale inverse.

[15] Equivalent met $\sqrt{1 + x} = 1 + \frac{1}{2}x + \cdots$.
[16] Lees: in een omgeving van, zie de appendix, Section 12.
[17] Voor Banachruimten X en Y, bijvoorbeeld $X = \mathbb{R}^n$ en $Y = \mathbb{R}^m$.

Voor we een (locale *impliciete functie*) stelling formuleren schetsen we eerst de *methode van Newton* voor het vinden van die oplossingen[18], waarbij we aannemen dat voor vaste x in de buurt van $x = 0$ de functie

$$y \to F(x, y)$$

differentieerbaar is in de buurt van $y = 0$. De afgeleide noteren we nu niet met $F'(x, y)$ maar met $F_y(x, y)$. Het subscript y geeft aan dat het om de afgeleide naar y gaat. Docent en student kunnen het onderstaande natuurlijk eerst in het speciale geval $F(x, y) = g(y) - x$ doen, met $F_y(x, y) = g'(y)$. Niet wezenlijk anders, maar lichter kwa notatie. Het eindresultaat is dan een locale *inverse functie* stelling[19].

Neem voor zo'n vaste x in de *startwaarde* $y_0 = 0$ de lineaire benadering van $F(x, y)$ als functie van y rond $y = y_0$. Stel nu niet $F(x, y)$ maar die lineaire benadering gelijk aan 0, los daaruit $y = y_1$ op, en gebruik de lineaire benadering van $F(x, y)$ rond $y = y_1$ om dan weer y_2 te vinden, en ga zo door. Als y_{n-1} al gevonden is en $F_y(x, y_{n-1})$ weer inverteerbaar[20] als afbeelding

$$h \to F_y(x, y_{n-1})h$$

van $\mathbb{R}$ naar $\mathbb{R}$, dan geeft dit een volgende y_n uniek gedefinieerd door

$$F(x, y_{n-1}) + F_y(x, y_{n-1})(y_n - y_{n-1}) = 0.$$

Voor $n = 1, 2, \ldots$ geldt dan dat

$$y_n = y_{n-1} - (F_y(x, y_{n-1}))^{-1} F(x, y_{n-1}), \quad \text{beginnend met} \quad y_0 = 0. \tag{4.12}$$

In het algemeen[21] breekt dit proces, dat de methode van Newton wordt genoemd, niet af.

Als y_n vervolgens convergeert, en de x-afhankelijke limiet y een zogenaamde impliciete functie

$$x \to y = f(x) \tag{4.13}$$

definieert waarvoor (hopelijk) geldt dat

$$F(x, f(x)) = 0, \tag{4.14}$$

dan kan daarna ook nog de vraag gesteld worden onder welke voorwaarden f continu of differentieerbaar is in $x = 0$.

[18] Convergentie van die methode bespreken we pas in Sectie 4.8.

[19] In de algemene situatie volgt de impliciete ook uit de inverse functiestelling.

[20] Hier hetzelfde als $F_y(x, y_{n-1}) \neq 0$ aannemen zodat $(F_y(x, y_{n-1}))^{-1} = \frac{1}{F_y(x, y_{n-1})}$.

[21] Probeer het maar met een voorbeeldje.

Een rechtstreeks maar lastig bewijs van (snelle) convergentie van de rij y_n waarmee $y = f(x)$ gemaakt moet worden is mogelijk via een schatting van de vorm

$$|y_{n+1} - y_n| \leq C|y_n - y_{n-1}|^2,$$

afgeleid uit o.a. een aanname op de tweede[22] afgeleide van $y \to F(x, y)$. Zo'n aanname doen we hier echter niet want voor het *aangepaste schema*

$$y_n = y_{n-1} - (F_y(0,0))^{-1} F(x, y_{n-1}) \tag{4.15}$$

kan (langzamere) convergentie eenvoudiger bewezen worden zonder zo'n aanname.

Behalve een voldoende kleine bovengrens op $|F(x,0|$ en de inverteerbaarheid van $F_y(0,0)$, gebruikt het convergentiebewijs dat nu volgt alleen de continuïteit[23] van

$$(x, y) \to F_y(x, y)$$

in $(x, y) = (0, 0)$. Dit zal voldoende blijken om in de buurt van $(0, 0)$ alle oplossingen van $F(x, y) = 0$ te vinden, als x-afhankelijke limieten van de rijen y_n gedefinieerd door (4.15), beginnende met

$$y_0 = 0 \quad \text{en} \quad y_1 = -(F_y(0,0))^{-1} F(x, 0).$$

Merk op dat als $y_1 = y_0 = 0$ meteen ook volgt dat $0 = y_0 = y_1 = y_2 = \cdots$ en dat in dat geval $y = y_0 = 0$ zelf al de oplossing is van $F(x, y) = 0$. We spellen nu eerst uit hoe convergentie van de rij y_n bewezen kan worden, en nemen daarbij aan dat $y_1 \neq y_0$, want $y_1 = y_0$ is alleen maar het geval als $F(x, y_0) = 0$. Hoe groot kan y_1 nu zijn als $F(x, y_0) = F(x, 0) \neq 0$? Wel, als $|(F_y(0,0))^{-1}| \leq M$ dan geldt voor y_1 dat

$$|y_1| = |(F_y(0,0))^{-1}|\,|F(x,0)| \leq M\,|F(x,0)|$$

en als $F(x, y_1)$ gedefinieerd is volgt ook dat

$$|y_2 - y_1| = |(F_y(0,0))^{-1}|\,|F(x,y_1)| \leq M\,|F(x,y_1)|,$$

als we (4.15) gebruiken met $n = 2$.

Als we echter (4.15) gebruiken met $n = 1$ *en* $n = 2$ dan volgt

$$y_2 - y_1 = y_1 - (F_y(0,0))^{-1} F(x, y_1) - y_0 + (F_y(0,0))^{-1} F(x, y_0)$$

$$= (F_y(0,0))^{-1} \left(F(x, y_0) - F(x, y_1) + F_y(0,0)y_1 - F_y(0,0)y_0\right),$$

[22] In feite: Lipschitz continuïteit van $y \to F_y(x, y)$, zie (4.10) en (4.11).

[23] Nog niet gedefinieerd wat dat is, maar dat kun je wellicht al bedenken nu.

waarin we $(F_y(0,0))^{-1}$ aan de voorkant buiten haakjes hebben gehaald en in de grote factor die overblijft in de eerste twee termen het verschil $F(x,y_0) - F(x,y_1)$ herkennen. Dat verschil kunnen we schrijven als

$$F(x,y_0) - F(x,y_1) = \int_0^1 F_y(x, ty_0 + (1-t)y_1)\, dt\, (y_0 - y_1),$$

de integraal die we krijgen door (4.9), de *middelwaardestelling in integraalvorm*[24], voor vaste x toe te passen op de functie $y \to F(x,y)$ met $a = y_1$ en $b = y_0$. Samen met de derde en vierde term herschrijft de hele factor nu als

$$\int_0^1 \left(F_y(x, ty_0 + (1-t)y_1) - F_y(0,0)\right)\, dt\, (y_0 - y_1),$$

door de andere twee termen binnen de integraal te brengen. We concluderen dat

$$y_2 - y_1 = (F_y(0,0))^{-1} \int_0^1 \left(F_y(x, ty_0 + (1-t)y_1) - F_y(0,0)\right)\, dt\, (y_0 - y_1).$$

Tot hier is alleen de continuïteit voor vaste x van $y \to F_y(x,y)$ gebruikt, op een nog te kiezen (x,y)-domein dat nu kan worden gekozen door continuïteit van $(x,y) \to F_y(x,y)$ in $(0,0)$ te eisen[25], en daarmee het bestaan van $F_y(x,y)$ voor (x,y) in de buurt van $(0,0)$. Om precies te zijn, we nemen dan aan dat er voor iedere $\eta > 0$ een $\varepsilon > 0$ is waarmee voor alle x en y in $\mathbb{R}$ de implicatie

$$|x| \le \varepsilon \text{ en } |y| \le \varepsilon \implies |F_y(0,0) - F_y(x,y)| \le \eta \tag{4.16}$$

geldt. Daarmee volgt dat voor x, y_0, y_1 met $|x| \le \varepsilon$, $|y_0| \le \varepsilon$ en $|y_1| \le \varepsilon$ de integraal in absolute waarde hoogstens gelijk kan zijn η en dus dat

$$|y_2 - y_1| \le M\eta\, |y_1 - y_0| \quad \text{met} \quad M = |(F_y(0,0))^{-1}| > 0,$$

i.e. de grootte van het tweede stapje is begrensd door

$$\theta = M\eta$$

keer de grootte van het eerste stapje.

Voor deze conclusie is in het bijzonder ook aangenomen dat $y \to F(x,y)$ voor alle x met $|x| \le \varepsilon$ differentieerbaar in elke y met $|y| \le \varepsilon$. Merk ook op

[24] Om later het bewijs te kunnen overschrijven voor $F : X \times Y \to Y$.

[25] In het geval dat $F(x,y) = g(y) - x$ is dit geen extra eis natuurlijk.

dat we hier i.p.v. een ε, δ-definitie een η, ε-definitie van continuïteit gebruiken. Dat is omdat uiteindelijk de limiet $y = f(x)$ van de rij y_n straks ook nog moet voldoen aan $|y| \leq \varepsilon$, voor x met $|x| \leq \delta$, $\delta > 0$ bij $\varepsilon > 0$ nog te kiezen[26], om f continu in $x = 0$ te krijgen.

Als vervolgens ook geldt dat $|y_2| \leq \varepsilon$, dan geeft hetzelfde argument dat

$$|y_3 - y_2| \leq \theta\,|y_2 - y_1|,$$

enzovoorts. Zolang $|y_n| \leq \varepsilon$ volgt dat[27]

$$|y_{n+1}| = |y_{n+1} - y_0| \leq |y_{n+1} - y_n| + \cdots + |y_1 - y_0| \leq$$

$$(\theta^n + \cdots + 1)|y_1 - y_0| < \frac{|y_1 - y_0|}{1-\theta} \leq \frac{M|F(x,0)|}{1-\theta} \leq \varepsilon$$

als $\theta < 1$, en $F(x,0)$ voldoet aan

$$|F(x,0)| \leq \frac{\varepsilon}{M}(1-\theta) = \frac{\varepsilon}{M}(1 - M\eta) = \varepsilon(\frac{1}{M} - \eta) = \tilde{\varepsilon}. \tag{4.18}$$

De laatste ongelijkheid wordt geforceerd door bij deze positieve $\tilde{\varepsilon}$ de bijbehorende $\tilde{\delta} > 0$ van de continuïteit van $x \to F(x,0)$ in $x = 0$ te nemen en x te beperken tot $|x| \leq \tilde{\delta}$. We nemen dus aan dat $x \to F(x,0)$ in $x = 0$ continu is[28].

Kies nu $\eta = \eta_0 < \frac{1}{M}$ vast en laat zoals in (4.16) hierboven ε_0 de bijbehorende ε zijn. Deze ene η_0 en ε_0 zijn genoeg om nu de convergentie te bewijzen. Neem daartoe $\tilde{\varepsilon}_0$ bij η_0 en ε_0 zoals in (4.18), en vervolgens $\tilde{\delta}_0$ bij $\tilde{\varepsilon}_0$ gegeven middels de continuïteit van $x \to F(x,0)$ in $x = 0$. Neem tenslotte voor δ_0 het minimum van ε_0 en $\tilde{\delta}_0$.

De (aannamen voor de) schattingen hierboven zijn nu hieronder alleen nog nodig voor x met $|x| \leq \delta_0$ en y met $|y| \leq \varepsilon_0$. Voor iedere x met $|x| \leq \delta_0$ volgt nu dat de x-afhankelijke som

$$\sum_{n=0}^{\infty}(y_{n+1} - y_n)$$

absoluut convergent is, en dus[29] ook gewoon convergent. Maar dat betekent dat de x-afhankelijke rij y_n, gedefinieerd door $y_0 = 0$ en (4.15) voor $n \in \mathbb{N}$,

[26] Dit *rijgen* van de kleine Griekse letters hebben we eerder gezien in Opgave 2.29.

[27] Gebruik de meetkundige reeks

$$1 + \theta + \theta^2 + \cdots = \frac{1}{1-\theta}. \tag{4.17}$$

[28] Geen extra eis in het speciale geval $F(x,y) = x - g(y)$ natuurlijk.

[29] De eerste keer dat we dit gebruiken!

convergeert naar een limiet $y = f(x)$ met $|y| \leq \varepsilon_0$. Deze y nu moet de unieke oplossing van $F(x, y) = 0$ in

$$\bar{B}_{\varepsilon_0} = \{y \in \mathbb{R} : \ |y| \leq \varepsilon_0\}$$

zijn.

In feite hebben we hierboven laten zien dat de x-afhankelijke afbeelding Φ gedefinieerd door

$$y \xrightarrow{\Phi} y - (F_y(0,0))^{-1} F(x, y), \tag{4.19}$$

voor x in een geschikt gekozen x-interval, op een geschikt gekozen maar niet van x afhankelijk y-interval precies één y op zichzelf afbeeldt. Zo'n y wordt een *vast punt* van Φ genoemd en, omdat $(F_y(0,0))^{-1}$ als inverse van $F_y(0,0)$ zelf inverteerbaar is, zijn vaste punten van $y \to \Phi(x, y)$ precies de oplossingen van $F(x, y) = 0$.

De facto blijkt dat Φ voldoet aan een schatting van de vorm

$$|\Phi(x, y) - \Phi(x, \tilde{y})| \leq \theta |y - \tilde{y}| \tag{4.20}$$

met $\theta < 1$. De afbeelding $y \to \Phi(x, y)$ heet dan strict contractief, i.e. uniform Lipschitz continu met Lipschitz constante $\theta < 1$. Dat kan hier met één en dezelfde constante voor alle x in de buurt van 0.

Merk nog een keer goed op dat $F(x, y) = 0$ via (4.19) equivalent is met $y = \Phi(x, y)$ omdat $(F(0,0))^{-1}$ als inverse van $F(0,0)$ inverteerbaar is. Daar $y \to \Phi(x, y)$ op $\bar{B}_{\varepsilon_0}$ voor alle x met $|x| \leq \delta_0$ uniform Lipschitz continu is met Lipschitz constante $M\eta_0 < 1$, is er voor iedere zulke x hoogstens één zo'n $y \in \bar{B}_{\varepsilon_0}$ met $y = \Phi(x, y)$. Voor de limiet $y = f(x)$ geldt vanwege de continuïteit[30] van $y \to \Phi(x, y)$ inderdaad dat

$$y = \lim_{n \to \infty} y_{n+1} = \lim_{n \to \infty} \Phi(x, y_n) = \Phi(x, y).$$

De formule direct boven (4.18) geeft via de nu geslaagde limietovergang $y_{n+1} \to y = f(x)$ dat $f : \bar{B}_{\delta_0} \to \bar{B}_{\varepsilon_0}$ welgedefinieerd is en voldoet aan

$$|f(x)| \leq \frac{M|F(x,0)|}{1-\theta} = \frac{|F(x,0)|}{\frac{1}{M} - \eta_0}. \tag{4.21}$$

Met de grafiek $y = f(x)$ van f op $\bar{B}_{\delta_0}$ hebben we dus precies alle oplossingen van $F(x, y) = 0$ in $\bar{B}_{\delta_0} \times \bar{B}_{\varepsilon_0}$ te pakken: *Teken plaatje!*

$$\forall (x, y) \in \bar{B}_{\delta_0} \times \bar{B}_{\varepsilon_0} : \quad F(x, y) = 0 \iff y = f(x). \tag{4.22}$$

[30] Die volgt uit de differentieerbaarheid.

Hiermee is het bewijs van het bestaan van de impliciete functie f wel klaar. Het enige dat we gebruikt hebben om in (4.22) de nulverzameling van F in een geschikt gekozen omgeving $\bar{B}_{\delta_0} \times \bar{B}_{\varepsilon_0}$ van $(x, y) = 0$ als een grafiek van de vorm $y = f(x)$ te karakteriseren met $f : \bar{B}_{\delta_0} \to \bar{B}_{\varepsilon_0}$ continu in $x = 0$, is dat

$F(0, 0) = 0$;
$x \to F(x, 0)$ is continu in $x = 0$;
$(x, y) \to F_y(x, y)$ is continu in $(x, y) = (0, 0)$;
$y \to F_y(x, y)$ is een continue functie voor x vast[31];
$F_y(0, 0)$ is inverteerbaar.

Omgekeerd is gegeven zo'n f de functie gedefinieerd door $(x, y) \to f(x) - y$ een i.h.a. andere (eenvoudigere) functie die precies aan de aannamen op de algemene F voldoet om f bij F te maken zoals hierboven.

4.5 Differentieerbare impliciete functies

Met het eindresultaat in Sectie 4.4 zijn we nog niet tevreden[32], want het enige dat we van $x \to f(x)$ nu weten is de continuiteit in $x = 0$, via de continuiteit van $x \to F(x, 0)$ en (4.21).

Om ook *differentieerbaarheid* van $x \to f(x)$ in $x = 0$ te krijgen schrijven we, wederom met (4.9),

$$\begin{aligned} 0 = F(x, f(x)) &= F(x, 0) + F(x, f(x)) - F(x, 0) \\ &= F(x, 0) + \int_0^1 F_y(x, tf(x)) f(x)\, dt \\ &= F(x, 0) + F_y(0, 0) f(x) + R(x), \end{aligned}$$

waarin

$$R(x) = \int_0^1 (F_y(x, tf(x)) - F_y(0, 0)) f(x)\, dt,$$

en waarmee de aanname dat $x \to F(x, 0)$ differentieerbaar is in $x = 0$ voor de hand ligt. In dat geval weten we dat $F(x, 0) = O(|x|)$ en via (4.21) ook dat

$$f(x) = O(|x|) \quad \text{als} \quad x \to 0,$$

hetgeen alvast een sterke uitspraak is dan de continuiteit van $x \to f(x)$ in $x = 0$.

[31] Waarbij $|x| \leq \delta_0$ en $|y| \leq \varepsilon_0$, met $\delta_0 > 0$ en $\varepsilon_0 > 0$ als in de constructie van (4.22).
[32] Hoewel resultaat nu al geschikt is voor (11.23) en het Lemma van Morse, Sectie 11.1.

Neem nu aan dat $(x,y) \to G(x,y) = F_y(x,y)$ continu is in $(x,y) = (0,0)$, hierboven geformuleerd met η en ε in (4.16), maar nu weer gewoon met een ε, δ-definitie.

Definitie 4.20. *Een functie $(x,y) \to G(x,y)$ heet continu in $(0,0)$ als er voor iedere $\varepsilon > 0$ een $\delta > 0$ is zodanig dat voor alle $x \in \mathbb{R}$ en $y \in \mathbb{R}$ de implicatie*

$$|x| \leq \delta \text{ en } |y| \leq \delta \implies |G(x,y) - G(0,0)| \leq \varepsilon$$

geldt.

Voor $x \in \mathbb{R}$ met $|x| \leq \delta$ en $|f(x)| \leq \delta$ geldt dus dat

$$|R(x)| \leq \varepsilon |f(x)|,$$

en de continuiteit van $f(x)$ in $x = 0$ geeft dat er $\hat{\delta} \in (0, \delta]$ is zodanig dat deze afschatting geldt voor $|x| \leq \hat{\delta}$. We concluderen nu dat aan

$$R(x) = o(|x|) \quad \text{als} \quad x \to 0$$

is voldaan, zie (4.4). Maar nu geldt

$$0 = F(x, f(x)) = F_x(0,0)x + r(x) + F_y(0,0)f(x) + R(x),$$

met ook $r(x) = o(|x|)$ voor $x \to 0$ omdat $x \to F(x,0)$ differentieerbaar is in $x = 0$. Links en rechts de inverse van $F_y(0,0)$ toepassen en differentieerbaarheid van f in $x = 0$ volgt.

Opgave 4.21. Onder de sterkere aanname dat $(x,y) \to F_y(x,y)$ continu is in $(x,y) = (0,0)$, en de extra aanname dat $x \to F(x,0)$ differentieerbaar is in $x = 0$, is de impliciete functie $x \to f(x)$ differentieerbaar in $x = 0$ met

$$f'(0) = -(F_y(0,0))^{-1} F_x(0,0).$$

Bewijs dit nu verder zelf.

In het convergentiebewijs hebben we gebruikt dat *absoluut convergente reeksen* convergent zijn. Per definitie wordt voor een rij $a_n \in \mathbb{R}$ de som S van de bijbehorende reeks gedefinieerd door

$$S = \sum_{n=1}^{\infty} a_n = \lim_{N \to \infty} \sum_{n=1}^{N} a_n,$$

als deze limiet bestaat. De volgende opgave moet je een keer gemaakt hebben. Wellicht is dit de eerste keer. Je hebt hem hierna nog vaker nodig. Zie ook Sectie 3.6 en Opgave 3.15.

Opgave 4.22. Neem aan dat voor een rij $a_n \in \mathbb{R}$ geldt dat er een $M \geq 0$ is waarvoor geldt dat

$$T_N = \sum_{n=1}^{N} |a_n| \leq M$$

voor alle $N \in \mathbb{N}$. Bewijs dat er een kleinste $M \geq 0$ is waarvoor dit geldt, en dat deze kleinste M de limiet is van T_N als $N \to \infty$, genoteerd als

$$T = \sum_{n=1}^{\infty} |a_n|.$$

Bewijs ook dat

$$S_N = \sum_{n=1}^{N} a_n$$

een Cauchyrij is, en dus convergent met een unieke limiet S, die genoteerd als

$$S = \sum_{n=1}^{\infty} a_n$$

voldoet aan

$$|\sum_{n=1}^{\infty} a_n| \leq \sum_{n=1}^{\infty} |a_n|.$$

Blijft nog de vraag hoe het in de andere x-waarden zit met de impliciete functie die we nu gemaakt hebben. Natuurlijk willen we een uitspraak die zegt dat, voor x eventueel wat dichter in de buurt van $x = 0$, dat

$$f'(x) = -(F_y(x, f(x)))^{-1} F_x(x, f(x)), \tag{4.23}$$

en dat het rechterlid in (4.23) continu van x afhangt. Het enige dat daarvoor nodig is dat $F_y(x, y)$ in de buurt van $(x, y) = (0, 0)$ continu en inverteerbaar is en dat $x \to F(x, y)$ voor y in de buurt van $y = 0$ differentieerbaar is in elke x in de buurt van 0. Daarover doen we nu moeilijk, met als belangrijkste hulpmiddel (4.17).

Opgave 4.23. De inverteerbaarheid van $A = F_y(x, y)$ in de buurt van van $(x, y) = (0, 0)$ volgt uit de inverteerbaarheid van $A_0 = F_y(0, 0)$. Bewijs dit met algebra en convergente *meetkundige reeksen*. Gebruik van de inverse B_0 van A_0 alleen dat $A_0 B_0 = B_0 A_0 = 1$. Delen is streng verboden evenals de regel $|AB| = |A|\,|B|$ (voor willekeurige A en B). Je mag wel de ongelijkheid $|AB| \leq |A|\,|B|$ gebruiken. Laat zien

dat er een B is met $AB = BA = 1$, onder de aanname dat $|A - A_0|$ klein genoeg is in termen van o.a. $|B_0|$, en dat

$$A \to A^{-1}$$

continu is in $A = A_0$.

Na deze opgave[33] kunnen we voor $x = x_0$ in de buurt van $x = 0$ rond (x_0, y_0) met $y_0 = f(x_0)$ alles hierboven herhalen onder de aanname dat $F_x(x_0, y_0)$ bestaat en F_y continu is in (x_0, y_0). Tenslotte is voor het continu zijn van

$$x \to f'(x) = -(F_y(x, f(x)))^{-1} F_x(x, f(x))$$

in $x = x_0$ ook de continuïteit van $(x, y) \to F_x(x, y)$ in (x_0, y_0) nodig. Kortom, *uiteindelijk* zegt de *impliciete functiestelling* dat in de buurt van een een willekeurig punt (x_0, y_0) waar $(x, y) \to F_x(x, y)$ en $(x, y) \to F_y(x, y)$ allebei bestaan en continu zijn, niet alleen in (x_0, y_0) maar in die hele buurt, in een iets kleinere buurt de oplossingen van $F(x, y) = F(x_0, y_0)$ worden gegeven als grafiek van een continu differentieerbare functie $x \to f(x)$, mits in dat éne punt (x_0, y_0) de partiële afgeleide$F_y(x_0, y_0)$ inverteerbaar is.

Stelling 4.24. *Hierboven staat de definitieve versie van de impliciete functiestelling die je nu zelf kunt formuleren.*

Opgave 4.25. Formuleer Stelling 4.24, ook voor het geval dat $F(x, y) = g(y) - x$ met $g(0) = 0$. Dit speciale geval is de inverse functiestelling, die zegt dat de inverse functie f van g in een omgeving van $x = 0$ als continu differentieerbare functie is gedefinieerd met de afgeleide die je verwacht.

Opgave 4.26. Alleen in het geval dat y en dus $F(x, y)$ in dezelfde 1-dimensionale $\mathbb{R}$ liggen, kan in plaats van (4.9) direct de gewone middelwaardestelling (Stelling 4.5) gebruikt worden, op alle plekken waar (4.9) gebruikt is, om tot dezelfde resultaten (onder wellicht iets zwakkere aannamen) te komen. Ga dit na.

Opgave 4.27. De formule (4.23) kan gezien worden als een differentaalvergelijking voor $x \to f(x)$. Neem een beetje mooie $F(x, y)$ naar keuze, bepaal F_x en F_y, zet $f(x) = y$ in (4.23), en schrijf (4.23) op een apart blaadje als

$$\frac{dy}{dx} = G(x, y)$$

[33] Waarom doen we zo moeilijk in deze opgave? Zie Sectie 6.3.

met $G(x,y)$ zo vereenvoudigd dat je niet meer ziet met welke $F(x,y)$ je begonnen bent. Geef aan een collega met de vraag: los even op als je kan.

4.6 Totale afgeleide, ketting- en Leibnizregel

In Opgave 4.21 is het bewijs voor het bestaan van een unieke impliciete functie f met $f(0)=0$, waarvoor geldt dat

$$x \to F(x, f(x))$$

constant is in de buurt van $x=0$, afgerond met de differentieerbaarheid van f in $x=0$. Voldoende hiertoe is de aanname dat

$x \to F(x,0)$ differentieerbaar[34] is in $x=0$;
$(x,y) \to F_y(x,y)$ continu is in een omgeving van $(0,0)$;
$F_y(0,0)$ inverteerbaar is.

De continuiteit van F_y is hier gemakshalve wat sterker geformuleerd dan aan het eind van Sectie 4.4. We bekijken nu wat de eerste twee aannames voor de functie $(x,y) \to F(x,y)$ zelf betekenen.

Met de definitie van differentieerbaarheid van $x \to F(x,0)$ in $x=0$, en de middelwaardestelling[35] schrijven we

$$\begin{aligned} F(x,y) &= F(x,0)+F(x,y)-F(x,0) = F(0,0)+F_x(0,0)x+R(x)+F_y(x,\eta)y \\ &= F(0,0)+F_x(0,0)x+F_y(0,0)y+\underbrace{R(x)+(F_y(x,\eta)-F_y(0,0))y}_{R(x,y)}, \end{aligned}$$

voor een η tussen 0 en y en met $R(x)=o(x)$ als $x \to 0$.

Dit leest zich nu als

$$F(x,y) = F(0,0)+F_x(0,0)x+F_y(0,0)y+R(x,y) \tag{4.24}$$

met

$$R(x,y)=o(r) \quad \text{als} \quad r=\sqrt{x^2+y^2} \to 0, \tag{4.25}$$

precies wat we voortaan de *(totale) differentieerbaarheid* van $(x,y) \to F(x,y)$ in $(0,0)$ noemen. Merk op dat in (4.24) het lineaire stuk te schrijven is als

$$F_x(0,0)x+F_y(0,0)y = \begin{pmatrix} F_x(0,0) \\ F_y(0,0) \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix},$$

waarin de vector met de partiële afgeleiden de *gradiënt*[36] van F in $(0,0)$ wordt genoemd.

[34] Differentieerbaarheid van de "partiële functie" $x \to F(x,0)$ in $x=0$.
[35] Hier dus met Stelling 4.5, maar zie Opgave 4.28.
[36] Notatie ∇F, spreek uit nabla F.

Opgave 4.28. Laat $r > 0$ en neem aan dat voor x met $|x| \leq r$ geldt dat $y \to F(x,y)$ differentieerbaar is voor $|y| \leq r$ met $y \to F_y(x,y)$ continu, en $(x,y) \to F_y(x,y)$ continu is in $(0,0)$. Bewijs de differentieerbaarheid van $(x,y) \to F(x,y)$ in $(x,y) = (0,0)$ met (4.9), dus zonder x-afhankelijke tussenpunten η te gebruiken.

In (4.24) staat een al precies gemaakte lineaire benadering van $F(x,y)$ rond $(0,0)$ met een restterm die voldoet aan (4.25). Natuurlijk kunnen uitdrukking als $x \to F(x, f(x))$ ook naar x gedifferentieerd worden met de nog te behandelen kettingregel, als $x \to f(x)$ en $(x,y) \to F(x,y)$ (totaal) differentieerbaar zijn. Die kettingregel komt uitgebreid aan de orde in Hoofdstuk 6, en is gebaseerd op Opgave 5.10 in Sectie 5.1. Het kan geen kwaad deze regel hier al in een speciaal geval te behandelen.

Voor functies van de vorm

$$t \to F(x(t), y(t))$$

luidt de kettingregel dat

$$\begin{aligned} \frac{d}{dt} F(x(t), y(t)) &= F_x(x(t), y(t))x'(t) + F_y(x(t), y(t))y'(t) \qquad (4.26) \\ &= x'(t)F_x(x(t), y(t)) + y'(t)F_y(x(t), y(t)), \end{aligned}$$

te onthouden als F differentiëren naar elke t-afhankelijke variabele, daarna doordifferentiëren naar t, en alles optellen. Als $t \to (x(t), y(t))$ toevallig een kromme parametrizeert waarlangs F constant is dan volgt in $t = 0$ dat

$$F_x(0,0)x'(0) + F_y(0,0)y'(0) = \begin{pmatrix} F_x(0,0) \\ F_y(0,0) \end{pmatrix} \cdot \begin{pmatrix} x'(0) \\ y'(0) \end{pmatrix} = 0,$$

waarin de tweede vector de raakvector aan de kromme in $(0,0)$ genoemd wordt[37], waar de gradiënt van F in $(0,0)$ dan loodrecht op staat.

Allebei de vormen in (4.26) zijn nuttig. Je kunt ervoor kiezen de afgeleide van de eerste schakel in de ketting systematisch[38] voor of systematisch na de afgeleide(n) van de tweede schakel in de ketting te zetten. Zie eventueel alvast (9.20) en (9.21) en wat daarna volgt. Het bijzonder geval dat $x(t) = t$, $y(t) = f(t)$ hebben we hierboven gezien met de impliciete functie f waarmee $F(t, f(t))$ constant is.

In het geval dat $F(x,y) = xy$ zien we

$$(x(t)y(t))' = \frac{d}{dt}(x(t)y(t)) = x'(t)y(t) + x(t)y'(t) \qquad (4.27)$$

[37] Tenzij deze vector de nulvector is.

[38] Willen we co- of willen we contravariante differentiaalrekening?

verschijnen, de produktregel van Leibniz[39].

Bewijs van deze kettingregel. Toch maar meteen een bewijsje van (4.26) hier, zonder beperking der algemeenheid in $t = 0$. Dus neem aan dat $t \to x(t)$ en $t \to y(t)$ differentieerbaar zijn in $t = 0$, en dat $(x, y) \to F(x, y)$ totaal differentieerbaar is in $(x(0), y(0))$ en, wederom zonder beperking der algemeenheid, dat $0 = x(0) = y(0) = F(0, 0) = 0$. Dan volgt met de differentieerbaarheid van $x(t)$ en $y(t)$ in $t = 0$, via de expansies

$$x(t) = x'(0)t + r(t) \quad \text{en} \quad y(t) = y'(0)t + s(t), \tag{4.28}$$

voor de samengestelde functie dat

$$F(x(t), y(t)) = F_x(0, 0)x(t) + F_y(0, 0)y(t) + R(x(t), y(t)) =$$
$$F_x(0, 0)(x'(0)t + r(t)) + F_y(0, 0)(y'(0)t + s(t)) + R(x(t), y(t))$$
$$= (F_x(0, 0)x'(0) + F_y(0, 0)y'(0))t + S(t),$$

waarin

$$S(t) = F_x(0, 0)r(t) + F_y(0, 0)s(t) + R(x(t), y(t))$$

kleine o van t moet zijn als $t \to 0$. Per definitie is dat het geval voor $r(t)$ en $s(t)$, de resttermen uit de lineaire benaderingen van $x(t)$ en $s(t)$ rond $t = 0$.

Opgave 4.29. Bewijs nu zelf met (4.28) en de aannames op $r(t), s(t)$ en $R(x, y)$ dat $R(x(t), y(t)) = o(t)$ als $t \to 0$. Daarmee is het bewijs dan klaar.

Zoals al opgemerkt volgt met $F(x, y) = xy$ de Leibniz regel (4.27). Met $x(t) = y(t)$ volgt natuurlijk ook de kettingregel voor

$$t \to x(t) \to F(x(t)) = \phi(t),$$

met

$$\phi'(t_0) = F'(x(t_0))x'(t_0) \tag{4.29}$$

als $t \to x(t)$ differentieerbaar is in t_0 en $x \to F(x)$ differentieerbaar is in $x(t_0)$. Ook de quotiëntregel is nu niet moeilijk meer, misschien eerst voor

$$t \to \frac{y(t)}{1 + x(t)}$$

met

$$x \to \frac{1}{1 + x}$$

als leuke oefening vooraf.

[39] Die schrijven we liever met x en y steeds in dezelfde volgorde.

4.7 Stationair onder randvoorwaarde

Als we twee functies Φ and F van x en y hebben die allebei (totaal) differentieerbaar zijn in $(x, y) = (0, 0)$, en een[40] functie f van x die differentieerbaar is in $x = 0$, waarvoor geldt dat

$$F_x(0,0) + F_y(0,0)f'(0) = 0, \tag{4.30}$$

hetgeen zeker geldt als $F(x, f(x))$ constant is, dan is

$$x \xrightarrow{\phi} \phi(x) = \Phi(x, f(x))$$

met de kettingregel differentieerbaar in $x = 0$, met

$$\phi'(0) = \Phi_x(0,0) + \Phi_y(0,0)f'(0). \tag{4.31}$$

Als $F_y(0,0)$ inverteerbaar is volgt nu uit (4.30) en (4.31) dat

$$\phi'(0) = 0 \iff \Phi_x(0,0) = \Phi_y(0,0)(F_y(0,0))^{(-1)}F_x(0,0) \tag{4.32}$$

de karakterisatie is voor het stationair zijn van $\phi(x)$ in $x = 0$.

Inverteerbaar betekent hier voor $F_y(0,0) \in \mathbb{R}$ gewoon dat $F_y(0,0) \neq 0$, waarmee

$$\phi'(0) = 0 \iff \Phi_x(0,0)F_y(0,0) = \Phi_y(0,0)F_x(0,0).$$

Het criterium voor $\phi'(0) = 0$ is dus dat de gradiënt van Φ in $(0,0)$ een veelvoud is van de gradiënt van F.

Opgave 4.30. Bewijs de laatste uitspraak, die zegt dat $\phi'(0) = 0$ in dit geval, dus onder de aanname dat $F_y(0,0) \neq 0$, equivalent is met het bestaan van een $\lambda \in \mathbb{R}$ waarmee

$$\begin{pmatrix}\Phi_x(0,0)\\ \Phi_y(0,0)\end{pmatrix} = \lambda\begin{pmatrix}F_x(0,0)\\ F_y(0,0)\end{pmatrix}.$$

Een soortgelijk uitspraak is te doen voor functies $x = g(y)$ met $y \to F(g(y), y))$ constant en het stationair zijn van

$$y \xrightarrow{\psi} \psi(y) = \Phi(g(y), y)$$

[40] Bijvoorbeeld met de inverteerbaarheid van $F_y(0,0)$ gemaakte impliciete functie.

in $y = 0$, onder de aanname dat $F_x(0,0) \neq 0$.

We komen nog terug op het stationair zijn van functies $\Phi : X \to \mathbb{R}$ onder randvoorwaarde(n) in Sectie 6.3: Opgave 4.30 is een eerste voorbeeld van de *multiplicatorenstelling* van Lagrange, die hier in het geval dat $X = \mathbb{R}^2$ zegt dat de gradiënt van Φ loodrecht moet staan op de door de randvoorwaarde gedefinieerde verzameling. De wiskundige formulering daarvan in Euclidische ruimten kan niet zonder de impliciete functies zoals gemaakt in de vorige sectie. De stelling zelf beperkt zich niet tot ruimten waarin een inprodukt voorhanden is en kent ook formuleringen in Banachruimten.

4.8 Zijstapje: convergentie van Newton's methode

We borduren nog wat verder op de methode van Newton waar we in Sectie 4.4 mee begonnen zijn, maar doen dat zonder x-afhankelijkheid in $F(x, y)$. En dus noemen we de y maar weer x en gebruiken maar eens een kleine f in plaats van een grote F: voor f van x, is $f(x) = 0$ nu de op te lossen vergelijking.

Voor zo'n functie $f : \mathbb{R} \to \mathbb{R}$ die voor een gegeven $\varepsilon_0 > 0$ differentieerbaar is op

$$\bar{B}_\varepsilon = \{x \in \mathbb{R} : \, |x| \leq \varepsilon\},$$

met $x \to f'(x)$ uniform Lipschitz continu op $\bar{B}_{\varepsilon_0}$, en een rij x_n in $\bar{B}_{\varepsilon_0}$ van Newton- of andere iteraties die de oplossing x van $f(x) = 0$ moeten benaderen, schrijven we (4.10) over als

$$f(x_n) = \underbrace{f(x_{n-1}) + f'(x_{n-1})(x_n - x_{n-1})}_{\text{lineaire benadering}} + R(x_n; x_{n-1}), \tag{4.33}$$

waarin

$$|R(x_n; x_{n-1})| \leq \frac{L}{2}|x_n - x_{n-1}|^2,$$

met L de Lipschitz constante van f' op $\bar{B}_{\varepsilon_0}$. We nemen nu ook aan dat voor alle $x \in \bar{B}_{\varepsilon_0}$ geldt dat

$$|(f'(x))^{-1}| \leq C$$

voor een positieve constante $C > 0$.

Introduceer nu

$$p_n = |x_n - x_{n-1}| \;\text{ and }\; q_n = |f(x_n)|, \tag{4.34}$$

en neem x_n gedefinieerd door het Newton schema[41]

$$x_n = x_{n-1} - (f'(x_{n-1}))^{-1} f(x_{n-1}) \quad (n \in \mathbb{N}), \tag{4.35}$$

[41] Het is hier niet nodig om aan raaklijnen te denken maar het mag wel.

met $x_0 = 0$. Dan zit $x_n \in \bar{B}_{\varepsilon_0}$ zolang

$$p_1 + p_2 + \cdots + p_n \leq \varepsilon_0, \tag{4.36}$$

en in dat geval volgt

$$p_n \leq C q_{n-1} \quad \text{en} \quad q_n \leq \frac{1}{2} L p_n^2, \tag{4.37}$$

omdat (4.35) de lineaire benadering in (4.33) nul stelt.

De ongelijkheden in (4.37) worden nu in stelling gebracht beginnende met

$$q_0 = |f(0)| \quad \text{en} \quad p_1 \leq C q_0 = C|f(0)|, \tag{4.38}$$

door (4.37) en (4.38) te combineren als

$$p_n \leq \mu p_n^2 \quad \text{met} \quad \mu = \frac{1}{2} LC \quad \text{and} \quad p_1 \leq C|f(0)|, \tag{4.39}$$

en de vraag is voor welke $\bar{P} = \bar{P}(\mu)$ we kunnen concluderen dat de implicatie

$$C|f(0)| \leq \bar{P} \implies \sum_{n=1}^{\infty} p_n \leq \varepsilon_0 \tag{4.40}$$

geldt.

Hoe groter $\bar{P}$, hoe sterker de uitspraak in de zin dat grotere $|f(0)|$ toegestaan zijn om een oplossing $x \in B_{\varepsilon_0}$ van $f(x) = 0$ te vinden via (4.35) startend vanaf $x_0 = 0$. De grootste $\bar{P}$ vinden we door in (4.39) en (4.40) overal gelijkheden te nemen. Dat geeft

$$p_n = \mu p_{n-1}^2 \quad \text{voor} \quad n \in \mathbb{N}; \quad p_1 = \bar{P}\,; \quad \sum_{n=1}^{\infty} p_n = \varepsilon_0. \tag{4.41}$$

Via $\xi_n = \mu p_n$ en $\xi_n = \xi_{n-1}^2$ is dit equivalent met

$$\mu\varepsilon_0 = G(\mu\bar{P}) \quad \text{met} \quad G(\xi) = \xi + \xi^2 + \xi^4 + \xi^8 + \xi^{16} + \cdots, \tag{4.42}$$

maar dit geeft geen simpele formule $\bar{P} = \bar{P}(\mu, \varepsilon_0)$.

Opgave 4.31. Gebruik nu

$$G(\xi) \leq \frac{\xi}{1-\xi}$$

om een simpele conditie van de vorm $|f(0)| \leq \ldots$ in termen van C en L te formuleren, die convergentie $x_n \to x \in \bar{B}_{\varepsilon_0}$ garandeert met $f(x) = 0$.

In Hoofdstuk 3 van [HM] wordt de bij de functie $x \to f(x)$ horende Newtonafbeelding

$$x \to x - \frac{f(x)}{f'(x)} = F(x)$$

in detail onderzocht voor simpele functies van $\mathbb{R}$ naar $\mathbb{R}$, bijvoorbeeld voor $f(x) = x^2 - 2$. Die analyse, die weer[42] terugvoert tot kleitabletten, zullen we hier niet herhalen. Over discrete dynamische systemen gedefinieerd door schema's van de vorm

$$x_n = F(x_{n-1}) \tag{4.43}$$

komen we nog wel te spreken in Sectie 5.4.

4.9 Integralen met parameters

Een niet onbelangrijk onderwerp is hoe integralen van parameters afhangen. Eerst maar even zien hoe het zit met de eigenschappen van een integraal met vaste grenzen, voor het gemak 0 en 1, en een integrand met een parameter. We beschouwen[43]

$$j(t) = \int_0^1 F(x,t)\,dx$$

met, voor elke t in een t-interval $[0,1]$, de functie $x \to F(t,x)$ continu op een x-interval $[0,1]$. Daan weten we zeker dat $j(t)$ welgedefinieerd is.

We gebruiken nu maar eens middelwaardestelling in de vorm van Stelling 4.5 zelf, toegepast voor iedere vaste $x \in [0,1]$ op $t \to F(x,t)$, die voor dit argument differentieerbaar moet zijn op $[0,1]$, of op een kleiner interval dat 0 bevat maar niet van x afhangt. Dan is

$$F(t,x) = F(0,x) + F_t(\tau,x)\,t,$$

met $\tau = \tau(x) \in (0,t)$. We kunnen dus schrijven

$$F(t,x) = F(0,x) + F_t(0,x)t + R(t,x), \tag{4.44}$$

met

$$R(t,x) = (F_t(\tau(x),x) - F_t(0,x))t.$$

Als nu in (4.44) voor vaste t alles continu is in x dan geldt

$$j(t) = \int_0^1 F(t,x)\,dx = \int_0^1 (F(0,x) + F_t(0,x)t + R(t,x))\,dx$$

[42] In [Eves §2-5,2.7,6-6] zien we Heron's methode voor $\sqrt{2}$ teruggaan tot YBC7289!
[43] Dit woord gebruiken we te weinig.

$$= j(0) + t\int_0^1 F_t(x,0)\,dx + \int_0^t R(t,x)\,dx \tag{4.45}$$

$$= j(0) + t\int_0^1 F_t(x,0)\,dx + r(t),$$

waarin

$$r(t) = \int_0^t R(t,x)\,dx.$$

Tot nu toe is daarbij alleen gebruikt dat $x \to F(t,x)$ en $x \to F_t(0,x)$ continu zijn op $[0,1]$ om de integralen te laten bestaan. De integraal $r(t)$ van $R(t,x)$ in (4.45) is dan ook continu. De tweede uitdrukking met $\tau(x) \in (0,t)$ boven (4.44) kan nu gebruikt worden om $r(t) = o(t)$ voor $t \to 0$ te krijgen.

Voor deze restterm $r(t)$ willen we dat $|r(t)| \leq \varepsilon t$ voor t klein genoeg, en daarvoor is (misschien niet nodig maar zeker wel) voldoende dat voor $(t,x) \to F_t(t,x) = f(t,x)$ geldt dat

$$|f(t,x) - f(0,x)| \leq \varepsilon \tag{4.46}$$

voor alle $x \in [0,1]$ tegelijk, en dat voor alle $t \in (0,\delta)$, met $\delta > 0$ bij ε te vinden.

Opgave 4.32. Neem aan dat $(t,x) \to f(t,x)$ in elk punt $(0,x)$ met $x \in [0,1]$ continu is. Bewijs dat er voor alle $\varepsilon > 0$ een $\delta > 0$ is zodanig dat (4.46) geldt voor alle $t \in (0,\delta)$ en alle $x \in [0,1]$. Hint: denk aan het bewijs dat uniforme continuïteit op gesloten begrensde intervallen uit continuïteit volgt en redeneer uit het ongerijmde zoals in Opgave 2.31. Bij de $\epsilon > 0$ waarvoor de uitspraak niet geldt is een rij punten (t_n, x_n) te vinden met $0 < t_n \to 0$. Neem een convergente deelrij van x_n, formuleer de tegenspraak en de volgende stelling is bewezen.

Stelling 4.33. *Niet eenvoudig om te onthouden, maar laat $(t,x) \to F(t,x)$ gedefinieerd zijn voor $x \in [a,b] \subset \mathbb{R}$ met $a < b$, en $t \in (t_0 - \delta, t_0 + \delta)$ met $t_0 \in \mathbb{R}$ en $\delta > 0$. Neem aan dat voor iedere vaste $t \in (t_0 - \delta, t_0 + \delta)$ de functie $x \to F(t,x)$ continu is op $[a,b]$ en*

$$j(t) = \int_a^b F(t,x)\,dx$$

dus bestaat. Als voor iedere vaste $x \in [a,b]$ de functie $t \to F(t,x)$ differentieerbaar is op $(t_0 - \delta, t_0 + \delta)$ en $(t,x) \to F_t(t,x)$ continu is in iedere (t_0,x) met $x \in [a,b]$, dan is $t \to j(t)$ differentieerbaar in t_0 met $j'(t_0) = \int_a^b F_t(t_0,x)\,dx$.

Stelling 4.34. *Makkelijker te onthouden is: als F en F_t bestaan en continu zijn op $I \times [a,b]$, met I een t-interval, dan is $j : I \to \mathbb{R}$ continu differentieerbaar met*

$$j'(t) = \int_a^b F_t(t,x)\,dx.$$

Opgave 4.35. Laat zien dat $t \to \int_a^b f(t,x)\,dx$ continu is op een interval I als $t \to f(t,x)$ continu is op $I \times [a,b]$. Hieruit volgt de continuïteit van j' in de laatste uitspraak van Stelling 4.34. Hint: gebruik een argument zoals in Opgave 4.32.

4.10 Partieel integreren en Taylorpolynomen

We verlaten hier nu de analyse om een stukje calculus te doen, waaruit vervolgens weer een belangrijke stelling volgt die tot de analyse gerekend kan worden. De regel voor partieel integreren is via Stelling 4.12 de tegenhanger van de produktregel van Leibniz, zie (4.27), waarin van alle drie de termen de integraal over een t-interval $I = [\alpha, \beta]$ kan worden genomen, als bijvoorbeeld $t \to x'(t)$ en $t \to y'(t)$ continu zijn op I. Na herordenen volgt dat

$$\int_\alpha^\beta x(t)y'(t)\,dt = [x(t)y(t)]_\alpha^\beta - \int_\alpha^\beta x'(t)y(t)\,dt. \tag{4.47}$$

We passen deze regel nu toe in een leerzaam voorbeeldje met andere letters voor de variabelen.

Voor gegeven $f \in C([0,1])$ is het randwaardeprobleem gegeven door

$$-u''(x) = f(x) \quad \text{voor alle} \quad 0 \le x \le 1, \quad \text{en} \quad u(0) = u(1) = 0 \tag{4.48}$$

uniek oplosbaar, door f twee keer te primitiveren en de twee vrije integratieconstanten zo te kiezen dat de dubbele primitieve u in $x = 0$ en $x = 1$ gelijk is aan 0. Het is instructief om dit in detail te doen.

Een keer primitiveren geeft

$$u'(x) = u'(0) - \underbrace{\int_0^x f(s)\,ds}_{F(x)},$$

met een nog onbepaalde $u'(0)$, en de primitieve van F van f die voldoet aan $F(0) = 0$, waarna nog een keer primitiveren geeft dat

$$u(x) = u'(0)x - \int_0^x F(s)\,ds,$$

$u'(0)$ nog steeds onbepaald, $x \to \int_0^x F(s)\,ds$ de primitieve van F die 0 is in $x = 0$, en $u(1) = 0$ nog niet gebruikt.

Met de produktregel van Leibniz is $F(s)$ te schrijven als

$$\underbrace{1\,F(s)}_{G'(s)F(s)} = \underbrace{(s-a)'}_{G'(s)}F(s) = (\underbrace{(s-a)}_{G(s)}F(s))' - \underbrace{(s-a)}_{G(s)}F'(s)$$

$$= \underbrace{((s-a)F(s))'}_{(G(s)F(s))'} - \underbrace{(s-a)f(s)}_{G(s)F'(s)}$$

(afgeleiden naar s). Onder de formules is aangegeven dat we 1 zien als $G'(s)$ met $G(s) = s - a$ en a nog vrij te kiezen.

De primitieve van $F(x)$ is dus te schrijven als

$$\int_0^x F(s)\,ds = [(s-a)F(s)]_0^x - \int_0^x (s-a)f(s)\,ds = \int_0^x (x-s)f(s)\,ds. \quad (4.49)$$

Met $a = x$ volgt dat

$$u(x) = u'(0)x - \int_0^x (x-s)f(s)\,ds$$

en door $x = 1$ in te vullen vinden we dat

$$u'(0) = \int_0^1 (1-s)f(s)\,ds.$$

Daarmee is

$$u(x) = \int_0^1 (1-s)f(s)\,ds\,x - \int_0^x (x-s)f(s)\,ds$$

$$= x\int_x^1 (1-s)f(s)\,ds + (1-x)\int_0^x sf(s)\,ds = \int_0^1 A(x,s)f(s)\,ds.$$

In deze laatste formule wordt

$$A(x,s) = \begin{cases} (1-x)s & \text{als } 0 \le s \le x \\ (1-s)x & \text{als } x \le s \le 1 \end{cases} \quad (4.50)$$

de kern van de oplossingsoperator[44] wordt genoemd. Deze operator geeft u in termen van f als

$$u(x) = \int_0^1 A(x,s)f(s)\,ds. \quad (4.51)$$

[44] De kern A is te vergelijken met een (symmetrische) matrix, maar dat voor nu terzijde.

We hebben hierboven (4.51) met (4.50) afgeleid als oplossing voor (4.48) via de produktregel voor de afgeleide van $s \to F(s)G(s)$. Ook als je de regel voor partieel integreren ben vergeten, Stelling 4.12 toepassen op de produktregel van Leibniz voor primitieven functies F en G van $f = F'$ en $g = G'$ in $C([a,b])$ geeft je snel dat

$$\int_a^b F(x)G'(x)\,dx = \left[F(x)G(x)\right]_a^b - \int_a^b F'(x)G(x)\,dx, \tag{4.52}$$

de bekende regel voor *integration by parts*.

Partieel integreren is voor alles en nog wat handig. Bijvoorbeeld voor herhaald primitiveren zoals hierboven. Daarover nu een serie opgaven die leiden tot de *Stelling van Taylor* over polynomiale benaderingen van voldoende vaak continu differentieerbare functies.

Opgave 4.36. Neem $f \in C([a,b])$ en definieer

$$F_1(x) = F(x) = \int_a^x f(s)\,ds \quad \text{en} \quad F_2(x) = \int_a^x F_1(s)\,ds.$$

Laat met (4.52) zien dat

$$F_2(x) = \int_a^x (x-s)f(s)\,ds.$$

Hint: de integratievariabele is s en de 1 die je niet ziet voor $F_1(s)$ is de afgeleide naar s van $s - x$.

Opgave 4.37. Definieer in de context van Opgave 4.36

$$F_{n+1}(x) = \int_a^x F_n(s)\,ds \quad (n = 1, 2, 3\ldots),$$

en laat met herhaald partieel integreren zien dat

$$F_n(x) = \frac{1}{(n-1)!}\int_a^x (x-s)^{n-1} f(s)\,ds.$$

Hint: voor F_3 moet je twee keer partieel integreren, voor F_4 drie keer, et cetera. En dan zie je wel hoe het gaat.

Opgave 4.38. De formule in Opgave 4.37 produceert de zoveelste primitieve van f als een gewone in plaats van een meervoudige integraal. Verander het schema nu in

$$F_0(x) = f(x), \quad F_n(x) = b_n + \int_a^x F_{n-1}(s)\,ds \quad (n = 1, 2, 3\ldots), \tag{4.53}$$

en geef weer een formule voor $F_n(x)$ zoals hierboven (met meer termen natuurlijk). Bij constructie is $F_n(a) = b_n$, $F_n'(a) = b_{n-1}$, $F_n''(a) = b_{n-2}$, $\ldots$, en je ziet hier een zogenaamde Taylorbenadering van orde $n-1$ staan voor een functie waarvan de eerste $n-1$ afgeleiden in a gegeven zijn door b-tjes. Verifieer nu dat voor elke n keer continue differentieerbare functie gedefinieerd op een interval I dat 0 bevat voor $x \in I$ geldt dat

$$f(x) = f(a) + f'(a)(x-a) + f''(a)\frac{(x-a)^2}{2!} + \cdots + f^{(n-1)}(a)\frac{(x-a)^{n-1}}{(n-1)!} + \frac{1}{(n-1)!}\int_a^x (x-s)^{n-1} f^{(n)}(s)\,ds.$$

De laatste term is de restterm. Als $M = M_n(x,a)$ en $m = m_n(x,a)$ het maximum respectievelijk het minimum zijn van $f^{(n)}(s)$ als s varieert van $s = a$ naar $s = x$, dan zit deze term tussen

$$\frac{M}{n!}(x-a)^n \quad \text{en} \quad \frac{m}{n!}(x-a)^n$$

in. Omdat $f^{(n)}(s)$ tussen $s = a$ en $s = x$ alle waarden tussen $M = M_n(x,a)$ en $m = m_n(x,a)$ aanneemt, is er een zekere maar i.h.a. verder niet nader te bepalen $s = \sigma$ tussen $s = a$ en $s = x$ waarvoor de restterm gelijk is aan

$$\frac{f^{(n)}(\sigma)}{n!}(x-a)^n.$$

Kortom,

$$f(x) = \sum_{k=0}^{n-1} \frac{f^{(k)}(0)}{k!}(x-a)^k + \underbrace{\frac{f^{(n)}(\sigma)}{n!}(x-a)^n}_{\frac{1}{(n-1)!}\int_a^x (x-s)^{n-1} f^{(n)}(s)\,ds.}, \tag{4.54}$$

met σ tussen a en x.

Het resultaat in formule (4.54) is ook te bewijzen *zonder* de aanname dat $f^{(n)}$ continu is, met een σ strict tussen a en x. In het geval dat $n = 1$ reduceert de uitspraak dan tot de middelwaardestelling (Stelling 4.5), en het is weer een slimme toepassing van diezelfde in Sectie 4.1 bewezen middelwaardestelling waarmee (4.54) kan worden bewezen. In sommige analyse colleges gebeurt dat wellicht nog.

4.11 Meer calculus: substitutieregels voor integralen

In de vorige secties was partieel integreren de tegenhanger voor de Leibnizregel voor de afgeleide van het produkt van twee differentieerbare functies. Merk op dat in (4.47), geschreven als

$$\int_\alpha^\beta x(t)\,\underbrace{y'(t)\,dt}_{dy} = [x(t)y(t)]_\alpha^\beta - \int_\alpha^\beta y(t)\,\underbrace{x'(t)\,dt}_{dx},$$

een tweede regel zich aandient als we gebruik maken van de notaties besproken na Stelling 4.12.

De integraal links suggereert een omschrijving naar een integraal met y als nieuwe variabele, en de integraal rechts een integraal met x als nieuwe variabele. De betreffende regel correspondeert via Stelling 4.12 met de kettingregel zoals geformuleerd in (4.29) en is af te lezen uit

$$\int_\alpha^\beta \underbrace{F'(x(t))}_{f(x(t))}\,x'(t)\,dt = F(x(\beta)) - F(x(\alpha)) = \int_{x(\alpha)}^{x(\beta)} F'(x)\,dx = \int_a^b \underbrace{F'(x)}_{f(x)}\,dx,$$

en geldt sowieso voor continue functies $f : [a,b] \to \mathbb{R}$ en continu differentieerbare surjectieve functies $t \to x(t)$ met $x(\alpha) = a$ en $x(\beta) = b$ in de vorm

$$\int_a^b f(x)\,dx = \int_\alpha^\beta f(x(t))x'(t)\,dt. \tag{4.55}$$

De voor hand liggende naam van (4.55) is de *substitutieregel*: is de integraal in het linkerlid slecht hanteerbaar, dan kan een substitutie van de vorm $x = x(t)$ soms uitkomst brengen. Omgekeerd kan het zijn dat in een integraal over een t-interval een factor van de vorm $x'(t)$ staat en dat $t \to x(t)$ inverteerbaar is, en een slecht hanteerbare t-integraal kan worden omgeschreven naar een beter hanteerbare x-integraal.

4.12 Conclusie

We hebben in dit hoofdstuk de belangrijkste aspecten van de theorie onder de differentiaalrekening voor functies van $x \in \mathbb{R}$ naar $\mathbb{R}$ wel behandeld. Met lineaire benaderingen en zonder differentiaalquotiënten. Omdat we ook het bestaan van differentieerbare impliciete functies hebben behandeld, hebben we de facto bijna alles dat essentieel is voor de theorie onder de differentiaalrekening voor functies van x en y ook al gezien.

Totale differentieerbaarheid van

$$(x,y) \to F(x,y)$$

in een (x_0, y_0) volgde, let wel, pas na het behandelen van de impliciete functiestelling voor oplossingen $y = f(x)$ van $F(x, y) = 0$, uit aannamen zwakker dan de makkelijker en bij voorkeur te onthouden aanname dat

$$F_x(x, y) \quad \text{en} \quad F_y(x, y)$$

in de buurt van (x_0, y_0) bestaan en continu zijn in (x_0, y_0). Vervolgens hebben we voor functies $F(x, y)$, $\Phi(x, y)$ en $f(x)$ met

$$x \to F(x, f(x))$$

constant, het stationair zijn van

$$x \to \Phi(x, f(x))$$

geformuleerd in termen van de gradiënten van Φ en F.

Daarmee is het belangrijkste[45] wel gedaan. Wel zullen we nog tweede orde afgeleiden nodig hebben voor een criterium dat het positief/negatief zijn van $F''(x_0)$ in een stationair punt van $F : \mathbb{R} \to \mathbb{R}$, waar per definitie $F'(x_0) = 0$, generaliseert, essentieel voor de studie van extreme waarden. In zo'n criterium voor $F : \mathbb{R}^2 \to \mathbb{R}$ in een punt waar $F_x(x_0, y_0) = F_y(x_0, y_0) = 0$ komen dan onvermijdelijk ook gemengde tweede orde partiële afgeleiden voor[46]. Dat de volgorde van zulke afgeleiden bij functies van meer variabelen niet uitmaakt is essentieel, en komt al eerder aan de orde als we van (complex) differentieerbare functies van $z = x+iy \in \mathbb{C}$ naar $w = u+iv \in \mathbb{C}$ willen laten zien dat hun reële en imaginaire delen u en v toch wel tamelijk bijzondere functies van x en y zijn.

De reële stelling waar het dan nog om gaat is dat de totale differentieerbaarheid van

$$(x, y) \to F_x(x, y) \quad \text{en} \quad (x, y) \to F_y(x, y)$$

in een (x_0, y_0) voldoende is voor

$$F_{xy}(x_0, y_0) = F_{yx}(x_0, y_0),$$

ook geformuleerd als

$$\frac{\partial^2 F}{\partial y \partial x} = \frac{\partial^2 F}{\partial x \partial y}$$

in (x_0, y_0), met kromme d's voor partiële afgeleiden. Hiermee hebben we het dan voorlopig wel gehad voor wat betreft de streng wiskundige onderbouwing van differentiaalrekening voor functies van meer variabelen[47] en kan

[45] TO DO: werk wat hierna komt goed uit voor ϕ in (4.31).

[46] Stelling 11.6 zegt wanneer $F(x, y)$ lokaal eigenlijk een kwadratische functie IS.

[47] Integraalrekening, en samenhang met differentiaalrekening, is een ander verhaal.

de calculus beginnen dan wel doorgaan, met het uitpakken van al bewezen resultaten.

In de laatste twee secties hebben de twee belangrijkste regels uit de calculus voor de 1-dimensionale integraalrekening afgeleid door de met analyse bewezen stellingen over differentiaalrekening via de hoofdstellingen van de integraalrekening toe te passen. Daar was verder geen analyse meer voor nodig.

4.13 Wat uitwerkingen

Opgave 4.15. Hoe doen we dit? Als $F(\xi) < c$ dan is $\xi < b$ omdat $F(b) > c$. Neem bij een $\epsilon > 0$ met $\epsilon < c - F(\xi)$ de bijbehorende $\delta > 0$ uit de definitie van continuiteit. Dan is $F(x) - F(\xi) \leq |F(x) - F(\xi)| \leq \epsilon < c - F(\xi)$ voor alle $x \in [a, b]$ met $|x - \xi| \leq \delta$ en in het bijzonder is dus $\xi + \delta < b$ en $F(\xi + \delta) < c$, een tegenspraak met de aanname dat ξ een bovengrens is. Dus $F(\xi) \geq c$. Stel $F(\xi) > c$. Neem dan bij $\epsilon > 0$ met $\epsilon < F(\xi) - c$ de bijbehorende $\delta > 0$ uit de definitie van continuiteit. Dan is $F(\xi) - F(x) \leq |F(x) - F(\xi)| \leq \epsilon < F(\xi) - c$ voor alle $x \in [a, b]$ met $|x - \xi| \leq \delta$. Voor die x-waarden geldt dus dat $F(x) > c$. Omdat $F(a) < c$ volgt $\xi - \delta > a$. Maar nu is $\xi - \delta$ ook een bovengrens, in tegenspraak met ξ zijnde de kleinste bovengrens. De enige mogelijkheid is dus dat $F(\xi) = c$.

Opgave 4.17. Zonder beperking der algemeenheid nemen we nu aan dat $x_0 = 0 = F(x_0)$ en $F'(0) > 0$. De differentieerbaarheid van F in $x = 0$ betekent dat er voor elke $\varepsilon > 0$ een $\delta > 0$ is zodanig dat voor alle $x \in \mathbb{R}$ met $|x| \leq \delta$ geldt dat

$$|F(x) - F'(0)x| \leq \varepsilon |x|,$$

hetgeen voor positieve x equivalent is met

$$(F'(0) - \varepsilon)x \leq F(x) \leq (F'(0) + \varepsilon)x.$$

De grafiek $y = F(x)$ ligt voor $x \in [0, \delta]$ dus in de driehoek begrensd door de lijnen $y = (F'(0) - \varepsilon)x$, $y = (F'(0) + \varepsilon)x$ en $x = \delta$.

We kunnen de ongelijkheden voor $x \in [0, \delta]$ schrijven als

$$\frac{1}{F'(0) + \varepsilon} F(x) \leq x \leq \frac{1}{F'(0) - \varepsilon} F(x),$$

en omdat we al weten dat F en G elkaars inverse functies zijn volgt nu dat

$$\frac{1}{F'(0) + \varepsilon} y \leq G(y) \leq \frac{1}{F'(0) - \varepsilon} y,$$

voor y in een interval $[0,\eta]$, waarbij we $\eta = (F'(0)-\varepsilon)\delta$ kunnen kiezen. Met die keuze van η ligt de grafiek $x = G(y)$ voor $y \in [0,\eta]$ in de driehoek begrensd door dezelfde twee "scheve" lijnen en de lijn $y = \eta$.

De ongelijkheden voor $G(y)$ gelden voor $y \in [0,\eta]$. Toewerkend naar de definitie van differentieerbaarheid van G in $y = 0$ volgt

$$(\frac{1}{F'(0)+\varepsilon} - \frac{1}{F'(0)})\, y \leq G(y) - \frac{y}{F'(0)} \leq (\frac{1}{F'(0)-\varepsilon} - \frac{1}{F'(0)})\, y,$$

i.e.

$$-\frac{\varepsilon}{F'(0)(F'(0)+\varepsilon)}\, y \leq G(y) - \frac{y}{F'(0)} \leq \frac{\varepsilon}{F'(0)(F'(0)-\varepsilon)}\, y$$

als

$$0 \leq y \leq \eta = (F'(0)-\varepsilon)\delta.$$

Om academische gezeur over delen door 0 te vermijden beperken we onze ε nu door

$$0 < \varepsilon \leq \frac{F'(0)}{2}$$

waarmee voor η volgt dat

$$\eta \geq \frac{F'(0)\delta}{2},$$

en vervolgens voor $G(y)$ dat

$$-\frac{\varepsilon}{F'(0)^2}\, y \leq G(y) - \frac{y}{F'(0)} \leq \frac{2\varepsilon}{F'(0)^2}\, y \quad \text{als} \quad 0 \leq y \leq \frac{F'(0)\delta}{2}.$$

Verifieer nu (voor $y < 0$ is het argument precies hetzelfde) dat

$$|G(y) - \frac{y}{F'(0)}| \leq \frac{2\varepsilon}{F'(0)^2}\, |y| \quad \text{als} \quad |y| \leq \frac{F'(0)\delta}{2},$$

met nog steeds dezelfde $\delta > 0$ bij $0 < \varepsilon \leq \frac{F'(0)}{2}$. We concluderen dat G in $y = 0$ differentieerbaar is met

$$G'(0) = \frac{1}{F'(0)},$$

de positieve voorfactoren bij ε en δ doen er immers niet toe.

5 Rekenregels en voorbeelden

We hebben de afgeleide van F in x_0 gedefinieerd als coëfficient van de lineaire benadering in (4.3) waarbij de restterm in verhouding to h willekeurig klein is voor $h \to 0$. Om de afgeleide ook als limiet van differentiequotiënten te schrijven herhalen we nu de definitie van het limietbegrip voor functies.

Definitie 5.1. *Als $I \subset \mathbb{R}$ een interval is en $f : I \to \mathbb{R}$ dan zeggen we voor*[1] *$x_0 \in \bar{I}$ en $L \in \mathbb{R}$ dat*

$$L = \lim_{x \to x_0} f(x)$$

als er voor elke $\varepsilon > 0$ een $\delta > 0$ bestaat zodanig dat voor alle $x \in I$ de implicatie

$$|x - x_0| \leq \delta \implies |f(x) - L| \leq \varepsilon$$

geldt. We schrijven ook $f(x) \to L$ als $x \to x_0$.

Opgave 5.2. Als $I \subset \mathbb{R}$ een interval is en $F : I \to \mathbb{R}$ differentieerbaar is in $x_0 \in I$, dan is

$$F'(x_0) = \lim_{x \to x_0} \frac{F(x) - F(x_0)}{x - x_0}.$$

Omgekeerd, als deze limiet bestaat dan is F differentieerbaar in x_0 met afgeleide $F'(x_0)$. Bewijs dit.

In de vorm

$$F'(x) = \frac{dF}{dx} = \lim_{h \to 0} \frac{F(x+h) - F(x)}{h} \tag{5.1}$$

is de formule in Opgave 5.2 de gebruikelijke definitie voor de afgeleide functie F' die meestal voorafgaat aan de behandeling van differentiaalrekening met lineaire benaderingen. De limiet, indien deze bestaat, heet dan het *differentiaalquotiënt*, ogenschijnlijk genoteerd als breuk met teller dF en noemer dx, hetgeen ook de notatie $dF = F'(x)dx$ in Sectie 4.2 suggereert. Met $h = \Delta x$ heet de breuk

$$\frac{F(x + \Delta x) - F(x)}{\Delta x} = \frac{\Delta F}{\Delta x}$$

een differentiequotiënt. Het differentiaalquotiënt is de limiet van het differentiequotiënt als $\Delta x \to 0$. We maken in het vervolg en waar dan ook zo min mogelijk gebruik van deze limieten van differentiequotiënten.

[1] $\bar{I} \subset \mathbb{R}$ is het gesloten interval dat je krijgt door de randpunten van I erbij te stoppen.

5.1 Rekenregels voor de afgeleide

De rekenregels voor de afgeleide functies volgen eenvoudig uit de definitie met lineaire benaderingen, en die lineaire benaderingen worden vaak met $x - x_0$ i.p.v. h geschreven. We behandelen de rekenregels hieronder in het geval van $\mathbb{R}$-waardige functies van $x \in \mathbb{R}$, maar de meeste regels en bewijzen kunnen later overschreven worden voor vectorwaardige functies van meer variabelen. Lees eventueel Sectie 4.6 nog even terug en merk op dat het uiteindelijk de kettingregel is waarvoor je dit zeker wil weten in je begrip van de differentiaalrekening. Het meeste hieronder is na Sectie 4.6 wat dubbelop, maar te leerzaam om over te slaan, ook omdat het hier alleen over functies van één variabele x gaat en de regels voor $F(x)+G(x)$, $F(x)G(x)$, $F(x)/G(x)$ en $F(G(x))$, respectievelijk de som-, produkt-, quotiënt- en kettingregel.

Opgave 5.3. Schrijf voor $F : I \to \mathbb{R}$ en $x_0 \in I$ de expansie (4.3) als

$$F(x) = F(x_0) + F'(x_0)(x - x_0) + R(x).$$

Herformuleer de conditie op de restterm[2] $R(x)$ die zegt dat F differentieerbaar is in x_0 met afgeleide $F'(x_0)$. Laat zien dat die conditie ook impliceert dat F continu is in x_0.

Als eerste voorbeeld van een rekenregel die we rechtstreeks uit de lineaire benadering halen bekijken we de functie

$$x \to \frac{1}{F(x)},$$

die natuurlijk wel eerst gedefinieerd moet zijn.

Opgave 5.4. Neem in Opgave 5.3 aan dat I een open interval is en dat $F(x_0) \neq 0$ met F continu in x_0. Bewijs dat

$$x \to \frac{1}{F(x)}$$

bestaat op een interval $[a, b] \subset I$ met $a < x_0 < b$ en continu is in x_0. Hint: gebruik eerst de definitie van continuïteit met $\varepsilon = \frac{1}{2}|F(x_0)|$ om het interval $[a, b]$ te maken.

In het bijzonder zegt deze opgave dat iedere $F \in C([a, b])$ die geen nulpunten heeft een "omgekeerde" functie definieert die ook in $C([a, b])$ zit.

[2] We schrijven $R(x)$ i.p.v. $R(x; x_0)$ nu.

Opgave 5.5. Neem in Opgave 5.4 aan dat F differentieerbaar is in x_0. Bewijs dat

$$x \to \frac{1}{F(x)}$$

differentieerbaar is in x_0 met afgeleide

$$-\frac{F'(x_0)}{F(x_0)^2}$$

door

$$\frac{1}{F(x)} = \frac{1}{F(x_0)} + \frac{F(x_0) - F(x)}{F(x)F(x_0)}$$

uit werken met behulp van de expansie in Opgave 5.3.

Opgave 5.6. Dezelfde vraag als in Opgave 5.5 maar nu voor

$$x \to \frac{1}{|F(x)|}.$$

Opgave 5.7. Schrijf voor $F : I \to \mathbb{R}$ en $x_0 \in I$ de expansie (4.3) als in Opgave 5.3, en gebruik eenzelfde expansie

$$G(x) = G(x_0) + G'(x_0)(x - x_0) + S(x).$$

voor $G : I \to \mathbb{R}$. Vermenigvuldig beide expansies, werk uit als

$$F(x)G(x) = F(x_0)G(x_0) + (F'(x_0)G(x_0) + F(x_0)G'(x_0))(x - x_0) + T(x),$$

en bewijs dat met F en G ook

$$x \to F(x)G(x)$$

differentieerbaar is in x_0. Hint: de afgeleide in x_0 leest zich onmiddellijk af, laat zien dat de restterm de juiste eigenschappen heeft.

De regel voor de afgeleide van $F(x)G(x)$ heet de *regel van Leibniz* of ook wel de *produktregel*: in elke x waar F en G differentieerbaar zijn is ook het produkt van F en G differentieerbaar met

$$(F(x)G(x))' = F'(x)G(x) + F(x)G'(x), \tag{5.2}$$

zie ook (4.27).

Opgave 5.8. Dezelfde (gaap) vraag als in Opgave 5.7 voor $x \to F(x) + G(x)$.

Opgave 5.9. Schrijf je bewijzen nog een keer over met $x - x_0$ vervangen door h en zie hoe in de produktregel de twee lineaire benaderingen $F(x_0) + F'(x_0)h$ en $G(x_0) + G'(x_0)h$ met elkaar vermenigvuldigd worden tot

$$F(x_0) + G(x_0) + (F'(x_0)G(x_0) + F(x_0)G'(x_0))h + o(h),$$

waarbij de term met h^2 in $o(h)$ is verdwenen.

Iets lastiger maar ook niet echt moeilijk is de rekenregel voor functies van de vorm $x \to G(F(x))$. Deze regel staat bekend als de *kettingregel*. We beperken ons nu tot open intervallen als definitiegebieden.

Opgave 5.10. Neem $F : (a, b) \to \mathbb{R}$ met beeld in (c, d) en $G : (c, d) \to \mathbb{R}$ met F differentieerbaar in $x_0 \in (a, b)$ en G differentieerbaar in $y_0 = F(x_0) \in (c, d)$. Vul de expansie voor $y = F(x)$ in in de expansie

$$G(y) = G(y_0) + G'(y_0)(y - y_0) + S(y)$$

en bewijs dat

$$x \to G(F(x))$$

differentieerbaar is in x_0. Hint: de afgeleide in x_0 leest zich af uit

$$G(F(x)) = G(y) = G(y_0) + G'(y_0)(F(x - F(x_0)) + S(y) =$$

$$G(y_0) + G'(y_0)F'(x_0)(x - x_0) + G'(y_0)R(x) + S(y)$$

Laat zien zien dat ook hier de (deels nog uit te werken) restterm de juiste eigenschappen heeft.

Opgave 5.11. Schrijf de lineaire benaderingen ook met $h = x - x_0$ en $k = y - y_0$ als $F(x_0) + F'(x_0)h$ en $G(y_0) + G'(y_0)k$ en zie hoe in de lineaire benadering

$$G(F(x_0 + h)) = G(F(x_0)) + G'(y_0)F'(x_0)h + o(h)$$

de term met h verschijnt door $k = F'(x_0)h$ te substitueren in $G'(y_0)k$. Anders gezegd, verifieer dat

$$h \xrightarrow{F'(x_0)} F'(x_0)h = k \xrightarrow{G'(y_0)} G'(y_0)k = G'(y_0)F'(x_0)h. \tag{5.3}$$

De coëfficiënt $G'(y_0)F'(x_0)$ van $h = x - x_0$ lezen we af als de afgeleide van $G(F(x))$ in x_0.

Met $x = x_0 + h$ is h een coördinaat om het gedrag van $y = F(x)$ en dus ook van $k = F(x) - F(x_0)$ rond $(h, k) = (0, 0)$ te beschrijven. In de kettingregel hebben we met $y = F(x)$ en $z = G(y)$ voor

$$k = F(x) - F(x_0) \quad \text{en} \quad l = G(F(x)) - G(F(x_0))$$

de grafieken

$$k = F'(x_0)h + o(h) \quad \text{en} \quad l = G'(y_0)k + o(k) = G'(y_0)F'(x_0)h + o(h)$$

in het h, k-vlak, het k, l-vlak en het h, l-vlak rond respectievelijk $(h, k) = (0, 0)$, $(k, l) = (0, 0)$ en $(h, l) = (0, 0)$.

Opgave 5.12. Leg uit waarom Opgave 5.5 achteraf overbodig was. Gebruik de vorige opgaven om de eveneens overbodige quotiëntregel voor de afgeleide van

$$x \to \frac{G(x)}{F(x)}$$

in x_0 af te leiden in het geval dat $F(x_0) \neq 0$.

Opgave 5.13. Differentieerbaar in x_0 impliceert continu in x_0 en daarmee dat alle "nieuwe" functies hierboven onder de aannamen ook continu zijn in x_0. Daarvoor was continuïteit van de ingrediënten natuurlijk voldoende. Formuleer en bewijs de voor de hand liggende uitspraken over continuïteit.

5.2 Mooie voorbeelden: machtreeksen

We doen even een voorbeeldje om te laten zien hoe we de analyse zo algebraïsch kunnen houden als maar kan. Neem $I = [-r, r]$ met $r > 0$, $F(x) = x^7$, vervang x_0 door a, $x_0 + h$ door x en maak voor dit geval (4.3) expliciet middels

$$x^7 = a^7 + 7a^6(x-a) + (x^5 + 2ax^4 + 3a^2x^3 + 4a^3x^2 + 5a^4x + 6a^5)(x-a)^2, \quad (5.4)$$

waarin we de afgeleide van x^7 in $x = a$ aflezen als de coëfficient van $x - a$.

Met $x - a$ in plaats van h is de algebra hier een stuk aangenamer, vooral omdat

$$x^7 - a^7 = (x - a)(x^6 + ax^5 + a^2x^4 + a^3x^3 + a^4x^2 + a^5x + a^6),$$

en de laatste factor met $x = a$ al de $7a^6$ geeft, waarna met een staartdeling het verschil van die factor met $7a^6$ is te schrijven als $(x - a)$ maal de grote factor in het rechterlid van (5.4). Leerzame *algebra* met een mooie uitkomst waar je meteen een patroon in ziet, en een formule voor de restterm herkent die je met inductie kunt bewijzen.

Opgave 5.14. Voor $n \in \mathbb{N}$, $r > 0$ en $x, a \in [-r, r]$ geldt dat

$$x^n = a^n + na^{n-1}(x - a) + R_{a,n}(x) \quad \text{met} \quad |R_{a,n}(x)| \leq \frac{n(n-1)}{2} r^{n-2}(x - a)^2.$$

Geef de formule voor $R_{a,n}(x)$ (bewijs die met inductie[3] als je dat leuk vindt) en bewijs de schatting. Hint: Gebruik $1 + 2 + \cdots + n = \frac{1}{2}n(n + 1)$ met $n - 1$.

We zien voor $n \geq 2$ dat x^n in a afgeleide na^{n-1} heeft omdat

$$(a + h)^n = a^n + na^{n-1}h + R(h; a),$$

met de expliciete resttermschatting

$$|R(h; a)| \leq \frac{n(n-1)}{2} r^{n-2}h^2 \leq \varepsilon|h| \quad \text{voor} \quad |h| \leq \delta = \frac{2\varepsilon}{n(n-1)r^{n-2}}.$$

Merk op dat voor $n = 0$ en $n = 1$ deze schatting minder relevant (en in deze vorm onjuist) is omdat in die twee gevallen $R(h; a) = 0$ voor alle h.

Voor een polynoom

$$P(x) = \sum_{n=0}^{k} \alpha_n x^n$$

van graad $k \geq 2$ gaat het net zo. Vermenigvuldig links en rechts in Opgave 5.14 met α_n en sommeer. Goed kijkend naar $n = 0, 1, 2$ volgt

$$P(x) = \sum_{n=0}^{k} \alpha_n x^n = \alpha_0 + \sum_{n=1}^{k} \alpha_n n a^{n-1}(x - a) + \sum_{n=2}^{k} \alpha_n R_{a,n}(x),$$

waarin we

$$P'(a) = \sum_{n=1}^{k} n\alpha_n a^{n-1}$$

aflezen omdat

$$\Big|\sum_{n=2}^{k} \alpha_n R_{a,n}(x)\Big| \leq \sum_{n=2}^{k} |\alpha_n| \frac{n(n-1)}{2} r^{n-2}(x - a)^2 \leq \varepsilon|x - a|$$

[3] Ontdekken zul je de formule daarmee zeker niet.

als

$$|x-a| \leq \delta \quad \text{met} \quad \delta \sum_{n=2}^{k} |\alpha_n| \frac{n(n-1)}{2} r^{n-2} = 2\varepsilon.$$

Maar ja, dan gaat het met (macht)reeksen precies hetzelfde. In

`(copy/paste in latex, replace ^k by ^\infty)`

$$P(x) = \sum_{n=0}^{\infty} \alpha_n x^n = \alpha_0 + \sum_{n=1}^{\infty} \alpha_n n a^{n-1}(x-a) + \sum_{n=2}^{\infty} \alpha_n R_{a,n}(x), \tag{5.5}$$

lezen we

$$P'(a) = \sum_{n=1}^{\infty} n\alpha_n a^{n-1} \tag{5.6}$$

af omdat

$$|\sum_{n=2}^{\infty} \alpha_n R_{a,n}(x)| \leq \sum_{n=2}^{\infty} |\alpha_n| \frac{n(n-1)}{2} r^{n-2}(x-a)^2 \leq \varepsilon |x-a| \tag{5.7}$$

voor

$$|x-a| \leq \delta \quad \text{met} \quad \delta \sum_{n=2}^{\infty} |\alpha_n| \frac{n(n-1)}{2} r^{n-2} = 2\varepsilon. \tag{5.8}$$

De enige eis om te stellen is dat de reeksen

$$\sum_{n=0}^{\infty} \alpha_n x^n, \quad \sum_{n=1}^{\infty} \alpha_n n a^{n-1}, \quad \sum_{n=2}^{\infty} |\alpha_n| \frac{n(n-1)}{2} r^{n-2}$$

convergeren en omdat $|x| \leq r$ en $|a| \leq r$ is dat via het gewoon convergent zijn van absoluut convergente reeksen en

$$1 \leq n \leq \frac{n(n-1)}{2} \leq \frac{1}{2} n^2 \quad (n \geq 2)$$

zeker het geval als

$$\sum_{n=1}^{\infty} n^2 |\alpha_n| r^n \leq \infty. \tag{5.9}$$

Opgave 5.15. Bepaal voor de gevallen $\alpha_n = 1$, $\alpha_n = \frac{1}{n!}$ en $\alpha_n = n!$ het maximale interval waarop de machtreeks gegeven door (5.5) gedefinieerd is, en doe hetzelfde voor $P'(x)$.

In (5.9) is $r > 0$ nog variabel. Is voor een $r > 0$ aan (5.9) voldaan, dan is

$$P(x) = \sum_{n=0}^{\infty} \alpha_n x^n$$

differentieerbaar op $[-r, r]$, met (vervang a in (5.6) door x)

$$P'(x) = \sum_{n=1}^{\infty} n\alpha_n x^{n-1} = \sum_{n=0}^{\infty} (n+1)\alpha_{n+1} x^n$$

voor alle $x \in [-r, r]$. De r-waarden waarvoor (5.9) geldt vormen een interval $I_2 \subset [0, \infty)$ met $0 \in I_2$.

Dit interval hangt niet heel erg sterk van de exponent 2 in (5.9) af. Voor elke $k \in \mathbb{N}$ is

$$I_k = \{r \geq 0 : \sum_{n=1}^{\infty} n^k |\alpha_n| r^n \leq \infty\}$$

zo'n interval en ze zijn allemaal hetzelfde, maar net niet helemaal.

Opgave 5.16. Het is evident dat

$$I_1 \supset I_2 \supset I_3 \supset \cdots ,$$

want met grotere k zijn de termen in de reeks groter. Bewijs dat er een $R \in [0, \infty]$ is met de eigenschap dat voor elke $k \in \mathbb{N}$ geldt dat $I_k = [0, R)$ of $I_k = [0, R]$ en geef voorbeelden van $R = 0$, $R = 1$ en $R = \infty$.

De volgende stelling vat samen wat we nu weten.

Stelling 5.17. *Elke machtreeks*

$$P(x) = \sum_{n=0}^{\infty} \alpha_n x^n$$

met $\alpha_n \in \mathbb{R}$ voor $n \in \mathbb{N}_0 = \{0\} \cup \mathbb{N}$ heeft een convergentiestraal $R \in [0, \infty]$. Voor $x \in \mathbb{R}$ met $|x| < R$ geldt dat de termen $\alpha_n x^n$ en $n\alpha_n x^{n-1}$ van de machtreeksen voldoende hard naar 0 gaan om absolute convergentie van zowel $P(x)$ als van

$$p(x) = \sum_{n=1}^{\infty} n\alpha_n x^{n-1} = \sum_{n=0}^{\infty} (n+1)\alpha_{n+1} x^n$$

te garanderen, waarbij p de afgeleide is van P op $\{x \in \mathbb{R} : |x| < R\}$ en P DE primitieve van p met $P(0) = \alpha_0$. Voor $x \in \mathbb{R}$ met $|x| > R$ zijn de termen van beide machtreeksen onbegrensd en de machtreeksen dus allebei niet convergent.

Op het met hoofdletters gedrukte DE komen we nog even terug. Zonder verdere voorkennis zouden we hier tevreden moeten zijn met de uitspraak dat $P(x)$, met een vrije keuze voor α_0, de enige machtreeks is die $p(x)$ als afgeleide machtreeks heeft. Merk op dat Stelling 5.17 eigenlijk een uitspraak is over P en p samen. Dat voor $|x| < R$ de enige vrijheidsgraad die er is in de primitieve van p overeenkomt met de waarde van (de constante term) $\alpha_0 \in \mathbb{R}$ is hier in eerste instantie een uitspraak over machtreeksen.

Het zou natuurlijk kunnen zijn dat er nog andere primitieve functies zijn. In het eenvoudigste voorbeeld dat $p(x) = 0$ voor alle $x \in \mathbb{R}$ leidt dit tot de vraag of er functies zijn die differentieerbaar zijn op een interval met afgeleide overal nul, maar zelf niet constant. Je bent het misschien vergeten, maar het antwoord op die vraag was gelukkig al nee, zie Stelling 4.5 en verder.

Met uitdrukkingen zoals $P(x)$ kunnen we nu eenvoudige differentiaalvergelijkingen oplossen. Bijvoorbeeld de *differentiaalvergelijking*

$$F'(x) = F(x). \tag{5.10}$$

Welke functies zijn differentieerbaar met afgeleide gelijk aan zichzelf? Probeer maar een machtreeks. Term voor term uitschrijven helpt om te zien wat er gebeurt. Als

$$P(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3 + \alpha_4 x^4 + \alpha_5 x^5 + \alpha_6 x^6 + \alpha_7 x^7 + \cdots$$

met convergentiestraal R dan is voor $|x| < R$ de afgeleide gelijk aan

$$P'(x) = \alpha_1 + 2\alpha_2 x + 3\alpha_3 x^2 + 4\alpha_4 x^3 + 5\alpha_5 x^4 + 6\alpha_6 x^5 + 7\alpha_7 x^6 + 8\alpha_8 x^7 + \cdots,$$

en dus is

$$P'(x) - P(x) = (\alpha_1 - \alpha_0) + (2\alpha_2 - \alpha_1)x + (3\alpha_3 - \alpha_2)x^2 + (4\alpha_4 - \alpha_3)x^3 + \cdots,$$

hetgeen alleen maar 0 is als

$$0 = \alpha_1 - \alpha_0 = 2\alpha_2 - \alpha_1 = 3\alpha_3 - \alpha_2 = 4\alpha_4 - \alpha_3 = \cdots!$$

Het uitroepteken staat er niet voor niets bij. Een waarheid als een koe. Toch? Of niet? Ja, wel!

Opgave 5.18. Neem aan dat $r > 0$ en dat $P(x)$ convergent is voor alle $x \in \mathbb{R}$ met $|x| < r$, en dat $P(x) = 0$ voor al die x. Waarom is $\alpha_0 = 0$? Haal vervolgens x buiten haakjes en concludeer dat $\alpha_1 = 0$, waarom? Etc.

Toegepast op de verschilmachtreeks zien we de recurrente betrekking

$$\alpha_n = \frac{\alpha_{n-1}}{n}$$

verschijnen. Als bijvoorbeeld $\alpha_0 = 1$ worden de volgende coëfficienten gegeven door

$$\alpha_1 = 1,\ \alpha_2 = \frac{1}{2},\ \alpha_3 = \frac{1}{2}\frac{1}{3},\ \alpha_4 = \frac{1}{2}\frac{1}{3}\frac{1}{4},\ \cdots,$$

m.a.w.

$$\alpha_n = \frac{1}{n!} \quad \text{voor alle} \quad n \in \mathbb{N}_0.$$

Stelling 5.19. *Laat $r > 0$. De enige machtreeks die voldoet aan zowel $P'(x) = P(x)$ voor alle $x \in \mathbb{R}$ met $|x| < r$ als aan $P(0) = 1$ is*

$$P(x) = \exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24} + \frac{x^5}{120} + \frac{x^6}{720} + \cdots.$$

Opgave 5.20. Laat zien dat de reeks in Opgave 5.19 voor elke $x \in \mathbb{R}$ convergent is. De machtreeks heeft dus convergentiestraal $R = \infty$ en is zijn eigen afgeleide.

Definitie 5.21. *Voor een functie $F : [a, \infty) \to \mathbb{R}$ zeggen we dat $F(x) \to 0$ als $x \to \infty$, als er voor elke $\varepsilon > 0$ een $\xi \in [a, \infty)$ bestaat waarmee geldt:*

$$x \geq \xi \implies |F(x)| \leq \varepsilon.$$

.

Opgave 5.22. Laat zien dat voor elke $n \in \mathbb{N}$ geldt dat

$$\frac{x^n}{\exp(x)} \to 0 \quad \text{als} \quad x \to \infty,$$

de bekende standaardlimiet die zegt dat $\exp(x)$ het wint van elke macht van x.

5.3 De natuurlijke logaritme

Expliciete schattingen zijn handig als je sommen wil nemen zoals in de stap van monomen x^n naar machtreeksen als in (5.5). Bijvoorbeeld om de vraag te beantwoorden voor welke $x \in \mathbb{R}$ de Laurentreeks

$$L(x) = \sum_{n=-\infty}^{\infty} \alpha_n x^n = \cdots + \frac{\alpha_{-2}}{x^2} + \frac{\alpha_{-1}}{x} + \alpha_0 + \alpha x + \alpha_2 x^2 + \cdots$$

convergent is en term voor term gedifferentieerd kan worden, met $|x|$ niet te groot voor de positieve machten en $|x|$ niet te klein voor de negatieve machten. Verzin hiertoe eerst maar een mooie variant op (5.4) in de vorm

$$\frac{1}{x^7} = \frac{1}{a^7} - \frac{7}{a^8}(x-a) + R_a(x).$$

Misschien doe je dit liever eerst voor de functie

$$x \to \frac{1}{x},$$

die bij de omgekeerde operatie (primitiveren) meteen het buitenbeentje is, want geen afgeleide van een algebraïsche functie!

Definitie 5.23. *Voor $x > 0$ wordt $\ln x$, de natuurlijke logaritme van x, wat onnatuurlijk gedefinieerd door*

$$\ln x = \int_1^x \frac{1}{s}\, ds.$$

Opgave 5.24. Leidt rechtsreeks uit de definities af dat voor $0 < a < b$, $f \in \mathrm{RI}([a,b])$ en $c > 0$ geldt dat

$$\int_a^b (f(x)\, dx = \frac{1}{c}\int_{ca}^{cb} f(\frac{x}{c})\, dx,$$

en gebruik dit om voor $x > 1$ en $y > 1$ de integraal voor $\ln y$ te schrijven als integraal van x tot xy. Concludeer dat

$$\ln xy = \ln x + \ln y.$$

Bewijs dat deze identiteit geldt voor alle $x, y \in \mathbb{R}^+$. Hint: laat apart zien dat

$$\ln x + \ln \frac{1}{x} = 0$$

voor alle $x > 0$.

Opgave 5.25. Uit de definitie volgt dat $\ln$ een strict stijgende functie is op $\mathbb{R}^+$. Bewijs en gebruik de ongelijkheid

$$\ln n \geq \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots + \frac{1}{n} = \sum_{k=2}^{n} \frac{1}{k}$$

om te laten zien[4] dat $\ln x \to \infty$ als $x \to \infty$. Geef daarvan en daartoe ook een definitie zoals Definitie 5.21. Wat geldt er voor $x \to 0$?

[4] Kijk nog even naar (22).

Opgave 5.26. Vanwege Opgave 5.25 en Stelling 4.16 heeft $\ln$ een inverse functie $F : \mathbb{R} \to \mathbb{R}^+$. Laat zie dat uit de stelling ook volgt dat $F(0) = 1$ en $F'(y) = F(y)$ voor alle $y \in \mathbb{R}$. Leg uit waarom $\ln$ maar één inverse functie kan hebben en concludeer dat $F = \exp$.

Opgave 5.27. Laat zien dat $\exp(x + y) = \exp(x)\exp(y)$ voor alle $x, y \in \mathbb{R}$, en dat met $e = \exp(1)$ gedefinieerd door

$$\ln e = \int_1^e \frac{1}{x}\, dx = 1$$

geldt dat

$$\exp(\frac{p}{q}) = e^{\frac{p}{q}} = \sqrt[q]{e^p}$$

voor alle $p \in \mathbb{Z}$ en alle $q \in \mathbb{N}$. Bij afspraak is $e^x = \exp(x)$ ook voor alle andere $x \in \mathbb{R}$.

Waarom die afspraak? In Opgave 5.27 zijn machten van e^x van e eerst voor $x = \frac{p}{q} \in \mathbb{Q}$ uitgewerkt. Voor zulke x vallen $\exp(x)$ en e^x dus samen. Omdat $x \to \exp(x)$ een differentieerbare en dus continue functie is geldt nu dat $e^r = \exp(r) \to \exp(x)$ als $r \to x$ langs $r \in \mathbb{Q}$, hetgeen de algemene consensus dat e^x als limiet van e^r gedefinieerd MOET worden wel rechtvaardigt: bij deze is e^x dus voortaan gelijk aan $\exp(x)$ voor alle $x \in \mathbb{R}$. Over de uitspraak e tot de macht x kan in rede[5] getwist worden.

Voor machten x^α van een willekeurige $x > 0$ is het verhaal precies hetzelfde. Met de hierboven gevonden rekenregels voor exp en ln is eerst

$$x^n = (e^{\ln x})^n = e^{n \ln x}$$

voor $n \in \mathbb{N}$, vervolgens leid je af dat $n \in \mathbb{N}$ door $r = \frac{p}{q} \in \mathbb{Q}$ vervangen mag worden en eenzelfde consensus als voor de e-macht is nu gerechtvaardigd:

$$x^\alpha = e^{\alpha \ln x} \quad \text{voor} \quad x > 0 \quad \text{en} \quad \alpha \in \mathbb{R}. \tag{5.11}$$

Het is leuk om daar even reëel mee te spelen. De volgende twee opgaven zijn een eerste kennismaking met *asymptotische formules*. De notatie

$$f(x) \sim g(x) \quad \text{voor} \quad x \to a \tag{5.12}$$

[5] Over e tot de macht πi niet natuurlijk, komt geen machtsverheffen bij van pas.

betekent bij afspraak dat

$$\frac{f(x)}{g(x)} \to 1 \quad \text{als} \quad x \to a,$$

waarbij voor a bijvoorbeeld 0 of ∞ mag worden gelezen. Zulke formules worden ook gebruikt met $n \in \mathbb{N}$ i.p.v. $x \in \mathbb{R}$. Bijvoorbeeld

$$n! \sim (\frac{n}{e})^n \sqrt{2\pi n} \quad \text{als} \quad n \to \infty, \tag{5.13}$$

de formule van *Stirling* waarover we hier iets minder uitgebreid dan in [HM] nog over komen te spreken.

Als wat links van het $\sim$ teken lastig te doorgronden is en de rechterkant wel transparant is dan kunnen dit soort uitspraken bijzonder nuttig zijn. Kwestie van smaak soms, want de keuze van $g(x)$ ligt uiteraard niet vast.

Opgave 5.28. Onderzoek de functie $F : x \to x^x$ met $x \in \mathbb{R}^+$ m.b.v. de definitie in (5.11). Schets de grafiek. Bepaal een zo simpel mogelijke functie $g : \mathbb{R}^+ \to \mathbb{R}$ zodanig dat

$$F(x) - 1 \sim xg(x)$$

als $x \to 0$, i.e.

$$\frac{F(x) - 1}{xg(x)} \to 1.$$

Als we $F(0) = 1$ stellen is F dan (rechts)differentieerbaar in $x = 0$?

Opgave 5.29. Omdat x^x strict stijgend is in x als x voldoende groot is (groter dan wat?) heeft $x \to x^x$ een inverse functie $y \to f(y)$ gedefinieerd voor y voldoende groot. Laat zien dat f gedefinieerd is door de vergelijking $x \ln x = \ln y$, breng $\ln x$ naar de andere kant en gebruik de formule die je dan krijgt voor x weer in de rechterkant teneinde een zo simpel mogelijke $g(y)$ te vinden zodanig dat

$$f(y) \sim \frac{\ln y}{g(y)}$$

als $y \to \infty$. Hint: uitdrukkingen met x en y die je naar 1 wil laten gaan kun je in x uitdrukken en vervolgens via $t = \ln x$ en Opgave 5.22 behandelen.

5.4 Zijstapje: tetratie

Dit is een verhaaltje over a tot de macht a tot de macht a tot de macht a en ga zo maar door, dat voor $a > 1$ niet alleen in verband staat met een echte[6] toepassing, maar ook voor $a \to 0$ laat zien hoe a tot de macht a tot de macht a tot huppeldepup maar niet kan kiezen tussen 0 en 1.

Neem daartoe voor $a > 0$ de rij gegeven door

$$x_0 = 1, \quad x_1 = a, \quad x_2 = a^a, \quad x_3 = a^{a^a}, \quad x_4 = a^{a^{a^a}}, \ldots,$$

dus x_n, notatie ook wel $x^n = {}^n a$, is gedefinieerd door

$$x_n = a^{x_{n-1}} = F_a(x_{n-1}) \quad (n \in \mathbb{N}) \quad \text{met} \quad x_0 = 1.$$

Een zogenaamd *discreet dynamisch systeem* waarbij de afbeelding

$$x \to F_a(x) = a^x = e^{x \ln a}$$

wordt geitereerd.

Opgave 5.30. Voor $a > 1$ is er een kritieke waarde $a = a_c$ waarvoor de grafiek $y = F_a(x)$ de diagonaal $y = x$ in het x, y-vlak raakt[7]. Bepaal a_c en de limiet van de rij ${}^n a_c$ voor $n \to \infty$. Bewijs de uitspraken die je doet.

Opgave 5.31. Laat zien dat de vergelijking $F_a(x) = x$ voor $1 < a < a_c$ twee oplossingen $\underline{a} < \bar{a}$ heeft, en dat de rij ${}^n a$ voor $n \to \infty$ naar $\underline{a}$ convergeert. Laat zien dat voor $a > a_c$ de rij ${}^n a$ naar oneindig gaat als $n \to \infty$.

Opgave 5.32. Laat zien dat de vergelijking $F_a(x) = x$ voor $0 < a < 1$ één oplossing $x = \underline{a}$ heeft, en dat de rij ${}^n a$ voor $n \to \infty$ naar $\underline{a}$ convergeert als $a > a^*$ met a^* gegeven door een conditie op $F'_a(\underline{a})$. Wat is die conditie? Bepaal a^* exact.

Opgave 5.33. Laat zien dat voor $0 < a < a^*$ de rij ${}^{2n} a$ convergeert naar één van de drie oplossingen van de vergelijking $F_a(F_a(x)) = x$ die geen oplossing is van $F_a(x) = x$, en dat de rij ${}^{2n-1} a$ naar de andere oplossing convergeert. Noem die limieten $\bar{b}$ en $\underline{b}$. Deze hangen van a af. Wat gebeurt er met $\bar{b}$ en $\underline{b}$ als $a \to 0$?

[6] Vraag het Henk Knoester van Tejin Twaron of anders Federico Camia (die zag het).
[7] Lees: $F_a(x) = x, F'_a(x) = 1$.

5.5 Meer differentiaalvergelijkingen

De machtreeks voor $\exp(x)$ als oplossing van een differentiaalvergelijking smaakt naar meer.

Stelling 5.34. *Laat $r > 0$. De enige machtreeks die voldoet aan zowel $P''(x) + P(x) = 0$ voor alle $x \in \mathbb{R}$ met $|x| < r$, als aan $P(0) = 0$ en $P'(0) = 1$, is*

$$P(x) = \sin x = x - \frac{x^3}{6} + \frac{x^5}{120} - \frac{x^7}{5040} + \cdots .$$

Opgave 5.35. Bewijs Stelling 5.34. Schrijf $P(x)$ in de fatsoenlijke somnotatie zonder stippels. Wat is de convergentiestraal? Wat is de afgeleide van $\sin$? Noem deze functie maar $\cos$. Wat is de afgeleide van $\cos$?

Dat $\exp(x)$, $\sin x$, $\cos x$ zijn[8] wat ze, moeten zijn, namelijk $e = \exp(1)$ tot de macht x, en sinus en cosinus van een hoek x in radialen, is nog niet meteen duidelijk[9] . Ook willen we wel even zeker weten dat er geen andere functies dan deze machtreeksen zijn met de eigenschappen die $\exp(x), \cos x, \sin x$ als machtreeks uniek definiëren. Eenzelfde vraag is reeds beantwoord voor functies met de eigenschap dat $F'(x) = 0$ voor alle x met $|x| < r$.

De eigenschappen van cos en sin zijn natuurlijk te halen uit het feit dat de reeksen van $\cos x$ en $\sin x$ net als die voor $\exp(x)$ absoluut convergent zijn voor alle $x \in \mathbb{R}$ en dus uitvermenigvuldigd[10] kunnen worden. Dat is wat gepriegel.

Opgave 5.36. Laat zien dat voor alle $x \in \mathbb{R}$ geldt dat

$$\cos^2 x + \sin^2 x = 1,$$

door voor $\cos x$ en $\sin x$ de machtreeksen in te vullen, de kwadraten[11] uit te werken en op te tellen.

---

[8] We zouden eigenlijk $\sin(x)$ en $\cos(x)$ moeten schrijven, maar dat doet vrijwel niemand.

[9] Voor exp inmiddels wel, maar alleen dankzij ln.

[10] Bewijs (met Stelling 3.18) en gebruik daarna dat voor absoluut convergente reeksen

$$(\sum_{n=0}^{\infty} a_n)(\sum_{m=0}^{\infty} b_m) = \sum_{k=0}^{\infty}(\sum_{j=0}^{k} a_j b_{k-j}). \tag{5.14}$$

[11] Lees $\cos^2 x = (\cos x)^2$ en idem voor sin.

Opgave 5.37. Bewijs de gelijkheid in Opgave 5.36 zonder al dat rekenwerk door op te merken dat $\cos^2 x$ en $\sin^2 x$ als kwadraten van machtreeksen weer machtreeksen zijn. Hint: hun som is dus ook een machtreeks. Wat weet je via Opgave 5.34 van de afgeleide van die machtreeks? Dus?

Bovenstaande opgaven zijn nuttige oefeningen[12], maar inzicht bieden ze niet echt. De machtreeksen zijn oplossingsformules voor een differentiaalvergelijking die met $F''(x) + F(x) = 0$ een notatie heeft die de tenen van een natuurkundige terecht doen krommen. Dit is namelijk de vergelijking voor een harmonische oscillator die we bij voorkeur schrijven als

$$\ddot{x} + x = 0,$$

waarin x de afhankelijke grootheid (plaats) is en t de onafhankelijke grootheid, die overeenkomt met wat wiskundigen de hoek in radialen noemen. Dus $x = x(t)$ en $\ddot{x}$ staat voor de tweede afgeleide van $t \to x(t)$. Oplossingen zijn periodiek met periode 2π. Het verband met cirkels en met name *eenparige* cirkelbewegingen[13] via $y = -\dot{x}$ in het x, y-vlak is natuurlijk de moeite van het bespreken waard en we verwijzen in dit verband naar [HM].

Met name omdat formules als

$$\cos(a + b) = \cos a \cos b - \sin a \sin b$$

en de vorm van de grafiek van $x = \sin t$ in het t, x-vlak bij voorkeur uit de differentiaalvergelijking gehaald worden en juist niet uit de oplossingsformules. Maar voor wie meer wil rekenen met zulke formules, het is zeker niet verboden, zeer leerzaam zelfs:

Opgave 5.38. Bewijs met de theorie tot nu toe door middel van o.a. machtsreeksberekeningen dat sin een periodieke functie is met een periode die we wel 2π mogen noemen, en dat $-\sin(-x) = \sin x = \sin(\pi - x) > 0$ voor $0 < x < \pi$. Schets de grafiek van sin.

Terug naar de differentiaalvergelijking $F'(x) = F(x)$: wat nog niet duidelijk is of er nog andere oplossingen zijn dan machtreeksoplossingen. Te prefereren is wellicht een directe uitspraak dat oplossingen van zulke differentiaalvergelijkingen altijd wel machtreeksen moeten zijn. Voor nu is het echter de

[12] En aanbevolen voor wie niet van formules houdt.

[13] Zie op Valentijnsdag 2017 de Bildungsscheurkalender van Henk Sissing.

(later wellicht overbodige[14]) vraag of een functie F die aan $F'(x) = F(x)$ voldoet op een interval I, op dat interval nulpunten kan hebben, zonder[15] op dat zelfde interval de nulfunctie te zijn.

We beantwoorden die vraag nu door juist te kijken naar oplossingen die ergens positief zijn (voor negatief is het argument hetzelfde door naar $-F(x)$ te kijken).

Opgave 5.39. Laat $I \subset \mathbb{R}$ een open interval zijn en $F : I \to \mathbb{R}$ differentieerbaar met $F'(x) = F(x)$ voor alle $x \in I$ en $(a, b) \subset I$ een maximaal open interval waarop $F(x) > 0$. Dan is $(a, b) = I$. Bewijs dit door de vergelijking op te lossen via scheiding van variabelen onder de aanname dat $F(x) > 0$, gebruikmakend van ondermeer Opgave 5.10 en

$$F'(x) = F(x) \iff \frac{F'(x)}{F(x)} = 1 \iff \ln(F(x)) = x + C \iff F(x) = e^{x+c}$$

Opgave 5.40. Dezelfde vraag als in Opgave 5.39 maar nu voor $F : I \to \mathbb{R}$ die voldoet aan $F'(x) = F(x)g(x)$ met $g : I \to \mathbb{R}$ continu. Los de differentiaalvergelijking ook weer (zoals dat op school genoemd wordt) exact op m.b.v. een primitieve G van g.

Opgave 5.41. Voor $\alpha \in \mathbb{R}$ voldoet de functie $F_\alpha : (-1, \infty) \to \mathbb{R}^+$ gedefinieerd door $F_\alpha(x) = (1 + x)^\alpha$ aan een differentiaalvergelijking zoals in Opgave 5.40. Schrijf deze in de vorm $(1 + x)F'(x) = \alpha F(x)$ en bepaal een machtreeksoplossing van de vorm

$$1 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots .$$

Schrijf de coëfficienten in een vorm waarmee je in het geval $\alpha = n \in \mathbb{N}$ het binomium van Newton herkent. De convergentiestraal van deze machtreeks (voor $\alpha \notin \mathbb{N}_0$) is $R = 1$. Kun je dat aantonen? Waarom volgt nu dat $F_\alpha(x)$ (alleen) voor $|x| < 1$ gelijk is aan de machtreeks[16] die je net hebt uitgerekend?

[14] Dat zijn soms de belangrijkste vragen...

[15] Als je weet dat $F(x + y) = F(x)F(y)$ ben je natuurlijk wel meteen klaar.

[16] Om echt te onthouden is $\alpha = -1$, ook $\alpha = \pm\frac{1}{n}$ komt vaak van pas.

Opgave 5.42. Schrijf een paar termen uit van[17]

$$\sqrt[n]{1+x} = 1 + \frac{x}{n} + \cdots \quad \text{en} \quad \frac{1}{\sqrt[n]{1+x}} = 1 - \frac{x}{n} + \cdots$$

Bovenstaande opgaven bevatten ad hoc argumenten om de oplossing van een eerste orde lineaire differentiaalvergelijking gegeven een randvoorwaarde als machtreeks te identificeren. Voor vergelijkingen als $F''(x)+F(x) = 0$ gaat het rekenwerk net zo, maar zijn de ad hoc argumenten wat anders, een reden te meer om de ad hoc redenaties te vervangen door een aanpak gebaseerd op integraalvergelijkingen, die we oplossen in de vectoruimte van continue functies op gesloten begrensde intervallen, in $C([a, b])$ dus. Dat doen we in Sectie 10.4.

Een alternatieve vraag om wellicht hier al te stellen blijft hoe je van een functie gedefinieerd als oplossing van zo'n integraalvergelijking, bijvoorbeeld

$$F(x) = 1 + \int_0^x F(s)\, ds, \tag{5.15}$$

meteen ziet dat de functie, in ieder geval voor $|x|$ kleiner dan een te bepalen en zo groot mogelijke r, gegeven wordt door een ondubbelzinnig gedefinieerde machtreeks.

5.6 Meer inverse functies

In Sectie 4.3 hadden we nog niet zoveel functies tot onze beschikking, maar nu wel. Hier volgen wat opgaven om de belangrijkste functies de revue te laten passeren.

Opgave 5.43. Laat zien dat de functie

$$x \to \frac{\sin x}{\cos x} = \tan x$$

strict stijgend is op $(-\frac{\pi}{2}, \frac{\pi}{2})$ en een inverse

$$y \to \arctan y$$

heeft op $\mathbb{R}$ met afgeleide

$$\frac{1}{1+y^2}$$

[17] Zie weer [Eves, §2.5] over het gebruik van zulke reeksen voor YBC7289?

en leid hieruit af dat

$$\arctan y = y - \frac{1}{3}y^3 + \frac{1}{5}y^5 - \cdots$$

voor $|y| < 1$.

Opgave 5.44. Laat zien dat de functie

$$x \to \sin x$$

strict stijgend is op $(-\frac{\pi}{2}, \frac{\pi}{2})$ en een inverse

$$y \to \arcsin y$$

heeft op $(-1, 1)$ met afgeleide

$$\frac{1}{\sqrt{1-y^2}}$$

en leid hieruit een machtreeks af voor $\arcsin y$ voor $|y| < 1$.

Opgave 5.45. Ook

$$x \to \cos x$$

is strict montonoon op een maximaal open interval, bijvoorbeeld op $(0, \pi)$. Laat zien dat voor de inverse geldt dat $\arccos y + \arcsin y$ constant is op $(-1, 1)$. Wat is de constante?

5.7 Rare voorbeelden

Er zijn ook functies $F : [a, b] \to \mathbb{R}$ die in elke $x_0 \in [a, b]$ differentieerbaar zijn met een afgeleide $F'(x_0)$, waarvoor $F' : [a, b] \to \mathbb{R}$ wel als functie

$$x \to F'(x) \tag{5.16}$$

is gedefinieerd maar niet in $C([a, b])$ zit. Dat komt omdat de meest *vreemde functies* differentieerbaar kunnen zijn.

Opgave 5.46. Als $g : \mathbb{R} \to \mathbb{R}$ een functie is die continu is in $x = 0$ met waarde $g(0) = 0$, dan is de functie $R : \mathbb{R} \to \mathbb{R}$ gedefinieerd door $R(x) = xg(x)$ differentieerbaar in $x = 0$ met afgeleide $R'(x) = 0$. Laat dit zien rechtstreeks vanuit Definitie 4.1.

Voor de functie g in Opgave 5.46 kun je bijvoorbeeld $g(x) = 0$ voor $x \in \mathbb{Q}$ en $g(x) = x$ voor $x \notin \mathbb{Q}$ nemen. Zo krijg je een functie $f : \mathbb{R} \to \mathbb{R}$ die in geen enkele $x \neq 0$ continu is maar wel differentieerbaar is in $x = 0$.

Opgave 5.47. Definieer $f : \mathbb{R} \to \mathbb{R}$ door $f(0) = 0$ en (voor $x \neq 0$)

$$f(x) = x^2 \sin \frac{1}{x^2}.$$

Laat zien dat f in elke $x \in \mathbb{R}$ differentieerbaar is maar dat $f'(x)$ niet eens begrensd is op het interval $[0, 1]$. Hint: gebruik dat de functies $\cos$ en $\sin$ periodiek zijn maar niet constant.

Opgave 5.48. Definieer $f : \mathbb{R} \to \mathbb{R}$ door $f(0) = 0$ en (voor $x \neq 0$)

$$\exp(-\frac{1}{x^2}).$$

Schets de grafiek van f. Laat met de theorie tot nu toe zien dat f afgeleiden $f', f'', \ldots$ van iedere orde heeft en dat in $x = 0$ alle afgeleiden van f gelijk zijn aan 0.

5.8 Stirling's formule via schalen en limieten

We hebben partieel integreren in Sectie 4.10 gebruikt om Taylorbenaderingen te maken, maar partieel integreren wordt natuurlijk ook gebruikt om gewoon integralen uit te rekenen. Een mooi voorbeeld is de integraalformule voor $n!$ met $n \in \mathbb{N}$.

Opgave 5.49. Bereken

$$\int_0^\infty \exp(-x)\,dx, \quad \int_0^\infty x \exp(-x)\,dx, \quad \int_0^\infty x^2 \exp(-x)\,dx, \quad \int_0^\infty x^3 \exp(-x)\,dx,$$

en geef voor $n!$ een integraalformule[18]. NB Dit zijn zogenaamde oneigenlijke integralen, zie Definitie 2.44. Lees dus

$$\int_0^\infty = \lim_{R \to \infty} \int_0^R$$

en doe daar verder vooral niet te moeilijk over (maar leg wel uit welke al behandelde standaardlimiet je steeds gebruikt).

[18] Het is niet verboden om die formule met inductie te bewijzen hoor.

Met Opgave 5.49 komen we nu terug op (5.13). Waar die formule vandaan komt wordt aanstonds duidelijk.

Opgave 5.50. Schets de grafiek $y = x^n e^{-x}$ (voor n nog niet te groot) in het x, y-vlak. Waar ligt de top van de berg? *Teken plaatje!*

Opgave 5.51. Door de in Opgave 5.49 gevonden integraal voor $n!$ met het maximum van de integrand te schalen en horizontaal te schuiven volgt dat

$$n! = (\frac{n}{e})^n \int_{-n}^{\infty} g_n(x)dx$$

met

$$g_n(x) = (1 + \frac{x}{n})^n e^{-x}$$

Laat dit zien m.b.v. (5.11) en schets de nieuwe grafiek $y = g_n(x)$.

Opgave 5.52. De eerste factor van $g_n(x)$ is de rente op rente benadering van e^x die in [HM] in detail wordt besproken alvorens $\exp(x)$ wordt gedefinieerd als de puntsgewijze limiet van

$$e_n(x) = (1 + \frac{x}{n})^n$$

voor $n \to \infty$. Verifieer deze limiet met wat je nu aan tools tot je beschikking hebt ,en ook dat $e_n(x)$ stijgt met n voor $x > -n$.

Bijgevolg stijgt $g_n(x)$ dus naar 1 voor elke $x \in \mathbb{R}$. Omdat $g_n(x)$ stijgend is in x voor $x \in [-n, 0]$ en dalend in x voor $x \geq 0$ is wel duidelijk dat

$$\int_{-n}^{\infty} g_n(x)dx \to \infty$$

als $n \to \infty$. Kunnen we het gebied waarvan deze integraal de oppervlakte geeft horizontaal schalen (met een n-afhankelijke schaalfactor) teneinde wel een eindige limiet te krijgen?

Opgave 5.53. Schrijf daartoe

$$g_n(x) = e^{-\psi_n(x)} \quad \text{met} \quad \psi_n(x) = -\ln(g_n(x)),$$

en verifieer dat

$$\psi_n(x) = x - n\ln(1+\frac{x}{n}) = n(\frac{x}{n} - \ln(1+\frac{x}{n})) = n\psi_1(\frac{x}{n}).$$

De rij functies ψ_n ontstaat dus door schalen uit één en dezelfde functie $\Psi = \psi_1$, en die functie heeft als formulevoorschrift

$$\Psi(x) = x - \ln(1+x). \tag{5.17}$$

Hoe ziet de grafiek van Ψ eruit? Je weet al dat $\Psi(0) = \Psi'(0) = 0$ omdat $g_n(x)$ een maximum ter grootte 1 heeft in $x = 0$. Omdat

$$\Psi'(x) = 1 - \frac{1}{1+x} = \frac{x}{1+x} \sim x$$

als $x \to 0$, volgt voor Ψ zelf dat

$$\Psi(x) \sim \frac{1}{2}x^2$$

als $x \to 0$. De grafiek van Ψ wordt verder gekenmerkt door een verticale asymptoot bij $x = -1$ en bijna-lineaire groei voor $x \to \infty$.

Teken plaatje!

Wel een schetsje waard deze grafiek maar belangrijker is dat voor ψ_n nu geldt dat

$$\psi_n(x) = n\Psi(\frac{x}{n}) \sim \frac{x^2}{2n},$$

niet alleen voor $x \to 0$, maar ook voor $\frac{x}{n} \to 0$, hetgeen steeds grotere x toelaat, maar waarbij in de limiet $n \to \infty$ niets overblijft.

Een schaling waarbij wel wat overblijft lezen we onmiddellijk af. We moeten x zo schalen dat de n in het rechterlid verdwijnt, en dus substituteren we $x = s\sqrt{n}$. Met deze schaling volgt nu dat

$$n! = (\frac{n}{e})^n\sqrt{n}\int_{-\sqrt{n}}^{\infty} e^{-n\Psi(\frac{s}{\sqrt{n}})}ds \tag{5.18}$$

In de integrand zien we een nieuwe anders geschaalde versie van Ψ, die we zo hebben gekozen dat voor elke n geldt dat

$$n\Psi(\frac{s}{\sqrt{n}}) \sim \frac{1}{2}s^2, \tag{5.19}$$

nu niet alleen voor $s \to 0$, maar ook voor $\frac{s}{\sqrt{n}} \to 0$. Het fraaie is dat met grotere n deze geschaalde $\Psi(s)$ steeds meer op zijn kwadratische benadering

bij $s = 0$ gaat lijken. We zoomen zó in op de oorsprong dat in de limiet alleen de parabool overblijft, juist omdat we (de grafiek van) Ψ verticaal en horizontaal anders schalen. De schaling is zo gekozen dat $\frac{1}{2}s^2$ van schalen niets merkt en in de limiet $n \to \infty$ wordt het linkerlid in (5.19) gelijk aan het rechterlid. De linkergrens van de integraal schuift daarbij op naar $-\infty$.

Wat we nu nog moeten doen om te kunnen concluderen dat

$$n! \sim (\frac{n}{e})^n \sqrt{n} \int_{-\infty}^{\infty} e^{-\frac{1}{2}s^2} ds \tag{5.20}$$

als $n \to \infty$ is bewijzen dat

$$\int_{-\sqrt{n}}^{\infty} e^{-n\Psi(\frac{s}{\sqrt{n}})} ds \to \int_{-\infty}^{\infty} e^{-\frac{1}{2}s^2} ds \tag{5.21}$$

als $n \to \infty$. Puntsgewijs geldt weliswaar dat

$$f_n(s) = e^{-n\Psi(\frac{s}{\sqrt{n}})} \to e^{-\frac{1}{2}s^2}$$

als $n \to \infty$, maar dat op zich is niet genoeg om (5.21) te kunnen concluderen.

Opgave 5.54. Bereken de afgeleide van $f_n(s)$ naar n voor vaste s en laat zien dat deze afgeleide een teken heeft dat alleen van het teken van s afhangt. Hint: gebruik de formule voor Ψ pas op het allerlaatst.

We zien nu dat, net als hun integranden, de integralen

$$\int_{-\sqrt{n}}^{0} e^{-n\Psi(\frac{s}{\sqrt{n}})} ds \quad \text{en} \quad \int_{0}^{\infty} e^{-n\Psi(\frac{s}{\sqrt{n}})} ds$$

stijgend respectievelijk dalend zijn in n. Zijn deze rijen convergent met limiet respectievelijk

$$\int_{-\infty}^{0} e^{-\frac{1}{2}s^2} ds \quad \text{en} \quad \int_{0}^{\infty} e^{-\frac{1}{2}s^2} ds?$$

Een positief antwoord vertelt ons dan dat (5.20) waar is, en alleen de limietintegraal zelf moet dan nog worden uitgerekend. Wat we nodig hebben zijn een paar niet te moeilijke stellingen over het *verwisselen van integralen en limieten*.

5.9 Makkelijke convergentiestellingen

Opgave 5.55. We bekijken nu eerst het geval van eindige intervallen $[a, b]$ en integralen van rijen functies $f_n : [a, b] \to \mathbb{R}$ met

$$\forall x \in [a, b] \quad 1 \leq n_1 \leq n_2 \implies f_{n_1}(x) \leq f_{n_2}(x),$$

en

$$\forall n \in \mathbb{N} \quad a \leq x_1 \leq x_2 \leq b \implies f_n(x_1) \leq f_n(x_2).$$

Neem aan dat zo'n rij f_n begrensd is. Bewijs rechtstreeks vanuit de definitie met onder- en bovensommen van de integraal voor monotone functies dat

$$\int_a^b f(x)dx = \lim_{n\to\infty} \int_a^b f_n(x)dx,$$

waarbij $f : [a, b] \to \mathbb{R}$ is gedefinieerd door

$$f(x) = \lim_{n\to\infty} f_n(x).$$

Er zijn nog drie versies van deze stelling. Welke drie?

Merk op dat we niets aannamen over het teken van $f_n(x)$. In de volgende opgave doen we dat wel.

Opgave 5.56. We bekijken nu het geval van het interval $[0, \infty)$ en integralen van rijen functies $f_n : [0, \infty) \to [0, \infty)$ met

$$\forall x \in [0, \infty) \quad 1 \leq n_1 \leq n_2 \implies f_{n_1}(x) \geq f_{n_2}(x),$$

en

$$\forall n \in \mathbb{N} \quad 0 \leq x_1 \leq x_2 \implies f_n(x_1) \geq f_n(x_2).$$

Neem aan dat

$$\int_0^\infty f_1(x)dx = \lim_{b\to\infty} \int_0^b f_1(x)dx$$

bestaat. Bewijs rechtstreeks vanuit de definities dat dan

$$\int_0^\infty f(x)dx = \lim_{n\to\infty} \int_0^\infty f_n(x)dx,$$

waarbij

$$f(x) = \lim_{n\to\infty} f_n(x).$$

Van deze stelling zijn er geen drie versies omdat positieve monotoon stijgende functies niet integreerbaar zijn op $[0, \infty)$ en ook omdat het geval f_n stijgend echt anders is.

Opgave 5.57. Formuleer en bewijs een bruikbare stelling als in Opgave 5.56 voor rijen functies met

$$\forall x \in [0, \infty) \quad 1 \leq n_1 \leq n_2 \implies f_{n_1}(x) \leq f_{n_2}(x),$$

en

$$\forall n \in \mathbb{N} \quad 0 \leq x_1 \leq x_2 \implies f_n(x_1) \geq f_n(x_2).$$

Opgave 5.58. Bewijs (5.20) met behulp van de laatste twee opgaven.

6 Analyse met meer variabelen

In dit hoofdstuk doen we vooral differentiaal- en wat integraalrekening voor functies van meer variabelen. We beginnen in $\mathbb{R}^2$ met rechthoekige coördinaten $x, y \in \mathbb{R}$ en maken veelvuldig gebruik van de transformatie

$$x = r\cos\theta;$$

$$y = r\sin\theta,$$

waarmee punten $(x, y) \neq (0, 0)$ in het vlak via hun afstand $r = \sqrt{x^2 + y^2}$ tot de oorsprong $(0, 0)$ en hoek θ tussen de halfrechte

$$\{(tx, ty) : t \geq 0\}$$

en de positieve x-as beschreven worden. Ook identificeren we indien gewenst $\mathbb{R}^2$ met de verzameling $\mathbb{C}$ van de *complexe getallen*

$$z = x + iy$$

en schrijven $|z| = r$ voor wat de *absolute waarde* van z genoemd wordt, de afstand van z tot de oorsprong $z = 0$. De hoek $\theta = \arg z$ heet het *argument* en is voor iedere $z \neq 0$ modulo 2π uniek bepaald.

Naast de gewone optelling

$$w+z = (u+iv)+(x+iy) = u+x+i(v+y) = (u+x, v+y) = (u, v)+(x, y)$$

is er de wat complexere vermenigvuldiging

$$wz = (u+iv)(x+iy) = ux-vy+i(uy+vx) = (ux-vy, uy+vx) = (u, v)(x, y)$$

gebaseerd op de rekenregel $i^2 = -1$, voor $w = u+iv = (u, v)$ en $z = x+iy = (x, y) \in \mathbb{R}^2 = \mathbb{C}$. Hiermee wordt $\mathbb{C}$ een lichaam[1] dat algebraisch gesloten is. Erg fijn is het regeltje dat bij vermenigvuldigen van twee complexe getallen de absolute waarden vermenigvuldigd en de argumenten opgeteld moeten worden om absolute waarde en argument van het produkt te krijgen.

Opgave 6.1. Schrijf $z_j = r_j(\cos\theta_j + i\sin\theta_j)$ voor $j = 1, 2$ en gebruik de somregels voor cos en sin om te laten zien dat

$$z_1 z_2 = r_1 r_2(\cos(\theta_1 + \theta_2) + i\sin(\theta_1 + \theta_2)).$$

Dit regeltje is een van de vele redenen om

$$\cos\theta + i\sin\theta = \exp(i\theta)$$

te schrijven en $\exp(z) = \exp(x)\exp(iy)$.

[1] Lees: heeft dezelfde rekenregels als $\mathbb{R}$. Ook geldt $|w+z| \leq |w|+|z|$ en $|wz| = |w|\,|z|$.

De poolcoördinaten zijn overigens niet nodig[2] om tot het inzicht te komen dat voor vaste $\gamma \in \mathbb{C}$ de afbeelding

$$z \to \gamma z \tag{6.1}$$

een draaivermenigvuldiging is, zie (6.11) en Opgave 6.4.

6.1 Zijstapje: de hoofdstelling van de algebra

De uitspraak dat $\mathbb{C}$ algebraisch gesloten is betekent dat elk polynoom

$$P(z) = \sum_{k=0}^{n-1} \alpha_k z^k + z^n \tag{6.2}$$

met $\alpha_0, \ldots, \alpha_{n-1} \in \mathbb{C}$ en $n \geq 2$ een nulpunt $z_1 \in \mathbb{C}$ heeft, en vervolgens geldt met dank aan de staartdeling dat

$$P(z) = \sum_{k=0}^{n-1} \alpha_k z^k + z^n = (z - z_1)Q(z)$$

waarin

$$Q(z) = \sum_{k=0}^{n-2} \beta_k z^k + z^{n-1},$$

met $\beta_0, \ldots, \beta_{n-2} \in \mathbb{C}$. Herhaald toepassen geeft

$$P(z) = \Pi_{k=1}^{n}(z - z_k) \quad \text{met} \quad z_1, \ldots, z_n \in \mathbb{C}. \tag{6.3}$$

Deze uitspraak wordt helaas de *hoofdstelling van de algebra* genoemd. Een sympathieke wiskundige geboren in Reeuwijk heeft meer dan 50 verschillende bewijzen van deze stelling verzameld. Geen daarvan is echt algebraisch en het bewijs dat nu volgt is er daar één van. Ik vond het toen ik probeerde het standaardbewijs met complexe functies te vermijden. Dat complexe bewijs zie je doorgaans bij een college over complexe functietheorie als toepassing van de Stelling van Liouville die zegt dat er geen begrensde complex differentieerbare functies $f : \mathbb{C} \to \mathbb{C}$ zijn behalve de constante functies. Zie ook

www-groups.dcs.st-and.ac.uk/history/HistTopics/Fund_theorem_of_algebra.html

[2] Waar zijn die hoeken eigenlijk wel voor nodig?

Een vergelijkbare stelling zegt dat er geen positieve harmonische[3] functies $f : \mathbb{R}^N \to \mathbb{R}$ zijn. Dat bewijs probeerde ik hier in stelling te brengen en toen ging het argument een eigen leven leiden. Uiteindelijk maakt het alleen gebruik van Stelling 12.11 in de appendix, die voor het geval N = 1 overeenkomt met Opgave 2.35, de opgave die de exercities na Definitie 2.25 afsluit.

Didaktisch wel aardig is dat je het bewijs eerst kunt doen voor een tweedegraads polynoom[4] zonder wortelformules[5] te gebruiken, en dan ziet dat het voor elke hogere graad hetzelfde gaat. Hieronder doe ik het echter meteen voor het algemene geval en dat is min of meer hoe Argand het deed.

Als $z \to P(z) = p(x,y) + iq(x,y)$ geen nulpunten heeft in $\mathbb{C}$ dan is

$$(x,y) = x + iy = z \to |P(z)| > 0$$

vanwege enig voorwerk als in Opgave 5.13 een continue strict positieve $\mathbb{R}$-waardige functie op $\mathbb{R}^2$ die op cirkels van de vorm $x^2+y^2 = r^2$ naar beneden begrensd is middels[6]

$$|P(z)| = |\sum_{k=0}^{n-1} \alpha_k z^k \, + z^n| \geq |z^n| - |\sum_{k=0}^{n-1} \alpha_k z^k| \geq r^n - \sum_{k=0}^{n-1} |\alpha_k| r^k$$

$$\geq r^{n-1}(r - \sum_{k=0}^{n-1} |\alpha_k|) \to \infty$$

als $r \to \infty$. Maar dan heeft $(x,y) \to |P(z)|$ op iedere disk

$$\{(x,y) : \, x^2 + y^2 \leq r^2\}$$

een strict positief globaal minimum dat voor grote r niet meer verandert als r nog groter genomen wordt, en dus een globaal minimum op $\mathbb{R}^2$, zeg in $z = z_0$.

Stel nu $w = z - z_0$ en

$$Q(w) = \frac{P(z)}{P(z_0)}.$$

Dan is

$$Q(w) = 1 + \sum_{k=1}^{n} \gamma_k w^k \tag{6.4}$$

[3] Functies met $\Delta f = 0$, zie (6.14) en Sectie 9.4.

[4] Met $z^2 + 1$ bijvoorbeeld, als je de lachers op je hand wil krijgen.

[5] Die zijn er niet voor graad vijf en hoger.

[6] Bewijs en gebruik $|a + b| \geq |a| - |b|$ voor $a, b \in \mathbb{C} = \mathbb{R}^2$.

met $\gamma_1, \gamma_2, \ldots, \gamma_n \in \mathbb{C}$, $\gamma_n \neq 0$, en $w \to |Q(w)|$ heeft nu een globaal minimum ter grootte $Q(0) = 1$.

Hetzelfde geldt[7] voor $z \to |Q(z)|$ en dus neemt $u + iv = w = Q(z) = Q(x + iy)$ geen w-waarden aan met $u^2 + v^2 < 1$. Maar met Opgave 6.1 is $\gamma_k = c_k \exp(i\phi_k)$ en dus

$$w = Q(z) = 1 + \sum_{k=1}^{n} c_k r^k \exp(i(\phi_k + k\theta)), \tag{6.5}$$

een uitdrukking[8] waarin de ϕ_k-tjes als parameters[9] gezien kunnen worden met r klein maar vast genomen. En dat geeft een tegenspraak:

Opgave 6.2. Leidt nu een tegenspraak af door aan te tonen dat er waarden met $u^2 + v^2 < 1$ worden aangenomen. Hint: er is een eerste c_k die niet nul is, en voor r klein is die term groter dan de rest van de som.

6.2 Differentiaalrekening, ook met complexe getallen

In Sectie 5.2 hebben we gezien, uitgaande van het voorbeeld $x \to x^7$, dat

$$x \to \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \cdots = \sum_{n=1}^{\infty} \alpha_n x^n$$

voor iedere keuze van $\alpha_n \in \mathbb{R}$ voor $n \in \mathbb{N}_0$ een functie definieert op een

$$B_R = \{x \in \mathbb{R} : |x| < R\}$$

met $R \in [0, \infty]$, een functie waarmee voor wat betreft differentiaalrekening mag worden gerekend als met polynomen. Het punt om hier te maken is dat, met een search en replace waarbij het linker gedeelte van

$$x, x_0, a, \alpha_n \in \mathbb{R} \quad \leftrightarrow \quad z = x+iy, z_0 = x_0+iy_0, c = a+bi, \gamma_n = \alpha_n+i\beta_n \in \mathbb{C}$$

wordt vervangen door het rechtergedeelte[10], de hele vrijwel algebraische behandeling van differentiaalrekening van machtreeksen gewoon doorgaat, met het begrip complex differentieerbaar voor functies

$$H : \mathbb{C} \to \mathbb{C},$$

[7] We komen letters te kort, dus hergebruiken we ze hier maar voor de *dummy* variabele.
[8] Niet zo fraai aan het bewijs is het gebruik van die niet-algebraische poolcoördinaten.
[9] Ptolemaeus had dit wel mooi gevonden.
[10] Ik probeer systematisch te zijn met het alphabet voor zolang het duurt.

overgeschreven van gewoon differentieerbaar voor $F : \mathbb{R} \to \mathbb{R}$.

We beginnen dus met de complexe versie van (4.3) in de vorm

$$w = H(z) = H(z_0) + \gamma(z - z_0) + T(z; z_0) \tag{6.6}$$

$$= H(z_0) + H'(z_0)(z - z_0) + o(|z - z_0|)$$

als $z \to z_0$, en pakken dit nu uit, waarbij $w = u + iv$ en u en v dus functies van x en y zijn, $u = F(x, y)$ en $v = G(x, y)$, en[11] $H(z) = F(x, y) + iG(x, y)$.

We kunnen H dan natuurlijk ook zien als vectorwaardige functie

$$H : \mathbb{R}^2 \to \mathbb{R}^2$$

via de identificatie $\mathbb{C} = \mathbb{R}^2$, met componenten $H_1 = F$ en $H_2 = G$ als functies van[12] $x_1 = x$ en $x_2 = y$.

Voor zulke functies hebben we ook de gewone aanpak van differentieerbaarheid, met lineaire expansies

$$H_1(x_1, x_2) = H_2(a_1, a_2) + A_{11}(x_1 - a_1) + A_{12}(x_2 - a_2) + R_1(x_1, x_2; a_1, a_2);$$

$$H_2(x_1, x_2) = H_2(a_1, a_2) + A_{21}(x_1 - a_1) + A_{22}(x_2 - a_2) + R_2(x_1, x_2; a_1, a_2),$$

en resttermen $R_{1,2}$ die klein moeten zijn in verhouding tot het lineaire stuk[13] als $x_1 \to a_1$ en $x_2 \to a_2$. Dat zeggen we zo wat precieser, maar eerst nog een opgave om te laten zien waar de observaties hierboven toe leiden als je een eenvoudige voorbeeldje uitwerkt.

Opgave 6.3. Werk $w = \exp(z) = \exp(x+iy)$ uit met de machtreeks voor $\exp(z)$ en verifieer dat $\exp(z) = \exp(x)\exp(iy)$ met $\exp(iy) = \cos x + i \sin x$. Leg uit waarom dit leidt het invoeren van de meerwaardige[14] functie

$$w \to \log w = \ln |w| + i \arg w.$$

In vectorvorm schrijven we de expansies voor H_1 en H_2 als $H(x_1, x_2) =$

$$\begin{pmatrix} H_1(x_1, x_2) \\ H_2(x_1, x_2) \end{pmatrix} = \begin{pmatrix} H_2(x_1, x_2) \\ H_2(a_1, a_2) \end{pmatrix} + \begin{pmatrix} A_{11}(x_1 - a_1) + A_{12}(x_2 - a_2) \\ A_{21}(x_1 - a_1) + A_{22}(x_2 - a_2) \end{pmatrix} + R,$$

[11] Met toevallig de H van holomorf.

[12] Variabelen nummeren we met alphabet en indices, al naar wat uitkomt.

[13] De termen met de coëfficiënten $A_{11}, A_{12}, A_{21}, A_{22}$.

[14] Een kniesoor die er op let: meerwaardige functies zijn geen functies.

waarin ook $R = R(x_1, x_2; a_1, a_2)$, net als $H(x_1, x_2)$ en $H(a_1, a_2)$, weer een 2-vector is. Met $h_i = x_i - a_i$ is het lineaire stuk in de expansie voor $H(x_1, x_2)$ hierboven in matrixnotatie[15] gelijk aan

$$Ah = \begin{pmatrix} A_{11}h_1 + A_{12}h_2 \\ A_{21}h_1 + A_{22}h_2 \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}, \tag{6.7}$$

en af te schatten via bijvoorbeeld

$$|A_{11}h_1 + A_{12}h_2| \leq \sqrt{A_{11}^2 + A_{12}^2}\,\sqrt{h_1^2 + h_2^2};$$

$$|A_{21}h_1 + A_{22}h_2| \leq \sqrt{A_{21}^2 + A_{22}^2}\,\sqrt{h_1^2 + h_2^2},$$

als

$$|A_{11}h_1 + A_{12}h_2|^2 + |A_{21}h_1 + A_{22}h_2|^2 \leq (\underbrace{A_{11}^2 + A_{12}^2 + A_{21}^2 + A_{22}^2}_{|A|^2})(h_1^2 + h_2^2),$$

ofwel

$$|Ah|_2 \leq |A|_2\,|h|_2, \tag{6.8}$$

waarin de "absolute waarde strepen" met subindex 2 nu voor Euclidische lengte van $h = x - a$, Ah en ook A staan[16], steeds de wortel uit de som van de kwadraten van de entries, ook voor de matrix A geintroduceerd in (6.7), die we steeds ook zien als (lineaire) afbeelding[17]

$$h \xrightarrow{A} Ah.$$

De conditie voor differentieerbaarheid[18] is met

$$H(x) = H(a) + A(x - a) + R(x; a) \quad \text{dat} \quad |R(x; a)|_2 = o(|x - a|_2) \tag{6.9}$$

als $|x - a|_2 \to 0$. Dit is typografisch nu niet meer te onderscheiden van de differentieerbaarheid van $H : \mathbb{R}^m \to \mathbb{R}^n$, met bijvoorbeeld $m = n = 1$. Ook voor $H : X \to Y$ met X en Y genormeerde ruimten, $A : X \to Y$ lineair en continu[19], ziet het er, afgezien van de subindex 2, precies hetzelfde uit, zie eventueel al Sectie 10.2.

[15] Spot the matrix.
[16] Dus $|h|_2^2 = h_1^2 + h_2^2$ en $|A|_2^2 = A_{11}^2 + A_{12}^2 + A_{21}^2 + A_{22}^2$.
[17] Wat moet je hier van matrices weten? Deze afspraak en de matrixnotatie als in (6.7).
[18] Vergelijk ook met (4.25) in Sectie 4.6, waar $a = (0, 0)$ genomen is in (4.24).
[19] Lineaire afbeeldingen zijn dan niet meer automatisch continu!

Door per component van H steeds één variabele tegelijk te variëren zien we dat A de afgeleiden van de partiële functies

$$x_1 \to H_1(x_1, x_2),\ x_2 \to H_1(x_1, x_2),\ x_1 \to H_2(x_1, x_2),\ x_2 \to H_2(x_1, x_2)$$

geëvalueerd in $(x_1, x_2) = (a_1, a_2)$ als entries heeft, dus

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \begin{pmatrix} \frac{\partial H_1}{\partial x_1} & \frac{\partial H_1}{\partial x_2} \\ \frac{\partial H_2}{\partial x_1} & \frac{\partial H_2}{\partial x_2} \end{pmatrix} = H'(x) = DH(x) = \frac{\partial H}{\partial x} \tag{6.10}$$

als $m = n = 2$. In Sectie 9.3 komen we hier nog op terug. We zien $A = H'(x)$ als lineaire afbeelding[20] van in dit geval $\mathbb{R}^2$ naar $\mathbb{R}^2$, die we identificeren met zijn Jacobimatrix $DH(x)$, een matrix die we ook graag noteren met $\frac{\partial H}{\partial x}$. In die Jacobimatrix staan in de rijen (horizontaal) de coördinaten van de gradiënten van de $\mathbb{R}$-waardige functies H_1 en H_2.

Vergelijk dit nu met (6.6), waarin, met $h = x - x_0$ en $k = y - y_0$, de lineaire term herschrijft als

$$\gamma(z - z_0) = (\alpha + i\beta)(h + ik) = \alpha h - \beta k + i(\beta h + \alpha k),$$

en correspondeert met

$$\begin{pmatrix} \alpha h - \beta k \\ \beta h + \alpha k \end{pmatrix} = \begin{pmatrix} \alpha & -\beta \\ \beta & \alpha \end{pmatrix} \begin{pmatrix} h \\ k \end{pmatrix}, \tag{6.11}$$

met daarin de matrix die de afbeelding (6.1) beschrijft. In het bijzonder staat in de eerste kolom het beeld van de eerste eenheidsvector, en in de tweede kolom het beeld van de tweede eenheidsvector.

Opgave 6.4. Overtuig jezelf ervan dat (6.1) een draaivermenigvuldiging is, i.e. de samenstelling van een rotatie rond de oorsprong en een puntvermenigvuldiging vanuit de oorsprong.

We zien nu dat (6.9) dan moet gelden met een bijzondere vorm voor de 2×2 matrix A. Dat formuleren we als stelling.

Stelling 6.5. *Laat $H : \mathbb{R}^2 \to \mathbb{R}^2$ differentieerbaar zijn in (a_1, a_2) met lineaire benadering gegeven door (6.7). Dan is H gezien als $H : \mathbb{C} \to \mathbb{C}$ ook complex differentieerbaar in (a_1, a_2) dan en slechts dan als voor de matrix A geldt dat*

$$\frac{\partial H_1}{\partial x_1} = A_{11} = A_{22} = \frac{\partial H_2}{\partial x_2} \quad \text{en} \quad \frac{\partial H_1}{\partial x_2} = A_{12} = -A_{21} = -\frac{\partial H_2}{\partial x_1},$$

de afgeleiden geëvalueerd in $(x_1, x_2) = (a_1, a_2)$.

[20] Wat moet je met zulke A kunnen? Vooral de inverse uitrekenen, zie Sectie 6.3.

Is H op deze manier complex differentieerbaar, dan schrijven we nu even liever $(x, y) = (x_0, y_0)$ dan $(x_1, x_2) = (a_1, a_2)$. Via

$$z = x + iy \in \mathbb{C} \leftrightarrow (x, y) \in \mathbb{R}^2 \quad \text{en} \quad w = u + iv \in \mathbb{C} \leftrightarrow (u, v) \in \mathbb{R}^2$$

en met

$$\alpha + i\beta = H'(z_0)$$

geldt dan dat α en β gegeven worden door

$$\alpha = \frac{\partial u}{\partial x} \quad \text{en} \quad \beta = -\frac{\partial u}{\partial y}, \tag{6.12}$$

waarbij u en v als functies van x en y voldoen aan de zogenaamde Cauchy-Riemann vergelijkingen

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \quad \text{en} \quad \frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y} \tag{6.13}$$

in $(x, y) = (x_0, y_0)$.

Zijn deze partiële afgeleiden zelf differentieerbaar[21] in $(x, y) = (x_0, y_0)$ dan volgt dat

$$\frac{\partial^2 u}{\partial x^2} = \frac{\partial}{\partial x}\frac{\partial v}{\partial y} = \frac{\partial}{\partial y}\frac{\partial v}{\partial x} = -\frac{\partial}{\partial y}\frac{\partial u}{\partial y} = -\frac{\partial^2 u}{\partial y^2},$$

als de volgorde van differentiëren er niet toe doet, en net zo voor $v(x, y)$. In dat geval geldt

$$\Delta u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0 = \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} = \Delta v \tag{6.14}$$

in (x_0, y_0).

Op de differentiaaloperator Δ, die de *Laplaciaan* heet, komen we nog uitgebreid terug, en juist daarom willen we nu wel even zeker weten hoe het zit met de volgorde van de partiële afgeleiden. Dit is daarom zo'n beetje de laatste stelling uit de analyse over differentiaalrekening die we echt nodig hebben.

Stelling 6.6. *Laat $v : \mathbb{R}^2 \to \mathbb{R}$ de eigenschap hebben dat*

$$(x, y) \to \frac{\partial v}{\partial x} = v_x(x, y) \quad \text{en} \quad (x, y) \to \frac{\partial v}{\partial y} = v_y(x, y)$$

differentieerbaar zijn in (x_0, y_0). Dan bestaan in (x_0, y_0) de tweede orde partiële afgeleiden en geldt

$$v_{yx}(x_0, y_0) = \frac{\partial}{\partial x}\frac{\partial v}{\partial y} = \frac{\partial}{\partial y}\frac{\partial v}{\partial x} = v_{xy}(x_0, y_0).$$

[21] Geldt onder de aanname dat f complex differentieerbaar is in de buurt van (x_0, y_0)!

Bewijs. Neem zonder beperking der algemeenheid $(x_0, y_0) = (0, 0)$. De aanname impliceert dat de eerste orde partiële afgeleiden bestaan in de buurt van $(0, 0)$. Voor $x \neq 0$ en $y \neq 0$ voldoende klein geldt vanwege de middelwaardestelling (Stelling 4.5), toegepast op

$$y \to v(x, y) - v(0, y)$$

voor zekere (ook x-afhankelijke) η tussen 0 en y, en de differentieerbaarheid van v_y in $(0, 0)$ dat

$$v(x, y) - v(0, y) - v(x, 0) + v(0, 0) = (v_y(x, \eta) - v_y(0, \eta))y$$

$$= (v_y(0, 0) + v_{yx}(0, 0)x + v_{yy}(0, 0)\eta + R(x, \eta) - v_y(0, 0) - v_{yy}(0, 0)\eta - R(0, \eta))y$$

$$= (v_{yx}(0, 0)x + R(x, \eta) - R(0, \eta))y,$$

waarin

$$R(x, \eta) = o(\sqrt{x^2 + \eta^2}) \quad \text{en dus ook} \quad R(0, \eta) = o(\eta) \quad \text{als} \quad \sqrt{x^2 + \eta^2} \to 0.$$

De differentieerbaarheid van

$$(x, y) \to v_y(x, y)$$

in $(0, 0)$ is hier dus twee keer gebruikt, met dezelfde "resttermfunctie" R. Omdat $|\eta| \leq |y|$ concluderen we dat

$$v(x, y) - v(0, y) - v(x, 0) + v(0, 0) = v_{yx}(0, 0)xy + y\, o(r)$$

$$= v_{yx}(0, 0)xy + o(r^2) = v_{xy}(0, 0)xy + o(r^2) \tag{6.15}$$

als $r = \sqrt{x^2 + y^2} \to 0$.

De tweede versie onder (6.15) volgt door de rollen van x en y te verwisselen. Merk op dat het eerste gelijkteken op de regel in (6.15) hierboven van links naar rechts moet worden gelezen. Als een restterm y keer kleine o van r is, dan is die term zeker kleine o van r^2, en zo bewijst (6.15) dat $v_{yx}(0, 0) = v_{yx}(0, 0)$.

Opgave 6.7. Zie (4.4) en gebruik de definitie van kleine o om de conclusie dat $v_{yx}(0, 0) = v_{yx}(0, 0)$ te rechtvaardigen.

Twee keer differentieerbare functies $u(x,y)$ en $v(x,y)$ die op een open[22] verzameling van $\mathbb{R}^2$ voldoen aan (6.14) heten harmonisch. Bijgevolg zijn de functies

$$(x,y) \to \mathrm{Re}(x+iy)^n \quad \text{en} \quad (x,y) \to \mathrm{Im}(x+iy)^n$$

harmonisch op heel $\mathbb{R}^2$. Dit zijn de zogenaamde homogene harmonische polynomen van graad $n = 1, 2, 3, \ldots$ die in (6.17) nog terugkomen.

Twee keer differentieerbaar betekent hier dat de afbeelding

$$(x,y) \to (\frac{\partial u}{\partial x} \; \frac{\partial u}{\partial y})$$

zelf weer differentieerbaar is. Met de kettingregel[23] zijn dan ook

$$(x,y) \to (\frac{\partial u}{\partial x} \; \frac{\partial u}{\partial y}) \to \frac{\partial u}{\partial x} \quad \text{en} \quad (x,y) \to (\frac{\partial u}{\partial x} \; \frac{\partial u}{\partial y}) \to \frac{\partial u}{\partial y}$$

differentieerbaar en dus heeft $\Delta u = 0$ betekenis als

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0, \tag{6.16}$$

zonder dat daar een v bij hoort[24] via een complexwaardige functie.

Niet-constante oplossingen van (6.16) zijn er in overvloed. Bijvoorbeeld de homogene *harmonische polynomen*

$$x, y, x^2-y^2, 2xy, x^3-3xy^2, 3x^2y-y^3, x^4-6x^2y^2+y^4, 4x^3y-4xy^3, \ldots \tag{6.17}$$

van respectievelijk graad $1, 1, 2, 2, 3, 3, 4, 4, \ldots$ hierboven.

6.3 Randvoorwaarden en de methode van Lagrange

Dit onderwerp is al begonnen in Sectie 4.7 waar we de formule

$$\Phi_x = \Phi_y (F_y)^{(-1)} F_x \tag{6.18}$$

in $(x,y) = (0,0)$ hebben afgeleid voor het stationair zijn van

$$x \xrightarrow{\phi} \phi(x) = \Phi(x, f(x))$$

in $x = 0$ met de in Sectie 4.6 gevonden impliciete functie

$$y = f(x)$$

[22] Zie Opgave 12.14 in Hoofdstuk 12 voor een definitie van open.
[23] Triviaal toegepast op een ketting waarin de tweede schakel lineair is.
[24] Althans niet a priori.

als lokale beschrijving van $F(x,y)=0$ in de buurt van $(x,y)=(0,0)$.

De continuiteit van de partiële afgeleiden

$$(x,y)\to F_x(x,y) \quad \text{en} \quad (x,y)\to F_y(x,y)$$

in de buurt van $(0,0)$, en de inverteerbaarheid van F_y in $(0,0)$ waren voldoende voor een bewijs dat de niveauverzameling

$$S=\{(x,y):\ F(x,y)=F(0,0)\} \tag{6.19}$$

in de buurt van $(x,y)=(0,0)$ beschreven wordt als de grafiek van een impliciet gedefinieerde continu differentieerbare functie f.

De verzameling S is zo lokaal geparametriseerd door

$$x\to X(x)=(x,f(x))$$

met een 2×1 Jacobimatrix $\frac{\partial X}{\partial x}$. De parameterisatie is lokaal een bijectie tussen S en een omgeving van $x=0$, en dat is dankzij de inverteerbaarheid van de 1×1 matrix

$$A=F_y \tag{6.20}$$

in $(0,0)$.

Differentieerbaarheid van

$$(x,y)\to\Phi(x,y)$$

was vervolgens voldoende voor (6.18) als noodzakelijke en voldoende voorwaarde voor $\phi'(x)=0$, niet alleen in $x=0$ maar in de hele omgeving van $x=0$ waar $f(x)$ gemaakt en $F_y(x,f(x))$ inverteerbaar is, en daar deden we een beetje moeilijk over in Opgave 4.23, waar *delen door* werd vermeden met meetkundige reeksen.

Dat had een reden want met bijvoorbeeld

$$x\in\mathbb{R}^2,\ y\in\mathbb{R}^3,$$

$$F:\mathbb{R}^5\to\mathbb{R}^3,\ \Phi:\mathbb{R}^5\to\mathbb{R},$$

$$f:\mathbb{R}^2\to\mathbb{R}^3,\ \phi:\mathbb{R}^2\to\mathbb{R},$$

zijn de stellingen en bewijzen hetzelfde als in Secties 4.4, 4.5, 4.6, 4.7. In het bijzonder geldt dat voor het bestaan van de functie f en conditie (6.18) als karakterisatie voor het stationair zijn van de functie

$$x\xrightarrow{\phi}\Phi(x,f(x)).$$

We laten nu eerst zien hoe dat uitpakt als de multiplicatoren methode van Lagrange en lezen (6.18) daartoe als een uitspraak voor Jacobimatrices en de bijbehorende lineaire afbeeldingen. Als je vertrouwd bent met matrixrekening kun je het volgende overslaan en verder lezen vanaf (6.23), de getransponeerde versie van (6.18).

In het algemeen is een[25] $m \times n$ reële matrix A een blok met reële entries a_{ij}, de verticale index i lopend van 1 tot en met m, de horizontale index j van 1 tot en met n. Als lineaire afbeelding stuurt A dan een n-vector $x \in \mathbb{R}^n$ met coördinaten $x_1, \ldots, x_n$ naar een een m-vector $y \in \mathbb{R}^m$ met coördinaten

$$y_i = \sum_{j=1}^{n} a_{ij} x_j.$$

We zeggen dat

$$A \in L(\mathbb{R}^n, \mathbb{R}^m),$$

de verzameling van lineaire afbeeldingen van $\mathbb{R}^n$ naar $\mathbb{R}^m$, die zelf weer een vectorruimte vormen.

Is B een reële $n \times p$ matrix met entries b_{jk}, de verticale index j lopend van 1 tot en met n, de horizontale index k van 1 tot en met p, dan is per definitie AB de $m \times p$ matrix met entries

$$\sum_{j=1}^{n} a_{ij} b_{jk}, \tag{6.21}$$

met bijbehorende lineaire afbeelding[26]

$$A \circ B : \mathbb{R}^p \xrightarrow{B} : \mathbb{R}^n \xrightarrow{A} \mathbb{R}^m.$$

Klappen we beide blokken om door met de eerste index juist horizontaal, en met de tweede index verticaal te nummeren, dan krijgen we *getransponeerde* matrices A^T en B^T met entries $a^T_{ji} = a_{ij}$ en $b^T_{kj} = b_{jk}$, leest (6.21) als de entries van $B^T A^T$, en is $(AB)^T = B^T A^T$.

In het bijzonder geval dat $m = n = p$ kan het voorkomen dat $AB = I$, de $n \times n$ matrix met op de diagonaal alleen maar 1-en en voor de rest alleen maar 0-en. Met I correspondeert de lineaire afbeelding $I = I_n$ die elke $x \in \mathbb{R}^n$ naar zichzelf stuurt. *Wat je van lineaire albegra nu echt moet weten*[27] is niet alleen dat in dat geval de lineaire afbeelding $A \circ B$ gelijk is aan de afbeelding

[25] Met m en n in $\mathbb{N}$ natuurlijk.

[26] A voorafgegaan door B.

[27] Een bewijs zou bij Lineaire Algebra in een van de eerste colleges moeten zitten.

I_n, maar ook dat A en B nu elkaars inverse zijn als lineaire afbeeldingen: $A \circ B = B \circ A = I_n$, $B = A^{(-1)}$, $A = B^{(-1)}$. En hetzelfde geldt dan voor de getransponeerden. Uitspraken die alleen te maken hebben met het oplossen van vergelijkingen van de vorm $Ax = y$ met A een vierkante matrix, en daarop komen we in Sectie 6.5 nog terug.

Met nog een derde $p \times r$ matrix C met entries c_{kl} is $(AB)C$ de matrix met entries

$$\sum_{k=1}^{p}(\sum_{j=1}^{n} a_{ij}b_{jk})c_{kl} = \sum_{k=1}^{p}\sum_{j=1}^{n} a_{ij}b_{jk}c_{kl}, \tag{6.22}$$

hetgeen ook de entries zijn van $A(BC)$ door de sommatie volgorde te verwisselen. Kortom $(AB)C = A(BC)$ en dus schrijven we ABC voor het produkt van A, B en C. De bijbehorende lineaire afbeelding is $A \circ B \circ C$. Transponeren geeft dan weer dat $(ABC)^T = C^T B^T A^T$, en dat gebruiken we nu voor (6.18).

We schrijven (6.18) getransponeerd als

$$\nabla_x F \, (\nabla_y F)^{(-1)} \, \nabla \Phi_y = \nabla_x \Phi, \tag{6.23}$$

waarin

$$\nabla_x F, \nabla_y F, \nabla_x \Phi, \nabla_y \Phi$$

de getransponeerde matrices zijn van de "partiële" Jacobimatrices

$$\frac{\partial F}{\partial x}, \frac{\partial F}{\partial y}, \frac{\partial \Phi}{\partial x}, \frac{\partial \Phi}{\partial y}$$

behorende bij F_x, F_y, Φ_x, Φ_y.

Uitpakken[28] van de notatie geeft

$$\nabla_x F = (\nabla_x F_1 \ \nabla_x F_2 \ \nabla_x F_3) = \begin{pmatrix} \frac{\partial F_1}{\partial x_1} & \frac{\partial F_2}{\partial x_1} & \frac{\partial F_3}{\partial x_1} \\ \frac{\partial F_1}{\partial x_2} & \frac{\partial F_2}{\partial x_2} & \frac{\partial F_3}{\partial x_2} \end{pmatrix}$$

en hetzelfde voor $\nabla_y F$, een bij aanname in $(0,0,0,0,0)$ vierkante inverteerbare 3×3 matrix, waarvan de inverse als afbeelding de gradiëntvectoren

$$\nabla_y F_1, \nabla_y F_2, \nabla_y F_3$$

terugstuurt[29] naar de eenheidsvectoren e_1, e_2, e_3 in $\mathbb{R}^3$.

Schrijf nu $\nabla_y \Phi \in \mathbb{R}^3$ als lineaire combinatie[30]

$$\nabla_y \Phi = \lambda_1 \nabla_y F_1 + \lambda_2 \nabla_y F_2 + \lambda_3 \nabla_y F_3, \tag{6.24}$$

[28] Meer is het niet, maar ga dat wel na!

[29] Ga na: de kolomvectoren van een matrix A zijn de beelden onder A van de e-tjes.

[30] Dit kan vanwege de inverteerbaarheidseis op F_y.

met $\lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}$. Dan volgt dat $\nabla_x F \left(\nabla_y F\right)^{(-1)}$ in het linkerlid van (6.23) werkt op (6.24) als

$$\nabla_y \Phi \xrightarrow{(\nabla_y F)^{(-1)}} \lambda_1 e_1 + \lambda_2 e_2 + \lambda_3 e_3 \xrightarrow{\nabla_x F} \lambda_1 \nabla_x F_1 + \lambda_2 \nabla_x F_2 + \lambda_3 \nabla_x F_3 = \nabla_x \Phi$$

vanwege (6.23), hetgeen met (6.24) combineert tot

$$\nabla \Phi = \lambda_1 \nabla F_1 + \lambda_2 \nabla F_2 + \lambda_3 \nabla F_3, \tag{6.25}$$

simpelweg[31] omdat het voor ∇_x en ∇_y apart geldt! Het stationair zijn van

$$\Phi : S \to \mathbb{R}$$

in $(0,0)$ is zo equivalent met het bestaan van multiplicatoren $\lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}$ waarvoor (6.25) geldt in $(0,0,0,0,0)$.

6.4 Toepassing: de ongelijkheid van Hölder

In (6.8) was

$$\left|Ah\right|_2 \leq \left|A\right|_2 \left|h\right|_2$$

een bijzonder geval van

$$\left|AB\right|_2 \leq \left|A\right|_2 \left|B\right|_2,$$

met A en B die op elkaar passen in de zin dat AB als produktmatrix bestaat, zie de uitleg rond (6.21), en (6.30) war verderop. In het speciale geval dat $A = a$ een rijmatrix is met entries a_i en $B = b$ een kolommatrix met entries b_i, staat hier dat

$$|\sum_{i=1}^{n} a_i b_i| \leq \sqrt[2]{\sum_i |a_i|^2} \sqrt[2]{\sum_i |b_i|^2},$$

een ongelijkheid die bekend staat als de ongelijkheid van Cauchy-Schwarz[32], en die zegt dat het inprodukt van twee vectoren in absolute waarde op zijn hoogst gelijk is aan het produkt van hun lengten uitgerekend met Pythagoras. Let wel, deze ongelijkheid wordt standaard bij elk eerste Lineaire Algebra college bewezen, en vervolgens gebruikt om te laten zien dat de Pythagorasnorm aan de driehoeksongelijkheid voldoet, en op die ongelijkheid is alles wat we tot nu doen wel zo'n beetje gebaseerd. Wat nu volgt voor algemenere exponenten is dus niet zomaar een alternatief bewijs van de ongelijkheid van

[31] Er hoeft dus geen 3×3 matrix geinverteerd te worden.....

[32] Die elkaar royaal niet gekend hebben.

Cauchy-Schwarz zelf, want we stellen nu alleen de vraag voor welke andere waarden van p en q ook geldt dat

$$|\sum_{i=1}^{n} a_i b_i| \leq |a|_p \, |b|_q, \tag{6.26}$$

als $|a|_p$ en $|b|_q$ gedefinieerd zijn door (we vermijden de wortels)

$$|a|_p^p = \sum_{i=1}^{n} |a_i|^p \quad \text{en} \quad |b|_q^q = \sum_{i=1}^{n} |b_i|^q.$$

Met $p = q = 2$ is (6.26) ongelijkheid van Cauchy-Schwarz.

Merk op dat voor elke $p > 0$ en $q > 0$ de ongelijkheid (6.26) schaalt met a en b. De vraag voor welke p en q de ongelijkheid geldt kan dus beantwoord worden door het linkerlid te bekijken op de verzameling van vectoren a en b gegeven door $|a|_p = |b|_q = 1$, ofwel

$$\phi(a_1, \ldots, a_n) = |a_1|^p + \cdots + |a_n|^p = 1;$$

$$\psi(b_1, \ldots, b_n) = |b_1|^q + \cdots + |b_n|^q = 1,$$

twee randvoorwaarden waaronder we het maximum en het minimum van de functie

$$(a_1, \ldots, a_n, b_1, \ldots, b_n) \xrightarrow{F} a_1 b_1 + \cdots + a_n b_n$$

zoeken. We weten dat dat maximum en minimum bestaan want de randvoorwaarden definiëren een begrensde gesloten deelverzameling van $\mathbb{R}^{2n}$ en F is continu.

De randvoorwaarden worden gegeven door de functies ϕ en ψ die echter alleen voor $p > 1$ en $q > 1$ continu differentieerbaar zijn, omdat de afgeleide van $x \to |x|^p$ gelijk is aan $x \to px^{p-1}$, waarbij we even de tijdelijke afspraak maken dat we voor $r > 0$ de functie $x \to x^r$ als oneven functie definiëren. Met twee Lagrange multiplicatoren λ en μ krijgen we nu de $2n$ vergelijkingen

$$b_i = \lambda p a_i^{p-1}; \quad a_i = \mu q b_i^{q-1} \quad (i = 1, \ldots, n)$$

om op te lossen samen met

$$\sum_{i=1}^{n} |a_i|^p = \sum_{i=1}^{n} |b_i|^q = 1.$$

Opgave 6.8. Neem aan dat $(p-1)(q-1) \neq 1$. Bewijs dat in dat geval voor de oplossingen van deze vergelijkingen geldt dat alle $|a_i|$ aan elkaar gelijk zijn en alle $|b_i|$ aan elkaar gelijk zijn, en dat bijgevolg

$$\sum_{i=1}^{n} |a_i b_i| = n(\frac{1}{n})^{\frac{1}{p}+\frac{1}{q}} = n^{1-\frac{1}{p}-\frac{1}{q}}. \tag{6.27}$$

Leid hieruit af dat (6.26) geldt voor $p > 1$ en $q > 1$ met $\frac{1}{p} + \frac{1}{q} = 1$.

6.5 Aftelbare sommen van vierkante matrices

Het enige dat we na Opgave 4.23 nog moeten doen om zeker te weten dat alles uit Secties 4.4, 4.5, 4.6, 4.7 hier overgeschreven kan worden, is hoe het zit met meetkundige reeksen als

$$I + A + A^2 + A^3 + \cdots, \tag{6.28}$$

met A een vierkante matrix, bijvoorbeeld een 2×2 matrix als in (6.7). Die vraag verdient nu deze aparte sectie. In de volgende sectie behandelen we dan de impliciete functiestelling voor $F : \mathbb{R}^{n+m} \to \mathbb{R}^m$.

Voor een 2×2 matrix als in (6.7) hebben we voor de bijhorende lineaire afbeelding $A : \mathbb{R}^2 \to \mathbb{R}^2$ een schatting van de vorm (6.8). Je gaat eenvoudig na[33] dat voor elke $n \times n$ matrix en elke reële n-vector h geldt dat

$$|Ah|_2 \leq M|h|_2, \tag{6.29}$$

waarin $M \geq 0$ gedefinieerd is door

$$M^2 = \sum_{i,j=1}^{n} a_{ij}^2,$$

zeg maar de Pythagoraslengte van A, dus $M = |A|_2$. De kleinste M waarvoor (6.29) geldt heet de operatornorm van A, notatie $|A|_{op}$, waarbij we A nu zien als lineaire afbeelding van $\mathbb{R}^n$ naar $\mathbb{R}^n$. Deze $|A|$ is dus de grootst mogelijke verhouding tussen de Euclidische lengtes van Ah en h.

We zeggen weer dat

$$A \in L(\mathbb{R}^n, \mathbb{R}^n) = L(\mathbb{R}^n),$$

[33] Voor mijn part met volledige inductie.

waarmee de verzameling $L(\mathbb{R}^n)$ gedefinieerd is, niet alleen een lineaire vectorruimte over $\mathbb{R}$, maar ook een genormeerde algebra, vanwege de extra bewerking

$$(A, B) \to AB,$$

en de eigenschappen van de operatornorm

$$A \to |A|_{op}.$$

Het is eenvoudig om rechtstreeks uit de definities te zien dat niet alleen

$$|A|_{op} = 0 \iff A = 0, \quad |\lambda A|_{op} = |\lambda|\,|A|_{op} \geq 0, \quad |A + B|_{op} \leq |A|_{op} + |B|_{op},$$

maar ook

$$|AB|_{op} \leq |A|_{op}\,|B|_{op}$$

voor alle $A, B \in L(\mathbb{R}^n)$ en $\lambda \in \mathbb{R}$.

Merk op dat $L(\mathbb{R}^n)$ als vectorruimte hetzelfde[34] is als $\mathbb{R}^{n^2}$, met de standaard Pythagorasnorm

$$A \to |A|_2$$

gegeven door

$$|A|_2^2 = \sum_{i,j=1}^{n} a_{ij}^2.$$

Aardig is dat ook

$$|AB|_2 \leq |A|_2\,|B|_2 \tag{6.30}$$

geldt[35], maar omdat $|A|_{op} \leq |A|_2$ voor alle $A \in L(\mathbb{R}^n)$ gebruiken we in Opgave 6.10 liever de kleinere norm $|\cdot|_{op}$ waarvoor deze ongelijkheid ook geldt.

.

Opgave 6.9. Bewijs dat er een $\mu_n \in (0, 1]$ bestaat waar mee geldt dat

$$\mu_n |A|_2 \leq |A|_{op} \leq |A|_2$$

voor alle $A \in L(\mathbb{R}^n)$. Hint[36]: stel niet, dan kun je op

$$\{A \in L(\mathbb{R}^n) : |A|_{op} = 1\}$$

de Pythagorasnorm $|A|_2$, en dus ook een van de lengtes van de kolomvectoren willekeurig groot krijgen, en dat zal toch een tegenspraak moeten geven met $|A|_{op} = 1$.

[34] Of je de entries nu in een blok of in een kolom of rij zet maakt niet echt uit.
[35] Bewijs dat dit zo is!
[36] Terzijde: $|A|_{op}$ is de wortel uit de grootste eigenwaarde van $A^T A$.

Opgave 6.10. Neem aan dat $A \in L(\mathbb{R}^n)$ voldoet aan $|A|_{op} < 1$. Laat zien dat voor de reeks in (6.28) geldt dat

$$(I - A)(I + A + A^2 + A^3 + \cdots) = I,$$

en leg uit waarom

$$(I + A)^{(-1)} = I - A + A^2 - A^3 + \cdots = \sum_{j=0}^{\infty} (-A)^j$$

als $|A|_{op} < 1$, hetgeen zeker het geval is als $|A|_2 < 1$.

De details van de impliciete functiestelling voor vectorwaardige functies van meer variabelen en de multiplicatorenmethode van Lagrange kun je nu wel invullen. Maar ook zie je met een wat suggestieve notatie dat de machtreeks

$$\sum_{j=0}^{\infty} x^j$$

van de functie

$$x \xrightarrow{f} (1 - x)^{-1}$$

correspondeert met

$$A \xrightarrow{f} \sum_{j=0}^{\infty} A^j = I + A + A^2 + A^3 + \cdots . \tag{6.31}$$

Voor welke functies f en welke $A \in L(\mathbb{R}^n)$ is $f(A)$ op een natuurlijke manier gedefinieerd met rekenregels die uit de rekenregels voor de functies volgens? Een vraag om op terug te komen, maar voor machtreeksen[37] kun je een partieel antwoord wel geven nu.

6.6 Impliciete functiestelling in Euclidische ruimten

We gebruiken de notatie

$$x \in X = \mathbb{R}^n,\ y \in Y = \mathbb{R}^m, \quad (x, y) \in X \times Y = \mathbb{R}^{n+m},$$

en formuleren de impliciete functiestelling in de buurt van een vast punt $(x, y) = (a, b)$. Omdat we net als in (4.23) continue differentieerbaarheid van de impliciete functie willen hebben eisen we nu meteen dat $(x, y) \to F_x(x, y)$ en $(x, y) \to F_y(x, y)$ continu zijn in een omgeving van $(x, y) = (a, b)$, en dat is hetzelfde[38] als F continu differentieerbaar op die omgeving.

[37] Denk als eerste aan $\exp(A)$.

[38] Daar komen we nog op terug in algemenere context.

Stelling 6.11. *(Impliciete functiestelling) Laat voor $r > 0$ de functie F continu differentieerbaar zijn op $\bar{B}_r(a) \times \bar{B}_r(b)$. Als $F_y(a,b)$ inverteerbaar is dan bestaan er $\delta_0 > 0$ en $\varepsilon_0 > 0$, en een continu differentieerbare functie $f : \bar{B}_{\delta_0}(a) \to \bar{B}_{\varepsilon_0}(b)$, zodanig dat*

$$\{(x,y) \in \bar{B}_{\delta_0}(a) \times \bar{B}_{\varepsilon_0}(b) : F(x,y) = F(a,b)\} = \{(x, f(x)) : x \in \bar{B}_{\delta_0}(a)\},$$

waarbij geldt dat $f'(x) = -(F_y(x,f(x)))^{-1} F_x(x,f(x))$ voor alle $x \in \bar{B}_{\delta_0}(a)$.

Het bewijs kan nu worden overgeschreven van de bewijzen in Secties 4.4 en 4.5. In het bijzonder wordt de samengestelde functie $x \to F(x, f(x))$ niet gedifferentieerd om de uitdrukking voor $f'(x)$ af te leiden.

Omdat in de constructie van $y = f(x)$ eerst $0 < \varepsilon_0 \leq r$ en daarna $\delta_0 > 0$ klein genoeg gekozen wordt, en vervolgens δ_0 nog kleiner moet worden gekozen om $f'(x) = -(F_y(x,f(x)))^{-1} F_x(x,f(x))$ gedefinieerd te hebben voor $|x| \leq \delta_0$, zal in het algemeen niet gelden dat $\delta_0 > \varepsilon_0$. De facto kan Stelling 6.11 dus gelezen worden als het bestaan van $0 < \delta_0 \leq \varepsilon_0 \leq r$ waarvoor de uitspraken in de stelling gelden.

Passen we Stelling 6.11 toe op

$$F(x,y) = x - g(y),$$

dan krijgen we de inverse functiestelling via de uitspraak

$$\{(x,y) \in \bar{B}_{\delta_0}(a) \times \bar{B}_{\varepsilon_0}(b) : g(y) = x\} = \{(x, f(x)) : x \in \bar{B}_{\delta_0}(a)\},$$

waarbij geldt dat $f'(x) = (g'(f(x))^{-1}$ voor alle $x \in \bar{B}_{\delta_0}(a)$. Merk op dat de oplossing $y = f(x)$ van $x = g(y)$ nu wordt geconstrueerd met

$$y_{n+1} = y_n - g'(0)^{-1}(g(y_n) - x),$$

startend vanaf $y_0 = 0$. We formuleren het resultaat voor $X = Y = \mathbb{R}^n$ en $g : Y \to Y$.

Stelling 6.12. *(Inverse functiestelling) Laat voor $r > 0$ de functie $g : Y \to Y$ continu differentieerbaar zijn op $\bar{B}_r(b)$ en laat $a = g(b)$. Als $g'(b)$ inverteerbaar is dan bestaan er $0 < \delta_0 \leq \varepsilon_0 \leq r$, en een continu differentieerbare injectieve functie $f : \bar{B}_{\delta_0}(a) \to \bar{B}_{\varepsilon_0}(b)$, zodanig dat voor alle $(x,y) \in \bar{B}_{\delta_0}(a) \times \bar{B}_{\varepsilon_0}(b)$ geldt: $x = g(y) \iff y = f(x)$. Bovendien geldt dat $f'(x) = (g'(f(x))^{-1}$ voor alle $x \in \bar{B}_{\delta_0}(a)$.*

N.B. Stelling 6.11 geeft de $f : \bar{B}_{\delta_0}(a) \to \bar{B}_{\varepsilon_0}(b)$ in Stelling 6.12 alleen als continu differentieerbare functie. Omdat $y = f(x)$ voor $x \in \bar{B}_{\delta_0}(a)$ geldt

echter $x = g(y) = g(f(x))$, dus f is injectief op $\bar{B}_{\delta_0}(a)$, en vanwege $f'(x) = (g'(f(x))^{-1}$ is $f'(x)$ inverteerbaar in elke $x \in \bar{B}_{\delta_0}(a)$.

Voor g werkt dit argument niet zonder meer: om $x = g(y)$ in $y = f(x)$ in te mogen vullen moet $g(y)$ wel in het domein van f liggen. Stelling 6.12 kan echter nog een keer worden toegepast (draai de rollen van x en y om) om $0 < \varepsilon_1 \leq \delta_1 \leq \delta_0$ en een continu differentieerbare $g_1 : \bar{B}_{\varepsilon_1}(b) \to \bar{B}_{\delta_1}(a)$ te vinden zodanig dat voor $(x, y) \in \bar{B}_{\delta_1}(a) \times \bar{B}_{\varepsilon_1}(b)$ weer geldt: $x = g_1(y) \iff y = f(x)$. Met de eerdere equivalentie $x = g(y) \iff y = f(x)$ voor alle $(x, y) \in \bar{B}_{\delta_0}(a) \times \bar{B}_{\varepsilon_0}(b)$ zien we dat $g_1 = g$ op $\bar{B}_{\varepsilon_1}(b)$. Net als eerder voor $f : \bar{B}_{\delta_0}(a) \to \bar{B}_{\varepsilon_0}(b)$ volgt nu dat g_1 en dus g injectief is op $\bar{B}_{\varepsilon_1}(b)$. Samenvattend concluderen we dat in de keten

$$\bar{B}_{\varepsilon_1}(b) \xrightarrow{g} \bar{B}_{\delta_1}(a) \to \bar{B}_{\delta_0}(a) \xrightarrow{f} \bar{B}_{\varepsilon_0}(b) \xrightarrow{g} X = Y = \mathbb{R}^n$$

behalve f ook g in de eerste schakel injectief is. Het tweede pijlte (zonder bovenschrift) is de inclusieafbeelding. De keten kan steeds naar links worden uitgebreid. Om precies te zijn, nu maar eens startend met een continu differentieerbare

$$\mathbb{R}^n \supset \bar{B}_{\delta_0}(a) \xrightarrow{f} \mathbb{R}^n \tag{6.32}$$

en $f'(a)$ inverteerbaar, hebben we met $b = f(a)$ een doorlopend diagram

$$\begin{array}{ccc}
\bar{B}_{\delta_0}(a) & \xrightarrow{f} & \mathbb{R}^n \\
\uparrow & & \uparrow \\
\bar{B}_{\delta_1}(a) & \xleftarrow{g} & \bar{B}_{\varepsilon_1}(b) \\
\uparrow & & \uparrow \\
\bar{B}_{\delta_2}(a) & \xrightarrow{f} & \bar{B}_{\varepsilon_2}(b) \\
\uparrow & & \uparrow \\
\bar{B}_{\delta_3}(a) & \xleftarrow{g} & \bar{B}_{\varepsilon_3}(b) \\
\uparrow & & \uparrow
\end{array}$$

De door Stelling 6.12 gemaakte continu differentieerbare

$$\bar{B}_{\varepsilon_1}(b) \xrightarrow{g} \bar{B}_{\delta_1}(a)$$

en iedere schakel behalve de bovenste f zijn injectief (maar in het algemeen niet surjectief), met inverteerbare $f'(x)$ en $g'(y)$ (vanwege $f'(x) = (g'(f(x))^{-1}$ en $g'(y) = (f'(g(y))^{-1})$. Teruglopend langs de horizontale functiepijltjes en

de verticale inclusiepijltjes worden de epsilons en delta's alleen maar kleiner[39]. Over het beeld van ieder bolletje onder f dan wel g is nog wel wat te zeggen natuurlijk, maar eerst kijken we in Sectie 6.8 naar andere vragen waarbij Stelling 6.11 een rol speelt.

Opgave 6.13. Leid Stelling 6.11 af uit Stelling 6.12. Hint: maak m.b.v. F een functie $\tilde{F} : \mathbb{R}^{\mathrm{N}} = \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n \times \mathbb{R}^m$, waarvan de laatste m componenten gegeven worden door $F(x, y)$ en de eerste n door x zelf.

6.7 Flashback

Exercise 4.17 concerned a bijection between two closed intervals $[a, b]$ and $[c, d]$. Compared to (6.32) and the discussion that follows now think of $n = 1$ and $[a, b]$ as a closed ball in $\mathbb{R}$, and a bijection

$$f : [a, b] \to [c, d]. \tag{6.33}$$

By definition $[c, d]$ is the image under f of $[a, b]$ and vise versa $[a, b]$ is the image under the inverse function g of $[c, d]$. The inverse function exists because f is assumed to be a bijection[40]. Thus, unlike the discussion of (6.32), the present discussion is *not* about the existence of the inverse function. It already exists by assumption. Instead we ask, in Opgave 4.17 that is, about the differentiability of g in some $y_0 = f(x_0)$ with $x_0 \in (a, b)$ and f differentiable in x_0 with $f'(x_0) \neq 0$. The positive answer to this question is that g is differentiable in y_0 and that

$$f'(x_0)g'(y_0) = 1, \tag{6.34}$$

a statement which is symmetric in f and g. The exercise calls for a proof of this statement that will not involve the continuity[41] of f and g and relies on a reformulation of the special case that $f(0) = 0$ and $f'(0) = 1$ which is symmetric in x and y, whence the diffferentiability of g in $y = 0$ with $g'(0) = 1$ will be immediate.

To establish the positive answer you should first make you life that easy by showing[42] that without loss of generality you may assume that $0 = x_0 = y_0 = 0 = f(0)$, and that $f'(x_0) = 1$, meaning that

$$f(x) = x + o(x) \quad \text{as} \quad x \to 0, \tag{6.35}$$

[39] In Section 6.7 below we discuss the case $n = 1$ in relation to Section 4.3.
[40] In dimension $n > 1$ we rarely have bijections between balls of course.
[41] But do prove that f and g are continuous without any further assumptions.
[42] You really should.

i.e.

$$\forall \epsilon > 0\, \exists \delta > 0: \quad |x| \le \delta \implies |f(x) - x| \le \epsilon |x|. \tag{6.36}$$

The inequality for $|f(x) - x|$ means that

$$(1-\epsilon)x \le y \le (1+\epsilon)x \quad \text{if} \quad 0 \le x \le \delta \quad \text{and} \quad y = f(x), \tag{6.37}$$

and the other way around for $-\delta \le x \le 0$. We want to replace this statement by an equivalent statement which is symmetric in x and y, and thereby also equivalent to

$$g(y) = y + o(y) \quad \text{as} \quad y \to 0. \tag{6.38}$$

How do we get the equivalent symmetric statement? Clearly the condition $y = f(x)$ already is symmetric because

$$y = f(x) \iff x = g(y),$$

but the inequalities with x and y are not. Note though that

$$(1-\epsilon)x \le y \le (1+\epsilon)x \implies (1-\epsilon)x \le y \le \frac{1}{1-\epsilon}x$$

if $x \ge 0$ and $0 < \epsilon < 1$. In other words (6.36) implies that

$$\forall \epsilon \in (0,1)\, \exists \delta > 0: \quad \begin{matrix} 0 \le x \le \delta \\ y = f(x) \end{matrix} \implies (1-\epsilon)x \le y \le \frac{1}{1-\epsilon}x, \tag{6.39}$$

and likewise[43] for $-\delta \le x \le 0$.

Next observe that (6.39) and its version for $x \le 0$ in turn imply

$$\forall \epsilon \in (0,1)\, \exists \delta > 0: \quad |x| \le \delta \implies |f(x) - x| \le \frac{\epsilon}{1-\epsilon}|x|, \tag{6.40}$$

since

$$\frac{1}{1-\epsilon} = 1 + \frac{\epsilon}{1-\epsilon}.$$

But (6.40) and (6.36) are equivalent, by setting

$$\tilde{\epsilon} = \frac{\epsilon}{1-\epsilon},$$

and thus (6.39) and its version for $x \le 0$ make up for an equivalent definition of (6.35) stating that

$$\forall \epsilon \in (0,1)\, \exists \delta > 0: \quad G_\delta = \{(x,y):\, 0 < |x| \le \delta,\, y = f(x)\} \subset S_\epsilon, \tag{6.41}$$

[43] With the same δ given $0 < \epsilon < 1$, and with reversed inequalities for y.

in which

$$S_\epsilon = \{(x,y) \neq (0,0) : \frac{1}{1-\epsilon} \leq \frac{y}{x} \leq 1-\epsilon\} \tag{6.42}$$

is clearly symmetric in x and y. Now choose $\tilde{\delta} > 0$ such that, for the same $\epsilon \in (0,1)$, it holds that

$$F_{\tilde{\delta}} = \{(x,y) : 0 < |y| \leq \tilde{\delta},\, x = g(y)\} \subset S_\epsilon.$$

How? Draw a picture to see that

$$\tilde{\delta} = (1-\epsilon)\delta$$

does the job. This completes the proof.

6.8 Karakterisaties van deelvarieteiten

Lineaire deelvarieteiten van $\mathbb{R}^{\mathrm{N}}$ zijn per definitie oplossingsverzamelingen van lineaire vergelijkingen. Denk aan punten, lijnen en vlakken als de niet-triviale voorbeelden in $\mathbb{R}^3$. Bij lineaire algebra heb je gezien dat, behalve in het geval van punten, deze niet-triviale lineaire deelvarieteiten, na eventueel hernummeren van de coördinaten, altijd te beschrijven zijn als grafiek van een lineaire functie

$$y = Ax + b, \tag{6.43}$$

waarbij $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, $A :\in \mathbb{R}^n \to \mathbb{R}^m$ lineair, $b \in \mathbb{R}^m$, en $N = n + m$ met $n, m \in \mathbb{N}$.

Aan (6.43) is nog niet te zien of we matrixnotatie gebruiken. Als we x en y zien als kolomvectoren dan is (6.43) te lezen met A een matrix bestaande uit m rijen en n kolommen. In (6.43) staan dan m vergelijkingen die geschreven kunnen worden in de vorm $Ax - y = b$ voor $(x,y) \in \mathbb{R}^{\mathrm{N}}$, i.e.

$$(A \;\; -I)\begin{pmatrix} x \\ y \end{pmatrix} = b \in \mathbb{R}^m,$$

met $C = (A \;\; -I)$ een (wat bijzondere) matrix met m rijen en N kolommen, waarvan de eerste n kolommen de matrix A vormen, en de laatste m kolommen een diagonaal matrix met -1 op de diagonaal. Die matrix C werkt op kolomvectoren

$$z = \begin{pmatrix} x \\ y \end{pmatrix}$$

in $\mathbb{R}^{\mathrm{N}}$. Zo gelezen is (6.43) dus een voorbeeld van een stelsel van m lineaire vergelijkingen voor N onbekenden $z_1, z_2, \ldots, z_{\mathrm{N}}$.

Uitgeschreven met puntjes is dat stelsel

$$C_{11}z_1 + C_{12}z_2 + \cdots + C_{1N}z_{\mathrm{N}} = b_1;$$

$$C_{21}z_1 + C_{22}z_2 + \cdots + C_{2\mathrm{N}}z_{\mathrm{N}} = b_2;$$

$$\vdots$$

$$C_{m1}z_1 + C_{m2}z_2 + \cdots + C_{m\mathrm{N}}z_{\mathrm{N}} = b_m.$$

In het voorbeeld heeft de coëfficiëntenmatrix C maximale rang, hetgeen equivalent is met het kunnen kiezen van m kolommen van C die samen een inverteerbare (vierkante) matrix vormen, hier de laatste m kolommen, en is y door inverteren van in dit geval $-I$ in x uit te drukken. Dat is net als in het algemene geval, waarin $C = (A\ B)$ en B inverteerbaar is, en oplossen via $y = B^{-1}(b - Ax)$ weer een grafiek van de vorm (6.43) produceert. We zien hier dus de equivalentie van $Cz = b$ en $y = Ax + b$ als beschrijvingen in $\mathbb{R}^{\mathrm{N}}$ van de niet-triviale lineaire deelvarieteiten anders dan de singletons, waarbij C bij aanname maximale rang heeft.

6.8.1 Niet per se lineaire deelvarieteiten

Voor niet-lineaire deelvarieteiten[44] stellen we nu de vraag of er ook zo'n equivalentie is, uitgaande van de niet-lineaire versie van $Cz = b$, geschreven zoals in de impliciete functiestelling, dus[45]

$$F(z) = F(x, y) = 0,$$

met $F : \mathbb{R}^{\mathrm{N}} \to \mathbb{R}^m$ continu differentieerbaar. Omdat (6.43) een beschrijving is zonder verdere condities op A gebruiken we de niet-lineaire versie van (6.43) om af te spreken wat we onder een deelvarieteit $M \subset \mathbb{R}^{\mathrm{N}}$ willen verstaan: we zeggen dat een n-dimensionale deelvarieteit $M \subset \mathbb{R}^{\mathrm{N}}$ een verzameling is die in de buurt van elk punt van M na hernummeren van de coördinaten te schrijven is zoals de verzameling waar $F(x, y) = F(a, b)$ in Stelling 6.11, dus als de grafiek van een continu differentieerbare functie[46]. Hierbij is de dimensie $n \in \{1, \ldots, \mathrm{N} - 1\}$ vast gekozen en sluiten we het analagon van singletons uit[47].

[44] We moet nog een definitie geven van wat we daaronder verstaan!

[45] Er is een reden waarom we hier liever y achter x schrijven.

[46] Preciezer: M heet dan een C^1-deelvarieteit, en een hypersurface if $n = \mathrm{N} - 1$.

[47] Dat analogon zou een deelverzameling van geisoleerde punten zijn.

Opgave 6.14. Laat $F : \mathbb{R}^{\mathrm{N}} \to \mathbb{R}^m$ continu differentieerbaar zijn. Als voor alle $z \in \mathbb{R}^{\mathrm{N}}$ met $F(z) = 0$ geldt dat $F'(z)$, gezien als matrix, maximale rang heeft, dan is $\{z \in \mathbb{R}^{\mathrm{N}} : F(z) = 0\}$ een n-dimensionale deelvarieteit van $\mathbb{R}^{\mathrm{N}}$, met $n + m = \mathrm{N}$. Bewijs dit.

Opgave 6.15. Geef een voorbeeld van een n-dimensionale deelvarieteit $M \subset \mathbb{R}^{\mathrm{N}}$ die niet gegeven wordt door een functie F zoals in Opgave 6.14.

Een standaardvoorbeeld bij Opgave 6.14 is de rand van een bol in $\mathbb{R}^n$ met middelpunt $(a_1, a_2, \ldots, a_n)$ en straal $\delta > 0$, beschreven door

$$(x_1 - a_1)^2 + \cdots + (x_n - a_n)^2 - \delta^2 = 0. \tag{6.44}$$

Opgave 6.16. Bewijs dat de verzameling van $x \in \mathbb{R}^n$ die voldoen aan (6.44) een $(n-1)$-dimensionale deelvarieteit is.

Er zijn drie equivalente manieren om te zeggen dat $M \subset \mathbb{R}^{\mathrm{N}}$ een n-dimensionale deelvarieteit is

(A) M is lokaal de grafiek van een continu differentieerbare

$$f : \mathbb{R}^n \to \mathbb{R}^m \quad (n + m = \mathrm{N}),$$

gegeven door $y = f(x)$ na hernummeren in $z = (x, y)$.

(B) M is lokaal de nulverzameling van een continu differentieerbare

$$F : \mathbb{R}^{\mathrm{N}} \to \mathbb{R}^m \quad (n + m = \mathrm{N}),$$

F_y inverteerbaar in de betreffende hernummerde $z = (x, y) \in M$.

(C) M is lokaal het beeld[48] van een continu differentieerbare injectieve

$$\Phi : \mathbb{R}^n \to \mathbb{R}^{\mathrm{N}},$$

met Φ' van maximale rang.

Stelling 6.11 liet al (B) $\Longrightarrow$ (A) zien, en (A) $\Longrightarrow$ (B) geldt omdat (A) een speciaal geval is van B met $F(x, y) = g(y) - x$. Evenzo is (A) een speciaal

[48] Als afbeelding naar dat beeld heeft Φ een inverse die een kaart op M genoemd wordt.

geval van C met $\Phi(x) = (x, f(x))$. Om de implicatie (C) $\implies$ (A) te laten zien en zo het kringetje rond te maken gebruiken we Stelling 6.12, maar hebben we ook de kettingregel nodig, de regel die meestal vóór de impliciete functiestelling wordt behandeld.

De *kettingregel* generaliseert de regel in Opgave 5.10, zie ook Sectie 4.6, Opgave 5.11 en later Opgave 10.19. Met de opmerking dat (5.3) gelezen moet worden met matrices[49] is de regel met bewijs en al over te schrijven en nu meteen[50] toepasbaar. We zien Φ daartoe nu als

$$\Phi : \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^m,$$

met, na hernummeren weer, $\Phi(x) = (\Psi(x), \chi(x))$, $\Psi : \bar{B}_r(a) \to \mathbb{R}^n$ en $\chi : \bar{B}_r(a) \to \mathbb{R}^m$ continu differentieerbaar, en $\Psi'(a)$ inverteerbaar in a, bij aanname van de maximale rang van $\Phi'(x)$ in $x = a$. Stelling 6.12, toegepast op $g = \Psi$ met $y = x$, geeft nu een continu differentieerbare injectieve functie f die we ϕ noemen, $\phi : \bar{B}_{\delta_0}(\Psi(a)) \to \mathbb{R}^n$, met $\phi'(\xi)$ inverteerbaar voor alle $\xi \in \bar{B}_{\delta_0}(\Psi(a))$, al gebruiken we dat verder niet, en $\Psi(\phi(\xi)) = \xi$ voor alle $\xi \in \bar{B}_{\delta_0}(\Psi(a))$. Zo parametriseert

$$\xi \to \Phi(\phi(\xi)) = (\Psi(\phi(\xi)), \chi(\phi(\xi))) = (\xi, f(\xi))$$

met $f(\xi) = \chi(f(\xi))$ nu ook M in de buurt van $b = \Psi(a)$ en is M lokaal dus gegeven door de grafiek van $f : \bar{B}_{\delta_0}(b) \to \mathbb{R}^{\mathrm{N}}$. De continue differentieerbaarheid van f volgt uit de kettingregel, en het is de eerste keer dat we die hier gebruiken. Hiermee is het bewijs van

$$\text{(A)} \iff \text{(B)} \iff \text{(C)}$$

klaar.

Opgave 6.14 betrof het geval dat M met één F als in (B) globaal te beschrijven is zoals bijvoorbeeld in Opgave 6.16. In Stelling 6.12 en de daaropvolgend discussie werd zo'n bol met rand en al injectief afgebeeld door een injectieve continu differentieerbare functie f (met $f'(x)$ inverteerbaar voor alle x in die gesloten bol) naar een deelverzameling van (een bol in) $\mathbb{R}^n$. Is het beeld van de rand van de bol nu ook weer een $(n-1)$-dimensionale deelvarieteit?

Opgave 6.17. Laat $M \subset \mathbb{R}^n$ een deelvarieteit zijn en $f : \mathbb{R}^n \to \mathbb{R}^n$ continu differentieerbaar in een omgeving van elk punt van M. Als $f'(x)$ inverteerbaar is voor elke $x \in M$ en f injectief is op M, dan is het beeld van f ook een deelvarieteit.

[49] Beter: lineaire afbeeldingen, in dit hele hoofdstuk de facto matrices.

[50] Nou ja, aan het eind van het argument dat nu komt.

Opgave 6.18. Precies als Opgave 6.17, maar met $f : \mathbb{R}^n \to \mathbb{R}^m$ en $f'(x)$ van maximale rang voor elke $x \in M$.

6.8.2 Beelden (van randen) van bollen

Stelling 6.12 leidde met $0 < \varepsilon_1 \leq \delta_1 \leq \delta_0 \leq \varepsilon_0$ tot een keten

$$\bar{B}_{\varepsilon_1}(b) \xrightarrow{g} \bar{B}_{\delta_1}(a) \xrightarrow{f} \bar{B}_{\varepsilon_0}(b)$$

waarin beide afbeeldingen injectief maar in het algemeen niet surjectief zijn. De kleinere δ_1 en ε_1 waren nodig voor de injectiviteit van g op de kleinere (gesloten) bol $\bar{B}_{\varepsilon_1}(b)$.

De randen $\partial B_{\varepsilon_1}(b)$ en $\partial B_{\varepsilon_0}(a)$ hebben als beeld de deelvarieteiten $g(\partial B_{\varepsilon_1}(b))$ en $f(\partial B_{\varepsilon_0}(a))$. In het geval dat g en f lineaire afbeeldingen zijn en $a = b = 0$, ga je eenvoudig na dat deze beelden te zien zijn als grafiek over de eenheidssfeer

$$S^{n-1} = \{x \in \mathbb{R}^n : |x| = 1\}.$$

Als $A : \mathbb{R}^n \to \mathbb{R}^n$ zo'n inverteerbare lineaire afbeelding is, dan kun je een gladde hoogtefunctie $h : S^{n-1} \to \mathbb{R}^+$ maken zodanig dat het beeld van (de bolrand) $\partial B_1(0)$ onder A van de vorm

$$S_h = \{h(x)x : x \in S^{n-1}\} \tag{6.45}$$

is, al moeten we nog precies maken wat we hier met gladheid van h bedoelen. Je maakt de functie h door voor elke $x \in S^{n-1}$ de halfrechte

$$\{\lambda x : \lambda > 0\}$$

met $A(\partial B_1(0))$ te snijden, waarbij je x misschien liever een andere naam wil geven als je denkt in termen van $y = Ax$.

Opgave 6.19. Laat $A : \mathbb{R}^n \to \mathbb{R}^n$ een inverteerbare lineaire afbeelding zijn. Bewijs dat er voor elke $\xi \in S^{n-1}$ precies een $\lambda > 0$ is waarvoor $\lambda\xi \in A(\partial B_1(0))$. Door $\lambda = h_A(\xi)$ te stellen is $h_A : S^{n-1} \to \mathbb{R}^+$ gedefinieerd. Het beeld van $\partial B_\delta(0)$ onder A heeft als hoogtefunctie $\xi \to \delta h_A(\xi)$.

De eerdere vragen over $g(\partial B_{\varepsilon_1}(b))$ en $f(\partial B_{\varepsilon_0}(a))$ geven aanleiding tot de vraag of de uitspraak in Opgave 6.19 ook waar is voor het beeld van de rand van een voldoende klein gekozen bol $\bar{B}_{\delta_1}(0)$ onder het beeld van een continue differentieerbare functie $F : \bar{B}_\delta(0) :\to \mathbb{R}^n$ van de vorm

$$F(x) = Ax + R(x) \quad \text{met} \quad R(x) = o(|x|) \quad \text{voor} \quad |x| \to 0.$$

Op grond van Stelling 6.12 weten we dat F injectief is op een kleinere bol $\bar{B}_{\delta_0}(0)$ met $F'(x)$ invertbaar (niet alleen voor $x = 0$ maar ook) voor alle $x \in \bar{B}_{\delta_0}(0)$. De volgende opgave is een klein project waarin ook Stelling 6.11 gebruikt moet worden, nog uit te werken in een definitieve versie van dit basisboek.

Opgave 6.20. Bewijs de uitspraak in Opgave 6.19 voor het beeld $F(\partial B_{\delta_1}(0))$ van de rand van een voldoende kleine bol $\bar{B}_{\delta_1}(0)$. Laat zien dat de hoogtefunctie h die je maakt continu differentieerbaar is in de buurt van ieder punt op S^{n-1}, als functie van geschikt gekozen coördinaten in een omgeving van dat punt.

6.8.3 Overgang op andere coördinaten

Als een punt P in n-dimensionale deelvarieteit M van $\mathbb{R}^{\mathrm{N}}$ zowel in het beeld van een Φ als van een Ψ ligt zoals in beschrijving (C) in Sectie 6.8.1, zeg met $\Phi(\xi)$ en $\Psi(\eta)$, en $P = \Phi(0) = \Psi(0)$, met 0 een inwendig punt van de in (C) gebruikte domeinen van Φ en Ψ, dan wordt het verband tussen ξ en η in de buurt van 0 gegeven door uitspraken zoals in Stelling 6.12.

6.8.4 Meer dan continu differentieerbaar

Om uit te werken nog. Als $F : X \to Y$ continu differentieerbaar is, met bijvoorbeeld $X = \mathbb{R}^m$ en $Y = \mathbb{R}^n$, dan is $x \to F'(x)$ een continue functie (of afbeelding) van X naar

$$L(X,Y) = \{A : X \to Y : A \quad \text{is lineair}\},$$

waarin we als norm op $L(X,Y)$ naar keuze de normen

$$|A|_{op} = \max_{|x|\leq 1} |Ax| \quad \text{of} \quad |A|_2 = \sqrt{\sum_{ij} a_{ij}^2}$$

kunnen gebruiken. Deze F' kan weer continu differentieerbaar zijn en in dat geval is ook

$$\tilde{F} : (x,h) \to (F(x), F'(x)h)$$

continu differentieerbaar. In het geval dat $X = Y$ is inverteerbaarheid van F' in $x = x_0$ equivalent met inverteerbaarheid van $\tilde{F}'$ in $(x,h) = (x_0, 0)$. Dit leidt tot een formulering van Stelling 6.12 voor twee keer continu differentieerbare functies. Ook de impliciete functiestelling krijgt zo een twee keer differentieerbare versie, en deelvarieteiten kunnen vervolgens C^2 zijn in plaats van C^1. Of C^k met $k \in \mathbb{N}$.

6.9 Integralen over rechthoeken

Integraalrekening voor continue functies

$$u : [a,b] \times [c,d] \to \mathbb{R},$$

met $[a,b] \times [c,d]$ een rechthoek in $\mathbb{R}^2$ gaat met partities P als in (2.4) voor $[a,b]$, en

$$c = y_0 \leq y_1 \leq \cdots \leq y_{\mathrm{M}} = b \tag{6.46}$$

voor $[c,d]$.

Om kort te gaan, met onder- en bovensommen, of met tussensommen van de vorm

$$\sum_{k=1}^{\mathrm{N}} \sum_{l=1}^{\mathrm{M}} u(\xi_k, \eta_l)(x_k - x_{k-1})(y_l - y_{l-1}),$$

met tussenwaarden $\xi_k \in [x_{k-1}, x_k]$ en $\eta_l \in [y_{l-1}, y_l]$, volgt het bestaan van één unieke $J \in \mathbb{R}$ die de integraal van u over $[a,b] \times [c,d]$ genoemd wordt, met notatie (en stelling over het vrij mogen kiezen van de integratievolgorde)

$$J = \int_{[a,b]\times[c,d]} u = \int_a^b \int_c^d u(x,y)\, dy\, dx = \int_c^d \int_a^b u(x,y)\, dx\, dy. \tag{6.47}$$

De herhaalde integralen worden uitgewerkt met de technieken en regels voor gewone integralen.

We maken daar hier nu verder geen woorden aan vuil maar wijzen wel nog even op de details in de hier gebruikte notaties in relatie tot de discussie na Stelling 4.12 en Opgave 2.48 onder Definitie 2.42.

6.9.1 Notationele kwesties met die d-tjes

Anders dan de notatie suggereert is (6.47) geen uitdrukking met de onder Stelling 4.12 aangekondigde 2-vormen. In (6.47) is de herhaalde integraal immers te lezen als

$$\int_c^d \left\{ \int_a^b u(x,y)\, dx \right\} dy,$$

een integraal van een integraal, net als de herhaalde integraal direct daarvoor, met de afsluitende accolade stevig tussen dx en dy in. Beide herhaalde integralen zijn gelijk aan

$$\int_{[a,b]\times[c,d]} u,$$

gedefinieerd via Riemannsommen en een limietovergang, met een notatie waarin $dxdy$ of $dydx$ als produkt van dx en dy vermeden is.

Opgave 6.21. Net als in Opgave 2.48 kun je je nu drie assen (nu in de ruimte in plaats van in het vlak) voorstellen die loodrecht op elkaar staan in hun gemeenschappelijke snijpunt $x = y = z = 0$, met richtingen nog te kiezen, maar dat pas nadat je op de x-as a en b gemarkeerd hebt, op de y-as c en d, en de grafiek van zo maar een continue functie zonder nulpunten en tekenwisselingen hebt gekozen. Nu zijn er acht gevallen waarin je de integraal kunt relateren aan de inhoud ingesloten door $x = a$, $x = b$, $y = c$, $y = d$, $z = 0$ en $z = u(x, y)$. Kijk nu naar de tekens van $b - a$ en $d - c$, zeg maar van dx en dy, in de twee gevallen dat zowel u als de integraal positief zijn, en laat zien dat ze hetzelfde teken hebben: zeg maar $dxdy > 0$.

Opgave 6.22. Neem nu eventueel voor het gemak $a = c = 0$, $|b| = |d| = 1$. In het geval dat $dxdy$ en $u(x, y)$ allebei positief zijn is er een orienterende ordening tussen de drie assen die overeenkomt met de correpondentie

positieve x-as $\leftrightarrow$ middelvinger
positieve y-as $\leftrightarrow$ wijsvinger
positieve z-as $\leftrightarrow$ duim

op je linkerhand. Overtuig jezelf daarvan en ook van het volgende: teken je een x-as en een y-as op het schoolbord die elkaar loodrecht snijden in $x = y = 0$ dan vertellen de positieve x-as en de positieve y-as als geordend assenpaar welke kant van het bord $z > 0$ heeft; bij de standaardkeuze met x naar rechts en y naarboven komt de positieve z-as naar voren het bord uit. Het bijbehorende assenkruis heet dan rechtsdraaiend[51].

Als je wil rekenen met formules waarin produkten van dx en dy voorkomen dan kun je in het licht van Opgave 6.21 en Opgave 6.22 tot de conclusie komen dat het handig is om af te spreken dat

$$dxdy = -dydx,$$

zonder de afsluitende accolade tussen dx en dy, in de zin dat

$$J = \int_{[a,b]\times[c,d]} u = \iint_{[a,b]\times[c,d]} u(x, y)\, dxdy,$$

gezien als

$$\iint_{[a,b]\times[c,d]} \text{ werkend op } \quad u(x, y)\, dxdy,$$

gelijk is aan

$$-\iint_{[a,b]\times[c,d]} u(x, y)\, dydx,$$

[51] Dit is moeilijk voor jongens.

gezien als

$$-\iint_{[a,b]\times[c,d]} \quad \text{werkend op} \quad u(x,y)\,dydx,$$

waarbij die werking dan precies gedefinieerd moet worden op zo'n manier dat onder coördinatentransformaties de integralen transformeren op een manier die consistent is met wat er gebeurt met de inhouden waar het verhaal mee begonnen is. Maar dat hebben we hier nog niet behandeld.

6.9.2 Formele d-algebra zonder betekenis?

Ondertussen spreekt de algebra voor differentiaalvormen deels voor zich. Na $du = u'(x)dx$ voor $u = u(x)$ hebben we natuurlijk ook

$$du = u_x dx + u_y dy = \frac{\partial u}{\partial x}dx + \frac{\partial u}{\partial y}dy \tag{6.48}$$

als de d van de 0-vorm $u = u(x, y)$, een uitdrukking van de vorm

$$fdx + gdy = f(x, y)dx + g(x, y)dy,$$

een 1-vorm waarvan de d via

$$d(f(x, y)dx + g(x, y)dy)$$

vraagt[52] om een som en produktregel met d van $f(x, y)$ en $g(x, y)$, en een nog te definiëren d van dx en dy.

Voor de hand als som- en daarna twee keer de produktregel ligt

$$d(fdx + gdy) = d(fdx) + d(gdy) =$$

$$dfdx + fddx + dgdy + gddy =$$

$$f_x dxdx + f_y dydx + g_x dxdy + g_y dydy + fddx + gddy.$$

Als $dxdy = -dydx$ volgt dan

$$d(fdx + gdy) = f_x dxdx + g_y dydy + (g_x - f_y)dxdy + fddx + gddy.$$

Met $y = x$ leest $dxdy = -dydx$ nu toch echt als $dxdx = -dxdx$, dus $dxdx = 0 = dydy$ als consistent voor te schrijven tweede rekenregel ligt dan ook voor de hand. Daarmee reduceert het bovenstaande tot

$$d(fdx + gdy) = (g_x - f_y)dxdy + fddx + gddy,$$

[52] Wat d is dat kun je straks weten.

en blijft de vraag wat ddx en ddy dan wel mogen zijn.

Keine blasse Ahnung zou ik zeggen dus doe dan maar NUL. Dan wordt de werking van d op 1-vormen

$$f(x,y)dx + g(x,y)dy \xrightarrow{d} (g_x(x,y) - f_y(x,y))dxdy,$$

en dus

$$u(x,y) \xrightarrow{d} u_x(x,y)dx + u_y(x,y)dy \xrightarrow{d} (u_{yx}(x,y) - u_{xy}(x,y))dxdy = 0.$$

Der blasse Ansatz $ddx = ddy = 0$ gibt $ddu = 0$ voor alle $u = u(x,y)$.

Ik zou zeggen: machen wir. Dus $d^2 = 0$[53], en met $f = f(x,y), g = g(x,y)$ hebben we in de notatie met partiële differentiaalquotiënten nu dat

$$f \xrightarrow{d} \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy, \quad fdx + gdy \xrightarrow{d} (\frac{\partial g}{\partial x} - \frac{\partial f}{\partial y})dxdy, \quad gdxdy \xrightarrow{d} 0 \quad (6.49)$$

Opgave 6.23. Doe de algebra voor

$$f \xrightarrow{d} \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial z}dz,$$

$$fdx + gdy + hdz \xrightarrow{d} (\frac{\partial h}{\partial y} - \frac{\partial g}{\partial z})dydz + (\frac{\partial f}{\partial z} - \frac{\partial h}{\partial x})dzdx + (\frac{\partial g}{\partial x} - \frac{\partial f}{\partial y})dxdy,$$

$$fdydz + gdzdx + hdxdy \xrightarrow{d} (\frac{\partial f}{\partial x} + \frac{\partial g}{\partial y} + \frac{\partial h}{\partial z})dxdydz,$$

$$hdxdydz \xrightarrow{d} 0,$$

met $f = f(x,y,z), g = g(x,y,z), h = h(x,y,z)$, gebruikmakend van de som- en produktregel voor d, $dxdy = -dydx, dxdz = -dzdx, dydz = -dzdy$, en $ddx = ddy = ddz = 0$. Verifieer dat niet alleen $ddf = 0$ maar ook $dd(fdx + gdy + hdz) = 0$: dd maakt alles nul en zo hoort dat kennelijk.

Opgave 6.24. Doe de algebra nog een keer voor $F \xrightarrow{d} F'(x)dx$ en $f(x)dx \xrightarrow{d} 0$ met $f(x)$ en $F(x)$.

[53] Daar kan i een puntje aan zuigen.

6.9.3 Transformaties en parametrisaties

Als $x \to f(x) = F'(x)$ en $t \to x'(t)$ continu zijn met bijvoorbeeld $x(0) = a$ en $x(1) = b$, dan is

$$\int_a^b f(x)\,dx = \int_a^b F'(x)\,dx = \int_a^b dF = F(b) - F(a) = F(x(1)) - F(x(0)) =$$

$$[F(x(t))]_0^1 = \int_0^1 F'(x(t))x'(t)\,dt = \int_0^1 f(x(t))x'(t)\,dt,$$

waarin we

$$dx = x'(t)dt \tag{6.50}$$

als een stukje van de formele d-algebra in Sectie 6.9.2 herkennen.

Een 1-vorm $f(x)dx$ in x wordt via $t \to x(t)$ zo teruggetrokken tot een 1-vorm

$$f(x)dx = f(x(t))x'(t)dt \tag{6.51}$$

in t. Evenzo wordt de 1-vorm

$$f(x,y)dx + g(x,y)dy$$

via $t \to x(t)$ en $t \to y(t)$ met dezelfde algebra teruggetrokken tot een 1-vorm

$$f(x,y)dx + g(x,y)dy = (f(x(t),y(t))x'(t) + g(x(t),y(t))y'(t))dt \tag{6.52}$$

in t.

De gelijktekens in (6.51) en zeker (6.52) zijn wat dubieus, het is wellicht beter om (6.51) en (6.52) nu als afbeeldingen te lezen, bijvoorbeeld

$$f(x,y)dx + g(x,y)dy \xrightarrow{terugtrekken} (f(x(t),y(t))x'(t) + g(x(t),y(t))y'(t))dt,$$

waarbij het terugtrekken gebeurd via de afbeelding in kwestie, in dit voorbeeld $t \to (x(t),y(t))$, die van 1 naar 2 variabelen gaat.

Nemen we in plaats van een 1-vorm een 2-vorm, zeg $dxdy$ zelf, dan geeft terugtrekken via dezelfde $t \to (x(t),y(t))$ dat

$$dxdy \to x'(t)dt\,y'(t)dt = x'(t)y'(t)dtdt = 0.$$

Weinig zinvol wellicht, maar met

$$(r,\theta) \to (x(r,\theta),y(r,\theta))$$

krijgen we

$$dxdy \to (\frac{\partial x}{\partial r}dr + \frac{\partial x}{\partial \theta}d\theta)(\frac{\partial y}{\partial r}dr + \frac{\partial y}{\partial \theta}d\theta) = (\frac{\partial x}{\partial r}\frac{\partial y}{\partial \theta} - \frac{\partial y}{\partial r}\frac{\partial x}{\partial \theta})drd\theta, \quad (6.53)$$

met de determinant[54] van de Jacobimatrix.

In het geval dat $x = r\cos\theta, y = r\sin\theta$ wordt dit

$$dxdy \to rdrd\theta,$$

op grond van een algebra die zich (ook) voor coördinatentransformaties genereert uit de regeltjes die we in de d-algebra hebben geponeerd.

Merk op dat (6.52) niet met een coördinatentransformatie kan corresponderen maar (6.53) wel, en lees de afleiding van (6.52) nog eens terug met t vervangen door ϕ in bijvoorbeeld $[0, 2\pi]$. Met $x(0) = x(2\pi)$ en $y(0) = y(2\pi)$ kun je dit voorbeeld nu goed vergelijken met (6.53) waarin r vast is genomen, alvorens verder te ontdekken hoe de terugtrekalgebra werkt voor

$$(\theta, \phi) \to (x(\theta, \phi), y(\theta, \phi), z(\theta, \phi) \quad (6.54)$$

en 2-vormen in x, y, z.

Opgave 6.25. Trek $f(x, y, z)dx + g(x, y, z)dy + h(x, y, z)dz$ terug naar een 2-vorm in θ en ϕ.

6.9.4 Een transformatiestelling

Voor $R \subset \mathbb{R}^2$ en $f : R \to \mathbb{R}$ geldt dat

$$\iint_R f(x, y)\, dxdy = \iint_Q f(x(r, \theta), y(r, \theta))\, (\frac{\partial x}{\partial r}\frac{\partial y}{\partial \theta} - \frac{\partial y}{\partial r}\frac{\partial x}{\partial \theta})drd\theta,$$

als

$$Q \xrightarrow{(r,\theta)\to(x,y)} R$$

aan wat aannamen voldoet. Voorbeeld: Sectie 9.1. Hoe doen we dit? Onderstaande is om de gedachten te bepalen.

Met een bijectie

$$(x, y) \xrightarrow{\Phi} (u, v)$$

[54] Plus of min de oppervlakte van de ruit opgespannen door de twee kolomvectoren.

tussen $R \subset \mathbb{R}^2_{x,y}$ en $A \subset \mathbb{R}^2_{u,v}$ is wat we willen een uitspraak die de integraal

$$\iint_A g(u,v)\,dudv$$

relateert aan een integraal met $g(u(x,y),v(x,y))$ en $dxdy$ over R, liefst meteen al met de conventie dat $dudv = -dvdu$ en $dxdy = -dydx$.

Als de bijectie niet-lineair is dan zijn R en A zeker niet allebei een rechthoek, laat staan van de vorm zoals in Sectie 6.9, dus neem in eerste instantie alleen R van die vorm. We nemen R gesloten, dus met de rand erbij, en Φ gedefinieerd[55] op heel $\mathbb{R}^2$, met continue partiële afgeleiden. Voor het gemak nemen we $R = [0,1] \times [0,1]$.

Zie nu Opgave 4.25 en lees i.p.v

$$F(x,y) = g(y) - x \quad \text{nu} \quad F(x,y,u,v) = \begin{pmatrix} \Phi_1(x,y) - u \\ \Phi_2(x,y) - v \end{pmatrix}.$$

De uitgepakte versie van de stelling die je daar bewezen hebt is de inverse functiestelling die we nu nodig hebben. Die stelling zegt dat als in (x_0, y_0) de Jacobimatrix

$$J(x,y) = \begin{pmatrix} \frac{\partial \Phi_1}{\partial x} & \frac{\partial \Phi_1}{\partial y} \\ \frac{\partial \Phi_2}{\partial x} & \frac{\partial \Phi_2}{\partial y} \end{pmatrix}$$

inverteerbaar is, in een omgeving van $(u_0, v_0) = (\Phi_1(x_0, y_0), \Phi_2(x_0, y_0))$ de inverse functie

$$(u,v) \xrightarrow{\Phi^{(-1)}} (x,y)$$

bestaat en continu differentieerbaar is. De Jacobimatrix van de inverse afbeelding is de inverse van de Jacobimatrix van Φ.

Voor de te formuleren transformatiestelling nemen we nu aan dat de Jacobimatrix $J(x,y)$ in elk punt van R inverteerbaar is. Daarmee is A een gebied in $\mathbb{R}^2_{u,v}$ met vier randen geparametriseerd[56] door

$$x \to \Phi(x,0), \quad y \to \Phi(1,y), \quad x \to \Phi(x,1), \quad y \to \Phi(0,y).$$

Bij partities

$$(P) \quad 0 = x_0 \le x_1 \le \cdots \le x_{\mathrm{N}} = 1 \quad \text{met} \quad \mathrm{N} \in \mathbb{N},$$

[55] Zonder beperking der algemeenheid, maar dat moet nog blijken.

[56] Het alternatief dat later nog komt wellicht is:

$$t \to \Phi(t,0), \quad t \to \Phi(1,t), \quad t \to \Phi(1-t,1), \quad t \to \Phi(1-t,0),$$

$$(Q) \quad 0 = y_0 \leq y_1 \leq \cdots \leq y_{\mathrm{M}} = 1 \quad \text{met} \quad \mathrm{M} \in \mathbb{N},$$

horen $(\mathrm{M}+1)(\mathrm{N}+1)$ parametrisaties

$$x \to \Phi(x, y_j) \quad \text{en} \quad y \to \Phi(x_i, y) \quad (i = 0, \ldots, \mathrm{M},\, j = 0 \ldots, \mathrm{N}),$$

die een rooster van vervormde rechthoekjes S_{ij} maken in A.

De goede definitie van Riemann integreerbaarheid van $g : A \to \mathbb{R}$ moet[57] nu geven dat met

$$M_{ij} = \sup_{S_{ij}} g \quad \text{en} \quad m_{ij} = \inf_{S_{ij}} g$$

geldt dat

$$\sum_{ij} m_{ij}|S_{ij}| \leq \iint_A g \leq \sum_{ij} M_{ij}|S_{ij}|,$$

waarin S_{ij} de oppervlakte is van S_{ij}, hetgeen we herschrijven als

$$\sum_{ij} m_{ij}\frac{|S_{ij}|}{|R_{ij}|}|R_{ij}| \leq \iint_A g \leq \sum_{ij} M_{ij}\frac{|S_{ij}|}{|R_{ij}|}|R_{ij}|,$$

en opmerken dat ook geldt dat

$$M_{ij} = \sup_{R_{ij}} f \quad \text{en} \quad m_{ij} = \inf_{R_{ij}} f$$

met $f = g \circ \Phi$.

Het is niet zo moeilijk[58] om precies te maken dat

$$\frac{|S_{ij}|}{|R_{ij}|} \sim |\mathrm{det}J(x_i, y_i)| \tag{6.55}$$

als $\mathrm{M}, \mathrm{N} \to \infty$ om op de Riemann integreerbaarheid van

$$(x, y) \to f(x, y)|J(x, y)|$$

over R uit te komen MET gelijkheid van de integralen:

$$\iint_R f|\mathrm{det}J| = \iint_A g.$$

[57] Moet nog blijken!

[58] Zie e.g. Sectie 5 van Hoofdstuk III in het Advanced Calculus boek van Edwards.

7 Standing at the crossroads

This chapter relates to the courses in Functional Analysis and Partial Differential Eqations as given in the bachelor pogrammes in Amsterdam, as well as courses run under the same name in the national mastermath programme. Have a look at Section 4.10, in particular at the solution method for (4.48), which clearly does not generalise to the problem of solving

$$-\Delta u = f \quad \text{in} \quad \Omega \quad \text{with} \quad u = 0 \quad \text{on} \quad \partial\Omega \tag{7.1}$$

for given $f : \Omega \to \mathbb{R}$ and $\Omega \subset \mathbb{R}^N$ on a bounded open set with boundary $\partial\Omega$.

Moreover, even is $\partial\Omega$ is smooth, it does not in general hold that (7.1) has a twice differentiable solution which solves the partial differential equation $-\Delta u = f$ for the given $f \in C(\bar{\Omega})$. Below we show another way of solving (4.48) which does generalise to a large class of problems including (7.1). This technique is based on integration by parts[1] and the theory of Hilbert spaces, mainly the unique and obvious generalisation[2] of the inner product space $\mathbb{R}^2$ with the dimension 2 replaced by the first cardinal to come over all the integers, see de Hoofdstukken 13 en 14. The chapters thereafter will hopefully fit in some future version of this book.

Als we de vergelijking in (4.48) vermenigvuldigen met een functie $v \in C^1[0,1]$ dan bestaan onder de aanname dat $u \in C^2[0,1]$ en $f \in C^0[0,1]$ beide integralen

$$-\int_0^1 u''(x)v(x)\,dx = \int_0^1 f(x)v(x)\,dx$$

en kan de linkerkant partieel geintegreerd worden. Het resultaat is

$$-[u'(x)v(x)]_0^1 + \underbrace{\int_0^1 u'(x)v'(x)\,dx}_{\text{symmetrisch in } u,v} = \underbrace{\int_0^1 f(x)v(x)\,dx}_{\text{symmetrisch in } f,v}, \tag{7.2}$$

waarbij de lelijke eerste term verdwijnt als we de extra aanname maken dat $v(0) = v(1) = 0$.

De oplossing $u \in C^2[0,1]$ van (4.48) heeft dus de eigenschap dat $u(0) = u(1) = 0$ en

$$\int_0^1 u'(x)v'(x)\,dx = \int_0^1 f(x)v(x)\,dx \quad \forall v \in C_0^1[0,1], \tag{7.3}$$

waarin

$$C_0^1[0,1] = \{v \in C^1[0,1] : v(0) = v(1) = 0\},$$

[1] Now have a look at (8.11).

[2] In different guises.

en de gelijkheid in (7.3) kan voor elke $u \in C_0^1[0,1]$ geverifieerd worden. Kortom, we zouden dus kunnen afspreken om $u \in C_0^1[0,1]$ een oplossing van (4.48) te noemen als aan (7.3) voldaan is.

Rechts in (7.3) zien we een integraal die te zien is als een inwendig product van f en v, dat we kunnen noteren als

$$f \cdot v = \int_0^1 f(x)v(x)\,dx, \tag{7.4}$$

waarmee (4.48) zich uiteindelijk herschrijft als

$$u \in C_0^1[0,1], \quad \underbrace{u' \cdot v'}_{((u,v))} = \underbrace{f \cdot v}_{(f,v)} \quad \forall v \in C_0^1[0,1], \tag{7.5}$$

een uitdrukking waarin twee inwendige producten voorkomen en $u \in C_0^1[0,1]$ en $f \in C^0[0,1]$ vast zijn, en $v \in C_0^1[0,1]$ willekeurig.

Opgave 7.1. Waarom is $(u,v) \to u' \cdot v'$ wel een inprodukt op $C_0^1[0,1]$ en niet op $C^1[0,1]$?

Opgave 7.2. Laat $L > 0$, bijvoorbeeld $L = 2\pi$ of $L = 1$. De vectorruimte van continue L-periodieke functies noemen we $C(\mathbb{R}_L) = C^0(\mathbb{R}_L)$, en de deelruimte van k keer ($k \in \mathbb{N}$) continu differentieerbare functies noemen we $C^k(\mathbb{R}_L)$. Voor welke $f \in C(\mathbb{R}_L)$ is de vergelijking $-u'' = f$ oplosbaar met u in $C^2(\mathbb{R}_L)$? Is de oplossing uniek? Hint: neem eerst $L = 1$ natuurlijk.

Opgave 7.3. Neem $L = 1$ en los de vergelijking $-u'' = f$ voor f continu, 1-periodiek met $\int_0^1 f(x)\,dx = 0$: geef een uitdrukking van de vorm (4.51) voor de oplossing u die (ook) voldoet aan $\int_0^1 u(x)\,dx = 0$. Hint: zonder deze laatste conditie is de oplossing niet uniek bepaald en evenzo is de (wederom symmetrische!) kern $A(x,s)$ niet uniek bepaald. Maar wel onder de conditie dat $\int_0^1 A(x,s)\,dx = \int_0^1 A(x,s)\,ds = 0$.

Opgave 7.4. Laat

$$\bar{C}^k(\mathbb{R}_L) = \{u \in C^k(\mathbb{R}_L) : \int_0^L u(x)dx = 0\}$$

en geef een herformulering van $-u'' = f$ voor $f \in \bar{C}^0(\mathbb{R}_L)$ en $u \in \bar{C}^1(\mathbb{R}_L)$ zoals in (7.5).

Hopelijk is duidelijk dat de twee laatste opgaven kwa bewerkelijkheid nogal uiteenliepen. Formuleringen als in (7.5) gebaseerd op *integration by parts*, zonder dat het doel daarvan het uitrekenen van getallen is, bieden een ander en vaak algemener perspectief om eigenschappen van oplossingsoperatoren te begrijpen dan expliciete oplossingsmethoden gebaseerd op primitiveren. In wat volgt zullen we daartoe v als een variabele zien en

$$v \to ((u, v)) \quad \text{en} \quad f \to (f, v)$$

als dezelfde lineaire functie, maar anders gepresenteerd.

Lineaire functies en inprodukten, hoe zit dat? Hoe weet je dat er bij f via dezelfde lineaire functie van v een u hoort? En kan dat algemener? Daarvoor gaan we in Hoofstuk 13 terug naar de makkelijkste niet-flauwe inproductruimte die we kennen, het platte vlak. Dit hoofdstuk besluiten we met een paar opgaven die laten zien hoe intrinsiek de herformulering als (7.5) verbonden is met de eigenschappen van de oplossingsoperator

$$A : f \xrightarrow{\forall v((u,v))=(f,v)} u, \tag{7.6}$$

waarbij het van het specifieke probleem afhangt welke inproducten en welke functieruimten gedefinieerd moeten worden om een A te maken die in simpele gevallen samenvalt met expliciet uitgerekende integraaloperatoren als in (4.51).

Opgave 7.5. Laat zien dat een solver als A in (7.6) de eigenschap heeft dat $((Au, v)) = ((u, Av))$ en $(Af, g) = (f, Ag)$ voor alle u, v en f, g in de nog te kiezen ruimten $V \subset H$ en H waarop de inproducten zijn gedefinieerd met de eigenschappen die we nodig hebben om alles precies te maken.

De operator A is dus symmetrisch[3] met betrekking tot twee inproducten, waaronder het 'gewone' inproduct (7.4) dat in eerste instantie was opgeschreven onder verschillende aannames voor f en v.

Opgave 7.6. Gebruik de symmetrie van A om te laten zien dat eigenvectoren[4] van A bij verschillende eigenwaarden van A loodrecht op elkaar staan.

[3] We praten nog niet over complexwaardige functies hier.

[4] $A\phi = \lambda\phi$, $\lambda \in \mathbb{R}$, ϕ een (eigen)functie.

Opgave 7.7. Gebruik de vorige opgave en Opgave 7.4 om zonder rekenwerk te laten zien dat

$$\int_{-\pi}^{\pi} \sin nx \sin mx\, dx = 0 = \int_{-\pi}^{\pi} \cos nx \cos mx\, dx \quad (m, n \in \mathbb{N},\ m \neq 0)$$

$$\int_{-\pi}^{\pi} \sin nx \cos mx\, dx = 0 \quad (m, n \in \mathbb{N})$$

Natuurlijk wist je dit al, waarschijnlijk via $\exp(inx) = \cos nx + i \sin nx$ en de gebruikelijke rekenregeltjes gebaseerd op de somformules[5] voor $\cos(a+b)$ en $\sin(a+b)$ die niet meer tot de tegenwoordig zelden precies gerechtvaardigde basiskennis van de gemiddelde β-student horen. De functie sin kan gedefinieerd worden als de unieke oplossing van het beginwaardeprobleem

$$u'' + u = 0; \quad u(0) = 0; \quad u'(0) = 1, \tag{7.7}$$

en cos als de afgeleide van sin. Alle eigenschappen van cos en sin, i.h.b. de somformules volgen uit deze definities en kunnen gebruikt worden voor de volgende opgave.

Opgave 7.8. Bepaal alle $\lambda > 0$ waarvoor $u'' + \lambda u = 0$ oplossingen van periode 2π heeft en bepaal alle even en oneven oplossingen voor die waarden λ.

De even oplossingen die je zo vindt zijn veelvouden van $c_1 : x \to \cos x$, $c_2 : x \to \cos 2x$, $c_3 : x \to \cos 3x$, ..., en de oneven oplossingen zijn veelvouden van $s_1 : x \to \sin x$, $s_2 : x \to \sin 2x$, $s_3 : x \to \sin 3x$, ..., en ieder tweetal van deze functies staat loodrecht op elkaar, zoals we in Opgave 7.7 gezien hebben. En ze zijn gemiddeld allemaal nul, hetgeen betekent dat ze loodrecht staan op de functie $\mathbf{1} : x \to 1$, bijvoorbeeld

$$(\mathbf{1}, s_1) = \mathbf{1} \cdot s_1 = \int_{-\pi}^{\pi} 1 \sin x\, dx = \int_{-\pi}^{\pi} \sin x\, dx = 0.$$

Wat we in vervolg gaan doen is $1, c_1, c_2, c_3, \ldots, s_1, s_2, s_3, \ldots$ zien als vectoren die lijnen door de oorsprong[6] definiëren. En die lijnen zien we als een assenkruis waarmee we een oneindig-dimensionale ruimte opspannen, een ruimte waarin we willen werken zoveel mogelijk als we dat in het platte vlak doen.

[5] Zie Wiskunde in je Vingers, sectie 10.4.

[6] Die oorsprong is de nulfunctie $\mathbf{0} : x \to 0$.

8 Multilinear algebra and integration

This chapter is at the crossroads of analysis and algebra[1]. Let us first state the main analytical result, which concerns a bounded open set $\Omega \subset \mathbb{R}^N$ with $\partial\Omega \in C^1$ and $v \in C^1(\bar{\Omega})$. We will establish that

$$\int_\Omega v_{x_i} = \int_{\partial\Omega} \nu_i \, v. \tag{8.1}$$

The integral on the left in (8.1) is a Riemann integral of the continuous partial derivative

$$v_{x_i} = \frac{\partial v}{\partial x_i} = D_i v = \partial_i v$$

of v with respect to the i^{th} variable. It is defined if $\partial\Omega$ is a set with zero N-dimensional measure, see Opmerking 2.1. This is the case if the bounded closed set $\partial\Omega$ is the zero level set of a function $F \in C^1(\mathbb{R}^N)$ with $\nabla F(x) \neq 0$ for every[2]

$$x \in \{x \in \mathbb{R}^N : \, F(x) = 0\} = \partial\Omega.$$

If the bounded open set Ω is given by

$$\Omega = \{x \in \mathbb{R}^N : \, F(x) < 0\}, \tag{8.2}$$

the vector

$$\nu = \frac{\nabla F}{|\nabla F|_2}$$

is defined in every point of $\partial\Omega$, and thereby we have a unit vector field on $\partial\Omega$ normal[3] to $\partial\Omega$, which is then a compact C^1-hypersurface[4] in $\mathbb{R}^N$, and the integral on the right in (8.1) is the integral of the continuous function

$$x \to \nu_j(x) \, v(x)$$

over $\partial\Omega$, as defined for continuous functions on compact n-dimensional manifolds $M \subset \mathbb{R}^N$ such as $M = \partial\Omega$ in Section 8.9 below.

This analytical result is later restated for continuously differentiable vector valued functions

$$\Omega \ni x \xrightarrow{V} \mathbb{R}^N$$

at the beginning of Section 8.13 as the Gauss Divergence Theorem in (8.68), while it is first proved in Section 8.1 for functions v that vanish outside an N-dimensional block $[a, b]$ in which $\partial\Omega \cap [a, b]$ is the graph of a C^1-function as explained at the beginning of that section, with $\Omega \cap [a, b]$ on one[5] side of the

[1] Written while teaching from Edwards' book Advanced Calculus of Several Variables.

[2] Zero is then called a regular value of F, google Sard's Theorem.

[3] To the local tangent space in every point of.

[4] See Opgave 6.14.

[5] Which easily leads to mistakes in the direction of ν so be careful.

graph. In the 2-dimensional case N = 2 this formulation avoids the Jordan Curve Theorem[6].

The local statement under (8.42) becomes a global statement dropping the tildes and skipping the intermediate term. It then reads

$$\int_{\partial\Omega} \nu \cdot v \; dS_1 = \iint_{\Omega} \left(\frac{\partial v_1}{\partial x_1} + \frac{\partial v_2}{\partial x_2} \right) dx_1 \, dx_2, \tag{8.3}$$

which putting $v_2 = -p$ and $v_1 = q$ rewrites as (8.43), the 2-dimensional of the Stokes Curl Theorem, in which the line integral is the integral over $\partial\Omega$ of the inner product of the vectorfield with components p and q and the tangent[7] vectorfield τ with components $-\nu_2$ and ν_1. Renaming p and q as v_1 and v_2 the result

$$\begin{aligned} \int_{\partial\Omega} \tau \cdot v \; dS_1 &= \iint_{\Omega} \left(\frac{\partial v_2}{\partial x_1} - \frac{\partial v_1}{\partial x_2} \right) dx_1 \, dx_2 \\ &= \int_{\partial\Omega} v_1(x_1, x_2) dx_1 + v_2(x_1, x_2) dx_2, \end{aligned} \tag{8.4}$$

is known as Green's Theorem, the second formulation being valid if $\partial\Omega$ is parameterised by continuously differentiable parameterisations

$$t \to x(t) = (x_1(t), x_2(t))$$

defining normalised tangent and normal vectors

$$\tau(t) = \frac{1}{\sqrt{x_1'(t)^2 + x_2'(t)^2}} \begin{pmatrix} x_1'(t) \\ x_2'(t) \end{pmatrix} \quad \text{and} \quad \nu(t) = \frac{1}{\sqrt{x_1'(t)^2 + x_2'(t)^2}} \begin{pmatrix} x_2'(t) \\ -x_1'(t) \end{pmatrix}$$

with ∇F in $x(t)$ being a positive multiple of $\nu(t)$ if Ω is defined by (8.2).

The local statement in Section 8.1 does indeed lead to the global statement via arguments which involve cut-off functions[8] and partitions of unity, which are discussed as an independent topic in Section 8.15. These are also needed for a proper definition of integrals over manifolds in Section 8.9, the other analytical tool being the transformation theorem discussed in Section 6.9.4 applied in Section 8.17 to the transformations in Section 8.19.

So far for analysis. In Section 8.10 we rewrite (8.1) as stated in (8.10) in the language of the differential forms introduced in Section 6.9.2 and below formula (8.4) just above. This language is re-presented in the language also used by Edwards in Section 8.12. In Section 8.13 we spell out the algebra

[6] Which I should discuss in the context of Section 9.6 and polygons, to do.

[7] Edwards writes $\nu = \mathbf{N}$ and $\tau = \mathbf{T}$.

[8] A misnomer, I would call them fading functions.

to explain how the general Stokes Theorem for integrals of differential forms then follows[9]. This general theorem then has the Stokes Curl Theorem in (8.74) for certain[10] closed closed curves in $\mathbb{R}^3$ as a reformulation without differential forms as a second most commonly occuring example.

8.1 Local integrals with the normal at the boundary

This section is influenced by all the local boundary flattening arguments in Chapter 5 on Sobolev spaces (see Chapter 18 below) in Evans' PDE book, which involve applications of the Gauss Divergence Theorem (in essence integration by parts) stated in an appendix, without proof. The flattening argument introduced here effectively puts that horse in front of the wagon again.

Let Ω be a bounded open set in $\mathbb{R}^{\mathrm{N}} = \mathbb{R}^{n+1}$ and $M = \partial\Omega$ the union of finitely many patches $P = M \cap (a, b)$, each of which, after renumbering, comes with a description of $[a, b] \cap \Omega$ as given by

$$a_{n+1} \leq x_{n+1} < f(x_1, \ldots, x_n) < b_{n+1} \tag{8.5}$$

or

$$a_{n+1} < f(x_1, \ldots, x_n) \leq x_{n+1} < b_{n+1}. \tag{8.6}$$

Then there exist finitely many patches as above such that

$$M \subset P_1 \cup \cdots \cup P_I$$

If $P_i \cap P_j \neq \emptyset$, the transformations used in Section 8.19 respect the local descriptions of Ω, which we can use to compute a unique normal ν which points out of Ω in every patch on $M = \partial\Omega$. Overlapping patches then have the same outward normal vector ν.

In the simplest[11] relevant case with $n < \mathrm{N}$ and $\mathrm{N} > 1$, which is $n = 1$ and $\mathrm{N} = 2$, and $\Omega \subset \mathbb{R}^2$ bounded and open, $M = \partial\Omega$ is a 1-dimensional compact manifold covered by finitely many graphs

$$\{(x, f_i(x)) : a_i < x < b_i\} \text{ and } \{(g_j(y), y) : c_j < y < d_j\},$$

with $f_i \in C^1([a_i, b_i])$, $g_j \in C^1([c_j, d_j])$. The question we ask is how the integrals

$$\int_\Omega v_x(x, y)\, dxdy \quad \text{and} \quad \iint_\Omega w_y(x, y)\, dxdy \tag{8.7}$$

[9] Still to be done: integrals over manifolds with boundaries.

[10] The ones obtained as the image of $\partial\Omega$ under injective C^1 maps $\Phi : \bar{\Omega} \to \mathbb{R}^3$.

[11] What follows will generalise to $\mathrm{N} = n + 1 > 1$.

evaluate in terms of the values of v and w on the boundary $M = \partial\Omega$ for v and w in $C^1(\bar{\Omega})$. Here I use the common notation $dxdy = dydx$. With x_1 and x_2 replacing x and y the notation

$$\int_\Omega v_{x_1} = \int_\Omega v_{x_1}(x)\,dx = \iint_\Omega v_{x_1}(x_1,x_2)\,d(x_1,x_2)$$

(and likewise for the other integral) is more to my liking, but everybody writes $dx_1\,dx_2$ and $dx],dy$, rather than $d(x,y)$.

To answer the question we first consider a piece of the boundary described by $y = f(x)$, with $f \in C^1([a,b])$ and $c < f(x) < d$ for all $x \in [a,b]$, such that

$$\tilde{\Omega} = \Omega \cap ([a,b]\times[c,d]) = \{(x,y) :\ a \le x \le b,\ f(x) < y \le d\}, \tag{8.8}$$

and multiply w by a function $\zeta \in C^1(\mathbb{R}^2)$ as in Section 8.9 which is zero outside a subset $[\tilde{a},\tilde{b}]\times[\tilde{c},\tilde{d}]$ of $(a,b)\times(c,d)$. Denoting the resulting product by $\tilde{w} = \zeta w$ we now evaluate

$$\begin{aligned}\int_\Omega \tilde{w}_y &= \iint_{\tilde{\Omega}} \tilde{w}_y(x,y)\,dx\,dy = \int_a^b \left(\int_{f(x)}^d \tilde{w}_y(x,y)dy\right) dx\\ &= -\int_a^b \tilde{w}(x,f(x))\,dx\\ &= \int_a^b \underbrace{\frac{-1}{\sqrt{1+f'(x)^2}}}_{\nu_y}\,\tilde{w}(x,f(x))\,\underbrace{\sqrt{1+f'(x)^2}dx}_{ds=dS_1}\\ &= \int_\Phi \nu_y\tilde{w}\,ds,\end{aligned}$$

in which we recognised the y-component of the normal vector

$$\nu = \frac{1}{\sqrt{1+f'(u)^2}}\begin{pmatrix}-f'(u)\\ 1\end{pmatrix}$$

and

$$dS_1 = ds = |\Phi'(u)|^2\,du = \sqrt{1+f'(u)^2}\,du$$

evaluated via the parameterisation $\Phi(u) = (u, f(u))$.

For the integral of $\tilde{v} = \zeta v$ we use the new coordinates

$$\xi = x,\ \eta = y - f(x) \quad \text{whence} \quad x = \xi,\ y = \eta - f(\xi) \quad \text{and} \quad dx\,dy = d\xi\,d\eta$$

when transforming the integral over $(x,y) \in \tilde{\Omega}$ to an integral over

$$(\xi,\eta) \in D = \{(x, y - f(x)) :\ a \le x \le b,\ f(x) \le y \le d\}.$$

Defining $\phi(\xi,\eta)$ by

$$\phi(\xi,\eta)=\tilde{v}(x,y) \quad \text{we have} \quad v_x(x,y)=\phi_\xi(\xi,\eta)-f'(\xi)\phi_\eta(\xi,\eta)$$

via the chainrule, whence[12]

$$\begin{aligned}\int_\Omega \tilde{v}_x &= \iint_{\tilde{\Omega}} \tilde{v}_x(x,y)\,dx\,dy = \iint_D \left(\phi_\xi(\xi,\eta)-f'(\xi)\phi_\eta(\xi,\eta)\right)\,d\xi\,d\eta \\ &= \int_0 \left(\int_a^b \phi_\xi(\xi,\eta)\,d\xi\right)\,d\eta - \int_a^b \left(\int_0 f'(\xi)\phi_\eta(\xi,\eta)\,d\eta\right)\,d\xi \\ &= \int_0 \left(\phi(b,\eta)-\phi(a,\eta)\right)\,d\eta - \int_a^b f'(\xi)\int_0 \phi_\eta(\xi,\eta)\,d\eta\,d\xi \\ &= \int_a^b f'(\xi)\phi(\xi,f(\xi))\,d\xi = \int_\Phi \nu_x\tilde{v}\,ds,\end{aligned}$$

just as before, after inserting $\sqrt{1+f'(\xi)^2}$ and recognising $ds=dS_1$ as well as the x-component of the normal vector ν. In conclusion we have

$$\int_\Omega \tilde{v}_x = \int_{\partial\Omega} \nu_x\tilde{v}\,dS_1 \quad \text{and} \quad \int_\Omega \tilde{w}_y = \int_{\partial\Omega} \nu_y\tilde{w}\,dS_1 \tag{8.9}$$

Now look at (8.33), and choose[13] functions $\zeta_0\in C_c^1(\Omega)$ and $\zeta_1,\cdots,\zeta_I$, such that $0\le\zeta_i\le 1$ for $i=0,\ldots,I$, $\zeta_1,\ldots,\zeta_I$ as ζ above, such that

$$\zeta_0+\zeta_1+\cdots+\zeta_I\equiv 1 \quad \text{on} \quad \bar{\Omega},$$

to conclude, via $v=\zeta_0 v+\zeta_1 v+\cdots+\zeta_I v$ that

$$\int_\Omega v_{x_i} = \int_{\partial\Omega} \nu_i v\,dS_n \tag{8.10}$$

for $v\in C^1(\bar{\Omega})$ with $M=\partial\Omega$ and ν a globally defined[14] normal vector field on M such that locally a description as above holds.

In (8.10) we stated the formula for $\Omega\subset\mathbb{R}^N$ open and $M=\partial\Omega$ a compact $(N-1)$-dimensional manifold. The local proof involves

$$dS_{N-1}=dS_n=\sqrt{1+(\frac{\partial f}{\partial u_1})^2+\cdots+(\frac{\partial f}{\partial u_n})^2}\;du_1\cdots du_n$$

[12] We drop a conveniently chosen fixed upper bound in the η-integrals from the notation.
[13] This requires a technique we take for granted now.
[14] This is no additional assumption on Ω and its boundary.

and the normal vector ν with

$$\nu_1 = \frac{1}{\sqrt{1+|\nabla f(u)|^2}}\,\frac{\partial f}{\partial u_1}\,,\ldots,\,\nu_n = \frac{1}{\sqrt{1+|\nabla f(u)|^2}}\,\frac{\partial f}{\partial u_n},$$

$$\nu_{\rm N} = \nu_{n+1} = \frac{-1}{\sqrt{1+|\nabla f(u)|^2}}.$$

Formula (8.10) is often called Green's Theorem. Applying it to the product of v and some other function $\zeta \in C^1(\bar{\Omega})$ we obtain the integration by parts formula

$$\int_\Omega v_{x_i}\zeta = \int_{\partial\Omega} \zeta v\,\nu_i dS_n - \int_\Omega \zeta_{x_i} v. \tag{8.11}$$

8.2 The length of curve

In the 1-dimensional case I now follow Edwards and write $x = \gamma(t)$ with $t \in [a,b]$ and $\gamma : [a,b] \to \mathbb{R}^{\rm N}$. For any such γ the natural definition of the length would be the smallest upper bound on the set of numbers obtained via

$$\sum_{j=1}^{m} |\gamma(t_j) - \gamma(t_{j-1})|_2$$

with

$$a = t_0 < t_1 < \cdots < t_m = b.$$

Clearly this definition of length is invariant under reparameterisation of γ via strictly monotone bijections $\phi : [a,b] \to [c,d]$ as in Section 4.3. It's not a very hard exercise to show that for continuously differentiable $\gamma : [a,b] \to \mathbb{R}^{\rm N}$ the length is given by

$$s(\gamma) = \int_a^b |\gamma'(t)|_2\, dt,$$

and the change of variables formula applied to $u = \phi(t)$ with $\phi \in C^1([a,b])$ with $\phi'(t) \neq 0$ confirms that

$$\Phi :\, u \to \gamma(\phi^{-1}(u)) \tag{8.12}$$

is in $C^1([c,d])$ with $[c,d] = \phi([a,b])$, and has the same length[15]. Also, if $f = f(x)$ is continuous on $\gamma([a,b]) = \Phi([c,d])$, it follows that

$$\int_\gamma f = \int_\gamma f ds = \int_a^b f(\gamma(t))|\gamma'(t)|_2\, dt = \int_c^d f(\Phi(u))|\Phi'(u)|_2\, du = \int_\Phi f ds. \tag{8.13}$$

[15] The condition that $\gamma'(t) \neq 0$ also carries over to $\Phi'(u) \neq 0$.

As a special case we have that

$$s = \phi(t) = \int_a^t |\gamma'(\tau)|_2 \, d\tau$$

defines a reparameterisation for which $\hat{\gamma} = \Phi$ defined by $\gamma(t) = \hat{\gamma}(s) = \Phi(s)$ has

$$|\hat{\gamma}'(s)|_2 = |\Phi'(s)|_2 = 1.$$

Such a reparametrised $\tilde{\gamma}$ is called a unit speed path.

8.3 Line integrals of vector fields along curves

Besides (8.13) as a 1-dimensional example of what is to come in (8.31) we can also define an integral for $F = F(x) \in \mathbb{R}^n$ continuous on $\gamma([a,b])$, namely

$$\int_\gamma F \cdot ds = \int_a^b F(\gamma(t)) \cdot \gamma'(t) \, dt = \int_a^b F(\gamma(t)) \cdot \underbrace{\frac{\gamma'(t)}{|\gamma'(t)|_2}}_{T(t)} |\gamma'(t)|_2 \, dt, \quad (8.14)$$

but Edwards avoids the commonly used notation in the left hand side of (8.14), and instead writes

$$\int_\gamma F \cdot T \, ds,$$

with T the unit tangent vector[16] defined by

$$T(t) = \frac{\gamma'(t)}{|\gamma'(t)|_2}.$$

For reparametrisations $u = \phi(t)$ with $\phi \in C^1([a,b])$ and $\phi'(t) > 0$ and Φ defined as in (8.12) above you easily verify that the work

$$W = \int_\gamma F \cdot ds = \int_\gamma F \cdot T \, ds = \int_\Phi F \cdot T \, ds = \int_\Phi F \cdot ds.$$

done by the force field F does not change under reparametrisations $u = \phi(t)$ with $\phi'(t) > 0$. Of course

$$W = \int_\gamma F \cdot ds = \int_\gamma F \cdot T \, ds = \int_a^b (F_1(\gamma(t))\gamma_1'(t) + \cdots + F_n(\gamma(t))\gamma_n'(t)) \, dt$$

[16] I will use $\tau = T$.

$$= \int_a^b F_1(\gamma(t))\,\underbrace{\gamma_1'(t)\,dt}_{dx_1} + \cdots + \int_a^b F_n(\gamma(t))\,\underbrace{\gamma_n'(t)\,dt}_{dx_n}$$

leads to the notational convention

$$\int_\gamma F \cdot ds = \int_\gamma F_1 dx_1 + \cdots + \int_\gamma F_n dx_n = \int_\gamma F_1 dx_1 + \cdots + F_n dx_n. \quad (8.15)$$

If $F = \nabla f$ it is then common to write

$$\int_\gamma df = \int_\gamma \underbrace{\frac{\partial f}{\partial x_1} dx_1 + \cdots + \frac{\partial f}{\partial x_n} dx_n}_{df} = \int_\gamma \nabla f \cdot ds =$$

$$\int_a^b \nabla f(\gamma(t)) \cdot \gamma'(t)\,dt = f(\gamma(t))\big|_a^b = f(\gamma(b)) - f(\gamma(a)),$$

a notation which generalises (4.8), after which d was seen[17] as acting on f to produce $df = f'(x)dx$. Here we have d acting on f as

$$df = \frac{\partial f}{\partial x_1} dx_1 + \cdots + \frac{\partial f}{\partial x_n} dx_n, \quad (8.16)$$

producing what is called a 1-form in Section 1 of Chapter V in Edwards.

These 1-forms act on vectors. Whereas the x-dependent vector

$$F(x) = F_1(x)e_1 + \cdots + F_n(x)e_n \quad (8.17)$$

and the vector

$$v = v_1 e_1 + \cdots + v_n e_n$$

have and x-dependent inner product

$$F(x) \cdot v = F_1(x)v_1 + \cdots + F_n(x)v_n,$$

the 1-form

$$\omega = F_1(x)dx_1 + \cdots + F_n(x)dx_n \quad (8.18)$$

assigns to the same vector v the same x-dependent scalar

$$F_1(x)v_1 + \cdots + F_n(x)v_n,$$

in which we can insert $x = \gamma(t)$ and $v_i = \gamma_i'(t)$ to get a t-dependent quantity that we can integrate from $t = a$ to $t = b$ to define

$$\int_a^b (F_1(\gamma(t))\gamma_1'(t) + \cdots + F_n(\gamma(t))\gamma_n'(t))\,dt = \int_\gamma \omega.$$

[17] Writing f instead of F again.

Thus, ω evaluated in $x = \gamma(t)$ acts on $\gamma'(t)$ and is integrated from $t = a$ to $t = b$ to define $\int_\gamma \omega$. Note a reparameterisation of γ with $u = \phi(t)$ and $\phi'(t) < 0$ changes the sign of the integral.

The notation for ω hides the x-dependence and may look like the abuse of notation in $f = f(x)$, but (13) and (14) in Section 1 of Chapter V in Edwards are mathematically precise[18]. In conclusion we have $\int_\gamma f = \int_\gamma f\, ds$ defined for continuous scalar functions $f = f(x)$ and $\int_\gamma \omega$ for 1-forms $\omega = F_1(x)dx_1 + \cdots + F(x)dx_n$.

8.4 Surface area

We need some linear algebra for integrals over more general surface patches then the ones encountered in Section 8.1 , a surface patch being a set in $\mathbb{R}^3$ parameterised by a continuously differentiable injective map

$$\Phi : [0,1] \times [0,1] \to \mathbb{R}^3, \tag{8.19}$$

with

$$\nabla\Phi = (\nabla\Phi_1\ \nabla\Phi_2\ \nabla\Phi_3) = \begin{pmatrix} \frac{\partial\Phi_1}{\partial u_1} & \frac{\partial\Phi_2}{\partial u_1} & \frac{\partial\Phi_3}{\partial u_1} \\ \frac{\partial\Phi_1}{\partial u_2} & \frac{\partial\Phi_2}{\partial u_2} & \frac{\partial\Phi_3}{\partial u_2} \end{pmatrix}$$

denoting the matrix of which the columns are the gradients of the N = 3 components Φ_1, Φ_2, Φ_3 of Φ with respect to the $n = 2$ variables[19] u_1, u_2 in $\Phi = \Phi(u) = \Phi(u_1, u_2)$, consistent with the notation in Section (6.3).

Momentarily switching to a notation with Φ_1, Φ_2, Φ_3 as functions of u, v, $\nabla\Phi$ is the transpose of the Jacobian matrix

$$(\ \frac{\partial\Phi}{\partial u}\ \frac{\partial\Phi}{\partial v}\ ),$$

which has column vectors $\frac{\partial\Phi}{\partial u}$, $\frac{\partial\Phi}{\partial v}$.

In the special linear case with

$$\Phi_i(u, v) = a_i u + b_i v \tag{8.20}$$

the Jacobian matrix is the transpose of

$$\nabla\Phi = A = \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{pmatrix},$$

the matrix example in (8.28) starting the discussion in Section 8.7 below on

[18] We also write $\int_a^b f$ for the integral of a function.

[19] Everything that follows should generalise or trivialise to $1 \le n \le$ N.

the area $\mathcal{M}_2(a,b)$ of a parallelogram spanned by two vectors a and b

with entries a_1, a_2, a_3 and b_1, b_2, b_3 respectively. This parallelogram is then the image of $[0,1]\times[0,1]$ under Φ defined by (8.20), and its area is then equal to

$$\int_0^1 \int_0^1 \mathcal{M}_2(\frac{\partial\Phi}{\partial u}, \frac{\partial\Phi}{\partial v})\, dudv, \tag{8.21}$$

the integrand being independent of u, v, as $a = \frac{\partial\Phi}{\partial u}$ and $b = \frac{\partial\Phi}{\partial v}$ are constant vectors is the linear case (8.20).

It will be no surprise that (8.21) will also be used to define the area of the surface patch defined by Φ if Φ is not a linear map from $[0,1]^2$ to $\mathbb{R}^3$, and that everything generalises to $\Phi : [0,1]^n \to \mathbb{R}^{\mathrm{N}}$ with $1 \leq n < \mathrm{N}$. We expand on the linear case of this generalisation next.

8.5 The transpose, quadratic forms and operator norms

In (6.21) we can put $B = A^T$, the transpose of the matrix A with entries a_{ij} used in

$$y_i = \sum_{j=1}^{n} a_{ij} x_j,$$

which defined $A \in L(\mathbb{R}^n, \mathbb{R}^m)$. This gives

$$S = AA^T \in L(\mathbb{R}^m, \mathbb{R}^m) \quad \text{with entries} \quad s_{ik} = \sum_{j=1}^{n} a_{ij}a_{kj} = s_{ki}. \tag{8.22}$$

Since

$$|A|_{op} = \max_{0\neq x\in\mathbb{R}^n} \frac{|Ax|_2}{|x|_2} = \max_{|x|_2=1} |Ax|_2,$$

and likewise for $|A^T|_{op}$, we have

$$|A^T|^2_{op} = \max_{|z|_2=1} \underbrace{|A^T z|^2_2}_{A^Tz\cdot A^Tz} = \max_{|z|_2=1} AA^T z\cdot z = \max_{|z|_2=1} Sz\cdot z = \max_{0\neq z\in\mathbb{R}^m} \frac{Sz\cdot z}{z\cdot z}, \tag{8.23}$$

and we note that the bilinear mapping

$$(z,w) \to Sz\cdot w$$

from $\mathbb{R}^m \times \mathbb{R}^m$ to $\mathbb{R}$ then satisfies all the axioms of an inner product, except that $Sz\cdot z = 0$ does not imply that $z = 0$.

Opgave 8.1. Rederive the Cauchy-Schwarz inequality for $z, w \in \mathbb{R}^m$ by inspection of the minimum of the nonnegative function

$$\lambda \to |\lambda w - z|_2^2 = (\lambda w - z) \cdot (\lambda w - z),$$

and show that the same reasoning leads to

$$|Sz \cdot w| \leq \sqrt{Sz \cdot z}\,\sqrt{Sw \cdot w}.$$

Note the special case $m = n$ and $S = A = I$ and don't forget to discuss the possibility that the function you use is not a quadratic but a linear function.

For $S = AA^T$ as above we set

$$M = \max_{|z|_2 = 1} Sz \cdot z,$$

whereby we note that S is a symmetric matrix for which $Sz \cdot z \geq 0$ holds for all $z \in \mathbb{R}^m$. Just like it is easy to prove from the definition of the 2-norm via

$$|w|_2 = \sqrt{w \cdot w}$$

that

$$|z + w|_2^2 + |z - w|_2^2 = 2\,|z|_2^2 + 2\,|w|_2^2,$$

you easily verify that

$$S(z + w) \cdot (z + w) + S(z - w) \cdot (z - w) = 2\,Sz \cdot z + 2\,Sw \cdot w, \qquad (8.24)$$

an identity to play with, with $S = AA^T$as above, but also with $S = I$ the identity:

Opgave 8.2. The Cauchy-Schwarz inequality and the definition of the operator norm in Section 6.5 immediately imply that $M \leq |S|_{op}$. Write

$$4\,Sz \cdot w = S(z + w) \cdot (z + w) - S(z - w) \cdot (z - w)$$

and estimate the right hand side in terms of M to obtain that in particular for all $z, w \in \mathbb{R}^m$ with $|z|_2 = |w|_2 = 1$ it holds that $|Sz \cdot w| \leq M$. Conclude that $|S|_p = M$.

The map

$$z \to Q(z) = Sz \cdot z$$

defined by the symmetric matrix S is called a quadratic form. You should never forget the remarkable fact that the maxima of $z \to |Q(z)|$ and $z \to |Sz|$ on the unit ball coincide.

8.6 Eigenvalues of compact linear symmetric operators

The above carries over to $S : H \to H$ when H is any inner product space and $S : H \to H$ is linear and symmetric with respect to that inner product, and has the property that $Sz \cdot z \geq 0$ for all $z \in H$, except that we no longer know that the maxima exist. Introducing

$$|S|_{op} = \sup_{0 \neq z \in H} \frac{|Sz|}{|z|} = \sup_{0 \neq z \in H} \sqrt{\frac{Sz \cdot Sz}{z \cdot z}} = \sup_{z \cdot z = 1} \sqrt{Sz \cdot Sz}, \tag{8.25}$$

and

$$M = \sup_{z \cdot z = 1} \sqrt{Sz \cdot z}, \tag{8.26}$$

it suffices to have that S is bounded on the unit ball in H to have $M = |S|_{op} < \infty$.

Ignoring the trivial case that $M = 0$ we now observe that the Cauchy-Schwarz inequality in Exercise 8.1 also holds with S replaced by $M - S = MI - S$, I being the identity map, and it thus holds that

$$|(M - S)z \cdot w| \leq \sqrt{(M - S)z \cdot z}\, \sqrt{(M - S)w \cdot w}, \tag{8.27}$$

whence (varying w over the unit ball)

$$|(M - S)z| \leq \sqrt{(M - S)z \cdot z}\, \sqrt{|M - S|_{op}} \leq \sqrt{(M - S)z \cdot z}\, \sqrt{M + |S|_{op}}$$

Taking a sequence $z_n \in H$ with $|z_n| = 1$ and $Sz_n \cdot z_n \to M$, it then follows that the right hand side goes to zero, and thus

$$Mz_n - Sz_n \to 0.$$

If the sequence z_n can be chosen to have Sz_n converging to a limit $y \in H$, it follows that also $Mz_n \to y$ and that $M = |y| > 0$. But then $w = \frac{y}{M}$ is a unit eigenvector of S with eigenvalue M. We have therefore proved the following Theorem.

Stelling 8.3. *If H is an inner product space and $S : H \to H$ is linear, symmetric with $Sz \cdot z \geq 0$ for all $z \in H$, and if for every bounded sequence z_n in H it holds that Sz_n has a convergent subsequence, then*

$$\lambda_1 = \max_{0 \neq z \in H} \frac{Sz \cdot z}{z \cdot z} > 0$$

exists, and λ_1 is an eigenvalue of S whose eigenvectors are the maximizers of the maximum.

Given an eigenvector w_1 with $|w_1| = 1$ it easily follows that S maps

$$H_1 = \{z \in H : \, z \cdot w_1 = 0\}$$

to itself. Unless $H_1 = \{0\}$ it then follows that

$$\lambda_2 = \max_{z \cdot w_1 = 0 \neq z \in H} \frac{Sz \cdot z}{z \cdot z} \geq 0$$

is also an eigenvalue of S with eigenvector w_2 with $|w_2| = 1$. Note that we cannot exclude that $\lambda_2 = 0$. If this happens than H_1 is the null space of S.

Repeating the argument with

$$H_2 = \{z \in H : \, z \cdot w_1 = z \cdot w_2 = 0\}$$

we obtain a sequence of eigenvalues

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$$

and a sequence of corresponding mutually perpendicular unit eigenvectors

$$w_1, w_2, \ldots,$$

which either terminate (if H is finite-dimensional) or have the property that $\lambda_n \to 0$ as $n \to \infty$. The latter statement is a consequence of the convergent subsequences assumption.

If we do *not* assume that $Sz \cdot z \geq 0$ for all $z \in H$ then the absolute value of the first eigenvalue is still obtained as

$$|\lambda_1| = \max_{0 \neq z \in H} \frac{|Sz \cdot z|}{z \cdot z} > 0,$$

because, changing from S to $-S$ if necessary, it is no restriction to assume that

$$M = \sup_{0 \neq z \in H} \frac{|Sz \cdot z|}{z \cdot z} = \sup_{0 \neq z \in H} \frac{Sz \cdot z}{z \cdot z},$$

and reason as above. With the Cauchy-Schwarz inequality in (8.27) still holding[20] while the version in Exercise 8.1 fails, the upshot is that we still obtain eigenvalues with

$$|\lambda_1| \geq |\lambda_2| \geq \cdots \geq 0,$$

with eigenvectors as before. This is essentially the spectral theorem for compact symmetric linear operators S from an inner product space H to itself. It does not require any knowledge of the determinants which will become important next in the finite-dimensional case.

[20] I first saw this Cauchy-Schwarz trick in the appendix of the PDE book of Craig Evans.

8.7 Singular values and measures of parallelotopes

In the case that $H = \mathbb{R}^m$ the subsequence argument is not needed as the maximizer w for the maximum in Theorem 8.3 exists in view of the compactness of the unit ball in $\mathbb{R}^m$. In particular, in relation to (8.22), we have achieved that $\mathbb{R}^m$ has an orthonormal basis basis of eigenvectors $w_1, \ldots, w_m$ corresponding to eigenvalues $\sigma_1 \geq \cdots \geq \sigma_m \geq 0$ of AA^T. If $n \geq m$ these eigenvalues are called the singular values of A.

As an example consider the case that

$$A = \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{pmatrix} \tag{8.28}$$

and

$$AA^T = \begin{pmatrix} a_1^2 + a_2^2 + a_3^2 & a_1b_1 + a_2b_2 + a_3b_3 \\ b_1a_1 + b_2a_2 + b_3a_3 & b_1^2 + b_2^2 + b_3^2 \end{pmatrix} = \begin{pmatrix} a \cdot a & a \cdot b \\ b \cdot a & b \cdot b \end{pmatrix} \tag{8.29}$$

The outer product $a \times b$ of these two 3-column vectors a and b with, respectively, entries a_1, a_2, a_3 and entries b_1, b_2, b_3, is defined as the 3-vector with entries

$$a_2b_3 - a_3b_2, \quad a_3b_1 - a_1b_3, \quad a_1b_2 - a_2b_1,$$

and has squared length

$$|a \times b|^2 = (a_2b_3 - a_3b_2)^2 + (a_3b_1 - a_1b_3)^2 + (a_1b_2 - a_2b_1)^2 = \det(\mathrm{AA^T}),$$

as you should verify. That is to say, $\det(\mathrm{AA^T})$ is the sum of all the squares of all the 2×2-determinants of 2×2 submatrices of A. Here we count these 2×2 submatrices modulo the column permutations in (8.28).

As you may know, the length of the outer product $a \times b$ of a and b equals the area of the parallellogram spanned by a and b. Thus this area is the square root of the sum of the squares of the three 2×2-determinants in (8.28). It is precisely this statement that generalises to the n-dimensional measure of a parallelotope spanned by n vectors $x_1, \ldots, x_n$ in $\mathbb{R}^{\mathrm{N}}$.

Stelling 8.4. *Let $1 \leq n \leq \mathrm{N}$. Consider the parallelotope P spanned by the vectors $x_1, \ldots, x_n$ in $\mathbb{R}^{\mathrm{N}}$. After putting these vectors in the columns*[21] *of a matrix A, the n-dimensional measure $\mathcal{M}_n(x_1, \ldots, x_n)$ of P is the square root of the determinant of A^TA, and this determinant in turn is the sum of all the squares of the determinants of all $n \times n$ submatrices, and also equals the product $\sigma_1 \cdots \sigma_n$ of the singular values of A.*

[21] NB! Compared to (8.28) we switch from A to A^T in the notation.

Let us sketch a proof of this statement, first for (8.28), without using the outer product, using the invariance of the area under shear transformations. That is to say, the area of the parallelogram spanned by the vectors a and b is the same as that of the parallelogram spanned by the vectors $a + tb$ and b with $t \in \mathbb{R}$ arbitrary. The same statement holds for the determinant of $S = A^T A$ and the determinant of $S_t = A_t^T A_t$ where A_t is the matrix with column vectors $a + tb$ and b. Indeed, writing $A_t = A + tB$ we have

$$A_t^T A_t = (A + tB)^T (A + tB) = A^T A + tA^T B + tB^T A + t^2 B^T B$$
$$= \underbrace{A^T A + tA^T B}_{C_t} + t(\underbrace{B^T A + tB^T B}_{D_t}) = S_t$$

The matrix C_t is the matrix obtained from $S = A^T A$ by adding t times the second (last) row of S to its first row. Therefore C_t and S have the same determinant. In turn, the matrix S_t is obtained from C_t by adding t times the second (last) column of C_t to its first column. Therefore S_t and C_t have the same determinant. It follows that S_t and S have the same determinant. So both the area and the determinant are invariant under this shear transformation, which allows us to restrict our proof to the case in which $a \cdot b = 0$. Then the square of the area is equal to the product of the squares of the lengths of a and b, which is also the determinant of the diagonal matrix with entries $a \cdot a$ and $b \cdot b$. To prove the general statement in the theorem we use repeated shear transformations which leave both the determinant and the measure invariant and reduce the statement to be proved to the case that $x_i \cdot x_j = 0$ if $i \neq j$ and a corresponding diagonal matrix S with entries $x_1 \cdot x_1, \ldots, x_n \cdot x_n$. But this should be obvious from any formal definition of the n-dimensional measure of parallelotopes spanned by n vectors, a definition we happily leave here to be for what it is.

It remains to show that the determinant of the matrix S defined in (8.22) is also equal to the sum of the squares of the determinants of all the maximal square submatrices of A. These are also invariant under the shear transformations used above. Rather than using these transformations to reduce the statement to be proved to the case that the column vectors satisfy $x_i \cdot x_j = 0$ for $i \neq j$ we now use them diagonalise a maximal square part of the matrix A. Note that if the matrix A has no $n \times n$ submatrix with nonzero determinant, then the sum of the squared $n \times n$ determinants is zero, while also it cannot be the case that the column vectors are independent. Then our reduction to the case that the column vectors satisfy $x_i \cdot x_j = 0$ leads to one of these vectors being zero making the n-dimensional measure of P, and thereby the determinant of $A^T A$ zero as well.

Thus we may as well assume that the upper $n \times n$ part of A has nonzero determinant. It is a straightforward linear algebra exercise to show that, most

likely after relabeling the first n coordinates, shear transformations bring A in the form

$$A = \begin{pmatrix} \Lambda \\ B \end{pmatrix}$$

where Λ is an $n \times n$ diagonal matrix with nonzero entries $\lambda_1, \ldots, \lambda_n$. Here we already assumed that $n < \mathrm{N}$ because otherwise there was nothing to prove[22] in the first place. It now follows that

$$A^T A = \Lambda^2 + B^T B = \Lambda^2 + S,$$

where B is an $m \times n$ matrix with entries b_{ik} and S has entries

$$s_{ij} = \sum_{k=1}^{m} b_{ik} b_{jk}.$$

We therefore have, writing $B = [B_1, \ldots, B_n]$ with $B_1, \ldots, B_n$ the column vectors of B and using product notation, that

$$\det(\mathrm{A^T A}) = \underbrace{\Pi_j \lambda_j^2}_{\lambda_\mathrm{j}^1 \cdots \lambda_\mathrm{j}^\mathrm{n}} + \mathrm{s}_{11} \underbrace{\Pi_{j \neq 1} \lambda_j^2}_{\lambda_\mathrm{j}^2 \cdots \lambda_\mathrm{j}^\mathrm{n}} + \cdots + \mathrm{s}_{\mathrm{nn}} \Pi_{\mathrm{j} \neq \mathrm{n}} \lambda_\mathrm{j}^2$$

$$+\det \begin{pmatrix} \mathrm{s}_{11}\, \mathrm{s}_{12} \\ \mathrm{s}_{12}\, \mathrm{s}_{22} \end{pmatrix} \Pi_{\mathrm{j} \neq 1,2} \lambda_\mathrm{j}^2 + \cdots + \mathrm{detS} =$$

$$\Pi_j \lambda_j^2 + (B_1 \cdot B_1) \Pi_{j \neq 1} \lambda_j^2 + \cdots + \det \begin{pmatrix} \mathrm{B}_1 \cdot \mathrm{B}_1 & \mathrm{B}_1 \cdot \mathrm{B}_2 \\ \mathrm{B}_1 \cdot \mathrm{B}_2 & \mathrm{B}_2 \cdot \mathrm{B2} \end{pmatrix} \Pi_{\mathrm{j} \neq 1,2} \lambda_\mathrm{j}^2 + \cdots,$$

in which we wrote the term of degree n and only the first terms of degree $2n-2$ and degree $2n-4$ in $\lambda_1, \ldots, \lambda_n$. It should be obvious what the remaining terms are.

On the other hand, the sum of the squared determinants of the $n \times n$ submatrices of A is

$$\Pi_j \lambda_j^2 + (b_{11}^2 + b_{21}^2 + \cdots + b_{m1}^2) \Pi_{j \neq 1} \lambda_j^2 + \cdots + \left(\det \begin{pmatrix} \mathrm{b}_{11}\, \mathrm{b}_{12} \\ \mathrm{b}_{12}\, \mathrm{b}_{22} \end{pmatrix}^2 + \cdots \right) \Pi_{j \neq 1,2} \lambda_j^2 + \cdots$$

It remains to show that

$$B_1 \cdot B_1 = b_{11}^2 + b_{21}^2 + \cdots + b_{m1}^2,$$

which is clearly the case, and then that

$$\det \begin{pmatrix} \mathrm{B}_1 \cdot \mathrm{B}_1 & \mathrm{B}_1 \cdot \mathrm{B}_2 \\ \mathrm{B}_1 \cdot \mathrm{B}_2 & \mathrm{B}_2 \cdot \mathrm{B2} \end{pmatrix} = \det \begin{pmatrix} \mathrm{b}_{11}\, \mathrm{b}_{12} \\ \mathrm{b}_{12}\, \mathrm{b}_{22} \end{pmatrix}^2 + \cdots + \det \begin{pmatrix} \mathrm{b}_{(\mathrm{m}-1)1}\, \mathrm{b}_{(\mathrm{m}-1)2} \\ \mathrm{b}_{\mathrm{m}2}\, \mathrm{b}_{\mathrm{m}2} \end{pmatrix}^2,$$

[22] If you know your determinants.

etcetera. These are the statements we set out to prove for A, before applying shear transformations, but with shorter columnvectors, namely of length $\mathrm{N}-n$, respectively for two such vectors, up to m such vectors. We can thus systematically reduce the statement we want to proof to lower dimensions of the matrix under consideration, until we reach the easy case that $m = 1$.

8.8 Surface integrals

The treatment of Theorem 8.4 above is somewhat different from Edwards' exposition. I now return to (8.21). Generalising to $1 \leq n \leq \mathrm{N}$ we consider

$$\int_{[0,1]^n} \mathcal{M}_n(\frac{\partial\Phi}{\partial u})\, du = \int_0^1 \cdots \int_0^1 \mathcal{M}_n(\frac{\partial\Phi}{\partial u_1}, \ldots, \frac{\partial\Phi}{\partial u_n})\, du_1 \cdots du_n \qquad (8.30)$$

in which

$$\mathcal{M}_n(\frac{\partial\Phi}{\partial u_1}, \ldots, \frac{\partial\Phi}{\partial u_n}) = \mathcal{M}_n(\Phi_{u_1}, \ldots, \Phi_{u_n})$$

is given by Theorem 8.4. Here $du = du_1 \cdots du_n$ and $\int_{[0,1]^n} = \int_0^1 \cdots \int_0^1$ are just notational conventions.

In the special case that $n = 1$ we have

$$\mathcal{M}_1(\Phi_u) = \sqrt{\Phi_1'(u)^2 + \cdots + \Phi_n'(u)^2},$$

and

$$ds = \mathcal{M}_1(\Phi_u)\, du = \sqrt{\Phi_1'(u)^2 + \cdots + \Phi_n'(u)^2}\, du$$

is a common notation, introduced in Edwards[23] after a change of coordinates defined by

$$\frac{ds}{du} = \sqrt{\Phi_1'(u)^2 + \cdots + \Phi_n'(u)^2}.$$

While not corresponding to a change of coordinates the notation

$$dS = \mathcal{M}_2(\Phi_u, \Phi_v) du dv,$$

with the S of surface, is also common. Here I will use dS_n for

$$dS_n = \mathcal{M}_n(\frac{\partial\Phi}{\partial u})\, du = \mathcal{M}_n(\Phi_{u_1}, \ldots, \Phi_{u_n})\, du_1 \cdots du_n$$

in (8.30), i.e.

$$\int_\Phi dS_n = \int_0^1 \cdots \int_0^1 \mathcal{M}_n(\frac{\partial\Phi}{\partial u_1}, \ldots, \frac{\partial\Phi}{\partial u_n})\, du_1 \cdots du_n,$$

[23] In Section V.1 his $\gamma(t)$ would correspond $\Phi(u)$.

and, for a function $f = f(x) = f(x_1, \cdots, x_n)$ which is continuous on

$$\{x = \Phi(u) : \, u \in [0,1]^n\},$$

write

$$\int_\Phi f dS_n = \int_0^1 \cdots \int_0^1 f(\Phi(u_1, \ldots, u_n)) \mathcal{M}_n(\frac{\partial \Phi}{\partial u_1}, \ldots, \frac{\partial \Phi}{\partial u_n}) \, du_1 \cdots du_n, \tag{8.31}$$

the subscript Φ on the integral being at least consistent with the case $n = 1$ and $ds = dS_1$, and in fact coinciding with the notation in the second part of (8.13). Personally I often drop the dS_n from the notation and just write $\int_\Phi f$ instead of $\int_\Phi f dS_n$, and $\int_\gamma f$ if $n = 1$ and $\gamma = \Phi$ is a path in $\mathbb{R}^{\mathrm{N}}$. Below we will also allow general closed blocks

$$[a, b] = [a_1, b_1] \times \cdots \times [a_n, b_n].$$

8.9 Integrals of functions over compact manifolds

Section 6.8.1 and Section 8.18 concern 3 descriptions of what it means for $M \subset \mathbb{R}^{\mathrm{N}}$ to be an n-dimensional manifold in $\mathbb{R}^{\mathrm{N}}$. We now use characterisation (C) just below Exercise 6.16, and assume in addition that there exist finitely many injective continuously differentiable

$$\Phi_i : \, [a_i, b_i] \to \mathbb{R}^{\mathrm{N}}$$

defined on blocks $[a_i, b_i]$ as in the elaboration on (C) in Section 8.18 above[24], such that

$$M = \Phi_1((a_1, b_1)) \cup \cdots \cup \Phi_I((a_I, b_I)) = \Phi_1([a_1, b_1]) \cup \cdots \cup \Phi_I([a_I, b_I]), \tag{8.32}$$

and moreover that there exist corresponding smooth functions

$$\zeta_i : \mathbb{R}^{\mathrm{N}} \to [0, 1]$$

with

$$\zeta_1 + \cdots + \zeta_I \equiv 1 \;\text{ on } M \;\text{ and }\; \text{supp } \zeta_i \circ \Phi_i \subset (a_i, b_i)$$

for every $i = 1, \ldots, I$. Here supp $\zeta_i \circ \Phi_i$ is the support of the function $u \to \zeta_i(\Phi_i(u))$, defined as the closure of the set

$$\{u \in (a_i, b_i) : \, \zeta_i(\Phi_i(u)) \neq 0\}.$$

[24] The index i numbering the blocks now.

We say that $u \to \zeta_i(\Phi_i(u))$ belongs to $C^1_c((a_i, b_i))$, the class of C^1-functions with support contained in the open set (a_i, b_i).

You can think of each function ζ_i as fading the patch $\Phi_i((a_i, b_i))$, making it fade away completely near its boundary where $\zeta_i \equiv 0$, while together the ζ_i leave the whole of M as bright as it was before. Such *brightness* functions ζ_i can be chosen to vanish outside a neighbourhood in $\mathbb{R}^N$ of the image $\Phi_i(K_i)$, and the collection $\zeta_1, \ldots, \zeta_I$ is called a finite partition of unity on M, which is then (turning[25] a theorem around which says that such partitions exist if M is compact) a closed and bounded subset of $\mathbb{R}^N$.

If $f : M \to \mathbb{R}$ is continuous we now wish to define

$$\int_M f dS_n = \int_{\Phi_1} f\zeta_1 dS_n + \cdots + \int_{\Phi_I} f\zeta_I dS_n, \tag{8.33}$$

which requires a theorem that says this is independent of the choice of patches and brightness functions. We leave this issue[26] for now.

Of course the exposition above involves the change of variables theorem and Section 6.8.3. At the end of the day every theorem that we may wish to prove involving integrals of functions over M may be proved by restating and proving a local form only.

Finally we note that if the blocks $[a_i, b_i]$ and the injective continuously differentiable functions $\Phi_i : [a_i, b_i] \to \mathbb{R}^N$ with $\Phi'(u)$ of maximal rank can be chosen such that[27]

$$M = \Phi_1([a_1, b_1]) \cup \cdots \cup \Phi_I([a_I, b_I]) \quad \text{with} \quad \Phi_i((a_i, b_i)) \cap \Phi_j((a_j, b_j)) = \emptyset \tag{8.34}$$

for $i \neq j$, then

$$\int_M f dS_n = \int_{\Phi_1} f\, dS_n + \cdots + \int_{\Phi_I} f dS_n \tag{8.35}$$

is the obvious definition which Edwards uses, and which is what you do in examples.

8.10 More integration of differential forms

We look again at the right hand side of (8.10) with $N = n + 1$, evaluated for $\tilde{v}_i = \zeta v_i$ with ζ a cut-off function vanishing outside and near the boundary of some window

$$[a, b] = [a_1, b_1] \times \cdots \times [a_N, b_N],$$

[25] Following Steenbrink in his exposition of the Poincaré conjecture in Noordwijkerhout.
[26] But see later sections.
[27] Edwards: a hard theorem says this can be done.

in which we now assume a local representation of $\Omega \cap [a,b]$ given by[28]

$$(x_1,\dots,x_n) \in [a_1,b_1]\times\cdots\times[a_n,b_n] \quad \text{and} \quad a_{n+1} \le x_{n+1} < f(x_1,\dots,x_n),$$

with $f \in C^1([a_1,b_1]\times\cdots\times[a_n,b_n])$ taking values in (a_{n+1},b_{n+1}), and

$$\Phi(u_1,\dots,u_n) = (u_1,\dots,u_n,f(u_1,\dots,u_n)) \tag{8.36}$$

parameterising $M \cap [a,b] = \partial\Omega \cap [a,b]$. We denote the unit basis vectors by $e_1,\dots,e_{n+1}$.

For $n=2$ the vector obtained by the formal determinant manipulation

$$\begin{vmatrix} \frac{\partial\Phi_1}{\partial u_1} & \frac{\partial\Phi_2}{\partial u_1} & \frac{\partial\Phi_3}{\partial u_1} \\ \frac{\partial\Phi_1}{\partial u_2} & \frac{\partial\Phi_2}{\partial u_2} & \frac{\partial\Phi_3}{\partial u_2} \\ e_1 & e_2 & e_3 \end{vmatrix} = \begin{vmatrix} \frac{\partial\Phi_1}{\partial u_1} & \frac{\partial\Phi_2}{\partial u_1} \\ \frac{\partial\Phi_1}{\partial u_2} & \frac{\partial\Phi_2}{\partial u_2} \end{vmatrix} e_3 + \begin{vmatrix} \frac{\partial\Phi_2}{\partial u_1} & \frac{\partial\Phi_3}{\partial u_1} \\ \frac{\partial\Phi_2}{\partial u_2} & \frac{\partial\Phi_3}{\partial u_2} \end{vmatrix} e_1 + \begin{vmatrix} \frac{\partial\Phi_3}{\partial u_1} & \frac{\partial\Phi_1}{\partial u_1} \\ \frac{\partial\Phi_3}{\partial u_2} & \frac{\partial\Phi_1}{\partial u_2} \end{vmatrix} e_2 \tag{8.37}$$

is commonly called the cross product of the vectors Φ_{u_1} and Φ_{u_2}, and for $\Phi(u_1,u_2) = (u_1,u_2,f(u_1,u_2))$ it evaluates as[29]

$$e_3 - \frac{\partial f}{\partial u_1}e_1 - \frac{\partial f}{\partial u_2}e_2 = -\frac{\partial f}{\partial u_1}e_1 - \frac{\partial f}{\partial u_2}e_2 + e_3, \tag{8.38}$$

which is a positive multiple of the unit vector ν characterised by having its last component positive and being perpendicular to the graph defined by $u_3 = f(u_1,u_2)$. For any continuously differentiable

$$\Phi : [a_1,b_1]\times[a_2,b_2] \to \mathbb{R}^3$$

with Φ_{u_1} and Φ_{u_2} linearly independent, the vector defined by (8.37) is perpendicular to the plane spanned by Φ_{u_1} and Φ_{u_2}, and can be normalised by dividing it by its length, which we recognise as

$$\mathcal{M}_2(\Phi_{u_1},\Phi_{u_2})$$

in view of Theorem 8.4. If we call this normalised vector ν, which in case of (8.38) is simply[30]

$$\nu = \frac{1}{\sqrt{1+f_{u_1}^2+f_{u_2}^2}}\left(-\frac{\partial f}{\partial u_1}e_1 - \frac{\partial f}{\partial u_2}e_2 + e_3\right), \tag{8.39}$$

[28] Like (8.5), whereas I did most local arguments using (8.6).

[29] Denoting the partial with subscripts u_1 and u_2.

[30] Please allow the simultaneous use of both expressions in $f_{u_i} = \frac{\partial f}{\partial u_i}$.

and consider $\tilde{v}_i$ as the i^{th} component of a vector field $\tilde{v} = \zeta v$ defined on $M \cap [a,b]$, with v a vector field on M, then

$$\int_M \nu \cdot \tilde{v}\ dS_2 =$$

$$\iint_{[a_1,b_1]\times[a_2,b_2]} \left(\tilde{v}_1 \begin{vmatrix} \frac{\partial \Phi_2}{\partial u_1} & \frac{\partial \Phi_3}{\partial u_1} \\ \frac{\partial \Phi_2}{\partial u_2} & \frac{\partial \Phi_3}{\partial u_2} \end{vmatrix} + \tilde{v}_2 \begin{vmatrix} \frac{\partial \Phi_3}{\partial u_1} & \frac{\partial \Phi_1}{\partial u_1} \\ \frac{\partial \Phi_3}{\partial u_2} & \frac{\partial \Phi_1}{\partial u_2} \end{vmatrix} + \tilde{v}_3 \begin{vmatrix} \frac{\partial \Phi_1}{\partial u_1} & \frac{\partial \Phi_2}{\partial u_1} \\ \frac{\partial \Phi_1}{\partial u_2} & \frac{\partial \Phi_2}{\partial u_2} \end{vmatrix} \right) \underbrace{du_1\, du_2}_{du},$$

which we may be inclined to write as

$$\int_\Phi \tilde{v}_1\, dx_2 dx_3 + \tilde{v}_2\, dx_3 dx_1 + \tilde{v}_3\, dx_1 dx_2 = \int_\Phi \omega, \tag{8.40}$$

with

$$\omega = \tilde{v}_1\, dx_2 dx_3 + \tilde{v}_2\, dx_3 dx_1 + \tilde{v}_3\, dx_1 dx_2,$$

using formal rules such as

$$dx_2 dx_3 = \begin{vmatrix} \frac{\partial x_2}{\partial u_1} & \frac{\partial x_3}{\partial u_1} \\ \frac{\partial x_2}{\partial u_2} & \frac{\partial x_3}{\partial u_2} \end{vmatrix} \underbrace{du_1\, du_2}_{du} = \begin{vmatrix} \frac{\partial \Phi_2}{\partial u_1} & \frac{\partial \Phi_3}{\partial u_1} \\ \frac{\partial \Phi_2}{\partial u_2} & \frac{\partial \Phi_3}{\partial u_2} \end{vmatrix} \underbrace{du_1\, du_2}_{du}.$$

We then have that (8.40) is equal to

$$\int_\Omega \nabla \cdot \tilde{v} = \int_\Omega \nabla \cdot \tilde{v}(x)\, dx = \iiint_\Omega \left(\frac{\partial \tilde{v}_1}{\partial x_1} + \frac{\partial \tilde{v}_2}{\partial x_2} + \frac{\partial \tilde{v}_3}{\partial x_3} \right) \underbrace{dx_1\, dx_2\, dx_3}_{dx},$$

which we will wish to write as an integral of the differential form

$$d\omega = \left(\frac{\partial \tilde{v}_1}{\partial x_1} + \frac{\partial \tilde{v}_2}{\partial x_2} + \frac{\partial \tilde{v}_3}{\partial x_3} \right) \underbrace{dx_1 dx_2 dx_3}_{\neq dx},$$

in which $dx_1 dx_2 dx_3$ is part of a 3-form and not be read as $dx = dx_1\, dx_2\, dx_3$.

All of the above generalises[31] to arbitrary $\mathrm{N} = n + 1$, e.g. we also have

$$\int_\Omega \left(\frac{\partial \tilde{v}_1}{\partial x_1} + \frac{\partial \tilde{v}_2}{\partial x_2} + \frac{\partial \tilde{v}_3}{\partial x_3} + \frac{\partial \tilde{v}_4}{\partial x_4} \right) dx \tag{8.41}$$

$$= \int_\Phi \tilde{v}_1\, dx_2 dx_3 dx_4 + \cdots \text{(cyclicly permutated terms)} \cdots = \int_\Phi \omega,$$

[31] This is why we put the unit vectors in the last row of the determimant in (8.37).

using rules like

$$dx_2dx_3dx_4 = \begin{vmatrix} \frac{\partial x_2}{\partial u_1} & \frac{\partial x_3}{\partial u_1} & \frac{\partial x_4}{\partial u_1} \\ \frac{\partial x_2}{\partial u_2} & \frac{\partial x_3}{\partial u_2} & \frac{\partial x_4}{\partial u_2} \\ \frac{\partial x_2}{\partial u_3} & \frac{\partial x_3}{\partial u_3} & \frac{\partial x_4}{\partial u_3} \end{vmatrix} \underbrace{du_1\, du_2\, du_3}_{du},$$

and (8.41) should be the integral of the 4-form

$$d\omega = \left(\frac{\partial \tilde{v}_1}{\partial x_1} + \frac{\partial \tilde{v}_2}{\partial x_2} + \frac{\partial \tilde{v}_3}{\partial x_3} + \frac{\partial \tilde{v}_4}{\partial x_4} \right) dx_1dx_2dx_3dx_4.$$

Clearly such a d-calculus requires rules such as $dx_idx_j = -dx_jdx_i$. I played with the formal rules that one might like to have in Section 6.9, see also the discussion after Stelling 4.12. This notation, used in Edwards, is cumbersome as the difference between spaces or no spaces between dx_i and dx_j is hardly visible, which is a reason to write $dx_i \wedge dx_j$ instead of dx_idx_j.

We conclude with the simplest but slightly confusing case, $n = 1$ and $\mathrm{N} = 2$, when (8.37) should be replaced by

$$\begin{vmatrix} \frac{\partial \Phi_1}{\partial u} & \frac{\partial \Phi_2}{\partial u} \\ e_1 & e_2 \end{vmatrix} = \frac{\partial \Phi_2}{\partial u}\, e_1 - \frac{\partial \Phi_1}{\partial u}\, e_2, \tag{8.42}$$

which for

$$\Phi(u) = (u, f(u))$$
$$e_2 - f'(u)e_1,$$

and leads to

$$\int_M \nu \cdot \tilde{v}\; dS_1 = \int_{[a,b]} \left(-\tilde{v}_1 \frac{\partial \Phi_2}{\partial u} + \tilde{v}_2 \frac{\partial \Phi_1}{\partial u} \right) du = \iint_\Omega \left(\frac{\partial \tilde{v}_1}{\partial x_1} + \frac{\partial \tilde{v}_2}{\partial x_2} \right) dx_1\, dx_2,$$

in which we dropped the subscripts in a_1, b_1, u_1. Here we have

$$\omega = -\tilde{v}_1\, dx_2 + \tilde{v}_2\, dx_1 \quad \text{with} \quad d\omega = \left(\frac{\partial \tilde{v}_1}{\partial x_1} + \frac{\partial \tilde{v}_2}{\partial x_2} \right) dx_1dx_2,$$

and

$$\int_{\partial\Omega} \omega = \int_\Omega d\omega.$$

In x, y notation for $\omega = p(x,y)dx + q(x,y)dy$ we have $d\omega = (q_x - p_y)dxdy$ and

$$\int_{\partial\Omega} p(x,y)dx + q(x,y)dy = \int_\Omega (q_x - p_y)dxdy, \tag{8.43}$$

which should make you wonder about

$$\int_\gamma p(x,y,z)dx + q(x,y,z)dy + r(x,y,z)dz,$$

for $\gamma : [a,b] \to \mathbb{R}^3$ as in Section 8.3. Section 8.11 below explores what's going on here.

Note that in all these examples the N-form $\omega = f(x)dx_1 \cdots dx_{\mathrm{N}}$ integrated over the domain Ω should sensibly be agreed to give[32]

$$\int_\Omega \omega = \int_\Omega f(x)dx_1 \cdots dx_{\mathrm{N}} = \int_\Omega f.$$

8.11 From Green's Theorem to Stokes' curl theorem

Now consider (8.36) as a local description of a manifold M and forget about Ω as being a domain with $M = \partial\Omega$. Instead let Ω be as in (8.1) with $\mathrm{N} = 2$ and let M be the graph of $f : \Omega \to \mathbb{R}$. Assume for simplicity that $\partial\Omega$ is parameterised by a 1-periodic continuously differentiable function $t \to u(t) = (u_1(t), u_2(t))$. Then

$$t \xrightarrow{\gamma} (u_1(t), u_2(t), f(u_1(t), u_2(t))) \tag{8.44}$$

parameterises the "boundary"

$$\partial M = \{\underbrace{(u, f(u))}_{\Phi(u)} : u \in \partial\Omega\},$$

and

$$u \xrightarrow{\Phi} (u, f(u)) \tag{8.45}$$

parameterises M, with $u = (u_1, u_2) \in \Omega$.

For

$$F(x) = F_1(x)e_1 + F_2(x)e_2 + F_3(x)e_3$$

we introduce

$$\omega = F_1(x)dx_1 + F_2(x)dx_2 + F_3(x)dx_3$$

as in (8.17) and (8.18) and consider the integral

$$\int_{\partial M} \omega$$

[32] Don't confuse this f with f in the local description above.

as in (8.15). It evaluates as

$$\int_{\partial M} \omega = \int_0^1 (F_1(\gamma(t))\gamma_1'(t) + F_2(\gamma(t))\gamma_2'(t)) + F_3(\gamma(t))\gamma_3'(t))) \, dt$$

$$= \int_0^1 (F_1(u(t), f(u(t))u_1'(t) + F_3(u(t), f(u(t)))f_{u_1}(u(t)u_1'(t)) \, dt$$

$$+ \int_0^1 (F_2(u(t), f(u(t))u_2'(t) + F_3(u(t), f(u(t)))f_{u_2}(u(t)u_2'(t)) \, dt =$$

$$\int_{\partial\Omega} \zeta = \int_\Omega d\zeta, \tag{8.46}$$

in which

$$\zeta = \left(F_1 + F_3 \frac{\partial f}{\partial u_1}\right) du_1 + \left(F_2 + F_3 \frac{\partial f}{\partial u_2}\right) du_2$$

Next we compute

$$d\zeta = \left(\frac{\partial F_1}{\partial x_2} + \frac{\partial F_1}{\partial x_3}\frac{\partial f}{\partial u_2} + \frac{\partial F_3}{\partial x_2}\frac{\partial f}{\partial u_1} + \frac{\partial F_3}{\partial x_3}\frac{\partial f}{\partial u_2}\frac{\partial f}{\partial u_1} + F_3\frac{\partial^2 f}{\partial u_2 \partial u_1}\right) du_2 du_1$$

$$+ \left(\frac{\partial F_2}{\partial x_1} + \frac{\partial F_2}{\partial x_3}\frac{\partial f}{\partial u_1} + \frac{\partial F_3}{\partial x_1}\frac{\partial f}{\partial u_2} + \frac{\partial F_3}{\partial x_3}\frac{\partial f}{\partial u_1}\frac{\partial f}{\partial u_2} + F_3\frac{\partial^2 f}{\partial u_1 \partial u_2}\right) du_1 du_2,$$

which in view of $du_2 du_1 = -du_1 du_2$ reduces to

$$d\zeta = \phi(u_1, u_2) du_1 du_2 \tag{8.47}$$

with $\phi(u_1, u_2)$ given by

$$\phi = -\underbrace{\left(\frac{\partial F_3}{\partial x_2} - \frac{\partial F_2}{\partial x_3}\right)}_{G_1}\frac{\partial f}{\partial u_1} - \underbrace{\left(\frac{\partial F_1}{\partial x_3} - \frac{\partial F_3}{\partial x_1}\right)}_{G_2}\frac{\partial f}{\partial u_2} + \underbrace{\left(\frac{\partial F_2}{\partial x_1} - \frac{\partial F_1}{\partial x_2}\right)}_{G_3} \tag{8.48}$$

$$= -G_1\frac{\partial f}{\partial u_1} - G_2\frac{\partial f}{\partial u_2} + G_3.$$

You should note that the *second order derivatives* of (8.45) are dropouts in the calculations that lead to (8.48).

Now compare (8.48) to ν in (8.39) and recall that for Φ given by (8.45) we know that

$$\mathcal{M}_2(\Phi_{u_1}, \Phi_{u_2}) = \sqrt{1 + f_{u_1}^2 + f_{u_2}^2}.$$

Summing up we thus have

$$\int_{\partial M} (F \cdot \tau)\, dS_1 =$$

(hello forms)

$$\int_{\partial M} \omega = \int_{\partial \Omega} \zeta = \int_{\Omega} d\zeta = \int_{\Omega} \underbrace{\phi\, du_1 du_2}_{d\zeta} =$$

(goodbye forms)

$$\int_{\Omega} \phi = \int_{\Omega} (G \cdot \nu)\, \mathcal{M}_2(\Phi_{u_1}, \Phi_{u_2}) = \int_{M} (G \cdot \nu)\, dS_2,$$

with G derived from F as indicated in (8.48), and commonly denoted as $G = \nabla \times F$, i.e.

$$\int_{\partial M} (F \cdot \tau)\, dS_1 = \int_{M} (G \cdot \nu)\, dS_2 \quad \text{with} \quad G = \nabla \times F, \tag{8.49}$$

using the parameterisations as indicated[33]. But don't say goodbye:

8.12 Pullbacks and the action of d

We already saw in the reasoning from (8.15) to (8.16) that d acting on a C^1-function $f = f(x_1, \ldots, x_{\mathrm{N}})$ produces a 1-form

$$df = \frac{\partial f}{\partial x_1} dx_1 + \cdots + \frac{\partial f}{\partial x_{\mathrm{N}}} dx_{\mathrm{N}} = \frac{\partial f}{\partial x_i} dx_i, \tag{8.50}$$

using the convention that we sum over repeated indices. With $f(x_1, \ldots, x_{\mathrm{N}})$ replaced by $u(x, y)$ this is (6.48) in Section 6.9.2. There I played with the d-algebra that emerges whenever you do integration using formal notations such as (4.8), which is just (8.50) with $n = 1$ and $f(x_1, \cdots, x_{\mathrm{N}})$ replaced by $F(x)$.

Now consider a parameterisation $x = \Phi(u)$ as in (C) in Section 6.8.1. We use Φ to pull back expressions with x and $dx_1, \ldots, dx_{\mathrm{N}}$ back to expressions with u and $du_1, \ldots, du_n$, in a way that is consistent with the discussion

[33] Figure out that annoying $\pm$ afterwards? We have, depending on the parameterisation:

$$\int_{\partial M} (F \cdot \tau)\, dS_1 = \pm \int_{M} (G \cdot \nu)\, dS_2.$$

leading to (8.40) and the formal rules that emerge in the calculations to do so. Thus we certainly want to deal with

$$f(x) = \phi(u) \quad \text{via} \quad x = \Phi(u). \tag{8.51}$$

A mathematician's way to do so is to introduce

$$\phi = \Phi^*(f) = f \circ \Phi, \tag{8.52}$$

the pullback of f via Φ, which then also provides us with

$$d\phi = \frac{\partial \phi}{\partial u_1} du_1 + \cdots + \frac{\partial \phi}{\partial u_n} du_n. \tag{8.53}$$

If g is another function of x then clearly

$$\Phi^*(f+g) = \Phi^*(f) + \Phi^*(g), \quad \Phi^*(fg) = \Phi^*(f)\Phi^*(g),$$

which suggests as a definition of the pullback of a 1-form $\omega = f_i dx_i$ that

$$\Phi^*(f_i dx_i) = \underbrace{\Phi^*(f_i)}_{\phi_i}\, \Phi^*(dx_i), \tag{8.54}$$

in which $\phi_i(u) = f_i(\Phi(u))$ as before. This definition would imply that

$$\Phi^*(df) = \underbrace{\frac{\partial f}{\partial x_i}((\Phi(u))}_{\Phi^*(D_i f)(u)}\, \Phi^*(dx_i). \tag{8.55}$$

Note that $D_i f$ as notation for the i^{th} first order partial derivative of f has the advantage of not using the variable x in the notation.

On the other hand (8.53) implies via the chain rule that

$$d(\Phi^*(f)) = \frac{\partial}{\partial u_j}(f(\Phi(u))\, du_j = \frac{\partial f}{\partial x_i}(\Phi(u)) \frac{\partial \Phi_i}{\partial u_j}\, du_j, \tag{8.56}$$

and comparing to (8.55) we see that, if we define the pullback of dx_i under Φ to be

$$\Phi^*(dx_i) = \frac{\partial \Phi_i}{\partial u_j}\, du_j, \tag{8.57}$$

it follows that

$$\Phi^*(df) = \Phi^*(df). \tag{8.58}$$

The definition of $\Phi^*(dx_i)$ by (8.57) is just a formalisation of the familiar "rule"

$$dx_i = \frac{\partial x_i}{\partial u_j}\, du_j$$

for expressing dx_i in u, $du_1, \ldots, du_n$, just like expressing $f(x)$ in u via (8.51) is formalised by (8.52). It implies that the pullback of the 1-form in (8.54) evaluates as

$$\underbrace{\Phi^*(f_i dx_i) = \phi_i \frac{\partial \Phi_i}{\partial u_j}\, du_j}_{with\ \phi_i(u)=f_i(\Phi(u))} = f_i(\Phi(u))\, \frac{\partial \Phi_i}{\partial u_j}\, du_j = f_i(\Phi(u))\, D_j\Phi_i(u)\, du_j. \quad (8.59)$$

Next we observe that d acting on the resulting 1-form in (8.59) may be evaluated, using the chain rule and $du_k du_j = -du_j du_k$, as

$$d(\Phi^*(f_i dx_i)) = d(f_i(\Phi(u))\, \frac{\partial \Phi_i}{\partial u_j}\, du_j) = \frac{\partial}{\partial u_k}(f_i(\Phi(u))\, \frac{\partial \Phi_i}{\partial u_j})\, du_k du_j$$

$$= (\frac{\partial}{\partial u_k}(f_i(\Phi(u)))\, \frac{\partial \Phi_i}{\partial u_j}\, du_k du_j + f_i(\Phi(u))\, \underbrace{\frac{\partial^2 \Phi_i}{\partial u_k \partial u_j}\, du_k du_j}_{zero\ the\ hero!} =$$

$$\frac{\partial f_i}{\partial x_k}(\Phi(u))\, \frac{\partial \Phi_k}{\partial u_k}\, \frac{\partial \Phi_i}{\partial u_j}\, du_k\, du_j = \Phi^*(D_k f_i)\, \frac{\partial \Phi_k}{\partial u_k}\, \frac{\partial \Phi_i}{\partial u_j}\, du_k\, du_j, \quad (8.60)$$

in which we used

$$d(f_i dx_i) = \frac{\partial f_i}{\partial x_k}\, dx_k\, dx_i \quad (8.61)$$

in the u-variables. Recall that this was the definition[34] in Section 6.9.2 of the action of d on 1-forms. With

$$\Phi^*(f_{ij} dx_i dx_j) = \underbrace{\Phi^*(f_{ij})}_{\phi_{ij}}\, \Phi^*(dx_i dx_j) \quad (8.62)$$

as the obvious defining analog of (8.54), we have that

$$\Phi^*(d(f_i dx_i)) = \Phi^*(\frac{\partial f_i}{\partial x_k}\, dx_k dx_i) = \Phi^*(D_k f_i)\, \Phi^*(dx_k dx_i). \quad (8.63)$$

Comparing to (8.63) to (8.60) it follows that

$$\Phi^*(d(f_i dx_i)) = d(\Phi^*(f_i dx_i)), \quad (8.64)$$

provided we define

$$\Phi^*(dx_k dx_i) = \underbrace{\frac{\partial \Phi_k}{\partial u_k}\, \frac{\partial \Phi_i}{\partial u_j}\, du_k\, du_j}_{sum\ over\ 1 \le k,j \le n} = \underbrace{(\frac{\partial \Phi_k}{\partial u_k}\, \frac{\partial \Phi_i}{\partial u_j} - \frac{\partial \Phi_k}{\partial u_k}\, \frac{\partial \Phi_i}{\partial u_j})\, \underline{du_k\, du_j}}_{sum\ over\ 1 \le k < j \le n}$$

[34] Recall the choice to set $ddx_i = 0$, leading to $dd\omega = 0$ for any form ω.

$$= \frac{\partial(\Phi_k, \Phi_i)}{\partial u_k \partial u_j} \underline{du_k\, du_j}, \tag{8.65}$$

in which the underline indicates that we sum over all k, j with $1 \leq k < j \leq n$. Just as in (8.58) we see that the actions of d and Φ^* commute.

Note that the second order derivatives have disapperared in (8.60). The derivation is typically done under the assumption that $\Phi \in C^2$, also in Edwards, and an additional analysis argument is needed[35] to give meaning to the results if Φ is only in C^1, because the determinants in (8.65) are exactly the determinants that showed up in (8.37) and the subsequent derivation of (8.40), where effectively $dx = dx_1\, dx_2\, dx_3$ is first replaced by a 3-form $dx_1 dx_2 dx_3$ pulled back to a 2-form $du_1 du_2$, which in turn is replaced by $du = du_1\, du_2$ again.

The step by step generalisation to the action of d and Φ^* on k-forms of any order k is easily made once the reasoning above is understood. For any k-form

$$\omega = f_{i_1,\ldots,i_k} dx_{i_1} \cdots dx_{i_k}$$

we have

$$\Phi^*(d\omega) = d(\Phi^*(\omega)) \tag{8.66}$$

Every such form may be written as

$$\omega = f_{i_1,\ldots,i_k} dx_{i_1} \cdots dx_{i_k} = \tilde{f}_{i_1,\ldots,i_k}\, \underline{dx_{i_1} \cdots dx_{i_k}}, \tag{8.67}$$

where in the second expression we sum only over those $i_1, \ldots, i_k$ for which $1 \leq i_1 < \cdots < i_k \leq \mathrm{N}$. For instance

$$\omega = f_{ij}\, dx_i dx_j = \underbrace{(f_{ij} - f_{ji})}_{\tilde{f}_{ij}}\, \underline{dx_i dx_j},$$

but this is not compulsory, as the examples

$$\omega = f_1\, dx_1 + f_2\, dx_2 + f_3\, dx_3$$

with cyclic notation for

$$d\omega = \underbrace{(\frac{\partial f_3}{\partial x_2} - \frac{\partial f_2}{\partial x_3})}_{g_1}\, dx_2 dx_3 + \underbrace{(\frac{\partial f_1}{\partial x_3} - \frac{\partial f_3}{\partial x_1})}_{g_2}\, dx_3 dx_1 + \underbrace{(\frac{\partial f_2}{\partial x_1} - \frac{\partial f_1}{\partial x_2})}_{g_3}\, dx_1 dx_2$$

and

$$\zeta = g_1\, dx_2 dx_3 + g_2\, dx_3 dx_1 + g_3\, dx_1 dx_2$$

[35] Using approximation arguments.

with

$$d\zeta = (\frac{\partial g_1}{\partial x_1} + \frac{\partial g_2}{\partial x_2} + \frac{\partial g_3}{\partial x_3})\, dxdydz$$

in Section 8.13 show.

Finally we observe that if we put the coefficients f_1, f_2, f_3 of this ω in a vector $F = f_1e_1 + f_2e_2 + f_3e_3$ and the coefficients g_1, g_2, g_3 in this cyclic representation of $d\omega$ in a vector $G = g_1e_1 + g_2e_2 + g_3e_3$, we obtain that

$$G = \nabla \times F,$$

the curl of F, whereas with the coefficients of η we obtain the coefficient of $d\zeta$ as

$$\frac{\partial g_1}{\partial x_1} + \frac{\partial g_2}{\partial x_2} + \frac{\partial g_3}{\partial x_3} = \nabla \cdot G,$$

the divergence of G. These appear in the Gauss divergence and the Stokes curl theorems for vectorfields in $\mathbb{R}^3$ in Section 8.13 below[36]. The general statement (8.77) is also called Stokes Theorem. It has both theorems in $\mathbb{R}^3$ and Green's Theorem in $\mathbb{R}^3$ as special cases.

8.13 From Gauss' to Stokes' Theorem

From Section 8.10 and partitions of unity arguments we have that for $\Omega \subset \mathbb{R}^N = \mathbb{R}^{n+1}$ open and bounded, with $\partial\Omega$ a compact $(N-1)$-dimensional C^1-manifold, and in every $p \in M$, after renumbering, a local description of $\Omega \cap [a, b]$ given by

$$a_{n+1} \leq x_{n+1} < f(x_1, \ldots, x_n) < b_{n+1}$$

or

$$a_{n+1} < f(x_1, \ldots, x_n) \leq x_{n+1} < b_{n+1},$$

with $f \in C^1$ and $p \in (a, b)$, that there exists a globally defined normal vectorfield $\nu : \partial\Omega \to \mathbb{R}^N$ with $\nu(p)$ pointing out of Ω in every patch as above. For every continuously differentiable $V : \bar{\Omega} \to \mathbb{R}^N$ it now holds that

$$\int_\Omega \nabla \cdot V = \int_{\partial\Omega} \nu \cdot V \, dS_{N-1}, \tag{8.68}$$

and this statement is called the Gauss Divergence Theorem.

[36] The statement that $dd\omega = 0$ corresponds to the div of a curl being always zero:

$$\nabla \cdot \nabla \times F = 0.$$

We now use the reformulation with differential forms and pullbacks of forms with $\Phi : \mathbb{R}^{n+1} \to \mathbb{R}^{\mathrm{N}}$ with $\mathrm{N} > n+1$ to formulate Stokes' Theorem for integral n-forms over $\Phi(M)$ considered as the boundary of $\Phi(\Omega)$, first for $n+1 = 2$ and $\mathrm{N} = 3$. So let

$$\omega = f_1(x)dx_1 + f_2(x)dx_2 + f_3(x)dx_3 \tag{8.69}$$

and $\Phi : \mathbb{R}^2 \to \mathbb{R}^3$. Then

$$\Phi^*(dx_1) = \frac{\partial \Phi_1}{\partial u_1}du_1 + \frac{\partial \Phi_1}{\partial u_2}du_2; \quad \Phi^*(dx_2) = \frac{\partial \Phi_2}{\partial u_1}du_1 + \frac{\partial \Phi_2}{\partial u_2}du_2;$$

$$\Phi^*(dx_3) = \frac{\partial \Phi_3}{\partial u_1}du_1 + \frac{\partial \Phi_3}{\partial u_2}du_2,$$

and with $\phi_1 = \Phi^* f_1$, $\phi_2 = \Phi^* f_2$, $\phi_3 = \Phi^* f_3$ we have

$$\Phi^*(F) = (\phi_1 \frac{\partial \Phi_1}{\partial u_1} + \phi_2 \frac{\partial \Phi_2}{\partial u_1} + \phi_3 \frac{\partial \Phi_3}{\partial u_1})\, du_1 + (\phi_1 \frac{\partial \Phi_1}{\partial u_2} + \phi_2 \frac{\partial \Phi_2}{\partial u_2} + \phi_3 \frac{\partial \Phi_3}{\partial u_2})\, du_2$$

$$= p_1(u_1, u_2)du_1 + p_2(u_1, u_2)du_2 = \zeta,$$

a 1-form that can be integrated over $M = \partial\Omega$, and to which (8.43) applies, whence

$$\int_{\partial\Omega} \zeta = \int_{\partial\Omega} p_1(u_1, u_2)du_1 + p_2(u_1, u_2)du_2 = \int_\Omega (\frac{\partial p_2}{\partial u_1} - \frac{\partial p_1}{\partial u_2})du_1 du_2 = \int_\Omega d\zeta. \tag{8.70}$$

Observe that the second equality in (8.70) holds in view of (8.43), which is a rewritten version of (8.68) with $\mathrm{N} = 2$, while the first and the third merely substitute $\omega = p_1 du_1 + p_2 du_2$ and evaluate $d\omega$ according to (8.61).

We need

$$\int_{\partial\Omega} \zeta = \int_{\partial\Omega} \Phi^*\omega = \int_{\Phi(\partial\Omega)} \omega, \tag{8.71}$$

and

$$\int_\Omega d\zeta = \int_\Omega d\, \Phi^*\omega = \int_\Omega \Phi^*(d\omega) = \int_{\phi(\Omega)} d\omega \tag{8.72}$$

to conclude for ω given by (8.69) that

$$\int_{dS} \omega = \int_{dS} f_1 dx_1 + f_2 dx_2 + f_3 dx_3 = \int_S d\omega, \tag{8.73}$$

in which $S = \Phi(\Omega)$. It is the last equality in each of (8.71) and (8.72) that has to be checked, the other equalities follow from our d-algebra and the commutation of d and Φ^*.

Let us once more spell out the d-algebra by which (8.61) evaluates as

$$d\omega = (\frac{\partial f_1}{\partial x_1}dx_1 + \frac{\partial f_1}{\partial x_2}dx_2 + \frac{\partial f_1}{\partial x_3}dx_3)dx_1$$

$$+(\frac{\partial f_2}{\partial x_1}dx_1 + \frac{\partial f_2}{\partial x_2}dx_2 + \frac{\partial f_2}{\partial x_3}dx_3)dx_2$$

$$+(\frac{\partial f_3}{\partial x_1}dx_1 + \frac{\partial f_3}{\partial x_2}dx_2 + \frac{\partial f_3}{\partial x_3}dx_3)dx_3 =$$

$$\frac{\partial f_1}{\partial x_2}dx_2dx_1 + \frac{\partial f_2}{\partial x_1}dx_1dx_2 + \frac{\partial f_1}{\partial x_3}dx_3dx_1 + \frac{\partial f_3}{\partial x_1}dx_1dx_3 + \frac{\partial f_2}{\partial x_3}dx_2 + \frac{\partial f_3}{\partial x_2}dx_2dx_3$$

$$= (\frac{\partial f_2}{\partial x_1} - \frac{\partial f_1}{\partial x_2})dx_1dx_2 + (\frac{\partial f_3}{\partial x_2} - \frac{\partial f_2}{\partial x_3})dx_2dx_3 + (\frac{\partial f_1}{\partial x_3} - \frac{\partial f_3}{\partial x_1})dx_3dx_1$$

$$(\frac{\partial f_3}{\partial x_2} - \frac{\partial f_2}{\partial x_3})dx_2dx_3 + (\frac{\partial f_1}{\partial x_3} - \frac{\partial f_3}{\partial x_1})dx_3dx_1 + (\frac{\partial f_2}{\partial x_1} - \frac{\partial f_1}{\partial x_2})dx_1dx_2$$

$$= g_1dx_2dx_3 + g_2dx_3dx_1 + g_3dx_1dx_2.$$

Comparing to (8.40) we recognise for $F(x) = f_1(x)e_1 + f_2(x)e_2 + f_3(x)e_3$ that

$$\int_{dS} f_1dx_1 + f_2dx_2 + f_3dx_3 =$$

$$\int_S (\frac{\partial f_3}{\partial x_2} - \frac{\partial f_2}{\partial x_3})dx_2dx_3 + (\frac{\partial f_1}{\partial x_3} - \frac{\partial f_3}{\partial x_1})dx_3dx_1 + (\frac{\partial f_2}{\partial x_1} - \frac{\partial f_1}{\partial x_2})dx_1dx_2$$

$$= \int_S G \cdot \nu = \int_S (\nabla \times F) \cdot \nu, \tag{8.74}$$

in which $g_1(x)e_1 + g_2(x)e_2 + g_3(x)e_3 = G(x) = \nabla \times F$ and ν is the normal vector on $S = \Phi(\Omega)$ defined by (8.37).

It thus remains to check the two analytical statements

$$\int_{\partial\Omega} \Phi^*\omega = \int_{\Phi(\partial\Omega)} \omega \quad \text{and} \quad \int_\Omega \Phi^*(d\omega) = \int_{\Phi(\Omega)} d\omega, \tag{8.75}$$

which complement the d-algebra presented above, and which are both of the form

$$\int_M \Phi^*\omega = \int_{\Phi(M)} \omega, \tag{8.76}$$

with respectively $M = \partial\Omega$ and $M = \Omega$. For this we need again Section 6.9.4 combined with the usual localisations via partitions of unity. Not very hard but still to be done.

It will be convenient here to have $\Phi(M)$ described by compositions of Φ and patches of M, see the remark at the end of Section 8.19. Also, we still have to deal with integrals over manifolds with boundaries, to obtain

$$\int_{\Phi(\partial\Omega)} \omega = \int_{\Phi(\Omega)} d\omega, \tag{8.77}$$

as the final result in which $M = \Phi(\Omega)$ is a manifold with boundary $\partial M = \Phi(\partial\Omega)$, with $\Omega \in \mathbb{R}^n$ as described at the beginning of this section, Φ a continuously differentiable injective map from $\bar{\Omega}$ to $\mathbb{R}^N$ with Jacobian matrix of rank n throughout $\bar{\Omega}$, and ω an n-form with continuously differentiable coefficients. Generalisations to piecewise C^1-boundaries then still have to be discussed.

8.14 More exercises

Laat

$$\Omega = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1\}. \tag{8.78}$$

Op $\partial\Omega$ is ν de normaalvector met componenten $\nu_x = x$ en $\nu_y = y$ in de notatie zoals in de uitleg in Sectie 8.1 en de route van (8.7) naar (8.9). Met $y = \sqrt{1-x^2}$, $y = -\sqrt{1-x^2}$, $x = \sqrt{1-y^2}$, en $x = -\sqrt{1-y^2}$ hebben we vier grafieken van functies f zoals in (8.5) and (8.6), op te kiezen domeinen, bijvoorbeeld $[-a, a]$ met $\frac{1}{\sqrt{2}} < a < 1$ if we think in terms of (8.32), or $a = \frac{1}{\sqrt{2}}$ if we think in terms of (8.34).

The graph parameterisations

$$\begin{aligned} &x \to (x, \sqrt{1-x^2}), \quad x \to (x, -\sqrt{1-x^2}), \\ &y \to (\sqrt{1-y^2}, y), \quad y \to (-\sqrt{1-y^2}, y) \end{aligned} \tag{8.79}$$

may look uglier than

$$\phi \to (\cos\phi, \sin\phi), \tag{8.80}$$

which requires only two ϕ-domains, e.g. $[-\frac{3\pi}{4}, \frac{3\pi}{4}]$ and $[\frac{\pi}{4}, \frac{7\pi}{4}]$ for (8.32), and $[-\frac{\pi}{2}, \frac{\pi}{2}]$ and $[\frac{\pi}{2}, \frac{3\pi}{2}]$ for (8.34).

Parameterisations obtained from substitutions like $y = tx$ in the defining equation $x^2 + y^2 = 1$ for $\partial\Omega$ are also handy: from $x^2 + t^2x^2 = 1$ we have

$$x = \frac{1}{\sqrt{1+t^2}};\ y = \frac{t}{\sqrt{1+t^2}} \quad \text{and} \quad x = -\frac{1}{\sqrt{1+t^2}};\ y = -\frac{t}{\sqrt{1+t^2}}$$

parameterising the two semicircles on the left and on the right, and likewise $x = ty$ for the upper and lower semicircle, with t running from $-\infty$ to $+\infty$, and you can put

$$t = \frac{s}{1-s} \quad \text{or} \quad t = -\frac{s}{1-s} \tag{8.81}$$

in each of them to obtain.

$$x = \frac{1-s}{\sqrt{1-2s+2s^2}};\ y = \frac{s}{\sqrt{1-2s+2s^2}}$$

parameterising $\{(x,y) \in \mathbb{R}^2 :\ x \geq 0,\ y \geq 0,\ x^2+y^2 = 1\}$ with $s \in [0,1]$.

Opgave 8.5. Use the t-parameterisations above to calculate the area of the unit disk via integrals such as $\int x dy$ of $\int y dx$ over $\partial\Omega$. You should get and evaluate integrands[37] like

$$\frac{t^2}{(1+t^2)^2} = \frac{1}{1+t^2} - \frac{1}{(1+t^2)^2}.$$

Opgave 8.6. Refering to (8.4) and the subsequent line integral notation with 1-forms, consider the form

$$\omega = (a_{20}x^2 + a_{11}xy + a_{02}y^2)dx + (b_{20}x^2 + b_{11}xy + b_{02}y^2)dy$$

and evaluate $\int_{\partial\Omega} \omega$ for $\Omega = \{(x,y) \in \mathbb{R}^2 :\ x^2+y^2 < 1\}$ with $\partial\Omega$ parameterised such that (8.42) defines a vector pointing out of Ω.

Opgave 8.7. Same as Exercise 8.6 but with

$$\omega = (a_{30}x^3 + a_{21}x^2y + a_{12}xy^2 + a_{03}y^3)dx + (b_{30}x^3 + b_{21}x^2y + b_{12}xy^2 + b_{03}y^3)dy$$

Which coefficients disappear in the calculations? Generalise to the obvious n^{th} order case.

Opgave 8.8. In physics results like (8.3) are usually taken for granted in view of the trivial case that

$$\Omega = (a,b) = (a_1,b_1) \times (a_2,b_2)$$

is a rectangle parallel to the axes[38]. Verify directly that (8.3) holds for $v :\ [a,b] \to \mathbb{R}^2$ continuously differentiable.

[37] Recall $\int_{-\infty}^{\infty} \frac{1}{1+t^2}\,dt = \pi$, $\int_{-\infty}^{\infty} \frac{1}{(1+t^2)^2}\,dt = \frac{\pi}{2}$, $\int_{-\infty}^{\infty} \frac{1}{(1+t^2)^3}\,dt = \frac{3\pi}{8}$, ...

[38] https://www.quora.com/What-is-the-plural-of-axis

Note that Ω in Exercise 8.8 is given by

$$F(x_1, x_2) = (x_1 - a_1)(x_1 - b_1)(x_2 - a_2)(x_2 - b_2) < 0,$$

which has a zero gradient in the corners of Ω, just as (8.84) for Exercise 8.10 below. Here is an example without an F to define Ω.

Opgave 8.9. Suppose that the boundary of a bounded open set $\Omega \subset \mathbb{R}^2$ is given by a periodic solution of a system of differential equations $\dot{x} = P(x, y)$ and $\dot{y} = Q(x, y)$, with $P, Q : \mathbb{R}^2 \to \mathbb{R}$ continuously differentiable on $\bar{\Omega}$. Show that

$$\iint_\Omega (\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y})\, dxdy = 0.$$

Edwards has a nice exercise about Descartes' Folium from which I lifted the $y = tx$-trick above. It allows to find the solutions of

$$F(x, y) = x^3 + y^3 - 3xy = 0, \tag{8.82}$$

in the form

$$x = x(t) = \frac{3t}{1 + t^3};\ y = y(t) = \frac{3t^2}{1 + t^3}, \tag{8.83}$$

with $t \in (0, \infty)$, $t \in (-1, 0)$ and $t \in (-\infty, -1)$ giving the smooth parts of the curve. The origin $(0, 0)$ is the intersection of two solution curves, one given by (8.83) with $t \in (-1, 1)$, the other by (8.83) with x and y interchanged. Exercise 2.3 in Chapter V of Edwards is about

$$\Omega = \{(x, y) \in \mathbb{R}^2 :\ x > 0,\ y > 0,\ F(x, y) = x^3 + y^3 - 3xy < 0\}. \tag{8.84}$$

with $\partial\Omega$ given by (8.83) and $t \in [0, \infty)$. You should examine the graphs of x and y as functions of t in (8.83). You can get the area of Ω as

$$-\int_0^\infty y(t)x'(t)\, dt = \int_0^\infty x(t)y'(t)\, dt, \tag{8.85}$$

or the average of the two integrals, which may turn out to be easier, using Green's Theorem the way we derived it. Edwards tells you to cut the folium along the diagonal $y = x$, in which case you have the boundary consisting of two curves, the part described by (8.83) with $0 \leq t \leq 1$, and the diagonal part given by $y = x = t$ with $0 \leq t \leq \frac{3}{2}$, which you should parameterise as $y = x = \frac{3}{2} - t$ if you think about it. Still, I wonder whether Edwards actually did the exercise:

Opgave 8.10. Substitute $y = t^{\frac{1}{3}}x$ in the equation for the folium to get x and y in terms of t and evaluate (8.85) above to obtain the value $\frac{3}{2}$ for the area of $\{(x, y) \in \mathbb{R}^2 : x > 0,\, y > 0,\, x^3 + y^3 - 3xy < 0\}$.

Opgave 8.11. As Exercise 8.10 but use (8.81) to get the boundary parameterised with $0 \leq s \leq 1$.

In the last exercise you see that the boundary of (8.84) is actually given by one single parameterisation with the parameter s in the unit interval $[0, 1]$, with $s = 0$ and $s = 1$ both mapped to the origin where the condition for the local description as used in Section 8.1 fails. The same issue occurs in the trivial case of Exercise 8.8.

Note that (8.84) is a special case of an obvious general question with two parameters, these being $p = 3$ and $n = 2$ here[39]. Dropping the coefficient of xy we have for general $p > 2$ that

$$x = s^{\frac{1}{p(p-2)}}\,(1-s)^{\frac{p-1}{p(p-2)}}\;;\; y = s^{\frac{p-1}{p(p-2)}}\,(1-s)^{\frac{1}{p(p-2)}}\,, \tag{8.86}$$

parameterises the loop in the solution set of $x^p + y^p = xy$, with

$$x\frac{dy}{ds} - y\frac{dx}{ds} = \frac{1}{p}\,s^{\frac{1}{p-2}-1}(1-s)^{\frac{1}{p-2}-1}, \tag{8.87}$$

which looks much better than the individual terms $x\frac{dy}{ds}$ and $y\frac{dx}{ds}$. With the β-function[40] defined by

$$B(x, y) = \int_0^1 s^{x-1}(1-s)^{y-1}\,ds,$$

the area surrounded by $[0, 1] \ni s \to (x(s), y(s))$, the loop in

$$x^p + y^p = xy \tag{8.88}$$

is thus equal to

$$A_p = \frac{1}{2p}\,B(\frac{1}{p-2}, \frac{1}{p-2}), \tag{8.89}$$

which gives $\frac{1}{6}$ for $p = 3$ and differs from Exercise 8.10 by a factor 3^2, consistent with (8.83).

[39] See Exercise 8.18.

[40] More on the β-function in [HM].

Note that in deriving (8.87) from (8.86) you may get lost if you don't introduce

$$\alpha = \frac{1}{p(p-2)} \quad \text{and} \quad \beta = \frac{p-1}{p(p-2)} = (p-1)\alpha$$

and continue your calculations with α and β. I also suggest to write derivatives such as

$$\frac{d}{ds}\, s^\alpha (1-s)^\beta = (\frac{\alpha}{s} - \frac{\beta}{1-s})\, s^\alpha (1-s)^\beta = (\alpha - (\alpha+\beta)s)\, s^{\alpha-1}(1-s)^{\beta-1},$$

which will help you to factor out common factors when such expressions have to be combined later on, as you will notice if you tackle this question: how about the volume V_p in $\{(x,y,z) \in \mathbb{R}^3 : x \geq 0,\, y \geq 0,\, z \geq 0\}$ surrounded by $x^p + y^p + z^p = xyz$ when $p > 3$?

Opgave 8.12. Substitute $y = s^{\frac{1}{p}}x$ and $z = t^{\frac{1}{p}}x$ in $x^p + y^p + z^p = xyz$ to obtain a parameterisation of the solutions with $x, y, z > 0$ in the form

$$x = s^\alpha t^\alpha (1+s+t)^{-p\alpha}, \quad y = s^{(p-2)\alpha} t^\alpha (1+s+t)^{-p\alpha}, \quad z = s^\alpha t^{(p-2)\alpha} (1+s+t)^{-p\alpha},$$

and evaluate

$$xdydz = x(\frac{\partial y}{\partial s}\frac{\partial z}{\partial t} - \frac{\partial y}{\partial t}\frac{\partial z}{\partial s})$$

as xyz times a factor that you have to computer carefully, to find the correct double integral in s and t that gives the desired volume. The integral is the difference of two similar terms each of which is st to some power times $(1+s+t)$ to some power. Substituting $t = (1+s)x$ both integrals reduce to products of single integrals that reduce to β-functions again.

Just in case, I arrived via

$$xyz = \frac{(st)^{\frac{1}{p-3}}}{(1+s+t)^{\frac{3}{p-3}}}$$

and

$$\frac{1}{yz}(\frac{\partial y}{\partial s}\frac{\partial z}{\partial t} - \frac{\partial y}{\partial t}\frac{\partial z}{\partial s}) = \frac{1}{p^2(p-3)st}\left(\frac{p}{1+s+t} - 1\right)$$

at

$$\frac{1}{p(p-3)} \underbrace{\int_0^\infty \int_0^\infty \frac{(st)^{\frac{1}{p-3}-1}\, dsdt}{(1+s+t)^{\frac{p}{p-3}}}}_{S(\frac{1}{p-3},\frac{p}{p-3})} - \frac{1}{p^2(p-3)} \underbrace{\int_0^\infty \int_0^\infty \frac{(st)^{\frac{1}{p-3}-1}\, dsdt}{(1+s+t)^{\frac{3}{p-3}}}}_{S(\frac{1}{p-3},\frac{3}{p-3})}.$$

These integrals are known. With

$$B(a,b) = \int_0^1 s^{a-1}(1-s)^{b-1}\,ds$$

we have[41]

$$T(a,b) = \int_0^\infty \frac{s^{a-1}\,ds}{(1+s)^b} = B(a,b-a)$$

and[42]

$$S(a,b) = \int_0^\infty \int_0^\infty \frac{(st)^{a-1}\,dsdt}{(1+s+t)^b} = T(a,b)T(a,b-a),$$

so V_p can be expressed in p via β-functions. It should lead to what we get in Exercise 8.17, which is really nice[43].

Opgave 8.13. How general is the $y = tx$-trick in $\mathbb{R}^2$? Let $F : \mathbb{R}^2 \to \mathbb{R}$ be continuously differentiable, and suppose that $F(x_0, y_0)$ for some $(x_0, y_0) \in \mathbb{R}^2$ with $x_0 \neq 0$. Define t_0 by $y_0 = t_0 x_0$ and apply the implicit function theorem to derive a condition that guarantees the existence of a C^1-solution curve of the form $t \to (x(t), y(t) = (x(t), tx(t))$ defined on an t-interval which has t_0 as an interior point.

Don't forget you want to have nonzero speed, which is a second condition on top of the usual condition from the the implicit function theorem. The latter condition will involve a simple combination of x, y, F_x, F_y in (x_0, y_0) with a clear (but local) geometric interpretation.

Verify that in the end the nonzero speed condition follows from $x \neq 0$ and the condition from the implicit function theorem. Note that if $(x_0, y_0) \neq (0, 0)$ you always realise at least one of $t \to (x(t), y(t) = (x(t), tx(t))$ and $t \to (x(t), y(t) = (ty(t), y(t))$ if this condition is satisfied. Relate your results to polar coordinates.

Opgave 8.14. Verify that computing the area of (8.84) using polar coordinates is even a bigger pain than using the $y = tx$-trick.

Opgave 8.15. In Exercise 8.13 you must have computed the time derivatives of $x(t)$ and $y(t) = tx(t)$. Verify[44] that the derivative of

$$\frac{y(t)}{x(t)}$$

[41] Via $s = \frac{t}{1-t}$, a substitution I avoided for (8.12).

[42] Via $t = (1+s)\tau$.

[43] There were mistakes in an earlier version and then it did not, but now it does.

[44] You should have got $\dot{x} = -\frac{x^2 F_y}{xF_x + yF_y}$, $\dot{y} = \frac{x^2 F_x}{xF_x + yF_y}$

is what it should be, and that the area of such a curve parameterised by $t \in \mathbb{R}$ with $(x(t), y(t) \to (0,0)$ as $t \to 0$ and $t \to \infty$ is given by[45]

$$\frac{1}{2}\int_0^\infty x(t)^2\,dt,$$

and compute again the area in Exercise 8.10 from the formula for $x(t)$ in (8.83).

Opgave 8.16. Verify (8.89) by putting $y = tx$ in (8.88), solve for x, and set $t^p = s$ in the integral you get from Exercise 8.15 and convert to β-functions.

Opgave 8.17. See Exercise 8.15. How would $F(x,y,z) = 0$ lead to

$$\frac{1}{3}\int_0^\infty \int_0^\infty x(s,t)^3\,dsdt?$$

Hint: in relation to

$$x^p + y^p + z^p = xyz$$

and for

$$x = x(s,t) = \left(\frac{st}{1+s^p+t^p}\right)^{\frac{1}{p-3}}$$

this integral is equal to[46]

$$V_p = \frac{1}{3p^2}\,B(\frac{1}{p-3},\frac{1}{p-3})\,B(\frac{1}{p-3},\frac{2}{p-3}),$$

and you might see a pattern emerge.

Opgave 8.18. Let $p > 4$. The 4-dimensional measure of the bounded open set in $\mathbb{R}^4$ with all coordinates positive and bounded by

$$x_1^p + x_2^p + x_3^p + x_4^p = x_1x_2x_3x_4$$

is

$$\frac{1}{4p^3}\,B(\frac{1}{p-4},\frac{1}{p-4})\,B(\frac{1}{p-4},\frac{2}{p-4})\,B(\frac{1}{p-4},\frac{3}{p-4}),$$

[45] Compare this to a similar formula with polar coordinates.

[46] Earlier mistakes have have been corrected....

and likewise[47] for

$$\sum_{j=1}^{n} x_j^p = \Pi_{j=1}^{n} x_j$$

in $\mathbb{R}^n$ for $p > n$.

8.15 Cut-off functions and partitions of unity

This section explains the basic tools for cutting up functions in smaller parts which are localised. This involves two tricks, each of which you can play with.

The first trick concerns an open set $O \subset \mathbb{R}^{\mathrm{N}}$ and a compact subset $K \subset O$ which should be non-empty[48]. Then every $a \in K$ is contained in an open ball B centered at a such that the closed ball with the same center but twice the radius is contained in O. We denote this larger ball by $2B$. Thus we have

$$K \ni a \in B \subset 2B \subset O.$$

These balls cover K and the (sequential) compactness[49] of K implies that K is covered bij finitely many of such balls, i.e.

$$K \subset B_1 \cup \cdots \cup B_k \subset 2B_1 \cup \cdots \cup 2B_k \subset O.$$

On each such ball $2B_i$ we choose a smooth function $\eta_i \in C_c^\infty(2B)$ with $0 \leq \eta_i \leq 1$ and $\eta_i \equiv 1$ on B_i, and we extend these functions[50] to the whole of $\mathbb{R}^{\mathrm{N}}$ by setting $\eta_i \equiv 0$ outside $2B_i$. Then $\eta_i \in C_c^\infty(\mathbb{R}^{\mathrm{N}})$ for $i = 1, \ldots, k$ and a new function $\chi \in C_c^\infty(\mathbb{R}^{\mathrm{N}})$ may be defined by[51]

$$1 - \chi(x) = (1 - \eta_1(x)) \cdots (1 - \eta_k(x)). \tag{8.90}$$

Indeed, if x is not contained in the union of the supports of $\eta_1, \ldots, \eta_k$ then all factors in the right hand side of (8.90) are equal to 1 and hence $\chi(x) = 0$. On the other hand, if x is contained in one of the balls B_i then the corresponding factor in the right hand side of (8.90) is equal to zero making the right hand side vanish whence $\chi(x) = 1$. In particular $\chi(x) \equiv 1$ on K. Moreover, since all factors take values in $[0, 1]$ the same holds for $\chi(x)$, for any $x \in \mathbb{R}^{\mathrm{N}}$. We conclude that

$$\chi \in C_c^\infty(O), \quad \forall x \in \Omega \quad \chi(x) \in [0, 1], \quad \forall x \in K \quad \chi(x) = 1, \tag{8.91}$$

[47] Exponents and dimensions, another story: Chapter 18.

[48] The set K could be the closure of a bounded domain Ω, or its boundary.

[49] This characterisation of compactness was not discussed yet in these notes.

[50] We can use the p-norm to our liking, the choice $p = 2$ allows radially symmetric η_i.

[51] I first saw this elegant trick in Folland's Real Analysis book.

and this is why χ is called a cut-off function for K in O.

The second trick applies the first trick to a finite collection of such sets

$$\emptyset \neq K_1 \subset O_1, \ldots, \emptyset \neq K_m \subset O_m \quad \text{with} \quad \eta_j \in C_c^\infty(O_j) \tag{8.92}$$

cut-off functions as in (8.91). We define $\zeta_j \in C_c^\infty(O_j)$ by

$$\zeta_j(x) = \frac{\chi_j(x)}{\chi_1(x) + \cdots + \chi_m(x)} \tag{8.93}$$

and extend ζ_j to $\mathbb{R}^N$ via $\zeta_j(x) \equiv 0$ outside O_i. Note that below we don't really use the last part of (8.91) as $\chi_j(x) > 0$ for all $x \in K_i$ suffices to obtain the essential properties of the collection $\zeta_1, \ldots, \zeta_m$, which is called a partition of unity. For every x for which one of the $\chi_j(x) > 0$ it follows that

$$\zeta_1(x) + \cdots + \zeta_m(x) = 1. \tag{8.94}$$

Certainly this holds for x in $K_1 \cup \cdots \cup K_m$. On the other hand, outside the union of $O_1, \ldots, O_m$ this sum is by definition equal to zero.

Any function

$$f : K_1 \cup \cdots \cup K_m \to \mathbb{R}$$

splits up via

$$f(x) = f_1 + \cdots + f_m = \zeta_1(x)f(x) + \cdots + \zeta_m(x)f(x),$$

with the smaller parts $f_j = \zeta_j f$ compactly supported in O_j, and ζ_j not harming any smoothness the original function f may enjoy. Adding more K_j to the collection changes the functions ζ_j only via (8.93), with (8.94) remaining valid.

8.16 Partitions of compact manifolds

This section was written before Section 8.18. For Ω and $M = \partial\Omega$ you can jump to the end of this section. We now apply the techniques in Section 8.15 to a non-empty compact set $M \subset \mathbb{R}^N$ for which (C) in Section 6.8.1 applies in a sense we made more precise in Section 8.18 specifying blocks $[\tilde{a}, \tilde{b}] \subset \mathbb{R}^N$ in which the description (A) of Section 6.8.1 can be given, see (8.102). Below we rather choose blocks $[\tilde{a}_i, \tilde{b}_i] \subset \mathbb{R}^n$, given Φ_i as in (8.32). Thus for each $p \in M$ there exists a continuously differentiable injective

$$\Phi_i : [a, b] = [a_1, b_1] \times \cdots \times [a_n, b_n] \to \mathbb{R}^N$$

with $\mathcal{M}(\frac{\partial\Phi}{\partial u}) > 0$ such that $p \in \Phi((\tilde{a}, \tilde{b}))$ for some $[\tilde{a}, \tilde{b}] \subset (a, b)$, and in some open neighbourhood O of the compact set $K = \Phi([\tilde{a}, \tilde{b}])$ it holds that

$$x \in M \iff x \in \Phi((a, b)) \tag{8.95}$$

We would now like to consider the sets $\Phi((\tilde{a}, \tilde{b}))$ as open sets covering M, so that by compactness

$$M \subset \Phi_1((\tilde{a}_1, \tilde{b}_1)) \cup \cdots \cup \Phi_m((\tilde{a}_m, \tilde{b}_m)), \tag{8.96}$$

for some finite collection Φ_j, but clearly the sets $\Phi((\tilde{a}, \tilde{b}))$ are not open[52] in $\mathbb{R}^{\mathrm{N}}$, unless $n = \mathrm{N}$. Nevertheless such a finite subcover exists.

To see this first choose $[\underline{a}, \underline{b}] \subset (\tilde{a}, \tilde{b})$ with $p \in \Phi((\underline{a}, \underline{b}))$ and a suitable open neighbourhood $\underline{O}$ of $\underline{K} = \Phi([\underline{a}, \underline{b}])$ with $\underline{O} \subset O$ to have the characterisation in (8.95) hold for all $x \in \underline{O}$ as well, and such that $\underline{O}$ does not intersect the (compact) image under Φ of the compact set $[a, b] \backslash (\tilde{a}, \tilde{b})$. It then follows that $M \cap \underline{O} \subset \Phi((\tilde{a}, \tilde{b}))$ because Φ is injective.

Varying $p \in M$ the open sets $\underline{O}$ cover M and by compactness there exists a finite collection $O_1, \ldots, O_m$ such that

$$M \subset \underline{O}_1 \cup \cdots \cup \underline{O}_m \subset \Phi_1((\tilde{a}_1, \tilde{b}_1)) \cup \cdots \cup \Phi_m((\tilde{a}_m, \tilde{b}_m)),$$

which is the desired finite covering (8.96) consisting of patches.

We can now put $K_j = \Phi_j([\tilde{a}_j, \tilde{b}_j])$ and the corresponding open neighbourhoods O_j of K_j in which (8.95) characterises the elements of M. The description following (8.32) in Section 8.9 with unit blocks then results from Section 8.15.

We note that we can also have our partition of unity defined using cut-off functions $\chi = \chi(u)$ for $[\tilde{a}, \tilde{b}] \subset (a, b)$, such as the blocks appearing in (8.96), but it is then slightly more complicated to formulate (8.93), because each χ_j is then a function of u. This allows us to deal with manifolds which are not necessarily embedded in $\mathbb{R}^{\mathrm{N}}$.

Finally we observe that any Ω and $M = \partial\Omega$ as in Section 8.1 allow a choice of functions $\zeta_1, \ldots, \zeta_I \in C_c^\infty((a_i, b_i))$ with $0 \leq \zeta_i \leq 1$ and $\zeta_1 + \cdots + \zeta_I \equiv 1$ on a neighbourhood of $\bar{\Omega}$ such that every for every i either $[a_i, b_i] \subset \Omega$ holds, or $P_i = M \cap (a_i, b_i)$ is a patch such as in Section 8.1.

8.17 Changing partitions

We still have to check that the integrals do not depend on the choice of the partitioning functions $\zeta_1, \ldots, \zeta_I$. We observe that (8.33) defines a linear map

[52] Of course they should be open in M.

$$f \xrightarrow{L} \int_M f\, dS_n \tag{8.97}$$

from $X = C(M)$, the space of continuous real valued functions on M, to $\mathbb{R}$. Note that L is bounded in the sense that $|Lf| \leq C|f|_\infty$, just as in Section 2.4, but we will not be using this below[53].

The partition naturally defines linear subspaces

$$X_i = \{\zeta_i f : f \in C(M)\},$$

and the same holds for any other partition of M, given by say $\eta_1, \ldots, \eta_J$, which also defines a linear map

$$f \xrightarrow{K} \int_M f\, dS_n \tag{8.98}$$

via (8.33), and corresponding linear subspaces Y_j. Now let $\underline{\zeta}_{ij} = \zeta_i \eta_j$, with $i = 1, \ldots, I$ and $j = 1, \ldots, J$. Then

$$f = f \sum_{i=1}^{I} \zeta_i = \sum_{i=1}^{I} \zeta_i f = \sum_{i=1}^{I} \zeta_i f \sum_{j=1}^{J} \eta_j = \sum_{i=1}^{I} \sum_{j=1}^{J} \zeta_i \eta_j f, \tag{8.99}$$

whence

$$Lf = L(\sum_{i=1}^{I} \zeta_i f) = \sum_{i=1}^{I} \int_{\Phi_i} \zeta_i f\, dS_n = \sum_{i=1}^{I} \sum_{j=1}^{J} \int_{\Phi_i} \zeta_i \eta_j f\, dS_n,$$

and likewise

$$Kf = \sum_{j=1}^{J} \sum_{i=1}^{I} \int_{\Psi_j} \eta_j \zeta_i f\, dS_n,$$

and thus it remains to show that

$$\int_{\Phi_i} \zeta_i \eta_j f\, dS_n = \int_{\Psi_j} \eta_j \zeta_i f\, dS_n \tag{8.100}$$

The integral on the left is defined via (8.31) as

$$\int_{\Phi_i} \zeta_i \eta_j f dS_n = \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} \zeta_i(\Phi_i(u)) \eta_j(\Psi_j(v)) f(\Phi_i(u)) \mathcal{M}_n(\frac{\partial \Phi_i}{\partial u})\, du_1 \cdots du_n.$$

[53] But we will need it to get rid of the annoying assumption $\Phi \in C^2$ in Section 8.12.

It should be equal to the integral on the right which is defined via (8.31) as

$$\int_{\Psi_j} \eta_j \zeta_i f dS_n = \int_{c_1}^{d_1} \cdots \int_{c_n}^{d_n} \eta_j(\Psi_j(v))\zeta_i(\Phi_i(u)) f(\Psi_j(v)) \mathcal{M}_n(\frac{\partial \Psi_j}{\partial v})\, dv_1 \cdots dv_n.$$

The coordinates v have to be expressed in u and vice versa via coordinate transformations such as the ones in Section 8.19. These were defined in neighbourhoods of a given points $p \in \Phi_i((a,b)) \cap \Psi_j((c,d))$ only. Therefore we need another localisation argument[54] before we can apply Section 6.9.4 to conclude that the two integrals are the same.

8.18 Again: local descriptions of a manifold

Let us be very precise in what we established for the local descriptions as in (A), (B) and (C) of Section 6.8.1, which correspond to (a,b,c) in III.4 of Edwards. Writing $z = (x, y)$ we take as a starting point that $F = F(z)$ is continuously differentiable on a block

$$[a, b] = [a_1, b_1] \times \cdots \times [a_n, b_n] \times [a_{n+1}, b_{n+1}] \times \cdots \times [a_{\mathrm{N}}, b_{\mathrm{N}}] \subset \mathbb{R}^{\mathrm{N}}$$

and that for some $p \in (a, b)$ the derivative $F'(p)$ is of maximal rank. Renaming and relabeling the variables in $z = (x, y)$ we can then arrange for the "partial"derivative $F_y(p)$ to be invertible. Theorem 6.11 then implies that there exists $(\tilde{a}, \tilde{b}) \subset (a, b)$ with $p \in (\tilde{a}, \tilde{b})$ and a continuously differentiable function

$$f : [\tilde{a}_x, \tilde{b}_x] \to (\tilde{a}_y, \tilde{b}_y)$$

such that $p = (p_x, p_y) \in (\tilde{a}, \tilde{b})$ and

$$F^{-1}(p) \cap [\tilde{a}, \tilde{b}] = \{(x, f(x)) : x \in [\tilde{a}_x, \tilde{b}_x]\} \subset [\tilde{a}_x, \tilde{b}_x] \times (\tilde{a}_y, \tilde{b}_y), \qquad (8.101)$$

with subscripts indicating the x and the y-parts of p, $\tilde{a}$ and $\tilde{b}$. Thus in the smaller block $[\tilde{a}, \tilde{b}]$ the level set of $F(p)$ coincides with the graph of f, and in the same block $[\tilde{a}, \tilde{b}]$ this graph then coincides with the zero-level set of $\tilde{F}(z) = \tilde{F}(x, y) = y - f(x)$.

As for (C), if we have, with subscripts denoting the x and the y-parts of Φ, that $\Phi(u) = (\Phi_x(u), \Phi_y(u))$ is continuously differentiable on $[a, b]$ with $0 \in (a, b)$ and $p = \Phi(0)$, then via Theorem 6.12 the invertibility of $\Phi'_x(0)$ is sufficient for the existence of $[\underline{a}_x, \underline{b}_x]$ with $p_x \in (\underline{a}_x, \underline{b}_x)$ and a continuously differentiable function $\phi : [\underline{a}_x, \underline{b}_x] \to (a, b)$ such that $\Phi_x(\phi(x)) = x$ for all

[54] Try this one by yourself.

$x \in [\underline{a}_x, \underline{b}_x]$. Moreover[55], we can choose $[\underline{a}_x, \underline{b}_x]$ such that $\phi((\underline{a}_x, \underline{b}_x))$ is an open set as the inverse image of $(\underline{a}_x, \underline{b}_x)$ under Φ_x.

The function f defined by $f(x) = \Phi_y(\phi(x))$ now defines a graph

$$\{(x, f(x)) : x \in [\underline{a}_x, \underline{b}_x]\}$$

which is a subset of $\Phi([a, b])$. If in addition Φ is injective on $[a, b]$ then the image under Φ of the closed bounded set $[a, b]\backslash\phi((\underline{a}_x, \underline{b}_x))$ is bounded and closed, and does not contain p. Thus there exists a block $[\tilde{a}, \tilde{b}]$ with $p \in (\tilde{a}, \tilde{b})$ such that $[\tilde{a}_x, \tilde{b}_x] \subset (\underline{a}_x, \underline{b}_x)$ with

$$\Phi([a, b]\backslash\phi((\underline{a}_x, \underline{b}_x))) \cap [\tilde{a}, \tilde{b}] = \emptyset.$$

The continuity of f implies that we can restrict $\tilde{a}_x$ and $\tilde{b}_x$ a bit further to ensure that $f([\tilde{a}_x, \tilde{b}_x]) \subset (\tilde{a}_y, \tilde{b}_y)$. We note we also have that

$$\Phi([a, b]\backslash\phi((\tilde{a}_x, \tilde{b}_x))) \cap [\tilde{a}, \tilde{b}] = \emptyset,$$

since the additional points in the larger image $\Phi([a, b]\backslash\phi((\tilde{a}_x, \tilde{b}_x)))$ are on the graph of f outside $[\tilde{a}_x, \tilde{b}_x]$. Thus we have arrived from (C) to exactly the same formulation of (A) as above starting from (B): $p \in (\tilde{a}, \tilde{b})$ and

$$\Phi([a, b] \cap [\tilde{a}, \tilde{b}] = \{(x, f(x)) : x \in [\tilde{a}_x, \tilde{b}_x]\} \subset [\tilde{a}_x, \tilde{b}_x] \times (\tilde{a}_y, \tilde{b}_y). \qquad (8.102)$$

The two statements (8.101) and (8.102) should be compared to the definition Edwards gives in Section 4 of his Chapter III for $M \subset \mathbb{R}^N$ to be an n-dimensional manifold. Every $p \in M$ should, after relabeling and renaming in $z = (x, y)$, be contained in an open set O in which

$$P = O \cap M = \{(x, f(x)) : x \in U\},$$

with $U \subset \mathbb{R}^n$ open and $f : U \to \mathbb{R}^m$ continuously differentiable, is called a C^1-patch of M. Of course it is then clear that $U \supset [\tilde{a}_x, \tilde{b}_x] \supset (\tilde{a}_x, \tilde{b}_x)$ and $O \supset [\tilde{a}, \tilde{b}] \supset (\tilde{a}, \tilde{b})$ for some $(\tilde{a}, \tilde{b}) \ni p$, and thus it is completely equivalent to ask that $p \in (\tilde{a}, \tilde{b})$ and

$$p \in M \cap [\tilde{a}, \tilde{b}] = \{(x, f(x)) : x \in [\tilde{a}_x, \tilde{b}_x]\} \subset [\tilde{a}_x, \tilde{b}_x] \times (\tilde{a}_y, \tilde{b}_y) \qquad (8.103)$$

for some closed block $[\tilde{a}, \tilde{b}]$ with $(\tilde{a}_x, \tilde{b}_x) \ni p_x$, and some continuously differentiable $f : [\tilde{a}_x, \tilde{b}_x] \to (\tilde{a}_y, \tilde{b}_y)$, exactly as in (8.101,8.102), the patch being

$$M \cap (\tilde{a}, \tilde{b}) = \{(x, f(x)) : x \in (\tilde{a}_x, \tilde{b}_x)\} \ni p = (p_x, f(p_x)). \qquad (8.104)$$

In the closed N-block $[\tilde{a}, \tilde{b}]$ there are no other points of M than the points on the graph of $f : [\tilde{a}_x, \tilde{b}_x] \to (\tilde{a}_y, \tilde{b}_y)$.

[55] See the discussion after Theorem 6.12.

8.19 Coordinate transformations

By definition every $p \in M$ is in such a patch as above and typically patches overlap. If p is in two such patches, say with functions f and g, it may happen that f and g are functions of the x-part of z. In that case the patches are parameterised by

$$u \to \Phi(u) = (u, f(u)) \quad \text{and} \quad v \to \Psi(v) = (v, g(v)) \tag{8.105}$$

defined on overlapping blocks with p_x in the interior of the intersection of the blocks, which is an open block itself. The common part of M is then contained in the intersection of the two N-blocks.

Viewing the n-tuples u and v as local coordinates on M near p, a transformation of these coordinates is simply given by $v = u$. In all other cases, we may renumber the variables of $\mathbb{R}^{\mathrm{N}}$ to have the patches parameterised as

$$u \to \Phi(u) = (u_1, u_2, f_3(u_1, u_2), f_4(u_1, u_2));$$

$$v \to \Psi(v) = (v_1, g_2(v_1, v_3), v_3, g_4(v_1, v_3)),$$

with (u_1, u_2) and (v_1, v_3) in some open block in $\mathbb{R}^n$, or as

$$u \to \Phi(u) = (u_1, f_2(u_1), f_3(u_1));$$

$$v \to \Psi(v) = (g_1(v_2), v_2, g_3(v_2)),$$

with u_1 and v_3 in some open block in $\mathbb{R}^n$. Note that the first case above cannot occur if $\mathrm{N} = n + 1$.

To rewrite Ψ in the form Φ we need the invertibility of respectively

$$\frac{\partial g_2}{\partial v_3} \quad \text{and} \quad \frac{\partial g_1}{\partial v_2}, \tag{8.106}$$

in which case we we obtain respectively

$$w \to \tilde{\Phi}(w) = (w_1, w_2, h_3(w_1, w_2), h_4(w_1, w_2))$$

and

$$w \to \tilde{\Phi}(w) = (w_1, h_2(w_1), h_3(w_1))$$

as local descriptions of the Ψ-patches near p. The definition of what a manifold is then implies that

$$\tilde{\Phi} \equiv \Phi$$

on an open block containing (p_1, p_2) in the first case and p_1 in the second case. It then follows as above that $u = w$ is a coordinate transformation just as

$u = v$ for (8.105) while w is obtained from v via a coordinate transformation just as x from u in the proof of (A) from (C) above.

It thus remains to establish the invertibility of the partial Jacobian matrices in (8.106) in p to conclude there exists a local C^1-transformation from u to v near p. Note that these are also the conditions for solving part[56] of $\Phi(u) = \Psi(v)$ via

$$v_1 = u_1,\ v_3 = f_3(u_1, u_2) \quad \text{and} \quad u_1 = v_1,\ u_2 = g_2(v_1, v_3) \tag{8.107}$$

in the first case, and

$$u_1 = g_1(v_2) \quad \text{and} \quad v_2 = f_2(u_1) \tag{8.108}$$

in the second case. The invertiblity of the partial Jacobian matrices in (8.106) in p follows because otherwise the Ψ-patch cannot achieve all respectively (u_1, u_2)-directions and u_1-directions that occur in the Φ-patch, contradicting the assumption that the Ψ-patch covers all of M in its defining neighbourhood.

The restriction to patches of the form (8.104) looks like an obvious choice for simplicity, but may bother us later when dealing with (8.76), we'll see.

[56] All equations but the last one, which then requires some argument to hold as well.

9 Toepassingen met een zuiver toetje

Wat bruggetjes naar hoe de natuurkundigen en scheikundigen het doen, en aan het eind wat complexe functietheorie, met de lijnintegralen alleen maar over rechte lijnstukjes. Voldoende voor de functional calculus waarmee voor z in $f(z)$ ook iets heel anders mag worden ingevuld, bijvoorbeeld een vierkante matrix.

9.1 Integraalrekening in poolcoördinaten

Merk op dat we *in het echte leven* over meer verzamelingen zullen willen integreren dan over rechthoeken. Bijvoorbeeld over heel $\mathbb{R}^2$. Voor niet-negatieve functies $u : \mathbb{R}^2 \to \mathbb{R}$ is

$$\iint_{\mathbb{R}^2} u = \lim_{R\to\infty} \underbrace{\iint_{[-R,R]\times[-R,R]} u(x,y)d(x,y)}_{J(R)} = \lim_{R\to\infty} J(R) \tag{9.1}$$

op natuurlijke manier gedefinieerd in $[0,\infty]$ als limiet van een niet-dalende functie $R \to J(R) \geq 0$.

Er is natuurlijk geen enkele reden om een integraal over heel $\mathbb{R}^2$ per se als een limiet van integralen in rechthoekige coördinaten over in dit geval vierkanten te introduceren. Poolcoördinaten zijn vaak veel handiger. Voor Riemannsommen in poolcoördinaten ten behoeve van de rechtstreekse definitie en uitwerking van

$$\begin{aligned}\iint_{x^2+y^2\leq R^2} u(x,y)d(x,y) &= \int_0^{2\pi}\int_0^R u(r\cos\theta, r\sin\theta)\, r dr\, d\theta \\ &= \int_0^R \int_0^{2\pi} u(r\cos\theta, r\sin\theta)\, d\theta\, r dr \end{aligned} \tag{9.2}$$

gebruiken we

$$0 = r_0 \leq r_1 \leq \cdots \leq r_{\mathrm{M}} = R \quad \text{met} \quad \mathrm{M} \in \mathbb{N} \tag{9.3}$$

en

$$0 = \theta_0 \leq \theta_1 \leq \cdots \leq \theta_{\mathrm{N}} = 2\pi \quad \text{met} \quad \mathrm{N} \in \mathbb{N}, \tag{9.4}$$

en tussensommen van de vorm

$$\sum_{k=1}^{M}\sum_{l=1}^{N} u(\rho_k\cos\phi_l, \rho_k\sin\phi_l)\, \underbrace{\frac{1}{2}(r_k^2 - r_{k-1}^2)(\theta_l - \theta_{l-1})}_{\text{waarom dit dan?}} =$$

$$\sum_{k=1}^{M}\sum_{l=1}^{N} u(\rho_k\cos\phi_l,\rho_k\sin\phi_l)\underbrace{\frac{r_k+r_{k-1}}{2}}_{\tilde{\rho}_k}(r_k-r_{k-1})(\theta_l-\theta_{l-1}),$$

met tussenwaarden $\rho_k, \tilde{\rho}_k \in [r_{k-1}, r_k]$ en $\phi_l \in [\theta_{l-1}, \theta_l]$. De details zijn zelf in te vullen. Leuker is deze mooie toepassing van (9.2) in de volgende stelling over harmonische functies.

Opgave 9.1. Een twee keer continu differentieerbare functie $(x, y) \to u(x, y) = u(r\cos\theta, r\sin\theta)$ heet harmonisch als $\Delta u = 0$. Laat zien dat

$$u(0,0) = \frac{1}{2\pi}\int_0^{2\pi} u(r\cos\theta, r\sin\theta)\, d\theta,$$

en dat harmonische functies dus in elk punt het gemiddelde van hun waarden op een diskvormige omgeving zijn. Hint: gebruik Stelling 4.34 als je de integraal van Δu over $\bar{B}_R$ hebt vertaald naar een integraal met alleen maar $d\theta$.

Ook leuk is dat voor niet-negatieve continue functies $u : \mathbb{R}^2 \to \mathbb{R}$ de integraal

$$\iint_{\mathbb{R}^2} u = \lim_{R\to\infty}\iint_{x^2+y^2\le R^2} u(x,y)d(x,y) \tag{9.5}$$

nu net zo natuurlijk gedefinieerd is in $[0, \infty]$ als door (9.1). Alleen een wiskundige vraagt zich dan af dit consistent is. Dat moet en dat mag hoor:

Opgave 9.2. Voor niet-negatieve continue $u : \mathbb{R}^2 \to \mathbb{R}$ geldt

$$\lim_{R\to\infty}\iint_{x^2+y^2\le R^2} u(x,y)d(x,y) = \lim_{R\to\infty}\iint_{[-R,R]\times[-R,R]} u(x,y)d(x,y).$$

Zoek maar uit waarom en onthoud dat

$$\iint_{\mathbb{R}^2} u$$

op twee (eigenlijk vier) manieren is uit te rekenen als dat nodig is, zie (6.47) en (9.2).

9.2 Integraalrekening en kansverdelingen

In de formule van Stirling, zie (5.20), stond nog een integraal die we nu netjes kunnen uitrekenen met behulp van Opgave 9.2 en de functie

$$(x,y) \xrightarrow{u} e^{-\frac{1}{2}(x^2+y^2)}$$

Kort door de bocht opgeschreven concluderen we dat

$$\left(\int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2}\,dx\right)^2 = \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2}\,dx \int_{-\infty}^{\infty} e^{-\frac{1}{2}y^2}\,dy = \iint_{\mathbb{R}^2} e^{-\frac{1}{2}(x^2+y^2)}\,d(x,y)$$

$$= \int_0^{\infty}\int_0^{2\pi} e^{-\frac{1}{2}r^2}\,rd\theta dr = \int_0^{\infty} 2\pi e^{-\frac{1}{2}r^2}\,rdr = 2\pi[-e^{-\frac{1}{2}r^2}]_0^{\infty} = 2\pi.$$

Opgave 9.3. Laat met Opgave 9.2 zien dat

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2}\,dx = \sqrt{2\pi}.$$

Bijgevolg hebben

$$f(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2} \quad \text{en} \quad u(x,y) = \frac{1}{2\pi}e^{-\frac{1}{2}(x^2+y^2)} \tag{9.6}$$

dus de eigenschap dat ze (positief zijn en) en totale integraal gelijk aan 1 hebben. We noemen zulke functies *kansdichtheden*. De dichtheid $f(x)$ hoort bij een stochastische grootheid X waarvoor geldt dat de kans op uitkomst $X \in [a,b]$ gelijk is aan

$$P(X \in [a,b]) = \int_a^b f(x)\,dx,$$

en

$$P(X \leq x) = \int_{-\infty}^{x} f(s)\,ds$$

wordt de cumulatieve verdelingsfunctie van X genoemd.

Een van X onafhankelijke stochastische grootheid Y kan een kansdichtheid $g(y)$ hebben die beschrijft dat de kans op $Y \in [c,d]$ gelijk is aan

$$P(Y \in [c,d]) = \int_c^d g(y)\,dy.$$

De simultane kansdichtheid $u(x,y) = f(x)g(y)$ geeft dan de kans op $X \in [a,b]$ en $Y \in [c,d]$ als

$$\iint_{\mathbb{R}^2} u = \int_a^b f(x)\,dx \int_c^d g(y)\,dy.$$

De kansdichtheden in (9.6) worden de 1-en 2-dimensionale standaard *normale verdeling* genoemd. Is de functie g hetzelfde als de functie f in (9.6), dan zijn X en Y allebei standaard normaal verdeeld. De twee stochastische grootheden X en Y kunnen op elkaar gedeeld worden. De kans op

$$Q = \frac{Y}{X} \in [a,b]$$

is dan gelijk aan de integraal van $u(x,y)$ over het gebied ingesloten door de lijnen $y = ax$ en $y = bx$.

In het geval dat X en Y standaard normaal verdeeld en onderling onafhankelijk zijn, bestaat die integraal uit twee identieke stukken waarvan er één gegeven wordt door

$$\{(x,y) : x \geq 0,\, ax \leq y \leq bx\},$$

een gebied dat in poolcoördinaten beschreven wordt door θ in een deelinterval van $(-\frac{\pi}{2}, \frac{\pi}{2})$.

We willen concluderen dat

$$P(Q \in [a,b]) = 2\int_0^\infty \int_{ax}^{bx} \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}\,dy\,dx$$

$$= \frac{1}{\pi}\int_0^\infty \int_{\arctan a}^{\arctan b} e^{-\frac{1}{2}r^2}\,r d\theta dr = \frac{1}{\pi}(\arctan b - \arctan a) = \int_a^b \frac{1}{\pi}\frac{1}{1+q^2}\,dq.$$

De stochastische grootheid Q heeft dan een kansdichtheid gegeven door de functie

$$q \to \frac{1}{\pi}\frac{1}{1+q^2}.$$

Opgave 9.4. Hierboven manipuleerden we met meervoudige oneigenlijke integralen over "taartpunten" in $\mathbb{R}^2$. De daarvoor benodigde theorie vraagt om een uitbreiding van de theorie van integralen over het hele vlak in poolcoördinaten. Dat kun je ook zelf proberen precies te maken nu.

9.3 Gradient, kettingregel, coördinatentransformaties

De *kettingregel* generaliseert de regel in Opgave 5.10, zie ook Sectie 4.6, Opgave 5.11 en later Opgave 10.19. Met de opmerking dat (5.3) gelezen moet worden met matrices[1] is de regel met bewijs en al over te schrijven en nu meteen toepasbaar.

We spellen een en ander nu uit in het geval van coördinatentransformaties, met als belangrijk voorbeeld de overgang op poolcoördinaten die we al gebruikten om $\mathbb{C}$ te beschrijven en in $\mathbb{C}$ te rekenen: ieder punt $(x, y) \in \mathbb{R}^2$ kunnen we via

$$x = r\cos\theta \quad \text{en} \quad y = r\sin\theta \tag{9.7}$$

zien als gegeven door poolcoördinaten $r, \theta \in \mathbb{R}$ voor $(x, y) \neq (0, 0)$.

Een differentieerbare scalaire functie $F(x, y)$ van x en y is zo automatisch ook een differentieerbare functie van r en θ. In wat volgt zien we (9.7) als transformatie

$$Z : \mathbb{R}^2 \to \mathbb{R}^2$$

van de onafhankelijke plaatsvariabelen, en $F(x, y) = F(Z(r, \theta))$ als de *afhankelijke* variabele. Buiten de wiskunde, met name in de natuurkunde, is het gebruikelijk om de afhankelijke variabele met hetzelfde symbool te noteren als alleen de onafhankelijke variabelen worden getransformeerd.

9.3.1 Gradient, divergentie en Laplaciaan

Voor $F : \mathbb{R}^2 \to \mathbb{R}$ is de definitie van differentieerbaarheid in de gewone rechthoekige coördinaten x en y en $h = x - x_0$, $k = y - y_0$ te lezen als

$$F(x_0 + h, y_0 + k) = F(x_0, y_0) + ah + bk + R(h, k; x_0, y_0), \tag{9.8}$$

met $a, b \in \mathbb{R}$ en

$$\frac{R(h, k; x_0, y_0)}{\sqrt{h^2 + k^2}} \to 0 \quad \text{als} \quad \sqrt{h^2 + k^2} \to 0, \tag{9.9}$$

zie ook Sectie 4.6 en vergelijk met het uitpakken van (6.9) hierboven. De volgende opgave is misschien nu wat dubbelop, maar dat kan geen enkel kwaad.

Opgave 9.5. Neem voor $F : \mathbb{R}^2 \to \mathbb{R}$, $(x_0, y_0) \in \mathbb{R}^2$ en $a, b \in \mathbb{R}$ aan dat (9.8) geldt met (9.9). Dan volgt dat

$$\frac{F(x_0 + h, y_0) - F(x_0, y_0)}{h} \to a \quad \text{en} \quad \frac{F(x_0, y_0 + k) - F(x_0, y_0)}{k} \to b$$

[1] Beter: lineaire afbeeldingen, in dit hele hoofdstuk de facto matrices.

als $h, k \to 0$. Laat dit zien.

Meerdere notaties worden gebruikt, zoals

$$a = F_x(x_0, y_0) = \frac{\partial F}{\partial x}(x_0, y_0) = (\delta_x F)(x_0, y_0) = (D_1 F)(x_0, y_0); \tag{9.10}$$

$$b = F_y(x_0, y_0) = \frac{\partial F}{\partial y}(x_0, y_0) = (\delta_y F)(x_0, y_0) = (D_2 F)(x_0, y_0), \tag{9.11}$$

waarbij (x_0, y_0) en haakjes vaak worden weggelaten want

$$a = F_x = \frac{\partial F}{\partial x} = \delta_x F = D_1 F \quad \text{en} \quad b = F_y = \frac{\partial F}{\partial y} = \delta_y F = D_2 F$$

ziet er gewoon fijner uit.

Als kolomvector schrijven we ook, met

$$e_x = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{en} \quad e_y = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{9.12}$$

$$\begin{pmatrix} a \\ b \end{pmatrix} = \nabla F = \frac{\partial F}{\partial x} e_x + \frac{\partial F}{\partial y} e_y = e_x \frac{\partial F}{\partial x} + e_y \frac{\partial F}{\partial y}, \tag{9.13}$$

de *gradient* van F in (x_0, y_0), geschreven zonder (x_0, y_0). Merk op dat het lineaire gedeelte in (9.8) te schrijven is als

$$ah + bk = \begin{pmatrix} a \\ b \end{pmatrix} \cdot \begin{pmatrix} h \\ k \end{pmatrix} = \begin{pmatrix} h \\ k \end{pmatrix} \cdot \begin{pmatrix} a \\ b \end{pmatrix} = h \frac{\partial F}{\partial x} + k \frac{\partial F}{\partial y}, \tag{9.14}$$

het inprodukt[2] van ∇F en de verschilvector $\binom{h}{k}$.

We zien dus hoe de gradiënt de vector is die de lineaire afbeelding $DF : \mathbb{R}^2 \to \mathbb{R}$ via het inprodukt representeert als

$$\begin{pmatrix} h \\ k \end{pmatrix} \xrightarrow{DF} \nabla F \cdot \begin{pmatrix} h \\ k \end{pmatrix},$$

maar ook dat (9.14) te lezen is als de *differentiaaloperator*

$$h \frac{\partial}{\partial x} + k \frac{\partial}{\partial y} \quad \text{werkend op} \quad F.$$

[2] Het inprodukt van twee vectoren in $\mathbb{R}^2$ wordt gegeven door $\binom{a}{b} \cdot \binom{h}{k} = ah + bk$.

Evenzo zien we ∇ als

$$\nabla = e_x \frac{\partial}{\partial x} + e_y \frac{\partial}{\partial y} \quad \text{werkend op} \quad F \quad \text{geeft} \quad \nabla F, \tag{9.15}$$

een vectorwaardige differentiaaloperator.

Middels het inprodukt kan ∇ ook werken op een vectorwaardige differentieerbare functie

$$(x, y) \to \begin{pmatrix} V_x(x, y) \\ V_y(x, y) \end{pmatrix} = \begin{pmatrix} V_x \\ V_y \end{pmatrix} = V_x e_x + V_y e_y,$$

en wel als

$$\nabla \cdot V = (e_x \frac{\partial}{\partial x} + e_y \frac{\partial}{\partial y}) \cdot (V_x e_x + V_y e_y) = \frac{\partial V_x}{\partial x} + \frac{\partial V_y}{\partial y}, \tag{9.16}$$

de *divergentie* van V.

We schrijven hier nu V met subscripten[3] x, y voor de x- en y-coördinaten V_x en V_y van V t.o.v. de orthonormale vectoren (9.12) die samen de standaardbasis van $\mathbb{R}^2$ vormen. Merk wel op dat V_x en V_y van x en y afhangen maar e_x en e_y niet. De indices x en y staan voor de x-richting en de y-richting, en die richtingen zijn overal in het x, y-vlak hetzelfde.

Elk van de twee termen in ∇ werkt nu alleen op V_x en V_y, en omdat

$$e_x \cdot e_x = e_y \cdot e_y = 1 \quad \text{en} \quad e_x \cdot e_y = e_y \cdot e_x = 0, \tag{9.17}$$

blijven er maar twee termen over in (9.16). Omdat e_x en e_y niet van x en y afhangen geeft elk van de vier termen

$$e_x \frac{\partial}{\partial x} \cdot V_x e_x, \quad e_x \frac{\partial}{\partial x} \cdot V_y e_y, \quad e_y \frac{\partial}{\partial y} \cdot V_x e_x, \quad e_y \frac{\partial}{\partial x} \cdot V_y e_y$$

die we krijgen bij het uitwerken van (9.16) maar één term, te weten

$$e_x \frac{\partial}{\partial x} \cdot V_x e_x = e_x \frac{\partial}{\partial x} \cdot V_x e_x = e_x \cdot \frac{\partial}{\partial x} V_x e_x = e_x \cdot \frac{\partial V_x}{\partial x} e_x = \frac{\partial V_x}{\partial x} e_x \cdot e_x = \frac{\partial V_x}{\partial x}$$

voor de eerste,

$$e_x \frac{\partial}{\partial x} \cdot V_y e_y = e_x \frac{\partial}{\partial x} \cdot V_y e_y = e_x \cdot \frac{\partial}{\partial x} V_y e_y = e_x \cdot \frac{\partial V_x}{\partial x} e_y = \frac{\partial V_x}{\partial y} e_x \cdot e_y = 0$$

voor de tweede, en

$$e_y \frac{\partial}{\partial y} \cdot V_x e_x = 0, \quad e_y \frac{\partial}{\partial x} \cdot V_y e_y = \frac{\partial V_y}{\partial y}$$

[3]Niet te verwarren met het gebruik van subscripten voor partiële afgeleiden!

voor de derde en vierde. Van de vier termen worden er dus nog twee nul vanwege $e_x \cdot e_y = 0$ in (9.17) en de andere twee vereenvoudigen en blijven in die vorm over in (9.16).

Als $V = \nabla F$ differentieerbaar is dan volgt zo dat

$$\nabla \cdot \nabla F = (e_x \frac{\partial}{\partial x} + e_y \frac{\partial}{\partial y}) \cdot (\frac{\partial F}{\partial x} e_x + \frac{\partial F}{\partial y} e_y) = \frac{\partial}{\partial x}\frac{\partial F}{\partial x} + \frac{\partial}{\partial y}\frac{\partial F}{\partial y} = \Delta F, \quad (9.18)$$

de Laplaciaan van F, die weer gezien kan worden als

$$\Delta F \quad \text{is} \quad \Delta = \frac{\partial}{\partial x}\frac{\partial}{\partial x} + \frac{\partial}{\partial y}\frac{\partial}{\partial y} = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \quad \text{werkend op} \quad F. \quad (9.19)$$

Omschrijven van gradiënt, divergentie en Laplaciaan naar poolcoördinaten is nu een nuttige oefening waarvoor de volgende subsecties van belang zijn. Het is handig om daarbij naar twee net iets anders uitgewerkte notaties voor de kettingregel te kijken, zie ook Sectie 4.6 waar dat voor $t \to F(x(t), y(t)$ al is gedaan.

9.3.2 Kettingregel uitgeschreven voor transformaties

We weten dat we de kettingregel toe mogen passen op

$$(r, \theta) \to (r\cos\theta, r\sin\theta) = (X(r,\theta), Y(r,\theta)) = (x, y) \to F(x, y) = G(r, \theta),$$

door de lineaire benadering van

$$(r, \theta) \to (r\cos\theta, r\sin\theta)$$

rond (r_0, θ_0) in te vullen in de lineaire benadering van

$$(x, y) \to F(x, y)$$

rond (x_0, y_0). We doen dit nu met $\tilde{h} = r - r_0$ en $\tilde{k} = \theta - \theta_0$, met weglating van (r_0, θ_0) in de partiële afgeleiden.

Omdat we in deze sectie $F(x, y) = G(r, \theta)$ als onbekende afhankelijke grootheid willen zien, bijvoorbeeld de oplossing van een partiële differentiaalvergelijking, kiezen we nu eerst voor de schrijfwijze zoals rechts in (9.14). De lineaire termen in de expansies

$$X(r_0 + \tilde{h}, \theta_0 + \tilde{k}) = X(r_0, \theta_0) + \tilde{h}\frac{\partial X}{\partial r} + \tilde{k}\frac{\partial X}{\partial \theta} + \cdots,$$

$$Y(r_0 + \tilde{h}, \theta_0 + \tilde{k}) = Y(r_0, \theta_0) + \tilde{h}\frac{\partial Y}{\partial r}\tilde{h} + \tilde{k}\frac{\partial Y}{\partial \theta} + \cdots$$

moeten dan als

$$h = \tilde{h}\frac{\partial X}{\partial r} + \tilde{k}\frac{\partial X}{\partial \theta} \quad \text{en} \quad k = \tilde{h}\frac{\partial Y}{\partial r} + \tilde{k}\frac{\partial Y}{\partial \theta}$$

in (9.14) worden ingevuld[4], en het resultaat

$$(\tilde{h}\frac{\partial X}{\partial r} + \tilde{k}\frac{\partial X}{\partial \theta})\frac{\partial F}{\partial x} + (\tilde{h}\frac{\partial Y}{\partial r} + \tilde{k}\frac{\partial Y}{\partial \theta})\frac{\partial F}{\partial y} =$$

$$\tilde{h}(\frac{\partial X}{\partial r}\frac{\partial F}{\partial x} + \frac{\partial Y}{\partial r}\frac{\partial F}{\partial y}) + \tilde{k}(\frac{\partial X}{\partial \theta}\frac{\partial F}{\partial x} + \frac{\partial Y}{\partial \theta}\frac{\partial F}{\partial y})$$

is dan volgens de kettingregel gelijk aan

$$\tilde{h}\frac{\partial G}{\partial r} + \tilde{k}\frac{\partial G}{\partial \theta}.$$

Er volgt dus dat

$$\frac{\partial G}{\partial r} = \frac{\partial X}{\partial r}\frac{\partial F}{\partial x} + \frac{\partial Y}{\partial r}\frac{\partial F}{\partial y} \tag{9.20}$$

$$\frac{\partial G}{\partial \theta} = \frac{\partial X}{\partial \theta}\frac{\partial F}{\partial x} + \frac{\partial Y}{\partial \theta}\frac{\partial F}{\partial y}, \tag{9.21}$$

in vector-matrixnotatie te schrijven als

$$\begin{pmatrix} \frac{\partial G}{\partial r} \\ \frac{\partial G}{\partial \theta} \end{pmatrix} = \begin{pmatrix} \frac{\partial X}{\partial r}\frac{\partial F}{\partial x} + \frac{\partial Y}{\partial r}\frac{\partial F}{\partial y} \\ \frac{\partial X}{\partial \theta}\frac{\partial F}{\partial x} + \frac{\partial Y}{\partial \theta}\frac{\partial F}{\partial y} \end{pmatrix} = \begin{pmatrix} \frac{\partial X}{\partial r} & \frac{\partial Y}{\partial r} \\ \frac{\partial X}{\partial \theta} & \frac{\partial Y}{\partial \theta} \end{pmatrix} \begin{pmatrix} \frac{\partial F}{\partial x} \\ \frac{\partial F}{\partial y} \end{pmatrix}, \tag{9.22}$$

waarin we links een 2 bij 1 matrix zien met de partiële afgeleiden van G, en rechts net zo'n matrix voor F, en een 2 bij 2 matrix voor

$$(r, \theta) \xrightarrow{Z} (X(r, \theta), Y(r, \theta)),$$

met $Z : \mathbb{R}^2 \to \mathbb{R}^2$ via (9.7) gedefinieerd door

$$Z(r, \theta) = (X(r, \theta), Y(r, \theta)) = (r \cos\theta, r \sin\theta).$$

Horizontaal worden deze matrices genummerd met de variabele grootheden in het beeld, verticaal met die in het domein van de betreffende afbeelding. Precies andersom als in (6.10) dus, omdat we de schrijfwijze rechts in (9.14) hebben gebruikt. Transponeren geeft natuurlijk de vorm die consistent is met (5.3).

De kolomvectoren in (9.22) zien er uit als gradiënten, maar dat is slechts misleidende schijn, zoals we in Sectie 9.3.4 zullen zien.

[4] We gaan er nu niet echt vanuit dat de lezer al met matrices heeft leren rekenen.

9.3.3 Kettingregel met Jacobimatrices

Mooie voorbeelden van matrixprodukten als in (6.22) zien we als we in (9.22) aan beide kanten links $(\tilde{h}\ \tilde{k})$ erbij zetten. Dan is

$$(\tilde{h}\ \tilde{k})\begin{pmatrix}\frac{\partial G}{\partial r}\\ \frac{\partial G}{\partial \theta}\end{pmatrix} = (\tilde{h}\ \tilde{k})\begin{pmatrix}\frac{\partial X}{\partial r} & \frac{\partial Y}{\partial r}\\ \frac{\partial X}{\partial \theta} & \frac{\partial Y}{\partial \theta}\end{pmatrix}\begin{pmatrix}\frac{\partial F}{\partial x}\\ \frac{\partial F}{\partial y}\end{pmatrix}, \tag{9.23}$$

nu links en rechts uit te werken tot een 1 bij 1 matrix, met daarin precies de twee lineaire stukken die we hierboven aan elkaar gelijkstelden bij het uitwerken van de kettingregel, om tot (9.20) en (9.21) te komen.

Via links en rechts transponeren is (9.23) equivalent met

$$(\frac{\partial G}{\partial r}\ \frac{\partial G}{\partial \theta})\begin{pmatrix}\tilde{h}\\ \tilde{k}\end{pmatrix} = (\frac{\partial F}{\partial x}\ \frac{\partial F}{\partial y})\begin{pmatrix}\frac{\partial X}{\partial r} & \frac{\partial X}{\partial \theta}\\ \frac{\partial Y}{\partial r} & \frac{\partial Y}{\partial \theta}\end{pmatrix}\begin{pmatrix}\tilde{h}\\ \tilde{k}\end{pmatrix}, \tag{9.24}$$

waarin we de *Jacobimatrices* van G, F en Z herkennen, waarin de beeldvariabelen niet horizontaal maar verticaal genummerd worden, zoals we in (6.10) al gezien hebben. Ook zien we dat de volgorde in (9.24) nu net is als in (5.3), hetgeen prettig is als we $\tilde{h}$ en $\tilde{k}$ zien als variabel.

Als $F(x,y) = G(r,\theta)$ een grootheid is met twee componenten

$$F_1(x,y) = G_1(r,\theta) \quad \text{en} \quad F_2(x,y) = G_2(r,\theta),$$

dan kan (5.3) voor beide componenten in één keer opgeschreven worden als

$$\begin{pmatrix}\frac{\partial G_1}{\partial r} & \frac{\partial G_1}{\partial \theta}\\ \frac{\partial G_2}{\partial r} & \frac{\partial G_2}{\partial \theta}\end{pmatrix}\begin{pmatrix}\tilde{h}\\ \tilde{k}\end{pmatrix} = \begin{pmatrix}\frac{\partial F_1}{\partial x} & \frac{\partial F_1}{\partial y}\\ \frac{\partial F_2}{\partial x} & \frac{\partial F_2}{\partial y}\end{pmatrix}\begin{pmatrix}\frac{\partial X}{\partial r} & \frac{\partial X}{\partial \theta}\\ \frac{\partial Y}{\partial r} & \frac{\partial Y}{\partial \theta}\end{pmatrix}\begin{pmatrix}\tilde{h}\\ \tilde{k}\end{pmatrix}, \tag{9.25}$$

en zien we hoe de kettingregel toegepast op

$$\mathbb{R}^2 \xrightarrow{Z} \mathbb{R}^2 \xrightarrow{F} \mathbb{R}^2$$

de Jacobimatrix van G produceert via het matrixprodukt van de Jacobimatrices van F en Z.

Deze notatie suggereert om de afhankelijke grootheid $F(x,y) = G(r,\theta)$ als 2-vector te zien, dus

$$F(x,y) = \begin{pmatrix}F_1(x,y)\\ F_2(x,y)\end{pmatrix} \quad \text{en} \quad G(r,\theta) = \begin{pmatrix}G_1(r,\theta)\\ G_2(r,\theta)\end{pmatrix},$$

en dus ook x, y en r, θ als componenten van de 2-vectoren

$$\begin{pmatrix}x\\ y\end{pmatrix} \quad \text{en} \quad \begin{pmatrix}r\\ \theta\end{pmatrix}.$$

We blijven echter $F = F(x,y)$ en $G = G(r,\theta)$ schrijven.

9.3.4 Omschrijven van differentiaaloperatoren

De notatie (9.23) is handiger als we zoals gebruikelijk in de natuurkunde aan $F(x,y) = G(r,\theta)$ denken als één en dezelfde afhankelijke grootheid, en niet als een functie zoals gebruikelijk in de wiskunde.

In dat geval ligt het voor de hand om die grootheid af te splitsen uit de notatie in (9.22) en de kettingregel voor coördinatentransformaties te schrijven als

$$\begin{pmatrix} \frac{\partial}{\partial r} \\ \frac{\partial}{\partial \theta} \end{pmatrix} = \begin{pmatrix} \frac{\partial X}{\partial r} & \frac{\partial Y}{\partial r} \\ \frac{\partial X}{\partial \theta} & \frac{\partial Y}{\partial \theta} \end{pmatrix} \begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \end{pmatrix}, \tag{9.26}$$

hetgeen de matrixnotatie is voor

$$\frac{\partial}{\partial r} = \frac{\partial X}{\partial r}\frac{\partial}{\partial x} + \frac{\partial Y}{\partial r}\frac{\partial}{\partial y};$$

$$\frac{\partial}{\partial \theta} = \frac{\partial X}{\partial \theta}\frac{\partial}{\partial x} + \frac{\partial Y}{\partial \theta}\frac{\partial}{\partial y},$$

waaruit de differentiaaloperatoren

$$\frac{\partial}{\partial x} \quad \text{en} \quad \frac{\partial}{\partial y}$$

kunnen worden opgelost in termen van de coëfficiënten

$$\frac{\partial X}{\partial r}, \frac{\partial Y}{\partial r}, \frac{\partial X}{\partial \theta}, \frac{\partial Y}{\partial \theta} \quad \text{en de differentiaaloperatoren} \quad \frac{\partial}{\partial r}, \frac{\partial}{\partial \theta}.$$

Opgave 9.6. In het concrete geval van poolcoördinaten geeft dit

$$\frac{\partial}{\partial x} = \cos\theta \frac{\partial}{\partial r} - \frac{\sin\theta}{r}\frac{\partial}{\partial \theta};$$

$$\frac{\partial}{\partial y} = \sin\theta \frac{\partial}{\partial r} + \frac{\cos\theta}{r}\frac{\partial}{\partial \theta}.$$

Laat dit zien.

Met Opgave 9.6 zijn we nog niet klaar als we in (9.15)

$$\nabla = \begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \end{pmatrix} = e_x \frac{\partial}{\partial x} + e_y \frac{\partial}{\partial y}$$

willen omschrijven naar r en θ. De vraag is ook hoe we e_x en e_y omschrijven naar e_r en e_θ, en daarvoor komt de vraag wat e_r en e_θ eigenlijk zijn.

Een natuurkundige zal hier niet lang over nadenken Teken maar een plaatje en het is evident dat

Teken plaatje!

$$e_r = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} \quad \text{en} \quad e_\theta = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix},$$

en

$$\nabla = e_x \frac{\partial}{\partial x} + e_y \frac{\partial}{\partial y} = e_r \frac{\partial}{\partial r} + \frac{1}{r} e_\theta \frac{\partial}{\partial \theta} \tag{9.27}$$

de gradiënt in poolcoördinaten geeft. Daar had'ie de hele kettingregel überhaupt niet voor nodig. Omdat

$$e_r \cdot e_r = e_\theta \cdot e_\theta = 1 \quad \text{en} \quad e_r \cdot e_\theta = e_\theta \cdot e_r = 0,$$

staan de vectoren e_r en e_θ in ieder punt onderling loodrecht[5], met elk lengte 1, en wijzen in de richtingen waarin het punt $(x, y) = (r\cos\theta, r\sin\theta)$ loopt als je r respectievelijk θ varieert. De voorfactor $\frac{1}{r}$ compenseert de met r evenredige snelheid bij gelijkmatige toename van θ.

Opgave 9.7. In (9.27) staan twee representaties van dezelfde operator. Door e_x en e_y in e_r en e_θ uit te drukken en Opgave 9.6 te gebruiken kun je zien dat ze inderdaad hetzelfde zijn. Doe dat. Schrijf ook $V = V_x e_x + V_y e_y$ om als $V = V_r e_r + V_\theta e_\theta$.

Opgave 9.8. Laat zien dat de divergentie in poolcoördinaten wordt gegeven door

$$\nabla \cdot V = (e_r \frac{\partial}{\partial r} + \frac{1}{r} e_\theta \frac{\partial}{\partial \theta}) \cdot (V_r e_r + V_\theta e_\theta) = \frac{\partial V_r}{\partial r} + \frac{V_r}{r} + \frac{1}{r} \frac{\partial V_\theta}{\partial \theta}.$$

Hint: Omdat e_r en e_θ van θ afhangen werkt met de produktregel van Leibniz de $e_\theta \frac{\partial}{\partial \theta}$ in de factor links nu ook op e_r en e_θ in de factor rechts, en één van die twee geeft na inprodukt met de voorfactor e_θ een bijdrage.

Opgave 9.9. Pas de regel in Opgave 9.8 nu toe op ∇ zelf en laat zien dat

$$\Delta = \nabla \cdot \nabla = \underbrace{\frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r}}_{\Delta_r} + \frac{1}{r^2} \underbrace{\frac{\partial^2}{\partial \theta^2}}_{\Delta_S}$$

Hint: wellicht eerst Opgave 9.8 toepassen op ∇ als werkend op de afhankelijke grootheid $G = F$, waarvoor de natuurkundige dezelfde letter gebruikt en de wiskundige dan

[5] Wiskundig is dit per definitie en consistent met wat je ziet als je pijltjes tekent.

met $G(r,\theta) = F(r\cos\theta, r\cos\theta)$ in de war raakt, omdat G en F niet dezelfde functies zijn.

In Opgave 9.9 zien we

$$\Delta = \Delta_r + \frac{1}{r^2}\Delta_S, \tag{9.28}$$

waarin Δ_r de radiële Laplaciaan is, die ook werkt op functies $R = R(r)$, en Δ_S de Laplace-Beltrami operator op

$$S = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 1\},$$

uitgedrukt in de hoekvariabele θ als

$$\Delta_S = \frac{\partial^2}{\partial\theta^2}.$$

Het aardige nu is dat de integraal van de Laplaciaan van een nette functie

$$u(x,y) = u(r\cos\theta, r\sin\theta)$$

over een disk B_R met straal $R > 0$ in poolcoördinaten meteen tot een belangrijke conclusie leidt, maar daarvoor moeten we eerst weten wat meervoudige integralen zijn.

9.4 Harmonische polynomen

We vinden deze polynomen ook als we de Laplace vergelijking

$$u_{xx} + u_{yy} = 0$$

voor $u = u(x,y)$ met *scheiding van variabelen* in poolcoördinaten oplossen door de operator in Opgave 9.9 los te laten op

$$u(x,y) = R(r)\Theta(\theta), \tag{9.29}$$

en het resultaat gelijk aan nul te stellen. Dit geeft

$$0 = (\frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial\theta^2})R(r)\Theta(\theta)$$

$$= \Theta(\theta)(\frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r})R(r) + \frac{R(r)}{r^2}\frac{\partial^2}{\partial\theta^2}\Theta(\theta)$$

$$= \Theta(\theta)(R''(r) + \frac{1}{r}R'(r)) + \frac{R(r)}{r^2}\Theta''(\theta).$$

Als Θ'' een veelvoud is van Θ, zeg

$$-\Theta'' = \mu\Theta \tag{9.30}$$

dan volgt Euler's vergelijking

$$R''(r) + \frac{1}{r}R'(r) = \mu\frac{R(r)}{r^2} \tag{9.31}$$

voor $R(r)$.

Merk op dat (9.30) gezien kan worden als een (eigenwaarde)probleem voor

$$-\Delta_S = -\frac{d^2}{d\theta}$$

op de eenheidscirkel waar bij Θ een 2π-periodieke functie moet zijn om een functie op de cirkel

$$S = \{(x, y) : x^2 + y^2 = 1\}$$

te definiëren.

Opgave 9.10. Welke μ zijn toegestaan in (9.30) voor oplossingen (9.29) die op heel $\mathbb{R}^2$ zijn gedefinieerd? Leg uit dat je die waarden ook meteen[6] aan de harmonische polynomen kunt zien zonder de precieze vorm van (9.30) te kennen. Schrijf die harmonische polynomen in gescheiden variabelen r en θ als $R(r)\Theta(\theta)$ en verifieer dat $R(r)$ een oplossing is van (9.31) met de bijbehorende μ.

Opgave 9.11. Voor elke $\mathrm{N} \in \mathbb{N}$ en $a_0, \ldots, a_\mathrm{N}$, $b_1, \ldots, b_\mathrm{N}$ in $\mathbb{R}$ is

$$\frac{a_0}{2} + \sum_{k=1}^{N}(a_k \cos k\theta + b_k \sin k\theta)r^n$$

via $x = r\cos\theta, y = r\sin\theta$ een harmonische functie. Overtuig jezelf van de juistheid van de informele uitspraak dat deze oplossing in $(0, 0)$ gelijk is aan zijn gemiddelde op elke disk met middelpunt $(0, 0)$.

[6] In $\mathbb{R}^3$ eigenwaarden en -functies van Laplace-Beltrami operator ook via polynomen.

Opgave 9.11 suggereert

$$u(x,y) = \frac{a_0}{2} + \sum_{k=1}^{\infty}(a_k \cos k\theta + b_k \sin k\theta) r^n$$

als een algemene oplossing voor de Laplacevergelijking op de eenheidsdisk met randvoorwaarde

$$u(\cos\theta, \sin\theta) = f(\theta) = \frac{a_0}{2} + \sum_{k=1}^{\infty}(a_k \cos k\theta + b_k \sin k\theta), \tag{9.32}$$

een zogenaamde *Fourierreeks*[7] voor een 2π-periodieke functie $\theta \to f(\theta)$. Ook deze $u(x,y)$ is dan in $(x,y) = (0,0)$ het gemiddelde van $u(x,y)$ op elke disk met middelpunt $(0,0)$ en straal voldoende klein, kleiner dan 1 in dit geval.

Opgave 9.12. In $\mathbb{R}^3$ gebruiken we *bolcoördinaten*

$$x = r\sin\theta\cos\phi;$$
$$y = r\sin\theta\sin\phi;$$
$$z = r\cos\theta,$$

en

$$e_r = \sin\theta\cos\phi\, e_x + \sin\theta\sin\phi\, e_y + \cos\theta\, e_z$$
$$e_\theta = \cos\theta\cos\phi\, e_x + \cos\theta\sin\phi\, e_y - \sin\theta\, e_z$$
$$e_\phi = -\sin\phi\, e_x + \cos\phi\, e_y.$$

Schrijf e_r, e_θ, e_ϕ al of niet als kolomvectoren, en verifieer dat

$$e_r \cdot e_r = e_\theta \cdot e_\theta = e_\phi \cdot e_\phi = 1; \quad e_r \cdot e_\theta = e_r \cdot e_\phi = e_\theta \cdot e_\phi = 0.$$

Overtuig jezelf van

$$\nabla = e_r \frac{\partial}{\partial r} + \frac{1}{r} e_\theta \frac{\partial}{\partial \theta} + \frac{1}{r\sin\theta} e_\phi \frac{\partial}{\partial \phi}, \tag{9.33}$$

en gebruik (9.33) om voor

$$V = V_r e_r + V_\theta e_\theta + V_\phi e_\phi$$

eerst af te leiden dat

$$\nabla \cdot V = \frac{\partial V_r}{\partial r} + \frac{2}{r} V_r + \frac{1}{r}\left(\frac{\partial V_\theta}{\partial \theta} + \frac{\cos\theta}{\sin\theta} V_\theta + \frac{1}{\sin\theta}\frac{\partial V_\phi}{\partial \phi}\right),$$

[7] Uitgebreid behandeld in de mamanotes van vorig jaar.

en vervolgens via $V = \nabla F$ dat

$$\Delta = \underbrace{\frac{\partial^2}{\partial r^2} + \frac{2}{r}\frac{\partial}{\partial r}}_{\Delta_r} + \frac{1}{r^2}(\underbrace{\frac{\partial^2}{\partial \theta^2} + \frac{\cos\theta}{\sin\theta}\frac{\partial}{\partial \theta} + \frac{1}{\sin\theta}\frac{\partial^2}{\partial \phi^2}}_{\Delta_S}).$$

Wederom zien we hier (9.9), maar nu met Δ_S gedefinieerd op

$$S = \{(x, y) \in \mathbb{R}^2 : \, x^2 + y^2 + z^2 = 1\},$$

De formules in $\mathbb{R}^n$ laten zich nu raden, afgezien wellicht van de exacte vorm van Δ_S in de hoekvariabelen $\theta_1, \ldots, \theta_{n-1}$, maar met

$$\Delta_r = \frac{\partial^2}{\partial r^2} + \frac{n-1}{r}\frac{\partial}{\partial r}$$

voor het radiële gedeelte.

Opgave 9.13. In $\mathbb{R}^3$ gebruiken we bolcoördinaten

$$x = r\sin\theta\cos\phi;$$

$$y = r\sin\theta\sin\phi;$$

$$z = r\cos\theta,$$

en

$$e_r = \sin\theta\cos\phi\, e_x + \sin\theta\sin\phi\, e_y + \cos\theta\, e_z$$

$$e_\theta = \cos\theta\cos\phi\, e_x + \cos\theta\sin\phi\, e_y - \sin\theta\, e_z$$

$$e_\phi = -\sin\phi\, e_x + \cos\phi\, e_y.$$

Schrijf e_r, e_θ, e_ϕ al of niet als kolomvectoren en verifieer dat

$$e_r \cdot e_r = e_\theta \cdot e_\theta = e_\phi \cdot e_\phi = 1; \quad e_r \cdot e_\theta = e_r \cdot e_\phi = e_\theta \cdot e_\phi = 0.$$

Overtuig jezelf van

$$\nabla = e_r\frac{\partial}{\partial r} + \frac{1}{r}e_\theta\frac{\partial}{\partial \theta} + \frac{1}{r\sin\theta}e_\theta\frac{\partial}{\partial \phi} \tag{9.34}$$

en gebruik (9.34) om voor

$$V = V_r e_r + V_\theta e_\theta + V_\phi$$

eerst af te leiden dat

$$\nabla V = \frac{\partial V_r}{\partial r} + \frac{2}{r}V_r + \frac{1}{r}(\frac{\partial V_\theta}{\partial \theta} + \frac{\cos\theta}{\sin\theta}V_\theta + \frac{1}{\sin\theta}\frac{\partial V_\phi}{\partial \phi}),$$

en vervolgens via $V = \nabla F$ dat

$$\Delta = \underbrace{\frac{\partial^2}{\partial r^2} + \frac{2}{r}\frac{\partial}{\partial r}}_{\Delta_r} + \frac{1}{r^2}(\underbrace{\frac{\partial^2}{\partial \theta^2} + \frac{\cos\theta}{\sin\theta}\frac{\partial}{\partial \theta} + \frac{1}{\sin\theta}\frac{\partial^2}{\partial \phi^2}}_{\Delta_S}).$$

Wederom zien we hier (9.9), maar nu met Δ_S gedefinieerd op

$$S = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 + z^2 = 1\},$$

De formules in $\mathbb{R}^n$ laten zich nu raden, afgezien wellicht van de exacte vorm van Δ_S in de hoekvariabelen $\theta_1, \ldots, \theta_{n-1}$, maar met

$$\Delta_r = \frac{\partial^2}{\partial r^2} + \frac{n-1}{r}\frac{\partial}{\partial r}$$

voor het radiële gedeelte.

9.5 Intermezzo: het waterstofatoom

Met

$$V(r) = -\frac{e^2}{r}$$

is de stationaire Schrödinger vergelijking voor het waterstofatoom

$$\frac{\hbar^2}{2m}\Delta\psi - \frac{e^2}{r}\psi = E\psi, \tag{9.35}$$

waarin m de massa van het electron is, e de lading van het electron, $\hbar$ de constante van Planck. De negatieve waarden van E waarvoor (9.35) een oplossing met

$$\iiint_{\mathbb{R}^3} |\psi(x, y, z)|^2 \, d(x, y, z) = 1$$

heeft zijn de energieniveaus die het electron in gebonden toestand kan aannemen.

We hebben gezien dat

$$\Delta = \underbrace{\frac{\partial^2}{\partial r^2} + \frac{2}{r}\frac{\partial}{\partial r}}_{\Delta_r} + \frac{1}{r^2}(\underbrace{\frac{\partial^2}{\partial \theta^2} + \frac{\cos\theta}{\sin\theta}\frac{\partial}{\partial \theta} + \frac{1}{\sin\theta}\frac{\partial^2}{\partial \phi^2}}_{\Delta_S}).$$

Via

$$\psi(x, y, z) = R(r)P_l(\frac{x}{r}, \frac{y}{r}, \frac{z}{r}),$$

waarin $P_l(x,y,z) = Y(\theta,\phi)$ een harmonisch homogeen polynoom van graad l in x,y,z is, en een nieuwe x en n gedefinieerd door

$$x = \frac{1}{\hbar}\sqrt{-2mE}\, r \quad \text{en} \quad -E = \frac{me^4}{2\hbar^2 n^2},$$

leidt dit tot

$$\frac{d^2R}{dx^2} + \frac{2}{x}\frac{dR}{dx} - \frac{l(l+1)}{x^2}R + \frac{2n}{x}R = R$$

met $R(x) \sim x^l$ voor $x \to 0$ en $R(x) \sim e^{-x}$ voor $x \to \infty$.

Substitueer daarom $R(x) = x^l e^{-x} u(x)$ en leidt voor $u(x)$ af dat

$$\frac{d^2u}{dx^2} + (\frac{4l}{x} - 2)\frac{du}{dx} = 2\frac{n-l-1}{x}u.$$

Opgave 9.14. Corrigeer eventuele typo's hierboven. De machtreeksoplossing[8]

$$u(x) = 1 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots$$

breekt af voor een n die van l afhangt. Welke n is dat?

9.6 Lijnintegralen over polygonen en Coursat

We beginnen met formule (4.9) uit Sectie 4.1, al of niet via de "verboden" operaties daarboven, geschreven als

$$F(x_1) - F(x_0) = \int_0^1 \underbrace{F'((1-t)x_0 + tx_1)}_{f(x(t))} \underbrace{(x_1 - x_0)dt}_{dx} = \int_{x_0}^{x_1} f(x)\, dx,$$

waarin, met $x, x(t), x_0, x_1, f(x)$ vervangen door $z, z(t), z_0, z_1, f(z)$,

$$t \to z(t) = (1-t)z_0 + tz_1 \tag{9.36}$$

het interval $[z_0, z_1]$ parametriseert, en

$$\int_0^1 f(z(t))z'(t)dt =$$

$$\int_0^1 f((1-t)z_0 + tz_1)dt\,(z_1 - z_0) = \int_{z_0}^{z_1} f(z)\, dz. \tag{9.37}$$

[8]Instructief om eerst $\frac{d^2R}{dx^2} + \frac{2}{x}\frac{dR}{dx} = R$ op te lossen.

Dit is een formule die we, zonder dat (4.9) daarvoor nog nodig is, nu kunnen lezen met $z_0, z_1 \in \mathbb{C}$ en $f : \mathbb{C} \to \mathbb{C}$, met het linkerlid als ondubbelzinnige definitie van het rechterlid: de lijnintegraal

$$\int_{z_0}^{z_1} f(z)\,dz$$

over het rechte lijnstuk van z_0 naar z_1, van de functie $z \to f(z)$. Niet meer praten over andere parametervoorstellingen van $[z_0, z_1]$ dan (9.36), tenzij het nodig[9] is zou ik zeggen. Merk op dat $[z_0, z_1]$ voor alle $z_0, z_1 \in \mathbb{C}$ is gedefinieerd, dus $[1, 0]$ heeft nu ook betekenis. Je moet er even aan wennen maar het spreekt vanzelf. Het ligt voor de hand om aan z_0 als het begin- en z_1 als het eindpunt van $[z_0, z_1]$ te denken. Daarmee wordt $[z_0, z_1]$ meer dan alleen een verzameling: $[z_0, z_1]$ is zo een georienteerd lijnstuk.

Opgave 9.15. Laat zien dat

$$\int_{z_0}^{z_1} z\,dz = \frac{1}{2}(z_1^2 - z_0^2).$$

Evalueer vervolgens $\int_{z_0}^{z_1} z^n\,dz$ voor alle $n \in \mathbb{N}$.

Opgave 9.16. Evalueer $\int_{z_0}^{z_1} z^n\,dz$ voor alle $n \in -\mathbb{N} = \{-1, -2, -3, \cdots\}$. Doe $n = -1$ als laatste[10]. Welke voorwaarde moet je leggen op z_0 en z_1?

Opgave 9.17. Laat zien dat

$$\left|\int_{z_0}^{z_1} f(z)\,dz\right| \leq |z_0 - z_1| \max_{z\in[z_0,z_1]} |f(z)|.$$

Integralen gedefinieerd als hierboven door (9.37) kunnen we rijgen tot een integraal over een *polygonaal pad*

$$z_0 \to z_1 \to \cdots \to z_n$$

[9] Quod non.

[10] De eersten zullen de laatsten zijn.

middels

$$\int_{z_0}^{z_1} f(z)\,dz + \int_{z_1}^{z_2} f(z)\,dz + \cdots + \int_{z_{n-1}}^{z_n} f(z)\,dz = \int_{z_0,\ldots,z_n} f(z)\,dz, \quad (9.38)$$

als $z \to f(z)$ een continue functie

$$[z_0, z_1] \cup \cdots \cup [z_{n-1}, z_n] \xrightarrow{f} \mathbb{C}$$

definieert. In het bijzondere geval dat $z_0 = z_n$ is eenvoudig na te gaan dat hier NUL uitkomt als $n = 2$.

Opgave 9.18. Ga dit na. Hint: neem eerst $z_0 = z_2, z_1 \in \mathbb{R}$ om te zien hoe het moet werken.

Wat deze opgave zegt is dat heen en weer lopen geen bijdrage aan een keten als in (9.38) geeft omdat vrijwel per definitie

$$\int_{z_0,z_1,z_0} f(z)\,dz = \int_{z_0}^{z_1} f(z)\,dz + \int_{z_1}^{z_0} f(z)\,dz = 0,$$

voor *iedere* $[z_0, z_1] \xrightarrow{f} \mathbb{C}$ continu. We kunnen integralen dus vereenvoudigen door heen en weer stukjes weg te laten, ook als die niet achter elkaar zitten in het polygonale pad.

Neem nu in je complexe vlak z_0, z_1, z_2, not all on a line[11], en neem aan dat

$$\Delta = \Delta_{z_0,z_1,z_2} = \{t_0 z_0 + t_1 z_1 + t_2 z_2 : t_0, t_1, t_2 \geq 0, t_0 + t_1 + t_2 = 1\} \xrightarrow{f} \mathbb{C}$$

continu is. De verzameling Δ bestaat dus uit alle convexe combinaties van z_0, z_1, z_2 gezien als punten in het complexe vlak. Dat is een driehoekige tegel waarvan de rand een driehoek[12] is.

Teken plaatje!

Laat z_3, z_4, z_5 de middens zijn van $[z_0, z_1]$, $[z_1, z_2]$, $[z_2, z_0]$, die je krijgt door achtereenvolgens t_2, t_0, t_1 nul, en steeds de andere twee t-tjes $\frac{1}{2}$ te kiezen. Dan is

$$\int_{z_0,z_1,z_2,z_0} f(z)\,dz = \int_{z_0,z_3,z_5,z_4,z_3,z_1,z_4,z_2,z_5,z_0} f(z)\,dz =$$

$$\int_{z_0,z_3,z_5,z_0} f(z)\,dz + \int_{z_3,z_4,z_5,z_3} f(z)\,dz + \int_{z_3,z_1,z_4,z_3} f(z)\,dz + \int_{z_5,z_4,z_2,z_5} f(z)\,dz.$$

[11] Sommigen horen hierbij de stem van Erdös.

[12] Vereniging van 3 lijnstukjes: $[z_0, z_1] \cup [z_1, z_2] \cup [z_2, z_0]$, met zo gewenst een orientatie.

De integraal over het gesloten[13] pad

$$z_0 \to z_3 \to z_5 \to z_4 \to z_3 \to z_1 \to z_4 \to z_2 \to z_5 \to z_0$$

is zo enerzijds gelijk aan de integraal over het gesloten pad

$$z_0 \to z_1 \to z_2 \to z_0$$

rond de grote driehoek, en anderzijds de som van vier integralen over de vier gesloten paden

$$z_0 \to z_3 \to z_5 \to z_0, \quad z_3 \to z_4 \to z_5 \to z_3,$$

$$z_3 \to z_1 \to z_4 \to z_3, \quad z_5 \to z_4 \to z_2 \to z_5$$

rond de vier kleinere driehoeken.

Als de oorspronkelijke integraal niet nul was, zeg gelijk[14] aan 1, dan is tenminste één van de vier integralen in absolute waarde minstens gelijk aan $\frac{1}{4}$, en kan daarna met die integraal het argument herhaald worden om een rij geneste driehoeken

$$\Delta = \Delta_{z_0,z_1,z_2} \supset \Delta_{z_0^{(1)},z_1^{(1)},z_2^{(1)}} \supset \Delta_{z_0^{(2)},z_1^{(2)},z_2^{(2)}} \supset \Delta_{z_0^{(3)},z_1^{(3)},z_2^{(3)}} \supset \cdots$$

te maken met

$$\left| \int_{z_0^{(k)},z_1^{(k)},z_2^{(k)},z_0^{(k)}} f(z)\,dz \right| \geq \frac{1}{4^k}$$

voor $k = 0, 1, 2, 3, \ldots$.

Opgave 9.19. Bewijs dat de rijen $z_0^{(k)}, z_1^{(k)}, z_2^{(k)}$ convergeren naar een limiet in Δ_{z_0,z_1,z_2} als $k \to \infty$.

Zonder beperking der algemeenheid mogen we nu wel aannemen[15] dat deze limiet gelijk is aan 0, i.e.

$$z_0^{(k)}, z_1^{(k)}, z_2^{(k)} \to 0.$$

Kan het zo zijn dat f complex differentieerbaar is in 0? Zo ja, dan geldt voor $z \in \Delta$ dat

$$f(z) = f'(0)z + R(z)$$

[13] Teken het in je plaatje.

[14] Door $f(z)$ te delen door zijn integraal over de rand van Δ_{z_0,z_1,z_2}.

[15] Schuif de boel anders op.

met $R(z) = o(|z|)$ als $|z| \to 0$.

Maar dan is

$$\int_{z_0^{(k)},z_1^{(k)},z_2^{(k)},z_0^{(k)}} f(z)\,dz = \int_{z_0^{(k)},z_1^{(k)},z_2^{(k)},z_0^{(k)}} f'(0)z\,dz + \int_{z_0^{(k)},z_1^{(k)},z_2^{(k)},z_0^{(k)}} R(z)\,dz$$

$$= \int_{z_0^{(k)},z_1^{(k)},z_2^{(k)},z_0^{(k)}} R(z)\,dz,$$

omdat de eerste integraal nul is vanwege (9.15) toegepast op $\int_{z_0}^{z_1} z\,dz$, $\int_{z_1}^{z_2} z\,dz$, $\int_{z_2}^{z_0} z\,dz$. Dus volgt

$$\frac{1}{4^k} \leq |\int_{z_0^{(k)},z_1^{(k)},z_2^{(k)},z_0^{(k)}} R(z)\,dz| \leq \frac{|z_0 - z_1| + |z_1 - z_2| + |z_2 - z_0|}{2^k} \max_{z\in\delta\Delta^{(k)}} |R(z)|,$$

waarin $\delta\Delta^{(k)}$ de rand is van

$$\Delta^{(k)} = \Delta_{z_0^{(k)},z_1^{(k)},z_2^{(k)},z_0^{(k)}} \ni 0.$$

Omdat $|z|$ op zijn hoogste gelijk is aan de grootste afstand tussen twee punten in $\Delta^{(k)}$ geldt

$$|z| \leq \frac{d}{2^k} \quad \text{met} \quad d = \max_{z,w\in\Delta} |z - w|,$$

en samen met de 2^k die we al hadden krijgen nu ook een bovengrens voor de integraal, met een 4^k in de noemer en een $\varepsilon > 0$ naar keuze in de teller:

Opgave 9.20. Gebruik (de definitie van) $|R(z)| = o(|z|)$ als $|z| \to 0$ om met de laatste twee ongelijkheden een tegenspraak te forceren.

Stelling 9.21. *Voor iedere $f : \Delta_{z_0,z_1,z_2} \to \mathbb{C}$ die complex differentieerbaar is op de gesloten*[16] *driehoek Δ_{z_0,z_1,z_2} met hoekpunten $z_0, z_1, z_2 \in \mathbb{C}$ geldt dat*

$$\oint_{z_{012}} f(z)\,dz = \int_{z_0,z_1,z_2,z_0} f(z)\,dz = 0,$$

nu ook met een notatie[17] *die wellicht hierboven al voor de hand lag.*

[16] Voor w op de rand $f(z) = f(w) + f'(w)(z - w) + R(z; w)$ met $z \in \Delta_{z_0,z_1,z_2}$ lezen.

[17] Zie het rondje door de integraalslang heen maar als driehoekje.

Eenzelfde uitspraak geldt voor iedere n-hoek ($n > 3$) gemaakt uit n driehoeken

$$\Delta_{w_0,z_0,z_1}, \Delta_{w_0,z_1,z_2}, \ldots, \Delta_{w_0,z_{n-1},z_0},$$

met

$$\Delta_{w_0,z_0,z_1} \cap \Delta_{w_0,z_1,z_2} = [w_0, z_1], \quad \Delta_{w_0,z_1,z_2} \cap \Delta_{w_0,z_2,z_3} = [w_0, z_2],$$

$$\ldots$$

$$\Delta_{w_0,z_{n-2},z_{n-1}} \cap \Delta_{w_0,z_{n-1},z_n} = [w_0, z_{n-1}], \quad \Delta_{w_0,z_{n-1},z_0} \cap \Delta_{w_0,z_0,z_1} = [w_0, z_0].$$

Als de lijnstukjes waarmee (9.38) is gemaakt de zijden zijn van zo'n door

$$z_n = z_0, \ldots, z_{n-1} \tag{9.39}$$

gemaakte n-hoek met een punt w_0 in de n-hoek waarvoor alle $[w_0, z_k]$ in de veelhoek liggen[18], dan is

$$\oint_{z_0,\ldots,z_n} f(z)\, dz = \sum_{k=1}^{n} \oint_{w_0,z_{k-1},z_k,w_0} f(z)\, dz.$$

Als $z \to f(z)$ complex differentieerbaar is in elk punt van

$$P_{z_0,\ldots,z_{n-1}} = \cup_{k=1}^{n} \Delta_{w_0,z_{k-1},z_k},$$

dan volgt zo dat

$$\oint_{z_n=z_0,\ldots,z_n} f(z)\, dz = 0. \tag{9.40}$$

Bovendien kan ieder punt z_k dan met de andere punten vastgehouden naar binnen geschoven worden naar een $\tilde{z}_k$ in de door (9.39) gemaakte veelhoek, waarbij (9.40) niet verandert met eenzelfde argument waarin driehoeken met hoekpunten $z_{k-1}, z_k, \tilde{z}_k, z_{k+1}$ voorkomen. Kortom, (9.40) geldt voor iedere complex differentieerbare

$$f : P_{z_0,\ldots,z_{n-1}} \to \mathbb{C}$$

op ieder domein $P_{z_0,\ldots,z_{n-1}}$ begrensd door lijnstukjes $[z_{k-1}, z_k] \subset P_{z_0,\ldots,z_{n-1}}$.

Merk op dat we op een vanzelfsprekende manier kunnen praten over het links- of rechtsom genummerd zijn van de hoekpunten (9.39), ook als de n-hoek niet gemaakt is zoals boven (9.39) beschreven is. Als er er (maar) eindig veel punten $\zeta_1, \ldots, \zeta_p$ zijn in $P_{z_0,\ldots,z_{n-1}}$ (maar niet op de rand van) waar f niet

[18] De veelhoek is dan convex.

complex differentieerbaar is, dan het niet moeilijk om na te gaan dat (9.40) gelijk is aan de som van de integralen over de randen van kleine driehoekjes

$$\Delta^{(j)} = \Delta_{z_0^{(j)},z_1^{(j)},z_2^{(j)},z_0^{(j)}}$$

in $P_{z_0,\ldots,z_{n-1}}$ waar ζ_j echt in ligt, als die allemaal maar met dezelfde orientatie genomen worden. In dat geval is dus

$$\oint_{z_n=z_0,\ldots,z_n} f(z)\,dz = \sum_{j=1}^{p} \oint_{z_2^{(j)}=z_0^{(j)},z_1^{(j)},z_2^{(j)},} f(z)\,dz, \tag{9.41}$$

en we werken dat idee in het geval dat $p = 1$ met een hele bijzondere integrand nu uit in de volgende sectie, teneinde uiteindelijk ook te zien dat de termen in het rechterlid van (9.41) een verrassend simpele vorm krijgen.

9.7 Machtreeksen via een Cauchy integraalformule

Neem nu aan dat

$$D = \{z \in \mathbb{C} : |z| < 1\} \xrightarrow{f} \mathbb{C}$$

een complex differentieerbare functie is op de open eenheidsdisk. Neem een $\zeta \in \mathbb{C}$ met $|\zeta| = \rho < 1$. We laten nu rechtstreeks zien dat

$$f(\zeta) = \sum_{n=0}^{\infty} a_n \zeta^n,$$

met integraalformules voor de coëfficiënten $a_n \in \mathbb{C}$. De integralen zijn daarbij over regelmatige polygonen in D, met hoekpunten dicht bij de cirkelvormige[19] rand van D.

We leiden de formules of door Stelling 9.21 toe te passen op integralen van

$$z \to \frac{f(z) - f(\zeta)}{z - \zeta}$$

over geschikt gekozen driehoeken. Voor iedere $r \in (0, 1)$ en $n \in \mathbb{N}$ met $n \geq 3$ definiëren de punten

$$z_k = r \exp(2\pi i \frac{k}{n})$$

de hoekpunten van een regelmatig n-hoek $C_{r,n}$ in D met middelpunt 0. Maak een plaatje met $0 < |\zeta| < r$ en $n = 8$ of zo. Laat k van 0 tot en met n lopen om het kringetje rond[20] te maken. *Teken plaatje!*

[19] Denk aan deze contouren: http://www.fi.uu.nl/publicaties/literatuur/7244.pdf

[20] Nou ja, rond...

Op dezelfde manier maken

$$w_k = \zeta + \rho \exp(2\pi i \frac{k}{n})$$

de hoekpunten van een regelmatig n-hoek $\zeta + C_{\rho,n}$, met middelpunt ζ dat binnen $C_{r,n}$ ligt als $\rho < r - |\zeta|$. Teken beide polygonen in je plaatje en merk op dat voor iedere complex differentieerbare functie

$$\{z \in D : z \neq \zeta\} \xrightarrow{g} \mathbb{C}$$

nu geldt dat

$$\oint_{z_{0-n}} g(z)\,dz =$$

$$\int_{z_0,z_1,\ldots,z_n=z_0} g(z)\,dz = \int_{w_0,w_1,\ldots,w_n=w_0} g(z)\,dz \tag{9.42}$$

$$= \oint_{w_{0-n}} g(z)\,dz,$$

voor elke $n \geq 3$, waarbij we ook hier de voor de hand liggende notatie[21] met $\oint$ gebruiken voor de integralen van $g(z)$ over de linksom[22] doorlopen n-hoeken.

Opgave 9.22. Bewijs (9.42) door de zigzagintegraal

$$\int_{w_0,z_0,w_1,z_1,\ldots,w_n,z_n} g(z)\,dz$$

mee te nemen in de beschouwingen en de stelling van Coursat toe te passen op in totaal $2n$ driehoekjes.

Nu nemen we voor $g(z)$ het differentiaalquotiënt

$$\frac{f(z) - f(\zeta)}{z - \zeta}$$

en concluderen dat

$$\oint_{w_{0-n}} \frac{f(z) - f(\zeta)}{z - \zeta}\,dz = \oint_{z_{0-n}} \frac{f(z) - f(\zeta)}{z - \zeta}\,dz. \tag{9.43}$$

[21] Voor grote n is de veelhoek bijna een rondje.

[22] Dat is maar een woord hier, de punten bepalen de richting.

Opgave 9.23. De integraal in het linkerlid van (9.43) hangt van ρ af. Gebruik de differentieerbaarheid van f in ζ om te laten zien dat

$$\oint_{w_{0-n}} \frac{f(z)-f(\zeta)}{z-\zeta}\,dz \to 0$$

als $\rho \to 0$.

Maar de integraal in het linkerlid van (9.43) is gelijk aan de integraal in het rechterlid en hing niet van ρ af als $\rho < r - |\zeta|$. Hij ging[23] dus niet naar 0 want hij was al 0. Zo reduceert (9.43) tot

$$0 = \oint_{z_{0-n}} \frac{f(z)}{z-\zeta}\,dz - \oint_{z_{0-n}} \frac{f(\zeta)}{z-\zeta}\,dz,$$

en volgt

$$f(\zeta) \oint_{z_{0-n}} \frac{1}{z-\zeta}\,dz = \oint_{z_{0-n}} \frac{f(z)}{z-\zeta}\,dz.$$

Opgave 9.24. Laat zien dat

$$\oint_{z_{0-n}} \frac{1}{z-\zeta}\,dz = \oint_{w_{0-n}} \frac{1}{z-\zeta}\,dz = 2\pi i.$$

Hint: de eerste gelijkheid volgt als in Opgave 9.22 en de tweede integraal hangt niet van ζ of ρ af. In het linkerlid kan dus $\zeta = 0$ genomen worden. Op elke $[z_{k-1}, z_k]$ heeft $\frac{1}{z}$ een primitieve: de meerwaardige functie gedefinieerd in Opgave 6.3 die je als het goed is in Opgave 9.16 als laatste hebt gebruikt.

Stelling 9.25. *Als ζ ligt binnen een n-hoek zoals hierboven in de open eenheidsdisk D, en $f : D \to \mathbb{C}$ complex differentieerbaar is, dan is*

$$f(\zeta) = \frac{1}{2\pi i} \oint_{z_{0-n}} \frac{f(z)}{z-\zeta}\,dz.$$

This is the Cauchy Integral Formula, maar dan met n-hoeken in plaats van de gebruikelijke cirkels met middelpunt 0 en straal $r < 1$ groot genoeg.

Tenslotte volgt na het invullen van[24] de meetkundige reeksontwikkeling

$$\frac{1}{z-\zeta} = \frac{1}{z}\,\frac{1}{1-\frac{\zeta}{z}} = \frac{1}{z} + \frac{\zeta}{z^2} + \frac{\zeta^2}{z^3} + \frac{\zeta^3}{z^4} + \cdots$$

[23] Letterlijk gesproken.

[24] We prefereren weer de notatie met puntjes natuurlijk.

de machtreeksontwikkeling in de vorm als aangekondigd, via

$$f(\zeta) = \frac{1}{2\pi i} \oint_{z_{0-n}} f(z)(\frac{1}{z} + \frac{\zeta}{z^2} + \frac{\zeta^2}{z^3} + \frac{\zeta^3}{z^4} + \cdots)\, dz =$$

$$\frac{1}{2\pi i} \oint_{z_{0-n}} \frac{f(z)}{z}\, dz + \frac{1}{2\pi i} \oint_{z_{0-n}} \frac{f(z)}{z^2}\, dz\, \zeta + \frac{1}{2\pi i} \oint_{z_{0-n}} \frac{f(z)}{z^3}\, dz\, \zeta^2 + \cdots,$$

met

$$a_j = \frac{1}{2\pi i} \oint_{z_{0-n}} \frac{f(z)}{z^{j+1}}\, dz \tag{9.44}$$

voor alle $j \in \mathbb{N}_0$.

Stelling 9.26. *Als $f : D \to \mathbb{C}$ complex differentieerbaar is, dan geldt*

$$f(z) = \sum_{j=0}^{\infty} a_n z^n$$

met a_j gegeven door (9.44), waarin de integraal nu door $z_k = r\exp(2\pi i \frac{k}{n})$ $(k = 0, 1, \ldots, n)$ wordt gedefinieerd. En achteraf zijn dan zowel $r \in (0, 1)$ als $n \geq 3$ arbitrair.

Het rechtvaardigen van het verwisselen van $\oint$ en $\sum$ is wezen niets anders dan opmerken dat voor elke α en β in $\mathbb{C}$ de verzameling

$$\{f : [\alpha, \beta] \to \mathbb{C} : \ f \text{ is continu}\}$$

een (complexe) Banachruimte is, net zoals $C([a, b])$ een reële Banachruimte is. Het wordt dus tijd voor het volgende hoofdstuk.

Voor hier is het nog de vraag of we de limiet $n \to \infty$ willen nemen in Stelling 9.25 en (9.44) teneinde de $\oint$ te nemen over de cirkel geparametriseerd door

$$z = r\exp(i\theta) \quad \text{met} \quad dz = ir\exp(i\theta)d\theta \tag{9.45}$$

und so weiter.

Dat laatste kan komen na de observatie dat voor $0 \leq \rho < R$ en complex differentieerbare

$$A_{\rho,R} = \{z \in \mathbb{C} : \ \rho < |z| < R\} \xrightarrow{f} \mathbb{C}$$

geldt dat $f(z)$ te schrijven is als een zogenaamde *Laurentreeks*, i.e.

$$f(z) = \sum_{j=-\infty}^{\infty} a_j z^j = \sum_{j=0}^{\infty} a_j z^j + \sum_{j=1}^{\infty} \frac{a_{-j}}{z^j}, \tag{9.46}$$

met a_j gegeven door (9.44) voor *alle* $j \in \mathbb{Z}$, maar $n \geq 3$ en $r \in (\rho, R)$ wel zo gekozen dat met de punten $z_k = r\exp(2\pi i\frac{k}{n})$ de n-hoek in de annulus $A_{\rho,R}$ ligt.

Je bewijst dit met drie veelhoeken in $A_{\rho,R}$, waar ζ dan tussen twee van de drie in moet liggen, en in de kleinste. Als je eenmaal op het idee[25] bent gekomen wijst het zich vanzelf. De integraal over de nieuwe veelhoek wordt ook weer via een net iets andere meetkundige reeks in een machtreeks vertaald, nu met $\frac{1}{\zeta}$ die naar buiten gehaald wordt uit $\frac{1}{z-\zeta}$.

Deze zo verkregen spectaculaire uitspraak wordt gewoonlijk bewezen *na* het invoeren van lijnintegralen over echte krommen[26] zoals gegeven door (9.45) en de hele machinerie die nodig is om netjes te beschrijven wat krommen[27] eigenlijk zijn, waarbij vaak ook de continuiteit van $z \to f'(z)$ wordt gebruikt om de nog te bespreken stellingen van Green te kunnen gebruiken.

Die laatste stellingen zijn dus hier niet nodig. En de veelhoeken bieden veel meer mogelijkheden. Bijvoorbeeld voor functies die gedefinieerd en complex differentieerbaar zijn op gebieden ingesloten door twee geneste veelhoeken. Dat is misschien nog leuk om uit te zoeken.

Opgave 9.27. Natuurlijk geldt Stelling 9.26 niet alleen voor de eenheidsdisk, en kan r zowel zo klein als zo groot mogelijk gekozen worden voor het polygon waarmee de coefficienten worden berekend. Gebruik dit om te bewijzen dat er geen niet-constante begrensde complex differentieerbare functies $f : \mathbb{C} \to \mathbb{C}$ zijn.

9.8 De Cauchy Integraal Transformatie

De formule in Stelling 9.25 schrijven we met $1 = I$ en $\zeta = A$ als

$$f(A) = \frac{1}{2\pi i}\oint_{z_{0-n}} f(z)(zI - A)^{-1}\,dz, \tag{9.47}$$

nu voor een willekeurig polygon waar A binnen ligt en waarop

$$z \to (zI - A)^{-1} \tag{9.48}$$

dus bestaat als zeker een continue functie. Het polygon hoeft ook niet per se in de eenheidsdisk te liggen. Van de functie $z \to f(z)$ hoeven we bij nadere

[25] Gauss en Cauchy gingen ons voor.....

[26] Denk aan die goal van nummer 14 in het Zuiderpark en de enige echte Kromme.

[27] Denk ook aan hoe Murre dit woord uitspreekt.

beschouwing alleen maar aan te nemen dat f complex differentieerbaar is op het gebied begrensd door een polygon, inclusief het polygon[28] zelf.

Let op, de hoekpunten van het polygon moeten wel "linksom" genummerd worden, hetgeen ondubbelzinnig gedefinieerd kan worden aan de hand van de vergelijkingen voor de lijnen door de opeenvolgende hoekpunten, met iedere $z_k = x_k + iy_k$ opgevat als $(x_k, y_k) \in \mathbb{R}^2$, waarbij je wil formuleren dat het binnengebied van het polygon steeds links van ieder georienteerde interval $[z_{k-1}, z_k]$ ligt.

Met deze notatie kunnen we (9.47) nu ook lezen lezen met A een vierkante eerst nog reële matrix gezien als een continue lineaire afbeelding van $X = \mathbb{R}^n$ naar zichzelf, afbeeldingen die een algebra[29] vormen. Hier is nog het een en ander mee te doen, met behulp ook van

$$(zI - A)^{-1} = \frac{1}{z}(I + \frac{1}{z}A + \cdots)$$

als $|z|$ voldoende groot is, misschien beter meteen maar voor algemene X in Sectie 10.6.

De vraag is natuurlijk wel eerst wat we precies onder *A binnen het polygon gedefinieerd door*

$$z_0 \to z_1 \to \cdots \to z_n = z_0$$

moeten verstaan, als we (9.47) zomaar overschrijven met ζ vervangen door een lineaire operator $A : X \to X$. Voor de hand ligt dat A zo moet zijn dat met een groter polygon de zigzagtruc weer werkt, en de integrand als $L(X)$-waardige functie complex differentieerbaar is op het gebied tussen de twee polygonen, en ook op de twee polygonen zelf, en dat weer voor ieder groter polygon.

Daartoe moeten X en ook $L(X)$ zelf eerst complex uitgebreid worden, hetgeen abstract een constructie vereist maar in voorbeelden automatisch[30] gaat. En daarna is dan de natuurlijke eis dat (9.48) op het polygon en zijn buitengebied moet bestaan in de complexe versie van $L(X)$. Lees wat dit betreft verder in Sectie 10.6.

9.9 Kromme lijnintegralen

Met behulp van (9.37) is in (9.38)

$$\int_{z_0,\ldots,z_n} f(z)\,dz = \sum_{k=1}^{n} \int_{z_{k-1}}^{z_k} f(z)\,dz \tag{9.49}$$

[28] Via een zigzagintegraal als in Opgave 9.22 volgt de geldigheid van (9.47).
[29] Een Banachalgebra zelfs, zie Charlie's teleurstelling in Flowers for Algernon.
[30] Denk hier even over na.

gedefinieerd voor een rij punten die we (nog) niet als partitie zien, waarvoor we ook (nog) niet Riemanntussensommen als

$$\sum_{k=1}^{n} f(\zeta_k)(z_k - z_{k-1}) \quad \text{met} \quad \zeta_k \in [z_{k-1}, z_k] \tag{9.50}$$

hebben ingevoerd, zie Sectie 2.5. Maar als de "incrementen" $z_k - z_{k-1}$ klein zijn ligt gezien (9.37) iedere term in (9.50) voor de hand als benadering voor de overeenkomstige term in het rechterlid van (9.49) via

$$\int_{z_{k-1}}^{z_k} f(z)\,dz = \int_0^1 f((1-t)z_{k-1} + tz_k)dt\,(z_k - z_{k-1}) \approx f(\zeta_k)(z_k - z_{k-1}).$$

De vraag wat er gebeurt als $n \to \infty$ is echter nog niet goed gesteld, want de rij "partities" kan in principe willekeurig zijn.

In iedere schatting die het limietgedrag onder controle moet krijgen zal, behalve het klein worden van de incrementen, ook het gedrag van

$$\sum_{k=1}^{n} |z_k - z_{k-1}|$$

een rol spelen, met

$$z_k = z_k^{(n)}$$

zinvol afhankelijk van n gekozen, maar wat is zinvol? Hieronder wat overwegingen en een aanzet tot een uitgewerkt antwoord.

Het stuksgewijs lineaire pad P_n van $z_0^{(n)}$ via $z_1^{(n)}, \ldots, z_{n-1}^{(n)}$, naar $z_n^{(n)}$ voor $n \to \infty$ moet een nog te formuleren limietgedrag hebben, waarmee in ieder geval voor continue $z \to f(z)$ volgt dat

$$\int_{P_n} f(z)\,dz = \sum_{k=1}^{n} \int_{z_{k-1}^{(n)}}^{z_k^{(n)}} f(z)\,dz \quad \text{en} \quad \sum_{k=1}^{n} f(\zeta_k^{(n)})(z_k^{(n)} - z_{k-1}^{(n)}) \tag{9.51}$$

convergeren naar een limiet die we $\int_P f(z)\,dz$ zouden willen noemen.

Voor het verschil van deze sommen geldt

$$|\sum_{k=1}^{n} \int_{z_{k-1}^{(n)}}^{z_k^{(n)}} f(z)\,dz - \sum_{k=1}^{n} f(\zeta_k^{(n)})(z_k^{(n)} - z_{k-1}^{(n)})| =$$

$$|\sum_{k=1}^{n} \int_0^1 f((1-t)z_{k-1}^{(n)} + tz_k^{(n)})dt\,(z_k^{(n)} - z_{k-1}^{(n)}) - \sum_{k=1}^{n} f(\zeta_k^{(n)})(z_k^{(n)} - z_{k-1}^{(n)})| ==$$

$$|\sum_{k=1}^{n}\int_0^1 (f((1-t)z_{k-1}^{(n)}+tz_k^{(n)})-f(\zeta_k^{(n)}))dt\,(z_k^{(n)}-z_{k-1}^{(n)})| \leq$$

$$\max_{k=1,\ldots,n}|f((1-t)z_{k-1}^{(n)}+tz_k^{(n)})-f(\zeta_k^{(n)})|\sum_{k=1}^{n}|z_k^{(n)}-z_{k-1}^{(n)}| \leq$$

$$\max_{k=1,\ldots,n}\sup_{z,w\in[z_{k-1}^{(n)},z_k^{(n)}]}|f(z)-f(w)|\sum_{k=1}^{n}|z_k^{(n)}-z_{k-1}^{(n)}|,$$

en dat zou klein moeten zijn als f uniform continu is op een geschikt gekozen domein dat alle paden P_n bevat. In dat geval zijn de aannames dat

$$\mu_n = \max_{k=1,\ldots,n}|z_k^{(n)}-z_{k-1}^{(n)}| \to 0 \tag{9.52}$$

en

$$L_n = \sum_{k=1}^{n}|z_k^{(n)}-z_{k-1}^{(n)}| \quad \text{begrensd} \tag{9.53}$$

is als $n \to \infty$ voldoende om het verschil tussen de termen in (9.51) naar 0 te doen gaan als $n \to \infty$.

Voor we een definitie geven bekijken we wat we langs deelrijen sowieso kunnen bereiken kwa convergentie van P_n onder de aanname dat (9.52) en (9.53) gelden, en

$$z_0^{(n)} = a \quad \text{en} \quad z_n^{(n)} = b \tag{9.54}$$

vastgehouden worden in $\mathbb{C}$. We kijken dus naar mogelijke limieten van stuksgewijs lineaire paden van a naar b.

Het ligt voor de hand meteen een deelrij te nemen waarlangs L_n convergent is, zeg $L_{n_k} \to L \geq |b-a|$ met n_k een stijgende rij in $\mathbb{N}$. Vanaf zekere zulke n is er dan steeds een eerste $j = j_n$ waarvoor geldt dat de totale lengte langs P_n van a tot $z_{j_n}^{(n)}$ minstens $\frac{L}{2}$ is, en langs een verdere deelrij convergeren dan zowel $z_{j_n}^{(n)}$ als $z_{j_n-1}^{(n)}$ naar een limiet $z_{\frac{1}{2}}$.

Maar dit argument werkt niet alleen voor $\frac{1}{2}$. Voor elke $t \in (0,1)$ kunnen we vanaf zekere n een eerste $j = j_n^t$ vinden waarvoor de totale lengte langs P_n van a tot $z_{j_n^t}^{(n)}$ minstens tL is. Doen we dit voor

$$t = \frac{1}{2}, \frac{1}{4}, \frac{3}{4}, \frac{1}{8}, \frac{3}{8}, \frac{5}{8}, \frac{7}{8}, \ldots,$$

dan geeft een diagonaalrijargument[31] dat, voor elke rationale $t \in (0,1)$ met een noemer die een pure macht van 2 is, dat langs de geconstrueerde deelrij geldt dat

$$z_{j_n^t}^{(n)}$$

[31] Zie Sectie 3.3.

convergeert naar een limiet z_t voor al zulke t. Dit definieert een afbeelding

$$t \to z(t) = z_t,$$

waarvoor per constructie geldt[32] dat

$$|z(t_1) - z(t_2)| \leq |t_1 - t_2|L, \tag{9.55}$$

en die uniek uitbreidt tot een afbeelding $z : [0, 1] \to \mathbb{C}$ met dezelfde eigenschap.

Onze eerste geparametriseerde kromme die niet per se van de vorm (9.36) is. Een kromme waarvan de lengte nog niet gedefinieerd maar wel gelijk aan L is, als alles goed is[33], en waarlangs we kunnen integreren, middels benaderingen met Riemannsommen van de vorm

$$\sum_j f((z(\tau_j))(z(t_j) - z(t_{j-1}).$$

Wat we van dit alles hier willen uitwerken is nog de vraag, maar voor continu differentieerbare zulke $t \to z(t)$ is

$$\int_P f(z)\, dz = \int_0^1 f(z(t))z'(t)\, dt$$

een uitspraak die we willen hebben, waarbij het linkerlid gedefinieerd is als

$$\lim_{n\to\infty} \int_{P_n} f(z)\, dz$$

en de limiet langs de deelrij wordt genomen en moet bestaan. Dat vergt nog een stelling voor bijvoorbeeld continue $z \to f(z)$.

9.10 Still to fit in: stable polynomials

This relates to

`https://nl.wikipedia.org/wiki/Hurwitz-matrix`

You may have seen that we often need to solve equations of the form

$$\lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0 = 0,$$

[32] Wel even nagaan!

[33] En de lengte van het stuk tussen t_1 en t_2 gelijk aan $|t_1 - t_2|L$.

with real coefficients $a_0, a_1 \ldots, a_{n-1}, a_n = 1$, and that we want to know if the solutions all have negative real part. I found this in some paper but I forgot which one.

Write

$$f(z) = z^n + a_{n-1}z^{n-1} + \cdots + a_1 z + a_0 = p(z^2) + zq(z^2)$$

and introduce

$$P(z) = p(-z^2), \quad Q(z) = zq(-z^2).$$

For example, if

$$f(z) = 1 + z + z^2 + z^3 + z^4 + z^5,$$

then

$$P(z) = 1 - z^2 + z^4, \quad Q(z) = z - z^3 + z^5.$$

Assume that $f(z)$ factorizes as

$$f(z) = (z - z_1)(z - z_2)\cdots(z - z_n)$$

(this is always true), in which all z_j have negative real part. In the course I showed that this implies that

$$\mathrm{Im}(Q(z)\overline{P(z)}) > 0 \quad \text{if} \quad \mathrm{Im} z > 0.$$

For nonreal z this implies that $P(z) \neq 0 \neq Q(z)$ so that we can write (without worrying about dividing by zero!)

$$\frac{Q(z)}{P(z)} = \frac{Q(z)\overline{P(z)}}{|P(z)|^2} \quad \text{and} \quad \frac{P(z)}{Q(z)} = \frac{P(z)\overline{Q(z)}}{|Q(z)|^2},$$

and conclude that for $\mathrm{Im} z > 0$ the first quotient has positive imaginairy part and the second negative imaginairy part.

We also see that $P(z)$ and $Q(z)$ have only real zero's. Except for $z = 0$ for $Q(z)$, these appear in pairs $z = \pm\zeta$ with $\zeta > 0$, and we cannot have $P(\zeta) = Q(\zeta) = 0$, because in such a case $z = i\zeta$ is a zero of both $p(z^2)$ and $q(z^2)$, and therefore of $f(z)$. This would contradict the assumption that all zero's of $f(z)$ have negative real part.

Moreover, we can write for any such ζ, e.g. with $P(\zeta) = 0$, using long division (in dutch: staartdeling) that

$$P(z) = (z - \zeta)R(z),$$

with $R(z)$ a polynomial with real coefficients. We say that $z = \zeta$ is a simple zero of $P(z)$ if $R(\zeta) \neq 0$. If ζ is not a simple zero you can divide out more factors $z - \zeta$ and obtain that

$$P(z) = (z - \zeta)^k R_k(z)$$

with $k > 1$ and $R_k(\zeta) \neq 0$, so that in the end

$$\frac{P(z)}{Q(z)} = (z-\zeta)^k \frac{R_k(z)}{Q(z)} = (z-\zeta)^k S(z),$$

with $S(\zeta) \neq 0$. You should be able to see that this cannot be true if the imaginary part of this expression is negative close to ζ in the upper half plane.

Thus

$$P(z) = (z-\zeta)R(z) \quad \text{with} \quad R(\zeta) = P'(\zeta) \neq 0$$

and the quotient is of the form

$$\frac{P(z)}{Q(z)} = (z-\zeta)S(z) \quad \text{with} \quad S(\zeta) = \frac{P'(\zeta)}{Q(\zeta)} \neq 0.$$

The sign of the last quotient tells you the sign of the imaginary part of this expression for z a little above ζ. Conclude that

$$P(\zeta) = 0 \quad \Longrightarrow \quad \frac{P'(\zeta)}{Q(\zeta)} < 0,$$

and, by the same reasoning,

$$Q(\zeta) = 0 \quad \Longrightarrow \quad \frac{Q'(\zeta)}{P(\zeta)} > 0.$$

It follows that the zeros of $P(z)$ and $Q(z)$ are interlaced!

A remarkable consequence is also that all coefficients of $f(z)$ have the same sign and are positive:

$$a_0 > 0,\ a_1 > 0,\ \ldots, a_n > 0,\ a_{n-1} = 1 > 0.$$

To see this consider for instance $p(-z^2)$, replace z^2 by w and write

$$p(w) = a_0 - a_2 w + a_4 w^2 - \ldots.$$

All zero's of $p(w)$ are real, positive and simple. Convince yourself that polynomials with this property have alternating coefficients, just like the series for the $\cos x$ and $\sin x$ that I wrote on the board.

The final conclusion is that

- all the coefficients in $f(z)$ have the same sign;
- the zero's of $P(z)$ and $Q(z)$ are all real, simple, and interlaced.

It is another nice exercise to show that if these two properties hold, all zero's of $f(z)$ have negative real part.

10 Analyse in genormeerde ruimten

We hebben het al gehad over vectorruimten van functies. In zulke vectorruimten willen we over limieten kunnen spreken net zoals we dat doen in $\mathbb{R}$. In $\mathbb{R}$ gebruiken we daarvoor absolute waarden van verschillen, in algemene vectorruimten zoeken we naar een geschikte norm waarmee we de verschillen meten. Een eerste voorbeeld is de zogenaamde 1-norm.

Definitie 10.1. *Voor iedere $f \in \mathrm{RI}[a,b]$ schrijven we*

$$|f|_1 = \int_a^b |f(x)|\,dx.$$

Definitie 10.2. *Als X een lineaire vectorruimte over $\mathbb{R}$ is, en*

$$x \in X \to |x|_X \in \mathbb{R}$$

een afbeelding met de eigenschap dat voor alle $x, y \in X$ en voor alle $\lambda \in \mathbb{R}$ geldt dat

$$(i) \quad |x|_X \geq 0,\ |x|_X = 0 \iff x = 0; \quad (ii) \quad |\lambda x|_X = |\lambda|\,|x|_X$$

$$(iii) \quad |x+y|_X \leq |x|_X + |y|_X,$$

dan heet X genormeerd met norm $|\cdot|_X$. De subscript X laten we graag weg.

Een Cauchyrij in X is een door $n \in \mathbb{N}$ genummerde rij $x_n \in X$ met $|x_n - x_m|_X \to 0$ als $m, n \to \infty$.

Als alle Cauchyrijen in X convergent zijn, met per definitie hun limiet in X, dan heet X volledig. Volledige ruimten worden Banachruimten genoemd.

Opgave 10.3. Geef de ε-definitie van $|x_n - x_m|_X \to 0$ als $m, n \to \infty$. Geef ook de definitie van het convergent zijn van de rij x_n in X.

Opgave 10.4. Waarom is $|\cdot|_1$ geen norm op $\mathrm{RI}([a,b])$ maar wel op $C([a,b])$?

De voor de liggende norm op $C([a,b])$ is de ∞-norm, die de limiet is van de p-normen[1]

$$|f|_p = \left(\int_a^b |f(x)|^p\,dx\right)^{\frac{1}{p}} \tag{10.1}$$

[1] $p \geq 1$.

voor $p \to \infty$, al zie je dat aan de definitie

$$|f|_\infty = \max_{x\in[a,b]} |f(x)| \tag{10.2}$$

misschien niet meteen. Deze norm wordt op $C([a,b])$ ook wel de maximumnorm genoemd.

Stelling 10.5. *Met de maximumnorm (10.2) is $C([a,b])$ een Banachruimte.*

Bewijs. We bewijzen alleen dat $C([a,b])$ met de maximum norm (wel) volledig is. Neem een Cauchyrij in f_n in $C([a,b])$. Je had in Opgave 10.3 al bedacht dat de uitspraak dat $|f_n - f_m|_\infty \to 0$ als $m,n \to \infty$ equivalent is met

$$\forall\varepsilon > 0\, \exists \mathrm{N} \in \mathbb{N} : m,n \geq \mathrm{N} \implies |f_n - f_m|_\infty \leq \varepsilon, \tag{10.3}$$

ofwel

$$\forall\varepsilon > 0\, \exists \mathrm{N} \in \mathbb{N} : \quad m,n \geq \mathrm{N} \implies |f_n(x) - f_m(x)| \leq \varepsilon \quad \forall x \in [a,b]. \tag{10.4}$$

Maar dan is voor elke vast gekozen $x \in [a,b]$ de rij $f_n(x)$ een Cauchyrij in $\mathbb{R}$ en dus convergent met een limiet die we $f(x)$ noemen. Daarmee is een functie $f : [a,b] \to \mathbb{R}$ gedefinieerd en te bewijzen is natuurlijk dat $f \in C([a,b])$.

Voor we dat doen merken we op dat in (10.4) nu de limiet $m \to \infty$ genomen kan worden waarmee de uitspraak

$$\forall\varepsilon > 0\, \exists \mathrm{N} \in \mathbb{N} : \quad n \geq \mathrm{N} \implies |f_n(x) - f(x)| \leq \varepsilon \quad \forall x \in [a,b] \tag{10.5}$$

zich aandient. Als nu f (en dus ook $f - f_n$) inderdaad in $C([a,b])$ zit dan is (10.5) equivalent met

$$\forall\varepsilon > 0\, \exists \mathrm{N} \in \mathbb{N} : m,n \geq \mathrm{N} \implies |f_n - f|_\infty \leq \varepsilon \tag{10.6}$$

en zijn we klaar.

Waarom zit f in $C([a,b])$? Neem weer $\varepsilon > 0$ en probeer $|f(x) - f(y)| \leq \varepsilon$ te krijgen door $|x - y| \leq \delta$ voor een bijbehorende $\delta > 0$ te eisen. Dat moet wel via de f_n want die zijn uniform continu. Schrijf daartoe

$$f(x) - f(y) = f(x) - f_n(x) + f_n(x) - f_n(y) + f_n(y) - f(y),$$

waarmee

$$|f(x) - f(y)| \leq \underbrace{|f(x) - f_n(x)|}_{\leq \varepsilon \text{ als } n \geq \mathrm{N}} + |f_n(x) - f_n(y)| + \underbrace{|f_n(y) - f(y)|}_{\leq \varepsilon \text{ als } n \geq \mathrm{N}}. \tag{10.7}$$

In (10.7) kiezen we nu $n = \mathrm{N}$ en bij die vaste N een $\delta > 0$ zodanig dat voor de middelste term geldt dat

$$|f_{\mathrm{N}}(x) - f_{\mathrm{N}}(y) \leq \varepsilon \quad \text{als} \quad |x - y| \leq \delta,$$

culminerend in

$$|f(x) - f(y) \leq 3\varepsilon \quad \text{als} \quad |x - y| \leq \delta$$

vanwege (10.7). Hiermee is het bewijs klaar omdat $\varepsilon > 0$ willekeurig gekozen was.

Opgave 10.6. De voorfactor 3 doet er niet toe. Waarom niet?

Definitie 10.7. *Als de rij $f_n : [a, b] \to \mathbb{R}$ voldoet aan (10.5) in het bewijs hierboven dan heet de rij f_n uniform convergent op $[a, b]$. In deze definitie kan $[a, b]$ natuurlijk vervangen worden door een willekeurig interval I.*

10.1 Calculus voor X-waardige functies

Als X een genormeerde ruimte is dan kunnen we voor $f : [a, b] \to X$ de definitie van uniform continu overschrijven met $|f(x) - f(y)|$ in plaats van $|f(x) - f(y)|$, zie Definitie 2.22. Als de aardige en nuttige oefening in Sectie 2.5 is gelukt dan heb je nu een bewijs van de volgende stelling.

Stelling 10.8. *Laat X een Banachruimte zijn en $f : [a, b] \to X$ uniform continu. Dan bestaat er een unieke $J \in X$ waarvoor geldt dat*

$$|S - J| \leq \varepsilon(b - a)$$

voor elke S van de vorm (2.7), mits $x_k - x_{k-1} \leq \delta$ voor alle $k = 1, \ldots, \mathrm{N}$ waarmee de bijbehorende partitie P in (2.4) genummerd wordt. Per definitie is

$$J = \int_a^b f(x)\, dx.$$

Dat uniforme continuïteit ook hier uit puntsgewijze continuïteit volgt is ook weer een kwestie van overschrijven.

Opgave 10.9. Als $g : X \to \mathbb{R}$ een lineaire functie is, hetgeen nog steeds betekent dat

$$g(x + y) = g(x) + g(y) \quad \text{en} \quad g(\lambda x) = \lambda g(x) \quad \text{voor alle} \quad x, y \in X,\ \lambda \in \mathbb{R},$$

dan is $x \to g(f(x))$ weer een $\mathbb{R}$-waardige functie. Met een schatting van de vorm

$$|g(x)| \leq L|x| \quad \text{voor alle} \quad x \in X, \tag{10.8}$$

en dat voor een vaste $L \geq 0$, moet uit je bewijs voor Stelling 10.8 nu volgen dat

$$g(J) = \int_a^b g(f(x))\, dx,$$

als integraal van een gewone continue $\mathbb{R}$-waardige functie op $[a, b]$. Verifieer dit eerst voor $L = 1$. Hint: de relevante schattingen voor ondermeer de Riemannsommen bij $x \to g(f(x))$ zijn voor $L = 1$ identiek aan die voor de sommen bij f.

Definitie 10.10. *Een lineaire afbeelding $g : X \to \mathbb{R}$ van een genormeerde ruimte X naar $\mathbb{R}$ heet begrensd als (10.8) geldt voor een zekere $L \geq 0$.*[2]

Opgave 10.11. Bewijs voor lineaire afbeeldingen $g : X \to \mathbb{R}$ van een genormeerde ruimte X naar $\mathbb{R}$ dat

$$g \text{ is uniform Lipschitz continu} \iff g \text{ is begrensd} \iff g \text{ is continu in } x = 0$$

Bewijs ook de volgende stelling.

Stelling 10.12. *Laat X^* de verzameling van alle begrensde lineaire functies $g : X \to \mathbb{R}$ zijn. Dan is X^* met de voor de hand liggende bewerkingen en de (operator)norm*

$$|g|_{op} = \sup_{0 \neq x \in X} \frac{|g(x)|}{|x|_X}$$

een Banachruimte, die de duale ruimte van X genoemd wordt. We schrijven meestal gewoon $|g|$ voor $|g|_{op}$ en $|x|$ voor $|x|_X$, maar in de formule staan drie verschillende normen.

Voor $F : [a, b] \to X$ kunnen we de definitie van differentieerbaarheid in een vaste $x_0 \in [a, b]$ overschrijven van Definitie 4.1 met de expansie

$$F(x_0 + h) = F(x_0) + hF'(x_0) + R(h; x_0), \tag{10.9}$$

waarin

$$h \xrightarrow{F'(x_0)} hF'(x_0)$$

[2] Op $\mathbb{R}^n$ zijn alle lineaire functies begrensd.

met $F'(x_0) \in X$ gezien wordt als continue lineaire afbeelding en de restterm moet voldoen aan

$$|R(h; x_0)| = o(|h|) \quad \text{als} \quad h \to 0. \tag{10.10}$$

Ook Stelling 4.3 kan nu met bewijs en al worden overgeschreven voor $f : [a, b] \to X$ (uniform) continu.

Stelling 10.13. *Laat X een Banachruimte zijn. Definieer voor een continue functie $f : [a, b] \to X$ de functie $F : [a, b] \to X$ door*

$$F(x) = \int_a^x f(s)\, ds.$$

Dan is F differentieerbaar in elke $x_0 \in [a, b]$, met afgeleide $F'(x_0) = f(x_0)$.

Merk op dat in het bewijs de 3-hoeksongelijkheid voor integralen werd gebruikt. Die geldt ook hier, rechtstreeks vanuit de definities. Voor f als in Stelling 10.8 geldt

$$|\int_a^b f(x)\, dx| \leq \int_a^b |f(x)|\, dx,$$

waarbij de continuïteit van $x \to |f(x)|$ volgt uit

$$||f(x)| - |f(y)|| \leq |f(x) - f(y)|,$$

de omgekeerde[3] 3-hoeksongelijkheid toegepast op $f(x)$ en $f(y)$.

Stelling 10.13 zegt weer dat F een primitieve is van kleine f en voor deze primitieve geldt met $x = b$ dat

$$\int_a^b f(s)\, ds = F(b) - F(a), \tag{10.11}$$

omdat $F(a) = 0$. Als $\tilde{F}$ een andere primitieve is van f dan geldt voor

$$G = F - \tilde{F} : [a, b] \to X$$

dat G differentieerbaar is met $G'(x) = 0$ voor alle $x \in [a, b]$.

Opgave 10.14. Laat zien dat voor elke $g \in X^*$ de reëelwaardige functie

$$x \xrightarrow{\phi} g(G(x))$$

[3] Onthouden als $x \to |x|$ is uniform Lipschitz continu met Lipschitz constante $L = 1$.

nu differentieerbaar is op $[a,b]$ met afgeleide in iedere $x \in [a,b]$ gedefinieerd door

$$h \xrightarrow{\phi'(x)} g(G(x))G'(x)h = 0$$

voor $h \in \mathbb{R}$. Dus $\phi(b) = \phi(a)$ via de middelwaardestelling (Stelling 4.5).

We concluderen hieruit dat $g(G(b)) - g(G(a)) = 0$ voor elke $g \in X^*$. Voor $y = G(b) - G(a)$ geldt dus dat $g(y) = 0$ voor alle $g \in X^*$. We zeggen dat de continue lineaire functies de punten in X scheiden als dit impliceert dat $y = 0$. In dat geval zijn we klaar want dan is $F(b) - F(a) = \tilde{F}(b) - \tilde{F}(a)$. Hiermee is de volgende stelling bewezen, waarin we $\tilde{F}$ weer F noemen.

Stelling 10.15. *Laat X een Banachruimte zijn waarop de continue lineaire functies de punten scheiden*[4]*. Als $f : [a,b] \to X$ continu is op $[a,b]$ en de continue differentieerbare functie $F : [a,b] \to X$ als primitieve heeft*[5]*, dan is*

$$\int_a^b f(s)\,ds = F(b) - F(a).$$

10.2 Differentiaalrekening voor functies op X

Ook als X een genormeerde ruimte is en

$$F : X \to \mathbb{R}$$

een reëelwaardige functie[6] op X, dan kunnen we de definitie van differentieerbaarheid in een vaste $x_0 \in X$ overschrijven van Definitie 4.1, beginnende met de expansie

$$F(x_0 + h) = F(x_0) + F'(x_0)(h) + R(h; x_0), \tag{10.12}$$

waarin

$$F'(x_0) : X \to \mathbb{R}$$

lineair is en de restterm moet voldoen aan

$$|R(x_0, h)| = o(|h|) \quad \text{als} \quad |h| \to 0. \tag{10.13}$$

Definitie 10.16. *Een functie $F : X \to \mathbb{R}$ op een genormeerde ruimte heet differentieerbaar in $x_0 \in X$ als er een $g = F'(x_0) : X \to \mathbb{R}$ in X^* bestaat*[7] *zodanig dat (10.12) geldt met (10.13).*

[4] Op Banachruimten scheiden de continue lineaire functies de punten (Hahn-Banach).
[5] Hetgeen ook hier betekent dat $F'(x) = f(x)$ voor alle $x \in [a,b]$.
[6] Iedere $F : A \to \mathbb{R}$ met $A \subset X$ breidt uit tot $F : X \to \mathbb{R}$ via $F(x) = 0$ voor $x \notin A$.
[7] Let op, de continuiteit van de lineaire $F'(x_0) : X \to \mathbb{R}$ is essentieel: Opgave 10.17!

Opgave 10.17. Als een reëelwaardige $F : X \to \mathbb{R}$ op een genormeerde ruimte differentieerbaar in $x_0 \in X$ dan is F ook continu in x_0. Bewijs dit.

Opgave 10.18. Formuleer en bewijs de regel van Leibniz voor $F, G : X \to \mathbb{R}$. Hint: schrijf Opgave 5.7 en je uitwerking van die opgave over met normstrepen in plaats van absolute waarde strepen.

Ook van $F' : X \to X^*$ willen we kunnen zeggen of F' differentieerbaar is in een $x_0 \in X$ en daartoe schrijven we Definitie 10.16 over voor $F : X \to Y$ als X en Y genormeerde ruimten zijn.

Een search op $\to \mathbb{R}$ en replace by $\to Y$ dus, waarbij absolute waarde strepen in $\mathbb{R}$ weer vervangen worden door normstrepen voor de norm in Y, en $F'(x_0) \in X^*$ moet worden vervangen door $F'(x_0) \in L(X, Y)$, de genormeerde ruimte van lineaire afbeeldingen $A : X \to Y$ waarvoor $|Ax|_Y \leq |A|_{op} |x|_X$ voor alle $x \in X$, nu met de norm van Ax in Y, en $|A|_{op}$ weer de kleinste constante M waarvoor $|Ax|_Y \leq M |x|_X$ voor alle $x \in X$. Met $Y = X^*$ hebben we dan ook de definitie van twee keer differentieerbaar, enzovoorts.

Zonder extra algebraische structuur is de Leibnizregel niet meer zo maar over te schrijven, maar de kettingregel gaat weer precies hetzelfde.

Opgave 10.19. Schrijf Opgaven 5.10 en 5.11 en je uitwerking over voor $F : X \to Y$ en $G : Y \to Z$ met X, Y, Z genormeerde ruimten.

Kortom, veel van de gewone differentiaalrekening is precies hetzelfde in de algemene context. Belangrijke uitzonderingen zijn de middelwaardestelling (Stelling 4.5) en de inverse functiestelling zoals geformuleerd voor monotone functies. Beide stellingen maken echt gebruikt gebruik van de 1-dimensionale structuur die $\mathbb{R}$ heeft en de bijbehorende argumenten waarin het woordje "tussen" essentieel is.

De middelwaardestelling kan echter wel mooi in stelling gebracht worden voor

$$F : X \to \mathbb{R},$$

door voor twee punten $x, y \in X$ een functie $t \to \xi(t) \in X$ te introduceren die x en y met elkaar "verbindt", bijvoorbeeld

$$\xi(t) = (1 - t)x + ty, \quad \xi : [0, 1] \to X, \tag{10.14}$$

waarbij we spreken over het interval

$$[x, y] = \{\xi(t) = (1-t)x + ty;\ 0 \leq t \leq 1\}, \tag{10.15}$$

de door $t \in [0, 1]$ geparametriseerde convexe combinaties van x en y, en ook over het interval

$$(x, y) = \{\xi(t) = (1-t)x + ty;\ 0 < t < 1\}. \tag{10.16}$$

Stelling 10.20. *Als $F : X \to \mathbb{R}$ differentieerbaar is op X, dan is er voor elke $x, y \in X$ een $\xi \in (x, y)$ waarvoor geldt dat*

$$F(y) - F(x) = F'(\xi)(y - x).$$

Bewijs. Volgens de kettingregel is $\phi : [0, 1] \to \mathbb{R}$ gedefinieerd door

$$t \xrightarrow{\phi} F(\xi(t)) \in \mathbb{R} \tag{10.17}$$

differentieerbaar op $[0, 1]$, met

$$\phi'(t) = F'(\xi(t))(y - x).$$

De middelwaardestelling (Stelling 4.5) geeft nu een $\tau \in (0, 1)$ waarvoor geldt dat $\phi(1) - \phi(0) = \phi'(\tau)$. Noem $\xi(\tau)$ nu ξ en het bewijs is klaar.

Neem nu een willekeurige Banachruimte Y en $F : X \to Y$ continu differentieerbaar. Als we Stelling 10.15 toepassen met $a = 0$ en $b = 1$ op de functie $\phi : [0, 1] \to Y$ gedefinieerd door (10.17), dan volgt uit

$$\phi(1) - \phi(0) = \int_0^1 \phi'(t)\, dt$$

en

$$\phi'(t) = F'((1-t)x + ty)(y - x)$$

dat

$$F(y) - F(x) = \int_0^1 F'((1-t)x + ty)(y - x)\, dt \tag{10.18}$$

als Y-waardige integraal.

In het geval dat $Y = \mathbb{R}$ wordt deze $\mathbb{R}$-waardige integraal begrensd door het maximum en het minimum van de integrand, en dus volgt dat $F(y) - F(x) = F'(\xi)(y - x)$ voor een zekere

$$\xi \in [x, y] = \{(1-t)x + ty,\ 0 \leq t \leq 1\},$$

een iets zwakkere uitspraak dan in Stelling 10.20.

Als F niet $\mathbb{R}$-waardig maar Y-waardig dan werkt dit argument niet. Echter, als we Stelling 10.15 toepassen met $a = 0$ en $b = 1$ op de functie $\phi : [0,1] \to Y$ gedefinieerd door

$$\phi : t \xrightarrow{\xi} (1-t)x + ty \xrightarrow{F} Y$$

met $x, y \in X$ en $F : X \to Y$ continu differentieerbaar, dan volgt uit

$$\phi(1) - \phi(0) = \int_0^1 \phi'(t)\, dt$$

en

$$\phi'(t) = F'((1-t)x + ty)(y - x)$$

dat

$$F(y) - F(x) = \int_0^1 F'((1-t)x + ty)(y-x)\, dt \tag{10.19}$$

als Y-waardige integraal, hetgeen ook te schrijven is als

$$F(y) - F(x) = \int_0^1 F'((1-t)x + ty)\, dt(y-x), \tag{10.20}$$

een $B(X,Y)$-waardige integraal werkend op $y - x \in X$.

Vanwaar al deze uitwijdingen? Omdat formulering en bewijs van de impliciete functiestelling voor $F : X \times Y \to Z$ nu kunnen worden overgeschreven van Stelling 4.24 in Sectie 4.4. Speciaal geval $Z = Y$. Nog specialer: $X = \mathbb{R}^m$ en $Y = Z = \mathbb{R}^n$. Ook interessant: $X = \mathbb{R}$ en $Y = Z = C([a,b])$. Denk aan Sectie 10.5.

10.3 Partieel differentieerbaar $\Longrightarrow$?

We hebben gezien dat $F : \mathbb{R}^2$ differentieerbaar is in x_0, y_0 als er $a, b \in \mathbb{R}$ bestaan zo dat (9.8) geldt en de restterm voldoet aan de conditie onder (9.8). In dat geval bestaan de partiële afgeleiden

$$a = (\frac{\partial F}{\partial x})(x_0, y_0) = F_x(x_0, y_0) \quad \text{en} \quad b = (\frac{\partial F}{\partial y})(x_0, y_0) = F_x(x_0, y_0),$$

en leest (9.8) met $x = x_0 + h$ en $y = y_0 + k$ als

$$F(x,y) = F(x_0, y_0) + F_x(x_0, y_0)(x - x_0) + F_y(x_0, y_0)(y - y_0) + \text{restterm}.$$

We gebruiken de subscripts hier dus weer voor de partiële afgeleiden die we in Opgave 9.5 gevonden hebben voor a en b als

$$a = F_x(x_0, y_0) = \lim_{h\to 0} \frac{F(x_0+h, y_0) - F(x_0, y_0)}{h} = \lim_{x\to x_0} \frac{F(x, y_0) - F(x_0, y_0)}{x - x_0},$$

$$b = F_y(x_0, y_0) = \lim_{k\to 0} \frac{F(x_0, y_0+k) - F(x_0, y_0)}{k} = \lim_{y\to y_0} \frac{F(x_0, y) - F(x_0, y_0)}{y - y_0},$$

door de eigenschap van de restterm te gebruiken.

Het kan natuurlijk zijn dat voor een gegeven $F : \mathbb{R}^2 \to \mathbb{R}$ en $x_0 \in \mathbb{R}$ en $y_0 \in \mathbb{R}$ de partiële afgeleiden $F_x(x_0, y_0)$ en $F_y(x_0, y_0)$ zo bestaan als twee limieten, maar op zich zegt dat weinig. Als bijvoorbeeld $F : \mathbb{R}^2 \to \mathbb{R}$ gedefinieerd is door $F(x, y) = 0$ als $xy = 0$ en $F(x, y) = 1$ als $xy \neq 0$ dan is $F_x(0, 0) = F_y(0, 0) = 0$, maar F is zeker niet differentieerbaar in $(0, 0)$, want de lineaire benadering zou dan de nulfunctie zijn en die komt voor $xy \neq 0$ niet de buurt van de functiewaarde 1.

Dus wat hebben we nodig van $x \to F(x, y)$ en $y \to F(x, y)$ als we alleen deze "partiële" functies mogen gebruiken om te kunnen concluderen dat

$$F : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$$

differentieerbaar is in (x_0, y_0)? We denken bij deze vraag aan het algemenere geval van

$$F : X \times Y \to \mathbb{R},$$

en $x_0 \in X, y_0 \in Y$, en nemen aan dat $x \to F(x, y)$ en $y \to F(x, y)$ voor respectievelijk vaste $y \in B_\delta(y_0)$ en vaste $x \in B_\delta(x_0)$ differentieerbaar[8] zijn op respectievelijk $B_\delta(x_0)$ en $B_\delta(y_0)$, met $\delta_0 > 0$ vast.

Via de middelwaardestelling (Stelling 10.20) voor functies van genormeerde ruimten naar $\mathbb{R}$ is

$$F(x, y) = F(x_0, y_0) + F(x, y) - F(x_0, y_0) =$$

$$F(x_0, y_0) + \underbrace{F(x, y) - F(x_0, y)}_{\text{vary } x} + \underbrace{F(x_0, y) - F(x_0, y_0)}_{\text{vary } x} =$$

$$F(x_0, y_0) + F_x(\xi(y), y)(x - x_0) + F_y(x_0, \eta)(y - y_0),$$

voor $x \in B_\delta(x_0)$ en $y \in B_\delta(y_0)$ met $\xi(y) \in (x_0, x)$ en $\eta \in (y_0, y)$. Bijgevolg is

$$F(x, y) = F(x_0, y_0) + F_x(x_0, y_0)(x - x_0) + F_y(x_0, y_0)(y - y_0) + R \quad (10.21)$$

[8] We gebruiken van $y \to F(x, y)$ alleen maar de differentieerbaarheid voor $y = y_0$.

met restterm

$$R = (F_x(\xi(y), y) - F_x(x_0, y_0))(x - x_0) + (F_y(x_0, \eta) - F_y(x_0, y_0))(y - y_0).$$

Als

$$(x, y) \to F_x(x, y) \quad \text{en} \quad y \to F_y(x_0, y)$$

continu zijn in respectievelijk (x_0, y_0) en y_0 dan voldoet R nu aan

$$|R| \leq |(F_x(\xi(y), y) - F_x(x_0, y_0))(x - x_0)| + |(F_y(x_0, \eta) - F_y(x_0, y_0))(y - y_0)| \leq$$
$$\underbrace{|F_x(\xi(y), y) - F_x(x_0, y_0)|}_{\leq \varepsilon} |x - x_0| + \underbrace{|F_y(x_0, \eta) - F_y(x_0, y_0)|}_{\leq \varepsilon} |y - y_0)|$$
$$\leq \varepsilon \max(|x - x_0|, |y - y_0|) = \varepsilon |(x, y) - (x_0, y_0)|$$

als $\delta > 0$ voldoende klein is. We gebruiken hier als norm op $X \times Y$ de norm

$$|(x, y)| = \max(|x|, |y|).$$

We onthouden dit resultaat als de volgende stelling waarin we de mildere voorwaarde van continuiteit van $y \to F_y(x_0, y)$ vervangen door de voor de hand liggende sterkere voorwaarde op $(x, y) \to F_y(x_0, y)$.

Stelling 10.21. *Laat X en Y genormeerde ruimten zijn. Als $F : X \times Y \to \mathbb{R}$ partiële functies*

$$x \to F(x, y) \quad \text{en} \quad y \to F(x, y)$$

heeft die voor $x \in B_\delta(x_0)$ en $y \in B_\delta(y_0)$ met $x_0 \in X, y_0 \in Y, \delta > 0$ gedefinieerd en differentieerbaar zijn, met

$$(x, y) \to F_x(x, y) \in X^* \quad \text{en} \quad (x, y) \to F_y(x, y) \in Y^*$$

continu in (x_0, y_0), dan is F differentieerbaar in (x_0, y_0), met $F'(x_0, y_0)$ gedefinieerd door

$$(h, k) \xrightarrow{F'(x_0, y_0)} F_x(x_0, y_0)h + F_y(x_0, y_0)k.$$

Opgave 10.22. De karakterisatie van $F'(x_0, y_0)$ wekt natuurlijk geen verbazing. Als $F : X \times Y$ differentieerbaar is in (x_0, y_0) dan zijn zowel $x \to F(x, y_0)$ als $y \to F(x_0, y)$ differentieerbaar in respectievelijk x_0 en y_0 en hun afgeleiden heten de partiële afgeleiden, notatie o.a. $F_x(x_0, y_0)$, $F_y(x_0, y_0)$. Verifieer dit rechtstreeks aan de hand van de definities, en ook dat

$$F'(x_0, y_0)\begin{pmatrix} h \\ k \end{pmatrix} = F_x(x_0, y_0)h + F_y(x_0, y_0)k$$

voor alle $h \in X$ en $k \in Y$.

Opgave 10.23. Als X, Y, Z genormeerde ruimten zijn en $F : X \times Y \to Z$ dan werkt de bewijstechniek met de middelwaardestelling niet. Schrijf echter weer

$$F(x,y) = F(x_0,y_0) + \underbrace{F(x,y) - F(x_0,y)}_{\text{vary } x} + \underbrace{F(x_0,y) - F(x_0,y_0)}_{\text{vary } x}$$

en behandel de twee variërende termen met (10.18), waarbij je moet aannemen dat Z volledig is. Neem hiertoe aan dat $x \to F(x,y_0)$ continu differentieerbaar is voor $x \in X$ met $|x - x_0| < \delta_x$, en dat voor elke van deze x de functie $y \to F(x,y)$ continu differentieerbaar is voor $y \in Y$ met $|y - y_0| < \delta_y$, $\delta_x, \delta_y > 0$, en bovendien dat $(x,y) \to F_y(x,y)$ continu is in (x_0,y_0). Bewijs dat F differentieerbaar is in (x_0,y_0).

Opgave 10.24. Als X, Y, Z genormeerde ruimten zijn, met Z volledig, en $F : X \times Y \to Z$ heeft partiële functies met partiële afgeleiden F_x en F_y die continu zijn op een open verzameling O in $X \times Y$, dan is F differentieerbaar in elk punt van O en $F' : O \to L(X \times Y, Z)$ is continu, gedefinieerd in elke $(x_0,y_0) \in O$ als in Opgave 10.22.

10.4 Differentiaal- en integraalvergelijkingen

Terug naar functies van $t \in \mathbb{R}$ naar $x \in \mathbb{R}$. In (5.15) is de differentiaalvergelijking die de functie exp definieert omgeschreven tot integraalvergelijking. Vanaf hier kiezen we zo veel mogelijk voor de t van tijd als onafhankelijke variabele in de differentiaalvergelijkingen onder consideratie, en voor x als de afhankelijke variabele. Daarbij we spreken over $x = x(t)$ als de oplossing, en de lezer wordt gevraagd deze "abuse of notation" voor lief te nemen. Dus $x = x(t)$ voor $t \in I$ is een verkorte schrijfwijze om een functie $x : I \to \mathbb{R}$ aan te duiden.

Bijvoorbeeld $x(t) = \exp(t)$, de unieke oplossing van

$$x' = x \quad \text{met} \quad x(0) = 1 \quad \text{op} \quad I = \mathbb{R},$$

een probleem van de vorm

$$x'(t) = f(x(t)) \quad \text{voor alle} \quad t \in I \quad \text{met} \quad x(0) = \xi. \tag{10.22}$$

Merk op dat een oplossing van (10.22) per definitie differentieerbaar is op I en dus continu.

Voor iedere $f : \mathbb{R} \to \mathbb{R}$ en $\xi \in \mathbb{R}$ is (10.22) een probleem om op te lossen voor t in een liefst maximaal interval I, met $0 \in I$ omdat anders de

voorwaarde $x(0) = \xi$ geen betekenis heeft. We beperken ons hier nu tot differentiaalvergelijkingen waarin de onafhankelijke variabele t niet expliciet voorkomt. Dus $x'(t) = f(x(t))$ en niet $x'(t) = f(t, x(t))$.

Zulke differentiaalvergelijkingen hebben de eigenschap dat een horizontaal verschoven grafiek van een oplossing in het t, x-vlak weer de grafiek van een oplossing is. Met één oplossing heb je er dankzij deze *translatie-variantie* dus meteen een heleboel en soms zelfs allemaal (als f geen nulpunten heeft).

Opgave 10.25. Teken in het x, y-vlak een grafiek $y = f(x)$ van een mooie functie f met een paar nulpunten en gebruik de informatie uit die grafiek om grafieken van oplossingen in t, x-vlak te schetsen. Hint: schets eerst de grafieken van de constante oplossingen.

Onder de aanname dat f continu is op $\mathbb{R}$ is (10.22) equivalent met

$$x(t) = \xi + \int_0^t f(x(s))\, ds \quad \text{voor alle} \quad t \in I. \tag{10.23}$$

Opgave 10.26. Bewijs de equivalentie van (10.22) en (10.23) voor continue $f : \mathbb{R} \to \mathbb{R}$ en $\xi \in \mathbb{R}$ gebruikmakend van de integraal- en differentiaalrekening ontwikkeld tot nu toe.

Opgave 10.27. Dezelfde vraag (met $f : \mathbb{R} \to \mathbb{R}$ continu en $\xi, \eta \in \mathbb{R}$) voor

$$x'' = f(x) \quad \text{met} \quad x(0) = \xi \quad \text{en} \quad x'(0) = \eta$$

en de integraalvergelijking die je krijgt door twee keer te integreren. Hint: gebruik je uitwerking van Opgave 4.36 om de herhaalde integraal om te schrijven tot een enkele integraal.

Om een oplossing van (10.23) te maken kunnen we eerst $a, b \in \mathbb{R}$ kiezen met $0 \in [a, b] \subset I$ en (10.23) herschrijven als een vergelijking

$$x = \xi + \Phi(x), \tag{10.24}$$

voor $x \in C([a, b])$, met $\Phi : C([a, b]) \to C([a, b])$ gedefinieerd door

$$(\Phi(x))(t) = \int_0^t f(x(s))\, ds, \tag{10.25}$$

een formulevoorschrift voor de "nieuwe" functie $\Phi(x) \in C([a,b])$ gegeven de "oude" functie $x \in C([a,b])$.

In (10.24) is ξ nu te lezen als een constante functie $x_0 : t \to \xi$ en met het schema

$$x_{n+1} = \xi + \Phi(x_n) \quad (n \in \mathbb{N}_0) \tag{10.26}$$

ontstaat dan een rij x_n in $C([a,b])$ waarvan we nu willen weten of deze via zijn limiet een unieke oplossing van (10.23) definieert. De vraag nu is onder welke aannames op f en $[a,b]$ dit lukt. Om dat te onderzoeken nemen we eerst zonder beperking der algemeenheid $\xi = 0$.

In wat volgt vermijden we notationeel gepriegel met bovengrenzen t die kleiner zijn dan de ondergrens 0 in (10.25), door de afbeelding Φ te bekijken op $X = C([0,T])$. Vaste punten van Φ zijn dan oplossingen van (10.23) met $\xi = 0$ en daarmee van (10.22). Oplossingen van (het beginwaardeprobleem)

$$x'(t) = f(x(t)) \quad \text{voor} \quad t \in [0,T] \quad \text{en} \quad x(0) = 0$$

corresponderen dus met vaste punten van $\Phi : X \to X$. En voor het vinden van zulke "dekpunten" is er een stelling die iedereen die wiskunde doceert gezien moet hebben[9].

Stelling 10.28. *Laat X een Banachruimte zijn. Als $\Phi : X \to X$ een afbeelding is die voor een vaste $\theta \in [0,1)$ voldoet aan*

$$|\Phi(x) - \Phi(y)| \leq \theta |x - y| \quad \forall x, y \in X,$$

dan heeft Φ een uniek vast punt $\bar{x} = \Phi(\bar{x}) \in X$. Gegeven iedere $x_0 \in X$ geldt voor de rij gedefinieerd door $x_n = \Phi(x_{n-1})$ voor alle $n \in \mathbb{N}$ dat $x_n \to \bar{x}$ als $n \to \infty$.

Bewijs. Dat Φ geen twee vaste punten kan hebben volgt meteen uit de aanname. Gegeven een willekeurige x_0 en $x_1 = \Phi(x_0)$ volgt ook uit de aanname dat

$$|x_2 - x_1| = |\Phi(x_1) - \Phi(x_0)| \leq \theta |x_1 - x_0|,$$
$$|x_3 - x_2| = |\Phi(x_2) - \Phi(x_1)| \leq \theta |x_2 - x_1| \leq \theta^2 |x_1 - x_0|,$$

etc. Voor de incrementen geldt dus dat[10]

$$|x_{n+1} - x_n| \leq \theta^n |x_1 - x_0|, \tag{10.27}$$

[9] We hebben deze stelling al een keer niet gebruikt toen het wel kon. In welke bewijs?
[10] Eigenlijk moet dit weer met inductie.....

en dus volgt

$$\sum_{n=0}^{\infty}|x_{n+1}-x_n| \le \sum_{n=0}^{\infty}\theta^n|x_1-x_0| = |x_1-x_0|\sum_{n=0}^{\infty}\theta^n = |x_1-x_0|\frac{1}{1-\theta}.$$

De reeks

$$\sum_{n=0}^{\infty}(x_{n+1}-x_n)$$

is dus convergent met een limiet $S \in X$ vanwege de uitspraak in Opgave 10.29 hieronder.

Omdat

$$x_{\mathrm{N}+1} = x_0 + \sum_{n=0}^{\mathrm{N}}(x_{n+1}-x_n)$$

volgt dat

$$x_{\mathrm{N}+1} = \Phi(x_{\mathrm{N}}) \to x_0 + S \quad \text{als} \quad \mathrm{N}\to\infty.$$

Met $\bar{x} = x_0 + S$ hebben we de limiet van de rij x_n dus te pakken en omdat

$$\begin{aligned}|\Phi(\bar{x})-\bar{x}| &\le |\Phi(\bar{x})-\Phi(x_n)| + |\Phi(x_n)-x_n| + |x_n-\bar{x}|\\ &\le |x_{n+1}-x_n| + 2|x_n-\bar{x}| \to 0\end{aligned}$$

is $\bar{x}$ een vast punt van Φ. Daarmee is het bewijs klaar.

Opgave 10.29. Laat X een Banachruimte zijn en $a_n \in X$ voor $n \in \mathbb{N}_0$. Schrijf[11] het bewijs van het convergent zijn van absoluut convergente reeksen in $\mathbb{R}$ over met normstrepen i.p.v. absolute waarde strepen en bewijs zo dat

$$\sum_{n=0}^{\infty}|a_n| < \infty \implies \sum_{n=0}^{\infty}a_n \quad \text{bestaat en} \quad |\sum_{n=0}^{\infty}a_n| \le \sum_{n=0}^{\infty}|a_n|.$$

Opgave 10.30. Neem aan dat $f : \mathbb{R} \to \mathbb{R}$ uniform Lipschitz continu is met Lipschitz constante $L > 0$. Dus

$$|f(x)-f(y)| \le L|x-y| \quad \forall x,y \in \mathbb{R}.$$

Bewijs dat Φ in (10.25) voldoet aan

$$|\Phi(x)-\Phi(y)|_\infty \le LT|x-y|_\infty \quad \forall x,y \in X = C([0,T]).$$

[11] Zie Stelling 3.18.

Stelling 10.31. *Neem aan dat $f : \mathbb{R} \to \mathbb{R}$ uniform Lipschitz continu is, en dat $\xi \in \mathbb{R}$. Dan heeft de differentiaalvergelijking*

$$x'(t) = f(x(t))$$

precies één oplossing met $x(0) = \xi$. Die oplossing is gedefinieerd voor alle $t \in \mathbb{R}$.

Bewijs. Met $\xi = 0$ volgt het bestaan van een unieke oplossing op $[0, T]$ via Opgave 10.26, Stelling 10.28 en Opgave 10.30 onder de aanname dat $LT < 1$.

Voor $\xi \neq 0$ is het bewijs precies hetzelfde, niet alleen op $[0, T]$ maar ook op $[-T, 0]$, of in één keer op $[-T, T]$. Maar vanwege de translatie-invariantie geldt dit ook voor oplossingen met $x(T) = \eta \in \mathbb{R}$ op $[T, 2T]$.

Kies nu $\eta = x(T)$ met x de oplossing die we al hadden op $[0, T]$. Dit geeft een oplossing op $[T, 2T]$ met $x(0) = \xi$. Enzovoort. Stap voor stap, naar links en naar rechts, steeds met dezelfde stapgrootte T, breiden we de oplossing zo uit tot uiteindelijk heel $\mathbb{R}$ en het bewijs is klaar.

Opgave 10.32. Neem aan dat $f : \mathbb{R} \to \mathbb{R}$ uniform Lipschitz continu is en $\tau, \xi \in \mathbb{R}$ willekeurig. Als we de unieke oplossing van $x'(t) = f(x(t))$ met $x(\tau) = \xi$ noteren als $x(t; \tau, \xi)$ dan geldt $x(t; \tau, \xi) = x(t - \tau; 0, \xi)$. Leg uit waarom. Definieer nu de afbeeldingen $S_t : \mathbb{R} \to \mathbb{R}$ door $S_t(\xi) = x(t; 0, \xi)$. Leg uit waarom voor alle $s, t \in \mathbb{R}$ geldt dat $S_{t+s} = S_t(S_s(\xi))$ voor alle $\xi \in \mathbb{R}$, i.e.

$$S_{t+s} = S_t \circ S_s.$$

We schrijven ook $S_t(\xi) = x(t; \xi)$ en dan luidt deze regel dat

$$x(t + s; \xi) = x(t; x(s; \xi)) = x(s; x(t; \xi)).$$

Gebruik dit met $f(x) = x$ om te concluderen dat $\exp(t + s) = \exp(t) \exp(s)$, waarbij $\exp(t)$ gedefinieerd is door $\exp(t) = x(t; 1)$.

Opgave 10.33. Neem aan dat $f : \mathbb{R} \to \mathbb{R}$ uniform Lipschitz continu is, en dat $\xi, \eta \in \mathbb{R}$. Dan heeft de differentiaalvergelijking

$$x''(t) = f(x(t))$$

precies één oplossing met $x(0) = \xi$ en $x'(0) = \eta$. Die oplossing is gedefinieerd voor alle $t \in \mathbb{R}$. Schets een bewijs van deze uitspraak gebruikmakend van Opgave 10.27 en de strategie in het voorafgaande.

Opgave 10.34. De oplossing in Opgave 10.33 kan genoteerd worden als $x(t;\xi,\eta)$. Herschrijf $x(t+s;\xi,\eta)$ zoals in Opgave 10.32. Gebruik dit met $f(x) = -x$ om de bekende regels voor $\cos(s+t)$ en $\sin(s+t)$ af te leiden rechtstreeks vanuit de definitie dat $\sin(t) = x(t;0,1)$ en $\cos(t) = x(t;1,0)$.

Opgave 10.35. Verifieer dat $\exp, \cos, \sin$ als machtreeksoplossingen samenvallen met de oplossingen hierboven door de iteratieschema's voor de betreffende integraalvergelijkingen expliciet uit te voeren.

10.5 Impliciete functiestelling: integraalvergelijkingen

We besluiten met het uitwerken onder welke voorwaarden x in (10.24) een continu differentieerbare functie van $\xi \in \mathbb{R}$ wordt. Zie Sectie 7 in

`http://www.few.vu.nl/~jhulshof/NOTES/fa.pdf`

Voor $a, b \in \mathbb{R}$ met $0 \in [a,b]$ en $\xi \in \mathbb{R}$ gaat dit om de vergelijking

$$x = \xi + \Phi(x) \quad \text{met} \quad (\Phi(x))(t) = \int_0^t f(x(s))\,ds, \tag{10.28}$$

als formulevoorschrift voor de (nieuwe) functie $\Phi(x) \in C([a,b])$ gegeven de ("oude") functie $x \in C([a,b])$, waarbij we aannemen dat $f : \mathbb{R} \to \mathbb{R}$ tenminste continu is. De impliciete functiestelling is toepasbaar als

$$\Phi : C([a,b]) \to C([a,b])$$

continu differentieerbaar is.

Immers, neem $h \in C([a,b])$ en schrijf, de neus volgend,

$$(\Phi(x+h))(t) = \int_0^t f(x(s)+h(s))\,ds = \int_0^t [f(x(s)+\tau h(s))]_0^1\,ds$$

$$= \int_0^t \int_0^1 f'(x(s)+\tau h(s))h(s)\,d\tau\,ds$$

$$= \int_0^t \int_0^1 f'(x(s))h(s)\,d\tau\,ds + \underbrace{\int_0^t \int_0^1 (f'(x(s)+\tau h(s)) - f(x(s))h(s)\,d\tau\,ds}_{R(h;x)(t)}$$

$$= (\Phi'(x)h)(t) + R(h;x)(t),$$

waarin

$$h \xrightarrow{\Phi'(x)} \Phi'(x)h \quad \text{met} \quad (\Phi'(x)h)(t) = \int_0^t f'(x(s))h(s)\,ds, \tag{10.29}$$

en

$$\begin{aligned}
|R(h;x)(t)| &= |\int_0^t \int_0^1 (f'(x(s)+\tau h(s)) - f(x(s))h(s)\,d\tau\,ds| \\
&\le |\int_0^t |\int_0^1 (f'(x(s)+\tau h(s)) - f(x(s))h(s)\,d\tau|\,ds| \\
&\le |\int_0^t |\int_0^1 |(f'(x(s)+\tau h(s)) - f(x(s))h(s)|\,d\tau|\,ds| \\
\le |\int_0^t |\int_0^1 &\underbrace{|f'(x(s)+\tau h(s)) - f(x(s)|}_{\le\varepsilon} \underbrace{|h(s)|}_{|h|_\infty}\,d\tau|\,ds| \le (b-a)\varepsilon|h|_\infty
\end{aligned}$$

als $|h|_\infty \le \delta$, met $\delta > 0$ bij $\varepsilon > 0$ uit de definitie van uniforme continuiteit van $f' : [-A, A] \to \mathbb{R}$ op een interval gegeven door bijvoorbeeld $A = |x|_\infty$ plus een beetje.

Opgave 10.36. Op zijn kortst door de bocht: pas de impliciete functiestelling toe om te laten zien dat voor de oplossing $x = x(\xi)$ van (10.28) geldt dat

$$\xi \to x(\xi)$$

continu differentieerbaar is, en x_ξ de oplossing is van de integraalvergelijking die hoort bij

$$y'(t) = f'(x(t))y(t) \quad \text{met} \quad y(0) = 1,$$

eerst onder de voorwaarde dat a en b dicht bij 0 zitten.

10.6 Calculus in Banachalgebra's van operatoren

Deze sectie is nog wat schetsmatig maar niettemin precies. We willen (9.47) uitwerken voor $A \in L(X)$ en schrijven met z vervangen door λ

$$f(A) = \frac{1}{2\pi i} \oint_P f(\lambda)(\lambda - A)^{-1}\,d\lambda, \tag{10.30}$$

nu voor een willekeurig polygon[12] met hoekpunten $\lambda_1, \ldots, \lambda_n = \lambda_0$, waarop en waarbuiten[13]

$$\lambda \to (\lambda - A)^{-1} = (\lambda I - A)^{-1} \tag{10.31}$$

[12] Of een vereniging daarvan.
[13] Wat bedoelen we daarmee?

gedefinieerd is. Het complement van het domein van (10.31) in $\mathbb{C}$ heet het spectrum van A, notatie $\sigma(A)$. Het domein zelf heet de resolvente verzameling, notatie $\rho(A)$, en de afbeelding in (10.31) heet de resolvente van A.

Opgave 10.37. Gebruik berekeningen met meetkundige reeksen om te laten zien dat iedere $\lambda \in \mathbb{C}$ met $\lambda > |A|$ in $\rho(A)$ ligt en dat $\rho(A)$ open is. Bewijs ook dat (10.31) complex differentieerbaar is op $\rho(A)$. Wat is de afgeleide?

Opgave 10.38. Kan het zijn dat $\rho(A) = \mathbb{C}$? Het antwoord is nee, maar dat vergt nog een argument dat we weer zo licht mogelijk willen houden. Uit het ongerijmde, we zouden dan hebben dat (10.31) een $L(X)$-waardige functie definieert die naar $0 \in L(X)$ gaat als $\lambda \to \infty$ en dat moet niet kunnen, met een argument dat over te schrijven zou moeten zijn van wat we voor gewone complexwaardige functies weten, zie Opgave 9.27.

In deze opgaven heb je niet gebruikt dat met $AB = BA = I$ en $A \in L(X)$ ook volgt dat $B \in L(X)$, een wat diepere stelling voor Banachruimten, die ook maar eens heel kort en clean moet worden uitgelegd. Dat komt nog wel een keer. Denk in het vervolg voorlopig bijvoorbeeld eerst aan $X = \mathbb{C}^2$ als complexe uitbreiding van $\mathbb{R}^2$ en A een lineaire afbeelding gegeven door een 2×2 matrix, met complexe of reële entries. In dat geval bestaat $\sigma(A)$ meestal uit 2 punten, en met twee disjuncte driehoekjes Δ_1 en Δ_2 om die punten heen kunnen we al aan de slag met ieder paar complex differentieerbare functies

$$f_1 : \Delta_1 \to \mathbb{C} \quad \text{en} \quad f_1 : \Delta_1 \to \mathbb{C}$$

die samen één functie

$$f : \Delta_1 \cup \Delta_2 \to \mathbb{C}$$

maken waarvan de twee stukken elkaar niet zien. Maar ook het rechterlid van (4.51) gezien als afbeelding van een gecomplexificeerde $X = C([0,1])$ naar zichzelf is een voorbeeld.

In het algemeen kan $\sigma(A)$ van alles zijn en daarom kijken we nu eerst wat voor gebieden we met eindig veel disjuncte polygonen kunnen maken. Elk polygon P heeft op natuurlijke manier een binnengebied C en een buitengebied U, waar we steeds de rand bijnemen, dus

$$P = U \cap C.$$

Als binnen een polygon P_0 een aantal kleinere polygonen $P_1, \ldots, P_n$ ligt, wier binnengebieden onderling disjunct zijn, dus

$$C_i \cap C_j = \emptyset \quad \text{als} \quad i \neq j \quad \text{voor} \quad i, j = 1, \ldots, n,$$

dan kan het zijn dat

$$\sigma(A) \subset K_{int} \subset K = C_0 \cap U_1 \cap \cdots \cap U_n, \tag{10.32}$$

waarbij we K zien als begrensd door de buitenkant P_0 naarbuiten en door binnenkanten $P_1, \ldots, P_n$ naar binnen, en

$$K_{int} = K \cap P_0^c \cap \cdots \cap P_n^c$$

de doorsnijding van K met de complementen van de polygonen $P_0, \ldots, P_n$ is, dus alles in K dat niet op de rand ligt. Als we polygonen *altijd* als linksom doorlopen zien dan schrijven we in dit geval

$$f(A) = \frac{1}{2\pi i} \oint_{\delta K} f(\lambda)(\lambda - A)^{-1}\, d\lambda =$$

$$\frac{1}{2\pi i} \left(\oint_{P_0} f(\lambda)(\lambda - A)^{-1}\, d\lambda - \sum_{j=1}^{n} \oint_{P_j} f(\lambda)(\lambda - A)^{-1}\, d\lambda \right) \tag{10.33}$$

voor $f : K \to \mathbb{C}$ complex differentieerbaar.

Ligt $\sigma(A)$ in een disjuncte eindige vereniging

$$K_1 \cup \cdots \cup K_m$$

van zulke K_j, en zijn

$$f_j : K_j \to \mathbb{C} \quad (j = 1, \ldots, m)$$

complex differentieerbaar, dan vormen die samen weer een complex differentieerbare functie

$$f : K = K_1 \cup \cdots \cup K_n \to \mathbb{C}$$

waarvoor we

$$f(A) = \frac{1}{2\pi i} \sum_{j=1}^{m} \oint_{\delta K_j} f(\lambda)(\lambda - A)^{-1}\, d\lambda \tag{10.34}$$

met iedere term in de som gedefinieerd als in (10.33) als definitie van $f(A)$ gebruiken.

Elk van de K_j kan van de vorm alleen maar $K_j = C_j$ zijn, en één $K = C$ is altijd mogelijk om dat $\sigma(A)$ begrensd is, maar hoe kleiner K gekozen wordt,

hoe meer speelruimte er is. De mogelijk steeds grotere[14] uitdrukkingen voor K moet daarbij graag op de koop toe worden genomen, en als $K_j \neq C_j$ kunnen de bijbehorende buitenkanten ook genest liggen. In het simpele geval dat $\sigma(A)$ een eindige discrete puntverzameling is kunnen we natuurlijk toe met $K = C_j = \Delta_j$, met de driehoekjes Δ_j zo klein als we maar willen en

$$\sigma(A) \subset K^{int} \subset K = \Delta_1 \cup \cdots \cup \Delta_m.$$

Wat we nu sowieso in alle gevallen willen is dat, als we de hoekpunten van de polygonen een beetje naar binnen schuiven, K in dus, de integralen die in (10.33) en (10.34) de nieuwe lineaire afbeelding $f(A) : X \to X$ moeten maken, niet veranderen. En hetzelfde als we K groter maken door de punten naar buiten te schuiven, zolang we maar niet uit het definitiegebied van de continu differentieerbare complexwaardige f lopen. Bij het verder kleiner of groter maken kan de structuur van K versimpelen als twee polygonen elkaar ontmoeten en vervolgens samen één polygon vormen. Strict genomen hebben we niet nodig hoe dat precies kan gaan, maar het is toch aardig om daar even over na te denken.

Opgave 10.39. Het is een aardige project om dat versimpelen precies te maken. Bij het groter maken van K kunnen twee buitenkanten van twee K_j-tjes elkaar ontmoeten waarna verder groter maken tot één nieuwe buitenkant leidt waarmee de bijbehorende binnenkanten dan samen de nieuwe binnenkanten van een nieuwe K_j worden. Ook kan uit een groeiende buitenkant die binnen een krimpende binnenkant ligt meteen na het eerste contact één nieuwe binnenkant ontstaan. Bij kleiner maken kunnen een binnen- en een buitenkant van eenzelfde K_j-tje elkaar ontmoeten en daarna een nieuwe buitenkant vormen, en ook kunnen twee binnenkanten elkaar ontmoeten en een nieuwe binnenkant vormen. Ga in alle gevallen na wat de nieuwe structuur wordt en welke andere scenario's er nog zijn, zoals ondermeer polygonen tot een punt laten krimpen en verdwijnen.

Via de inmiddels vertrouwde zigzagkrommen vernandert bij het geschuif met de hoekpunten (10.34) niet, mits de Stelling van Coursat geldt voor driehoekjes waarop en waarbinnen (10.31) complex differentieerbaar is. De betreffende integralen bestaan weer uit integralen over lijnstukjes. Integralen die gedefinieerd zijn dankzij het werk in Sectie 10.1, waarvoor het voorwerk al in Sectie 2.5 was gedaan: continue $L(X)$-waardige functies van $t \in [0, 1]$ zijn integreerbaar via de tussensommen van Riemann, en

$$t \to f((1-t)\lambda_{k-1} + t\lambda_k)((1-t)\lambda_{k-1} + t\lambda_k - A)^{-1}$$

[14] Als we alle zijden van alle poygonen dicht bij $\sigma(A)$ willen hebben.

is zo'n functie waarmee $L(X)$-waardige integralen als

$$\int_{\lambda_{k-1}}^{\lambda_k} f(\lambda)(\lambda - A)^{-1}\, d\lambda$$

nu gedefinieerd zijn.

Mooi, dan kan voor

$$\lambda \to f(\lambda)(\lambda - A)^{-1}$$

de Stelling van Coursat met bewijs en al worden overgeschreven[15] en is (10.33) een goede definitie van $f(A)$. Voorlopig houden we nu A vast en kijken naar nog zo'n f, een g dus, waarbij we eerst aannemen dat we het allersimpelste geval hebben, één polygon rond $\sigma(A)$ waarmee de berekeningen gedaan worden. In dat geval is de samenstelling van de afbeeldingen $f(A)$ en $g(A)$ te schrijven als

$$f(A)g(A) = \frac{1}{2\pi i} \oint_{\lambda_{0-n}} f(\lambda)(\lambda - A)^{-1}\, d\lambda\, \frac{1}{2\pi i} \oint_{\mu_{0-n}} g(\mu)(\mu - A)^{-1}\, d\mu,$$

met in de Cauchyintegraal voor $g(A)$ de hoekpunten μ_l een klein beetje naar binnen geschoven hebben, niet omdat het moet, maar omdat het kan, iets minder ver naar binnen dan de hoekpunten λ_k. Het μ-polygon komt zo binnen het λ-polygon te liggen.

Omdat

$$f(A) = \frac{1}{2\pi i} \sum_{k=1}^{n} \int_{\lambda_{k-1}}^{\lambda_k} f(\lambda)(\lambda - A)^{-1}\, d\lambda,$$

$$g(A) = \frac{1}{2\pi i} \sum_{l=1}^{n} \int_{\mu_{l-1}}^{\mu_l} g(\mu)(\mu - A)^{-1}\, d\mu,$$

wordt $f(A)g(A)$ afgezien van de voorfactoren dankzij overwegingen als bij (6.47) een som van produkten

$$\int_{\lambda_{k-1}}^{\lambda_k} f(\lambda)(\lambda - A)^{-1}\, d\lambda \int_{\mu_{l-1}}^{\mu_l} g(\mu)(\mu - A)^{-1}\, d\mu =$$

$$\int_{\mu_{l-1}}^{\mu_l} \int_{\lambda_{k-1}}^{\lambda_k} f(\lambda)g(\mu)(\lambda - A)^{-1}(\mu - A)^{-1}\, d\lambda\, d\mu =$$

$$\int_{\lambda_{k-1}}^{\lambda_k} \int_{\mu_{l-1}}^{\mu_l} f(\lambda)g(\mu)(\lambda - A)^{-1}(\mu - A)^{-1}\, d\mu\, d\lambda.$$

[15] Op detail nog te bespreken.

Dankzij wat fraaie algebra, te weten

$$(\lambda - A)^{-1}(\mu - A)^{-1} = \frac{1}{\mu - \lambda}(\lambda - A)^{-1} + \frac{1}{\lambda - \mu}(\mu - A)^{-1},$$

kunnen de integralen gesplitst worden in

$$\int_{\lambda_{k-1}}^{\lambda_k} f(\lambda)(\lambda - A)^{-1} \int_{\mu_{l-1}}^{\mu_l} \frac{g(\mu)}{\mu - \lambda}\, d\mu\, d\lambda$$

en

$$\int_{\mu_{l-1}}^{\mu_l} g(\mu)(\mu - A)^{-1} \int_{\lambda_{k-1}}^{\lambda_k} \frac{f(\lambda)}{\lambda - \mu}\, d\lambda\, d\mu,$$

en in beide herhaalde integralen zien we een bij sommeren over de index in de binnenste integraal een gewone complexwaardige lijnintegraal verschijnen waar nul uit komt als de noemer niet nul is in het binnengebied, en een functiewaarde anders, kijk maar naar de Cauchy integraalformule. Sommeren over l in de eerste geeft derhalve 0, en sommeren over k in de tweede

$$\int_{\mu_{l-1}}^{\mu_l} g(\mu)(\mu - A)^{-1}\, 2\pi i\, f(\mu)\, d\mu = 2\pi i \int_{\mu_{l-1}}^{\mu_l} f(\mu)\, g(\mu)(\mu - A)^{-1}\, d\mu,$$

en nog een keer sommeren vervolgens $(2\pi i)^2 (fg)(A)$. We concluderen dat

$$(fg)(A) = \frac{1}{2\pi i} \oint_{\lambda_{0-n}} f(\lambda) g(\lambda)(\lambda - A)^{-1}\, d\lambda = f(A)g(A) = g(A)f(A), \tag{10.35}$$

en daar is nog veel mee te spelen.

Opgave 10.40. Ga na dat in het algemene geval (10.34), wanneer $f(A)$ en $g(A)$ de som zijn van een eindig aantal integralen over links- dan wel rechtsom[16] doorlopen polygonen P_j, er in de compositie alleen bijdragen zijn van de vorm zoals juist behandeld en dat ook in dat geval volgt dat $(fg)(A) = f(A)g(A)$.

De tweede gelijkheid in (10.35) is een gelijkheid in de niet-commutatieve Banachalgebra $L(X)$ van continue lineaire afbeeldingen van X naar zichzelf, en $f \to f(A)$ is een afbeelding die gedefinieerd is voor een klasse van functies gedefinieerd op een omgeving van het $\sigma(A)$. Die omgeving mag van f afhangen, dus met f en g moeten we ons beperken tot de doorsnede van de twee definitiegebieden. Wat we nog willen laten zien is dat een schrijfwijze met (10.33) en (10.34) altijd mogelijk is met alle polygonen zo dicht bij $\sigma(A)$ als we maar willen. Daarmee bewijzen we dan ook meteen de volgende stelling.

[16] Lees: linksom, maar met een min voor het integraalteken.

Stelling 10.41. *Laat voor $A \in L(X)$ en een complexwaardige f de operator $f(A)$ gedefinieerd zijn via (10.33) en (10.34). Dan geldt*

$$\sigma(f(A)) = f(\sigma(A)).$$

Om deze stelling te bewijzen maken we nu precies hoe we K kiezen. Kies daartoe een triangulatie van het complexe vlak opgespannen door $\rho > 0$ en $\rho \exp(\frac{\pi i}{6})$. De verzameling van al deze driehoekjes noemen we I. Voor elke $\Delta \in I$ maken we onderscheid tussen

$$\Delta \cap \sigma(A) = \emptyset, \quad \Delta \cap \sigma(A) \neq \emptyset = \delta\Delta \cap \sigma(A), \quad \delta\Delta \cap \sigma(A) \neq \emptyset,$$

waarmee $I = I_0 \cup I_1 \cup J$, met I_0, I_1, J de onderling disjuncte deelverzamelingen waarvoor respectievelijk de eerste, tweede dan wel derde karakterisatie geldt. Zowel I_1 als J hebben maar eindig veel elementen omdat $\sigma(A)$ begrensd is. Iedere $\Delta \in I_1$ kan als een K_j genomen worden in (10.34).

De driehoekjes in I_0 zijn niet relevant voor (10.34), maar iedere $\Delta \in J$ heeft 12 buren[17] waarvan er tenminste één ook in J ligt, zeg $\tilde{\Delta}$, gekarakteriseerd door

$$\delta\tilde{\Delta} \cap \delta\Delta \cap \sigma(A) \neq \emptyset,$$

en in dat geval noemen we Δ en $\tilde{\Delta}$ fijne buren in J. Twee zulke fijne buren die verder geen andere fijne buren hebben vormen samen een fijn duo verenigd in

$$\Delta \cup \tilde{\Delta},$$

en I_2 is per definitie de verzameling van zulke verder geisoleerde fijne buren, die verenigd steeds een ruit vormen, een ruit die als een K_j kan worden meegenomen in (10.34).

Een paar niet geisoleerde fijne buren kan nog 1 of meerdere fijne buren hebben, en als het maar 1 is, zeg $\hat{\Delta}$, dan kan het zijn dat die verder zelf geen fijne buren meer heeft. Dan vormen ze een fijn triootje waarbij verschillende standjes denkbaar zijn. Dit definieert de verzameling I_3, alle driehoeken Δ die onderdeel vormen van een fijn trio verenigd in

$$\Delta \cup \tilde{\Delta} \cup \hat{\Delta},$$

dat een parallelogram of een halve zeshoek is.

En zo gaat dat door met fijne quatrootjes, fijne quintootjes, etc totdat J op is, waarbij het aantal standjes flink maar niet oneindig toe kan nemen. Kortom, met I gepartioneerd als

$$I = I_0 \cup I_1 \cup I_2 \cup \cdots \cup I_p$$

[17] Waarvan er drie een zijde met Δ gemeen hebben en de rest alleen een hoekpunt.

is het nu nog de vraag wat de mogelijke onderlinge standjes zijn: als $\Delta_1 \in I_k$ met $k-1$ andere driehoeken in I_k een fijn k-stel vormt hoe kan de vereniging

$$\Delta_1 \cup \Delta_2 \cup \cdots \cup \Delta_k$$

er dan uitzien?

Antwoord: als een binnengebied van een polygon, of als het rechterlid van (10.32). Dat moet dus nog door iemand[18] bewezen worden, als dat niet al eens gebeurd is. Maar verder zijn we nu wel klaar met de beschrijving van $f(A)$. Dat kan altijd met eindig veel polygonen die willekeurig dicht bij $\sigma(A)$ liggen door ρ klein te kiezen. Hoe dichter bij $\sigma(A)$ hoe meer je er nodig hebt en hoe wilder de standjes kunnen worden.

We zijn nu klaar voor het bewijs van Stelling 10.41. Neem een $\mu \notin f(\sigma(A))$ en definieer g door

$$\lambda \xrightarrow{g} \frac{1}{\mu - f(\lambda)},$$

met f complex differentieerbaar op een omgeving van $\sigma(A)$. Kies een mogelijk kleinere omgeving waarop $f(\lambda) \neq \mu$. Uit de functional calculus volgt nu dat $g(A)$ gedefinieerd is en de algebra geeft

$$g(A)(\mu - f(A) = (\mu - f(A)g(A) = I,$$

waarmee $\mu \in \rho(f(A))$. Dus $\sigma(f(A)) \subset f(\sigma(A))$.

Kan de inclusie strict zijn? In dat geval is er een $\mu_0 = f(\lambda_0) \in \sigma(f(A))$ waarvoor $\mu_0 - f(A)$ inverteerbaar is terwijl $\lambda_0 - A$ het niet is. Door schuiven en schalen van f, en schuiven van A en λ kunnen we zonder beperking der algemeenheid wel aannemen dat $\lambda_0 = 0 = \mu_0$ en dat de machtreeks van f begint met λ^n voor zekere $n \in \mathbb{N}$ omdat $f(0) = 0$. In dat geval is

$$f(\lambda) = \lambda^n g(\lambda) \quad \text{met} \quad g(\lambda) = 1 + b_1\lambda + b_2\lambda^2 + \cdots$$

en dus is $g(A)$ inverteerbaar, net als $f(A)$. Maar de algebra geeft

$$f(A) = A^n g(A).$$

Voor $n = 1$ is de tegenspraak onmiddellijk. Voor $n > 1$ niet helemaal. Pas daarom het argument hierboven aan en concludeer eerst dat $\mu_0 = f(\lambda_0)$ zo gekozen kan worden dat $f'(\lambda_0) \neq 0$. Hiermee is het bewijs van de stelling wel klaar. Als g een andere functie is die complex differentieerbaar is op een omgeving van $\sigma(f(A)) = f(\sigma(A))$ dan volgt ook vrij direct uit de definities dat

$$g(f(A)) = (g \circ f)(A).$$

[18] Ik pas, maar dat is voor even.

Opgave 10.42. Bewijs dit.

Nog een expliciet voorbeeld. Als

$$\lambda = \lambda \sum_{j=1}^{\mathrm{N}} \chi_j(\lambda) = \sum_{j=1}^{\mathrm{N}} \lambda \chi_j(\lambda),$$

met

$$\chi_j(\lambda) = \delta_{ij} \quad \text{voor} \quad \lambda \in K_i,$$

dan

$$I = \sum_{j=1}^{\mathrm{N}} I\chi_j.$$

Definieer de "spectraalprojecties"

$$P_j = \chi_j(A).$$

Opgave 10.43. Laat zien dat $P_iP_j = \delta_{ij}P_j$, $AP_j = P_jA$, $\sigma(AP_j) = \sigma(A) \cap K_j$,

$$I = \sum_{j=1}^{\mathrm{N}} P_j \quad \text{en} \quad A = \sum_{j=1}^{\mathrm{N}} AP_j = \sum_{j=1}^{\mathrm{N}} P_jA.$$

Zo wordt

$$X = R(P_1) \oplus \cdots \oplus R(P_n),$$

en beeld A iedere $X_i = R(P_i)$) op zich zelf af, en volgt voor

$$A_j : X_i \xrightarrow{AP_j} X_i$$

dat $\sigma(A_j) = \sigma(A) \cap K_j$.

Zo, en dat alles met een beetje lijnintegreren.

11 Extremisme

Nu we de differential calculus voor afbeeldingen van Banachruimten naar Banachruimten min of meer in onze vingers hebben is de te beantwoorden vraag hoe het zit met extrema[1] van reëelwaardige functies op Banachruimten. De eerste probleemstelling in dit hoofdstuk is hoe we gegeven een reële Banachruimte X, een functie $F : X \to \mathbb{R}$, en een deelverzameling $A \subset X$, de extremen van F beperkt tot A vinden door vergelijkingen op te schrijven, en die op te lossen.

Als $A = X$ en f differentieerbaar is op heel X dan neemt f eventuele extrema alleen maar aan in *stationaire punten* van f, oplossingen van

$$F'(x) = 0,$$

dus stel nu dat $F'(x_0) = 0$ met $x_0 \in X$. De eerste vraag is dan of F in x_0 een maximum of een minimum aanneemt, en voor functies $F : \mathbb{R} \to \mathbb{R}$ doe je dit met wat vroeger het functieonderzoek heette, dat culmineerde in het zonder hulpmiddelen tekenen van de grafiek van F.

11.1 Tweede afgeleide, Hessiaan, Lemma van Morse

Welnu, dat functieonderzoek is afgeschaft. Er is bijna geen schoolboek meer dat vraagstellingen over een functie presenteert zonder je de grafiek kado te doen. Wat nog wel behandeld wordt is het criterium van de tweede afgeleide. Het teken van $F''(x_0)$ vertelt je of de grafiek er lokaal als $\smile$ of als $\frown$ uitziet[2].

Op zoek naar de happy minimizers is nu de eerste vraag hoe voor

$$F : X \to \mathbb{R}$$

de restterm in

$$F(x) = F(x_0) + F'(x_0)(x - x_0) + R(x; x_0)$$

er expliciet uitziet. In het geval dat $F : \mathbb{R} \to \mathbb{R}$ weten we dat

$$F(x) = F(x_0) + F'(x_0)(x - x_0) + \frac{1}{2}F''(x_0)(x - x_0)^2 + o((x - x_0)^2)$$

als $F'' : \mathbb{R} \to \mathbb{R}$ continu is in x_0, en zien we waarom het teken van $F''(x_0)$ in een punt waar $F'(x_0) = 0$ het verschil maakt tussen happy minima and sad maxima: $\smile$ or $\frown$.

[1] maxima EN minima.

[2] Happy or sad parabolas in case of $F(x) = \alpha(x - x_0)^2 + \beta$.

Kijk nog even naar Opgave 4.14 en Opgave 4.38 voor we een eerste Taylorbenadering maken van $F : X \to \mathbb{R}$ rond x_0, en vervolgens kijken wat we dan kunnen zien in stationaire punten, punten waar $F'(x_0) = 0$. Voor het gemak nemen we zo $x_0 = 0$ en brengen Opgave 4.38 in stelling via

$$[0,1] \xrightarrow{g} \mathbb{R} \quad \text{gedefinieerd door} \quad g(t) = F(tx)$$

als

$$g(t) = g(0) + g'(0) + \int_0^1 (1-t)g''(t)\,dt,$$

waarin $g(0) = F(0)$ en $g'(0) = F'(0)x$. Als $x = 0$ een stationair punt is van F dan gaat het dus om

$$R(x) = \int_0^1 (1-t)g''(t)\,dt \quad \text{met} \quad g'(t) = F'(tx)x, \tag{11.1}$$

waarbij de uitdrukking voor $g'(t)$ rechtstreeks uit de definities volgt maar natuurlijk ook consistent is met de kettingregel.

Opgave 11.1. Verifieer dit!

In (11.1) komt de afgeleide van $t \to F'(tx)x$ voor, dus we hebben nu voor het eerst de differentieerbaarheid van

$$x \to F'(x) \in X^*$$

nodig, die in $x = x_0 \in X$ per definitie geformuleerd wordt als

$$F'(x) = F(x_0) + F''(x_0)(x - x_0) + S(x; x_0),$$

met $F''(x_0) : X \to X^*$ begrensd en lineair en $|S(x;x_0)|_{X^*} = o(|x - x_0|)$ als $|x - x_0| \to 0$.

Met $x_0 = t_0 x$ volgt dus dat

$$\begin{aligned} F'(tx) &= F'(t_0x) + F''(t_0x)(tx - t_0x) + S(tx; t_0x) \\ &= F'(t_0x) + (t - t_0)F''(t_0x)x + S(tx; t_0x), \end{aligned}$$

en dus

$$g'(t) = g'(t_0) + (t - t_0)(F''(t_0x)x)x + S(tx; t_0x)x,$$

waaruit we aflezen dat $g'(t)$ uit (11.1) differentieerbaar is in t_0, met

$$g''(t_0) = (F''(t_0x)x)x. \tag{11.2}$$

Merk op dat $F''(x_0)h \in X^*$ voor alle $h \in X^*$ en $F''(x_0)hk \in \mathbb{R}$, dus

$$(h,k) \xrightarrow{F''(x_0)} F''(x_0)hk \in \mathbb{R} \tag{11.3}$$

is een kwadratische vorm die natuurlijk symmetrisch is, met een bewijs dat uit Stelling 6.6 volgt. Neem daartoe zonder beperking der algemeenheid $x_0 = 0$, maak de volgende opgave en lees meteen door!

Opgave 11.2. Definieer voor h en k in X en $F : X \to \mathbb{R}$, met $x \to F'(x)$ differentieerbaar in $x = 0$, de functie

$$(s,t) \xrightarrow{g} F(sh + tk)$$

en bewijs rechtstreeks uit de definities dat $g_{st}(0,0) = F''(0)k\,h$ en $g_{ts}(0,0) = F''(0)h\,k$.

Via Opgave 11.2 is (11.3) te schrijven als

$$\langle F''(0)h, k\rangle = \langle F''(0)k, h\rangle, \tag{11.4}$$

waarin de haken nu staan voor de onderlinge actie[3] tussen X^* en X, dus

$$\langle \phi, x\rangle = \phi(x) \quad \text{voor} \quad \phi \in X^* \quad \text{en} \quad x \in X.$$

In (11.4) zien we hoe $A = F''(x_0) \in L(X, X^*)$ voldoet aan

$$\langle Ah, k\rangle = \langle Ak, h\rangle,$$

hetgeen we als definiërende eigenschap voor

$$A \in S(X, X^*) \tag{11.5}$$

en $S(X, X^*) \subset L(X, X^*)$ zien.

Met ook $F''(tx) \in S(X, X^*)$ volgt nu dat in (11.1) de restterm gegeven wordt door

$$R(x) = \int_0^1 (1-t)F''(tx)x\,x\,dt = \int_0^1 (1-t)F''(tx)x\,dt\,x$$

$$= \int_0^1 (1-t)F''(tx)\,dt\,x\,x = \langle \int_0^1 (1-t)F''(tx)\,dt\,x, x\rangle$$

[3] The French notation, on ecrit aussi $X' = X^*$.

$$= \frac{1}{2}\langle F''(0)x, x\rangle + \langle \int_0^1 (1-t)(F''(tx) - F''(0))\, dt\, x, x\rangle$$

$$= \frac{1}{2}Q_0(x,x) + T(x), \tag{11.6}$$

met $T(x) = o(|x|^2)$ als $|x| \to 0$ omdat F'' continu is in $x = 0$.

Opgave 11.3. Ga na dat F'' continu in $x = 0$ betekent dat er voor elke $\varepsilon > 0$ een $\delta > 0$ is zo dat de implicatie

$$|x| \leq \delta \implies |(F''(x) - F''(0))y| \leq \varepsilon|y|$$

geldt voor alle $y \in X$.

Opgave 11.4. Werk het bovenstaande uit voor $X = \mathbb{R}^2$ en laat zien dat

$$Q_0(x,x) = F_{x_1x_1}(0,0)x_1^2 + 2F_{x_1x_2}(0,0)x_1x_2 + F_{x_2x_2}(0,0)x_2^2$$

voor $x = (x_1, x_2) \in \mathbb{R}^2$.

Opgave 11.5. Laat zien dat er $r > 0$ is zo dat voor elke $x \in X$ met $|x| \leq r$ geldt dat er een $\theta = \theta(x) \in [0,1]$ is waarvoor

$$T(x) = \int_0^1 (1-t)(F''(tx) - F''(0))x\, x\, dt = \frac{1}{2}F''(\theta(x))x\, x$$

(dit geldt onder de zwakkere aanname dat F twee keer differentieerbaar is in de buurt van $x = 0$).

In (11.6) is hebben we evenzo dat

$$|Q_0(x,x)| \leq |F''(0)|\, |x|^2.$$

Als[4] er een $\mu > 0$ is met

$$Q_0(x,x) \geq \mu|x|^2$$

voor alle $x \in X$ kunnen we dus concluderen dat $F(x)$ in $x = 0$ lokaal minimaal is.

[4] Maar kan dat?

We merken op dat in plaats van (11.6) we ook kunnen schrijven dat

$$R(x) = (\Phi(x)x)x = \langle \Phi(x)\, x, x\rangle$$

met

$$\Phi(x) = \int_0^1 (1-t)F''(tx)\, dt \in S(X, X^*) \subset L(X, X^*), \tag{11.7}$$

en dat vervolgens de vraag gesteld kan worden of via een coördinatentransformatie

$$x \to \langle \Phi(x)x, x\rangle \quad \text{en} \quad y \to \langle \Phi(0)y, y\rangle$$

in wezen dezelfde functies zijn.

Bob Rink liet me zijn oude dictaat van Hans Duistermaat en Joop Kolk zien, en daarin wordt daartoe zo'n transformatie gezocht in de vorm

$$y = A(x)x,$$

hetgeen leidt tot de vraag of $A = A(x) \in L(X, X)$ zo gevonden kan worden dat

$$\langle \Phi(x)x, x\rangle = (\Phi(x)x)\, x = (\Phi(0)y)\, y = \langle \Phi(0)y, y\rangle \tag{11.8}$$

voor x in de buurt van $x = 0$. Dus voor $y = A(x)x$ willen we

$$x \xrightarrow{A(x)} A(x)x = y \xrightarrow{\Phi(0)y} (\Phi(0)y)y = (\Phi(0)A(x)x) \circ A(x))\, x = (\Phi(x)x)x,$$

waarin x aan beide kanten twee keer lineair op zichzelf voorkomt.

Hieraan is zeker voldaan als

$$(\Phi_0 A\, x) \circ A = \Phi(x)x \tag{11.9}$$

in X^*, waarin we $A = A(x)$ zoeken bij $x \neq 0$ en $\Phi(x)$ met $\Phi(0) = \Phi_0$. Aan (11.9) is vervolgens zeker voldaan als de afbeelding

$$h \to \Phi(x)h \quad \text{in} \quad L(X, X^*) \quad \text{gelijk is aan} \quad h \to \Phi_0 Ah \circ A, \tag{11.10}$$

een kwadratische $L(X, X^*)$-waardige vergelijking voor $A \in L(X, X)$, abstract te schrijven als

$$\kappa_0(A, A) = \Phi(x). \tag{11.11}$$

In (11.11) is

$$X \times X \xrightarrow{\kappa_0} L(X, X^*) \quad \text{in} \quad L(X \times X, L(X, X^*))$$

gedefinieerd door

$$h \to \kappa_0(A, B)\, h = \Phi_0 Ah \circ B$$

als afbeelding in $L(X, X^*)$ voor iedere $A, B \in L(X)$. Voor $x = 0$ is $A = I$ een oplossing van $\kappa_0(A, A) = \Phi(x)$ omdat $\Phi(0) = \Phi_0$.

De impliciete functiestelling is toepasbaar als de afgeleide van de kwadratische functie

$$A \to \kappa_0(A, A)$$

in $A = I$ inverteerbaar is, en $x \to \Phi(x)$ continu is in $x = 0$. Dat laatste is bij aanname dat F'' continu is in 0 het geval.

Met betrekking tot de afgeleide naar A in I schrijven we $A = I + H$ en merken voordat we gaan expanderen in de kleine H op dat (11.11) nu herschrijft als

$$\chi_0(H)h = \Phi_0 Hh + \Phi_0 h \circ H + \Phi_0 Hh \circ H = (\Phi(x) - \Phi_0)h \tag{11.12}$$

voor alle $h \in X$. Het linkerlid definieert een X^*-waardige functie

$$H \xrightarrow{\chi_0} \chi_0(H)$$

van H in X, met een $\chi_0(H)$ die twee lineaire termen en een kwadratisch term heeft, waarin Φ_0 steeds als coefficient gezien kan worden.

Omdat de afgeleide in $H = 0$ evident gegeven wordt door

$$h \xrightarrow{\chi_0'(I)H} \Phi_0 Hh + \Phi_0 h \circ H,$$

lijkt de inverteerbaarheidsconditie te zijn dat

$$\Phi_0 Hh + \Phi_0 h \circ H = Ch \tag{11.13}$$

voor alle $h \in X$, oplosbaar moet zijn voor elke $C \in L(X, X^*)$. Merk echter op dat het rechterlid van de op te lossen vergelijking

$$\chi_0(H) = \Phi(x) - \Phi_0 \tag{11.14}$$

in $S(X, X^*)$ zit, zie (11.5). Nadere beschouwing van het linkerlid van (11.12) middels

$$\begin{aligned}\langle \chi_0(H)h, k\rangle &= \langle \Phi_0 Hh, k\rangle + \langle \Phi_0 h \circ H, k\rangle + \langle \Phi_0 Hh \circ H, k\rangle \\ &= \langle \Phi_0 Hh, k\rangle + \langle \Phi_0 h, Hk\rangle + \langle \Phi_0 Hh, Hk\rangle \\ &= \langle \Phi_0 Hh, k\rangle + \langle \Phi_0 Hk, h\rangle + \langle \Phi_0 Hh, Hk\rangle \end{aligned} \tag{11.15}$$

voor h en k in X omdat $\Phi_0 \in S(X, X^*)$, leert ons dat $\chi_0'(0)H \in L(X, X^*)$ gekarakteriseerd wordt door

$$\langle \chi_0'(0)Hh, k\rangle = \langle \Phi_0 Hh, k\rangle + \langle \Phi_0 Hk, h\rangle, \tag{11.16}$$

en daarom volgt dat $\chi_0'(0)H \in S(X, X^*)$.

De kwadratische restterm $R_0(H) \in L(X, X^*)$ in

$$\chi_0(H) = \chi_0'(0)H + R_0(H)$$

is af te lezen in(11.15) als gekarakteriseerd door

$$\langle R_0(H)h, k\rangle = \langle \Phi_0 Hh, Hk\rangle, \tag{11.17}$$

en derhalve is ook $R_0(H) \in S(X, X^*)$. We kunnen en moeten (11.14) dus lezen als de vergelijking

$$\chi_0(H) = \chi_0'(0)H + R_0(H) = \Phi(x) - \Phi_0 \quad \text{in} \quad S(X, X^*). \tag{11.18}$$

Daarmee wordt ook (11.13) beperkt tot $C \in S(X, X^*)$ en herschrijft als

$$\langle \Phi_0 Hh, k\rangle + \langle \Phi_0 Hk, h\rangle = \langle Ch, k\rangle \tag{11.19}$$

voor alle $h, k \in X$.

In het geval dat $X = \mathbb{R}^2$ zijn Φ_0 en C symmetrische 2×2 matrices en staan hier vier vergelijkingen voor een 2×2 matrix H in termen van maar drie coefficienten in C. Zonder extra aannamen op H is (11.19) dan een onderbepaald stelsel. De voor hand liggende extra conditie op H is nu dat $\Phi_0 H \in S(X, X^*)$. In dat geval is (11.19) equivalent met

$$2\Phi_0 H = C, \tag{11.20}$$

een vergelijking die weer equivalent is met

$$H = \frac{1}{2}\Psi_0 C, \tag{11.21}$$

als

$$\Psi_0 = \Phi_0^{(-1)} \tag{11.22}$$

bestaat, en in dat geval is $\Phi_0 H = \Phi_0(\frac{1}{2}\Phi_0^{(-1)})C = \frac{1}{2}C \in S(X, X^*)$.

Aan de conditie voor het toepassen van de impliciete functiestelling op (11.18) in $S(X, X^*)$ rond $H = 0$ en $x = 0$ is dus voldaan als (11.18) als een vergelijking voor

$$\{H \in L(X, X^*) : \Phi_0 H \in S(X, X^*)\} \xrightarrow{\chi_0} S(X, X^*)$$

gezien wordt. Met $B = \Phi_0 H$ als nieuw onafhankelijke variabele verandert dit in

$$2Bh + Bh \circ (\Psi_0 B) = (\Phi(x) - \Phi_0)h$$

voor alle $h \in X$, ofwel

$$2B + T_0(B) = C(x) = \Phi(x) - \Phi_0 \tag{11.23}$$

in $S(X, X^*)$, met $T_0 : S(X, X^*) \to S(X, X^*)$ gedefinieerd door

$$T_0(B)h = Bh \circ (\Psi_0 B) \tag{11.24}$$

voor alle $h \in X$, en

$$X \xrightarrow{C} S(X, X^*)$$

continu en $C(0) = 0$.

De impliciete functiestelling kan nu gebruikt worden om, zonder meer[5] te eisen dan continuiteit van $\Phi(x)$, een unieke oplossing $B(x)$ te vinden voor x in de buurt van $x = 0$, die daar continu van x afhangt. Hiermee is de volgende stelling bewezen.

Stelling 11.6. *Laat X een Banachruimte zijn en $F : X \to \mathbb{R}$ twee keer continu differentieerbaar in de buurt van $x = 0$. Neem aan dat $F'(0) = 0$ en dat $F''(0) \in L(X, X^*)$ inverteerbaar is, met een inverse in $L(X, X^*)$. Dan is er een coördinatentransformatie van de vorm*

$$y = A(x)x = (I + \Phi_0^{(-1)} B(x))x,$$

waarin

$$\Phi_0 = \frac{1}{2} F''(0)$$

en

$$x \xrightarrow{B} S(X, X^*)$$

continu is met $B(0) = 0$, zodanig dat

$$F(x) = \langle \Phi_0 A(x)x, A(x)x \rangle,$$

in de buurt van $x = 0$.

De aanname dat $F''(0)$ inverteerbaar is, is nogal zwaar en sluit een aantal standaardruimten meteen uit. Is dit eigenlijk een Hilbertruimtestelling, vraag je[6] dan achteraf af[7]. Ook is op te merken dat de impliciete functiestelling hier achteraf helemaal niet[8] nodig was. Merk daartoe op dat via $\phi = \Phi_0 h$ en $\psi = \Phi_0 k$ geldt dat

$$\langle \phi, \Psi_0 \psi \rangle = \langle \psi, \Psi_0 \phi \rangle \quad \text{voor alle} \quad \phi, \psi \in X^*,$$

[5] Zie (4.21) en formulering en bewijs daarboven.

[6] To check: X en X^* lineair homeomorf $\implies$ X is a Hilbert space in disguise?

[7] Via André Ran kwam van Jan Fourie antwoord: nee, $X \oplus X^*$, met X reflexief.

[8] Maar de stelling zette ons wel op het juiste spoor.

i.e. $\Psi_0 \in S(X^*, X)$, en enig spelen met deze definitie en die van $B, \Phi_0 \in S(X, X^*)$ geeft dat (11.23) equivalent is met

$$2\Psi_0 B + \Psi_0 B \Psi_0 B = \Psi_0 C(x). \tag{11.25}$$

Voor H zelf betekent dit dat

$$I + 2H + H^2 = I + P(x) \quad \text{in} \quad L(X) = L(X, X), \tag{11.26}$$

waarbij[9] zowel $H(x)$ als $P(x)$ (moeten) voldoen aan $\Phi_0 P(x), \Phi_0 H(x) \in S(X, X^*)$. Met de machtreekstechnieken die we gezien hebben[10] voor matrices volgt dat

$$H = \frac{1}{2}\,P - \frac{1}{2!}\,\frac{1}{2}\,\frac{1}{2}\,P^2 + \frac{1}{3!}\,\frac{1}{2}\,\frac{1}{2}\,\frac{3}{2}\,P^3 - \frac{1}{4!}\,\frac{1}{2}\,\frac{1}{2}\,\frac{3}{2}\,\frac{5}{2}\,P^4 + \cdots \tag{11.27}$$

als $|P| < 1$, en dus $y = A(x)x$ met[11]

$$A(x) = I + E(x) - \frac{1}{2!}\,E(x)^2 + \frac{1 \cdot 3}{3!}\,E(x)^3 - \frac{1 \cdot 3 \cdot 5}{4!}\,E(x)^4 + \cdots$$

en

$$E(x) = \int_0^1 (1-t)(F''(0)^{(-1)} F''(tx) - I)\,dt,$$

wel aardig. Modulo rekenfouten wellicht maar ik denk het niet. Aan de formules zie je niet meer dat het niet alleen over $X = \mathbb{R}$ gaat.

11.2 Nog een keer de methode van Lagrange

In Sectie 6.3 hebben we essentieel gebruikt gemaakt van de gradiënten van Φ, F_1, F_2, F_3. Dat was niet echt nodig, en dat laten we nu zien in de abstracte setting met $x \in X$, $y \in Y$, $F : X \times Y \to Y$ en $\Phi : X \times Y \to \mathbb{R}$.

Onder de aanname dat

$$(x, y) \xrightarrow{F_x} F_x(x, y) \quad \text{en} \quad (x, y) \xrightarrow{F_y} F_y(x, y)$$

continu zijn in de buurt van $(x, y) = (0, 0)$ met F_y inverteerbaar, volgt weer het bestaan van een continu differentieerbare impliciete functie $y = f(x)$ en een lokale beschrijving van S zoals voor (6.19). Voor differentieerbare Φ, met dus $(x, y) \to \Phi_x(x, y) \in X^*$ en $(x, y) \to \Phi_y(x, y) \in Y^*$, is (6.18) weer over te schrijven als criterium voor het stationair zijn van Φ beperkt tot S.

[9] De omweg via H was ook niet nodig. De vergelijking was gewoon $A^2 = I + \Psi_0 C(x)$.
[10] Zie Opgave 5.42.
[11] Er vallen een heleboel 2-en weg bij de terugvertaling naar F''.

Als nu $\Phi_y \in Y^*$ te schrijven is als

$$\Phi_y = \Lambda \circ F_y,$$

dan volgt

$$\Phi_x = \Phi_y(F_y)^{(-1)}F_x = \Lambda \circ F_y(F_y)^{(-1)}F_x = \Lambda \circ F_x,$$

en samengenomen dat

$$\Phi' = \Lambda \circ F'$$

als criterium voor het stationair zijn.

Wat we hierboven nodig hebben is een stelling die zegt dat bij iedere begrensde lineaire $A : Y \to Y$ en $\psi \in Y^*$ een (unieke) $\Lambda \in Y^*$ hoort met $\psi = \Lambda \circ A$. Dat is voor een college functionaalanalyse.

12 Appendix: al of niet metrische topologie

Een aardig dictaatje is hier te vinden, uit de tijd dat Leiden de R nog in de naam had:

`http://www.few.vu.nl/~jhulshof/NOTES/anal.pdf`

Hieronder neem ik het over met wat aanvullingen en correcties.....

12.1 Metrische ruimten

Onze genormeerde ruimten X, waaronder $\mathbb{R}, \mathbb{R}^2$ en ook $C([a,b])$ met de maximumnorm, maar helaas niet $R([a,b])$ met de 1-norm, zijn voorbeelden van metrische ruimten met het afstandsbegrip gedefinieerd door de metriek

$$(x,y) \xrightarrow{d} d(x,y) = |x-y|, \tag{12.1}$$

een afbeelding[1] d van $X \times X$ naar $[0,\infty)$ met de eigenschappen dat voor alle $x,y,z \in X$ geldt dat

$$(i) \quad d(x,y) = 0 \iff x = y; \quad (ii) \quad d(x,y) = d(y,x);$$

$$(iii) \quad d(x,y) \leq d(x,z) + d(z,x). \tag{12.2}$$

Iedere niet-lege deelverzameling A van X is zo een metrische ruimte, waarbij we de algebraische vectorruimte operaties nu vergeten.

Definitie 12.1. *Een metrische ruimte is een niet-lege verzameling X met een afbeelding $d : X \times X \to [0,\infty)$ waarvoor (i),(ii) en (iii) uit (12.2) hierboven gelden voor alle $x,y,z \in X$.*

Opgave 12.2. Zie Opgave 10.3. De ε, N-definitie van $d(x_n, x_m) \to 0$ als $m,n \to \infty$ definieert wat een Cauchyrij in X is. Geef die definitie. Geef ook de definitie van het convergent zijn van de rij x_n in X.

We gebruiken hieronder de notatie $x_n \to x$ voor $x_1, x_2, x_3, \ldots, x \in X$ zonder er steeds $n \to \infty$ bij te zetten en spreken over ook een rij x_n zonder te vermelden dat $n \in \mathbb{N}$ (of een andere deelverzameling van $\mathbb{Z}$ van de vorm $m + \mathbb{N}$ met $m \in \mathbb{Z}$, bijvoorbeeld $\mathbb{N}_0$).

[1] De d van distance, a van afstand doen we maar niet.

Opgave 12.3. Een flauwe opgave om aan de de notaties, definities en axioma's te wennen: laat zien dat als $x_n \to x$ en $x_n \to y$ (alles in X) voor de limieten x en y geldt dat $x = y$. De limiet van een convergente rij is dus uniek.

Met convergente rijen kunnen we voor metrische ruimten X en Y zeggen wat het voor een afbeelding

$$F : X \to Y$$

betekent om continu te zijn in $a \in X$.

Definitie 12.4. *Een afbeelding F van een metrische ruimte X naar een (niet per se andere) metrische ruimte Y heet continu in $a \in X$ als de implicatie*

$$x_n \to a \implies F(x_n) \to F(a)$$

geldt voor elke rij x_n in X. Als dit het geval is voor elke $a \in X$ dan zeggen we dat $F : X \to Y$ continu is.

Opgave 12.5. Als X, Y, Z metrische ruimten en

$$X \xrightarrow{F} Y \quad \text{en} \quad Y \xrightarrow{G} Z$$

afbeeldingen dan is de afbeelding

$$X \xrightarrow{G \circ F} Z \quad \text{gedefinieerd door} \quad X \xrightarrow{F} Y \xrightarrow{G} Z$$

continu in $a \in X$ als F continu is in a en G continu is in $b = F(a)$. Hint: triviaal, leg uit.

Definitie 12.6. *Een metrische ruimte heet rijkompakt als elke rij in X een convergente deelrij heeft, en volledig als elke Cauchyrij in X convergent is (in beide gevallen met limiet in X dus).*

Opgave 12.7. Bewijs dat rijkompakte metrische ruimten volledig zijn.

Opgave 12.8. Als X en Y metrische ruimten zijn, met X rijkompakt, dan is iedere continue $F : X \to Y$ uniform continu, i.e.

$$\forall \varepsilon > 0 \, \exists \delta > 0 \, \forall x, a \in X : \, d(x, a) \leq \delta \implies d(F(x), F(a)) \leq \varepsilon.$$

Bewijs dit door een eerder bewijs (van Stelling 2.32) over te schrijven.

Opgave 12.9. Als X een rijkompakte metrische ruimte is, dan heeft iedere continue $F : X \to \mathbb{R}$ een globaal maximum en een globaal minimum op X. Bewijs ook dit door een eerder bewijs over te schrijven.

12.2 Omgevingen, open en gesloten verzamelingen

Continuiteit kunnen we ook met open verzamelingen beschrijven. In het standaardjargon heet een deelverzameling $G \subset X$ van een metrische ruimte X gesloten als voor iedere rij x_n in G met $x_n \to x \in X$ de limiet x in G zit (je kan G niet uit door limieten te nemen). Een verzameling $O \subset X$ heet open[2] als zijn complement gesloten is.

Opgave 12.10. Bewijs dat in een Banachruimte iedere rijkompakte deelverzameling begrensd en gesloten is en dat in $\mathbb{R}^n$ ook de omgekeerde uitspraak geldt.

Uit Opgave 12.9 en Opgave 12.10 volgt dat de stelling over maxima en minima van continue functies op gesloten begrensde deelverzamelingen van $\mathbb{R}^{\mathrm{N}}$.

Stelling 12.11. *Laat $K \subset \mathbb{R}^{\mathrm{N}}$ begrensd en gesloten zijn en $F : K \to \mathbb{R}$ continu zijn. Dan zijn er $a, b \in K$ met $F(a) \leq F(x) \leq F(b)$ voor alle $x \in K$. De punten a en b heten de minimizer en de maximizer voor F, en de waarden $F(a)$ en $F(b)$ het minimum en het maximum van F.*

Opgave 12.12. De collectie $\mathcal{G}$ van alle gesloten deelverzamelingen van een metrische ruimte X heeft drie belangrijke eigenschappen:

$$(i) \quad \emptyset \in \mathcal{G},\, X \in \mathcal{G}; \quad (ii) \quad G_1, G_2 \in \mathcal{G} \implies G_1 \cup G_2 \in \mathcal{G};$$

en (voor elke indexverzameling I)

$$(iii) \quad G_i \in \mathcal{G}\ \forall i \in I \implies \cap_{i \in I} G_i \in \mathcal{G}.$$

Bewijs dit via de definitie dat $G \in \mathcal{G}$ als voor iedere rij x_n in G met $x_n \to x \in X$ voor de limiet geldt $x \in G$.

[2] Minder gelukkige naamgeving, sorry, is niet anders.

Opgave 12.13. De collectie $\mathcal{O}$ van alle open deelverzamelingen van een metrische ruimte X heeft de volgende eigenschappen:

$$\emptyset \in \mathcal{O},\, X \in \mathcal{O}; \quad O_1, O_2 \in \mathcal{O} \implies O_1 \cap O_2 \in \mathcal{O};$$

en (voor elke indexverzameling I)

$$O_i \in \mathcal{O}\ \forall i \in I \implies \cup_{i\in I} O_i \in \mathcal{O}.$$

Bewijs dit via de definitie dat $O \in \mathcal{O}$ als

$$O^c = \{x \in X :\, x \notin O\} \in \mathcal{G}.$$

Opgave 12.14. Laat zien dat in een metrische ruimte X een deelverzameling $O \subset X$ open is dan en slechts dan als voor elke $a \in O$ er een $r > 0$ is zo dat

$$\bar{B}_r(a) = \{x \in A :\, d(x,a) \leq r\} \subset O.$$

Bewijs ook dat $\bar{B}_r(a)$ gesloten is.

Om te weten welke verzamelingen open zijn moet je dus weten wat de gesloten bollen $\bar{B}_r(a)$ zijn maar niet eens dat. Heb je bijvoorbeeld twee normen en noemen we de bijbehorende bollen $\bar{B}_r(a)$ en $\bar{K}_s(a)$ dan krijgen we precies dezelfde open verzamelingen als elke $\bar{B}_r(a)$ met $r > 0$ altijd een $\bar{K}_s(a)$ bevat met $s > 0$ en omgekeerd. Is X een vectorruimte over $\mathbb{R}$ met twee normen dan noemen we die normen equivalent als ze dezelfde collectie $\mathcal{O}$ definieren. Via Opgave 12.12 leidt dat tot deze karakterisatie van equivalente normen op X.

Opgave 12.15. Als twee normen

$$x \to |x|_1 \quad \text{en} \quad x \to |x|_2$$

dezelfde collectie $\mathcal{O}$ van open verzamelingen definieren dan zijn er constanten A_1 en A_2 zo dat voor alle $x \in X$ geldt

$$|x|_1 \leq A_2 |x|_2 \quad \text{en} \quad |x|_2 \leq A_1 |x|_1.$$

Bewijs dit. Terzijde, omgekeerd geldt ook en is makkelijker.

Opgave 12.16. Laat zien dat in een metrische ruimte X een deelverzameling $O \subset X$ open is dan en slechts dan als voor elke $a \in O$ er een $r > 0$ is zo dat

$$B_r(a) = \{x \in A : d(x,a) < r\} \subset O.$$

Bewijs ook dat $B_r(a)$ open is.

Stelling 12.17. *Laat X en Y metrische ruimten zijn en $F : X \to Y$. Dan is F continu dan en slechts dan als alle inverse beelden van open verzamelingen in Y open zijn in X.*

Opgave 12.18. Wel een kluifje: bewijs Stelling 12.17. Triviaal daarna is dat als X, Y, Z metrische ruimten zijn en

$$X \xrightarrow{F} Y \quad \text{en} \quad Y \xrightarrow{G} Z$$

continue afbeeldingen, dat de afbeelding

$$X \xrightarrow{G \circ F} Z \quad \text{gedefinieerd door} \quad X \xrightarrow{F} Y \xrightarrow{G} Z$$

continu is. Waarom? Zie nog even Opgave 12.5.

In $\mathbb{R}^2$ hebben we behalve the standaardnorm

$$|x| = \sqrt{x_1^2 + x_2^2} = \sqrt{x \cdot x} \quad \text{voor} \quad x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix},$$

afkomstig van het standaardinprodukt

$$x \cdot y = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = x_1 y_1 + x_2 y_2,$$

de normen

$$|x|_p = \sqrt[p]{|x_1|^p + |x_2|^p} \quad \text{voor} \quad p \geq 1 \quad \text{en} \quad |x|_\infty = \max(|x_1|, |x_2|).$$

Als deze normen zijn equivalent.

Opgave 12.19. Bewijs dat al deze p-normen equivalent zijn en teken in het x_1, x_2-vlak de gesloten eenheidsbollen $\bar{B}^p = \{x \in \mathbb{R}^2 : |x|_p \leq 1\}$ voor $p = 1, 2$ en $p = \infty$, en voor nog twee p's naar keuze. Blader nog even terug naar Opgave 12.15 en de karakterisatie daaronder en boven van open verzamelingen met behulp bollen, gesloten of open, zoals $B_\varepsilon^p(\xi) = \{x \in \mathbb{R}^2 : |x - x_0|_p < \varepsilon\}$ met $\xi \in \mathbb{R}^2$ en $\varepsilon > 0$.

Opgave 12.20. De bollen B^1 en B^∞ zijn ook te beschrijven als doorsnijdingen van open halfvlakken van de form $K = \{x \in \mathbb{R}^2 : f(x) < b\}$ met $f : \mathbb{R}^2 \to \mathbb{R}$ lineair gegeven door $f(x) = a_1x_1 + a_2x_2$ en $a_1, a_2, b \in \mathbb{R}$. Laat dat zien.

Opgave 12.21. Een alternatieve manier om te zeggen dat een $O \in \mathbb{R}^2$ open is te zeggen dat er voor elke $\xi \in O$ drie[3] open halfvlakken K_1, K_2, K_3 zijn zoals in Opgave 12.20, waarvoor geldt

$$\xi \in K_1 \cap K_2 \cap K_3 \subset O.$$

Waarom definieert dit dezelfde open verzamelingen? Geef ook zo'n definitie van open in $\mathbb{R}^3$.

Opgave 12.22. Een verzameling W in een genormeerde ruimte X heet zwak open als er voor elke $\xi \in W$ geldt dat er er eindig veel open halfvlakken zijn zo dat geldt

$$\xi \in K_1 \cap \cdots \cap K_n \subset W.$$

Bewijs dat voor deze zwak open verzamelingen W dezelfde eigenschappen gelden als in Opgave 12.13. Met eindige doorsnijdingen van open halfvlakken is dus een topologie te maken: een collectie van "open" verzamelingen die voldoet aan de "axioma's" in Opgave 12.13. In het geval dat $X = \mathbb{R}^n$ zijn alle normen op X en deze topologie equivalent.

`https://www.youtube.com/watch?v=fmTcSGukO4o`

Opgave 12.23. Bewijs dat iedere norm $x \to |x|$ op $\mathbb{R}^2$ equivalent is met de 2-norm. Hint: laat eerst zien dat $x \to |x|$ op $S = \{x \in \mathbb{R}^2 : x_1^2 + x_2^2 = 1\}$ een positief minimum en maximum heeft.

Opgave 12.24. Laat X_1 en X_2 genormeerde ruimten zijn. Bewijs dat

$$X_1 \times X_2 = \{x = (x_1, x_2) : x_1 \in X_1, x_2 \in X_2\}$$

met de voor de hand liggende bewerkingen weer een genormeerde ruimte is met (equivalente) normen (voor $p \geq 1$)

$$x \to \sqrt[p]{|x_1|^p + |x_2|^p} \quad \text{en} \quad x \to \max(|x_1|, |x_2|).$$

[3] $3 = 2 + 1$.

Opgave 12.25. Laat X_1 en X_2 genormeerde ruimten zijn en $X = X_1 \times X_2$. Bewijs dat iedere $f \in X^*$ van de vorm

$$x = (x_1, x_2) \xrightarrow{f} f_1(x_1) + f_2(x_2)$$

is met $f_1 \in X_1^*, f_2 \in X_2^*$. Met andere woorden $X^* = X_1^* \times X_2^*$.

Opgave 12.26. Laat X_1 en X_2 genormeerde ruimten zijn en $f \in X^* = X_1^* \times X_2^*$. Bepaal de norm van f in X^* als voor de norm op $X = X_1 \times X_2$ de norm $x \to |x_1| + |x_2|$ genomen wordt. Zelfde vraag voor $x \to \max(|x_1|, |x_2|)$.

12.3 Metrische ruimten; continue afbeeldingen

Aanvullend materiaal voor het college Analyse 1

J. Hulshof (destijds RUL)

Cursief wat opmerkingen van 24 jaar later ingevoegd tijdens het geven van het college Analyse 3 in het tweede jaar, met verwijzingen naar het boek Principles of Topology van Croom.

1. Inleiding. In deze syllabus behandelen we een aantal fundamentele onderwerpen uit de analyse. Uitgangspunt hierbij is de volgende algemene probleemstelling:

Laat X een puntverzameling zijn en A een niet-lege deelverzameling van X, eventueel X zelf. Zij $f : A \to \mathbb{R}$ een reëelwaardige functie. Hoe en onder wat voor veronderstellingen kunnen we dan concluderen dat de functie f op A een globaal maximum aanneemt, m.a.w. bestaat er een punt $x_0 \in A$, zo dat

$$f(x) \leq f(x_0) \quad \forall x \in A?$$

Om deze vraag te beantwoorden beschouwen we het supremum van f op A,

$$M = \sup\{f(x) : \ x \in A\} \ \in \ \mathbb{R} \cup \{+\infty\}.$$

Wat we van M nu willen weten is ten eerste of M eindig is, en zo ja, of de waarde M ook door de functie f wordt aangenomen. Omdat M het supremum is van alle door f aangenomen functiewaarden, kunnen we M benaderen met deze functiewaarden. We onderscheiden twee gevallen.

(i) $M < +\infty$. Dan bestaat er voor elk natuurlijk getal n een $x_n \in A$, zodat

$$f(x_n) > M - \frac{1}{n}.$$

(ii) $M = +\infty$. Dan bestaat er voor elk natuurlijk getal n een $x_n \in A$, zodat

$$f(x_n) > n.$$

In beide gevallen geeft dit ons een rij punten, genoteerd als $(x_n)_{n=1}^{\infty}$, in A. Als we nu kunnen concluderen dat, eventueel door een aantal van deze punten uit de rij weg te laten, deze punten voor grote waarden van $n \in \mathbb{N}$ steeds dichter komen te liggen bij een "limietpunt"dat zelf ook in A ligt, en dat de waarde van f in dat limietpunt gelijk is aan

$$\lim_{n\to\infty} f(x_n),$$

dan hebben we in één klap de beide bovenstaande vragen met ja beantwoord. In deze syllabus zullen we de voor bovenstaande probleemstelling relevante begrippen behandelen, met hier en daar een zijstapje.

In de vorige alinea staat de zinsnede "dichterbij". Aangezien we van X alleen maar hebben aangenomen dat X een puntverzameling is, heeft dit zonder verdere veronderstellingen geen betekenis. Een natuurlijke manier om dit de verhelpen is de invoering van een zogenaamde afstandsfunctie of metriek op X. Dit leidt dan tot de definitie van een metrische ruimte (Sectie 2). De deelverzamelingen waarvoor altijd een limietpunt van een rij bestaat blijken de zogenaamde rijkompakte verzamelingen te zijn (Sectie 3). Voor functies op (deelverzamelingen van) metrische ruimten kan het begrip "continu"worden gedefinieerd (Sectie 5), waarmee de bovenstaande limietovergang kan worden gerechtvaardigd. Als voorbeeld behandelen we de gevallen dat $X = \mathbb{R}$ en $X = \mathbb{R}^N$ (Sectie 6). In de appendix komen nog enige iets meer geavanceerde onderwerpen met betrekking tot kompaktheid aan de orde.

De schrijver van deze syllabus is van mening dat iedere student in de wiskunde of theoretische natuurkunde, onafhankelijk van wat hij/zij in de doctoraalfase als afstudeerrichting kiest, zich de basisstof in Sectie 1 tot en met 6 van deze syllabus moet eigen maken. In de meeste theoretische analyse boeken is deze stof, althans voor het geval $X = \mathbb{R}^N$, terug te vinden in de inleidende hoofdstukken. Zie bijvoorbeeld het boek "Mathematical Analysis, a modern approach to advanced calculus"van T.M. Apostol (Addison-Wesley 1957), waarin bijna alle analyse die in de eerste twee jaar van de studie aan de orde komt, is terug te vinden, of "Principles of Mathematical Analysis"van

W. Rudin (McGraw Hill 1964). Voor meer algemene metrische (en topologische) ruimten zijn er o.a. de boeken "Topology and Normed Spaces"van G.J.O. Jameson (Wiley 1974) en "Introduction to Topology and Modern Analysis"van G.F. Simmons (McGraw Hill 1963).

Als voorkennis wordt verondersteld dat de lezer bekend is met de elementaire verzamelingsleer, begrippen als aftelbaar oneindig en overaftelbaar oneindig, en met de axioma's voor de natuurlijke getallen $\mathbb{N}$ en de reële getallen $\mathbb{R}$. Hoofdstuk 1 uit het boek "Calculus 1 2nd edition"van T.M. Apostol (Wiley 1967) is ruim voldoende.

2. Metrische ruimten. Laat X weer een puntverzameling zijn.

Definitie 2.1. Een functie $d: X \times X \to \mathbb{R}$ heet een *metriek* op X als
(i) $\forall x, y \in X$:

$$d(x,y) \geq 0 \text{ en } d(x,y) = 0 \iff x = y.$$

(ii) $\forall x, y \in X$:

$$d(x,y) = d(y,x).$$

(iii) $\forall x, y, z \in X$:

$$d(x,z) \leq d(x,y) + d(y,z) \quad \text{(driehoeksongelijkheid)}.$$

Als $d: X \times X \to \mathbb{R}$ een metriek is, dan heet het paar (X,d) een *metrische ruimte*.

Het is duidelijk dat op een verzameling X meerdere metrieken kunnen zijn gedefinieerd. Toch spreekt men vaak over de metrische ruimte X i.p.v. over de metrische ruimte (X,d). Aangenomen is dan dat over de stilzwijgend gemaakte keuze van de metriek d geen misverstand kan bestaan. In het vervolg is X nu steeds een metrische ruimte met metriek d.

Voorbeeld 2.2. $X = \mathbb{R}$ is de verzameling van de reële getallen, met $d(x,y) = |x-y|$.

Opmerking 2.3. Iedere deelverzameling van een metrische ruimte is met dezelfde metriek weer een metrische ruimte.

Definitie 2.4. Laat $A \subset X$.
(i) Een punt $a \in A$ heet een *inwendig punt* van A als

$$\varepsilon xists \delta > 0: \quad B_\delta(a) = \{x \in X: \ d(x,a) < \delta\} \ \subset \ A.$$

De verzameling $B_\delta(a)$ heet de open bol met straal δ en middelpunt a.

(ii) Een punt $a \in A$ heet een *geisoleerd punt* van A als

$$\exists \delta > 0 : \quad B_\delta(a) \cap A = \{a\}.$$

(iii) Een punt $p \in X$ heet een *ophopingspunt* van A als

$$\forall \delta > 0 \; \exists a \in A : \quad a \neq p \;\text{ en }\; d(a,p) < \delta.$$

(iv) Als elk punt van A een inwendig punt van A is, dan heet A een *open deelverzameling* van X.
(v) Als het complement van A,

$$A^c = X - A = \{x \in X : x \notin A\},$$

open is, dan heet A een *gesloten deelverzameling* van X.

In het vervolg zullen we kortweg zeggen dat A open (gesloten) is als A een open (gesloten) deelverzameling van X is.

LET OP! *Definitie 2.4 benoemt eigenschappen van punten $a \in A$ en $p \in X$ met $A \subset X$ en X een metrische ruimte. Maar A is zelf ook weer een metrische ruimte en dus te zien in de rol van X hierboven. Je kunt de grotere X dan verder vergeten bij het doen van uitspraken over deelverzamelingen van A. Bijvoorbeeld over $A \subset A$, net zoals je uitspraken over X kunt doen, gezien als deelverzameling van de metrische ruimte X.*

Voorbeeld: *De gehele getallen vormen een verzameling waarvoor geldt dat $\mathbb{Z} \subset \mathbb{R}$. In $\mathbb{Z}$ is elk punt een geisoleerd punt en elke bol met straat $\frac{1}{2}$ een singleton, dus $\{0\}$ is een open deelverzameling van $\mathbb{Z}$ maar geen open deelverzameling van $\mathbb{R}$.*

Stelling 2.5. (i) Iedere open bol is een open verzameling.
(ii) Een verzameling A in X is open dan en slechts dan als A een vereniging van open bollen is.

Bewijs. Opgave.

Stelling 2.6. (i) X is open, de lege verzameling is open.
(ii) De vereniging van elke collectie open deelverzamelingen is open.
(iii) De doorsnede van elk eindig aantal open deelverzamelingen is open.

Bewijs. Opgave.

Stelling 2.7. (i) X is gesloten, de lege verzameling is gesloten.
(ii) De doorsnede van elke collectie gesloten deelverzamelingen is gesloten.
(iii) De vereniging van elk eindig aantal gesloten verzamelingen is gesloten.

Bewijs. Opgave.

Stelling 2.8. Laat $A \subset X$. Dan is A gesloten dan en slechts dan als A al zijn ophopingspunten bevat.

Bewijs. Neem eerst aan dat A gesloten is en laat p een ophopingspunt zijn van A. We moeten laten zien dat $p \in A$. Stel niet. Dan $p \in A^c$. Maar A^c is open, dus p is een inwendig punt van A^c. Zodoende is er een $\delta > 0$ waarvoor $B_\delta(p) \subset A^c$, in tegenspraak met de veronderstelling dat p een ophopingspunt is van A. Dus p ligt wel in A.

Neem vervolgens aan dat A een verzameling is die al zijn ophopingspunten bevat. We tonen aan dat A gesloten is door te bewijzen dat A^c open is. Zij $p \in A^c$. Dan is p geen ophopingspunt van A, dus er is een $\delta > 0$ zo dat $B_\delta(p)$ geen punten van $A - \{p\}$ bevat, en wegens $p \in A^c$ betekent dit dat $B_\delta(p) \subset A^c$. m.a.w. p is een inwendig punt van A^c. Dit geldt voor elke $p \in A^c$, en dus is A^c open. Q.e.d.

Definitie 2.9. Een rij $(x_n)_{n=1}^\infty$ in X heet *convergent* als er een $x_0 \in X$ is zo dat

$$\forall \varepsilon > 0 \; \exists n_\varepsilon \in \mathbb{N} : \quad n \geq n_\varepsilon \implies d(x_n, x_0) < \varepsilon.$$

Het punt x_0 heet de *limiet* van de rij.

Stilzwijgend zijn n, n_ε, m en met andere letters in het midden van het alphabet genoteerde variabelen vaak elementen van $\mathbb{N}$. De definitie wordt vaak geschreven zonder de subindex ε:

$$\forall \varepsilon > 0 \; \exists N : \quad n \geq N \implies d(x_n, x_0) < \varepsilon.$$

Stelling 2.10. Als de rij $(x_n)_{n=1}^\infty$ in X convergent is, dan is de limiet x_0 eenduidig bepaald, notatie

$$\lim_{n\to\infty} x_n = x_0 \quad \text{of} \quad x_n \to x_0.$$

Bewijs. Neem aan dat de rij twee verschillende limieten heeft, zeg x_0 en x_0'. Omdat $x_0 \neq x_0'$, is $d(x_0, x_0') > 0$. Kies $0 < \varepsilon < \frac{1}{2} d(x_0, x_0')$. Dan is er een n_ε zo dat $d(x_n, x_0) < \varepsilon$ voor alle $n \geq n_\varepsilon$. Evenzo is er een n_ε' zo dat $d(x_n, x_0') < \varepsilon$ voor alle $n \geq n_\varepsilon'$. Met behulp van de driehoeksongelijkheid volgt nu

$$d(x_0, x_0') \leq d(x_0, x_n) + d(x_n, x_0') < \varepsilon + \varepsilon < d(x_0, x_0').$$

voor $n \geq \max(n_\varepsilon, n_e')$, tegenspraak. Q.e.d.

In plaats van "de rij $(x_n)_{n=1}^\infty$ is convergent", zegt men ook wel dat "$\lim_{n\to\infty} x_n$ bestaat".

Stelling 2.11. Als de rij $(x_n)_{n=1}^{\infty}$ in X convergent is, dan is de rij begrensd, d.w.z. bevat in een vaste (open) bol $B_\delta(a) \subset X$.

Bewijs. Opgave.

Stelling 2.12. Laat $A \subset X$ gesloten zijn. Als de rij $(x_n)_{n=1}^{\infty}$ in A convergent is in X, dan ligt de limiet in A.

Bewijs. Opgave.

Stelling 2.13. Laat $A \subset X$ en $p \in X$. De volgende drie uitspraken zijn equivalent:
(i) p is een ophopingspunt van A.
(ii) Er bestaat een rij $(a_n)_{n=1}^{\infty}$ in $A - \{p\} = \{x \in A : \ x \neq p\}$ die convergent is met p als limiet.
(iii) Iedere (niet-lege) open bol met middelpunt p bevat oneindig veel punten van A.

Bewijs. $(i \implies ii)$. Neem aan dat p een ophopingspunt van A is. We construeren de rij $(a_n)_{n=1}^{\infty}$ in A. Kies $\varepsilon_1 > 0$. Dan is er een $a_1 \in A - \{p\}$ met $d(a_1, p) < \varepsilon_1$. De rij $(a_n)_{n=1}^{\infty}$ wordt nu verder inductief gedefinieerd door voor $n = 1, 2, \ldots$, nadat a_n gekozen is, $\varepsilon_{n+1} = \frac{1}{2} d(a_n, p)$ te stellen, en $a_{n+1} \in A - \{p\}$ met $d(a_{n+1}, p) < \varepsilon_{n+1}$ te kiezen. Opgave: laat zien dat $\varepsilon_{n+1} < \frac{1}{2}\varepsilon_n$ en dat de rij naar p convergeert.
$(ii \implies iii)$. Opgave.
$(iii \implies i)$. Triviaal.

Correctie! *In de 1993-versie stond in Stelling 2.13 en in het bewijs daarvan hier en daar nog een a waar een p moest staan.*
Het gebruik van definities met de quantor voor alle ($\forall$). *Belangrijk is om te onthouden dat een definitie die begint met $\forall \delta > 0$ geldt "iets" waarbij δ een rol speelt, pas echt gebruikt is als voor een rij $\delta_n \downarrow 0$ dat "iets" gebruikt is. En in plaats van δ's kunnen natuurlijk ook ε's gebruikt worden.*
Variaties op het bewijs. *Het dalend kiezen van de rij $\varepsilon_n > 0$ in $(i \implies ii)$ kan natuurlijk op vele manieren. Met $\varepsilon_1 = 1$ en vervolgens*

$$\varepsilon_{n+1} = \min(d(a_n, p), \frac{1}{n+1})$$

voor $\varepsilon_2, \varepsilon_3, \ldots$ werkt het bewijs net zo goed. Met die constructie volgt dan meteen dat $\varepsilon_n \leq \frac{1}{n}$. In het als opgave te geven bewijs van $a_n \to p$ moet je voor alle $\varepsilon > 0$ een N vinden zoals onder Definitie 2.9. Kies daartoe N met $\frac{1}{N} < \varepsilon$ (waarom kan dat?) en gebruik de bijbehorende ε_N gedefinieerd zoals hier direct boven.

De Archimedische eigenschap van de verzameling van de reële getallen. *Waarom bestaat er voor elke ε eigenlijk een n met $\frac{1}{n} < \varepsilon$? Wel, indien niet dan zou er een $\varepsilon > 0$ zijn met $\frac{1}{n} \geq \varepsilon$ en dus $n \leq \frac{1}{\varepsilon}$ voor alle $n \in \mathbb{N}$. De (niet-lege) verzameling $\mathbb{N}$ zou dan naar boven begrensd zijn in $\mathbb{R}$ en in $\mathbb{R}$ een kleinste bovengrens S hebben. Dan is $S - \frac{1}{2}$ geen bovengrens en dus bestaat er een $N \in \mathbb{N}$ met $S - \frac{1}{2} < N \leq S$ en bijgevolg $N + 1 > S + \frac{1}{2}$. Om dat $N + 1 \in \mathbb{N}$ kan S dus geen bovengrens zijn van $\mathbb{N}$, laat staan de kleinste.*

Zonder epsilons kan het dus ook. *De nu bewezen uitspraak dat onder elke $\varepsilon > 0$ altijd een $\frac{1}{n}$ zit met $n \in \mathbb{N}$ wordt de Archimedische eigenschap van $\mathbb{R}$ genoemd en maakt dat iedere definitie die begint met $\forall \varepsilon > 0$ en eindigt met $< \varepsilon$ kan worden vervangen door een definitie die begint met $\forall n \in \mathbb{N}$ en eindigt met $< \frac{1}{n}$. Alleen komen we dan al snel letters in het midden van het alphabet te kort.*

Om welk axioma voor, of eigenschap van de verzameling van de reële getallen ging het? *De Archimedische eigenschap geldt dankzij het axioma over het bestaan van kleinste bovengrenzen in $\mathbb{R}$ voor naar bovengrensde niet-lege deelverzamelingen van $\mathbb{R}$. In het boek van Croom wordt dit besproken in Sectie 2.1.*

Croom *gebruikt in zijn Sectie 3.2 het woord limietpunt als ander woord voor ophopingspunt. Ik reserveer de term limietpunt voor het gebruik zoals in Definitie 3.1 hieronder: limiet van een convergente deelrij van een gegeven rij. Dus rijen kunnen limietpunten hebben en verzamelingen ophopingspunten. Een rij in (A of) X is strict genomen ook geen deelverzameling van (A of) X maar een functie of afbeelding van $\mathbb{N}$ naar (A of) X. Croom's index verwijst voor limit point naar pagina 66 waar Stelling 3.6 komt na de definitie onderaan pagina 65, en daar zie je dat limietpunt bij Croom een andere naam is voor ophopingspunt. Een naamgeving wellicht verdedigbaar door de uitspraak dat bij een ophopingspunt p van A in X altijd een rij $(a_n)_{n=1}^{\infty}$ te vinden is waarvoor $d(a_n, p)$ strict dalend is in n en convergeert naar 0. Merk op dat voor $n \to \infty$ de equivalentie*

$$a_n \to p \iff d(a_n, p) \to 0$$

vanuit de definitie van convergentie vanzelfsprekend is.

Definitie A-1. *Laat weer $A \subset X$ met X een metrische ruimte. De afsluiting van A is de vereniging van A met al zijn ophopingspunten, notatie $\bar{A}$. Dus $p \in \bar{A}$ betekent dat de implicatie*

$$p \notin A \implies p \text{ is een ophopingspunt van } A$$

moet gelden.

Opgave A-2. *Bewijs dat $\bar{A}$ gesloten is in X. Hint: bewijs dat een ophopingspunt van $\bar{A}$ ook een ophopingspunt van A is gebruik Stelling 2.8.*

Opgave A-3. *De sterkere uitspraak is dat $\bar{A}$ de kleinste gesloten verzameling in X is die A bevat: bewijs dat $\bar{A}$ de doorsnede is van alle gesloten deelverzamelingen F van X met $A \subset F$. Hint: die doorsnede D is van vanwege Stelling 2.7 gesloten dus vanwege Opgave A-2 is $D \subset \bar{A}$. Kan A ophopingspunten hebben die niet in die doorsnede liggen?*

Croom noemt de verzameling van alle ophopingspunten de afgeleide verzameling van A, zonder verdere notatie, en schrijft $cl(A)$ voor $\bar{A}$.

We merken nog eens op dat de bovenstaande begrippen in principe afhangen van de keuze van de metriek op X. Toch kunnen verschillende metrieken tot hetzelfde leiden.

Definitie 2.14. Laat X een metrische ruimte zijn met metriek d, en laat $\bar{d}$ een andere metriek op X zijn. Dan heten d en $\bar{d}$ *equivalent* op X als er een reëel getal $\lambda > 0$ bestaat zo dat

$$\frac{1}{\lambda} d(x,y) \leq \bar{d}(x,y) \leq \lambda d(x,y) \quad \forall x, y \in X.$$

Opgave 2.15. Laat zien dat bij overgang op een equivalente metriek op X de in deze Sectie geintroduceerde begrippen (open, gesloten, convergent, etc.) niet veranderen. Hetzelfde geldt voor de begrippen die in de volgende drie secties worden behandeld (rijkompaktheid, volledigheid en continuiteit).

Op grond van het voorafgaande is het duidelijk dat voor het welslagen van de in de inleiding geschetste bewijsmethode, de geslotenheid van A vereist is. Dit staat echter los van de vraag of de in de inleiding geconstrueerde rij een limietpunt heeft.

3. Rijkompakte verzamelingen. We gaan nu de deelverzamelingen A karakteriseren waarvoor de in de inleiding geschetste bewijsmethode zal slagen.

Definitie 3.1. Laat $(x_n)_{n=1}^{\infty}$ een rij zijn in X.
(i) Als $(n_k)_{k=1}^{\infty}$ een strict stijgende rij natuurlijke getallen is, dan heet de rij $(x_{n_k})_{k=1}^{\infty}$ een *deelrij* van $(x_n)_{n=1}^{\infty}$.
(ii) Als $x_0 \in X$ de limiet is van een convergente deelrij van $(x_n)_{n=1}^{\infty}$, dan heet x_0 een *limietpunt* (ook wel: rijophopingspunt) van $(x_n)_{n=1}^{\infty}$.

Stelling 3.2. Een rij $(x_n)_{n=1}^{\infty}$ in X heeft een convergente deelrij met limiet x_0 dan en slechts dan als

$$\forall \varepsilon > 0 \ \forall n \ \exists k \geq n : \quad d(x_k, x_0) < \varepsilon.$$

Bewijs. Opgave.

Definitie 3.3. Laat $A \subset X$. A heet *rijkompakt* als iedere rij in A een limietpunt in A heeft.

Er is nog een andere definitie van kompaktheid die niet uitgaat van de metriek op X, maar van de open deelverzamelingen van X. In de appendix zullen we deze definitie behandelen, en laten zien dat de beide definities voor deelverzamelingen van metrische ruimten hetzelfde betekenen.

Stelling 3.4. Gesloten deelverzamelingen van rijkompakte verzamelingen zijn rijkompakt.

Bewijs. Merk eerst op: als $G \subset A \subset X$, dan kan het gesloten zijn van G op twee manieren worden opgevat: gesloten in A of gesloten in X. Opgave: laat zien dat, als A gesloten is in X, dan

$$G \text{ is gesloten in } A \iff G \text{ is gesloten in } X.$$

Stel dat $(a_n)_{n=1}^{\infty}$ een rij is in G. Omdat A rijkompakt is, bestaat er een convergente deelrij met limiet in A. Daar G gesloten is ligt de limiet in G. Conclusie: elke rij in G heeft een convergente deelrij met limiet in G. Q.e.d.

Stelling 3.5. (i) Laat (X_1, d_1) en (X_2, d_2) twee metrische ruimten zijn. Dan is het Cartesisch produkt X van X_1 en X_2,

$$X = X_1 \times X_2 = \{(x_1, x_2) : \quad x_1 \in X_1,\ x_2 \in X_2\},$$

weer een metrische ruimte t.a.v. de metriek gedefinieerd door

$$d(x, y) = d_1(x_1, y_1) + d_2(x_2, y_2) \ \ \forall x = (x_1, x_2),\ y = (y_1, y_2) \in X = X_1 \times X_2.$$

(ii) Als A_1 rijkompakt is in X_1 en A_2 rijkompakt is in X_2, dan is $A = A_1 \times A_2$ rijkompakt in $X = X_1 \times X_2$.

Bewijs. Opgave.

Stelling 3.6. Laat A een rijkompakte deelverzameling zijn van X. Dan is A gesloten en begrensd.

Bewijs. Opgave.

4. Volledigheid.

Definitie 4.1. Een rij $(x_n)_{n=1}^{\infty}$ in X heet *een Cauchyrij* als

$$\forall \varepsilon > 0\ \exists n_\varepsilon : \quad m, n \geq n_\varepsilon \implies d(x_m, x_n) < \varepsilon.$$

Stelling 4.2. Iedere convergente rij is een Cauchyrij.

Bewijs. Zij $(x_n)_{n=1}^{\infty}$ een convergente deelrij met limiet x_0. Laat $\varepsilon > 0$, en gebruik de definitie van convergentie met $\frac{1}{2}\varepsilon$. Dan,

$$\forall m, n \geq n_{\frac{1}{2}\varepsilon} : \quad d(x_m, x_n) \leq d(x_m, x_0) + d(x_0, x_n) < \frac{1}{2}\varepsilon + \frac{1}{2}\varepsilon = \varepsilon.$$

$\varepsilon > 0$ was willekeurig, dus de rij is een Cauchyrij. Q.e.d.

Stelling 4.3. Iedere Cauchyrij is begrensd.

Bewijs. Opgave.

Stelling 4.4. Als een Cauchyrij een limietpunt heeft, dan is de Cauchyrij convergent.

Bewijs. Laat $(x_n)_{n=1}^{\infty}$ een Cauchrij zijn die een convergente deelrij heeft. Laat $\varepsilon > 0$ en gebruik de definitie van Cauchyrij. Dus

$$\forall m, n \geq n_{\varepsilon} : \quad d(x_m, x_n) < \frac{1}{2}\varepsilon.$$

Anderzijds, omdat de rij een convergente deelrij heeft, zeg met limiet x_0, bestaat er een $k_{\varepsilon} > n_{\varepsilon}$ zodat

$$d(x_{k_{\varepsilon}}, x_0) < \frac{1}{2}\varepsilon.$$

Beide ongelijkheden combinerend vinden we

$$\forall n \geq n_{\varepsilon} : \quad d(x_n, x_0) \leq d(x_n, x_{k_{\varepsilon}}) + d(x_{k_{\varepsilon}}, x_0) < \frac{1}{2}\varepsilon + \frac{1}{2}\varepsilon = \varepsilon.$$

Definitie 4.5. Als iedere Cauchyrij in X convergent is, dan heet X *volledig*.

Om na te gaan of een rij $(x_n)_{n=1}^{\infty}$ in een volledige metrische ruimte X convergent is, hoeft men slechts na te gaan dat de rij een Cauchyrij is. Daarbij is het niet nodig om a priori de limietwaarde te kennen. Het bewijs van een belangrijke stelling in de analyse, de Banach contractie stelling, berust op dit principe:

Stelling 4.6. Laat X een volledige metrische ruimte zijn, en $T : X \to X$ een contractie, d.w.z. een afbeelding met de eigenschap dat

$$\exists \theta \in [0, 1) \ \forall x, y \in X \quad d(T(x), T(y)) \leq \theta d(x, y).$$

Dan is er precies één vast punt van T, d.w.z. een punt $\bar{x}$ in X met de eigenschap dat $T(\bar{x}) = \bar{x}$. Bovendien geldt voor elke $x \in X$ dat de rij gedefinieerd door

$$x_1 = T(x), \quad x_{n+1} = T(x_n) \ (n = 1, 2, \dots),$$

convergent is met limiet $\bar{x}$.

Bewijs. Merk eerst op dat er hoogstens een vast punt kan zijn, immers als $\bar{x}$ en $\bar{y}$ vaste punten zijn, dan

$$d(\bar{x},\bar{y}) = d(T(\bar{x}),T(\bar{y})) \leq \theta d(\bar{x},\bar{y}) \implies d(\bar{x},\bar{y}) = 0 \implies \bar{x} = \bar{y}.$$

Laat nu $x \in X$ willekeurig en laat de rij $(x_n)_{n=1}^{\infty}$ de in de stelling gedefinieerde rij zijn. We gaan bewijzen dat dit een Cauchyrij is. Neem hiertoe twee natuurlijke getallen m, n met $m < n$. Dan, door herhaald toepassen van de driehoeksongelijkheid,

$$d(x_m,x_n) \leq d(x_m,x_{m+1}) + d(x_{m+1},x_{m+2}) + \cdots + d(x_{n-1},x_n).$$

We gebruiken de notatie

$$T^0(x) = x, \quad T^1(x) = T(x), \quad T^2(x) = T(T(x)), \quad T^3(x) = T(T(T(x))), \quad \text{etc.}$$

Omdat

$$d(x_k,x_{k+1}) = d(T^k(x),T^{k+1}(x)) \leq \theta^k d(x,T(x)),$$

volgt nu dat

$$d(x_m,x_n) \leq (\theta^m+\theta^{m+1}+\cdots+\theta^{n-1})d(x,T(x)) \leq \frac{\theta^m}{1-\theta}d(x,T(x)) \to 0 \quad \text{als } m \to \infty.$$

Dus $(x_n)_{n=1}^{\infty}$ is een Cauchyrij. Omdat X volledig is heeft deze rij een limiet $\bar{x}$, dus $T^n(x) \to \bar{x}$. Maar dan geldt ook dat $T^{n+1}(x) \to \bar{x}$, terwijl

$$d(T^{n+1}(x),T(\bar{x})) = d(T(T^n(x)),T(\bar{x})) \leq \theta d(T^n(x),\bar{x}) \leq d(T^n(x),\bar{x}) \to 0.$$

Dit impliceert dat $T^{n+1}(x) \to T(\bar{x})$. Omdat de limiet van een convergente rij uniek bepaald is, kunnen we dus concluderen dat $\bar{x} = T(\bar{x})$. Q.e.d.

Er is nog een andere karakterisatie van het begrip volledigheid, die wordt gegeven door de volgende stelling.

Stelling 4.7. (Cantor) Een metrische ruimte X is volledig dan en slechts dan als voor elke dalende rij gesloten deelverzamelingen

$$F_1 \supset F_2 \supset F_3 \ldots,$$

met de eigenschap dat

$$\text{diam}(F_n) = \sup_{x,y \in F_n} d(x,y) \to 0 \text{ als } n \to \infty,$$

geldt dat de doorsnede

$$\cap_{n=1}^{\infty} F_n$$

precies één punt bevat.

5. Continue afbeeldingen.

Definitie 5.1. Laat X en Y metrische ruimten zijn, $A \subset X$, en $f : A \to Y$ een afbeelding. (i) f heet *continu in* $a \in A$, als

$$\forall \varepsilon > 0 \; \exists \delta > 0 \; \forall x \in A : \quad d(x,a) < \delta \implies d(f(x), f(a)) < \varepsilon.$$

(ii) f heet continu in A als f continu is in elke $a \in A$.

Stelling 5.2. Laat X en Y metrische ruimten zijn, $A \subset X$, $f : A \to Y$ een afbeelding, en $a \in A$. Dan is f continu in a dan en slechts dan als voor elke rij $(a_n)_{n=1}^{\infty}$ in A met $a_n \to a$ in X geldt dat $f(a_n) \to f(a)$ in Y.

Bewijs. Neem aan dat f continu is in a en laat $(a_n)_{n=1}^{\infty}$ een rij zijn in A met $a_n \to a$. Kies $\varepsilon > 0$ willekeurig, en laat $\delta > 0$ de bijbehorende δ uit de definitie van continuiteit van f in a zijn. Gebruik deze δ nu als ε in de definitie van convergentie. Dan

$$n \geq n_\delta \implies d(a_n, a) < \delta \implies d(f(a_n), f(a)) < \varepsilon.$$

Dus de rij $(f(a_n))_{n=1}^{\infty}$ convergeert naar $f(a)$.

Anderzijds, als f niet continu is in a, dan

$$\exists \varepsilon > 0 \; \forall \delta > 0 \; \exists a_\delta \in A : \quad d(a_\delta, a) < \delta \text{ en } d(f(a_\delta), f(a)) \geq \varepsilon.$$

Kies nu $\delta = \frac{1}{n}$, en noem de bijbehorende a_δ nu a_n, dan is $(a_n)_{n=1}^{\infty}$ een rij in A met $a_n \to a$, terwijl de rij $(f(a_n))_{n=1}^{\infty}$ niet convergeert naar $f(a)$. Q.e.d.

Stelling 5.3. Laat X en Y metrische ruimten zijn, en $f : X \to Y$ een afbeelding. Dan is f continu op X dan en slechts dan als het inverse beeld onder f van iedere open deelverzameling van Y weer een open deelverzameling van X is.

Bewijs. Neem aan dat f continu is, en zij B een open verzameling in Y. We moeten bewijzen dat het inverse beeld onder f van B open is. We laten zien dat elk punt van $A = f^{-1}(B)$ een inwendig punt is. Laat hiertoe $a \in A$ en $b = f(a) \in B$. Omdat B open is, is b een inwendig punt van B, dus er bestaat een $\varepsilon > 0$ met $B_\varepsilon(b) \subset B$. Kies de bij ε horende δ uit de definitie van continuiteit. Dan $f(B_\delta(a)) \subset B_\varepsilon(b)$ waardoor $B_\delta(a) \subset A = f^{-1}(B)$. Dit geldt voor elke $a \in A$, dus A is open.

Omgekeerd, als het inverse beeld van elke open verzameling open is, is te bewijzen dat f continu is in elk punt van X. Laat hiertoe $a \in X$ en $b = f(a)$, en zij $\varepsilon > 0$ willekeurig. Dan is $B_\varepsilon(b)$ open in Y, dus $A = f^{-1}(B_\varepsilon(b))$ is open in X, en omdat $a \in A$, is er een $\delta > 0$ zo dat $B_\delta(a) \subset A$. Met deze δ is dan aan de uitspraak in de definitie van continuiteit voldaan. Q.e.d.

Tot nu toe hebben we gesproken over afbeeldingen $f : X \to Y$. Vaak worden afbeeldingen ook functies genoemd, met name in het geval dat $Y = \mathbb{R}$. We keren nu terug naar de vraagstelling in de inleiding.

Stelling 5.4. Laat $A \subset X$, en $f : A \to \mathbb{R}$ een continue functie. Als A rijkompakt is, dan heeft f een maximum in A.

Bewijs. Met dit bewijs waren we al begonnen in de inleiding. Dus laat $(x_n)_{n=1}^{\infty}$ de rij zijn waarvan de functiewaarden het supremum M van f op A benaderen, zoals precies gemaakt in de inleiding. Omdat A rijkompakt is, heeft deze rij een convergente deelrij $(x_{n_k})_{k=1}^{\infty}$ met limiet $a \in A$. Vanwege de continuiteit van f in a is de rij $(f(x_{n_k}))_{k=1}^{\infty}$ convergent met limiet $f(a)$. Omdat een convergente rij begrensd is sluit dit de mogelijkheid $M = +\infty$ uit, zo dat

$$M - \frac{1}{n} < f(x_n) \leq M.$$

Dit impliceert dat de rij $(f(x_n))_{n=1}^{\infty}$ convergeert naar M. Maar een deelrij convergeert naar $f(a)$. Dus $f(a) = M$. Q.e.d.

Stelling 5.5. Laat X en Y metrische ruimten zijn, $A \subset X$, en $f : A \to Y$ een continue afbeelding. Als A rijkompakt is in X, dan is het beeld van A onder f,

$$R(f) = \{f(a) : \quad a \in A\},$$

rijkompakt in Y.

Bewijs. Opgave.

Definitie 5.6. Laat X en Y metrische ruimten zijn, $A \subset X$, en $f : A \to Y$ een afbeelding. Dan heet f *uniform continu in* A, als

$$\forall \varepsilon > 0 \; \exists \delta > 0 \; \forall x, y \in A : \quad d(x, y) < \delta \implies d(f(x), f(y)) < \varepsilon.$$

Stelling 5.7. Laat X en Y metrische ruimten zijn, $A \subset X$, en $f : A \to Y$ een continue afbeelding. Als A rijkompakt is in X, dan is f uniform continu in A.

Bewijs. Stel niet, dan

$$\exists \varepsilon > 0 \; \forall \delta > 0 \; \exists x, y \in A : \quad d(x, y) < \delta \;\text{ en }\; d(f(x), f(y)) \geq \varepsilon.$$

Kies $\delta = \frac{1}{n}$ en laat x_n en y_n de bijbehorende x en y zijn zoals in de regel hierboven. Omdat A rijkompakt is heeft de rij $(x_n)_{n=1}^{\infty}$ een convergente deelrij $(x_{n_k})_{k=1}^{\infty}$, zeg met limiet $a \in A$. Dan is ook de rij $(y_{n_k})_{k=1}^{\infty}$ convergent met dezelfde limiet. (Waarom?) Maar nu geldt

$$\lim_{k \to \infty} f(x_{n_k}) = \lim_{k \to \infty} f(y_{n_k}) = f(a),$$

terwijl $d(f(x_{n_k}), f(y_{n_k})) \geq \varepsilon$, tegenspraak. Q.e.d.

6. Het geval $X = \mathbb{R}$ en $X = \mathbb{R}^N$. We hebben gezien dat rijkompakte deelverzamelingen van X altijd gesloten en begrensd zijn. Als $X = \mathbb{R}$, geldt ook het omgekeerde. We laten dit eerst zien voor een gesloten begrensd interval.

Stelling 6.1. (Heine-Borel) Laat $a, b \in \mathbb{R}$. Dan is het gesloten begrensde interval

$$[a, b] = \{x \in \mathbb{R} : \quad a \leq x \leq b\}$$

rijkompakt.

Bewijs. Laat $(x_n)_{n=1}^{\infty}$ een rij zijn in $[a, b]$. We moeten bewijzen dat deze rij een convergente deelrij heeft met limiet in $[a, b]$. We delen hiertoe het interval in twee gelijke stukken, d.w.z. in

$$[a, \frac{a+b}{2}] \text{ en } [\frac{a+b}{2}, b].$$

Dan bevat tenminste één van deze twee intervallen voor oneindig veel waarden van n het rijelement x_n. Als we dit interval $[a_1, b_1]$ noemen, kunnen we dus een deelrij van $(x_n)_{n=1}^{\infty}$ kiezen die volledig bevat is in $[a_1, b_1]$. Noteer deze rij als $(x_n^1)_{n=1}^{\infty}$. Dit argument herhalende, krijgen we een dalende rij intervallen

$$[a, b] \supset [a_1, b_1] \supset [a_2, b_2] \supset [a_3, b_3] \supset \ldots,$$

en bijbehorende steeds verdere deelrijen, genoteerd als $(x_n^j)_{n=1}^{\infty}$, met dezelfde eigenschap, namenlijk dat elke deelrij $(x_n^j)_{n=1}^{\infty}$ steeds volledig bevat is in $[a_j, b_j]$.

Vervolgens nemen we de zogenaamde diagonaalrij, dat is de rij $(x_n^n)_{n=1}^{\infty}$. Dit is zelf weer een deelrij van de oorspronkelijke rij $(x_n)_{n=1}^{\infty}$, en er geldt dat $x_n^n \in [a_n, b_n]$.

Nu is $(a_n)_{n=1}^{\infty}$ een begrensde niet-dalende rij getallen, en $(b_n)_{n=1}^{\infty}$ een begrensde niet-stijgende rij getallen. Dus bestaan

$$\alpha = \sup_n a_n \text{ en } \beta = \inf_n b_n,$$

$$a \leq \alpha \leq \beta \leq b, \text{ en } \alpha, \beta \in [a_n, b_n] \text{ voor elk natuurlijk getal } n.$$

Bovendien is $|\beta - \alpha| \leq b_n - a_n = 2^{-n}(b - a)$ voor elk natuurlijk getal n, dus $\alpha = \beta$.

Zij $\varepsilon > 0$. Dan is er een n_ε zodat

$$\alpha - \varepsilon < a_{n_\varepsilon} \leq \alpha = \beta \leq b_{n_\varepsilon} < \beta + \varepsilon.$$

Maar dan geldt voor elke $n \geq n_\varepsilon$ dat

$$\alpha - \varepsilon < a_{n_\varepsilon} \leq a_n \leq x_n^n \leq b_n \leq b_{n_\varepsilon} < \beta + \varepsilon.$$

Dus

$$x_n^n \in (\alpha - \varepsilon, \alpha + \varepsilon), \text{ m.a.w. } |x_n^n - \alpha| < \varepsilon.$$

Omdat $\varepsilon > 0$ willekeurig was betekent dit dat de (deel)rij $(x_n^n)_{n=1}^\infty$ convergent is met limiet $\alpha \in [a, b]$. Q.e.d.

Gevolg 6.2. (Bolzano-Weierstrass) Iedere begrensde rij in $\mathbb{R}$ heeft een convergente deelrij.

Stelling 6.3. Laat $A \subset \mathbb{R}$. Dan geldt

$$A \text{ is rijkompakt } \iff A \text{ is gesloten en begrensd.}$$

Bewijs. Opgave.

Stelling 6.4. Laat A een gesloten begrensde deelverzameling van $\mathbb{R}$ zijn. Als $f : A \to \mathbb{R}$ een continue functie is, dan heeft f een maximum in A.

Bewijs. Opgave.

Stelling 6.5. Iedere Cauchyrij in $\mathbb{R}$ is convergent (m.a.w. $\mathbb{R}$ is volledig).

Bewijs. Opgave.

Stelling 6.6. (absoluut convergente reeksen zijn convergent) Als $(a_n)_{n=1}^\infty$ een rij is in $\mathbb{R}$, en

$$\sum_{n=1}^{\infty} |a_n| = \lim_{k\to\infty} \sum_{n=1}^{k} |a_n|$$

bestaat, dan bestaat ook

$$\sum_{n=1}^{\infty} a_n = \lim_{k\to\infty} \sum_{n=1}^{k} a_n.$$

Bewijs. Opgave. Hint: laat zien dat de rij $(s_k)_{k=1}^\infty$, gedefinieerd door

$$s_k = \sum_{n=1}^{k} a_n,$$

een Cauchyrij is.

De laatste vijf stellingen gelden ook voor de metrische ruimte

$$\mathbb{R}^N = \{x = (x_1, x_2, \ldots, x_N) : \ x_1, x_2, \ldots, x_N \in \mathbb{R}\}, \quad (N \in \mathbb{N})$$

met de standaard Euclidische metriek

$$d(x,y) = \sqrt{(x_1-y_1)^2 + (x_2-y_2)^2 + \cdots + (x_N-y_N)^2}.$$

Dit kunnen we inzien door $\mathbb{R}^N$ op te vatten als het herhaald Cartesisch produkt van $\mathbb{R}$ met zich zelf. Echter, de produktmetriek zoals gedefinieerd in Sectie 3, is niet de Euclidische metriek, maar

$$d_{prod}(x,y) = |x_1-y_1| + |x_2-y_2| + \cdots + |x_N-y_N|.$$

Opgave 6.7. Laat zien dat deze twee metrieken equivalent zijn, en bewijs Stelling 6.2 tot en met 6.6 voor $\mathbb{R}^N$.

Appendix. Kompaktheid en rijkompaktheid.

Definitie. Laat $A \subset X$.
(i) Een door een verzameling I geindexeerde collectie (open) deelverzamelingen $\{O_i : \ i \in I\}$ van X heet een *(open) overdekking* van A als

$$A \ \subset \ \cup_{i\in I} O_i.$$

(ii) Als iedere open overdekking $\{O_i : \ i \in I\}$ van A een eindige deeloverdekking van A bevat, d.w.z.

$$\exists i_1, i_2, \ldots, i_m \in I : \quad A \ \subset \ \cup_{j=1,2,\ldots,m} O_{i_j},$$

dan heet A *kompakt* .

Stelling. Laat $A \subset X$. Dan

$$A \text{ is kompakt } \iff A \text{ is rijkompakt.}$$

Bewijs. ($\Longrightarrow$). Stel A is kompakt en laat $(a_n)_{n=1}^{\infty}$ en rij zijn in A. We moeten bewijzen dat A een convergente deelrij met limiet in A heeft. Stel niet, dan is er voor elke $p \in A$ een $\varepsilon_p > 0$ en een n_p zodanig dat $a_n \notin B_{\varepsilon_p}(p)$ voor alle $n > n_p$. Neem nu als open overdekking van A de zojuist gevonden open bollen, dus

$$\{B_{\varepsilon_p}(p) : \ p \in A\}.$$

Omdat A kompakt is heeft deze overdekking een eindige deeloverdekking, dus er zijn $p_1, p_2, \ldots, p_m$ in A zodat

$$A \subset B_{\varepsilon_{p_1}}(p_1) \cup B_{\varepsilon_{p_2}}(p_2) \cup \cdots \cup B_{\varepsilon_{p_m}}(p_m),$$

waardoor A op zijn hoogst $n_1 + n_2 + \cdots + n_m$ punten van de rij kan bevatten, m.a.w. de rij bevat slechts eindig veel verschillende punten van A,

en tenminste een punt moet dus oneindig vaak voorkomen. Dus kunnen we een deelrij maken waarin dit punt alleen maar voorkomt, en dit is dan de gezochte deelrij.

($\Leftarrow$) Dit heeft wat meer voeten in de aarde. Daartoe eerst het volgende.

Definitie. Laat A een deelverzameling zijn van een metrische ruimte X. Dan heet A *aftelbaar kompakt* als voor elke rij open verzamelingen $(O_n)_{n=1}^{\infty}$ met

$$A \subset \cup_{n=1}^{\infty} O_n,$$

er een index k is zo dat

$$A \subset \cup_{n=1}^{k} O_n.$$

Stelling. Als $A \subset X$ rijkompakt is, dan is A aftelbaar kompakt.

Bewijs. Neem aan dat A rijkompakt is maar niet aftelbaar kompakt. Dan is er een rij open verzamelingen $(O_n)_{n=1}^{\infty}$ van A die A geheel bedekt, en die geen eindige deeloverdekking heeft. Dat betekent dat er voor elk natuurlijk getal k een punt p_k in A is met

$$p_k \notin \cup_{n=1}^{k} O_n.$$

Omdat A rijkompakt is heeft de rij $(p_n)_{n=1}^{\infty}$ een convergente deelrij met limiet p in A. Dus moet er een m zijn waarvoor $p \in O_m$. Omdat O_m open is zijn er dus oneindig veel punten van de rij $(p_n)_{n=1}^{\infty}$ die in O_m liggen. Voor $n > m$ is dit in tegenspraak met de keuze van p_n. Q.e.d.

Definitie. Laat A een deelverzameling zijn van een metrische ruimte X. Dan heet A *totaal begrensd* als er voor elke $\varepsilon > 0$ een eindige deelverzameling $\{p_1, p_2, \ldots, p_n\}$ van A bestaat met

$$A \subset B_\varepsilon(p_1) \cup B_\varepsilon(p_2) \cup \cdots \cup B_\varepsilon(p_n).$$

Stelling. Als $A \subset X$ rijkompakt is, dan is A totaal begrensd.

Bewijs. Stel A is niet totaal begrensd. Dan is er een $\varepsilon > 0$ waarvoor geen eindige verzameling als in de definitie kan worden gevonden. Kies nu $p_1 \in A$. Inductief kiezen we nu voor $n = 1, 2, \ldots$ een punt $p_{n+1} \in A$ met de eigenschap dat

$$p_{n+1} \notin B_\varepsilon(p_1) \cup B_\varepsilon(p_2) \cup \cdots \cup B_\varepsilon(p_n).$$

Maar dan is $d(p_i, p_j) \geq \varepsilon$ voor alle $i \neq j$. Dus de rij $(p_n)_{n=1}^{\infty}$ kan geen convergente deelrij hebben, in tegenspraak met de rijkompaktheid van A. Q.e.d.

Definitie. Een metrische ruimte X heet *separabel* als er een rij is in X zo dat elk punt van X een limietpunt is van deze rij.

Stelling. Als $A \subset X$ totaal begrensd is, dan is A separabel.

Bewijs. Voor elke k bestaat er een eindige deelverzameling A_k van A zo dat elk punt van A dichter dan $\frac{1}{k}$ bij een punt van A_k ligt. Maak nu een rij door eerst de elementen van A_1 te kiezen, dan de elementen van A_2, A_3, enzovoort. Dan is elk punt van A een limietpunt van de aldus verkregen rij. Q.e.d.

Stelling. Als $A \subset X$ separabel is, dan heeft elke open overdekking van A een aftelbare deeloverdekking.

Bewijs. Laat $(p_n)_{n=1}^{\infty}$ een rij zijn in A met de eigenschap dat elk punt van A limietpunt is van deze rij, en laat $\{O_i : \ i \in I\}$ een open overdekking zijn van A. Dan is elk punt p in A bevat in een O_i. Kies nu een natuurlijk getal m zo groot dat $B_{\frac{2}{m}}(p) \subset O_i$, en daarna een n zo dat $d(p_n, p) < \frac{1}{m}$. Dan is

$$p \in B_{\frac{1}{m}}(p_n) \subset O_i.$$

Laat J de verzameling zijn van de paren natuurlijke getallen (m, n) die we zo tegenkomen als p de verzameling A doorloopt. Dan is J aftelbaar en

$$A \subset \cup_{(m,n)\in J} B_{\frac{1}{m}}(p_n).$$

Maar voor elke $(m, n) \in J$ is er tenminste één O_i die $B_{\frac{1}{m}}(p_n)$ bevat. Kies er voor elke $(m, n) \in J$ precies één. Dit geeft een aftelbare deelcollectie die A overdekt. Q.e.d.

Uit bovenstaande stellingen volgt dat als $A \subset X$ rijkompakt is, dat A ook "gewoon"kompakt is.

Leiden, juni 1993.

TUSSENWOORD

Dit boek is een redelijk systematische inleiding in de analyse geworden.

Er is (nog) geen literatuurlijst. Alles is uit de lucht geplukt. Alleen aan het Lemma van Morse had ik niet gedacht om het op te nemen. De formulering en het bewijs van Stelling 11.6 zijn een aanpassing van wat de Utrechters kennen uit het vermaarde Analyse dictaat van Duistermaat en Kolk. Bob Rink gaf me dat dictaat, dat daarna een boek geworden is, in verband met wat iedereen over sub- en immersies zou moeten weten. Nu, dat wordt door dit nieuwe manuscript wel gedekt, al worden de woorden nog niet expliciet gebruikt. En toen zag ik daar dat Lemma van Morse als toepassing van de impliciete functiestelling. Aardig nu is dat je die impliciete functiestelling helemaal niet nodig hebt omdat de matrixvergelijking in het bewijs van het Lemma van Morse makkelijker is dan'ie eruit ziet. Maar dat zag ik pas bij het in stelling brengen van die stelling ten behoeve van (11.10). Wel leerzaam, in meerdere opzichten.

Het nieuwe Hoofdstuk 8 is geschreven terwijl ik uit de boeken van Edwards en Evans college gaf. In Hoofdstuk 12 heb ik ook een oude Leids dictaatje gehangen waarvan ik de texfile onlangs teruggevonden heb. Dat dictaatje heeft flinke overlap met de eerste twee secties van die appendix. Uit de tijd dat ik nog braaf $< \varepsilon$ schreef i.p.v. $\leq \varepsilon$.

De index moet nog bijgewerkt worden. Wat er na de index komt wordt onderdeel van een groter boek, zie Hoofdstuk 7.

Index

$\forall \varepsilon > 0$, zie voetnoten, 49
e tot de macht x, 119

absolute waarde van $z \in \mathbb{C}$, 130
absoluut convergent, 86, 501
afgeleide functie, 73
afgeleide in punt: $F'(x_0) = f(x_0)$, 72
afhankelijke variabelen, 220
algebra voor afgeleide van x^7, 110
alles via convergente rijen, 51
arctangens en zo, 123
argument van $z \in \mathbb{C}$: $\arg z$, 130
asymptotische formules, 116
axioma's voor de natuurlijke getallen, 504

Banachruimte $C([a, b])$, 58
Basisboek Wiskunde, 501
bijecties komen in paren, 78
bolcoördinaten, 230
Boolese algebra, 504
bovengrens, 48
bovensom, 44

Cauchy integraalformule, 241
Cauchy integraaltransformatie, 243
Cauchyrij, 501
cijferen, 485, 486, 490, 493
commutatieve Banachalgebra, 58
complexe getallen, 130
continu in een punt, 50
continuïteit: regels, 109
convergent, 501
convergentiestraal, 112
cosinus, 119

delen vermijden., 140
delta expliciet bij 2ε, 111
differentiaaloperator, 221
differentiaalquotiënt, 105
differentiaalvergelijking, 113
differentieerbaarheid impliciete f, 85
differentieerbare functie, 73
differentiequotiënten, 73
differenties als $\int_0^1$, 77
divergentie, 222
divergentie in poolcoördinaten, 227
domein, 507
driehoeksongelijkheid integralen, 57
dummy, 46
dynamisch systeem, 118

echte toepassing, 118
eenparige cirkelbeweging, 120
epsilons en delta's rijgen, 52, 83
equivalentieklasse, 506
equivalentierelatie, 506
exp, 119

Flowers for Algernon, 244
Fourierreeks, 230

gehele getallen $\mathbb{Z}$, 480
gemeenschappelijke verfijning, 50
gradiënt, 89, 221
gradiënt in poolcoördinaten, 227
grafiek, 38, 507
grote O van, 78

Hölder continu, 50
harmonische polynomen, 139
hoofdstelling van de algebra, 131

iedere begrensde rij in $\mathbb{R}$ heeft....., 53
implicatie, 503
impliciete functiestelling, 88
in de buurt van, 79
in het echte leven, 216

index nog niet klaar, 11
infimum, 48
injectief, 78
inleveren, 506
integraal $\int_a^b f(x)\,dx$, 46
integraal met parameter, 97
integraalrekening: hoofdstelling 1, 73
integraalrekening: hoofdstelling 2, 75
integreerbaar: definitie, 54
inverse functies op intervallen, 78
inverse via meetkundige reeks, 87
IS, 102

Jacobimatrix, 225

kansdichtheden, 218
kettingregel, 11, 106, 108, 155, 220, 223
kleine o van, 72
kleinste bovengrens, 499
kopen dat boekje, 480
kwadratische vorm, 278

Laplaciaan, 137
Laurentreeks, 242
Lemma van Morse, 276
limiet van een functie voor $x \to a$, 51
limietpunt, 52, 63
lineaire benadering met restterm, 71
Lipschitz continu, 50, 71, 77, 84, 93

machtreeksen, 111, 508
matrix, 135
matrixrekening, 141
matrixvergelijking, 511
maxima en minima, Opgave 2.35, 53
meerwaardige functie, 134
meetkundige reeks, 83, 87, 140, 145, 147, 241, 243
middelwaardestelling, 73, 82, 508
mulitplicatorenstelling, 93

natuurlijke getallen $\mathbb{N}$, 480
Newton schema aangepast, 81
Newton's methode voor nulpunten, 80
normale verdeling, 219

omgekeerde 3-hoeksongelijkheid, 254
ondersommen, 44
onvoorwaardelijk convergent, 501
oppervlakte, 505

partitie, 44
polygonaal pad, 234
primitieve functie, 73
probleemstelling, 276, 508, 509
produktregel, 107

quotiëntregel, 109

radialen, 119, 120
rationale getallen $\mathbb{Q}$, 507
reële getallen $\mathbb{R}$, 480
reeks, 499
regel van Leibniz, 107
relatie, 506
resolvente van A, 268
restterm, 72, 90, 96, 100, 107, 134
rijnotatie, 502
Ronald, 506

scheiding van variabelen, 228
sinus, 119
spectrum $\sigma(A)$ van A, 268
sport, 509
standaardlimiet, 114
startwaarde voor iteratie, 80
stationair (punt), 74
stationair punt, 276
Stirling uitroepteken, 117
substitutieregel voor integralen, 101
supremum, 48
surjectief, 78

tautologie, 499
Taylor benaderingen, 99

techniek, 509
totaal differentieerbaar, 89
transponeren, 141, 225
tussen, 47
tussen (al of niet strict), 44
tussenpunten, 47
tweede orde afgeleiden, 102, 137

uniform continu, 49

vast punt van stricte contractie, 84
vectorruimte over $\mathbb{R}$, 56
verfijnen van een partitie, 48
verwisselen integralen limieten, 58, 127
volledige inductie, 503
voor alle $\varepsilon > 0$, zie voetnoten, 49
vreemde functies, 123

waardenverzameling, 47
wat praktischer: $\leq \varepsilon$, 502
wiskundig netjes, 504

zigzagintegraal, 240, 244
zonder ordening boven- en onder, 60
zonder woorden alleen kwantoren, 51

13 Terug naar het platte vlak

In dit hoofdstuk verzamelen we op informele wijze onze basiskennis over het platte vlak, met in ons achterhoofd de gedachte dat we later niet in twee maar in meer dimensies willen denken en werken: $3, 4, \ldots$, tot en met aftelbaar oneindig. Bij het schrijven van dit hoofdstuk beginnen we in taal die hopelijk ook aansluit bij de schoolles, en nemen we soms ook dat perspectief als het gaat om wat we met inproducten van vectoren formuleren. Wie voor de klas staat of gaat staan heeft daar wellicht profijt van. De meeste opgaven zijn bedoeld als onderdeel van de uitleg. Convexe en gesloten deelverzamelingen, Cauchyrijen, en projecties zijn de belangrijkste begrippen die langskomen.

13.1 Punten en vectoren in het platte vlak

Opgave 13.1. Neem pen en blanco papier en teken een xy-vlak[1].

Zo, nu kunnen we aan de slag. Met en in een plat vlak waarin elk punt P gegeven is door 2 reële coördinaten, zeg $a \in \mathbb{R}$ en $b \in \mathbb{R}$. De assen labelen we met x en y. Het punt P is dus het punt met $x = a$ en $y = b$. We nummeren in deze notatie dus met het alfabet en zolang we in het vlak zitten is dat geen probleem. Ook in de 3-dimensionale ruimte kunnen we met 3 assen en $x = a, y = b, z = c$ prima uit de voeten maar vanaf dimensie 4 is het alfabet op als we beginnen bij x.

Op enig moment zullen we dus liever vanaf het begin met $x_1 = a_1$ en $x_2 = a_2$ willen werken. Een punt P gegeven door $x_1 = a_1$ en $x_2 = a_2$ kunnen we dan gewoon x noemen, soms dik gedrukt als $\mathbf{x}$, hetgeen met pen en papier weer vervelend is. Daarom ook vaak de notatie $\underline{x} = (x_1, x_2)$ voor een willekeurig, onbekend of variabel punt in het vlak, en vaak $\underline{a} = (a_1, a_2)$ voor een gegeven (vast) punt[2] in het vlak. De assen zijn dan de x_1-as en de x_2-as.

De punten $(1, 0)$ en $(0, 1)$ markeren we door er een 1 bij te zetten waarmee de schaalverdeling op de assen vast ligt. Beide punten zien we als liggend op afstand 1 tot de oorsprong $(0, 0)$, zonder fysische eenheid[3]. Het punt $(1, 1)$ heeft met Pythagoras dan afstand $\sqrt{2}$ tot $(0, 0)$.

Van een punt kun je een vector maken. In de tekening door een lijntje te trekken van de oorsprong $O = (0, 0)$ naar een punt $\underline{a} = (a_1, a_2)$ met een

[1] Suggestie: x-as horizontaal naar rechts, y-as verticaal omhoog.

[2] Dat we ook weer kunnen variëren natuurlijk.

[3] In de schoolpraktijk wordt vaak 1 cm als afstand tussen $(0, 0)$ en $(1, 0)$ aangehouden.

pijlkopje in $\underline{a}$. Het pijltje associëren we met de vector

$$\vec{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix},$$

en de lengte van het pijltje is met Pythagoras weer gelijk aan $\sqrt{a_1^2 + a_2^2}$. Correspondentie met de tekening of niet, de (Euclidische) norm van $\underline{a}$ en $\vec{a}$ is bij afspraak gelijk aan en genoteerd als

$$|\underline{a}| = |\vec{a}| = \sqrt{a_1^2 + a_2^2},$$

en voldoet aan de driehoeksongelijkheid. Er geldt voor alle $\vec{a}, \vec{b} \in \mathbb{R}^2$ dat

$$|\vec{a} + \vec{b}| \leq |\vec{a}| + |\vec{b}|,$$

het derde axioma voor de eigenschappen waar normen aan moeten voldoen.

Opgave 13.2. De eerste twee norm-axioma's zijn $|\vec{a}| > 0$ als $\vec{a}$ niet de nulvector is en $|t\vec{a}| = |t||\vec{a}|$ voor $t \in \mathbb{R}$ en $\vec{a} \in \mathbb{R}^2$. Verifieer dat de Euclidische norm aan de norm-axioma's voldoet.

We *denken* aan $\vec{a}$ als een pijltje dat we op kunnen schuiven[4] zodat de staart in een ander punt komt te liggen. Bijvoorbeeld in het punt $\underline{b}$, zodat de kop van het pijltje in het punt

$$\underline{c} = \underline{a} + \underline{b} = (a_1, a_2) + (b_1, b_2) = (a_1 + b_1, a_2 + b_2)$$

komt te liggen, waarbij we dan de vector

$$\vec{c} = \vec{a} + \vec{b} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \end{pmatrix}$$

hebben. De vector $\vec{a}$ ligt dan met zijn staart in $\underline{b}$ en met zijn kop in $\underline{c}$. Dat kan natuurlijk ook andersom, met de staart van $\vec{b}$ in $\underline{a}$ en de kop van $\vec{b}$ in $\underline{c}$. De afstand tussen $\underline{c}$ en $\underline{b}$ is dus de lengte van het pijltje $\vec{a} = \vec{c} - \vec{b}$: de norm van de vector $\vec{a} = \vec{c} - \vec{b}$.

We switchen regelmatig heen en weer tussen rij- en kolomnotatie en tussen punten en vectoren, al naar gelang het zo uitkomt. Een in de tijd bewegend punt $\underline{x}$ heeft op elk moment een snelheid $\vec{v}$ die we ons vanwege de fysische

[4] In het *platte* vlak geen probleem maar google op Gauss en kromming.

interpretatie het liefst met de staart in $\underline{x}$ voorstellen. En als het handig is dan zien we $\underline{x}$ ook als $\vec{x}$. Bijvoorbeeld in

$$\vec{x} = \vec{s} + t\vec{v} = \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} + t\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} + \begin{pmatrix} tv_1 \\ tv_2 \end{pmatrix} = \begin{pmatrix} s_1 + tv_1 \\ s_2 + tv_2 \end{pmatrix},$$

de formule[5] voor een punt dat beweegt over een rechte lijn l door het punt $\underline{s}$ met snelheidsvector $\vec{v}$.

Opgave 13.3. De lijn l door $\underline{s} \in \mathbb{R}^2$ met richtingsvector $\vec{v} \in \mathbb{R}^2$ kan ook gegeven worden door een vergelijking van de vorm

$$a_1x_1 + a_2x_2 = c$$

voor de punten $\underline{x} = (x_1, x_2)$ op de lijn l. Voor welke lijnen kan dat met $c = 1$? Bepaal voor die lijnen de bijbehorende a_1 en a_2.

Naast de vectoroptelling is in de vectorvoorstelling van een rechte lijn met steunvector $\vec{s}$ en richtingsvector $\vec{v}$ ook de scalaire vermenigvuldiging gebruikt. Voor iedere $t \in \mathbb{R}$ en $\vec{v} \in \mathbb{R}^2$ is $t\vec{v}$ gedefinieerd zoals je zou verwachten. De formule voor $\vec{c} = \vec{a} + \vec{b}$ gaat via $\vec{c} = \vec{x}$, $\vec{a} = \vec{s}$ en $\vec{b} = t\vec{v}$ over in de vectorvoorstelling van de lijn, waarin $\vec{x}$ de met t variërende vector is bij het punt $\underline{x}$.

In de formules mogen alle punten in het platte vlak voorkomen. En alle punten dat zijn alle punten van de vorm $\underline{x} = (x_1, x_2)$ met $x_1, x_2 \in \mathbb{R}$. Het platte vlak past daarmee weliswaar niet in ons universum maar gelukkig wel in ons hoofd, waar het de naam $\mathbb{R}^2$ gekregen heeft, met de 2 van 2-dimensionaal.

Ieder element uit de verzameling $\mathbb{R}^2$ wordt gegeven door een geordend reëel getallenpaar dat we aan kunnen geven met de letters die we willen, en met de notatie die we willen. Nummerend met het alfabet of met indices 1 en 2, achter elkaar of boven elkaar als

$$v_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + v_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = v_1\vec{e}_1 + v_2\vec{e}_2$$

geschreven, of eventueel ook als

$$v_1 + iv_2,$$

[5] Vectorvoorstelling van een lijn.

als maar duidelijk is dat v_1 de eerste, en v_2 de twee coördinaat is. De laatste twee vormen suggereren alvast de correspondentie

$$1 \quad \leftrightarrow \quad \vec{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$i \quad \leftrightarrow \quad \vec{e}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

en de representatie van de complexe getallen $\mathbb{C}$ als het (complexe) vlak $\mathbb{R}^2$ met een wat rare notatie[6].

13.2 Kortste afstanden

De kortste verbinding tussen twee punten in het vlak is de rechte lijn. In welk vlak? In het vlak dat we in ons hoofd hebben via de introductie van $\mathbb{R}^2$ in Sectie 13.1. Welke punten? Iedere $\underline{a}$ en $\underline{b}$ in die $\mathbb{R}^2$. Welke rechte lijn? Geen rechte lijn, maar het lijnstuk

$$\{t\underline{a} + (1-t)\underline{b} : \, 0 \leq t \leq 1\},$$

een stuk van de rechte lijn door steunvector $\underline{b}$ met richtingsvector $\vec{a} - \vec{b}$.

Er zijn geen andere paden van $\underline{b}$ naar $\underline{a}$ met een kortere afgelegde weg, een in het dagelijks leven op het Groningse platte land geboren uitspraak over *alle* paden van $\underline{b}$ naar $\underline{a}$, waarin twee begrippen voorkomen die wiskundig gezien hier nog niet eens gedefinieerd[7] zijn. Maar die kortste afgelegde weg moet natuurlijk wel gelijk zijn aan wat we de afstand tussen $\underline{a}$ en $\underline{b}$ noemen. Kortom, kortste afstanden gaan hier niet nog even niet over de weg van $\underline{a}$ naar $\underline{b}$. Er is maar een afstand tussen $\underline{a}$ en $\underline{b}$ en dat is

$$d(\underline{a}, \underline{b}) = |\underline{a} - \underline{b}| = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2} = |\vec{a} - \vec{b}|,$$

de lengte van de vector $\vec{a} - \vec{b}$.

Over de kortste afstand tussen $\underline{a}$ en $\underline{b}$ hoeven we het dus in het platte vlak niet te hebben. Daar is een formule voor die we als vanzelfsprekend zien. En die formule definieert een afstandsbegrip dat voldoet aan axioma's: de axioma's van een metriek[8].

Maar wat is de kortste afstand tussen een niet-lege deelverzameling A van $\mathbb{R}^2$ en een punt $\underline{b}$? Met andere woorden, als de functie $f_b : \mathbb{R}^2 \to \mathbb{R}$ gedefinieerd wordt door

$$f_{\underline{b}}(\underline{x}) = d(\underline{x}, \underline{b}) = |\vec{x} - \vec{b}|,$$

[6] En extra algebra gebaseerd op de afspraak dat i keer i is $i^2 = -1$.

[7] Om welke twee begrippen gaat het?

[8] Wat is een metriek? Zoek op.

wat kun je dan zeggen over de waardenverzameling

$$W = \{f_{\underline{b}}(\underline{x}) : \underline{x} \in A\}\,?$$

Heeft deze deelverzameling van $\mathbb{R}$ een kleinste element?

Wel, de waardenverzameling W is niet leeg en naar beneden begrensd door 0. Op grond van de axioma's (of eigenschappen) van de reële getallen heeft W dus een grootste ondergrens[9] d die we vanaf nu de afstand van $\underline{b}$ to A noemen:

$$d = d(\underline{b}, A) = \inf W = \inf_{\underline{x} \in A} d(\underline{x}, \underline{b}).$$

Dus ook als de kleinste waarde niet bestaat, of als we dat niet a priori weten, is zo de afstand d tussen $\underline{b}$ en A wiskundig gedefinieerd. Of d nu wordt aangenomen door $d(\underline{x}, \underline{b})$ voor een x in W of niet.

De wiskundige definitie vertelt ons dat voor iedere[10] positieve gehele n er een $x_n \in A$ is met

$$d(\underline{b}, A) \leq d(\underline{b}, \underline{x}_n) < d(\underline{b}, A) + \frac{1}{n},$$

want iedere n waarvoor zo'n $\underline{x}$ niet bestaat zou een grotere ondergrens voor W zijn. Of je de wiskundige de afstand d ook echt kan vinden als horende bij een $\underline{a} \in A$ via $d = d(\underline{a}, \underline{b})$ is maar de vraag natuurlijk.

Een strategie om aan de kleinste waarde d te komen is om de rij $\underline{x}_n$ convergent te kiezen. Als dat kan dan heeft de rij een limiet $\underline{a}$. Als vervolgens blijkt dat $\underline{a}$ in A ligt volgt hopelijk ook dat $d(\underline{b}, A) = d(\underline{b}, \underline{a})$. En blijft vervolgens nog de vraag of het punt in A waarin de kleinste afstand aangenomen wordt uniek is. Het gaat dus om twee zaken. Het vinden van convergerende minimaliserende rijen in A en daarna de vraag om daar altijd dezelfde limiet bij hoort.

Maar soms kun je d meteen uitrekenen. Hoewel?

Opgave 13.4. Wat is de kortste afstand tussen $\underline{a} = (1, 1)$ en de lijn met vergelijking $3x_1 + x_2 = 1$?

Opgave 13.5. De kortste afstand tussen $\underline{a} = (1, 1)$ en de deelverzameling $E \subset \mathbb{R}^2$ gegeven door $9x_1^2 + x_2^2 \leq 1$ is niet zo eenvoudig uit te rekenen. Probeer het maar. Maar is het punt in E met minimale afstand tot $\underline{a}$ uniek denk je? Waarom? Maak een plaatje.

[9] Ander woord: infimum.

[10] We mijden hier de $\epsilon > 0$, for all practical purposes is $\frac{1}{n}$ net zo goed.

Opgave 13.6. Reflecteer[11] op wat het begrip loodrecht met het begrip afstand te maken heeft.

Opgave 13.7. Teken voor verschillende (reële) waarden van a en b in je xy-vlak de vectoren[12]

$$\begin{pmatrix} a \\ b \end{pmatrix} \quad \text{en} \quad \begin{pmatrix} -b \\ a \end{pmatrix}$$

en reflecteer op het begrip loodrecht. Kun je andere paren vectoren in het vlak bedenken waarop het begrip loodrecht van toepassing is?

Opgave 13.8. Een deelverzameling $K \subset \mathbb{R}^2$ heet convex als met elk tweetal punten $\underline{a}$ en $\underline{b}$ in K ook het lijnstuk

$$\{t\underline{a} + (1-t)\underline{b} : \, 0 \leq t \leq 1\}$$

dat $\underline{a}$ en $\underline{b}$ verbindt in K ligt. Kunnen er twee punten in K zijn die $f_O(\underline{x}) = |\underline{x}|$ minimaliseren op K? Maak een plaatje dat je helpt om de vraag te beantwoorden.

13.3 Vlakke meetkunde met het inprodukt

Bij het maken van deze opgaven heb je ongetwijfeld rechte hoeken en driehoeken getekend en de (Stelling van) Pythagoras weer gebruikt, en wellicht al het inwendige produkt van vectoren gebruikt. Het *standaard inwendige produkt* in $\mathbb{R}^2$ wordt gedefinieerd door

$$\begin{pmatrix} a \\ b \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} = ax + cy,$$

hetgeen voor elke keuze van de 2-vectoren

$$\begin{pmatrix} a \\ b \end{pmatrix} \quad \text{en} \quad \begin{pmatrix} x \\ y \end{pmatrix}$$

een reëel getal definieert, dat vastgelegd wordt door de vier reële getallen a, b, x, y. De opgaven hebben je overtuigd dat twee vectoren in $\mathbb{R}^2$ loodrecht op elkaar staan precies dan als hun inwendig produkt nul is.

[11] Minimum op de rand, denk ook aan multiplicatoren van Lagrange.
[12] Al of niet met de staart in de oorsprong O.

Loodrecht is hier een begrip dat je buiten de wiskunde kende en nu in de wiskunde van betekenis hebt voorzien, en wel in het abstracte platte vlak in je hoofd, en de meetkunde die je daarin hebt leren bedrijven, al of niet gebruikmakend van twee onderling loodrecht voorgestelde coördinaatassen, gemarkeerd met 0 en 1.

De afstand van $(0,0)$ tot $\underline{a} = (a_1, a_2)$ is met Pythagoras gelijk aan $\sqrt{\vec{a} \cdot \vec{a}}$, de wortel uit het inwendige produkt van de bijbehorende vector $\vec{a}$ met zichzelf. Zo hebben we de begrippen afstand en loodrecht die we uit de dagelijkse werkelijkheid kennen in verband gebracht met het standaard inwendig produkt in $\mathbb{R}^2$, ons model voor het platte vlak. Dit verband zit stevig tussen onze oren, wat het verder ook moge betekenen. Wiskundige uitspraken doen we vanaf nu in termen van $\mathbb{R}^2$ met zijn vectoroptelling en het standaard inwendige produkt.

Opgave 13.9. Bewijs dat $|\vec{a} \cdot \vec{b}| \leq |\vec{a}||\vec{b}|$, met andere woorden, dat

$$(a_1 b_1 + a_2 b_2)^2 \leq (a_1^2 + a_2^2)(b_1^2 + b_2^2).$$

Hint: breng alles naar de rechterkant, doe de algebra en herken het kwadraat. Doe vervolgens ook

$$(a_1 b_1 + a_2 b_2 + a_3 b_3)^2 \leq (a_1^2 + a_2^2 + a_3^2)(b_1^2 + b_2^2 + b_3^2),$$

en overtuig jezelf ervan dat (even wat combinatoriek)

$$(\sum_{k=1}^{n} a_k^2)(\sum_{k=1}^{n} b_k^2) - (\sum_{k=1}^{n} a_k b_k)^2$$

de som is van $\frac{n(n-1)}{2}$ kwadraten.

Opgave 13.10. Teken twee vectoren $\vec{a}$ en $\vec{b}$ waarvoor $\vec{a} \cdot \vec{b} = 0$ en schuif een van de twee vectoren op en wel zó dat de kop van deze ene vector in de staart van de andere vector ligt (en een rechthoekige driehoek ontstaat). Werk $(\vec{a} + \vec{b}) \cdot (\vec{a} + \vec{b})$ uit tot de bekende formule voor $|\vec{a}|$, $|\vec{b}|$ en $|\vec{a} + \vec{b}|$.

Opgave 13.11. Leid met Opgave 13.9 en Opgave 13.10 nog een keer af dat de norm aan de driehoeksongelijkheid $|\vec{a} + \vec{b}| \leq |\vec{a}| + |\vec{b}|$ voldoet, ook voor $\vec{a} \cdot \vec{b} \neq 0$.

Opgave 13.12. Teken twee vectoren $\vec{a}$ en $\vec{b}$ waarvoor niet per se $\vec{a} \cdot \vec{b} = 0$ en schuif een van de twee op zó dat de kop van deze ene in de staart van de ander vector ligt (en een driehoek ontstaat). Werk $(\vec{a} + \vec{b}) \cdot (\vec{a} + \vec{b})$ en doe hetzelfde voor $\vec{a}$ en $-\vec{b}$. Beide uitdrukkingen bevatten $\vec{a} \cdot \vec{b}$ maar na sommatie vallen deze kruistermen weg. Formuleer wat bekend staat als de parallellogramwet.

Opgave 13.13. Een elegant bewijs van de Stelling van Pythagoras zonder vectoren maar met bijvoorbeeld vierkanten heeft iedereen wel eens gezien natuurlijk. Zie bijvoorbeeld

`http://www.few.vu.nl/~jhulshof/RYB.mov`

Is er ook zo'n elegant bewijs[13] van de parallellogramwet?

13.4 Projecteren op convexe verzamelingen

Vlakke en Euclidische meetkunde betreffen tamelijk expliciete zaken. Denk aan lijnen, vlakken etc. Teken een lijn in het vlak en doe wat. Het plaatje is altijd hetzelfde. Projecteren op een lijn, iedereen kan het. Bij projecteren op convexe verzamelingen gaat over een veel grotere klasse van verzamelingen maar met de algebra van het inprodukt is goed te begrijpen hoe dat gaat. Die algebra is niet beperkt tot het platte vlak. Maar nu eerst even wel.

Opgave 13.14. Als $\underline{b} \in \mathbb{R}^2$ en $K \subset \mathbb{R}^2$ niet leeg en convex is, dan heeft iedere minimaliserende rij $\underline{x}_n \in K$ met $d(\underline{x}_n, \underline{b}) \to d$ de eigenschap dat

$$d(\underline{x}_n, \underline{x}_m) \to 0 \quad \text{als} \quad m, n \to \infty$$

en dat kun je algemeen bewijzen. Neem zonder beperking der algemeenheid $\underline{b} = O$ en $d(\underline{x}_n, O)$ dalend, en laat dit zien door voor $m > n$ met de parallellogramwet $|\underline{x}_n - \underline{x}_m|^2$ af te schatten op $\epsilon_n = 4(d + \frac{1}{n})^2 - d^2)$. Hint: je hebt alleen nodig dat het midden van elk lijnstuk tussen twee punten in K weer in K zit ($t = \frac{1}{2}$ in de definitie).

Onze meetkundige kennis is in de opgaven hierboven in uitspraken over vectoren en inwendige produkten vertaald, met als opmerkelijk conclusie het resultaat in Opgave 13.14 dat zegt dat de minimaliserende rij een Cauchyrij[14]

[13] Vast wel, maar ik heb het zelf nog nooit gezien.

[14] Wat was dat ook al weer?

is. Net als in $\mathbb{R}$ zijn in $\mathbb{R}^2$ Cauchyrijen convergent. De limiet $\underline{a}$, waarvoor geldt dat

$$d(\underline{x}_n, \underline{a}) \to 0 \quad \text{als} \quad n \to \infty,$$

hoeft natuurlijk niet per se in A te liggen, maar doet dat wel als A gesloten is.

Opgave 13.15. $A \subset \mathbb{R}^2$ heet gesloten als iedere convergente rij x_n in A ook zijn limiet in A heeft. Als A niet gesloten is dan zijn er dus convergente rijen in A waarvan de limiet niet in A ligt. Bewijs dat de afsluiting $\overline{A}$, dat is A verenigd met al die limieten, altijd gesloten is.

Opgave 13.16. Voor iedere niet-lege convexe $K \subset \mathbb{R}^2$ en voor iedere $b \in \mathbb{R}^2$ bestaat er een $a \in \overline{K}$ met $d(\underline{b}, \underline{a}) = d(\underline{b}, K)$. Bewijs dit met de voorafgaande resultaten en laat zien dat $\underline{a}$ uniek is. Concludeer dat $\underline{b} \to \underline{a}$ een afbeelding $P_K : \mathbb{R}^2 \to \overline{K}$ definieert. Laat ook zien $(P_K(\vec{a}) - \vec{a}) \cdot (\vec{x} - P_K(\vec{a})) \geq 0$ voor alle $\underline{x} \in K$ en maak een plaatje om de betekenis van deze uitspraak meetkundig te begrijpen.

Opgave 13.17. Laat zien dat de afbeelding P_K een contractie is in de zin dat voor alle $\underline{x}, \underline{y} \in \mathbb{R}^2$ geldt dat $d(P_K(\underline{x}), P_K(\underline{y})) \leq d(\underline{x}, \underline{y})$. Hint: deze is lastig, spelen met het inprodukt, te leuk om voor te zeggen. Let op, voor variabele punten in K heb je nu een andere letter nodig.

Opgave 13.18. Pas de vorige opgave toe op het geval $K = l$, met l de lijn door $\underline{s}$ met richtingsvector $\vec{v}$ en geef een formule voor P_l. Hint: waarom wordt de ongelijkheid in Opgave 13.16 nu een gelijkheid voor alle $\underline{x} \in l$? Gebruik dit en reken $P_l(\underline{b})$ gewoon uit voor gegeven $\underline{b}$.

Opgave 13.19. Neem in de vorige opgave $\underline{s} = O$ en laat zien dat de nulverzameling

$$N(P_l) = \{\underline{x} \in \mathbb{R}^2 : \, P_l(\underline{x}) = \underline{0}\}$$

van P_l weer een lijn is, zeg lijn m, en dat m en l loodrecht op elkaar staan in dat vlak in je hoofd.

13.5 Andere inprodukten en bilineaire vormen

Het standaard inwendig produkt van

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \quad \text{en} \quad \vec{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

is een voorbeeld van een bilineaire functie, ook wel bilineaire vorm genoemd. Zulke *bilineaire vormen* $B : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ zijn altijd te schrijven als

$$B(\vec{x}, \vec{y}) = a_{11}x_1y_1 + a_{12}x_1y_2 + a_{21}x_2y_1 + a_{22}x_2y_2,$$

dit vanwege wat je in de volgende opgave nu uitwerkt.

Opgave 13.20. Laat zien dat als $B : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ voldoet aan

$$B(\vec{x}_1 + \vec{x}_2, \vec{y}) = B(\vec{x}_1, \vec{y}) + B(\vec{x}_2, \vec{y});$$

$$B(\vec{x}, \vec{y}_1 + \vec{y}_2) = B(\vec{x}, \vec{y}_1) + B(\vec{x}, \vec{y}_2);$$

$$B(t\vec{x}, \vec{y}) = B(\vec{x}, t\vec{y}) = tB(\vec{x}, \vec{y}),$$

voor alle $t \in \mathbb{R}$ en $\vec{x}, \vec{x}_1, \vec{x}_2, \vec{y}, \vec{y}_1, \vec{y}_2 \in \mathbb{R}^2$, dat B gegeven wordt door[15]

$$B(\vec{x}, \vec{y}) = \sum_{i,j=1}^{2} a_{ij}x_iy_j,$$

en dat $B(\vec{x}, \vec{y}) = B(\vec{y}, \vec{x})$ voor alle $\vec{x}, \vec{y} \in \mathbb{R}^2$ gelijkwaardig is met $a_{ij} = a_{ji}$ voor alle $i, j \in \{1, 2\}$.

Kortom, $B(\vec{x}, \vec{y})$ is van de vorm

$$B(\vec{x}, \vec{y}) = A\vec{x} \cdot \vec{y},$$

waarbij $A : \mathbb{R}^2 \to \mathbb{R}^2$ de lineaire afbeelding is gegeven is door

$$A\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix},$$

en de symmetrie van B equivalent is met de symmetrie van de lineaire afbeelding A en de bijbehorende matrix (a_{ij}):

$$B(\vec{x}, \vec{y}) = B(\vec{y}, \vec{x}) \quad \Leftrightarrow \quad A\vec{x} \cdot \vec{y} = \vec{x} \cdot A\vec{y} \quad \Leftrightarrow \quad a_{ij} = a_{ji}$$

[15] Let op: x_i en y_j zijn nu componenten van $\vec{x}$ en $\vec{y}$.

Een symmetrische bilineaire vorm definieert een inwendig produkt als de bijbehorende kwadratische vorm positief definiet is, dat wil zeggen

$$A\vec{x}\cdot\vec{x}>0 \quad \text{als} \quad \vec{x}\neq\vec{0}=\begin{pmatrix}0\\0\end{pmatrix},$$

en in dat geval heet A zelf ook positief[16] definiet[17]. Voorlopig zullen we in de notatie geen onderscheid maken tussen A als lineaire afbeelding en A als matrix. We schrijven dus ook

$$A=\begin{pmatrix} a_{11} & a_{12}\\ a_{21} & a_{22}\end{pmatrix}$$

en spreken over ook positief definiete (symmetrische) matrices.

Kwadratische vormen zijn homogene poynomen van graad twee in de variabelen. Een kwadratische vorm $Q:\mathbb{R}^2\to\mathbb{R}$ wordt dus gegeven door

$$Q(\vec{x})=Q(\underline{x})=Q(x_1,x_2)=q_{11}x_1^2+q_{12}x_1x_2+q_{22}x_2^2=\sum_{1\leq i\leq j=2}^{2}q_{ij}x_ix_j$$

en is altijd te schrijven als $Q(x_1,x_2)=B(\vec{x},\vec{x})=A\vec{x}\cdot\vec{x}$, met

$$a_{ii}=q_{ii} \quad \text{en} \quad a_{ij}=a_{ji}=\frac{1}{2}q_{ij} \quad (i<j).$$

Omdat

$$m=\min_{|\underline{x}|\leq 1}Q(\underline{x})=\min_{|\underline{x}|=1}Q(\underline{x}) \quad \text{en} \quad M=\max_{|\underline{x}|\leq 1}Q(\underline{x})=\max_{|\underline{x}|=1}Q(\underline{x})$$

bestaan als (op de rand aangenomen[18]) minimum en maximum van Q op de gesloten disk gegeven door

$$x_1^2+x_2^2\leq 1,$$

definieert een symmetrische A dus een (niet-standaard) inwendig produkt als $m>0$.

Opgave 13.21. Neem aan dat $0\leq m\leq M$. Laat zien dat

$$m\,\vec{x}\cdot\vec{x}\leq A\vec{x}\cdot\vec{x}\leq M\,\vec{x}\cdot\vec{x}$$

voor alle $\vec{x}\in\mathbb{R}^2$. Wat kun je zeggen zonder de aanname op de tekens van m en M?

[16] Echt iets anders dan $a_{ij}>0$ voor $i,j=1,2$.

[17] Impliciet is A dus symmetrisch verondersteld.

[18] Mini- en maximaliserende rijen $\underline{x}_1,\underline{x}_2,\ldots$ kunnen convergent gekozen worden.

De rand van de disk is een cirkel die kunnen we parametriseren met

$$x_1 = \cos(t) \quad \text{en} \quad x_2 = \sin(t),$$

waarin de functies cos en sin uniek gedefinieerd zijn door bijvoorbeeld[19]

$$\cos t = \cos(t) = \sum_{n=0}^{\infty} \frac{(-t)^{2n}}{(2n)!} = 1 - \frac{t^2}{2!} + \frac{t^4}{4!} - \frac{t^6}{6!} + \cdots$$

$$\sin t = \sin(t) = \sum_{n=0}^{\infty} \frac{(-t)^{2n+1}}{(2n+1)!} = t - \frac{t^3}{3!} + \frac{t^5}{5!} - \frac{t^7}{7!} + \cdots,$$

met[20] $\sin' = \cos$, $\cos' = -\sin$, $\cos(0) = 1$, $\sin(0) = 0$.

Opgave 13.22. Bereken het maximum M en het minimum m van de functie $q : \mathbb{R} \to \mathbb{R}$ gedefinieerd door

$$q(t) = Q(\cos t, \sin t) = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} \cos(t) \\ \sin(t) \end{pmatrix} \cdot \begin{pmatrix} \cos(t) \\ \sin(t) \end{pmatrix}$$

Hint, herschrijf als $q(t) = a\cos^2 t + b\cos t \sin t + c\sin^2 t$, neem eerst $b \neq 0$ en herleid $q'(t) = 0$ tot een vierkantsvergelijking voor $\tan t$. Verifieer dat in de *minimizers*

$$A\begin{pmatrix} \cos t \\ \sin t \end{pmatrix} = m\begin{pmatrix} \cos t \\ \sin t \end{pmatrix}$$

geldt, en in de *maximizers*

$$A\begin{pmatrix} \cos t \\ \sin t \end{pmatrix} = M\begin{pmatrix} \cos t \\ \sin t \end{pmatrix}.$$

Deze opgave laat zien dat m en M de twee reële eigenwaarden zijn van de symmetische matrix A. In het geval dat A positief definiet is nummeren we deze eigenwaarden $\lambda_1 = M \geq \lambda_2 = m > 0$. Je ziet[21] dat de bijbehorende eigenvectoren loodrecht staan. In het geval dat $M = m$ zijn alle vectoren eigenvectoren en kunnen ze loodrecht gekozen worden, $\vec{e}_1$ en $\vec{e}_2$ bijvoorbeeld.

[19] Zie [HM, hoofdstuk 10].
[20] De twee differentiaalvergelijkingen en beginvoorwaarden definiëren sin en cos.
[21] Misschien niet meteen.

Opgave 13.23. Bewijs direct, dus zonder cosinussen en sinussen, dat voor elke symmetrische (hier nog twee bij twee) matrix A geldt dat het maximum μ van de *absolute waarde* van de bijbehorende kwadratische vorm Q op $|\vec{x}| = 1$ wordt aangenomen in een eigenvector, en dat iedere *maximizer* een eigenvector is, bij μ of bij $-\mu$ (of bij allebei in bijzonder gevallen).

Opgave 13.24. De eigenvector in Opgave 13.23 bij $\lambda_1 = \pm\mu$ noemen we $\vec{v}_1$. De lijn door O met richtingsvector $\vec{v}_1$ noemen we l_1. Pas nu Opgave 13.19 toe[22] op $l = l_1$ en noem $m = l_2$. Laat zien dat A deze l_2 op zichzelf afbeeldt.

13.6 Om te onthouden

Symmetrische twee bij twee matrices komen met paren onderling loodrechte lijnen die we, zo we willen, als nieuwe coördinaatassen kunnen gebruiken. Met in die lijnen (eigen)vectoren $\vec{v}_1$ en $\vec{v}_2$ die onderling loodrecht staan en lengte 1 hebben,

$$\vec{v}_1 \cdot \vec{v}_1 = \vec{v}_2 \cdot \vec{v}_2 = 1 \quad \text{en} \quad \vec{v}_1 \cdot \vec{v}_2 = 0,$$

bij eigenwaarden λ_1 en λ_2,

$$A\vec{v}_1 = \lambda_1\vec{v}_1 \quad \text{en} \quad A\vec{v}_2 = \lambda_2\vec{v}_2.$$

In het bijzondere geval dat A een diagonaalmatrix

$$\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$$

is, krijgen we als eigenvectoren de standaardbasisvectoren $\vec{e}_1$ en $\vec{e}_2$.

Het andere belangrijke resultaat is dat we op (gesloten) convexe verzamelingen kunnen projecteren, Opgave 13.16. Niet benadrukt nog is wat de essentie was van het bewijs dat je in Opgave 13.23 hebt gegeven. Waar het resultaat in Opgave 13.16 via Opgave 13.14 en een convergente minimaliserende rij tot stand kwam, is in Opgave 13.23 een maximaliserende rij *niet* automatisch convergent en moet eerst een convergente deelrij genomen worden. En iedere begrensde rij in $\mathbb{R}^2$ heeft zo'n convergente deelrij. Alles wat we hier behandeld hebben gaat dus door voor $\mathbb{R}^3$, $\mathbb{R}^4$,..., met een kleine aanpassing bij Opgave 13.19. Pas in $\mathbb{R}^\infty$ gaat het een beetje anders.

[22] De notaties $\underline{x}$ en $\vec{x}$ liepen al door elkaar heen, liever $x = \underline{x} = \vec{x}$ vanaf nu?

13.7 Poolcoördinaten in het (complexe) vlak

We besluiten dit hoofdstuk met een korte herhaling van $\mathbb{R}^2$ gezien als de verzameling van complexe getallen $\mathbb{C}$. Het punt $(1,0)$ zien we als het getal 1 en het punt $(0,1)$ als het imaginaire getal i. We introduceren $\mathbb{C}$ door de correspondentie

$$(x,y) \in \mathbb{R}^2 \quad \leftrightarrow \quad z = x + yi = x + iy \in \mathbb{C}$$

met in $\mathbb{C}$ de gebruikelijke rekenoperaties: de complexe optelling en de complexe vermenigvuldiging. Die krijg je door te rekenen met uitdrukkingen als $z = x + iy$ en $c = a + bi$ alsof het eerstegraads polynomen in i zijn, met de afspraak dat $i^2 = -1$. De rollen van i en $-i$ zijn daarbij uitwisselbaar want ook $(-i)^2 = 1$. De coëfficiënten x, y, a, b zijn zelf reëel, en x en a heten de reële delen van respectievelijk z en c. De *imaginaire* delen zijn y en b en zijn net zo reëel als de reële delen.

We gaan ervan uit dat de lezer vertrouwd[23] is met deze complexe getallen en het waarom van de notatie en correspondentie

$$(\cos(t), \sin(t)) \quad \leftrightarrow \quad \exp(it) = \cos(t) + i\sin(t)$$

voor het over de eenheidscirkel bewegende punt $(\cos(t), \sin(t))$.

Die eenheidscirkel wordt gegeven door $|z| = 1$, waarbij de absolute waarde van $z = x + iy$ per definitie gelijk is aan

$$|z| = \sqrt{x^2 + y^2},$$

meestal r genoemd. Voor elke $r > 0$ doorloopt het punt

$$(r\cos(t), r\sin(t)) \quad \leftrightarrow \quad r\exp(it) = r(\cos(t) + i\sin(t))$$

een cirkel met straal r in het al of niet complexe vlak, en de (tijd) t is *per definitie* de hoek in radialen die de met dit punt corresponderende vector maakt met de positieve x-as. Ieder punt in het vlak wordt zo gegeven door een r en een t, en elke 2-vector is van de vorm

$$\vec{x} = \begin{pmatrix} r\cos(t) \\ r\sin(t) \end{pmatrix} = r\begin{pmatrix} \cos(t) \\ \sin(t) \end{pmatrix}, \quad \text{het scalaire product van } r \text{ en } \begin{pmatrix} \cos(t) \\ \sin(t) \end{pmatrix}.$$

Behalve de oorsprong heeft ieder punt $\underline{x}$ en iedere vector $\vec{x}$ een unieke r en een unieke t, waarbij je moet afspreken dat de t-waarden module 2π worden gerekend. En 2π per definitie het reële getal is waarvoor deze laatste karakterisatie correct is. In (tijd) $t = 2\pi$ ga je de cirkel rond.

[23] Zie anderes eventueel [HM, hoofdstuk 11].

Met behulp van deze *poolcoördinaten* volgt voor

$$\vec{c} = p\begin{pmatrix}\cos(s)\\ \sin(s)\end{pmatrix} \quad \text{en} \quad \vec{x} = r\begin{pmatrix}\cos(t)\\ \sin(t)\end{pmatrix}$$

dat

$$\vec{c}\cdot\vec{x} = pr(\cos(s)\cos(t) + \sin(s)\sin(t)) = pr\cos(s-t),$$

het product van de twee lengten en de cosinus van wat de *ingesloten hoek* wordt genoemd. Is die hoek gelijk aan $\pm\frac{\pi}{2}$ dan is het inproduct nul en staan de vectoren loodrecht op elkaar.

Opgave 13.25. De complexe afbeelding $z \to \frac{1}{z}$ laat zich in rechthoekige coördinaten x, y en in poolcoördinaten r en t bestuderen. Verifieer dat deze afbeelding de samenstelling is van $z \to \overline{z}$, een spiegeling in de x-as, en een andere afbeelding die spiegeling in de eenheidscirkel wordt genoemd, gegeven door $r \to \frac{1}{r}$. Construeer gegeven een punt binnen de cirkel zijn spiegelbeeld in de cirkel met behulp van een bij het gegeven punt geschikt gekozen raaklijn aan de cirkel.

Opgave 13.26. Merk op dat de uitkomst voor het *inwendig* product te vergelijken is met het gewone *complexe* product van de met de vectoren $\vec{c}$ en $\vec{x}$ corresponderende c en z. Verifieer dat voor

$$c = p\exp(is) \quad \text{en} \quad z = r\exp(it)$$

geldt dat

$$cz = p(\cos(s) + i\sin(s))r(\cos(t) + i\sin(t)) = pr(\cos(s+t) + i\sin(s+t)),$$

en bepaal het reële deel van $c\overline{z}$, waarin $\overline{z} = x - iy$ de complex geconjugeerde is van $z = x + iy$.

14 De Hilbertruimte in

Met $\mathbb{R}^3$, $\mathbb{R}^4$,.... kunnen we algebraïsch precies hetzelfde doen als wat we in Hoofdstuk 13 deden met $\mathbb{R}^2$. Typografisch wat lastig omdat we 2-vectoren gewoon met het LaTeχ-commando (backslash) *choose* voor binomiaalcoëfficiënten maakten en nu echt *array* nodig hebben, maar verder geen wezenlijk verschil. In $\mathbb{R}^3$ bijvoorbeeld met

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} = ax + by + cz \quad \text{of} \quad \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \sum_{i=1}^{3} a_i x_i,$$

in $\mathbb{R}^{42}$ met

$$\vec{a} \cdot \vec{x} = \sum_{i=1}^{42} a_i x_i \quad \text{voor} \quad \vec{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_{42} \end{pmatrix} \quad \text{en} \quad \vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{42} \end{pmatrix},$$

en tenslotte in $\mathbb{R}^\infty$, met

$$a \cdot x = \sum_{i=1}^{\infty} a_i x_i \quad \text{voor} \quad a = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \end{pmatrix} \quad \text{en} \quad x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \end{pmatrix},$$

waarbij we de pijltjes boven a en x in $\vec{a}$ en $\vec{x}$ maar hebben weggelaten. Punten of vectoren, we maken het onderscheid in de notatie tussen x als $\vec{x}$ en $\underline{x}$ steeds vaker alleen als het echt nodig is[1].

De laatste uitdrukking definieert $a \cdot x$ soms wel en soms niet, want zonder restricties op $a, x \in \mathbb{R}^\infty$ kan het met

$$a \cdot x = \sum_{i=1}^{\infty} a_i x_i = \sum_{i \in \mathbb{N}} a_i x_i$$

alle kanten op. En het wordt nog spannender als we de $i \in \mathbb{N}$ in vervangen door bijvoorbeeld $t \in \mathbb{R} = (-\infty, \infty)$ of[2] $t \in [-\pi, \pi]$. In zulke gevallen is ook $\sum$ aan vervanging toe. Overaftelbare[3] sommen gaan niet werken en sommeren moet hier dus wel integreren worden, wat anders? Met de notatie

[1] Als we niet meer recht kunnen praten wat krom is en rechte pijltjes niet passen.
[2] Denk ook aan de poolcoördinaten in het platte vlak.
[3] Waarom niet?

$t \to a_t = a(t)$ en $t \to x_t = x(t)$ voor functies $a : \mathbb{R} \to \mathbb{R}$ en $x : \mathbb{R} \to \mathbb{R}$ wordt een voor de hand liggend inwendig produkt van de functies a en x nu gedefinieerd met behulp van de formule

$$a \cdot x = \int_{-\infty}^{\infty} a(t)x(t)dt,$$

waarin *alle* $a(t)$ en $x(t)$ waarden gelijkwaardig voorkomen maar, paradoxaal wellicht, individueel geen invloed hebben op de uitkomst van de integraal die $a \cdot x$ definieert. Ook met die uitkomst kan het, bijvoorbeeld voor continue functies, alle kanten op, net als met $a \cdot x$ voor $a, x \in \mathbb{R}^{\infty}$.

Voor 2π-periodieke continue functies heeft deze integraalformule geen betekenis maar de formule

$$a \cdot x = \int_{-\pi}^{\pi} a(t)x(t)dt$$

vaak wel, het standaard inwendig produkt waarmee we werken in het geval van 2π-periodieke functies a en x, (goed) gedefinieerd voor continue functies als gewone Riemann integraal[4].

Opgave 14.1. Voor $n = 1, 2, 3, \ldots$ zijn de 2π-periodieke functies c_n en s_n gedefinieerd door $c_n(t) = \cos(nt)$ en $s_n(t) = \sin(nt)$. Bereken nog eens $c_n \cdot c_m$, $c_n \cdot s_m$, $s_n \cdot s_m$, voor $m, n = 1, 2, 3, \ldots$.

Je ziet het niet meteen, maar al deze cosinussen en sinussen staan "loodrecht" op elkaar, en ze hebben ook allemaal dezelfde "lengte", de wortel uit het inprodukt van de functie met zichzelf.

Opgave 14.2. Er is nog een functie die loodrecht staat op al deze cosinussen en sinussen. Welke functie?

14.1 Standaardassenkruizen

Tja[5], wat zijn dat? In het vlak waar we mee begonnen zijn wordt het assenkruis gevormd door 2 lijnen: de x-as door de oorsprong O en het punt $(1, 0)$ en de y-as door O en het punt $(0, 1)$, of wellicht liever de x_1-as en de x_2-as.

[4] En later via een subtiel proces voor nog veel meer functies.

[5] Een vraag voor de woensdagmiddag wellicht.

Een punt dat zich over zo'n as beweegt heeft een lange weg te gaan en kwam van ver. De x-as wordt geparametriseerd door $(x, y) = (t, 0)$, en de y-as door $(x, y) = (0, t)$, met bijbehorende snelheidsvectoren[6]

$$\vec{v}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{en} \quad \vec{v}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

die samen de standaardbasis van $\mathbb{R}^2$ als vectorruimte vormen.

Evenzo bestaat in $\mathbb{R}^3$ het standaardassenkruis uit 3 lijnen, de x- of x_1-as, de y- of x_2-as, en de z- of x_3-as, met bijbehorende vectoren[7]

$$\vec{v}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \vec{v}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \vec{v}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

die samen de standaardbasis van $\mathbb{R}^3$ genoemd worden. Drie vectoren met lengte 1 die onderling loodrecht staan.

En, we zouden het bijna vergeten, standaard of niet, een basis vormen ze. Iedere vector $\vec{v} \in \mathbb{R}^3$ is vanzelfsprekend uniek te schrijven als

$$\vec{v} = v_1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + v_2 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + v_3 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}.$$

Precies zoals in $\mathbb{R}^2$ waar iedere $\vec{v}$ van de vorm

$$\vec{v} = v_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + v_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$$

is, met een unieke correspondentie

$$\vec{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \quad \leftrightarrow \quad (v_1, v_2) = \underline{v},$$

waarin links en rechts v_1 en v_2 *hetzelfde* zijn.

De vectoren

$$\vec{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{en} \quad \vec{e}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

hebben lengte 1 en staan onderling loodrecht. In termen van het *standaard inwendig produkt*:

$$\vec{e}_1 \cdot \vec{e}_1 = \vec{e}_2 \cdot \vec{e}_2 = 1 \quad \text{en} \quad \vec{e}_1 \cdot \vec{e}_2 = \vec{e}_2 \cdot \vec{e}_1 = 0.$$

[6] In de wiskundeles meestal richtingsvectoren genoemd.

[7] Snelheidsvectoren, richtingsvectoren, het zijn maar woorden.

Met de gebruikelijke rekenregels volgt nu dat

$$\vec{v} \cdot \vec{e}_1 = (v_1\vec{e}_1 + v_2\vec{e}_2) \cdot \vec{e}_1 = v_1\vec{e}_1 \cdot \vec{e}_1 + v_2\vec{e}_2 \cdot \vec{e}_1 = v_1;$$

$$\vec{v} \cdot \vec{e}_2 = (v_1\vec{e}_1 + v_2\vec{e}_2) \cdot \vec{e}_2 = v_1\vec{e}_1 \cdot \vec{e}_2 + v_2\vec{e}_2 \cdot \vec{e}_2 = v_2,$$

en

$$\vec{v} = (\vec{v} \cdot \vec{e}_1)\vec{e}_1 + (\vec{v} \cdot \vec{e}_2)\vec{e}_2 = \sum_{i=1}^{2} (\vec{v} \cdot \vec{e}_i)\vec{e}_i$$

voor vectoren $\vec{v} \in \mathbb{R}^2$. In iedere $\mathbb{R}^n$ met n positief en geheel gaat het hetzelfde,

$$\vec{v} = \sum_{i=1}^{n} (\vec{v} \cdot \vec{e}_i)\vec{e}_i,$$

en pas in $\mathbb{R}^\infty$ wordt het wat lastiger.

14.2 Symmetrische matrices

Net als in de twee-dimensionale context heeft iedere symmetrische $n \times n$ matrix

$$A = (a_{ij})_{i,j=1,\ldots,n}$$

(een basis van) eigenvectoren $\vec{v}_1, \ldots, \vec{v}_n \in \mathbb{R}^n$ met

$$\vec{v}_i \cdot \vec{v}_j = \delta_{ij} = \begin{cases} 1 & \text{als } i = j \\ 0 & \text{als } i \neq j, \end{cases}$$

bij (reële) eigenwaarden

$$\lambda_1, \ldots, \lambda_n,$$

die gemaakt worden door Opgave 13.23 herhaald toe te passen. Dit geeft in ieder stap zowel een nieuwe λ_k als een nieuwe $\vec{v}_k$ via

$$|\lambda_k| = \max_{\substack{\vec{v} \cdot \vec{v} \leq 1 \\ \vec{v} \cdot \vec{v}_1 = \cdots = \vec{v} \cdot \vec{v}_{k-1} = 0}} |A\vec{v} \cdot \vec{v}|,$$

waarbij $k = 1, \ldots, n$.

Details van deze constructie komen aan de orde in de context van de standaard aftelbaar oneindig-dimensionale Hilbertruimte die we in $\mathbb{R}^\infty$ maken. *Daarvóór* komen we *voor het eerst* over abstracte Hilbertruimten[8] H te praten die we zo snel mogelijk gelijk[9] praten aan het standaardvoorbeeld

[8] Inprodukt ruimten waarin Cauchy rijtjes convergent zijn.

[9] Lees: isomorf.

in $\mathbb{R}^\infty$, onder de aanname van separabiliteit van H: het bestaan van een rij $x_1, x_2, x_3, \ldots$ in H die als limieten van zijn convergente deelrijen *alle* elementen in H heeft.

Kortom, in termen van de dimensie van onze ruimten maken we in één keer de stap van $n = 2$ en concreet ($\mathbb{R}^2$) naar $n = \infty$ en abstract (niet concreet). Let wel, dat kan *alleen* voor ruimten met een inwendig produkt.

14.3 Reële Hilbertruimten

Een reële Hilbertruimte H is een vectorruimte over $\mathbb{R}$ die naast de vectoroptelling en scalaire vemenigvuldiging ook een inwendig produkt

$$(x, y) \in H \times H \to (x, y)_H = x \cdot y$$

heeft, met de standaardrekenregels, en de eigenschap dat alle Cauchyrijtjes (dat zijn rijtjes waarvoor

$$(x_n - x_m) \cdot (x_n - x_m) \to 0$$

als $m, n \to \infty$) in H ook convergent zijn met limiet $\overline{x} \in H$ (i.e.

$$(x_n - \overline{x}) \cdot (x_n - \overline{x}) \to 0$$

als $n \to \infty$).

De norm wordt gegeven door $|x|_H^2 = (x, x)_H = x \cdot x$ en de onderlinge afstand van bijvoorbeeld x_n en x_m is

$$d_H(x_n, x_m) = |x_n - x_m|_H = \sqrt{(x_n - x_m) \cdot (x_n - x_m)},$$

waarin $d_H : H \times H \to \mathbb{R}^+ = [0, \infty)$ de *metriek* is op H. De subscripts H laten we voortaan weg, tenzij dat verwarring geeft.

Opgave 14.3. Formuleer en bewijs de ongelijkheid van Cauchy-Schwarz[10] (inclusief de karakterisatie van het geval van gelijkheid), bewijs de driehoeksongelijkheid, en formuleer en bewijs nog een keer Pythagoras en de parallellogramwet. Hint: overschrijven uit willekeurig Lineaire Algebra boek. Formuleer ook de axioma's voor metrische ruimten en bewijs deze voor d.

[10] De ongelijkheid in Opgave 13.9.

Opgave 14.4. Laat H een Hilbertruimte zijn, $K \subset H$ een gesloten convexe verzameling, en $a \in H$. Bewijs dat er een unieke $p \in K$ is die de afstand $d(a, K)$ van a tot K realiseert middels

$$|p - a| = \inf_{x \in K} |x - a| = d(a, K)$$

en laat zien dat $(p - a) \cdot (x - p) \geq 0$ voor alle $x \in K$. Hint: geef eerst de definities van gesloten, convex en afstand, en gebruik daarna de parallellogramwet, net zoals in Opgave 13.14. Bewijs ook dat de afbeelding $P_K : H \to K$ gedefinieerd door $P_K(a) = p$ de eigenschap heeft dat $|P_K(a) - P_K(b)| \leq |a - b|$ voor alle $a, b \in H$.

Opgave 14.5. Laat H een Hilbertruimte zijn, $L \subset H$ een gesloten lineaire deelruimte. Bewijs dat $P_L : H \to L$ lineair is en dat

$$M = N(P_L) = \{x \in H : P_L(x) = 0\} = L^{\perp} = \{x \in H : x \cdot y = 0 \ \forall\, y \in L\}^{11},$$

de *kern* of *nulruimte* van P_L, ook een gesloten lineaire deelruimte is met $M \cap L = \{0\}$. Laat zien dat $M + L = H$ en concludeer dat $L \oplus M = H$: iedere $x \in H$ is uniek te schrijven als $x = p + q$ met $p \in L$ en $q \in M$.

De uitspraak over het bestaan van p in Opgave 14.4 is natuurlijk equivalent met de uitspraak over het bestaan van het minimum van

$$(x - a) \cdot (x - a) = x \cdot x - 2a \cdot x + a \cdot a,$$

en daarmee dus equivalent met een uitspraak over minima op K van wat je parabolische functies zou kunnen noemen:

Opgave 14.6. Laat H een Hilbertruimte zijn, $K \subset H$ een gesloten convexe verzameling. Dan neemt voor iedere $b \in H$ de kwadratische uitdrukking[12]

$$|x|^2 + b \cdot x$$

op K in precies één punt een minimum[13] aan.

[11] $x \cdot y = 0$ voor alle y in L wordt kort geschreven als: $x \cdot y = 0 \ \forall\, y \in L$.

[12] Het kwadratische stuk kan algemener, het lineaire stuk niet!

[13] K is i.h.a. *niet* begrensd, laat staan rijkompakt (iets met convergente deelrijen).

Let op de eerste voeTnoot in Opgave 14.6. Het standaardinprodukt in $\mathbb{R}^2$ geeft via

$$\begin{pmatrix} a \\ b \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} = \underbrace{ax + cy = \begin{pmatrix} a & b \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}}_{matrix\ notatie}$$

een representatie van de lineaire functie

$$\begin{pmatrix} x \\ y \end{pmatrix} \to \underbrace{ax + cy = \begin{pmatrix} a & b \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}}_{matrix\ notatie}$$

en omgekeerd is *iedere* lineaire[14] functie van deze vorm. De correspondentie

$$\begin{pmatrix} a \\ b \end{pmatrix} \quad \leftrightarrow \quad \begin{pmatrix} a & b \end{pmatrix}$$

is evident bijectief en lineair. Links staat een 2-vector, rechts een 1 bij 2 matrix waarmee een lineaire afbeelding van $\mathbb{R}^2$ naar $\mathbb{R}$ gemaakt wordt.

In een willekeurige Hilbertruimte is er a priori geen matrixnotatie voor het maken van lineaire afbeeldingen. Welke lineaire afbeeldingen hebben we op zo'n Hilbertruimte H als van H verder niets gegeven is, behalve dan dat het een H is? Wel, in ieder geval is voor elke $y \in H$ de afbeelding $\phi_y : H \to \mathbb{R}$ gedefinieerd door[15]

$$x \to y \cdot x = \phi_y(x) = \phi_y x =< \phi_y, x > .$$

Kijk even goed, in de 1 na laatste notatie hebben we de haken weggelaten, zoals vaker bij lineaire afbeeldingen[16], en in de laatste staan ϕ_y en x zo te zien *gelijkwaardig* tussen strange brackets[17], waarbij net als in $y \cdot x$ de rollen van de tegenspelers verwisseld kunnen worden. Dualiteit heet dat met een mooi woord.

Voorlopig gebruiken we de notatie die het meest op de schoolnotatie lijkt. Een functie f van x, in dit geval ϕ_y, maak je expliciet[18] via $f(x)$, in dit geval $\phi_y(x) = y \cdot x$. Dat lijkt expliciet maar is het natuurlijk niet echt als we niet zeggen wat H is. Expliciet of niet, uit de ongelijkheid van Cauchy-Schwarz volgt nu dat

$$|\phi_y(x)| = |y \cdot x| \leq |y||x|$$

[14] De constante functie is NIET lineair, tenzij de constante nul is.
[15] Meteen maar met drie notaties.
[16] en bij $\cos, \sin, \tan, \ldots$.
[17] Tussen bra en ket, zoals fysici soms zeggen.
[18] Of niet, en dat veroorzaakt vaak veel verwarring.

en dus ook, *vanwege de lineairiteit* , dat

$$|\phi_y(x_1) - \phi_y(x_2)| = |y \cdot (x_1 - x_2)|| \leq L|x_1 - x_2| \quad \text{met} \quad L = |y|.$$

Een reëelwaardige functie f op een vectorruimte met een norm, die voldoet aan

$$|f(x_1) - f(x_2)| \leq L|x_1 - x_2|$$

voor alle x_1 en x_2 in die genormeerde vectorruimte, heet *Lipschitz continu*. Een mooi begrip, dat differentiaalrekening noch epsilons en delta's nodig heeft.

Opgave 14.7. Als zo'n (niet per se lineaire) functie een L heeft dan heeft hij ook een kleinste L. Bewijs dit. Hint: denk aan grootste ondergrenzen (infima).

Die kleinste L is dus voor alle Lipschitz continue functies op onze genormeerde ruimte (laten we die X noemen) gedefinieerd. Daar hoort een zijstapje bij:

Opgave 14.8. Voor elke genormeerde ruimte X vormen de Lipschitz continue functies $f : X \to \mathbb{R}$ een vectorruimte $Lip(X)$ met de vectorbewerkingen gedefinieerd door

$$(f + g)(x) = f(x) + g(x) \quad \text{en} \quad (tf)(x) = tf(x).$$

Voor elke f is de kleinste L zoals boven per definitie een soort norm van f, die we noteren met $L = [f]_{Lip}$. Waarom definieert

$$f \to [f]_{Lip}$$

geen norm op $Lip(X)$? En waarom wel op

$$Lip_0(X) = \{f \in Lip(X) : f(0) = 0\}?$$

Bewijs dat met deze norm elke Cauchyrij $f_n \in Lip_0(X)$ convergent is. Hint: bewijs dit eerst voor $X = \mathbb{R}$ en schrijf je bewijs nog een keer over voor $X = X$.

Klein probleempje is natuurlijk dat er misschien maar weinig van die al of niet lineaire Lipschitz functies op X zijn, als je verder niks van X weet. Maar op H is dat probleempje er niet. Elke $y \in H$ geeft je een ϕ_y in $Lip_0(H)$ die nog lineair is ook, en je ziet meteen wat de kleinste L is: op zijn hoogst $|y|$ en kleiner kan niet, vul maar $x = y$ in. Dat betekent dat we met $y \to \phi_y$ een afbeelding

$$\Phi := H \to Lip_0(H)$$

hebben, en het beeld van Φ is bevat in H^*, de (genormeerde) ruimte van Lipschitz continue *lineaire* functies $f : H \to \mathbb{R}$, en Φ is zelf weer lineair[19]:

Opgave 14.9. Verifieer dat $\Phi : H \to H^*$ voldoet aan

$$\Phi(x_1 + x_2) = \Phi(x_1) + \Phi(x_2) \quad \text{en} \quad \Phi(tx) = t\Phi(x)$$

voor alle $t \in \mathbb{R}$ en $x, x_1, x_2 \in H$, en dat $[\Phi(x)]_{Lip} = |x|$.

De vraag nu is of Φ surjectief is: is elke $f \in H^*$ van de vorm ϕ_y? Bekijk daartoe[20]

$$N_f = \{x \in H : f(x) = 0\}.$$

Opgave 14.10. Bewijs dat $N_f \subset H$ een gesloten lineaire deelruimte is.

In het bijzonder bestaat nu dankzij Opgave 14.5 de projectie

$$P_{N_f} : H \to N,$$

ook weer een lineaire afbeelding, en in de volgende opgave gaat het om de nulruimte van deze projectie op de nulruimte van f.

Opgave 14.11. Bewijs dat $M = N(P_{N_f})$ een gesloten lineaire deelruimte is die gegeven wordt door $M = \{te : t \in \mathbb{R}\}$ waarin $e \in N_f^\perp$ met $|e| = 1$. Laat zien dat f een veelvoud is van ϕ_e: $f(x) = f(e)e \cdot x$.

Opgave 14.12. Leg uit waarom met het resultaat in Opgave 14.10 de afbeelding $\Phi : H \to H^*$ een lineaire isometrie is.

Lineaire isometriën zijn de mooiste continue afbeeldingen die er bestaan. De inverse van Φ wordt de Riesz representatie van H^* genoemd, en via deze isometrie erft H^* ook het inwendig produkt van H: de reële Hilbertruimten H en H^* zijn als Hilbertruimten hetzelfde, al is het in concrete situaties niet altijd even handig om hier de nadruk op te leggen.

[19] Nu maar eens de axioma's noemen en verifiëren.

[20] We schrijven nu N_f i.p.v. $N(f)$, t.b.v. het onderscheid tussen f en P_L.

Het resultaat geldt zonder enige verdere restrictie op H en het is ook niet nodig om aan te nemen dat H separabel is. We noteren de inverse van Φ als

$$R_H,$$

met de ruimte H als subscript aan $R = \Phi^{-1}$ gehangen. Het domein van R_H is zo de deelruimte

$$H^* \subsetneq Lip_0(H).$$

Opgave 14.13. Gebruik Opgave 14.4 om aan te tonen dat er plenty niet-lineaire functies in $Lip_0(H)$ zijn.

14.4 De standaard Hilbertruimte

De wat informeel geintroduceerde ruimte $\mathbb{R}^\infty$ bestaat uit alle functies $f, x, a : \mathbb{N} \to \mathbb{R}$, hoe je ze ook wil noemen[21]. We kunnen deze functies zien als kolomvectoren $\vec{f}$ met daarin de waarden van f, al protesteert LaTeχ daarbij zo te zien een beetje. Helemaal op dezelfde hoogte lukt typografisch niet,

$$f = \begin{pmatrix} f(1) \\ f(2) \\ f(3) \\ \vdots \end{pmatrix} = \begin{pmatrix} f_1 \\ f_2 \\ f_3 \\ \vdots \end{pmatrix} = \vec{f},$$

en ook voor functies $f, x, a : \{1, 2, 3\} \to \mathbb{R}$ oogt

$$f = \begin{pmatrix} f(1) \\ f(2) \\ f(3) \end{pmatrix} = \begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix} = \vec{f}$$

niet echt lekker.

Wiskundig gezien praten we de facto over functies $f : A \to \mathbb{R}$, waarbij A hier een *discrete* verzameling is, en de verzameling van deze functies wordt ook wel genoteerd als $\mathbb{R}^A$. Als je $n \in \mathbb{N}$ gedefinieerd hebt als een wat rare verzameling, via[22]

$$1 = \{\emptyset\},\ 2 = \{\emptyset, \{\emptyset\}\},\ 3 = \{\emptyset, \{\emptyset, \{\emptyset\}\}\}, \ldots$$

[21] Het zijn er meer dan 26.

[22] Ik schuif wat ik naief in Halmos las wellicht eentje op, boekje niet bij de hand.

of zoiets, dan is de notatie voor $\mathbb{R}^A$ consistent[23] met die voor $\mathbb{R}^n$.

Elke $f \in \mathbb{R}^A$ heeft als Pythagoras norm

$$|f| = \sqrt{\sum_{a\in A} |f(a)|^2},$$

hetgeen voor $A = \{1, 2, 3\}$ overeenkomt met de Euclidische lengte

$$\sqrt{f_1^2 + f_2^2 + f_3^3}$$

van de vector $\vec{f}$ hierboven.

Het ligt voor de hand om $\mathbb{R}^A$ en $\mathbb{R}^B$ als dezelfde ruimte te zien als er een bijectie bestaat tussen A en B. Voor eindige verzamelingen A is de Pythagoras norm natuurlijk op heel $\mathbb{R}^A$ gedefinieerd, maar als A oneindig veel elementen bevat[24] dan is dat niet meer het geval.

Opgave 14.14. Stel dat A overaftelbaar is en $f \in \mathbb{R}^A$ eindige Pythagorasnorm heeft. Bewijs dat de verzameling

$$\{a \in A : \, f(a) \neq 0\}$$

aftelbaar is en ga na dat het in de somnotatie dan niet nodig is de volgorde van sommeren vast te leggen[25].

In het licht van deze opgave beperken we de aandacht voor oneindige A tot aftelbare A en die zijn allemaal bijectief met $\mathbb{N}$. We kunnen de functiewaarden van elementen $x, f, \ldots \in \mathbb{R}^{\mathbb{N}}$ dan op wat voor manier dan ook weer (niet allemaal) opschrijven, genummerd als $f(n)$ of x_n met $n = 1, 2, \ldots$, in bijvoorbeeld een kolomvector of rijvector met puntjes.

Onze standaard aftelbaar oneindig-dimensionale Hilbertruimte is nu

$$l^{(2)} = \{x = (x_1, x_2, \ldots) \in \mathbb{R}^{\mathbb{N}} : \, \sum_{n=1}^{\infty} x_n^2 < \infty\},$$

spreek uit: *(kleine) el twee*. Er is ook een *grote el twee* , namelijk de verzameling van kwadratisch integreerbare meetbare functies op een maatruimte,

[23] Let wel, $0 = \emptyset$ doet niet mee.

[24] We zeggen dan gemakshalve dat A oneindig is.

[25] Dit heet onvoorwaardelijke convergentie.

bijvoorbeeld[26] $\mathbb{R}$, voorzien van de gewone (Lebesque) lengtemaat[27]. Die *el twee* wordt genoteerd met

$$L^2(\mathbb{R}),$$

strict genomen geen functieruimte maar een ruimte van equivalentieklassen. We zeggen dat een (meetbare) functie f en een andere (meetbare) gunctie g equivalent zijn, notatie $f \sim g$, als de verzameling waarop ze verschillen (uitwendige) maat NUL heeft, en met f bedoelen we stiekem $[f]$, de equivalentieklasse van alle g waarvoor $g \sim f$.

De inwendige produkten zijn, respectievelijk,

$$x \cdot y = (x, y)_{l^{(2)}} = \sum_{n=1}^{\infty} x_n y_n \quad \text{en} \quad f \cdot g = (f, g)_{L^2(\mathbb{R})} = \int_{\infty}^{\infty} f(x)g(x)dx,$$

waarbij de integraalnotatie bij (niet ieders) voorkeur hetzelfde gekozen wordt als die van de Riemann integraal.

Opgave 14.15. Bewijs dat $l^{(2)}$ *volledig* is. Dat wil zeggen, laat zien dat Cauchy rijtjes in $l^{(2)}$ convergent zijn met limiet in $l^{(2)}$.

Als we $\mathbb{N}$ zien als maatruimte voorzien van de telmaat dan wordt kleine l weer groot. En met recht, want iedere separabele Hilbertruimte H is met $l^{(2)}$ te identificeren[28]. Hoe gaat dat? Wel, neem een rijtje $a_1, a_2, a_2, \ldots$ in H dat als limietpunten alle elementen van H heeft. Zet

$$e_1 = \frac{1}{|a_1|} a_1$$

als $a_1 \neq 0$ maar gooi a_1 weg als $a_1 = 0$. Hernummer in dat geval de rij en herhaal deze stap, net zolang[29] tot je een $a_1 \neq 0$ hebt. Stel vervolgens

$$y_2 = a_2 - (a_2, e_1)e_1 \quad \text{en} \quad e_2 = \frac{1}{|y_2|} y_2$$

als $y_2 \neq 0$, maar gooi a_2 weg als $y_2 = 0$ en hernummer in dat geval weer de rij. Herhaal deze stap, net zolang tot je een $y_2 \neq 0$ hebt en daarmee ook een e_2. Stel vervolgens

$$y_3 = a_3 - (a_3, e_2)e_2 - (a_3, e_1)e_1 \quad \text{en} \quad e_3 = \frac{1}{|y_3|} y_3,$$

[26] Ander voorbeeld: $\mathbb{R}$ modulo 2π, de facto de eenheidscirkel in $\mathbb{R}^2$.
[27] Zie "Wiskunde in je vingers" van H&M voor snelle intro maattheorie.
[28] Indien gewenst.
[29] Nou ja, als er geen dubbelen in de rij voorkomen dan...

als $y_3 \neq 0$, maar ..., enzovoorts. Dit produceert een rij $e_1, e_2, e_3, \ldots$ van vectoren waarvoor

$$(e_i, e_j) = \delta_{ij},$$

en deze vectoren spannen een lineaire deelruimte op in H.

Opgave 14.16. Bewijs dat

$$H = \{x = \sum_{n=1}^{\infty} x_n e_n : \sum_{n=1}^{\infty} x_n^2 < \infty\},$$

waarmee H dus met de standaard Hilbertruimte $l^{(2)}$ geidentificeerd *kan* worden.

14.5 Riesz of geen Riesz?

In alle Hilbertruimten is de Riesz representatiestelling van toepassing, en dus in het bijzonder ook in $l^{(2)}$.

Opgave 14.17. Geef een rechtstreeks bewijs voor de representatiestelling van Riesz voor het geval dat $H = l^{(2)}$.

Onze standaard Hilbertruimte is $H = l^{(2)} = L^2(\mathbb{N})$ met op $\mathbb{N}$ de telmaat, maar er zijn ook andere maten op $\mathbb{N}$. Ieder rijtje positieve getallen[30]

$$0 < \lambda_1 \leq \lambda_2 \leq \lambda_2 \leq \cdots$$

(hernummerd om het rijtje monotoon te maken) definieert een maat op $\mathbb{N}$ door aan het n-de element van $\mathbb{N}$ maat λ_n toe te kennen. De bijbehorende integraal van het produkt van twee functies $x, y : \mathbb{N} \to \mathbb{R}$ is dan

$$((x, y)) = (x, y)_V = \sum_{n=1}^{\infty} \lambda_n x_n y_n,$$

een inwendig produkt op een deelruimte V van onze standaard $H = l^{(2)}$, die niet gesloten is in H in het geval dat $\lambda_n \to \infty$.

Opgave 14.18. Waarom niet? Laat zien dat V met $((\cdot, \cdot))$ een Hilbertruimte is, en dat $V = H$ dan en slechts dan als λ_n een *begrensd* monotoon rijtje positieve getallen is.

[30] De letter λ wordt doorgaans gebruikt voor eigenwaarden.

Kortom $V \subset H$, en de norm op V wordt gegeven door

$$|x|_V = ||x|| = \sqrt{\sum_{n=1}^{\infty} \lambda_n x_n^2},$$

waarmee

$$||x||^2 \geq \lambda_1 |x|^2$$

voor alle $x \in V$. Hierin is

$$|x| = \sqrt{\sum_{n=1}^{\infty} x_n^2}$$

de gewone norm van x, een x die immers ook in H zit. De afbeelding[31]

$$i : x \in V \to x \in H$$

is lineair en continu. Er geldt immers voor alle $x \in V \subset H$ dat

$$\underbrace{||i(x)||}_{x \in V} = \underbrace{|x|}_{x \in H} \leq \frac{1}{\sqrt{\lambda_1}} \underbrace{||x||}_{x \in V},$$

en

$$||i|| = \frac{1}{\sqrt{\lambda_1}}$$

is de norm van i: de kleinste constante C waarmee geldt dat

$$|i(x)| \leq C||x||$$

voor alle $x \in V$. De notatie met i maakt onderscheid tussen x als $x \in V$ en x als $x \in H$. De afbeelding $i : V \to H$ stuurt $x \in V$ naar zichzelf, maar dan opgevat als $x \in H$, en heeft daarom een paar vanzelfsprekende eigenschappen.

Opgave 14.19. Ga na dat deze i injectief is en dat als $i(V) = V$ niet gelijk is aan H wel geldt dat $\overline{i(V)} = \overline{V} = H$.

Als $\overline{V} = H$ dan zeggen we dat V *dicht* ligt in H. Dat is hier dus altijd het geval, trivialerwijs ook als $V = H$. Met de Riesz representatiestelling (Opgave 14.12) is iedere continue lineaire functie $F : H \to \mathbb{R}$ van de vorm

$$F(x) = (f, x)$$

[31] i is hier niet die wortel uit -1 maar de inbedding van V in H.

met $f = R_H(F)$, en uiteraard is $F(x) = (f, x)$ ook gedefinieerd voor $x \in V$. De afbeelding

$$F \circ i : x \in V \xrightarrow{i} x \in H \xrightarrow{F} (f, x) \in \mathbb{R}$$

is zo continu en lineair, en dus vanwege de Riesz representatiestelling, toegepast nu op V, te representeren met een $u = R_V(F \circ i) \in V$, waarvoor nu geldt dat

$$F(x) = (f, x) = ((u, x)) \quad \forall x \in V.$$

Met andere woorden, de afbeelding

$$V \ni x \xrightarrow{f \in H} (f, x)_H$$

is precies dezelfde afbeelding als

$$V \ni x \xrightarrow{u \in V} (u, x)_V$$

(maar anders gerepresenteerd).

Opgave 14.20. Neem aan dat $0 < \lambda_1 \leq \lambda_2, \ldots$ een onbegrensd rij is. Waarom is niet elke continue lineaire $G : V \to \mathbb{R}$ van de vorm $F(x) = (f, x)$ met $f \in H$?

Het is voor een lineaire functie op V dan[32] dus makkelijker om continu te zijn met betrekking tot de norm op V dan met betrekking tot de norm op H: er zijn dus *meer* continue lineaire functies op V dan de functies

$$x \in V \to (y, x)_H \in \mathbb{R}.$$

Als we er voor kiezen om H^* met H te *identificeren* via R_H, dan is

$$V \subsetneq H = H^* \subsetneq V^*,$$

hetgeen dan niet meer spoort met een identificatie van V^* en V via R_V. Het is één van de twee (of geen van beide), maar niet allebei.

Hoe het ook zij, de afbeelding

$$B : f \in H \xrightarrow{R_H^{-1}} \underbrace{F \in H^* \xrightarrow{i^*} F \circ i \in V^*}_{i^*(F) = F \circ i} \xrightarrow{R_V} u \in V,$$

is lineair en continu, omdat de eerste en de derde schakel in deze keten allebei een isometrie zijn, en de middelste schakel i^*, die de geadjungeerde van i heet, continu is.

[32] Als $\lambda_n \to \infty$.

Opgave 14.21. Voor lineaire afbeeldingen tussen genormeerde ruimten is continuiteit equivalent met wat begrensdheid wordt genoemd. Een lineaire afbeelding $T : X \to Y$ heet begrensd als er een constante C is waarvoor geldt dat[33]

$$||Tx||_Y \leq C||x||_X$$

voor alle $x \in X$. Bewijs dit, en ook dat de kleinste C waarvoor dit geldt bestaat. Die kleinste C is per definitie de norm $||T||$ van T, en komt voor lineaire afbeeldingen overeen met wat we eerder $[\ ]_{Lip}$ genoemd hebben bij reëlewaardige functies, zie Opgave 14.8.

In het bijzonder is een begrensde lineaire T niet begrensd op X maar wel op

$$B_X = \{x \in X : \ ||x||_X \leq 1\},$$

de gesloten eenheidsbol op X.

Opgave 14.22. Bewijs dat $i^* : H^* \to V^*$ lineair en continu (begrensd) is.

Interessant is dat als V en H Hilbertruimten zijn en $i : V \to H$ een *injectieve* continue lineaire afbeelding is met $i(V) \subsetneq \overline{i(V)} = H$, het verhaal vaak precies hetzelfde is. Het is leerzaam om te kijken hoe de injectiviteit van i en het dicht liggen van het beeld van i dan een rol spelen.

Opgave 14.23. Neem nu aan dat H en V Hilbertruimten zijn en dat $i : V \to H$ lineair en begrensd is. Bewijs dat $B : H \to V$ gedefinieerd door $f \in H \to u = Bf \in V$ en

$$(u, x)_V = (f, i(x))_H$$

voor alle $x \in H$ goed gedefinieerd is als lineaire begrensde afbeelding

$$B = R_V \circ i^* \circ (R_H)^{-1},$$

en laat zien dat

$$||B|| = ||i^*|| = ||i||.$$

Hint: de eerste gelijkheid volgt omdat de eerste en de derde schakel in B isometrisch zijn, het gaat dus om de tweede gelijkheid. Dat $||i^*|| \leq ||i||$ volgt rechstreeks uit de definities. Om $||i^*|| \geq ||i||$ te laten zien moet je de definitie van $||i||$ gebruiken door een rij $x_n \in V$ te nemen met $||x_n||_V = 1$ en $||i(x_n)|| \to ||i||$, en bij die $y_n = i(x_n)$ elementen F_n uit H^* met $||F_n||_{H^*} = 1$ and $F_n(y_n) = ||y_n||_H$.

[33] Met notatie $T(x) = Tx$.

Opgave 14.24. Bewijs ook dat B injectief is als de afsluiting $\overline{i(V)}$ van $i(V)$ in H gelijk is aan H, dus als $\overline{i(V)} = H$. Hint: het gaat dus om de tweede equivalentie in B injectief $\iff$ i^* injectief $\iff$ $\overline{i(V)} = H$. De eerste is vanzelfsprekend om dat de andere twee factoren in B isometrisch zijn.

Opgave 14.25. Neem alleen aan dat H en V Hilbertruimten zijn en dat $i : V \to H$ lineair en begrensd is. Laat $B : H \to V$ gedefinieerd zijn via Opgave 14.23 door $f \in H \to u = Bf \in V$ en

$$(u, x)_V = (f, i(x))_H.$$

Laat zien dat

$$N(i) = \{x \in V : i(x) = 0\} = (B(H))^{\perp}.$$

In het bijzonder ligt het beeld van B dicht in V precies dan als i injectief is.

Opgave 14.26. Neem aan dat $i : V \to H$ lineair en begrensd is. Bewijs dat

$$A_0 = i \circ B : H \to H$$

symmetrisch is, i.e.

$$(A_0 f_1, f_2)_H = (f_1, A_0 f_2)_H$$

voor alle $f_1, f_2 \in H$, en

$$(A_0 f, f)_H \geq 0$$

voor alle $f \in H$. Welke conditie op $i : V \to H$ garandeert

$$(A_0 f, f)_H > 0 \quad \text{voor} \quad f \neq 0 \in H?$$

Hint: voor het antwoord op de vraag neem $x = u = Bf$ in de definitie van B in Opgave 14.23 en gebruik Opgave 14.24.

Opgave 14.27. Neem aan dat $i : V \to H$ lineair en begrensd is. Bewijs dat

$$A_1 = B \circ i : V \to V$$

symmetrisch is, i.e.

$$(A_1 u_1, u_2)_V = (u_1, A_1 u_2)_V$$

voor alle $u_1, u_2 \in V$, en

$$(A_1 v, v)_V \geq 0$$

voor alle $v \in H$. Welke conditie op $i : V \to H$ garandeert

$$(A_1 v, v)_V > 0 \quad \text{voor} \quad v \neq 0 \in V?$$

Hint: voor het antwoord op de vraag neem $f = i(v)$ en $u = Bf$ in Opgave 14.23.

Opgave 14.28. Laat zien dat

$$\frac{(A_0 f, f)_H}{(f, f)_H} = \frac{(Bf, Bf)_H}{(f, f)_H} \quad \text{en} \quad \frac{(A_1 v, v)_V}{(v, v)_V} = \frac{(i(v), i(v))_H}{(v, v)_V}$$

en concludeer dat beide quotiënten *dezelfde* suprema hebben. Hint: de suprema van de rechterleden zijn normen.

14.6 Convergente deelrijen

In het concrete voorbeeld dat $V \subset H = l^{(2)}$ gedefinieerd is via een (onbegrensde) rij

$$0 < \lambda_1 \leq \lambda_2 \leq \cdots$$

kunnen we subscripts wel weglaten en zien we A zowel als $A : V \to V$ en $A : H \to H$. Vanwege de aanname dat

$$\lambda_n \to \infty$$

heeft $i : V \to H$ nu een bijzonder eigenschap.

Opgave 14.29. Als $x_1, x_2, x_3, \ldots$ een begrensde rij in V is dan heeft $x_1, x_2, x_3, \ldots$ een convergente deelrij rij in H.

Met ander woorden, als $|x_n|_V = ||x_n||$ begrensd is dan is er een $x \in H$ zodanig dat

$$|x_{n_k} - x| \to 0$$

als $k \to \infty$, voor tenminste één rij indices

$$n_1 < n_2 < n_3 < \cdots$$

Deze conclusie kan *niet* getrokken worden voor een rij die alleen begrensd is in de H-norm. Immers, de basisvectoren $e_1, e_2, \ldots$ staan onderling loodrecht en hebben lengte 1. Bijgevolg is (voor $i \neq j$)

$$|e_i - e_j|^2 = (e_i - e_j, e_i - e_j) = (e_i, e_i) + (e_j, e_j) = 2$$

en heeft de rij $e_1, e_2, \ldots$ geen enkele convergente deelrij. Wel is het zo dat voor elke $y \in H$ geldt dat $(y, e_i) \to 0 = (y, 0)$. De rij $e_1, e_2, \ldots$ heet daarom *zwak* convergent met limiet $0 \in H$.

Opgave 14.30. Laat $x_1, x_2, \ldots$ een begrensde rij zijn in de standaard Hilbertruimte $H = l^{(2)}$. Dan heeft $x_1, x_2, \ldots$ *wel een zwak* convergente deelrij, i.e. er is een rij indices $n_1 < n_2 < \cdots$ en een $a \in H$ zodanig dat

$$(y, x_{n_k}) \to (y, a)$$

als $k \to \infty$ voor elke $y \in H$. Notatie:

$$x_{n_k} \rightharpoonup a \quad \text{als} \quad k \to \infty.$$

Hint: in $\mathbb{R}$ heeft iedere begrensde rij een convergente deelrij. Gebruik een *diagonaal-argument* om te laten zien dat er een rij indices $n_1 < n_2 < \cdots$ is waarvoor

$$a_j = \lim_{k \to \infty} (e_j, x_{n_k})$$

als limiet in $\mathbb{R}$ bestaat voor iedere $j \in \mathbb{N}$, en laat zien dat

$$a = \sum_{j=1}^{\infty} a_j e_j \in H$$

de zwakke limiet is van de rij $x_{n_1}, x_{n_2}, \ldots$.

Opgave 14.31. Bewijs dat in iedere separabele Hilbertruimte elke begrensde rij een zwak convergente deelrij heeft. Hint: gebruik een basis van H om H isometrisch met $\mathbb{R}^n$ of $l^{(2)}$ te maken en Opgave 14.30.

Opgave 14.32. Bewijs dat in iedere Hilbertruimte, separabel of niet, elke begrensde rij een zwak convergente deelrij heeft. Hint: laat L de kleinste gesloten deelruimte van H zijn die de hele rij bevat, schrijf $H = L \oplus L^{\perp}$ en gebruik het resultaat in Opgave 14.31 voor L.

14.7 Positieve eigenwaarden van wat?

De naam van een stelling past niet altijd in een kopje. Voor welke positief definiete symmetrische lineaire afbeeldingen $A : H \to H$ hebben we een uitspraak over positieve eigenwaarden met eigenvectoren die een orthonormale

basis vormen van H? De extra eigenschap die we nodig hebben is *kompaktheid*, en dat begrip formuleren we *eerst in termen van rijen*. Een verzameling K in een ruimte voorzien van een metriek heet *rijkompakt* als iedere rij in K door elementen uit de rij weg te laten[34] convergent gemaakt kan worden met limiet in K.

In de vorige subsectie hebben we gezien dat de gesloten eenheidsbol

$$B \subset H = l^{(2)}$$

niet rijkompakt is. Immers, de rij $e_1, e_2, \ldots$ heeft geen convergente deelrij. Terwijl de rij wel *zwak convergent* is met limiet 0. Maar onder de aanname dat de rij

$$0 < \lambda_1 \leq \lambda_2 \leq \cdots$$

de eigenschap heeft dat $\lambda_n \to \infty$ liet Opgave 14.29 zien dat voor de door deze rij gedefinieerde V in Opgave 14.18 geldt dat de *lineaire afbeelding*

$$i : V \to H = l^{(2)}$$

kompakt is in de volgende zin: als $x_1, x_2, x_3, \ldots$ een begrensde rij is in V dan heeft de rij $i(x_1), i(x_2), i(x_3), \ldots$ een convergente deelrij in H.

Opgave 14.33. Neem aan dat $i : V \to H$ zoals in Opgave 14.23 kompakt is. Laat zien dat $A_0 = i \circ B : H \to H$ en $A_1 = B \circ i : V \to V$ ook kompakt zijn: iedere begrensde rij originelen wordt na weglaten van voldoende veel rijelementen afgebeeld op een convergente rij.

Elke lineaire, injectieve, begrensde[35] en kompakte

$$i : V \to H \quad \text{met} \quad \overline{i(V)} = H$$

definieert nu dus via Opgaven 14.23, 14.26 en 14.27 twee[36] strict positief definiete symmetrische kompakte begrensde lineaire afbeeldingen $A_0 : H \to H$ en $A_1 : V \to V$, door respectievelijk de eerste en de laatste schakel in

$$V \xrightarrow{i} H \xrightarrow{(R_H)^{-1}} H^* \xrightarrow{i^*} V^* \xrightarrow{R_V} V \xrightarrow{i} H$$

weg te laten.

[34] Think of cheating experimentalists.
[35] Dit eisen is niet meer nodig, volgt uit kompaktheid i, ga na!
[36] Het past ook niet op 1 regel.

Opgave 14.34. Laat H een Hilbertruimte zijn en $A : H \to H$ lineair, kompakt, symmetrisch en positief strict definiet. Bewijs dat

$$\mu_1 = \sup_{0 \neq x \in H} \frac{(Ax, x)}{(x, x)}$$

als *strict positief maximum* bestaat, en dat iedere het quotiënt maximaliserende x een eigenvector is van A bij deze (eigenwaarde) $\mu_1 > 0$. Hint: μ_1 wordt benaderd door (Ax_n, x_n) met $(x_n, x_n) = 1$ waarbij de rij x_n zwak convergent met limiet x kan worden gekozen, en, vanwege de kompaktheid van A, de rij Ax_n convergent met limiet y.

Er is een eleganter bewijs van de stelling in Opgave 14.34 waarbij het begrip zwakke convergentie niet wordt gebruikt. Als $A : H \to H$ een symmetrische continue lineaire afbeelding is, dan is $|(Ax, x)| \leq ||Ax||\,||x|| \leq ||A||\,||x||^2$ en bestaan dus

$$M = \sup_{0 \neq x \in H} \frac{|(Ax, x)|}{(x, x)}, \quad M_+ = \sup_{0 \neq x \in H} \frac{(Ax, x)}{(x, x)}, \quad M_= \inf_{0 \neq x \in H} \frac{(Ax, x)}{(x, x)},$$

en M, het maximum van $|M_+|$ en $|M_-|$, voldoet dus aan $M \leq ||A||$. Via de parallelogramwettruc volgt ook de omgekeerde ongelijkheid. Voor $x, y \in H$ geldt dat

$$(Ax, y) = \frac{1}{4}(A(x+y), x+y) - \frac{1}{4}(A(x-y), x-y) \leq \frac{1}{4}(M_+||x+y||^2 - M_-||x-y||^2)$$

en dus dat

$$|A(x, y)| \leq \frac{M}{4}(||x + y||^2 + ||x - y||^2) = \frac{M}{2}(||x||^2 + ||y||^2).$$

Opgave 14.35. Neem aan dat $A : H \to H$ lineair is en dat voor alle $x, y \in H$ geldt dat

$$|(Ax, y)| \leq \frac{M}{2}(||x||^2 + ||y||^2).$$

Bewijs dat dan ook voor alle $x, y \in H$ geldt dat

$$|(Ax, y)| \leq M||x||\,||y||,$$

en dat dus $M = ||A||$. Hint: pas voor $a > 0$ de ongelijkheid toe op $\frac{x}{a}$ en ay en minimaliseer het rechterlid.

Opgave 14.36. Neem aan dat $A : H \to H$ een symmetrische continue lineaire afbeelding is en dat er een rij $x_n \in H$ is met $||x_n|| = 1$ en $(Ax_n, x_n) \to ||A|| = M$ waarvoor de rij beelden Ax_n convergent is. Bewijs dat de limiet een eigenvector is bij eigenwaarde M. Hint: laat zien dat $Ax_n - Mx_n \to 0$.

Opgave 14.37. De spectraalstelling voor $A : H \to H$ kompakt, strict positief definiet, symmetrisch en lineair: bewijs dat H een orthonormale basis heeft van eigenvectoren

$$\phi_1, \phi_2, \ldots$$

van A bij eigenwaarden *(uniquely) identified as*

$$\mu_n = \max\{(Ax, x) : x \in H, (x, x) = 1, (x, \phi_j) = 0 \text{ voor } 1 \leq j < n\},$$

voor $n = 1, 2 \ldots$. Dit zijn ze! Hoe je er ook verder aankomt.

Opgave 14.38. Kun je een voorbeeld bedenken van de stelling in Opgave 14.37 met een niet-separabele Hilbertruimte? Hint: zoja, heel knap? Zonee, heel goed! Waarom niet?

Opgave 14.39. Kun je een voorbeeld bedenken van de stelling in Opgave 14.37 met een separabele Hilbertruimte? Hint: hoe zijn we hierboven eigenlijk begonnen met λ's?

Wellicht wat cryptisch, maar dat scherpt de geest. Het springende punt is dat we in $\mathbb{R}^\infty$ zijn begonnen op basis van de voor de hand liggende standaardbasis, die de prachtige eigenschap heeft dat

$$e_i \cdot e_j = (e_i, e_j) = \delta_{ij}$$

en vervolgens een ander inprodukt hebben geïntroduceerd aan de hand van dezelfde standaardbasis en een rijtje λ's. De twee operatoren A_0 en A_1 hebben (*simultaan*) precies de e-tjes van de standaardbasis als eigenvectoren bij wat we μ-tjes genoemd hebben.

Opgave 14.40. Voor de μ-tjes en de λ-tjes geldt

$$\lambda_n \mu_n = 1,$$

en dat moet je wel even begrijpen nu!

Zijn er minder flauwe voorbeelden? Ja! Begin maar met een $i : V \to H$ met een concrete $V \subset H$ met $\overline{V} = H$ die niet al van te voren in $\mathbb{R}^\infty$ lagen (maar uiteindelijk via hun coordinaten weer wel in $\mathbb{R}^\infty$ te leggen zijn als je de basis van eigenvectoren eenmaal hebt). Nodig is de restrictie tot $V \subset H$ zeker niet. In alle gevallen zijn er twee symmetrische afbeeldingen waarvan het spectrum kan worden bepaald, $A_0 : H \to H$ en $A_1 : V \to V$, of we nu de extra aannamen[37] op i hebben gemaakt of niet. Net als tussen de normen (zie Opgave 14.28) is er een verband tussen de eigenwaarden en eigenvectoren.

Opgave 14.41. Als f een eigenvector is van A_0 bij eigenwaarde λ wat is dan een eigenvector van A_1 bij dezelfde λ? Als u een eigenvector is van A_1 bij eigenwaarde λ wat is dan een eigenvector van A_0 bij dezelfde λ? Hint: bekijk een langer stuk van de keten

$$V \xrightarrow{i} H \xrightarrow{(R_H)^{-1}} H^* \xrightarrow{i^*} V^* \xrightarrow{R_V} V \xrightarrow{i} H \xrightarrow{(R_H)^{-1}} H^* \xrightarrow{i^*} V^* \xrightarrow{R_V} V \xrightarrow{i} H \xrightarrow{(R_H)^{-1}}$$

en neem aan dat een $f \in H$ of een $u \in V$ op een λ-veelvoud van zichzelf terecht komt na 4 stapjes in de keten. In beide gevallen zou de eigenvector die je vindt bij de andere A de nulvector kunnen zijn. Welke conditie op i voorkomt dat?

Bij het toepassen van Opgave 14.37 kunnen we dus kiezen voor A_0 of A_1.

14.8 The French Connection

Het tripel

$$V \subset H = H^* \subset V^*$$

met V en H Hilbertruimten, en $i : V \to H$ kompakt en $V = i(V)$ dicht in H is het standaardframework dat in de Franse school voor partiële differentiaalvergelijkingen is ontwikkeld, zie bijvoorbeeld het originele franstalige boekje van Brézis over functionaalanalyse. Een grote klasse van concrete eigenwaardenproblemen[38] reduceert in abstracto tot een en hetzelfde probleem.

De eerste toepassing is die waarin cosinussen sinussen verschijnen als eigenfuncties van de oplossingsoperator voor het probleem

$$-u'' = f,$$

waarin we voor f en u nu eerst 2π-periodieke functies nemen. In het volgende hoofdstuk beperken we ons daarbij aanvankelijk tot functies die gemiddeld

[37] Injectiviteit van i en dichtheid van (het beeld van) i.

[38] Denk aan de frequenties van een trillend (muziek)instrument.

nul zijn, en zijn

$$(f,g) = \int_{-\pi}^{\pi} fg \quad \text{en} \quad ((u,v)) = \int_{-\pi}^{\pi} u'v'$$

de twee inprodukten die de rol van (f,g) en overnemen $((u,v))$ in de concrete voorbeelden hierboven, en zullen we die concrete voorbeelden weer terug zien komen uit de abstractie.

15 Lebesgue's Differentiatie Stelling

Wat we nodig hebben is dat meetbare lokaal integreerbare functies $f : \mathbb{R} \to \mathbb{R}$ voor bijna alle x voldoen aan

$$f(x) = \lim_{\epsilon \to 0} \frac{1}{2\epsilon} \int_{x-\epsilon}^{x+\epsilon} f(s)\, ds,$$

en die uitspraak bekijken we nu eerst voor meetbare functies $f : \mathbb{R}^N \to \mathbb{R}$ waarvoor

$$\int_{\mathbb{R}^N} |f| < \infty.$$

Zulke functies vormen de ruimte $L^1(\mathbb{R}^N)$, met norm

$$||f||_1 = \int_{\mathbb{R}^N} |f|.$$

Omdat $||f||_1 = 0$ niet impliceert dat $f(x) = 0$ voor alle $x \in \mathbb{R}^N$ maar wel dat de maat van

$$\{x \in \mathbb{R}^N : f(x) \neq 0\}$$

gelijk is aan nul, moeten we $f \in L^1(\mathbb{R}^N)$ zien als een equivalentieklasse $[f]$ van integreerbare meetbare functies $f : \mathbb{R}^N \to \mathbb{R}$ onder de relatie

$$f \sim g \quad \text{als} \quad |\{x \in \mathbb{R}^N : f(x) \neq g(x)\}| = 0.$$

We gebruiken hier $|A|$ als notatie voor de maat van een meetbare verzameling $A \subset \mathbb{R}^N$.

Voor iedere $f \in L^1(\mathbb{R}^N)$ en iedere bol

$$B(\xi, r) = \{x \in \mathbb{R}^N : (x_1 - \xi_1)^2 + \cdots + (x_n - \xi_n)^2 < r^2\}$$

bestaat

$$(A_r f)(x) = \frac{1}{|B(x,r)|} \int_{B(x,r)} f$$

onafhankelijk van de keuze van $f \in [f]$ en de verzameling

$$\mathcal{N}_f = \{x \in \mathbb{R}^N : \lim_{r \downarrow 0} (A_r f)(x) \text{ bestaat niet}\}$$

dus ook. Als de limiet wel bestaat dan ligt het voor de hand om de representant $f \in [f]$ zo te kiezen dat de limiet gelijk is aan $f(x)$ voor alle $x \notin \mathcal{N}_f$, en dat voor alle zulke x, maar daarvoor moeten de x-waarden waarvoor de limiet wel bestaat maar

$$\lim_{r \downarrow 0} (A_r f)(x) \neq f(x),$$

wel een meetbare verzameling van maat 0 vormen.

Een alternatieve manier is om $f \in L^1(\mathbb{R}^N)$ als equivalentklasse van Cauchyrijen van continue functies f_n die elk buiten een bol B_n identiek nul zijn. Van zo'n rij functies eisen we dan dat

$$||f_n - f_m|| = \int_{\mathbb{R}^N} |f_n - f_m| \to 0$$

als $m, n \to \infty$, waarbij de integralen de facto gewone Riemann integralen over bollen $B_{m,n}$ zijn waarbuiten f_n en f_m nul zijn. Twee zulke Cauchyrijen heten equivalent als $||f_n - g_n||_1 \to 0$ als $n \to \infty$, en

$$(A_r f)(x) = \lim_{n\to\infty} \frac{1}{|B(x,r)|} \int_{B(x,r)} f_n$$

is weer ondubbelzinning gedefinieerd, en daarmee ook $\mathcal{N}_f$.

We bekijken nu hoe $A_r f(x)$ zich gedraagt als $r \to 0$ en dat doen we eerst voor een gegeven integreerbare meetbare functie $f : \mathbb{R}^N \to \mathbb{R}$ met behulp van een te kiezen functie $g : \mathbb{R}^N \to \mathbb{R}$ die continu is en nul buiten een bol. We schrijven

$$A_r f(x) - f(x) = A_r f(x) - A_r g(x) + \underbrace{A_r g(x) - g(x)}_{\to 0 \text{ als } r\downarrow 0} + g(x) - f(x),$$

$$\limsup_{r\downarrow 0} |A_r f(x) - f(x)| \leq \underbrace{\limsup_{r\downarrow 0} |A_r f(x) - A_r g(x)|}_{\leq (H(f-g))(x)} + |g(x) - f(x)|, \quad (15.1)$$

waarin we Hardy-Littlewood functie (van $f - g$) gebruiken, die voor iedere $f \in L^1(\mathbb{R}^N)$ wordt gedefinieerd als

$$(Hf)(x) = \sup_{r>0}(A_r|f|)(x) = \sup_{r>0} \frac{1}{|B(x,r)|} \int_{B(x,r)} |f| \in [0, \infty],$$

met andere woorden, $(H(f-g))(x)$ is de grootst mogelijke gemiddelde waarde van $|f(x) - g(x)|$ op een bollen gecentreerd in x. Voor elke $\epsilon > 0$ geldt nu dat

$$\{x \in \mathbb{R}^N : \limsup_{r\downarrow 0} |A_r f(x) - f(x)| > 2\epsilon\} \subset$$

$$\{x \in \mathbb{R}^N : H(f-g))(x) > \epsilon\} \cup \{x \in \mathbb{R}^N : |f(x) - g(x)| > \epsilon\}, \quad (15.2)$$

en dus dat (met verkorte notatie)

$$|\{\limsup_{r\downarrow 0}|A_rf(x)-f(x)|>2\epsilon\}|\leq \underbrace{|\{H(f-g))(x)>\epsilon\}|}_{\text{nog af te schatten}}+\frac{1}{\epsilon}||f-g||,$$

hetgeen we door g geschikt te kiezen willekeurig klein willen praten teneinde te concluderen dat

$$\forall\epsilon>0:\ |\{\limsup_{r\downarrow 0}|A_rf(x)-f(x)|>2\epsilon\}|=0. \tag{15.3}$$

In dat geval is ook

$$|\{\limsup_{r\downarrow 0}|A_rf(x)-f(x)|>0\}|=0, \tag{15.4}$$

met andere woorden, de verzamelingen van punten waarvoor

$$\lim_{r\downarrow 0}A_rf(x) \tag{15.5}$$

niet bestaat of niet samenvalt met $f(x)$ heeft maat nul.

Voor het afschatten van $H(f-g))(x)$ merken we op dat (neem even $g\equiv 0$) de verzameling

$$E_\epsilon=\{x\in\mathbb{R}^{\mathrm{N}}:\ (Hf)(x)>\epsilon\}$$

open is omdat

$$x\in E_\epsilon\iff\exists r>0:\ \underbrace{\int_{B(x,r)}|f|}_{\text{continu in } r,x}>\epsilon|B(x,r)|. \tag{15.6}$$

De continuiteit volgt uit de gedomineerde convergentiestelling.

Hoe groot kan E_ϵ zijn? Iedere kompakte $K\subset E_\epsilon$ wordt overdekt door slechts eindig veel bolletjes zoals in (15.6), zeg

$$K\subset B_1\cup\cdots\cup B_m,$$

en als deze bolletje voor elke kompakte $K\subset E_\epsilon$ disjunct gekozen kunnen worden dan volgt voor iedere kompakte $K\subset E_\epsilon$ dat

$$\epsilon|K|\leq\epsilon(|B_1+\cdots+|B_m|)\leq\int_{B_1}|f|+\cdots+\int_{B_m}|f|=\int_{B_1\cup\cdots\cup B_m}|f|\leq\int_{\mathbb{R}^{\mathrm{N}}}|f|,$$

en daarmee dat $\epsilon|E_\epsilon|\leq||f||_1$.

Zo'n disjuncte keuze is in het algemeen natuurlijk niet mogelijk maar met een extra factor geldt er wel zo'n schatting. Kies daartoe de grootste bol, zeg B_{j_1}. De bollen B_i waarvoor $B_i \cap B_{j_1} \neq \emptyset$ liggen allemaal in $3B_{j_1}$ waarbij $3B_{j_1}$ staat voor de bol met hetzelfde middelpunt als B_{j_1} maar met drie keer zo grote straal. Als er nog andere bollen zijn neem dan daarvan weer de grootste, zeg B_{j_2}. Die andere bollen hebben allemaal $B_i \cap B_{j_1} = \emptyset$, en die waarvoor $B_i \cap B_{j_2} \neq \emptyset$ liggen allemaal in $3B_{j_2}$. Als er nog andere bollen zijn waarvoor $B_i \cap B_{j_1} = B_i \cap B_{j_2} = \emptyset$ neem dan daarvan weer de grootste, zeg B_{j_3}. Enzovoorts. Dit proces breekt af, mogelijk na de keuze van j_1, mogelijk na de keuze van j_2, of misschien pas na de keuze van j_k. Als B_{j_k} met $1 \leq k \leq m$ de laatste gekozen grootste bol is, dan zijn[1]

$$B_{j_1}, \ldots, B_{j_k}$$

onderling disjunct en geldt

$$B_1 \cup \cdots \cup B_m \subset 3B_{j_1} \cup \cdots \cup 3B_{j_k},$$

zodat

$$\begin{aligned} ||f||_1 &\geq \int_{B_1 \cup \cdots \cup B_m} |f| \geq \int_{B_{j_1} \cup \cdots \cup B_{j_k}} |f| = \int_{B_{j_1}} |f| + \cdots + \int_{B_{j_k}} |f| \\ &\geq \epsilon(|B_{j_1}| + \cdots + |B_{j_k}|) = 3^{-\mathrm{N}}\epsilon(|3B_{j_1}| + \cdots + |3B_{j_k}|) \\ &\geq 3^{-\mathrm{N}}\epsilon|3B_{j_1} \cup \cdots \cup 3B_{j_k}| \geq 3^{-\mathrm{N}}\epsilon|B_1 \cup \cdots \cup B_m| \geq 3^{-\mathrm{N}}\epsilon|K| \end{aligned}$$

voor alle kompakte $K \in E_\epsilon$ en dus volgt dat

$$\epsilon|E_\epsilon| \leq 3^{\mathrm{N}}||f||_1, \tag{15.7}$$

hetgeen we gebruiken in (15.2) om te concluderen dat

$$|\{x \in \mathbb{R}^{\mathrm{N}} : \limsup_{r \downarrow 0} |A_r f(x) - f(x)| > 2\epsilon\}| \leq \frac{1}{\epsilon}(3^{\mathrm{N}} + 1)||f - g||_1. \tag{15.8}$$

Als elke $f \in L^1(\mathbb{R}^{\mathrm{N}})$ willekeurig goed in 1-norm te benaderen is met continue functies g die nul zijn buiten een bol, volgt (15.4).

Per definitie van meetbaarheid en integreerbaarheid is iedere $f \in L^1(\mathbb{R}^{\mathrm{N}})$ zowel de puntsgewijze limiet als limiet in 1-norm van een rij functies van de vorm

$$\sum_{k=1}^{\mathrm{K}} a_k \chi_{E_k}$$

[1] De uitspraak die nu volgt bewijst Vitali's covering lemma.

waarin $E_1, \ldots, E_{\rm K}$ onderling disjuncte meetbare verzamelingen zijn met eindige maat, en

$$\chi_E(x) = \begin{cases} 1 & \text{als } x \in E \\ 0 & \text{als } x \notin E, \end{cases}$$

de karakteristieke functie van E definieert. De maat van zo'n E wordt als uitwendige maat gedefinieerd via karakteristieke functies van blokken die we met blipfuncties kunnen benaderen in 1-norm. Bijgevolg kan in (15.8) het rechterlid willekeurig klein gekozen worden.

Zo hebben we Lebesgue's differentiatiestelling nu bewezen. Iedere lokaal integreerbare meetbare functie heeft de eigenschap voor bijna elke x geldt dat

$$f(x) = \lim_{r \downarrow 0} A_r f(x).$$

Ook zonder maat- en integratie theorie is zo'n uitspraak te formuleren en bewijzen, rechtstreeks vanuit de definitie van $f \in L^1(\mathbb{R}^{\rm N})$ als equivalentklasse van Cauchyrijen van continue functies f_n die elk buiten een bol B_n identiek nul zijn. Van zo'n rij functies eisen we dan dat

$$||f_n - f_m|| = \int_{\mathbb{R}^{\rm N}} |f_n - f_m| \to 0$$

als $m, n \to \infty$, waarbij de integralen de facto gewone Riemann integralen over bollen $B_{m,n}$ zijn waarbuiten f_n en f_m nul zijn. Twee zulke Cauchyrijen heten equivalent als $||f_n - g_n||_1 \to 0$ als $n \to \infty$, en

$$(A_r f)(x) = \lim_{n \to \infty} \frac{1}{|B(x,r)|} \int_{B(x,r)} f_n$$

is weer ondubbelzinning gedefinieerd, en daarmee ook $\mathcal{N}_f$, maar de puntgewijze limiet

$$\lim_{n \to \infty} f_n(x)$$

hoeft in het algemeen niet te bestaan. Maar wel geldt

$$|f_n(x) - f_m(x)| \leq \underbrace{|f_n(x) - A_r f_n(x)|}_{\to 0 \quad \text{als } r \downarrow 0} + \underbrace{|A_r f_n(x) - A_r f_m(x)|}_{\leq H(f_n - f_m)(x)} + \underbrace{|A_r f_m(x) - f_m(x)|}_{\to 0 \quad \text{als } r \downarrow 0},$$

en dus dat we, gegeven $\epsilon > 0$ en $\eta > 0$,

$$|\{x \in \mathbb{R}^{\rm N} : |f_n(x) - f_m(x)| > \epsilon\}| \leq \frac{1}{\epsilon} 3^{\rm N} ||f_n - f_m|| \leq \eta$$

kunnen forceren door $m, n \geq {\rm N}$ te kiezen waarbij N afhangt van $\eta \epsilon 3^{-{\rm N}}$ via de Cauchy-eigenschap van de rij f_n.

Neem nu $\epsilon_k \downarrow 0$ en $\eta_k \downarrow 0$ met

$$\sum_{k=1}^{\infty} \epsilon_k = \epsilon, \quad \sum_{k=1}^{\infty} \eta_k = \eta$$

en bijbehorende N_k zoals hierboven dan convergeert f_{N_k} uniform op het complement F van de vereniging

$$E = \cup_{k=1}^{\infty} \{x \in \mathbb{R}^{\mathrm{N}};\ |f_{\mathrm{N}_{k+1}}(x) - f_{\mathrm{N}_k}(x)| > \epsilon_k\},$$

en

$$|E| \leq \sum_{k=1}^{\infty} \eta_k = \eta.$$

Noem $\eta = \delta_1$, $E = E_1$, $F = F_1$ en herhaal het argument met $\eta = \delta_2 < \delta_1$. Dit geeft E_2 met $|E_2| \leq \delta_1$ en een op $F_2 = E_2^c$ uniform convergente verdere deelrij. Met $\delta_n \downarrow 0$ geeft dit $G_n = F_1 \cup \cdots \cup F_n$ met $G_n^c \subset E_n$ en dus $|G_n^c| \leq \delta_n \to 0$, met steeds verdere deelrijen uniform convergent op G_n. De diagonaal rij convergeert nu op het complement van de nulverzameling $\cap_{n=1}^{\infty} E_n$.

16 Functies op de cirkel

Als we afspreken dat twee getallen in $\mathbb{R}$ eigenlijk hetzelfde zijn als ze een geheel veelvoud van 2π verschillen dan maken $\mathbb{R}$ en 2π de verzameling $\mathbb{R}_{2\pi}$, een verzameling waarin op natuurlijke manier de optelling is gedefinieerd. Dat gaat net als in $\mathbb{Z}_n$, de verzameling die we krijgen uit de verzameling $\mathbb{Z}$ van gehele getallen en een vast getal $n \in \mathbb{N}$, door af te spreken dat twee gehele getallen gelijk zijn als ze een geheel veelvoud van n verschillen. Zoals vaak

$$\mathbb{Z}_n = \{0, 1, \ldots, n-1\}$$

wordt geschreven, met $0 = n$, kunnen we ook

$$\mathbb{R}_{2\pi} = [0, 2\pi)$$

schrijven, maar we geven er de voorkeur om $\mathbb{R}_{2\pi}$ in de schrijfwijze te laten corresponderen met $[-\pi, \pi)$, waarbij $-\pi = \pi$. Deze π is hier een positief reëel getal, waarvoor we op enig moment de π die we van de cirkel kennen zullen nemen, *maar dat hoeft nu nog even niet*. Net als $\mathbb{Z}_n$ is $\mathbb{R}_{2\pi}$ met de voor de hand liggende optelling een commutatieve groep[1].

Functies $f : \mathbb{R} \to \mathbb{R}$ die 2π-periodiek zijn kunnen we ook opvatten als functies $f : \mathbb{R}_{2\pi} \to \mathbb{R}$, en omgekeerd. De verzameling van continue 2π-periodieke functies noemen we

$$C(\mathbb{R}_{2\pi}).$$

Ieder tweetal functies gedefinieerd op dezelfde verzameling, dus ook f en g in $C(\mathbb{R}_{2\pi})$, kunnen we bij elkaar optellen[2] middels

$$x \xrightarrow{f+g} f(x) + g(x)$$

als definitie van $f + g \in C(\mathbb{R}_{2\pi})$. Met

$$x \xrightarrow{tf} tf(x)$$

voor $t \in \mathbb{R}$ en $f \in C(\mathbb{R}_{2\pi})$ is ook de scalaire vermenigvuldiging gedefinieerd en zo is $C(\mathbb{R}_{2\pi})$ een vectorruimte[3] over $\mathbb{R}$, waarop

$$f \cdot g = (f, g) = \int_{-\pi}^{\pi} f(x)g(x)\,dx$$

[1] Google: Abelian group.

[2] Evenzo is natuurlijk ook fg gedefinieerd via $x \xrightarrow{fg} f(x)g(x)$.

[3] En met de vermenigvulding een algebra.

een inwendig produkt[4] definieert, maar $C(\mathbb{R}_{2\pi})$ is met dit integraalinprodukt geen Hilbertruimte, zoals de volgende opgave laat zien.

Opgave 16.1. De zaagtandfunctie Z wordt gedefinieerd door

$$Z(x) = \begin{cases} \pi - x & \text{als } x \in (0, \pi] \ ; \\ -x - \pi & \text{als } x \in [-\pi, 0), \end{cases}$$

en door $Z(0) = 0$. Met deze keuze voor $Z(0)$ behoort Z tot $\mathcal{C}_0(\mathbb{R}_{2\pi})$, de ruimte[5] van functies $f : \mathbb{R}_{2\pi} \to \mathbb{R}$ die continu zijn in iedere $x \in \mathbb{R}_{2\pi}$, behalve eventueel in $x = 0$, maar waarvoor wel geldt dat

$$f(0) = \frac{1}{2}\left(\lim_{x \downarrow 0} f(x) + \lim_{x \uparrow 0} f(x)\right),$$

waarbij linker- en rechterlimiet dus allebei bestaan. Op $\mathcal{C}_0(\mathbb{R}_{2\pi})$ is $(f, g) \to f \cdot g$ ook een inprodukt. Teken de grafieken van een rij functies $Z_1, Z_2, \ldots$ in $C(\mathbb{R}_{2\pi})$ waarvoor geldt dat $(Z_n - Z, Z_n - Z) \to 0$ als $n \to \infty$.

De rij $Z_1, Z_2, \ldots$ is convergent in $\mathcal{C}_0(\mathbb{R}_{2\pi})$ met betrekking tot de inproduktnorm

$$f \to |f| = \sqrt{(f, f)} = \left(\int_{-\pi}^{\pi} |f|^2\right)^{\frac{1}{2}}$$

omdat $|Z_n - Z| \to 0$ als $n \to \infty$, en dus is de rij $Z_1, Z_2, \ldots$ ook een Cauchyrij in $C(\mathbb{R}_{2\pi})$ met die inproduktnorm, die echter in $C(\mathbb{R}_{2\pi})$ geen limiet heeft[6]. Om van $C(\mathbb{R}_{2\pi})$ met de inproduktnorm een Hilbertruimte te maken, die we de naam

$$L^2(\mathbb{R}_{2\pi})$$

willen geven, moeten we alle limieten van Cauchyrijtjes aan $C(\mathbb{R}_{2\pi})$ toevoegen, maar hoe doe je dat?

16.1 Een Hilbert ruimte voor (periodieke) functies?

De abstracte constructie van een Hilbertruimte H uit $C(\mathbb{R}_{2\pi})$ met zijn inprodukt is dat je een 2-Cauchyrijtje $f_1, f_2, \ldots$, een rijtje waarvoor geldt dat

$$|f_n - f_m|_2 \to 0$$

[4] Let op, met $(f, g) \dot{\to} f \cdot g = (f, g)$ is de haakjesnotatie soms verwarrend.
[5] De notatie $\mathcal{C}_0$ is alleen voor nu even.
[6] Waarom niet?

als $m, n \to \infty$, ziet als een benadering van een functie f in de te construeren H. Dit is analoog aan de vertrouwde gewoonte om decimale of binaire ontwikkelingen te zien als benaderingen van reële getalen, waarbij verschillende ontwikkelingen hetzelfde reële getal kunnen maken. Bij die reële getallen denk je doorgaans aan punten op een getallenlijn, al hoeft dat natuurlijk niet: als je het jezelf moeilijk wil maken kan de abstracte constructie van $\mathbb{R}$ prima zonder.

De standaardvisualisatie van een functie is de grafiek van die functie, in het geval $f : \mathbb{R} \to \mathbb{R}$ een deelverzameling G van

$$\mathbb{R}^2 = \{(x, y) : x, y \in \mathbb{R}\}$$

met de eigenschap dat

$$\forall x \in \mathbb{R}\ \exists_1 y \in \mathbb{R} : (x, y) \in G.$$

In deze notatie is $\forall$ *short* voor *voor alle* , en staat $\exists_1$ voor *er is precies één*[7]. Die ene y in de uitspraak wordt in deze context bij afspraak genoteerd als $f(x)$, en daarmee is de formele definitie van een grafiek in $\mathbb{R}^2$ de facto equivalent met de definitie van een functie van $\mathbb{R}$ naar $\mathbb{R}$. Omdat $\mathbb{R}^2$ als vlak in ons hoofd past, en de grafiek G van $f : \mathbb{R} \to \mathbb{R}$ daar weer een deelverzameling van is, zien we de grafiek G van f nu in het platte xy-vlak als de verzameling gegeven door $y = f(x)$.

Het ligt voor de hand om de abstracte constructie van H uit $C(\mathbb{R}_{2\pi})$ te zien als gebeurende in het platte vlak $\mathbb{R}^2$, waarbij de grafiek G van een functie $f : \mathbb{R}_{2\pi} \to \mathbb{R}$ dus de eigenschap moet hebben dat

$$\frac{x_1 - x_2}{2\pi} \in \mathbb{Z} \implies f(x_1) = f(x_2),$$

hetgeen overeenkomt met het oprolbaar[8] zijn van het oneindige platte vlak tot een cylinder waarin de grafiek G keurig over zichzelf heen ligt.

Met of zonder voorstelling, twee verschillende 2-Cauchyrijtjes $f_1, f_2, \ldots$ en $g_1, g_2, \ldots$ in $C(\mathbb{R}_{2\pi})$ moeten hetzelfde element uit de te maken H zijn als geldt dat

$$|f_n - g_n| \to 0$$

voor $n \to \infty$. Opgave 16.1 laat bijvoorbeeld zien dat de zaagtandfunctie Z in de te maken H moet zitten, maar niet iedereen zal dezelfde rij $Z_1, Z_2, \ldots$ als grafieken getekend hebben. Het is goed om dat nog wat preciezer te bekijken.

[7] $\exists$ staat voor *er bestaat een.*

[8] Stel je de problemen bij het oprollen even voor....

Opgave 16.2. Maak Opgave 16.1 nog een keer maar anders. Teken de grafieken van een rij functies $\tilde{Z}_n \in C(\mathbb{R}_{2\pi})$ waarvoor geldt dat $(\tilde{Z}_n - Z, \tilde{Z}_n - Z) \to 0$ als $n \to \infty$. Kies de rij functies $\tilde{Z}_1, \tilde{Z}_2, \ldots$ nu zo dat voor alle $n \in \mathbb{N}$ geldt dat $\tilde{Z}_n(0) = 1$.

Opgave 16.3. Maak Opgave 16.2 maar nu met $\tilde{Z}_n(0) = 0$.

Opgave 16.4. Maak Opgave 16.3 maar nu met $\tilde{Z}_n(0) = 2$.

Opgave 16.5. Maak Opgave 16.3 maar nu met $\tilde{Z}_n(0) = n$.

Opgave 16.6. Laat in Opgaven 16.2,16.3,16.4,16.5 hierboven zien dat $|Z_n - \tilde{Z}_n| \to 0$ als $n \to \infty$, waarbij Z_n is als in Opgave 16.2.

Wat deze opgaven laten zien is dat functies in H geen gewone functies kunnen zijn. Abstract gezien zouden alle benaderende rijen dezelfde $Z : \mathbb{R}_{2\pi} \to \mathbb{R}$ moeten maken, maar dat lijkt via de opgaven hierboven te leiden tot de conclusie dat

$$0 = Z(0) = 1$$

en meer verwarring. Soortgelijke spelletjes kunnen we spelen met de nulfunctie

$$x \xrightarrow{\mathbf{0}} 0$$

zelf.

Opgave 16.7. Maak een rij functies $f_1, f_2, \ldots$ in $C(\mathbb{R}_{2\pi})$ waarvoor geldt dat $f_n(0) = 1$ en $|f_n| = |f_n - \mathbf{0}| \to 0$.

Iedere rij echte functies f_n die we gebruiken om een f in H te maken kan veranderd worden in een rij $\tilde{f}_n$ die in een gegeven punt gek gedrag vertoont, zoals convergeren naar een 'verkeerde' limiet, maar wel de eigenschap heeft dat $|f_n - \tilde{f}_n| \to 0$. We moeten kennelijk af van het idee dat een functie in elk punt gedefinieerd is. In sommige punten is dat wellicht een artefact, zoals in Opgave 16.7, maar bij functies als Z is er echt een keuze die gemaakt moet

worden. Of niet, als we afspreken dat functies niet per se in elk punt van hun definitiegebied gedefinieerd hoeven zijn. Let op, met Z zitten ook alle verschoven zaagtandfuncties Z_p (met $p \in \mathbb{R}$)

$$x \xrightarrow{Z_p} Z(x-p)$$

in H, en daarmee ook een grote klasse van functies van de vorm

$$S = \sum_{n=1}^{\infty} a_n Z_{p_n},$$

waarbij $p_1, p_2, \ldots$ een willekeurige rij punten in $\mathbb{R}$ mag zijn, en elke p_n een probleempunt is voor wat betreft de definitie van $S(p_n)$.

Opgave 16.8. Neem aan dat H geconstrueerd is zoals hierboven beschreven. Neem aan dat $p_1, p_2, \ldots$ en $a_1, a_2, \ldots$ rijen in $\mathbb{R}$ zijn, en dat

$$\sum_{n=1}^{\infty} |a_n| < \infty.$$

Waarom moet gelden dat $S \in H$? Hint: laat eerst zien dat S een begrensde functie is.

Kortom, behalve de mooie periodieke functies

$$x \xrightarrow{c_n} \cos nx$$

en

$$x \xrightarrow{s_n} \sin nx$$

($n \in \mathbb{N}$), waarvan we ook sommen van de vorm

$$\sum_{n=1}^{\infty} a_n c_n + \sum_{n=1}^{\infty} b_n s_n \tag{16.1}$$

kunnen nemen, met coefficiënten als in Opgave 16.8, moeten er in de H die we zoeken een heleboel lelijke functies zitten. Daarbij moeten evident verschillende functies soms (of vaak) als element van H als dezelfde functie gezien worden. Waarom? Omdat twee Cauchyrijen $f_1, f_2, \ldots$ en $\tilde{f}_1, \tilde{f}_2, \ldots$ in $C(\mathbb{R}_{2\pi})$ met de eigenschap dat $f_n - \tilde{f}_n \to 0$ dezelfde f in H moeten maken, en we in de voorbeelden gezien hebben dat bijvoorbeeld $f_n(0)$ en $\tilde{f}_n(0)$ verschillende of helemaal geen limieten kunnen hebben.

Opgave 16.9. Maak een Cauchyrij $f_1, f_2, \ldots$ die naar de nulfunctie $\mathbf{0}$ convergeert in de inproduktnorm maar waarvoor de rij $f_1(x), f_2(x), \ldots$ niet convergeert, welke $x \in \mathbb{R}_{2\pi}$ je ook kiest.

De vraag is dus niet alleen welke functie je kiest als de meest natuurlijke functie binnen een equivalentieklasse van functies die in H niet van elkaar te onderscheiden zijn, maar ook hoe je überhaupt aan zo'n functie komt als f in H gedefinieerd is via een Cauchyrij $f_1, f_2, \ldots$ in $C(\mathbb{R}_{2\pi})$.

16.2 Standaard Hilbertruimten voor 'functies'

In wat volgt maken we enerzijds precies welke functies f op te vatten zijn als $f \in H$ en anderzijds waarom we die functies nog wel als functies zien. Iedere $f \in H$ moet daartoe voor bijna[9] alle $x \in \mathbb{R}_{2\pi}$ een natuurlijke waarde hebben, waarbij het gedrag van f in de buurt van elk zulk een x leidend moet zijn[10]. Voor de zaagtand Z leidt dit bij het gelijkwegen van wat $Z(x)$ is voor $x < 0$ en $x > 0$ onherroepelijk tot $Z(0) = 0$ als de natuurlijke keuze voor $Z(0)$, het gemiddelde van de linker- en rechterlimiet. Maar of zulke limieten voor iedere f in de H die we maken altijd in genoeg punten bestaan is (zeker a priori) niet zo duidelijk.

Hoe het ook zij, de waarde van $f \in H$ in $\pi = -\pi \in S$ doet er niet toe. Voor iedere $a \in \mathbb{R}$ en iedere functie $f : (a - \pi, a + \pi)$ die we toe willen laten in H na periodieke uitbreiding van f tot $\mathbb{R} \to \mathbb{R}$ is het niet belangrijk of en hoe $f(a - \pi)$ en $f(a + \pi)$ gedefinieerd zijn. In het bijzonder is de functie $\tilde{Z}$ gedefinieerd

$$x \in (0, 2\pi) \xrightarrow{\tilde{Z}} \pi - x$$

na periodieke uitbreiding tot $Z : \mathbb{R} \to \mathbb{R}$ in H gelijk aan de Z uit Opgave 16.1. Waar bij de functies c_n en s_n het periodiek uitbreiden vanzelf gaat, is het bij functies als Z vervelend om de formules überhaupt op te schrijven.

De functie Z heeft in ieder geheel veelvoud van 2π een sprong. De even-eens oneven blokfunctie *blok* $\in H$, gedefinieerd door

$$\mathit{blok}(x) = \begin{cases} 1 & \text{als } x \in (0, \pi) \ ; \\ -1 & \text{als } x \in [-\pi, 0), \end{cases}$$

heeft in ieder geheel veelvoud van π een sprong. De even *kartelrandfunctie*

[9] Wat *bijna* betekent is de hamvraag.

[10] Waarom eigenlijk? Wel, we zijn uitgegaan van continue functies.

Ka $\in H$ daarentegen, gedefinieerd door

$$Ka(x) = \begin{cases} \frac{\pi}{2} - x & \text{als } x \in (0, \pi) \ ; \\ \frac{\pi}{2} + x & \text{als } x \in [-\pi, 0), \end{cases}$$

heeft geen sprongen als we de definitie van *Ka* uitbreiden met $Ka(2\pi n) = \frac{\pi}{2}$ in de gehele veelvouden $2\pi n$ van 2π ($n \in \mathbb{Z}$). Al deze functies zijn instructief als voorbeeld bij de vraag of ze te schrijven zijn als een oneindige som van de vorm (16.1). Met name de zaagtand is een bron van leerzaam vermaak zoals we zullen zien.

Opgave 16.10. Schets de grafieken van $Z, blok, Ka$, en ook van $c_1 = \cos$ en $s_1 = \sin$.

Ga nog eens na dat de functies

$$\frac{c_n}{\sqrt{\pi}}, \frac{s_n}{\sqrt{\pi}} \quad (n \in \mathbb{N}), \frac{1}{\sqrt{2\pi}}$$

een orthonormaal stelsel vormen, en dat H dus alle 'functies' f van de vorm

$$f = a_0 \frac{1}{\sqrt{2\pi}} + \sum_{n=1}^{\infty} a_n \frac{c_n}{\sqrt{\pi}} + \sum_{n=1}^{\infty} b_n \frac{s_n}{\sqrt{\pi}} \tag{16.2}$$

zou moeten bevatten, meestal geschreven als

$$f = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n c_n + b_n s_n),$$

als de (net iets andere) rijtjes van coëfficiënten a_n en b_n maar kwadratisch sommeerbaar zijn.

De vraag of je zo alle f in H krijgt kan beginnen met de vraag of de oneven functies Z en *blok* te schrijven zijn als

$$\sum_{n=1}^{\infty} b_n s_n,$$

en de even functie *Ka* als

$$\sum_{n=1}^{\infty} a_n c_n.$$

De sommen moeten hierbij convergent zijn in de 2-norm die hoort bij het standaard inproduct

$$f \cdot g = (f, g) = \int_{-\pi}^{\pi} f(x)g(x)\,dx.$$

Ons doel is te laten zien dat iedere f in H inderdaad van de vorm (16.2) met

$$\sum_{n=0}^{\infty} a_n^2 < \infty, \quad \sum_{n=1}^{\infty} b_n^2 < \infty,$$

en een karakterisatie van H als $L^2(\mathbb{R}_{2\pi})$ die los staat van de specifieke keuze die we met (16.2) maken.

16.3 Fourierreeksen

De functie

$$f_7(x) = \sin x - \frac{\sin 2x}{2} + \frac{\sin 3x}{3} - \frac{\sin 4x}{4} + \frac{\sin 5x}{5} - \frac{\sin 6x}{6} + \frac{\sin 7x}{7}$$

is periodiek met periode 2π. Op het interval $(-\pi, \pi)$ ligt de grafiek van f_7 vlakbij de grafiek van de oneven functie $f(x) = \frac{1}{2}x$. Kennelijk is

$$\frac{x}{2} = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{\sin kx}{k}, \tag{16.3}$$

maar alleen voor $-\pi < x < \pi$.

Opgave 16.11. Er is een verband met machtreeksen: het rechterlid in (16.3) is het imaginaire deel van

$$\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} \zeta^k, \quad \zeta = e^{ix}.$$

Bepaal de som van deze machtreeks voor $|\zeta| < 1$. Hint: differentieer naar ζ, sommeer en primitiveer.

Opgave 16.12. Volgens de complexe versie van het criterium van Leibniz convergeert de machtreeks in in Opgave 16.11 voor alle ζ met $|\zeta| = 1$ behalve $\zeta = -1$. Aangenomen dat de somfunctie die je in Opgave 16.11 hebt gevonden ook geldig is voor alle ζ met $|\zeta| = 1$ waar de reeks convergeert, verifieer (16.3). Op de complexe versie van het criterium van Leibniz en de aanname komen we nog terug in Sectie 5.2.

De reeks in (16.3) heet een Fouriersinusreeks. Maken we van de "minnen" plussen dan vinden we dat de grafiek van de functie

$$h_7(x) = \sin x + \frac{\sin 2x}{2} + \frac{\sin 3x}{3} + \frac{\sin 4x}{4} + \frac{\sin 5x}{5} + \frac{\sin 6x}{6} + \frac{\sin 7x}{7}$$

op het interval $(0, 2\pi)$ vlakbij de grafiek van de functie $f(x) = \frac{\pi - x}{2}$ ligt.

De functie

$$g_7(x) = \cos x - \frac{\cos 2x}{4} + \frac{\cos 3x}{9} - \frac{\cos 4x}{16} + \frac{\cos 5x}{25} - \frac{\cos 6x}{36} + \frac{\cos 7x}{49}$$

heeft een grafiek op het interval $(-\pi, \pi)$ vlakbij de grafiek van de even functie

$$g(x) = \frac{\pi^2}{12} - \frac{x^2}{4}$$

ligt. Kennelijk is

$$\frac{x^2}{4} = \frac{\pi^2}{12} + \sum_{k=1}^{\infty} (-1)^k \frac{\cos kx}{k^2}.$$

Het rechterlid, inclusief de constante term, heet een Fouriercosinusreeks. Invullen van $x = 0$ geeft

$$\frac{\pi^2}{12} = 1 - \frac{1}{4} + \frac{1}{9} - \frac{1}{16} + \frac{1}{25} - \cdots .$$

We zullen zien dat bij even 2π-periodieke functies Fouriercosinusreeksen horen, en bij oneven 2π-periodieke functies Fouriersinusreeksen. Omdat elke functie te splitsen is in een even en een oneven functie,

$$f(x) = \frac{f(x) + f(-x)}{2} + \frac{f(x) - f(-x)}{2},$$

hoort zo bij een algemene 2π-periodieke functie de som van een Fouriercosinusreeks en een Fouriersinusreeks. Zo'n som heet een Fourierreeks. Schrijven we de cosinussen en sinussen uit in complexe e-machten,

$$\cos x = \frac{e^{ix} + e^{ix}}{2}, \quad \sin x = \frac{e^{ix} - e^{ix}}{2i},$$

dan wordt een algemene Fourierreeks een reeks van de vorm

$$\sum_{k=-\infty}^{\infty} c_k e^{ikx} = \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx) \tag{16.4}$$

Uiteindelijk zullen we er voor kiezen om met de complexe vorm te werken, het linkerlid in (16.4) dus, een Laurentreeks in $\zeta = e^{ix}$.

Uit de Cauchy integraal formules hebben we Laurentreeksen van de vorm

$$\sum_{k=-\infty}^{\infty} c_k \zeta^k \tag{16.5}$$

zien verschijnen voor complex differentieerbare functies $f(\zeta)$ op een annulus

$$\{\zeta \in \mathbb{C} : R_1 < |\zeta| < R_2\},$$

waarbij het plusstuk

$$\sum_{k=0}^{\infty} c_k \zeta^k$$

convergent is voor $|\zeta| < R_2$ en het minstuk

$$\sum_{k=1}^{\infty} c_{-k} \zeta^{-k}$$

convergent is voor $|\zeta| > R_1$. Als $R_1 = 0$ dan heeft $f(\zeta)$ in $\zeta = 0$ een al dan niet ophefbare singulariteit. Omgekeerd, als we met een willekeurige Laurentreeks van de vorm (16.5) beginnen dan zijn er bijbehorende R_1 en R_2 zodat het plusstuk convergeert naar een complex differentieerbare functie op $\{\zeta \in \mathbb{C} : |\zeta| < R_2\}$ en het minstuk naar een complex differentieerbare functie op $\{\zeta \in \mathbb{C} : |\zeta| > R_1\}$. A priori kunnen zowel R_1 als R_2 ook 0 of ∞ zijn, en in het algemeen kan R_1 groter of kleiner zijn dan R_2. Laurentreeksen die verschijnen als Fourierreeksen van 2π-periodieke functies $f(x)$ via $\zeta = e^{ix}$ hebben meestal $R_1 = R_2 = 1$. In het vervolg zullen we zulke 2π-periodieke functies zien als functies $f : (-\pi, \pi) \to \mathbb{C}$.

Fourierreeksen gaan terug tot Daniel Bernouilli, die de golfvergelijking

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2}$$

voor bijvoorbeeld $0 < x < \pi$ en met randvoorwaarde $u = 0$ voor $x = 0$ en $x = \pi$, met Fouriersinusreeksen probeerde op te lossen. Later was Fourier de eerste die voor een gegeven functie f de coefficienten in integralen wist uit de drukken, toen hij Fouriersinusreeksen gebruikte om de warmtevergelijking

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}$$

op te lossen.

Tegenwoordig zien we de functies

$$\frac{1}{2}, \cos x, \sin x, \cos 2x, \sin 2x, \ldots,$$

en

$$\ldots, e^{-3ix}, e^{-2ix}, e^{-ix}, e^{0ix} = 1, e^{ix}, e^{i2x}, e^{3ix}, \ldots$$

als orthonormale bases in een (Hilbert)ruimte van functies, en de Fouriercoefficienten als coordinaten ten opzichte van deze basis. Voor een grote klasse van functies $f : (-\pi, \pi) \to \mathbb{R}$ zijn de Fouriercoefficienten a_k, b_k en c_k als coordinaten van f ten opzichte van deze bases gedefinieerd. De N-de partiële som van de Fourierreeks van f is

$$S_{\mathrm{N}} f(x) = \sum_{k=-\mathrm{N}}^{\mathrm{N}} c_k e^{ikx} = \frac{a_0}{2} + \sum_{k=1}^{\mathrm{N}} (a_k \cos kx + b_k \sin kx), \tag{16.6}$$

met Fouriercoefficienten

$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos nx \, dx, \quad b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin nx \, dx, \tag{16.7}$$

$$c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-inx} \, dx. \tag{16.8}$$

Het volgende programma is bedoeld om vertrouwd te raken met Fourierreeksen. Gebruik Maple/Mathematica voor de plotjes. De integralen kun je beter met de hand doen.

Opgave 16.13. Bereken $\int_{-\pi}^{\pi} \cos nx \cos mx \, dx$ en $\int_{-\pi}^{\pi} \cos nx \sin mx \, dx$ voor gehele m en n.

Opgave 16.14. Laat $f : (0, \pi) \to \mathbb{R}$ gegeven zijn door $f(x) = 1$ en kies een 2π-periodieke uitbreiding $f : \mathbb{R} \to \mathbb{R}$ die even is (i.e. $f(x) = f(-x)$). Bereken alle Fouriercoefficienten a_n en b_n.

Opgave 16.15. Laat $f : (0, \pi) \to \mathbb{R}$ gegeven zijn door $f(x) = 1$ en kies een 2π-periodieke uitbreiding $f : \mathbb{R} \to \mathbb{R}$ die oneven is (i.e. $f(x) = -f(-x)$).

1. Bereken alle Fouriercoefficienten a_n en b_n.
2. Plot f en $S_{\mathrm{N}}f$ (voor een aantal waarden van N) in een grafiek.
3. Onderzoek numeriek wat er gebeurt met de grootte en plaats van het maximum van $S_{\mathrm{N}}f$ als $\mathrm{N} \to \infty$.
4. Vereenvoudig $S_{\mathrm{N}}f$ in $x = \frac{\pi}{2}$. Vergelijk dit met $f(\frac{\pi}{2})$. Van welke gewone reeks is, aangenomen dat $S_{\mathrm{N}}f(\frac{\pi}{2})$ naar $f(\frac{\pi}{2})$ convergeert, nu de som te bepalen?
5. Idem voor $x = \frac{\pi}{4}$.

Opgave 16.16. Laat $f : (0, \pi) \to \mathbb{R}$ gegeven zijn door $f(x) = \sin x$. Kies een even 2π-periodieke uitbreiding $f : \mathbb{R} \to \mathbb{R}$. Bereken alle Fouriercoefficienten a_n en b_n.

1. Bereken alle Fouriercoefficienten a_n en b_n.
2. Plot f en $S_{\mathrm{N}}f$ (voor een aantal waarden van N) in een grafiek.
3. Vereenvoudig $S_{\mathrm{N}}f$ in $x = 0$. Vergelijk dit met $f(0)$. Van welke gewone reeks is, aangenomen dat $S_{\mathrm{N}}f(0)$ naar $f(0)$ convergeert, nu de som te bepalen?
4. Idem voor $x = \frac{\pi}{2}$.
5. Idem voor $x = \frac{\pi}{4}$.

Opgave 16.17. Laat $f : (0, \pi) \to \mathbb{R}$ gegeven zijn door $f(x) = \cos x$ en kies een oneven 2π-periodieke uitbreiding $f : \mathbb{R} \to \mathbb{R}$.

1. Bereken alle Fouriercoefficienten a_n en b_n en plot f en $S_{\mathrm{N}}f$ (voor een aantal waarden van N) in een grafiek.
2. Vergelijk het gedrag bij $x = 0$ voor N groot met dat in som 16.15.
3. Neem nu de oneven 2π-periodieke uitbreiding $f(x) = 1 - \cos x$ (het verschil van de functie in som 16.15 en de functie in deze som). Onderzoek numeriek het gedrag van $S_{\mathrm{N}}f$ bij $x = 0$ voor N.

Opgave 16.18. Laat $f : (0, \pi) \to \mathbb{R}$ gegeven zijn door $f(x) = \pi - x$ en kies een oneven 2π-periodieke uitbreiding $f : \mathbb{R} \to \mathbb{R}$.

1. Bereken alle Fouriercoefficienten a_n en b_n en plot f en $S_{\rm N}f$ (voor een aantal waarden van $\rm N$) in een grafiek.

2. Differentieer $S_{\rm N}f(x)$ naar x en noem de afgeleide $d_{\rm N}(x)$. Zijn er waarden van x waarvoor $d_{\rm N}(x)$ convergeert als ${\rm N}\to\infty$?

Opgave 16.19. Laat $f:(0,\pi)\to\mathbb{R}$ gegeven zijn door $f(x)=x(\pi-x)$ en kies een oneven 2π-periodieke uitbreiding $f:\mathbb{R}\to\mathbb{R}$.

1. Bereken alle Fouriercoefficienten a_n en b_n en plot f en $S_{\rm N}f$ (voor een aantal waarden van $\rm N$) in een grafiek.

2. Vereenvoudig $S_{\rm N}f$ in $x=\frac{\pi}{2}$. Vergelijk dit met $f(\frac{\pi}{2})$. Van welke gewone reeks is, aangenomen dat $S_{\rm N}f(x)$ naar $f(x)$ convergeert, nu de som te bepalen?

3. Differentieer $S_{\rm N}f(x)$ naar x en noem de afgeleide $g_{\rm N}(x)$. Laat zien dat $g_{\rm N}(x)$ op $\mathbb{R}$ uniform convergeert naar een limietfunctie.

4. Bepaal die limietfunctie numeriek.

5. Vergelijk $g_{\rm N}(0)$ met zijn limietwaarde. Welke som van welke gewone reeks vinden we nu?

16.4 Convergentie van Fourierreeksen

Bij de vraag of, en in welke zin, de Fourierreeks convergeert, en ook als limiet f heeft, m.a.w. of

$$f(x)=\sum_{k=-\infty}^{\infty}c_ke^{ikx}=\frac{a_0}{2}+\sum_{k=1}^{\infty}(a_k\cos kx+b_k\sin kx),$$

speelt het begrip convolutie een belangrijke rol.

Opgave 16.20. Laat zien dat

$$S_{\rm N}f(x)=\frac{1}{2\pi}\int_{-\pi}^{\pi}D_{\rm N}(y)f(x-y)dy,$$

de convolutie van de functies D_n en f op $(-\pi,\pi)$, waarbij

$$D_{\rm N}(x)=\frac{\sin({\rm N}+\frac{1}{2})x}{\sin\frac{1}{2}x}=\sum_{k=-{\rm N}}^{\rm N}e^{ikx}. \tag{16.9}$$

Maak plotjes van D_{N} voor een aantal waarden van N.

De functie D_{N} heet de Dirichlet kern. Voor grote N concentreert deze functie zich bij 0 met een steeds smallere piek waarvan de oppervlakte naar 2π gaat. Dat is de reden dat $S_{\mathrm{N}}f(x)$ naar $f(x)$ convergeert als f voldoende netjes is. Omdat D_{N} voor grotere N steeds meer tekenwisselingen heeft is dit lastig om te bewijzen. Het gemiddelde van D_0,..,D_{N},

$$F_{\mathrm{N}}(x) = \frac{1}{\mathrm{N}+1}\left(D_0(x) + \cdots + D_{\mathrm{N}}(x)\right) = \frac{1}{\mathrm{N}+1}\frac{\sin^2\frac{(\mathrm{N}+1)x}{2}}{\sin^2\frac{x}{2}}, \qquad (16.10)$$

is een veel mooiere functie. Geen tekenwisselingen, integraal 2π, en gepiekt in 0.

Opgave 16.21. Leidt de laatste gelijkheid in (16.10) door

$$\sin\frac{x}{2} + \cdots + \sin\frac{(\mathrm{N}+1)x}{2}$$

als imaginair deel van een eindige meetkundige reeks te schrijven. Verifieer ook dat

$$\int_{-\pi}^{\pi} F_{\mathrm{N}}(x)dx = 2\pi,$$

en dat $F_{\mathrm{N}}(x) \to 0$ als $\mathrm{N} \to \infty$ behalve in veelvouden van 2π. Preciezer:

$$0 < \delta \leq x \leq \pi \implies 0 \leq F_{\mathrm{N}}(x) \leq \frac{1}{\mathrm{N}+1}\frac{1}{\sin^2\frac{\delta}{2}}.$$

Voor vaste δ is de bovengrens klein te maken door N groot te maken. Merk op dat $F_{\mathrm{N}}(x)$ even en 2π periodiek is. Maak plotjes van F_{N} voor een aantal waarden van N.

Als een rij getallen a_n convergeert naar een limiet A, dan convergeren ook de gemiddelden

$$\frac{a_1 + a_2 + \cdots + a_n}{n}$$

naar A. De laatste limiet kan ook bestaan als de rij a_n zelf niet convergeert. Als we de rij a_n gebruiken om een A te benaderen dan kunnen we dus net zo goed eerst naar de gemiddelden kijken. Dat is het idee achter de Cesarosommen:

Opgave 16.22. Definieer

$$\sigma_{\mathrm{N}} f = \frac{1}{\mathrm{N}+1}(S_0 f + S_1 f + \cdots + S_{\mathrm{N}} f),$$

de Cesarosommen van f. Laat zien dat

$$\sigma_{\mathrm{N}} f(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} F_{\mathrm{N}}(y) f(x-y) dy.$$

Opgave 16.23. Laat f een integreerbare (lees: begrensde stuksgewijs continue) 2π-periodieke functie zijn. Laat zien dat dan

$$\sigma_{\mathrm{N}} f(x) - f(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} F_{\mathrm{N}}(y)\left(f(x-y) - f(x)\right) dy$$

Opgave 16.24. Laat f een integreerbare 2π-periodieke functie zijn, met $|f(x)| \leq M$ voor alle x, waarbij $M \geq 0$ vast is. Neem aan dat voor x vast en $|y| \leq \delta$ geldt dat $|f(x-y) - f(x)| \leq \epsilon$. Laat zien dat dan

$$|\sigma_{\mathrm{N}} f(x) - f(x)| = \frac{1}{2\pi}\int_{-\pi}^{\pi} F_{\mathrm{N}}(y)|f(x-y) - f(x)| dy \leq \epsilon + \frac{2M}{\mathrm{N}+1}\frac{1}{\sin^2\frac{\delta}{2}}.$$

Hint: splits de integraal in 3 integralen.

Opgave 16.25. Laat f een 2π-periodieke continue functie zijn. Dan is f uniform continu en begrensd. Waarom? Bewijs dat $\sigma_{\mathrm{N}} f$ uniform naar f convergeert als $\mathrm{N} \to \infty$.

Opgave 16.26. Laat f een 2π-periodieke begrensde stuksgewijs continue functie zijn met de eigenschap dat in elk punt de linker en de rechterlimiet bestaan. Bewijs dat voor elke x de rij $\sigma_{\mathrm{N}} f(x)$ convergeert als $\mathrm{N} \to \infty$. Wat is de limiet? Hint: splits de integraal in 4 integralen.

Opgave 16.27. Laat $f : [-\pi, \pi] \to \mathbb{R}$ 2 keer continu differentieerbaar zijn met $f(\pm\pi) = f'(\pm\pi) = f''(\pm\pi) = 0$. Bewijs dat f in elk punt de som van zijn (uniform convergente) Fourierreeks is. Hint: laat met partieel integreren en schatten zien dat de Fouriercoefficienten a_n en b_n sommeerbare rijen vormen.

Opgave 16.25 laat zien dat in de ruimte van continue functies 2π-periodieke functies voorzien van de maximumnorm

$$||f||_\infty = \max_{x \in myR} |f(x)|$$

geldt dat de Cesarosommen van f naar f convergeren: $||\sigma_{\rm N} f - f||_\infty \to 0$ als ${\rm N} \to \infty$. Maar hoe zit het met $S_{\rm N} f$ zelf? Daartoe is een andere norm veel geschikter. Bij lineaire algebra of topologie zijn ongetwijfeld verschillende normen van 2-vectoren $x = \binom{x_1}{x_2}$ behandeld, bijvoorbeeld

$$||x||_\infty = \max(x_1, x_2), \quad ||x||_1 = |x_1| + |x_2|, \quad ||x||_2 = (|x_1|^2 + |x_2|^2)^{\frac{1}{2}}.$$

Als we in de laatste norm elke 2 door p vervangen (niet de subscript!) dan krijgen we de p-norm. De p-norm is een norm als $p \geq 1$. Je kunt er mee rederen zoals je dat gewend bent bij de absolute waarde. De normaxioma's zijn, voor vectoren x en y, en scalairen λ:

$$||x|| \geq 0; \quad ||x|| = 0 \Leftrightarrow x = 0; \quad ||\lambda x|| = |\lambda|\,||x||; \quad ||x + y|| \leq ||x|| + ||y||.$$

Alleen de 2-norm is een inprodukt norm. Met $x \cdot y = x_1 y_1 + x_2 y_2$ geldt

$$||x||_2^2 = x \cdot x.$$

Om ook voor functies de 2-norm in te voeren maken we eerst een inprodukt. Hieronder zijn f en g steeds 2π-periodieke stuksgewijs continue functies $f, g : \mathbb{R} \to \mathbb{R}$. We doen dus alles eerst nog reeel. Met stuksgewijs continu bedoelen we dat op elk begrensd interval er slechts eindig veel punten zijn waar de functie niet continu is en dat in die punten linker- en rechterlimieten bestaan. De integraal

$$f \cdot g = \int_{-\pi}^{\pi} f(x) g(x) dx \tag{16.11}$$

heet het inprodukt van de functies f en g. We zien f en g als vectoren. Voor elke x zou je dan $f(x)$ en $g(x)$ als coordinaten van f en g kunnen

zien. Daar heb er dan wel heel veel van, voor elke x één van f en één van g. Overeenkomstige coordinaten vermenigvuldigen en sommeren gaat niet, maar integreren wel. Vandaar de inproduktnotatie. Omdat we alles nog reeel doen kunnen we nu gebruik maken van onze intuitie voor gewone reele vectorruimten en inprodukten daarop.

Als $f \cdot g = 0$ dan zeggen we dat f en g *loodrecht op elkaar* staan. De 2-norm van f wordt gedefinieerd door

$$|f|_2 = \sqrt{f \cdot f}, \tag{16.12}$$

zeg maar de lengte van f gezien als vector. Er geldt de volgende implicatie (Pythagoras)

$$f \cdot g = 0 \quad \Rightarrow \quad |f+g|_2^2 = |f|_2^2 + |g|_2^2. \tag{16.13}$$

Hieronder schrijven we

$$S_{\mathrm{N}} g(x) = \frac{c_0}{2} + \sum_{k=1}^{\mathrm{N}} (c_k \cos kx + d_k \sin kx), \tag{16.14}$$

dus f heeft reele Fouriercoefficienten a_k en b_k, en g heeft reele Fouriercoefficienten c_k en d_k. Je kunt nu het volgende programma afwerken.

Opgave 16.28. De Cauchy-Schwartz ongelijkheid zegt dat $|f \cdot g| \leq |f|_2 |g|_2$.

1. Bewijs deze ongelijkheid voor functies f en g met $|f|_2 = |g|_2 = 1$ door $0 \leq \int_{-\pi}^{\pi} (f(x) - g(x))^2 dx = \ldots$ uit te werken.
2. Bewijs de Cauchy-Schwartz ongelijkheid. Hint: pas onderdeel 1 toe op $f(x)/|f|_2$ en $g(x)/|g|_2$.
3. Bewijs de driehoeksongelijkheid voor de 2-norm

$$|f+g|_2 \leq |f|_2 + |g|_2. \tag{16.15}$$

Opgave 16.29. Laat zien dat

$$|S_{\mathrm{N}} f|_2^2 = \pi \left(\frac{1}{2} a_0^2 + \sum_{k=1}^{\mathrm{N}} (a_k^2 + b_k^2) \right)$$

Opgave 16.30. Laat zien dat

$$S_{\rm N} f \cdot S_{\rm N} g = \pi \left(\frac{1}{2} a_0 c_0 + \sum_{k=1}^{\rm N} (a_k c_k + b_k d_k) \right)$$

Opgave 16.31. Definieer $R_{\rm N} f = f - S_{\rm N} f$ en, met

$$\sigma_{\rm N} f = \frac{1}{{\rm N}+1} (S_0 f + S_1 f + \cdots S_{\rm N} f),$$

ook $\rho_{\rm N} f = f - \sigma_{\rm N} f$.

1. Laat zien dat $R_{\rm N} f \cdot S_{\rm N} f = 0$.
2. Laat zien dat $R_{\rm N} f \cdot \sigma_{\rm N} f = 0$.
3. Laat zien dat
$$|S_{\rm N} f|_2^2 + |R_{\rm N} f|_2^2 = |f|_2^2,$$
zodat $|S_{\rm N} f|_2 \leq |f|_2$, en dat (Bessel's ongelijkheid)
$$\frac{1}{2} a_0^2 + \sum_{k=1}^{\rm N} (a_k^2 + b_k^2) \leq \frac{1}{\pi} \int_{-\pi}^{\pi} f(x)^2 dx. \tag{16.16}$$
4. Laat zien dat
$$|R_{\rm N} f|_2^2 + |\sigma_{\rm N} f - S_{\rm N} f|_2^2 = |\rho_{\rm N} f|_2^2.$$
5. In Opgave 16.25 is bewezen dat als f continu en 2π-periodiek is dat dan $\sigma_{\rm N} f \to f$ uniform op $\mathbb{R}$ als ${\rm N} \to \infty$. Bewijs dat in dat geval ook $|R_{\rm N} f|_2 \to 0$ en dat dus (gelijkheid van Parceval)
$$\frac{1}{2} a_0^2 + \sum_{k=1}^{\infty} (a_k^2 + b_k^2) = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x)^2 dx. \tag{16.17}$$
Hint: gebruik onderdeel 4.
6. Laat zien dat
$$\begin{aligned} f \cdot g &= (S_{\rm N} f + R_{\rm N} f) \cdot (S_{\rm N} g + R_{\rm N} g) \\ &= S_{\rm N} f \cdot S_{\rm N} g + R_{\rm N} f \cdot R_{\rm N} g. \end{aligned}$$

7. Bewijs dat als f en g continu en 2π-periodiek zijn, dat

$$\frac{1}{2}a_0c_0 + \sum_{k=1}^{\infty}(a_kc_k + b_kd_k) = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)g(x)dx = \frac{1}{\pi}\, f \cdot g. \tag{16.18}$$

   Hint: gebruik onderdeel 6, som 16.30 en pas de Cauchy-Schwartz ongelijkheid toe op $R_{\mathrm{N}}f \cdot R_{\mathrm{N}}g$.

Opgave 16.32. In deze som laten we zien dat de gelijkheid van Parceval (16.17) ook geldt voor 2π-periodieke stuksgewijs continue functies. Laat $f : \mathbb{R} \to \mathbb{R}$ zo'n functie zijn. Hint: laat zien dat er een rij 2π-periodieke continue functies $f_k : \mathbb{R} \to \mathbb{R}$ bestaat met

$$|f_k - f|_2^2 = \int_{-\pi}^{\pi}(f_k(x) - f(x))^2dx \to 0$$

als $k \to \infty$. Als f discontinu is in x_0, vervang $f(x)$ dan op het interval $(x_0 - \frac{1}{k}, x_0 + \frac{1}{k})$ door een lineaire functie, zó dat de nieuwe functie f_k continu is en lineair op $(x_0 - \frac{1}{k}, x_0 + \frac{1}{k})$.

Opgave 16.33. Bewijs de gelijkheid van Parceval (16.17) voor f. Hint: de gelijkheid van Parceval is equivalent met $|R_{\mathrm{N}}f|_2 \to 0$. Schrijf

$$R_{\mathrm{N}}f = f - f_k + f_k - S_{\mathrm{N}}f_k + S_{\mathrm{N}}f_k - S_{\mathrm{N}}f = (f - f_k) + R_{\mathrm{N}}f_k + S_{\mathrm{N}}(f_k - f)$$

en gebruik de driehoeksongelijkheid (16.15) en som 3 voor $S_{\mathrm{N}}(f_k - f)$ om $|R_{\mathrm{N}}f|_2$ klein te krijgen. Kies hiertoe, gegeven een $\epsilon > 0$, eerst een vaste k groot genoeg, en redeneer dan verder.

Opgave 16.34. (Het Gibbs verschijnsel) De Fouriersinusreeks van $f(x) = \pi - x$ heeft $b_n = \frac{2}{n}$. De oneven 2π-periodieke uitbreiding van f heeft $f(0^+) = \pi$ en $f(0^-) = -\pi$. De N-de Fouriersinussom is

$$S_{\mathrm{N}}f(x) = \sum_{n=1}^{\mathrm{N}} \frac{2}{n}\sin nx.$$

1. Laat zien dat

$$x + S_{\mathrm{N}}f(x) = \int_0^x D_{\mathrm{N}}(s)ds.$$

2. De integraal in het rechterlid heeft extrema in de nulpunten van $D_{\rm N}$. Het eerste maximum $M_{\rm N}$ rechts van $x = 0$ is dus in

$$x = x_{\rm N} = \frac{\pi}{{\rm N} + \frac{1}{2}}.$$

Laat zien dat

$$M_{\rm N} = \int_0^{\frac{\pi}{{\rm N}+\frac{1}{2}}} \frac{\sin((\mathrm{N}+\frac{1}{2})s}{\sin\frac{1}{2}s} ds = \int_0^{\pi} \frac{\sin t}{\sin(\frac{1}{2}\frac{t}{{\rm N}+\frac{1}{2}})} \frac{1}{\mathrm{N}+\frac{1}{2}} dt$$

$$= 2\int_0^{\pi} \frac{\sin t}{t} \frac{\frac{t}{2{\rm N}+1}}{\sin(\frac{t}{2{\rm N}+1})} dt.$$

3. Laat zien dat

$$\frac{\frac{t}{2{\rm N}+1}}{\sin(\frac{t}{2{\rm N}+1})} \to 1,$$

uniform op $t \in [0, \pi]$.

4. Laat zien dat

$$M_{\rm N} \to 2\int_0^{\pi} \frac{\sin t}{t} dt$$

als ${\rm N} \to \infty$.

5. De functie $S_{\rm N} f(x)$ heeft in $x = x_{\rm N}$ een negatieve afgeleide. Leg uit waarom ook het eerste maximum van $S_{\rm N} f(x)$ rechts van $x = 0$ naar

$$2\int_0^{\pi} \frac{\sin t}{t} dt$$

convergeert als ${\rm N} \to \infty$. Dat is groter dan $\pi = f(0^+)$ met een factor 1.178979744.

Opgave 16.35. De integraal $\int_0^{\infty} \frac{\sin t}{t} dt$ is en wel uit te rekenen, met behulp van de complexe functie $\frac{e^{iz}}{z}$, zie Sectie 81 en 82 van Churchill & Brown.

16.5 Dat andere inprodukt met afgeleiden

De deelruimte V van H die de rol gaat spelen zoals in de eerdere voorbeelden met $H = l^{(2)}$ wordt gedefinieerd door het inprodukt

$$((f, g)) = (f', g'),$$

hetgeen niet voor alle f en g in H gedefinieerd is, net zoals het inprodukt in Opgave 14.18 niet voor alle x en y in $l^{(2)}$ gedefinieerd is. Informeel wordt V gegeven door

$$V = \{f \in H : f' \in L^2(-\pi, \pi)\},$$

waarbij met f ook f' steeds 2π-periodiek wordt uitgebreid tot een functie gedefinieerd op heel $\mathbb{R}$.

Dat uitbreiden is makkelijk, en komt in de opgaven hieronder eerst nog aan de orde, ook ter voorbereiding van wat een stuk lastiger is: *wat betekent het dat f' als meetbare en kwadratisch integreerbare functie bestaat?*

Opgave 16.36. Ga na dat (ook) voor functies f in $L^2(-\pi, \pi)$ met $f(-\pi) \neq f(\pi)$ er geen problemen zijn met de uitbreiding f naar een $f \in H$.

Opgave 16.37. Zijn er functies f in $L^2(-\pi, \pi)$ waarvoor aan $f(0)$ geen betekenis[11] kan worden gegeven?

Opgave 16.38. Er is maar één 2π-periodieke oneven[12] functie Ka die voldoet aan $Ka(x) = 1$ voor $0 < x < \pi$. Schets de grafiek van Ka en maak een rij 2π-periodieke oneven continue functies $Ka_1, Ka_2 \ldots$ waarvoor geldt dat $|Ka_n - Ka|_2 \to 0$ als $n \to \infty$. Hint: schets eerst de grafieken van Ka_n.

Opgave 16.39. Bewijs dat een oneven 2π-periodieke functie wordt vastgelegd door zijn functiewaarden op het interval $(0, \pi)$. Hint: gebruik de regels $f(-x) = -f(x)$ en $f(x) = f(x + 2\pi)$. Wat is $f(0)$? En $f(\pi)$?

Leuke functies om over na te denken, maar zulke functies komen we niet tegen als we een zinvolle definitie van de uitspraak dat f' bestaat in bijvoorbeeld $L^2(0, \pi)$ kunnen geven. Wel is het zo f' best zelf zo'n functie kan zijn. Bijvoorbeeld als je f definieert als

$$f(x) = \int_0^x S(s)\, ds,$$

[11] Lees: een betekenisvolle waarde kan worden toegekend?

[12] $Ka(-x) = -Ka(x)$ voor alle $x \in \mathbb{R}$.

met een begrensde S zoals eerder gemaakt in Opgave 16.8. Iedere primitieve functie

$$F(x) = \int_0^x f(s)\,ds$$

van een f in H is natuurlijk in principe kandidaat om tot V te behoren.

Opgave 16.40. Verifieer dat zo'n F een begrensde (2π-periodieke) functie is als $f \in H$, en dat het essentieel is dat in de definitie van H is opgenomen dat voor $f \in H$ moet gelden dat[13]

$$\int_{-\pi}^{\pi} f(x)dx = 0\,!$$

De ruimte V krijgen we nu als bestaande uit de primitieve functies van functies in H, waarbij de spreekwoordelijke constante wel goed gekozen moet worden.

Opgave 16.41. Als $f \in H$ dan is F periodiek. Waarom? Ga na dat er voor elke $f \in H$ precies één constante C is waarvoor $x \to F(x) - C$ in H zit.

We weten nu dus wat V moet zijn. De ruimte

$$\{F \in L^2_{loc}(\mathbb{R}) :\ (\forall x \in \mathbb{R})\ F(x) = F(x + 2\pi),\ f = F' \in L^2_{loc}(\mathbb{R})\}$$

is gelijk aan

$$\{F \in L^2_{loc}(\mathbb{R}) :\ f = F' \in H\}$$

de ruimte van *alle* primitieven F van functies $f \in H$, en V krijgen we door voor iedere primitieve F precies die constante te nemen waarmee de primitieve gemiddeld nul wordt. Dus

$$V = \{F \in L^2_{loc}(\mathbb{R}) :\ f = F' \in H,\ \int_{-\pi}^{\pi} F(x)\,dx = 0\}$$

De kwadratisch integreerbare periodieke functies $f : \mathbb{R} \to \mathbb{R}$ vormen een nul-dimensionale vectorruimte waarover nog wel het een en ander te vertellen is. Dat zullen we hier niet doen. Periodieke functies kunnen natuurlijk wel *lokaal* kwadratisch integreerbaar zijn. We schrijven

$$L^2_{loc}(\mathbb{R}) = \{f : \mathbb{R} \to \mathbb{R} :\ f \in L^2(I) \text{ voor elke begrensd interval } I \subset \mathbb{R}\},$$

[13] $0! = 1$, maar hier roept het uitroepteken wel.

maar de periodieke functies in $L^2_{loc}(\mathbb{R})$ vormen geen vectorruimte[14]. En $L^2_{loc}(\mathbb{R})$ zelf is wel een vectorruimte maar geen genormeerde ruimte, althans niet met een natuurlijke[15] norm. Maar

$$H = \{f \in L^2_{loc}(\mathbb{R}) : \ (\forall x \in \mathbb{R}) \ f(x) = f(x+2\pi); \ \int_{-\pi}^{\pi} f(x)dx = 0\}$$

wel, de ruimte van 2π-periodieke kwadratisch integreerbare[16] periodieke functies, met inprodukt

$$(f, g) = \int_{-\pi}^{\pi} f(x)g(x)\, dx,$$

waarbij we de ruimte nu beperken tot functies die gemiddeld nul zijn.

De constante functies zijn in deze H buitengesloten, omdat ze in het verhaal dat gaat volgen een vervelend buitenbeentje zijn. Bijgevolg van deze keuze zitten er in H ook geen positieve functies trouwens. Wel in H zitten de functies c_n en s_n uit Opgave 13.22 en net als elke functie in H zijn deze door beperking tot het interval $(-\pi, \pi)$ op te vatten als element van

$$\tilde{H} = \{f \in L^2(-\pi, \pi) : \ \int_{-\pi}^{\pi} f(x)\, dx = 0\},$$

een ruimte die we voor gemak met H identificeren door iedere $f \in \tilde{H}$ weer uit te breiden tot heel $\mathbb{R}$ middels $f(x) = f(x + 2\pi)$ voor alle x.

16.6 Blipfuncties

Het formulevoorschrift

$$x \xrightarrow{\textit{blip}} \begin{cases} \exp(-\frac{1}{x}) & \text{for } x > 0 \\ 0 & \text{for } x \leq 0, \end{cases}$$

definieert de functie $\textit{blip} : \mathbb{R} \to [0, 1)$ die met goed recht zowel oogverblindend mooi als gruwelijk lelijk genoemd[17] mag worden.

Opgave 16.42. Schets de grafiek van *blip* en onderzoek het gedrag van $\textit{blip}'(x)$ als $x \downarrow 0$. En van $\textit{blip}''(x)$. En van alle afgeleiden van *blip*. Concludeer dat *alle* afgeleiden van *blip* als continue functies van $\mathbb{R}$ naar $\mathbb{R}$ bestaan!

---

[14] Waarom niet?

[15] Bas, weet je nog, de discussie na je voordracht op de VU?

[16] D.w.z. f is meetbaar en $\int_{-\pi}^{\pi} f(x)^2 dx < \infty$.

[17] De naam *blip* heb ik gezien in een mooi boek, weet niet meer welk.

Opgave 16.43. Je kunt *blip* ook schalen. Definieer $blip_n$ door

$$blip_n(x) = blip(nx) = \exp(-\frac{1}{nx})$$

en bepaal $\lim_{n\to\infty} blip_n(x)$ voor elke $x \in \mathbb{R}$. De limietfunctie heet de Heaviside[18] functie, hier genoteerd als $He(x)$. Deze functie is niet continu in $x = 0$. De waarde van $He(0)$ als limietwaarde van $blip_n(0)$ is 0, maar net zo vaak wordt $He(0) = \frac{1}{2}$ of $He(0) = 1$ genomen. Of zelfs $He(0) = [0, 1]$.

Opgave 16.44. De functies *blip* en *He* zitten niet in $L^2(\mathbb{R})$. Waarom niet? Maar $He - blip$ wel. Waarom? En $He - blip_n$ ook. De affiene ruimte

$$He + L^2(\mathbb{R}) = \{f = He + g : \, g \in L^2(\mathbb{R})\}$$

is voorzien van de 2-metriek

$$d(f, g) = \left(\int_{-\infty}^{\infty} |f(x) - g(x)|^2 \, dx\right)^{\frac{1}{2}}$$

een volledige metrische ruimte[19]. Laat zien zien dat $d(blip_n, He) \to 0$ als $n \to \infty$.

Opgave 16.45. Definieer de functies $blok_n$ door

$$blok_n i(x) = blip_n(x) blip_n(\pi - x)$$

en laat zien dat

$$\lim_{n\to\infty} blok_n(x) = \chi_{(0,\pi)}(x)$$

voor alle $x \in \mathbb{R}$. Waarom geldt dat $blok_n \to \chi_{(0,\pi)}$ in 2-norm?

Opgave 16.46. Dezelfde vragen als in Opgave 16.45 maar nu voor $blok_n$ gedefinieerd door

$$blok_n i(x) = blip_n(x - \frac{1}{n}) blip_n(\pi - \frac{1}{n} - x),$$

[18] Google Heaviside.
[19] Wat is dat?

Opgave 16.46 laat zien dat $\chi_{(0,\pi)}$, opgevat als

Opgave 16.47. Het is goed om op een rijtje te zetten hoe je zeker weet dat elke $f \in L^2(-\pi,\pi)$ te benaderen is met een rij functies $f_1, f_2, \ldots$ in

$$C_c^\infty(-\pi,\pi),$$

de ruimte van van functies $f : (-\pi,\pi) \to \mathbb{R}$ die oneindig vaak differentieerbaar zijn en identiek nul zijn in de buurt van $x = 0$ en $x = 2\pi$. Benaderen betekent hier dat $f_n \to f$ in de 2-norm. Het speciale geval om eerst te begrijpen is

$$f(x) = \chi_I(x) = \begin{cases} 1 & \text{als } x \in I \\ 0; & \text{als } x \notin I, \end{cases}$$

met I een interval.

16.7 Intermezzo: out of Hilbertspace

De 2-norm is een bijzonder geval van

$$f \to |f|_p = \left(\int_{-\pi}^{\pi} |f(x)|^p \, dx \right)^{\frac{1}{p}},$$

waarmee voor $1 \leq p < \infty$ de p-norm op $C[-\pi,\pi]$ wordt gedefinieerd, en

$$|f|_\infty = \max_{x \in [-\pi,\pi]} |f(x)|,$$

de maximumnorm van f. Deze p-normen ($1 \leq p \leq \infty$) zijn te vergelijken met

$$|x|_p = \left(\sum_{j=1}^{\mathrm{N}} |x_j|^p \right)^{\frac{1}{p}},$$

de p-norm van $x = (x_1, \ldots, x_n) \in \mathbb{R}^{\mathrm{N}}$.

Opgave 16.48. Terug naar de overgeslagen calculussommetjes, bewijs (de ongelijkheid van Hölder)

$$|x \cdot y| \leq |x|_p |y|_q$$

voor $1 \leq p, q \leq \infty$ die voldoen aan

$$\frac{1}{p} + \frac{1}{q} = 1,$$

en $x = (x_1, \ldots, x_n)$ en $y = (y_1, \ldots, y_n)$ in $\mathbb{R}^{\mathrm{N}}$. Hint: leg eerst uit waarom het *geen* beperking is om aan te nemen dat $|x|_p = |y|_q = 1$.

Opgave 16.49. Bewijs dat $x \to |x|_p$ een norm is op $\mathbb{R}^{\mathrm{N}}$.

Opgave 16.50. Bewijs dat $|x|_p \to |x|_\infty$ als $p \to \infty$.

Opgave 16.51. Verzin en maak de analoge opgaven voor

$$f \to \left(\int_{-\pi}^{\pi} |f(x)|^p \, dx \right)^{\frac{1}{p}},$$

de p-norm op $C[-\pi, \pi]$.

17 A very partial reader for Olver's PDE book

The introductionary chapter contains the basic example of the Navier-Stokes system, which I discussed in Dutch for high school teachers here:

`http://www.few.vu.nl/~jhulshof/handoutNS.pdf`

The system in Olver's (1.4) is the unpacked version of

$$(\mathbf{NS}) \qquad u_t + (u \cdot \nabla)u + \nabla p = \nu \Delta u \quad \text{with} \quad \nabla \cdot u = 0,$$

in which the fluid velocity $u = u(t, x)$ is the unknown 3-vector valued function of x in a domain Ω in 3-dimensional real space, and of time t, and the unknown pressure $p = p(t, x)$ a scalar function. The positive parameter ν is called the kinematic viscosity, which differs from the dynamic viscosity μ by a factor ρ, the density of the fluid, which for water is effectively a constant. With ρ normalised to 1 we have $\mu = \nu$, but people do get annoyed if you write (**NS**) with μ like I did in that handout.

Anyway, with $\rho = 1$ the velocity u is a mass flux, often denoted by $q = \rho v$ if v is the velocity. In (**NS**) the velocity is called u. Many PDE solvers (persons who solve PDE's) prefer u as the unknown function in their PDE's, whatever u is. With constant density the second equation says that

$$\nabla \cdot q = 0,$$

which is a special case of the (mass) conservation law

$$(\mathbf{CL}) \qquad \rho_t + \nabla \cdot q = 0,$$

an equation you should have seen in the context of the treatment of the Gauss Divergence Theorem in your vector calculus course. Such conservation laws may be supplemented with various flux laws, e.g.

$$q = -\nabla \rho,$$

leading to the linear diffusion equation

$$\rho_t = \nabla \cdot (\nabla \rho) = (\nabla \cdot \nabla)\rho = \Delta \rho$$

for $\rho = \rho(t, x)$ with x in n-dimensional space and t a 1-dimensional time variable.

Another, example, in dimension $n = 1$, is

$$F(\rho) = \frac{1}{2}\rho^2 \quad \text{leading to} \quad \rho_t + \rho \rho_x = 0,$$

sometimes called the inviscid Burgers' equation, the viscous Burgers' equation being

$$\rho_t + \rho\rho_x = \nu\rho_{xx}.$$

With ρ replaced by u all these equation are treated in Olver's book (with $\gamma = \nu$ in case of the Burgers' equation). The linear diffusion equation is invariably called the heat equation for temperature $u = T$, derived from a conservation for thermal energy and a heat flux law. Also of interest, but not mentioned in the book is the Porous Medium Equation

$$\rho_t = \Delta\rho^\alpha,$$

first derived in the context of gas flow in porous media using a flux law of the form $q = -\nabla\rho^\alpha$ with $\alpha > 1$ a positive parameter.

With $\rho = u$ the viscous Burgers' equation looks like a 1-dimensional version of (**NS**) but physically this is nonsense of course, even before you observe that the second equation in (**NS**) reduces to $u_x = 0$: (**NS**) a system to avoid in dimension 1. The 2-dimensional case is of much more interest, as is explained in handoutNS.pdf which I recommend you to read. Not explained but implicitly used in handoutNS.pdf is that the solution $u = u(t, x)$, once known, defines a time dependent vector field, and thereby a system of Ordinary Differential Equations (ODE's)

$$\dot{x}(t) = u(t, x(t))$$

for particle trajectories $t \to x(t)$ in the (flow of) the fluid. For any (smooth) function $(t, x) \to F(t, x)$ such a trajectory defines a function

$$t \to F(t, x(t))$$

whose (time) derivative is given by

$$F_t(t, x(t)) + \sum_{i=1}^{n} F_{x_i}(t, x(t))\, \dot{x}_i(t) = F_t(t, x(t)) + \sum_{i=1}^{n} u_i(t)\, F_{x_i}(t, x(t)),$$

i.e. by

$$\frac{\partial}{\partial t} + u \cdot \nabla \quad \text{acting on} \quad F,$$

and $x(t)$ substituted for x in $u(t, x)$ and $F(t, x)$. The quantity

$$\frac{\partial F}{\partial t} + u \cdot \nabla F$$

is called the convective derivative of F given the velocity field $u = u(t, x)$, and with $F = u$ it is the acceleration field of the fluid flow. This explains

the first two terms in (**NS**), which relates the acceleration field to the forces acting on the fluid (in zero gravity): the gradient of the pressure and the viscous (friction) forces.

PDE's can be classified. They can be linear in u with a right hand side independent of u. Such equations are called inhomegenous linear PDE's (homegenous in case of zero right hand side) or nonlinear (if they are not linear). Linear PDE's allow superposition of solutions, much like your linear equations from linear algebra do, as well as linear ODE's.

The order of a PDE refers to the order of the highest order derivative in the PDE.

PDE's come with initial conditions at $t = 0$ for u if they are of the form u_t equals a right hand side without time derivatives of u, much like you have seen this for ODE's, also for higher order (in t) equations. For instance, the wave equation $u_{tt} = u_{xx}$ comes with initial conditions for u and u_t.

If the spatial variables are confined to a domain Ω typically also boundary conditions are needed. The ones that make sense to impose come from physical considerations. Examples: prescribing temperature or density at the boundary, or the normal component of flux (in- or outflow of heat), or the flux in terms of the boundary values of u (Newton's cooling law for instance). These conditions are named after respectively Dirichlet, Neumann and Robin for various reasons you may be interested in to know.

17.1 First order equations

Chapter 2 in Olver's book is about the Cauchy initial value problem for first order PDE's. I will concentrate on 2 very simple PDE's for $u = u(t, x)$:

$$u_t + c(x)u_x = 0 \quad \text{and} \quad u_t + c(u)u_x = 0,$$

for example with $c(x) = x$ and $c(u) = u$. Both are first order evolution equations for the unknown function $u = u(t, x)$. We shall mainly consider the Cauchy initial value problem in which the PDE is to be solved with initial data

$$u(0, x) = f(x)$$

given for all $x \in \mathbb{R}$. Is there a unique solution and what does it do as t varies? Think of a movie in the x, u-plane (which we can also run backwards).

Both PDE's have as a special case the PDE

$$u_t + cu_x = 0$$

in which $c \in \mathbb{R}$ is a constant. If f is a C^1-function (a continuously differentiable function), then clearly a solution of the Cauchy problem for this PDE is given by

$$u(t,x) = f(x - ct).$$

This is a *classical* solution, meaning that $u(t,x)$ and all the derivatives which appear in the equation are jointly continuous in t and x. You should be able to use your calculus skills to show that the *only* classical solution of $u_t + cu_x = 0$ is $u(t,x) = f(x - ct)$. Even when f is not C^1, the solution formula still makes perfect sense, so we may like to consider also non-classical solutions, in particular as limits of classical solutions u_n by taking a sequence of smooth functions f_n with $f_n \to f$. However below I assume that f is C^1.

The above solution is constant along the lines $x - ct = \xi$ where $\xi \in \mathbb{R}$ is an arbitrary constant. These lines are called characteristics along which the initial data propagate and these characteristics are solutions of

$$\dot{x} = \frac{dx}{dt} = c,$$

a simple ODE with solutions $x(t) = ct + \xi$, in which ξ is a constant of integration that coincides with the initial value for $x = x(t)$.

17.1.1 The method of characteristics

All 3 PDE's above are special cases of the PDE

$$u_t + c(x,u)u_x = h(x,u).$$

This PDE can be solved using the method of characteristics. Introduce $x = x(t)$ and use the PDE with $u = u(t) = u(t, x(t))$. You may prefer to be more strict in the notation here and write $X(t), U(t)$ rather than $x(t), u(t)$ if you get confused, but I will write $\dot{x}$ and $\dot{u}$ for the time derivatives of $x(t)$ and $u(t)$ to distinguish between $u(t)$ and $u(x,t)$. The result is that we should have

$$\dot{u} = u_t + u_x\dot{x} = h(x,u) + (-c(x,u) + \dot{x})u_x.$$

If we impose that $\dot{x} = c(x,u)$ we conclude that we must have $\dot{u} = h(x,u)$. An ODE system to study in relation to the PDE is therefore

$$\frac{dx}{dt} = c(x,u), \quad \frac{du}{dt} = h(x,u).$$

Systems of this type are studied in every ODE course so you know that we can solve this system for given initial values of x and u at $t = 0$ if the

coefficients $c(x,u)$ and $h(x,u)$ are C^1. You also know that these systems come with many possible behaviours of solutions. Solutions may run away to infinity in finite time, they may converge to equilibrium, they may spiral, they may become periodic and what have you.

In relation to the Cauchy problem for the PDE we have to consider all solutions of the ODE system which start on the graph of the initial data. Thus, as t evolves we see the graph deform, since each point on the graph starts to move around in the x,u-plane following the flow of the ODE system. As long as the deformed graph is a graph without vertical tangent lines, we should have a unique classical solution of the Cauchy initial value problem for the PDE, but a priori there is no reason why this graph property should be preserved. Indeed, the second example, with $h \equiv 0$ and $c = c(u)$ makes that points move only in the horizontal x-direction, with a velocity that depends on the height u, so pulse/bell type initial data will always lose the graph property and give rise to multivalued solutions.

The first example, with $h \equiv 0$ and $c = c(x)$ does not have this feature, but depending on the solutions $x = x(t)$ (*the characteristics*) of the first order ODE

$$\frac{dx}{dt} = c(x),$$

there is a large variety of behaviours of solutions of the Cauchy problem for the PDE. One can solve this ODE using a primitive of $\frac{1}{c(x)}$ (if it exists) and a constant of integration, in the book denoted by ξ. WARNING: in general this ξ is *not* the initial value of $x(t)$ though sometimes it is. Olver denotes the initial value sometimes by y and sometimes by k, see Fig. 2.6 and below. Observe that a qualitative analysis (which most of you are familiar with, see also Section 10.4) of the first order ODE for $x = x(t)$ based on the graph of the function $x \to c(x)$ already tells most of the story on the solution $u(t,x)$!

In this summary I have restricted the attention to autonomous PDE's: the time variable t does not occur explicitly in the examples above. Thus it is no restriction to assume that the initial data for $u(t,x)$ are given at $t = 0$. Once we are familiar with the solution methods for the Cauchy initial value problem for such autonomous PDE's we can also think of nonautonomous PDE's and other boundary conditions, say prescribing u on some curve in the t,x-plane, or on both of the positive coordinate axis.

Finally we note that the ODE system above in itself does not distinguish between the roles of x and u. On any point of the solution graph of u we can locally write $u = u(t,x)$ or $x = x(t,u)$. The PDE for $u(t,x)$ transforms to a PDE for $x = x(t,u)$. Below are some additional considerations in which I use this trick, with a more careful notation, to see what one can and should do if one insists on looking at the second PDE, e.g. with $c(u) = u$, as a (physical)

conversation law. This leads to classical solutions developing shocks which propogate obeying a so-called Rankine-Hugoniot condition. Such shocks have the property that in forward time our happy characteristics can only go into a shock and never come out of a shock. This property then also appears as an imposed *entropy* condition for the construction of a unique *entropy* solution starting from piecewise continuous initial data.

17.1.2 Shocks, mass conservation, Rankine-Hugoniot condition

Refering to Olver's book again consider Figure 2.16 where a solution of (2.31/2.48) constructed using the method of characteristics has become multivalued. Choose $x = a$ to the left and $x = b$ to the right of the multivalued part. The first point to make is that the integral

$$M(t) = M_{a,b}(t) = \int_a^b u(x,t)dx,$$

redefined in the appropriate sense after the solution has become multi-valued, still satisfies

$$\frac{dM}{dt} = \frac{1}{2}u(t,a)^2 - \frac{1}{2}u(t,b)^2.$$

Redefine M as the area of the region bounded by the solution curve and the lines $x = a$, $x = b$ and $u = 0$. We can write M as an integral with respect to u by inverting:

$$u = U(t,x) \quad \Leftrightarrow \quad x = X(t,u).$$

Momentarily I write upper cases for the unknown (possibly multivalued) solutions.

Exercise JH1. Check directly that

$$U_t + UU_x = 0 \quad \Leftrightarrow \quad X_t = u.$$

You already knew this from the method of characteristics of course.

The equation on the right is an easy PDE for $X(t,u)$, with initial data given by the possibly multivalued inverse function g of f as it appears in:

$$U(0,x) = f(x) \quad \Leftrightarrow \quad X(0,u) = g(u).$$

The solution is trivially given by

$$X(t,u) = g(u) + tu,$$

so it is easy to see that the multivaluedness of $g(u)$ is inherited by $X(t,u)$, and that different laps of the graph $u = U(t,x)$ do not overtake one another.

Below we reason from the assumption that $f(x)$ is positive, decreasing and continuous.

The redefined $M(t) = M_{a,b}(t)$ is given by

$$M = \int_{U(t,b)}^{U(t,a)} (X(t,u) - a)du + (b-a)U(b,t),$$

from which one easily recovers the above ODE for M by differentiating with respect to t. We then ask for a shock solution as indicated in Fig. 2.16 by writing an equation for the location of the vertical segment $x = \sigma(t)$ and asking the two areas to be the same. Why? Because such a shock solution has the same $M(t)$ as the multivalued solution. The equal area condition gives an equation involving the function $g(u)$, which supplements the two equations for the upper and lower values $u^-(t)$ and $u^+(t)$ at the shock:

$$g(u^-(t)) + tu^-(t) = g(u^+(t)) + tu^+(t) = \sigma(t).$$

Exercise JH2. Derive this third equation and verify (2.53) by applying the implicit function theorem. Then do the same for (2.64) and (2.59), replacing u by $c(u)$ and $\frac{u^2}{2}$ by a primitive $C(u)$ of $c(u)$.

Now the equation you found for $\dot{\sigma}(t)$ holds after the last moment $t = t_*$ that

$$x = g(u) + tc(u)$$

is invertible as $u = u(t,x)$, that is, as $u \to g(u) + tc(u)$ is no longer strictly monotone in u. Assuming that both $c(u)$ and $c'(u)$ are positive for the range of u-values given by $u = f(x)$ with x ranging over the real line, and $g'(u) < 0$ for those values, $t = t_*$ is characterised as the first t for which $g'(u)+tc'(u) = 0$ has a solution, say $u = u_*$, and $x = \sigma_* = g(u_*) + t_*c(u_*)$ in the point on the real line where the shock appears. If you understand the Implicit Function Theorem[1] here is a neat application for it:

17.1.3 Appearance of the shock

For $t > t_*$ but close to t_* there should be a small range of σ near σ^* for which $g(u) + tc(u) = \sigma$ has 3 solutions u in such a way that the equal area rule is satisfied for one particular σ in this small range. We write $t = t_* + \epsilon^2$ with ϵ small and try to solve for the 3 solutions u first. The reason to put $t = t_* + \epsilon^2$ and not $t = t_* + \epsilon$ is because the latter will lead to a sharp distinction between positive and negative ϵ, which makes the Implicit Function Theorem unlikely

[1] See Section 6.6.

to be applicable in $\epsilon = 0$. Since at $t = t_*$ the level u_* propagates with speed $c(u_*)$ to the right, the σ-range for the analysis to be carried out will be near $\sigma_* + c(u_*)\epsilon^2$, so it makes sense to put $\sigma = \sigma_* + c(u_*)\epsilon^2 + \nu\epsilon^3$ with ν a new unknown replacing σ. Finally we set $u = u_* + \epsilon v$.

Dividing the resulting equation by ϵ^3 you should get an equation for v in terms of ϵ and ν that replaces $g(u) + tc(u) = \sigma$ for u, and that has 3 distinct solutions for $\epsilon = \nu = 0$ around all 3 of which the Implicit Function Theorem can be applied to obtain 3 solutions $\underline{v} < v_0 < \overline{v}$ as implicit functions of ϵ and ν. The equal area rule to impose then becomes an integral from $\underline{v}$ to $\overline{v}$ which defines a new function to which the Implicit Function Theorem can be applied to find ν as an implicit function of ϵ, and thereby also the values of u at the top and bottom of the shock in terms of an expansion in ϵ.

17.1.4 ODE-system for the shock

From $g(u^-) + tc(u^-) = g(u^+) + tc(u^+) = \sigma$ and the equal area condition it follows, with Exercise JH2, that

$$(u^+ - u^-)\dot{\sigma} = C(u^+) - C(u^-), \quad (g'(u^-) + tc'(u^-))\dot{u}^- + c(u^-) = \dot{\sigma},$$

$$(g'(u^+) + tc'(u^+))\dot{u}^+ + c(u^+) = \dot{\sigma},$$

so that u^+, u^- solve

$$(g'(u^+) + tc'(u^+))\dot{u}^+ = \frac{C(u^+) - C(u^+)}{u^+ - u^-} - c(u^+)$$

$$= \frac{1}{u^+ - u^-}\int_{u^-}^{u^+} (c(s) - c(u^+))ds < 0,$$

$$(g'(u^-) + tc'(u^-))\dot{u}^- = \frac{C(u^+) - C(u^+)}{u^+ - u^-} - c(u^-)$$

$$= \frac{1}{u^+ - u^-}\int_{u^-}^{u^+} (c(s) - c(u^-))ds < 0,$$

if $c(u)$ is increasing in u. It's a very nice ODE exercise to examine how solutions $(u^-(t), u^+(t))$ of the system start from local minima of

$$t = -\frac{g'(u)}{c'(u)}$$

in the (t, u)-plane, and subsequently define $\sigma(t)$ via the shock condition, i.e. the equation for $\dot{\sigma}$. This produces curves

$$t \to (\sigma(t), u^+(t)) \quad \text{and} \quad t \to (\sigma(t), u^+(t))$$

in the (x,u)-plane from which we can trace back the solution of the PDE to $t=0$, thereby recovering the initial data.

For $c(u)=u$ the equations simplify to

$$(g'(u^+)+t)\dot{u}^+ = \frac{u^- - u^+}{2},$$

$$(g'(u^-)+t)\dot{u}^- = \frac{u^+ - u^-}{2}.$$

I did the example where the initial data are

$$f(x)=\frac{1}{1+x^2}, \quad g(u)=\sqrt{\frac{1}{u}-1}, \quad u_*=\frac{3}{4}, \quad x_*=\sqrt{3}, \quad t_*=\frac{8}{3\sqrt{3}},$$

put

$$t=\frac{8}{3\sqrt{3}}(1+s), \quad u^\pm=\frac{3+w^\pm}{4},$$

introduced

$$\phi(w)=(1+\frac{w}{3})^{\frac{3}{2}}(1+w)^{\frac{1}{2}}=\frac{1}{3}w^2+\frac{4}{27}w^3+\cdots$$

and arrived at

$$(s-\phi(w^+))\frac{dw^+}{ds}=\frac{w^- - w^+}{2}, \quad (s-\phi(w^-))\frac{dw^-}{ds}=\frac{w^+ - w^-}{2}.$$

Then I put $s=\tau^2$, $w^\pm=\tau z^\pm$ to get

$$(1-\frac{\phi(\tau z^+)}{\tau^2})(z^+ + \tau\frac{dz^+}{d\tau}) = z^- - z^+,$$

$$(1-\frac{\phi(\tau z^-)}{\tau^2})(z^- + \tau\frac{dz^-}{d\tau}) = z^+ - z^-,$$

in which $\tau=e^t$ (a new t) put this system into a form which to leading order (small $z^\pm$) is linear with a saddle point structure.

17.2 The one-dimensional wave equation

The last section of Chapter 2 concerns the one-dimensional inhomogeneous linear wave equation

$$u_{tt}-c^2u_{xx}=F(x,t)$$

which is solved explicitly in terms of its initial data for u and u_x at $t=0$. In the case that the initial data as well as F are 2π-periodic in x, the main result,

Theorem 2.18, must turn out to be consistent with the Fourier series approach in Chapters 3 and 4. Read and do the derivation of the homogeneous case ($F \equiv 0$) in Theorem 2.15 yourself. It shows how the solutions is constructed form signals to propagate along forward and backward characteristics. Here we will not attempt to take the concept of characteristics much further than that discovered in (2.71) and (2.74) for $u_{tt} - c^2 u_{xx} = 0$.

17.3 Fourier series

Olver's Chapter 3 is of independent interest, see also Hoofdstuk 16. Fourier series are not just important in the context op PDE's, but an introduction from PDE perspective is most appropriate. Section 3.1 rediscusses some of Chapter 1 and then derives (3.27) as the general form for 2π-periodic (in x) solutions $u(t, x)$ of the one-dimensional heat equation with unit diffusion coeffient:

$$u_t = u_{xx}$$

New and very important in relation to Chapter 1 is the eigenvalue perspective comparing (3.1) and (3.14), culminating in the table on page 68 and, in the 2π-periodic context, (3.27). Note that (3.18) is a solution of *separated variables*, the topic of Chapter 4.

The initial data of the solution in (3.27) is $f(x)$ given by (3.28). Section 3.2 explains how the left hand side $f(x)$ and the Fourier series on the right hand side correspond to one another. The starting point is (3.34) and (3.35) with

$$f(x) \sim \frac{a_0}{2} + \sum \cdots$$

This notation is introduced because a priori we do not know that or in what sense

$$f(x) = \frac{a_0}{2} + \sum \cdots,$$

except when f is a trigonometric polynomial, see Exercise 3.2.4.

Note that we interested in both directions. If we compute solutions $u(t, x)$ we find t-dependent coefficients of the solution and ask how well the solution is defined and how smooth it is. If we fit the general solution $u(t, x)$ to initial data f we do this by fitting the coefficients to the coefficients computed from f.

There are two equivalent forms of the Fourier series, real and complex. The real cos/sin form of the Fourier series allows a nice and useful distinction between even and odd functions $f(x)$: Prop. 3.14. The complex form (3.64) allows for smoother formula's and proofs and a much smoother transition

to the Fourier integral transform in Chapter 7. Both forms are intrinsically related to a suitable inner product for functions: (3.30)/(3.61).

One can prove that C^1 2π-periodic functions f are sums of *uniformly convergent* Fourier series computed through (3.35). A stronger localised statement is given in Thm 3.30. The uniform convergence result remains valid if $f'(x)$ has finitely many jump discontinuities (corners in the graph of f) on its interval of periodicity. However, if f itself has jump discontinuities the convergence *cannot be uniform*. It still holds pointwise, but only under the strong assumption that f' is piecewise continuous: Theorem 3.8, proved at the very end of Section 3.5. A slight variant of that proof, using the mean value theorem, shows the convergence is indeed uniform if f is continuous and f' is (piecewise) continuous, see e.g. The Way of Analysis by Strickarzt, or exercise JH4 below.

Remark: jump discontinuities may be removed by subtracting suitably scaled and shifted saw tooth functions. The resulting continuous piecewise C^1-function has a uniformly convergent Fourier series. For saw tooths we can examine the behaviour of their Fourier series by direct calculations which are somewhat reminiscent of the calculations in the convergence proof, and which also clarify the Gibbs overshoot phenomenon illustrated in Figure 3.7. The nicest saw tooth to illustrate what's going on is the odd 2π-periodic extension of

$$\pi - x = \frac{2\sin x}{1} + \frac{2\sin 2x}{2} + \frac{2\sin 3x}{3} + \frac{2\sin 4x}{4} + \frac{2\sin 5x}{5} + \dots \quad (0 < x < \pi).$$

If you differentiate the right hand side you get an expression which diverges! Its truncation after n terms is easily related to the outcome of (3.128) in the convergence proof.

Exercise JH3. Integrate the resulting expression to conclude that

$$\frac{2\sin x}{1} + \dots + \frac{2\sin nx}{n} = \int_0^x D_n(s)ds - x, \quad D_n(s) = \frac{\sin(n+\frac{1}{2})s}{\sin\frac{s}{2}}$$

This D_n is called the Dirichlet kernel. Examine the integral using the n-dependent scaling $y = (n+\frac{1}{2})x$ (and likewise for s). Identify π as a limit of the integral for $n \to \infty$ when $0 < x \leq \pi$ and

$$2\int_0^\pi \frac{\sin t}{t}dt > \pi = 2\int_0^\infty \frac{\sin t}{t}dt$$

as the limit of the largest maximum of $\int_0^x D_n(s)ds$, which occurs in $\frac{\pi}{n+\frac{1}{2}}$.

Basis general facts about uniform convergence are Thm 3.26 and 3.27 (which implies Thm 3.29), Prop. 3.28, and the discussion about interchanging sums and integrals just above Prop. 3.28. I will assume these facts

are known. Not mentioned here in the book is that uniform convergence, illustrated in Fig. 3.11, also comes with a norm, called the maximum or supremum norm,

$$||f||_m = \sup_x |f(x)|,$$

which is certainly defined if f is 2π-periodic and piecewise continuous in the sense of Def. 3.6. Note that the inner product norm is controlled by the maximum norm: if the truncation error goes to zero in the maximum norm, it certainly goes to zero in the inner product norm (but not the other way around!). It is in general not true that the partial sums s_n of the Fourier series of a 2π-periodic continuous function f satisfy

$$||s_n - f||_m \to 0,$$

as it is not even clear that $s_n(x) \to f(x)$ pointwise.

Section 3.5 should be studied in mathematical detail. It discusses not only uniform convergence, and pointwise convergence, but also convergence in the mean. The latter is by definition equivalent to convergence in the inner product norm (3.102) and allows one to think of and work with 2π-periodic functions (for which the inner product norm is defined) as column vectors, and the Fourier modes as unit base vectors. Think of straight angles and the Pythagorean theorem here, we will see expressions like

$$1 + \frac{1}{4} + \frac{1}{9} + \frac{1}{16} + \cdots = \frac{\pi^2}{6}.$$

The relevant formula's are given in the complex notation (3.124), the real version is discussed in 3.5.38. The proof relies on the essential observation made in Thm 3.36 that in terms of the inner product norm the n-th partial sum s_n of the Fourier series is the best trigonometric approximation of degree n to f, and that the error goes to zero as $n \to \infty$, which amounts to convergence in the mean.

Convergence in the mean is in fact equivalent to the first Plancherel (Pythagoras) formula in (3.124). On page 115 the author avoids the consideration of different norms with a direct proof of the Plancherel formula for the sum function of a uniformly convergent Fourier series. For general square integrable functions the Plancherel formula then follows by approximation arguments (the details are not given in the book). Thus, although for continuous functions even pointwise convergence may fail, as complicated examples show, one can be content that convergence in the mean always holds.

Section 3.2 emphasises that Fourier series are *not* like power series. Therefore Section 3.3 on what is allowed in differentiation and integration of

Fourier series should be read with care. The statements are only about the coefficients, not about convergence of the Fourier series, and they follow from integration by parts. Indeed, the formula's for the Fourier coefficients can be integrated by parts whenever $f'(x)$ exists, giving similar formula's with prefactors $\frac{1}{k}$ and thus better (and faster) convergence of the Fourier series than expected from the defining formula's. The smoother f, the more times we can integrate by parts and gain a prefactor $\frac{1}{k}$, and thus the better and faster the convergence of the Fourier series. However, for uniform convergence we need (3.99) with $\alpha > 1$, which requires more than one derivative for f, so this trick cannot compete with the direct convergence proof which only needs the first order derivative.

Nevertheless, the larger we can choose n in (3.100), the smoother the sum function of the Fourier series: Thm 3.31. Solving PDE's, we compute Fourier series solutions

$$u(t,x) \sim \sum \dots$$

Solutions of the heat equation $u_t = u_{xx}$ are easily seen to be very well behaved in this respect as soon as $t > 0$, even with ugly initial data, because of the exponentials in (3.27). However, interesting (fractal) issues with respect to lack of smoothness will appear in Chapter 7 when we solve $u_t = u_{xxx}$ with e.g. piecewise constant initial data.

Read also Section 3.4 yourself. It discusses simple changes of scale needed to deal with l-periodic functions $f(x)$, $l > 0$. The limit $l \to \infty$ leads to the Fourier integral transform in Chapter 7.

Exercise JH4. Assume that f and f' are piecewise continuous, so that f' is bounded by a fixed constant M. Show that the function $g(y)$ defined immediately after (3.131) is bounded in terms of M. Then split the integral of $g(y)\sin(n+\frac{1}{2})y$ in $\int_0^\delta$ and $\int_\delta^\pi$. The first is bounded in terms of δ and M. Integrate the second one by parts and estimate in terms of δ, M and n which appears in the denominator. Conclude the pointwise convergence proof without using the Riemann-Lebesgue lemma and show that the convergence is uniform if f is continuous and f' is (piecewise) continuous.

17.4 The integral Fourier transform

Next I slightly modify the presentation in 7.1 of Olver. Let $f = f(x)$ be defined on the real line, smooth and $f(x) = 0$ for $|x| \geq l$. Then the (uniformly convergent) Fourier series of f on the interval $[-l, l]$ is obtained via scaling:

$$F(y) = f(x), \quad \frac{x}{l} = \frac{y}{\pi}.$$

We have

$$f(x) = F(y) = \sum_{n=-\infty}^{\infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} F(\eta) e^{-in\eta} d\eta \, e^{iny}$$

$$= \frac{1}{\sqrt{2\pi}} \sum_{n=-\infty}^{\infty} \frac{\pi}{l} \underbrace{\frac{1}{\sqrt{2\pi}} \int_{-l}^{l} f(\xi) e^{-i\frac{n\pi}{l}\xi} d\xi}_{\widehat{f}(\frac{n\pi}{l}) = \widehat{f}(k_n)} \, e^{i\frac{n\pi}{l}x}.$$

Set

$$k_n = n\Delta k, \quad \Delta k = \frac{\pi}{l},$$

and define

$$\widehat{f}(k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{-ikx} dx,$$

then

$$f(x) = \frac{1}{\sqrt{2\pi}} \sum_{n=-\infty}^{\infty} \widehat{f}(k_n) e^{ik_n x} \, \Delta k.$$

In other words, $f(x)$ is equal to a uniformly convergent sum, which actually is a Riemann sum for

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \widehat{f}(k) e^{ikx} dk,$$

and likewise

$$\int_{-\infty}^{\infty} |f(x)|^2 dx = \int_{-l}^{l} |f(x)|^2 dx = \sum_{n=-\infty}^{\infty} |\widehat{f}(k_n)|^2 \, \Delta k,$$

which is a Riemann sum for

$$\int_{-\infty}^{\infty} |\widehat{f}(k)|^2 dk.$$

Note that these identities remain valid if we increase l. In particular both Riemann sums are independent of Δk in the limit $\Delta k \to 0$.

Can we conclude that both $f(x)$ and $\int |f(x)|^2$ are also equal to the integrals themselves? The answer will be yes if $\widehat{f}(k)$ is continuous and decays sufficiently fast as $|k| \to \infty$, to make the tails of both integrals and Riemann sums small. Then we can restrict the convergence argument to integrals and Riemann sums on bounded k-intervals. Since

$$\widehat{f}(k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{-ikx} dx =$$

$$\frac{1}{\sqrt{2\pi}}\left(\left[f(x)\frac{e^{-ikx}}{-ik}\right]_{-\infty}^{\infty}-\int_{-\infty}^{\infty}f'(x)\frac{e^{-ikx}}{-ik}dx\right)=\frac{\widehat{f'}(k)}{ik}=\frac{\widehat{f^{(m)}}(k)}{(ik)^m}$$

and

$$|\widehat{f^{(m)}}(k)|\leq 2l\max_{-l\leq x\leq l}|f^{(m)}(x)|,$$

we can conclude that indeed

$$f(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\widehat{f}(k)e^{ikx}dk,\quad \int_{-\infty}^{\infty}|f(x)|^2dx=\int_{-\infty}^{\infty}|\widehat{f}(k)|^2dk$$

for all f in C_c^∞, the class of smooth compactly supported functions.

Denoting the class of measurable complex valued functions f with bounded $\int|f(x)|dx$ by L^1 and the class with bounded $\int|f(x)|^2dx$ by L^2 we have that C_c^∞ is dense in L^2, and that the map

$$\mathcal{F}:C_c^\infty\to L^1\cap L^2\cap C^\infty\subset L^2,\quad \mathcal{F}(f)=\widehat{f}$$

uniquely extends to an isometry

$$\mathcal{F}:L^2\to L^2.$$

Only if $f\in L^1\cap L^2$ we can write $\widehat{f}(k)$ as an absolutely convergent integral. The inverse map is the same, up to a reflection in x, that is, for $f=f(k)$,

$$(\mathcal{F}^{-1})f(x)=\widehat{f}(-x).$$

More to be done here from a different perspective (Folland' s Real Analysis book).

17.5 The fast Fourier transform

These are some notes I prepared for a class at AUC that I took over once from JB. Not related to Olver's book, but of the same flavor as the section just above. Condsider an L-periodic function $f=f(t)$. Write

$$f(t)=\sum_{k=-\infty}^{\infty}c_ke^{\frac{2\pi kit}{L}},\quad c_k=\frac{1}{L}\int_0^L f(t)e^{-\frac{2\pi kit}{L}}dt$$

If we only know $f(t)$ in the points $t=0,\frac{L}{N},\frac{2L}{N},\frac{3L}{N},\dots$ then the obvious approximation for c_k is

$$\tilde{c}_k=\frac{1}{L}\sum_{j=0}^{N-1}\frac{L}{N}f(\frac{jL}{N})e^{-\frac{2\pi ki\frac{jL}{N}}{L}}=\frac{1}{N}\sum_{j=0}^{N-1}f(\frac{jL}{N})e^{-\frac{2\pi jki}{N}}$$

This involves precisely N samples of f.

An approximation of $f(t)$ with c_k replaced by $\tilde{c}_k$ does not make sense since the $\tilde{c}_k$ are N-periodic in k. Therefore

$$f(t) \not\approx \sum_{k=-\infty}^{\infty} \tilde{c}_k e^{\frac{2\pi kit}{L}}$$

since the latter is not defined. However, the natural finite sum with c_k replaced by $\tilde{c}_k$ does a perfect job in $t = \frac{mL}{N}$:

$$\sum_{k=0}^{N-1} \tilde{c}_k e^{\frac{2\pi kim\frac{L}{N}}{L}} = \sum_{k=0}^{N-1} \frac{1}{N} \sum_{j=0}^{N-1} f(\frac{jL}{N}) e^{-\frac{2\pi jki}{N}} e^{\frac{2\pi kmi}{N}} = f(\frac{mL}{N})$$

This follows by changing the order of summation and the fact that

$$\sum_{k=0}^{N-1} e^{-\frac{2\pi jki}{N}} e^{\frac{2\pi kmi}{N}} = N\delta_{jm}$$

NB. Summing over a sequence k symmetric around $k = 0$ would perhaps look more natural in terms of the usual convergence results for

$$\sum_{k=-n}^{n} c_k e^{\frac{2\pi kit}{L}} \to f(t),$$

as $n \to \infty$. But for the sample points there is no difference.

For MatLab reasons we number the values of f in $t = 0, \frac{L}{N}, \frac{2L}{N}, \frac{3L}{N}, \ldots$ as $f_1, f_2, \ldots, f_N$ and write $F_k = N\tilde{c}_{k-1}$. The relation between the N-vectors F and f is then given by

$$F = \Omega f, \quad \Omega_{jk} = \omega^{(j-1)(k-1)} \quad (j, k = 1, \ldots, N), \quad \omega = e^{-\frac{2\pi i}{N}}$$

To go from f to F thus takes N^2 multiplications and additions, but this can be improved!

First note that if $N = 2n$ is even, we can set

$$\tilde{\omega} = \omega^2, \quad \tilde{\Omega}_{jk} = \tilde{\omega}^{(j-1)(k-1)} \quad (j, k = 1, \ldots, n)$$

and decompose the vector f in $f_o = (f_1, f_3, \ldots, f_{N-1})$ and $f_e = (f_2, f_4, \ldots, f_N)$. Reshuffling the rows of Ω accordingly we find for $F^+ = (F_1, F_2, \ldots, F_n)$ and $F^- = (F_{n+1}, F_{n+2}, \ldots, F_{2n})$ that

$$F^+ = \tilde{\Omega} f_o + D\tilde{\Omega} f_e, \quad F^- = \tilde{\Omega} f_o - D\tilde{\Omega} f_e,$$

in which D is a diagonal matrix with entries ω^{j-1} with $j = 1, \ldots, n$. We used $\omega^n = -1$ here.

If $N = 2^p$ then we can use this reduction to recursively compute F in p steps, each of which contains approximately 2^p operations. This is because the lenght of the vectors involved in each step times the number of vectors involved in each step remains the same. This way the order of the number of computations required for F is about $N \log_2 N$.

18 Sobolev spaces

This is an overview of how we deal with the Sobolev spaces

$$W^{1,p}(U) = \{u \in L^p(U) : \frac{\partial u}{\partial x_i} = u_{x_i} \in L^p(U),\, i = 1 \dots N\},$$

in which U is an open connected subset of $\mathbb{R}^N$, and in which the partial derivatives are taken in the distributional sense. I'm adapting the approach in Chapter 5 of Evans' PDE book. Restricting the attention to first order weak derivatives to begin with, I will turn Chapter 5 around and begin with the main results for the space $W_0^{1,p}(U)$ that is defined as the closure of the $C_c^1(U)$, the closure taken in $W^{1,p}(U)$ with respect to the obvious norm, and the main result being the (compact) embeddings in spaces of functions which less regularity. This will involve only the additional assumption that U is bounded.

To extend the results for $W_0^{1,p}(U)$ to $W^{1,p}(U)$ we need a well behaved extension operator that maps $W^{1,p}(U)$ into $W_0^{1,p}(\tilde{U})$ with $\tilde{U}$ only slightly larger than U. This extension operator is first defined for $u \in C^1(\bar{U})$ and requires the boundary ∂U to be bounded and C^1 (locally the graph of a C^1-function). Partitions of unity play a crucial role here, and they are are also used to establish that $W^{1,p}(U)$ itself is the closure of $C^1(\bar{U})$ if ∂U is bounded and C^1. The elements of the partition are suitably chosen $\zeta_i \in C_c^\infty(\mathbb{R}^N)$ which, when multiplied by $u \in W^{1,p}(U)$, produce products $\zeta_i u \in W^{1,p}(U)$ to which the Leibniz rule applies[1].

A basic tool is mollifying nonsmooth functions by convolution with non-negative compactly supported smooth functions η_ϵ which approximate the δ-function.

18.1 Sobolev spaces

For $1 \leq p < \infty$ the Sobolev norm is defined by

$$|u|_{1,p}^p = |u|_p^p + |u_{x_1}|_p^p + \cdots + |u_{x_N}|_p^p, \tag{18.1}$$

and it makes $W^{1,p}(U)$ a Banach space. For $p = \infty$ one takes the maximum of the ∞-norms of $u, u_{x_1}, \dots, u_{x_N}$. Which functions are in $W^{1,p}(U)$? Well, certainly the functions in $C_c^1(U)$, the space of continuously differentiable functions with compact support in U. This space also acts as the space of test functions for verifying the definition of the weak first order partial derivatives

[1] Evans' stronger assumption $\zeta \in C_c^\infty(U)$ forces us into some cumbersome details.

of $u \in W^{1,p}(U)$. Note that every $u \in C_c^1(U)$ extends to a $u \in C_c^1(\mathbb{R}^N)$ by setting u equal to zero outside U.

The closure of $C_c^1(U)$ in $W^{1,p}(U)$ is a Banach space, denoted by

$$W_0^{1,p}(U),$$

a definition which makes sense only for $1 \le p < \infty$.

18.2 Estimates and embeddings for $W_0^{1,p}(U)$

Estimates derived for functions in $C_c^1(U)$ carry over to functions in $W_0^{1,p}(U)$, and I derived two basic estimates to which this principle is applied. The first (Gagliardo-Nirenberg-Sobolev) estimate was

$$|u|_q \le C_{p,N} |\nabla u|_p \quad \text{for} \quad \frac{1}{q} = \frac{1}{p} - \frac{1}{N} \quad \text{if} \quad 1 \le p < N, \tag{18.2}$$

in which the norm of ∇u is evaluated via an integral over the whole of U, and over the whole of $\mathbb{R}^N$ in the derivation, via repeated application of the one-dimensional estimate

$$|u|_\infty \le \frac{1}{2} |u|_1 \quad \text{for} \quad u \in C_c^1(\mathbb{R}),$$

first[2] in the special case that $p = 1$, with a clever use of the Hölder's inequality with exponents satisfying

$$\frac{1}{p_1} + \cdots + \frac{1}{p_{n-1}} = 1.$$

The general case in (18.2) followed from putting u^γ for u and a follow your nose estimate invoking Hölder's inequality for the integral of $\gamma |u|^{\gamma-1} u_{x_i}$, which involved a particular choice of γ to get the exponents right. The constant $C_{p,N}$ blows up as $p \to N$ (from below). For the p-norm of ∇u the p-norm of any norm of the vectors $\nabla u(x)$ can be used.

The second (Morrey) estimate[3] is

$$|u(x_1) - u(x_2)| \le C_{p,N} |\nabla u|_p |x_1 - x_2|^\alpha \quad \text{for} \quad \alpha = 1 - \frac{N}{p} \quad \text{if} \quad p > N, \tag{18.3}$$

the p-norm of ∇u defined via an integral over the intersection of the two balls centered in x_1 and x_2 with radius $|x_1 - x_2|$. This estimate was derived from the inequality

$$\int_{C_R} |u - u(0)| \le \frac{R^N}{N} \int_{C_R} \frac{|u_r|}{r^{N-1}}$$

[2] See Section 18.9.

[3] See Section 18.8.

for cones described in polar coordinates as

$$C_R = \{x = r\omega : 0 \leq r \leq R,\, \omega \in A\},$$

with A a nice subset of the unit sphere, and u_r denoting the radial derivative. The integral on the right hand side was estimated using a follow your nose estimate invoking Hölder's inequality, and (18.3) was proved estimating

$$|u(x_1) - u(x_2)| \leq |u(x_1) - u(x)| + |u(x) - u(x_2)|,$$

and integrating over the intersection of the two cones C_1 and C_2 centered in x_1 and x_2, chosen to have $C_1 \cup C_2$ equal to the union of the two balls mentioned earlier. Again the constant $C_{p,\mathrm{N}}$ blows up as $p \to \mathrm{N}$ (from above).

Opgave 18.1. Let $1 \leq p < \mathrm{N}$. Prove that $W_0^{1,p}(U) \subset L^q(U)$ if $\frac{1}{q} \geq \frac{1}{p} - \frac{1}{\mathrm{N}}$ if U is bounded, and that in that case $|\nabla u|_p$ defines an equivalent norm on $W_0^{1,p}(U)$.

Opgave 18.2. Let $p > \mathrm{N}$. Prove that $W_0^{1,p}(U) \subset C^\alpha(U)$ for $\alpha = 1 - \frac{\mathrm{N}}{p}$, in which

$$C^\alpha(U) = \{u \in C(U) : [u]_\alpha < \infty\} \quad \text{where} \quad [u]_\alpha = \sup_{x_1 \neq x_2} \frac{|u(x_1) - u(x_2)|}{|x_1 - x_2|^\alpha},$$

the supremum taken over the whole of U.

Opgave 18.3. Show that for U bounded and $\alpha \in (0, 1]$, every $u \in C^\alpha(U)$ extends to a continuous function on $\bar{U}$, and that the space $C^\alpha(U)$ is a Banach space with norm defined by $||u||_\alpha = |u|_\infty + [u]_\alpha$.

18.3 Compactness for $W_0^{1,p}(U)$

The Ascoli-Arzelà theorem states for compact metric spaces X that a sequence u_n in $C(X)$, the space of $\mathbb{R}$-valued continuous functions on X with norm

$$|u|_\infty = \max_{x \in X} |u(x)|,$$

has a convergent subsequence u_{n_k}, provided the sequence is bounded in $C(X)$ and has the property[4] that

$$\forall \epsilon > 0\, \exists \delta > 0\, \forall n \in \mathbb{N}\, \forall x, y \in X : d(x, y) \leq \delta \implies |f_n(x) - f_n(y)| \leq \epsilon.$$

Here convergence means uniform convergence.

[4] This is called the (uniform) equicontinuity of the sequence u_n.

Opgave 18.4. Let $p > \mathrm{N}$, U bounded, $\alpha = 1 - \frac{\mathrm{N}}{p}$. Use the Ascoli-Arzelà theorem to prove that every bounded sequence u_n in $W_0^{1,p}(U)$ has a subsequence that, considered as a sequence in $C(\bar{U})$, converges uniformly to a limit u, which is also in $C_0^\alpha(U)$, the subspace consisting of functions $u \in C^\alpha(U)$ which vanish on ∂U. Verify that this subspace has the seminorm $[\cdot]_\alpha$ as an equivalent norm.

In view of Exercise 18.4 the embedding

$$W_0^{1,p}(U) \to C(\bar{U})$$

is compact for $p > \mathrm{N}$ if U is bounded. Since $C(\bar{U}) \subset L^p(U)$, with the obvious bound on the norms, it then also holds that

$$W_0^{1,p}(U) \to L^p(U) \quad \text{is compact if} \quad U \quad \text{is bounded,} \tag{18.4}$$

and this holds for all $p \geq 1$.

Stelling 18.5. *The statement in (18.4) holds for all $p \in [1, \infty)$ if U is bounded. NB: the weaker statement that the embedding is bounded is characterised by the Poincaré inequality: $|u|_p \leq C_{U,p} |\nabla u|_p$.*

For the proof consider a bounded sequence $u_n \in W_0^{1,p}(U)$, choose $v_n \in C_c^1(U)$ and extend v_n to $v_n \in C_c^1(\mathbb{R}^\mathrm{N})$, such that $|u_n - v_n|_{1,p} \to 0$, and consider mollified versions of v_n defined by

$$v_n^\epsilon(x) = (\eta_\epsilon * v)(x) = \int_{\mathbb{R}^\mathrm{N}} \eta_\epsilon(x-y) v_n(y)\, dy = \int_{\mathbb{R}^\mathrm{N}} \eta_\epsilon(y) v_n(x-y)\, dy$$

$$= \int_{|y| \leq \epsilon} \eta_\epsilon(y) v_n(x+y)\, dy,$$

in which

$$\eta_\epsilon(y) = \frac{1}{\epsilon^\mathrm{N}} \eta(\frac{y}{\epsilon}) \quad \text{with} \quad 0 \leq \eta \in C_c^\infty(B) \quad \text{radial,} \quad \int_B \eta = 1,$$

where B is the open unit ball. Then split $|u_n - u_m|_p$ as

$$|u_n - u_m|_p \leq |u_n - v_n|_p + |v_n - v_n^\epsilon|_p + |v_n^\epsilon - v_m^\epsilon|_p + |v_m^\epsilon - v_m|_p + |v_m - u_m|_p.$$

The first and fifth term converge to zero in view of $|u_n - v_n|_{1,p} \to 0$, the second and fourth are controlled by, with $v = v_n$ and $v = v_m$, the estimate

$$\int_{\mathbb{R}^\mathrm{N}} |v^\epsilon(x) - v(x)|^p\, dx \leq \epsilon^p \int_{\mathbb{R}^\mathrm{N}} |\nabla v(x)|^p\, dx, \tag{18.5}$$

and therefore bounded by $C\epsilon$, with C depending only on the bound for the sequence $|v_n|_{1,p}$, and thus only on the original bound for the sequence u_n in $W_0^{1,p}(U)$, while for every $\epsilon > 0$ fixed the third term converges to zero along a subsequence in view of the Ascoli-Arzelà theorem applied to the sequence v_n^ϵ in $C(\bar{V})$, with V bounded and slight larger than U so as to have all v_n^ϵ in $C_c^\infty(V)$. We thus have that $|u_n - u_m|_p$ is asymptotically controlled by $C\epsilon$ along that subsequence, along which $|v_n^\epsilon - v_m^\epsilon|_\infty \to 0$. A standard diagonal argument now produces a subsequence which is a Cauchy sequence in $L^p(V)$. The following exercises fill in the details and are all that's needed to conclude the proof of Theorem 18.5.

Opgave 18.6. Show that

$$\int_{\mathbb{R}^N} |v^\epsilon(x) - v(x)|^p \, dx \le \int_{\mathbb{R}^N} \eta(y) \int_{\mathbb{R}^N} |v(x+\epsilon y) - v(x)|^p \, dx \, dy. \tag{18.6}$$

Hint: write the difference $v^\epsilon(x) - v(x)$ as a single y-integral over $|y| \le \epsilon$, scale y to have the integral over the unit ball, use

$$\eta(y) = \eta(y)^{\frac{1}{p'}} \eta(y)^{\frac{1}{p}},$$

and apply Hölder's inequality. Then use this estimate for the integral of $|v^\epsilon(x) - v(x)|^p$ and change the order of integration to conclude.

Uitwerking. Recall that

$$\eta_\epsilon(x) = \frac{1}{\epsilon^N} \eta(\frac{x}{\epsilon}), \quad \eta(x) = \eta(|x|) \ge 0, \quad \eta \in C_c^\infty(B), \quad \int_{\mathbb{R}^N} \eta = 1.$$

Since $v \in L^p(\mathbb{R}^N)$ *and* $\eta_\epsilon(x) = \eta_\epsilon(-x)$ *we have*

$$v^\epsilon(x) = \int_{\mathbb{R}^N} \eta_\epsilon(x-y) v(y) \, dy = \int_{\mathbb{R}^N} \eta_\epsilon(y) v(x-y) \, dy = \int_{\mathbb{R}^N} \eta_\epsilon(y) v(x+y) \, dy,$$

so

$$v^\epsilon(x) - v(x) = \int_{\mathbb{R}^N} \eta_\epsilon(y) v(x+y) \, dy - v(x) = \int_{\mathbb{R}^N} \eta_\epsilon(y) \left(v(x+y) - v(x)\right) \, dy$$

$$= \int_{\mathbb{R}^N} \eta(y) \left(v(x+\epsilon y) - v(x)\right) \, dy = \int_{\mathbb{R}^N} \eta(y)^{\frac{1}{p'}} \eta(y)^{\frac{1}{p}} \left(v(x+\epsilon y) - v(x)\right) \, dy,$$

whence Hölder's inequality gives

$$|v^\epsilon(x) - v(x)| \le \underbrace{\left(\int_{\mathbb{R}^N} \eta(y) \, dy \right)^{\frac{1}{p'}}}_{=1} \left(\int_{\mathbb{R}^N} \eta(y) |v(x+\epsilon y) - v(x)|^p \, dy \right)^{\frac{1}{p}},$$

whence (18.6) follows as indicated. Note that for $p = 1$ the estimate is immediate.

Opgave 18.7. Write the difference in the right hand side of (18.6) as an s-integral over the interval $[0, 1]$, use Hölder's inequality and $|y| \leq 1$ to arrive at an integral over s and x only, and obtain (18.5).

Uitwerking. In the proof we use the mean value theorem in integral form, see (4.9), and the (Hölder) estimate

$$|\int_0^1 f|^p \leq |\int_0^1 |f|^p \qquad (1 \leq p < \infty).$$

The x-integral in the right hand side of (18.6) is estimated for $|y| \leq \epsilon$ as

$$\int_{\mathbb{R}^N} |v(x+\epsilon y) - v(x)|^p \, dx = \int_{\mathbb{R}^N} |[v(x+t\epsilon y)]_{t=0}^{t=1}|^p \, dx$$

$$= \int_{\mathbb{R}^N} |\int_0^1 \nabla v(x+\epsilon t y) \cdot \epsilon y \, dt|^p \, dx$$

$$\leq \epsilon^p \int_{\mathbb{R}^N} \left(\int_0^1 |\nabla v(x+\epsilon t y)| \, dt \right)^p dx \leq \epsilon^p \int_{\mathbb{R}^N} \int_0^1 |\nabla v(x+\epsilon t y)|^p \, dt \, dx$$

whence (18.6) gives

$$\int_{\mathbb{R}^N} |v^\epsilon(x) - v(x)|^p \, dx \leq \int_{\mathbb{R}^N} \eta(y) \int_{\mathbb{R}^N} |v(x+\epsilon t y) - v(x)|^p \, dx \, dy \leq$$

$$\leq \int_{\mathbb{R}^N} \eta(y) \, \epsilon^p \int_{\mathbb{R}^N} \int_0^1 |\nabla v(x+\epsilon t y)|^p \, dt \, dx \, dy,$$

whence (18.5) follows by changing the order of integration from $dt\,dx\,dy$ to $dx\,dt\,dy$.

Opgave 18.8. Show that for fixed $\epsilon > 0$ the sequence v_n^ϵ has a convergent subsequence in $C(\bar{V})$.

Uitwerking. We apply the A-A Theorem. The bounds

$$|v_n^\epsilon(x)| = |\int_{\mathbb{R}^N} \eta_\epsilon(x-y) v_n(y) \, dy| \leq |\eta_\epsilon|_{p'} |v_n|_p$$

and

$$|\nabla v_n^\epsilon(x)| = |\int_{\mathbb{R}^N} \nabla\eta_\epsilon(x-y)v_n(y)\,dy| \le |\nabla\eta_\epsilon|_{p'}|v_n|_p$$

imply that the conditions of the A-A Theorem are satisfied if v_n is merely a bounded sequence in $L^p(U)$ extended by $v_n(x) = 0$ for $x \notin U$, whence v_n has a convergent subsequence on the compact closure of an ϵ-neighbourhood of U. Since v_n is in fact bounded in $W_0^{1,p}(U)$ we are done (also for $p = 1$).

18.4 Statements for $W^{1,p}(U)$ via extension

Given a bounded domain U we look for a slightly larger domain $\tilde{U}$ such that every $u \in W^{1,p}(U)$ extends to a $\tilde{u} \in W_0^{1,p}(\tilde{U})$ in the sense that $\tilde{u}(x) = u(x)$ for (almost) all $x \in U$. The extension map

$$u \in W^{1,p}(U) \xrightarrow{E} \tilde{u} \in W_0^{1,p}(\tilde{U})$$

should be linear and bounded. The extensions are first defined for $u \in C^1(\bar{U})$ and require U bounded and $\partial U \in C^1$. I usually followed Evans' approach in which the extensions are first defined locally for u and then glued together using a partition of unity, but it came to me that it is in fact simpler to first cut up u in smaller pieces $\zeta_i u$ and choose globally defined extensions of $\zeta_i u$ rather than locally defined extensions of u. This requires a suitable set of functions

$$\zeta_0 \in C_c^\infty(O_0), \zeta_1 \in C_c^\infty(O_1), \ldots, \zeta_n \in C_c^\infty(O_n),$$

with $0 \le \zeta_i \le 1$ and

$$\zeta_0(x) + \zeta_1(x) + \cdots + \zeta_n(x) = 1 \quad \text{for all} \quad x \in \bar{U}$$

with $O_0 \subset \bar{O}_0 \subset U$ and $O_1, \ldots, O_n$ chosen so as to allow globally defined extensions $\tilde{u}_1, \ldots, \tilde{u}_n$ of $u_1 = \zeta_1 u, \ldots, u_n = \zeta_n u$ which define

$$u \xrightarrow{E_i} \tilde{u}_i$$

as a linear map with

$$|\tilde{u}_i|_{1,p} \le C_i\,|u|_{1,p},$$

allowing to define

$$u \in C_c^1(\bar{U}) \xrightarrow{E} C_c^1(\tilde{U}) \quad \text{by} \quad u \to \zeta_0 u + \tilde{u}_1 + \cdots + \tilde{u}_n.$$

In view of the Leibniz rule

$$(\zeta u)_{x_j} = \zeta u_{x_j} + \zeta_{x_j} u$$

it follows that

$$|Eu|_{1,p} \leq C\,|u|_{1,p},$$

with C some horrible constant depending on $\tilde{U}$ and U via the norms of ζ_i in C^1. The functions $\zeta_1, \ldots, \zeta_n$ are chosen to allow a C^1-coordinate transformation similar to the ones used by Evans. It all looks a bit cleaner if the O_i are taken, after a permutation of coordinates, in cylindrical form as $C_i = B_i \times I_i$, with B_i an open ball, I_i a bounded interval, and

$$U \cap C_i = \{x = (x_1, \ldots, x_{n-1}, x_n) \in C : \, x_n > \gamma_i(x_1, \ldots, x_{n-1})\},$$

in which $\gamma_i : \bar{B}_i \to I_i$ is C^1, and the supports of the ζ_i are contained in smaller cylinders $\tilde{C}_i = \tilde{B}_i \times \tilde{I}_i \subset\subset C_i$. The extensions of $\zeta_i u$ are then defined via the transformations in Appendix C.1 and the higher order reflection in Section 5.4.

Finally

$$u \in C^1(\bar{U}) \xrightarrow{E} \tilde{u} \in C_c^1(\tilde{U}),$$

as then defined with the desired $W^{1,p}$-estimates, has to be extended as a map from $W^{1,p}(U)$ to $W_0^{1,p}(\tilde{U})$ using the density of $C^1(\bar{U})$ in $W^{1,p}(U)$.

18.5 Density of smooth functions in $W^{1,p}(U)$

We need $C^1(\bar{U})$ to be dense in $W^{1,p}(U)$, just like $C_c^1(U)$ is, by definition, dense in $W_0^{1,p}(U)$. I will simplify Evans' approach to achieve this. Note that in Section 5.2 he uses (1) as motivation for the definition of weak derivatives. Both u and u_{x_i} have to be locally integrable for the identity to be varified with testfunctions in $C_c^1(U)$ and the uniqueness of weak derivatives folllowed easily in the subsequent lemma via (4). Section 5.2.3 then formulates the Leibniz rule for products of testfunctions and weakly differentiable functions, but inspection of the proof shows that

$$(u\zeta)_{x_i} = u_{x_i}\zeta + u\zeta_{x_i} \tag{18.7}$$

for $\zeta \in C^1(\bar{U})$ and $u \in W_{loc}^{1,1}(U)$. In particular (18.7) holds for $u \in W^{1,p}(U)$ and $\zeta \in C^1(\bar{U})$. So any statement that we want make about a function $u \in W^{1,p}(U)$ can be localised using partitions of unity as in Section 18.4 above, splitting u via $\zeta_0, \zeta_1, \ldots, \zeta_n \in C_c^1(\mathbb{R}^N) \subset C_c^\infty(\mathbb{R}^N)$ with $\zeta_0 + \zeta_1 + \cdots + \zeta_n \equiv 1$ on $\bar{U}$, writing

$$u = u_0 + u_1 + \cdots + u_n = \zeta_0 u + \zeta_1 u + \cdots + \zeta_n u,$$

in which we take $\zeta_0 \in C_c^1(U)$ and $\zeta_i \in C_c^1(C_i)$ with C_i as in Section 2 above. We now use mollifiers to approximate each of these by smooth functions.

Opgave 18.9. Mollifiers are special cases of integral operators of the form $(Tf)(x) = \int K(x,y)f(y)\,dy$ in which both $\int |K(x,y)|\,dx$ and $\int |K(x,y)|\,dy$ are bounded by one fixed constant C. Modify the argument of part 4 of the proof of Theorem 7 in Appendix C.5 to show that $|Tf|_p \leq C|f|_p$ holds for all $f \in L^p$. The estimate holds for all p with $1 \leq p \leq \infty$.

Uitwerking. Note that we replace $\eta_\epsilon(x-y)$ by $K(x,y)$. The assumption that both $\int |K(x,y)|\,dx$ and $\int |K(x,y)|\,dy$ are bounded by one fixed constant C is satisfied by $\eta_\epsilon(x-y)$ with $C=1$. For $1<p<\infty$ we have via Hölder's inequality that

$$|(Tf)(x)| \leq \int |K(x,y)|\,|f(y)|\,dy = \int |K(x,y)|^{\frac{1}{p'}} |K(x,y)|^{\frac{1}{p}}\,|f(y)|\,dy$$

$$\leq \left(\int |K(x,y)|\,dy\right)^{\frac{1}{p'}} \left(\int |K(x,y)|\,|f(y)|^p\,dy\right)^{\frac{1}{p}}$$

$$\leq C^{\frac{1}{p'}} \left(\int |K(x,y)|\,|f(y)|^p\,dy\right)^{\frac{1}{p}},$$

whence

$$|\int |Tf(x)|^p\,dx \leq C^{\frac{p}{p'}} \iint |K(x,y)|\,|f(y)|^p\,dy\,dx$$

$$= C^{\frac{p}{p'}} \iint |K(x,y)|\,|f(y)|^p\,dx\,dy \leq C^{\frac{p}{p'}}\, C \int |f(y)|^p\,dy = C^p \int |f(y)|^p\,dy,$$

and the taking the p^{th} root the estimate $|Tf|_p \leq C|f|_p$ follows. The cases $p=1$ and $p=\infty$ are easier and require only one of the C-bounds.

Opgave 18.10. The function $u_0 = \zeta_0 u$ is in $W^{1,p}(U)$ and vanishes outside the support of ζ_0. With η_ϵ defined as in Appendix C.5 show that $u_0^\epsilon = \eta_\epsilon * u_0 \to u_0$ in $W_0^{1,p}(U)$ as $\epsilon \to 0$. This statement holds only if $1 \leq p < \infty$.

Uitwerking. In view of the Leibniz rule we have that u_0 and its first order weak derivatives are in $L^p(\mathbb{R}^N)$ and vanish outside the support of ζ_0. Since mollifying commutes with taking weak derivatives it suffices to show that $u_0^\epsilon \to u_0$ in $L^p(\mathbb{R}^N)$ because the argument can then be copied for the first order weak derivatives of u_0^ϵ and u_0. The estimate in Exercise 18.9 implies that $|u_0^\epsilon|_p \leq |u_0|_p$. Now take a continuous function v_δ with compact support

in some open neighbourhood V of U that is δ-close to u_0 in $L^p(\mathbb{R}^N)$ and let $v_\delta^\epsilon = \eta_\epsilon * v_\delta$. Then the same estimate applied to $v_\delta - u_0$ implies that

$$|u_0^\epsilon - u_0|_p \leq \underbrace{|u_0^\epsilon - v_\delta^\epsilon|_p}_{\leq |u_0 - v_\delta|_p \leq \delta} + |v_\delta^\epsilon - v_\delta|_p + \underbrace{|v_\delta - u_0|_p}_{\leq \delta} \leq 2\delta + \underbrace{|v_\delta^\epsilon - v_\delta|_p}_{\leq \delta} \leq 3\delta$$

since $\epsilon > 0$ can be chosen to have $|v_\delta^\epsilon - v_\delta|_p \leq \delta$, in view of the uniform convergence of v_δ to v_δ as $\epsilon \to 0$. The latter follows from[5]*, dropping the subscript δ,*

$$|v^\epsilon(x) - v(x)| = |\int_{\mathbb{R}^N} \eta(y)\,(v(x+\epsilon y) - v(x))\,dy| \leq$$

$$\int_{\mathbb{R}^N} \eta(y)\,\underbrace{|v(x+\epsilon y) - v(x)|}_{\leq \delta}\,dy| \leq \delta$$

by using the δ, ϵ-definition of the uniform continuity of v on $\mathbb{R}^N$.

For the $u_1, \ldots, u_n$ we need some regularity of the boundary to get the approximations in $C^1(\bar{U})$. In view of the C^1-regularity of ∂U needed[6] for the extension operator in Section 2 above we assume $\partial U \in C^1$, but uniform Lipschitz continuity suffices, as I explained in the course.

Opgave 18.11. Assume that $\zeta_i \in C_c^1(C_i)$ is as in Section 2. Assume that the function γ_i which describes $\partial U \cap C_i$ is Lipschitz continuous. Modify the argument in the proof of Theorem 3 in Section 5.3.3 to show that the ϵ-mollified version of the $\lambda\epsilon$-translate of $u_i = \zeta_i u$ is in $C_c^1(\bar{U})$ and converges to u_i in $W^{1,p}(U)$ as $\epsilon \to 0$.

We conclude that $C^1(\bar{U})$ is dense in $W^{1,p}(U)$, provided of course that we can find a partition of unity as above.

18.6 The extension and trace operators

We needed the boundedness of U and the C^1-regularity of ∂U to define an extension operator

$$C^1(\bar{U}) \xrightarrow{E} C_c^1(\tilde{U})$$

[5] As below Exercise (18.6).

[6] To define the extension for $u \in C^1(\bar{U})$ and to get the $W^{1,p}$-estimate.

with the $W^{1,p}(\tilde{U})$-norm bound of Eu controlled by the $W^{1,p}(U)$-norm of u. The domain $\tilde{U}$ can be taken as close to U as desired by taking the cylinders C_i small, and E extends to

$$W^{1,p}(U) \xrightarrow{E} W_0^{1,p}(\tilde{U})$$

via the density result in Section 18.5. The other important operator is the bounded linear trace operator

$$W^{1,p}(U) \xrightarrow{T} L^p(\partial U)$$

in Section 5.4, which extends

$$u \in C^1(\bar{U}) \to u|_{\partial U} \in C^1(\partial U).$$

Evans defines it locally, first under the assumption that ∂U is flat and $u \in C^1(\bar{U})$. The same splitting as in Section 18.5 can be used to first define $T(\zeta_i u)$ instead, for $u \in C^1(\bar{U})$, so

$$u \in C^1(\bar{U}) \to \zeta_i u = u_i \in C^1(\bar{U} \cap \bar{C}_i) \to u_i|_{\partial U} \in C^1(\partial U).$$

The local coordinate transformation flattening $\partial U \cap \bar{C}_i$ is not even needed, as u_i is defined for all $x_n \geq \gamma(x_1, \ldots, x_{\mathrm{N}-1})$ with $(x_1, \ldots, x_{\mathrm{N}-1}) \in B_i$ and vanishes for x_{N} large. Thus

$$Tu_i(x_1, \ldots, x_{\mathrm{N}-1}) = u_i(x_1, \ldots, x_{\mathrm{N}-1}, \gamma(x_1, \ldots, x_{\mathrm{N}-1})) = -\int_\gamma^\infty (u_i)_{x_{\mathrm{N}}},$$

and the p-norm on B_i is estimated by the p-norm of ∇u_i, the factor

$$\left(1 + \gamma_{x_1}^2 + \cdots + \gamma_{x_{\mathrm{N}-1}}^2\right)^{\frac{1}{2}}$$

being irrelevant for the estimate. The characterisation of the kernel of T as in Theorem 2 of Section 5.4 is also done locally then, as Evans observes in (6), in which the flattening avoids cumbersome notation in the already technical proof that follows. Actually the proof is not so hard. It relies on this estimate, formulated in $\mathbb{R}^2$ without loss of generality for $u \in C_c^2(\mathbb{R}^2)$:

$$\int_{-\infty}^\infty |u(x,y)|^p \, dx \leq 2^p \left(\int_{-\infty}^\infty |u(x,0)|^p \, dx + y^{p-1} \int_0^y \int_{-\infty}^\infty |u_y|^p \right). \qquad (18.8)$$

Opgave 18.12. Prove (18.8) and explain why it holds for $u \in W^{1,p}(\mathbb{R}^2_+)$ with compact support in $\mathbb{R} \times [0, \infty)$.

Opgave 18.13. If such a u has $Tu = 0$, then the functions u_m defined by $u_m(x,y) = (1 - \zeta(my))u(x,y)$ with $\zeta \in C_c^\infty([0,2))$ and $\zeta \equiv 1$ on $[0,1]$, $\zeta' \leq 0$ on $[0,2)$ are in $W_0^{1,p}(\mathbb{R}^2_+)$ and converge to u in $W^{1,p}(\mathbb{R}^2_+)$. Prove this and conclude that $u \in W_0^{1,p}(\mathbb{R}^2_+)$. Hint: you have to use Exercise 18.8.

18.7 Additional exercises

1. Let $u \in C_c^2(\mathbb{R}^n)$. The norms below are the standard p-norms taken over the whole space.
   (a) Take $n = 2$ and prove that $|u(x_1,x_2)| \leq \frac{1}{4} \int\int |u_{x_1x_2}|$.
   (b) Take $n = 3$ and use (a) to derive an estimate of the form

$$|u|_3 \leq C \max_{i \neq j} |u_{x_i x_j}|_1$$

   (c) Take $n = 4$ and use (a) to derive an estimate of the form

$$|u|_2 \leq C \max_{i \neq j} |u_{x_i x_j}|_1$$

   (d) Generalise to arbitrary $n > 2$. Use scaling arguments to show that

$$|u|_p \leq C_n \max_{i \neq j} |u_{x_i x_j}|_1$$

   can only hold for one particular p which depends on n.
   (e) Generalise to $u \in C_c^3(\mathbb{R}^n)$ and 1-norms of the mixed third order derivatives for $n > 3$.

2. Let U be a bounded domain in $\mathbb{R}^2 = \{(x,y) : x,y \in \mathbb{R}\}$ and $\zeta \in C^1(\mathbb{R}^2)$.
   (a) Prove that $\zeta u \in W^{1,p}(U)$ if $u \in W^{1,p}(U)$.
   (b) Introduce new coordinates ξ, η by

$$x = x_0 + a\xi + b\eta, \quad y = y_0 + c\xi + d\eta,$$

   with $ad \neq bc$, define V by $(\xi,\eta) \in V \iff (x,y) \in U$,
   and write $v(\xi,\eta) = u(x,y)$. Show that

$$u \in W^{1,p}(U) \iff v \in W^{1,p}(V)$$

   and that this correspondence defines a linear homeomorphism between the two Sobolev spaces.
   (c) Assume $\Phi : \mathbb{R}^2 \to \mathbb{R}^2$ is C^1, injective on $\bar{U}$, with invertible Jacobian matrix in every $(x,y) \in \bar{U}$.

Then $u \to u \circ \Phi^{-1} = v$ defines a bijective map from $C^1(\bar{U}) \to C^1(\bar{V})$ where $V = \Phi(U)$. Show that

$$\frac{1}{C}|v|_{W^{1,p(V)}} \leq |u|_{W^{1,p(U)}} \leq C|v|_{W^{1,p(V)}}$$

for some $C > 1$.
(d) Explain why this map uniquely extends to a bijection from $W^{1,p}(U)$ to $W^{1,p}(V)$ if $\partial U \in C^1$.
The intersection of $W^{1,p}(U)$ and $C(\bar{U})$ is a Banach space with norm e.g. $|u|_\infty + |u_x|_p + |u_y|_p$.
(f) Explain why this Banach space is the closure of $C^1(\bar{U})$ with respect to this norm.
NB: These two results also hold if ∂U is Lipschitz in view of the improved version of Theorem 3 in 5.3.3.

18.8 Hölder continuity via Morrey's estimate

The fundamental estimate, for u a C^1-function on $\mathbb{R}^n$, requires $p > n > 1$. Evans derives the estimate

$$|u(x_1) - u(x_2)| \leq C||Du||_{L^p(B(x_1,2r))}\, r^{1-\frac{n}{p}}, \quad r = |x_1 - x_2|,$$

and mentions that this estimate can be improved, but the improved version is not optimal yet. Follow the reasoning below to show that

$$|u(x_1)-u(x_2)| \leq C_{np}||Du||_{L^p(W_{x_1x_2})}\, |x_1-x_2|^{1-\frac{n}{p}}, \quad W_{x_1x_2} = B(x_1,r)\cap B(x_2,r), \quad r = |x_1-x_2|.$$

1. The boundary of $W_{x_1x_2}$ is given by the union of

$$A_1 = \delta B(x_1,r) \cap B(x_2,r) \;\text{ and }\; A_2 = B(x_1,r) \cap \delta B(x_2,r).$$

Let

$$C_1 = \{sx_1+(1-s)z,\, 0 \leq s \leq 1,\, z \in A_1\} \;\text{ and }\; C_2 = \{sx_2+(1-s)z,\, 0 \leq s \leq 1,\, z \in A_2\}.$$

Show that

$$\int_{z\in C_i} |u(z) - u(x_i)|dz \leq \frac{r^n}{n} \int_{z\in C_i} \frac{|\nabla u(z)|}{|z - x_i|^{n-1}} dz, \quad (i = 1, 2).$$

Hint. Without loss of generality assume that $x_i = 0$, $u(0) = 0$. Using polar coordinates $z = \rho\omega$, $0 \leq \rho \leq r$ $|\omega| = 1$, define the cone C_r by $0 \leq \rho \leq r$, $\omega \in A$. What you have to show is that

$$(\textbf{first}) \quad \int_{C_r} |u(z)|dz \leq \frac{r^n}{n} \int_{C_r} \frac{|u_r(z)|}{|z|^{n-1}} dz,$$

in which u_r is the radial derivative defined by

$$u_r(x) = \frac{x}{|x|} \cdot \nabla u(x),$$

and A is an appropriate subset of the unit sphere. Which subset? Show that

$$|u(\rho\omega)| \leq \int_0^\rho |u_r(s\omega)| ds,$$

integrate this inequality with respect to ω over A, change the order of integration on the right hand side to obtain the integral of

$$\frac{|u_r(x)|}{|x|^{n-1}}$$

over the cone $C_\rho = \{x = s\omega : 0 \leq s \leq \rho, \omega \in A\}$. Multiply by ρ^{n-1} and conclude. Then estimate the integral on the right hand side of (**first)** using Hölder's inequality with p for u_r.

2. Show that

$$|u(x_1) - u(x_2)| \leq C_{p,n} ||Du||_{L^p(W)} \, r^{1-\frac{n}{p}}, \quad r = |x_1 - x_2|, \quad W = B(x_1, r) \cap B(x_2, r).$$

Hint. The inequalities in 1 give estimates for the averages of $|u(z) - u(x_i)|$ over the cones C_i. Integrate the inequality

$$|u(x_1) - u(x_2)| \leq |u(x_1) - u(z)| + |u(z) - u(x_2)|$$

over $C_1 \cap C_2$ and estimate in terms of integrals of $|Du|^p = |\nabla u|^p$ over C_1 en C_2 and finally W. Give a precise description of the constant $C_{p,n}$ that you find.

18.9 GNS-estimates

In the third hour I proved

$$|f(x)| \leq \frac{1}{2} \int_{-\infty}^{\infty} |f'(x)| dx$$

for $f \in C_c^1(\mathbb{R})$ and applied it to $x \to u(x, y)$ and $y \to u(x, y)$ to derive an estimate for the 2-norm of $u \in C_c^1(\mathbb{R}^2)$ in terms of the 1-norms of u_x and u_y. The result was

$$\iint_{\mathbf{R}^2} |u|^2 \leq \frac{1}{4} \iint_{\mathbf{R}^2} |u_x| \iint_{\mathbf{R}^2} |u_y| \quad \text{from which} \quad |u|_2 \leq \frac{1}{2} \max_{i=1,2} |u_{x_i}|_1$$

follows (I'm writing single bars with subscript p for the p-norm in L^p).

The same trick with $x \to u(x,y,z)$, $y \to u(x,y,z)$ and $z \to u(x,y,z)$ and Hölder's inequality applied 3 times with exponents $p_1 = p_2 = \frac{1}{2}$ applied to the successive integrations with respect to x, y, z gives

$$\begin{aligned}
\iiint_{\mathbf{R}^3} |u|^{\frac{3}{2}} &\le \iiint_{\mathbf{R}^3} (\frac{1}{2}\int_x |u_x|)^{\frac{1}{2}} (\frac{1}{2}\int_y |u_y|)^{\frac{1}{2}} (\frac{1}{2}\int_z |u_z|)^{\frac{1}{2}} \\
&= (\frac{1}{2})^{\frac{3}{2}} \int_z \int_y \int_x (\int_x |u_x|)^{\frac{1}{2}} (\int_y |u_y|)^{\frac{1}{2}} (\int_z |u_z|)^{\frac{1}{2}} \\
&\le (\frac{1}{2})^{\frac{3}{2}} \int_z \int_y (\int_x |u_x|)^{\frac{1}{2}} (\int_x \int_y |u_y|)^{\frac{1}{2}} (\int_x \int_z |u_z|)^{\frac{1}{2}} \\
&\le (\frac{1}{2})^{\frac{3}{2}} \int_z (\int_y \int_x |u_x|)^{\frac{1}{2}} (\int_x \int_y |u_y|)^{\frac{1}{2}} (\int_y \int_x \int_z |u_z|)^{\frac{1}{2}} \\
&\le (\frac{1}{2})^{\frac{3}{2}} (\int_z \int_y \int_x |u_x|)^{\frac{1}{2}} (\int_z \int_x \int_y |u_y|)^{\frac{1}{2}} (\int_y \int_x \int_z |u_z|)^{\frac{1}{2}}
\end{aligned}$$

(in each integration one of the 3 factors does not depend on the integration variable), so

$$|u|_{\frac{3}{2}} \le \frac{1}{2} |u_x|_1^{\frac{1}{3}} |u_y|_1^{\frac{1}{3}} |u_z|_1^{\frac{1}{3}} \le \frac{1}{2} \max_{i=1,2,3} |u_{x_i}|_1,$$

which generalises to

$$|u|_{\frac{n}{n-1}} \le \frac{1}{2} \max_{i=1,\dots,n} |u_{x_i}|_1$$

for $u \in C_c^1(\mathbb{R}^n)$ via n integrations and Hölder's inequality with $p_1 = \cdots = p_n = \frac{1}{n-1}$ applied in every step (in each integration one of the n factors does not depend on the integration variable). Applied to u^γ it follows via Hölder's inequality with

$$\frac{1}{p} + \frac{1}{p'} = 1$$

that

$$|u|^\gamma_{\frac{\gamma n}{n-1}} = |u^\gamma|_{\frac{n}{n-1}} \le \frac{1}{2} \max_{i=1,\dots,n} |\gamma u^{\gamma-1} u_{x_i}|_1 \le \frac{\gamma}{2} \max_{i=1,\dots,n} |u^{\gamma-1}|_{p'} |u_{x_i}|_p = \frac{\gamma}{2} |u|^{\gamma-1}_{\frac{(\gamma-1)p}{p-1}} \max_{i=1,\dots,n} |u_{x_i}|_p$$

in which γ can be chosen to have equal subscripts of $|u|$ in the first and last expression in this chain.

For $1 \le p < n$ you should check that this gives

$$q = \frac{\gamma n}{n-1} = \frac{(\gamma-1)p}{p-1} = \frac{pn}{n-p},$$

which you may prefer to memorise as

$$\frac{1}{q} = \frac{1}{p} - \frac{1}{n}.$$

What's the value of γ? Dividing by $|u|_q^{\gamma-1}$ on both sides you get

$$|u|_q \leq C_{np} \max_{i=1,\ldots,n} |u_{x_i}|_p$$

with an explicit constant C_{np}. Check again that $1 \leq p < n$ is the assumption to make here.

At the end I also discussed how to obtain the estimate

$$|u|_{\frac{n}{n-2}} \leq \frac{1}{4} \max_{i \neq j} |u_{x_i x_j}|_1$$

for $u \in C_c^2(\mathbb{R}^n)$ and similar estimates for $u \in C_c^3(\mathbb{R}^n)$, $u \in C_c^4(\mathbb{R}^n)$,.... A nice exercise.

19 Evans' Chapter 6 and Navier-Stokes

The two large exercises in Section 19.4 relate to the Navier-Stokes equations. You should compare Problem (**N**) in the first exercise to Exercise 6.6.4 in Evans. Exercises 3,4,5,6 in Section 6.6 are variants on the general theme in Section 6.2. Note that 4 requires Theorem 1 in Section 5.8.1. Make sure you understand the proof. Each of these 4 exercises has $B(u, v)$ symmetric and the solution operator S compact and symmetric with respect to both the L^2-inner product, and an inner product defined by B.

In the manuscript I posted earlier this week have a look at Section 6.6.1 and Stelling 6.13 and thereafter for the eigenvalue formula's. Viewing S as the inverse of L the formula's for the eigenvalues of L follow as in in Theorem 2 of Section 6.5.1. Evaluate these eigenvalue formula's for the problems in Exercises 3,4,5,6.

19.1 Existence of weak solutions via Lax-Milgramm

To sum it up what we did consider first the equation

$$Lu = -(a_{ij}u_{x_i})_{x_j} + cu = f \quad \text{with boundary condition} \quad u = 0 \tag{19.9}$$

for $u = u(x)$, $x \in U$, U a bounded domain in $\mathbb{R}^N$, ∂U at least continuous, i.e. locally the graph of a continuous function.

Compared to (1) in Section 6.1, I drop the summation signs, use subscripts for the coefficients, and omit the first order terms. Existence of classical solutions, i.e. solutions u with $u \in C^2(U)$ to have equation (19.9) make sense in U, and $u \in C(\bar{U})$ to have the homogeneous (Dirichlet) boundary condition $u = 0$ have a meaning, requires conditions on the coefficients $a_{ij} = a_{ij}(x)$ and $c = c(x)$, and on the right hand side f.

19.1.1 Weak solutions

In the weak solution approach we multiply (19.9) by a $v \in C^1(\bar{U})$, integrate over U and use integration by parts to rewrite the terms with a_{ij} as

$$-\int_U \underbrace{(a_{ij}u_{x_i})_{x_j}}_{w_{x_j}} v = -\int_{\partial U} \nu_j \underbrace{a_{ij}u_{x_i}}_{w} v + \int_U \underbrace{a_{ij}u_{x_i}}_{w} v_{x_j}, \tag{19.10}$$

in which ν_j is the j^{th} component of the outward normal on ∂U. This requires ∂U to be piecewise C^1, the coefficients $a_{ij} \in C^2(\bar{U})$, $c \in C(\bar{U})$, the right

hand side $f \in C(\bar{U})$, and $u \in C^2(\bar{U})$. The boundary integral disappears if $v = 0$ on ∂U, leading to the identity

$$\underbrace{\int_U a_{ij} u_{x_i} v_{x_i} + \int_U cuv}_{B[u,v]} = \underbrace{\int_U fv}_{(f,v)} \tag{19.11}$$

for all $v \in C^1(\bar{U})$ with $v = 0$ on ∂U.

This identity that make sense for $u \in C^1(\bar{U})$, with u still required do have $u = 0$ on ∂U. The assumptions of a_{ij}, c, f can of course be weakened now. The weak solution approach works with weak solutions which have their first order derivatives in $L^2(U)$, the natural (Hilbert) space for u, v to live is $H_0^1(U)$ and $u \in H_0^1(U)$ is called a weak solution if (19.11) holds for all $v \in H_0^1(U)$.

This formulation requires $a_{ij}, c \in L^\infty(U)$ only, and $f \in L^2(U)$ then suffices for the right hand side of (19.11) to makes sense because $H_0^1(U) \subset L^2(U)$. Recall that we defined $H_0^1(U)$ as the closure of $C_c^1(U)$ in $H^1(U) = W^{1,2}(U)$. The right hand side of (19.11) is equal to the inner product of f and v in $L^2(U)$ and it defines a linear functional

$$v \in H_0^1(U) \xrightarrow{F} \int_U fv = (f,v)_{L^2(U)} = \underbrace{F(v) = \langle F, v \rangle}_{\substack{\text{different notations} \\ \text{for same functional } F}}, \tag{19.12}$$

the latter being the notation used in the Lax-Milgramm Theorem in Section 6.2.1.

In (19.12) it is tempting to write f for F in $\langle F, v \rangle$, as it is really f that acts on v, but *not* via the $H^1(U)$ inner product, as the $H^1(U)$ inner product is defined by the left hand side of (19.11) with $a_{ij} = \delta_{ij}$ and $c = 1$, i.e.

$$(u,v)_{H^1(U)} = \underbrace{\int_U u_{x_i} v_{x_i}}_{\substack{\text{highest order} \\ \text{terms}}} + \int_U uv = \int_U \nabla u \cdot \nabla v + \int_U uv, \tag{19.13}$$

the bilinear form corresponding to the partial differential equation $\Delta u + u = f$.

19.1.2 The Lax-Milgramm Theorem

The f in Theorem 1 in Section 6.2.1 is really the F in (19.12) if the theorem is applied to the bilinear form in (19.11) with $H = H_0^1(U)$.

If the bilinear form is bounded on H, i.e.

$$\forall u, v \in H \quad |B[u,v]| \leq \alpha ||u||\, |v||,$$

them for each $u \in H$ the map

$$v \in H \xrightarrow{Au} B[Au, v] = \underbrace{(Au)(v) = \langle Au, v \rangle}_{\substack{\text{different notations} \\ \text{for same functional } Au}} \tag{19.14}$$

is linear and bounded, since

$$|\langle Au, v \rangle| = |B[u, v]| \leq \alpha \, ||u|| \, |v||,$$

implying that

$$||Au|| \leq \alpha \, ||u||$$

for all $u \in H$. Thus $A : H \to H^*$, where

$$H^* = \{f : H \to \mathbb{R} : \ f \text{ is linear and bounded}\}$$

normed by

$$||f|| = \sup_{0 \neq v \in H} \frac{|\langle f, v \rangle|}{||v||},$$

is linear and bounded. This dual space H^* of H can be identified with H via the Riesz Representation Theorem and $\langle f, v \rangle = f \cdot v = (f, v)_H$, considering $f \in H = H^*$, but in the application to $H = H_0^1(U)$ this is not the inner product in the right hand side to (19.11).

If the bilinear form is also coercive on H, i.e.

$$\forall u \in H \quad B[u, u] \geq \beta \, ||u||^2,$$

then

$$\beta \, ||u||^2 \leq B[u, u] = \langle Au, u \rangle \leq ||Au|| \, ||u|| \quad \text{whence} \quad ||Au|| \geq \beta \, ||u||$$

for all $u \in H$ and it follows that A is a bijection between H and $A(H)$, a subspace of H^*, bounded in both directions. Thus $A(H)$ is complete, and thereby a closed subspace of H^* which[7] coincides with H^*. The (linear) solution operator is then defined by

$$F \in H^* \xrightarrow{S} u \in H \quad \text{defined by} \quad B[u, v] = \langle F, v \rangle \quad \text{for all} \quad v \in H \tag{19.15}$$

and has the property that

$$||u||_H = ||SF||_H \leq \frac{1}{\beta} \, ||F||_{H^*}$$

In the application to boundary value problems any right hand side of (19.9) that defines an F in the dual of the Sobolev space used is allowed. In the case of $H = H_0^1(U)$ this dual space is denoted by $H^{-1}(U)$ and may be viewed as the space consisting of functions in L^2 as well as their first order distributional derivatives, see Section 5.9.1 in Evans.

[7] Via the Riesz Representation or the Hahn-Banach Theorem and the reflexivity of H.

19.1.3 Lax-Milgramm; boundedness condition

It is usually easy to show that the bilinear form derived from the boundary value problem formulation via integration by parts is bounded, also for other boundary conditions, such as the Neumann condition

$$\nu_j a_{ij} u_{x_i} = 0 \quad \text{on} \quad \partial U, \tag{19.16}$$

which is a special case of the Robin boundary condition

$$\nu_j a_{ij} u_{x_i} + bu = 0 \quad \text{on} \quad \partial U, \tag{19.17}$$

in which $b = b(x)$ is assumed to bounded and integrable for instance. See Exercises 6.6.4 and 6.6.5. The latter condition is called Newton's cooling law in the case that a_{ij} is a positive multiple of the identity matrix and u is the temperature in a body U with heat exchange at the boundary[8]. In (19.10) this condition gives the additional term

$$\int_{\partial U} buv$$

which should now be included in the left hand side of (19.11), and the natural Sobolev space to pose

$$\underbrace{\int_U a_{ij} u_{x_i} v_{x_i}}_{\text{highest order terms}} + \int_U cuv + \int_{\partial U} buv = \underbrace{\int_U fv}_{(f,v)} \tag{19.18}$$

in is now $H^1(U)$.

In the case of the Neumann condition (19.16) this extra term is not there and the only difference with the Dirichlet problem is the choice of the Sobolev space. Boundary conditions which are used in the integration by parts derivation of the weak formulation are sometimes called natural boundary conditions. The Dirichlet boundary condition is not such a natural boundary condition, it has to be forced on the solution by the choice of the smaller Sobolev space $H_0^1(U)$.

19.1.4 Lax-Milgramm; coercivity

It is usually more delicate to show the coercivity of the bilinear form. The basic (ellipticity) assumption on the coefficients a_{ij} is (4) in Section 6.1.1 of Evans. With $v = u$ it bounds the highest order terms from below by the

[8] This physical context forces the exchange coefficient to be positive.

highest order terms in (19.13). In the case of $H_0^1(U)$ the Poincaré inequality

$$\int_U u^2 \leq C_U \int_U |\nabla u|^2 \tag{19.19}$$

helps. In particular the bilinear form

$$B[u,v] = \int_U \nabla u \cdot \nabla v$$

used for solving

$$-\Delta u = f \quad \text{with boundary condition} \quad u = 0$$

is coercive on $H_0^1(U)$ considered as a subspace of $H^1(U)$ with the norm derived from (19.13).

The Neumann problem for $-\Delta u = f$ is very instructive. It requires a condition on f for solvability, as well as the same condition on u to have a unique condition, choosing

$$\tilde{H}^1(U) = \{u \in H^1(U) : \int_U u = 0\}$$

as the Sobolev space to be used in the weak formulation.

You should compare the role of b in the Robin boundary condition to that of c in the partial differential equation, as should be clear from (19.18). Coercivity requires some positivity. It's easy to cook up exam questions on this theme.

Also, the higher order problem for the bi-Laplacian in Exercise 6.6.3 is only one of the problems of this type. It has two ünnatural"boundary conditions, which are forced upon the solution by the choice of $H_0^2(U)$. Can you think of natural boundary conditions that lead to a formulation in $H^2(U)$, or a mix of natural and unnatural boundary conditions that require $H^2(U) \cap H_0^1(U)$ as the space to be used? Note that for the coercivity of the bilinear form you need the regularity theory in Section 6.3.

19.1.5 The general case with first order terms

The treatment of the Dirichlet problem in Section 6.2.2 should be easy to follow after the discussion above. The main issue is how to deal with the terms in $B[u,v]$ that come from the first order derivatives in the Lu. I did not discuss Section 6.2.3 with the adjoint operator and the Fredholm alternative but read Theorem 4. It is proved via an application of the Fredholm alternative to the solution operator S_μ for the bilinear form

$$B_\mu[u,v] = B[u,v] + \gamma \int_U uv$$

with γ chosen to make B_m coercive.

19.2 The selfadjoint case

The first order terms in L typically prevent the bilinear form from being symmetric. Without these first order terms the symmetry of a_{ij} makes the bilinear form symmetric. This symmetry is usually assumed, see the opening statements in Section 6.5.1. In the case that $B[u, v]$ is a symmetric bounded coercive bilinear form, it defines an equivalent norm on the Sobolev space (used in in the weak formulation) via

$$||u|| = \sqrt{B[u, u]}.$$

The solution operator

$$f \xrightarrow{S} u$$

then satisfies both

$$(Sf, g)_{L^2(U)} = (f, Sg)_{L^2(U)} \quad \text{and} \quad B(Su, v) = B(u, Sv)$$

as you should satisfy, and it is compact form $L^2(U)$ to $L^2(U)$ as well as from the Sobolev space to itself. The eigenvalue formula's I discussed for the solution operator using $B[u, v]$ then lead to eigenvalue formula's of which the first is stated in the remark following Theorem 2 in Section 6.5.1.

19.2.1 Second hand in homework set

I will restrict the second homework set to Exercises 6.6.3, 6.6.4, 6.6.5, 6.6.6 in the second edition of Evans. In all exercises also write down the eigenvalue formula for the first eigenvalue when f is replaced by λu. I advise you to also do the exercises in Section 19.4 below.

19.2.2 Maximum principles

You should have a look at Section 6.4. I did not have time to deal with them in this course. More on those principles in Chapters 5 and 10 in

`http://www.few.vu.nl/~jhulshof/NOTES/ellpar.pdf`

19.3 The Navier-Stokes equations

I usually spend a full afternoon on these equations, but as we lost one session there will be no time for it this year. You can read about these equations in Dutch on a very introductionary and informal level in

http://www.math.vu.nl/~jhulshof/handoutNS.pdf

Consider the Navier-Stokes equations on a bounded domain $\Omega \subset \mathbb{R}^2$ for $t \geq 0$ with smooth boundary $\partial\Omega$, given initial data for the velocity

$$u = \begin{pmatrix} u_1(t, x_1, x_2) \\ u_2(t, x_1, x_2) \end{pmatrix}$$

at $t = 0$ and no-slip boundary conditions $u = 0$ on $\partial\Omega$ for all $t \geq 0$. For the exercises below you may restrict your attention to the case that

$$\Omega = \{(x_1, x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 < 1\} \quad \text{with outer normal} \quad n = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \quad \text{in} \quad x \in \partial\Omega.$$

The Navier-Stokes equations read (with kinematic viscosity equal to unity)

$$u_t + (u \cdot \nabla)u + \nabla p = \Delta u, \quad \nabla \cdot u = 0.$$

The second zero divergence equation has to be imposed on the initial data for u at $t = 0$ as well. In view of the Laplacian in the equation and the boundary condition $u = 0$ on $\partial\Omega$ the natural spaces for solutions to live in as functions of t are

$$H_0^1(\Omega)^2 = \{ \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} : \, u_1, u_2 \in H_0^1(\Omega)\} \subset (L^2(\Omega))^2 = \{ \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} : \, u_1, u_2 \in L^2(\Omega)\},$$

but the zero divergence equation imposes an a priori restriction as explained next.

If $u \in (L^2(\Omega))^2$ satisfies $\nabla \cdot u \in L^2(\Omega)$ then the normal component $n \cdot u$ of the velocity is well defined in $L^2(\partial\Omega)$ by a theorem similar to the trace theorems in Evans, and the Gauss divergence formula

$$\int_\Omega \nabla \cdot u = \int_{\partial\Omega} n \cdot u$$

holds true for such u. Solutions with finite kinetic energy

$$E(u) = \frac{1}{2} \int_\Omega (u_1^2 + u_2^2)$$

actually live in

$$H = \{u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} : \, u_1, u_2 \in L^2(\Omega), \, \nabla \cdot u = 0 \text{ on } \Omega, \, n \cdot u = 0 \text{ on } \partial\Omega\}.$$

If also the first order spatial weak derivatives exist with

$$\mathcal{E}(u) = \int_\Omega |Du|^2 = \int_\Omega \left((\frac{\partial u_1}{\partial x_1})^2 + (\frac{\partial u_1}{\partial x_2})^2 + (\frac{\partial u_2}{\partial x_1})^2 + (\frac{\partial u_2}{\partial x_2})^2 \right) < \infty,$$

then $u \in H^1(\Omega)^2$ and it is possible to speak of u on $\partial\Omega$ as the trace of u and in particular of its tangential component $n \times u = n_1 u_2 - n_2 u_1$ in the usual sense.

19.4 Navier-Stokes related exercises

1. This exercise concerns the projection of

$$L^2_{div}(\Omega) = \{w \in (L^2(\Omega))^2 : \nabla \cdot w \in L^2(\Omega)\}$$

on the space H above (the subscript *div* stands for divergence). For $w \in L^2_{div}(\Omega)$ let $f = -\nabla \cdot w \in L^2(\Omega)$ and $g = n \cdot w \in L^2(\partial\Omega)$, and consider the Neumann problem

$$(\mathbf{N}) \qquad -\Delta p = f \quad \text{in} \quad \Omega \quad \text{with} \quad \frac{\partial p}{\partial n} = g \quad \text{on} \quad \partial\Omega.$$

You may think of p in (**N**) as related to the pressure in the Navier-Stokes equations.

   (a) What is the natural condition on arbitrary $f \in L^2(\Omega)$ and $g \in L^2(\partial\Omega)$ to have a solution of (**N**)? Hint: use the divergence theorem, you may argue as if f, g and p are smooth. Does your condition hold for the particular choice of f and g above? If so, why? Explain why then the solution p is never unique and can be chosen to have $\int_\Omega p = 0$.

   (b) Explain why for $f \in L^2(\Omega)$ and $g \in L^2(\partial\Omega)$ we say that $p \in H^1(\Omega)$ is a weak solution of (**N**) if

$$(\mathbf{N_{weak}}) \qquad \int_\Omega \nabla p \cdot \nabla \phi = \int_\Omega f\phi + \int_{\partial\Omega} g\phi \quad \text{for all} \quad \phi \in H^1(\Omega).$$

   Check that this can only hold if your condition in (a) is satisfied, in which case it suffices to show that that the identity in ($\mathbf{N_{weak}}$) holds for every $\phi \in \tilde{H}^1(\Omega) = \{p \in H^1(\Omega) : \int_\Omega p = 0\}$.

   (c) Let $\tilde{H}^1(\Omega)$ be as in (b). Show that

$$((p, \phi)) = \int_\Omega \nabla p \cdot \nabla \phi$$

   defines an inner product on $\tilde{H}^1(\Omega)$ with an inner product norm that is equivalent on $\tilde{H}^1(\Omega)$ to the full H^1-norm defined by

$$(p, \phi)_{H^1(\Omega)} = \int_\Omega p\phi + \int_\Omega \nabla p \cdot \nabla \phi, \quad ||p||^2_{H^1(\Omega)} = (p, p)_{H^1(\Omega)},$$

   (d) Explain why for every $f \in L^2(\Omega)$ and every $g \in L^2(\partial\Omega)$ satisfying your condition in (a) there is a unique $p \in \tilde{H}^1(\Omega)$ that satisfies ($\mathbf{N_{weak}}$).

(e) Recall that

$$H = \{u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} : \, u_1, u_2 \in L^2(\Omega), \, \nabla \cdot u = 0 \text{ on } \Omega, \, n \cdot u = 0 \text{ on } \partial\Omega\}.$$

Explain why every $w \in L^2_{div}(\Omega)$ can be written as $w = \nabla p + u$ with $u \in H$ and $p \in H^1(\Omega)$, and that u is uniquely determined by w. This u is called the Leray projection of w.

2. In this exercise we consider smooth solutions of the Navier-Stokes equations with zero slip boundary conditions as above (so you can forget about weak derivatives and all that now).

(a) Write u_0 for the initial velocity field of a smooth solution u with pressure p: then $u(x, 0) = u_0(x)$ and u_0 must satisfy $\nabla \cdot u_0 = 0$. We write $u(t)$ for the function $x \to u(x, t)$. Integrate the inner product of

$$u_t + (u \cdot \nabla)u + \nabla p - \Delta u$$

with u over Ω and derive that

$$\frac{d}{dt} E(u(t) + \mathcal{E}(u(t)) = 0,$$

where $E(u)$ and $\mathcal{E}(u)$ are as in the introduction above. In other words, show that

$$\frac{1}{2}\frac{d}{dt}\int_\Omega |u|^2 + \int_\Omega |Du|^2 = 0.$$

Why does it follow that

$$\int_0^T \int_\Omega |Du|^2 \leq \frac{1}{2}\int_\Omega |u_0|^2 \, ?$$

Hint: write terms out in coordinates, e.g.

$$u \cdot (u \cdot \nabla)u = \sum_{j,k=1}^{2} u_k u_j \frac{\partial u_k}{\partial x_j},$$

and use integration by parts (the boundary terms disappear, as well as $\nabla \cdot u$).

(b) For smooth solutions u and v with pressures p and q respectively let $w = u - v$. Subtract the equations for u and v, take the inner product with w and integrate over Ω to derive that

$$\frac{1}{2}\frac{d}{dt}\int_\Omega |w|^2 + \int_\Omega |Dw|^2 = -\int_\Omega w \cdot (w \cdot \nabla)v \leq \int_\Omega |Dv||w|^2.$$

Hint: (i) use integration by parts for the equality, the boundary terms disappear, as well as $\nabla \cdot u$, $\nabla \cdot v$, $\nabla \cdot w$, if they show up. Check that the terms coming from the nonlinear terms in the equations may be rewritten as a term giving the integral with w and v, and another integral with u and w which disappears; (ii) for the subsequent inequality use $Ax \cdot x \leq |A||x|^2$ for 2×2 matrices A and 2-vectors x, with $|A|^2 = A_{11}^2 + A_{12}^2 + A_{21}^2 + A_{22}^2$.

(c) Derive from (b) that

$$\frac{d}{dt}\int_\Omega |w|^2 + 2\int_\Omega |Dw|^2 \leq 2(\int_\Omega |Dv|^2)^{\frac{1}{2}}(\int_\Omega |w|^4)^{\frac{1}{2}}.$$

(d) Insert the inequality

$$\int_\Omega |w|^4 \leq \int_\Omega |w|^2 \int_\Omega |Dw|^2$$

in (c) to derive that

$$\frac{d}{dt}\int_\Omega |w|^2 \leq \frac{1}{2}\int_\Omega |Dv|^2 \int_\Omega |w|^2.$$

Hint: in the right hand side you get a product which contains the factor $a = \int_\Omega |Dw|^2$. Use the inequality $2ab \leq a^2 + b^2$ and observe that a^2 als appears on the left hand side.

(e) Derive from (d) and (a) with u replaced by v that

$$\int_\Omega |w(t)|^2 \leq \int_\Omega |w_0|^2 e^{\frac{1}{4}\int_\Omega |v_0|^2}.$$

(f) Prove the inequality used in (d) for compactly supported smooth vectorfields on $\mathbb{R}^2$ by first showing that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u(x_1, x_2)^4 \, dx_1 dx_2 \leq$$

$$(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u^2)\,(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u_{x_1}^2)^{\frac{1}{2}}\,(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u_{x_2}^2)^{\frac{1}{2}}$$

for compactly supported smooth functions u. In short:

$$||u||_4^4 \leq ||u||_2^2 \, ||u_{x_1}||_2 \, ||u_{x_2}||_2.$$

Hint: write $u(x_1, x_2)^4 = u(x_1, x_2)^2 u(x_1, x_2)^2$ and show first that

$$u(x_1, x_2)^2 \leq \int_{-\infty}^{\infty} u(\xi, x_2) u_{x_1}(\xi, x_2) d\xi$$

and likewise

$$u(x_1, x_2)^2 \leq \int_{-\infty}^{\infty} u(x_1, \eta) u_{x_2}(x_1, \eta) d\eta.$$

20 Geostuff

I will use L for the Lagrangian and not F. We assume that $L = L(t, u, p)$ is as smooth as we need. Chapter 1 of [J&J] concerned Euler-Lagrange equations for $u = u(t) \in \mathbb{R}^n$. We saw how minimizing

$$I(u) = \int_a^b L(t, u(t), \dot{u}(t))\, dt \tag{20.1}$$

for sufficiently smooth functions $u : [a, b] \to \mathbb{R}^n$ (with $u(a)$ and $u(b)$ prescribed) leads to the Euler-Lagrange system of differential equations:

$$\frac{d}{dt}\frac{\partial L}{\partial p^i} - \frac{\partial L}{\partial u^i} = 0 \quad (i = 1, \ldots, n) \tag{20.2}$$

We also saw the Jacobi equations, obtained from (1.3.6) and the linearised Lagrangian

$$\phi = \frac{\partial^2 L}{\partial p^i \partial p^j}\pi^i\pi^j + 2\frac{\partial^2 L}{\partial u^i \partial p^j}\pi^i\eta^j + \frac{\partial^2 L}{\partial u^i \partial u^j}\eta^i\eta^j \tag{20.3}$$

The Euler-Lagrange equations of (20.3) are the Jacobi equations

$$\frac{d}{dt}\frac{\partial \phi}{\partial \pi^i} - \frac{\partial \phi}{\partial \eta^i} = 0 \quad (i = 1, \ldots, n) \tag{20.4}$$

These Jacobi equations are the linearised Euler-Lagrange equations. Verify this!

For Lagrangians independent of t we noticed a conservation law. When you multiply (20.2) by $p^i(t) = \dot{u}^i(t)$ you get

$$0 = p^i(t)\frac{d}{dt}\frac{\partial L}{\partial p^i} - \dot{u}^i(t)\frac{\partial L}{\partial u^i} = \frac{d}{dt}\left(p^i\frac{\partial L}{\partial p^i}\right) \underbrace{-\dot{p}^i(t)\frac{\partial L}{\partial p^i} - \dot{u}^i(t)\frac{\partial L}{\partial u^i}}_{-\frac{dL}{dt}}$$

$$= \frac{d}{dt}\left(p^i\frac{\partial L}{\partial p^i} - L\right)$$

20.1 Submanifolds of $\mathbb{R}^d$ are Riemannian

Chapter 2 deals with the problem of finding the shortest connecting curve between two given points in an n-dimensional submanifold M of $\mathbb{R}^d$ with $d > n$. For this will need knowledge of the concept of covariant differentiation

on M. The nonabstract introduction with submanifolds below provides a machinery that also works in the abstract setting of general Riemannian manifolds.

Locally M is given by smooth parameterisations

$$x = f(u)$$

(coordinate charts) defined on open connected sets $U \subset \mathbb{R}^n$ with smooth transitions between u and $\tilde{u}$ on $U \cap \tilde{U}$ if $f : U \to M$ and $\tilde{f} : \tilde{U} \to M$ are two different coordinate patches. A (preferably finite[1]) collection with this property that describes the whole of M is called an atlas for M.

Every such parameterisation provides us with locally defined tangent vector fields

$$x_1 = \frac{\partial x}{\partial u^1}, \cdots, x_n = \frac{\partial x}{\partial u^n},$$

since for every $u \in U$ the vectors $x_i(u)$ are tangent to M in $x(u) \in M$. The inner products

$$g_{ij} = g_{ij}(u) = x_i \cdot x_j$$

are locally defined scalar fields, the coefficients of the Riemannian metric on M inherited from the inner product in the ambient space $\mathbb{R}^d$.

In terms of local coordinates $u^1, \ldots, u^n$ tangent vector fields V on M are described by

$$V = V^i x_i = V^i(u) x_i(u) = V^1(u) x_1(u) + \cdots + V^n(u) x_n(u), \tag{20.5}$$

in which we use a summation convention for repeated lower and upper indices. Two such vectors fields have inner product

$$V \cdot W = V^i x_i \cdot W^j x_j = V^i W^j x_i \cdot x_j = V^i W^j g_{ij}$$

Don't forget the u-dependence which is usually dropped from the notation and pay attention to the double use of subscripts: as indices in g_{ij} and as derivatives in x_i. The inner product of two tangent vector fields on M defines a scalar field[2] on M. The map

$$(V, W) \to V \cdot W$$

is well defined, independent of the choice of coordinates, and multilinear over the scalar fields[3]. In particular, if $\phi, \psi : M \to \mathbb{R}$ are (smooth) functions, then

$$(\phi V) \cdot (\psi W) = \phi\psi \, (V \cdot W)$$

[1] This is related to the concept of compactness

[2] A real valued function

[3] Tensor property

20.2 Covariant differentiation

If we differentiate a vector field V as given by (20.5) we get contributions from u-dependence in $V^i(u)$ and from u-dependence in $x_i(u)$. The tangential part of the resulting derivative is what is by definition the covariant derivative. The partial derivative of (20.5) with respect to u^j can be written as

$$\frac{\partial V}{\partial u^j} = \frac{\partial V^i}{\partial u^j} x_i + V^i x_{ij}, \quad x_{ij} = \frac{\partial x_i}{\partial u^j} = \frac{\partial^2 x}{\partial u^j \partial u^i} = \frac{\partial^2 x}{\partial u^i \partial u^j} = x_{ji} \tag{20.1}$$

In the case that $M = \mathbb{R}^n = \mathbb{R}^d$ with $x^i = u^i$, the tangent vectors x_i are the unit base vectors e_i so that $x_{ij} = 0$ and the covariant partial derivatives of V are just the partial derivatives V. The same holds if $x(u)$ in linear in u. In all other cases we decompose x_{ij} as

$$x_{ij} = \Gamma^l_{ij} x_l + \textit{ normal parts}$$

and take the inner product with x_k to get

$$\Gamma_{ijk} := x_{ij} \cdot x_k = \Gamma^l_{ij} x_l \cdot x_k = \Gamma^l_{ij} g_{lk}$$

Thus Γ_{ijk} is obtained from Γ^l_{ij} using g_{lk}. Introducing $g^{kl} = g^{lk}$ by

$$g_{lk} g^{km} = \delta^m_l,$$

we also obtain Γ^m_{ij} from Γ_{ijk}:

$$g^{mk} \Gamma_{ijk} = \Gamma^l_{ij} g_{lk} g^{km} = \Gamma^l_{ij} \delta^m_l = \Gamma^m_{ij}$$

The relation between both Γ-symbols is given by

$$\Gamma_{ijk} = \Gamma^l_{ij} g_{lk}, \quad \Gamma^m_{ij} = g^{mk} \Gamma_{ijk}$$

The metric coefficients are used to raise and lower the exponents[4].

Next we determine Γ_{ijk}. Differentiating g_{ij} with respect to u^k we get

$$g_{ij,k} = \frac{\partial g_{ij}}{\partial u^k} = \frac{\partial}{\partial u^k}(x_i \cdot x_j) = x_{ki} \cdot x_j + x_{jk} \cdot x_i = \Gamma_{kij} + \Gamma_{jki}$$

Note the two cyclic permutations kij and jki of ijk on the right. Using cyclic permutation, we have the following three equivalent forms of the resulting statement:

$$g_{ij,k} = \Gamma_{kij} + \Gamma_{jki}$$

[4]Just as with tensor coefficients, though the Γ's are not tensor coefficients

$$g_{jk,i} = \Gamma_{ijk} + \Gamma_{kij}$$
$$g_{ki,j} = \Gamma_{jki} + \Gamma_{ijk}$$

Multiplying by $-\frac{1}{2}$, $\frac{1}{2}$ and $\frac{1}{2}$ and adding up we get

$$\Gamma_{ijk} = \frac{1}{2}\left(g_{jk,i} + g_{ki,j} - g_{ij,k}\right)$$

Using the symmetry $g_{ij} = g_{ji}$ it follows that

$$\Gamma_{ijk} = \frac{1}{2}\left(g_{jk,i} + g_{ik,j} - g_{ij,k}\right), \quad \Gamma^m_{ij} = \frac{1}{2}g^{mk}\left(g_{jm,i} + g_{im,j} - g_{ij,m}\right) \tag{20.2}$$

These formula's define the *Christoffel symbols* $\Gamma^k_{ij} = \Gamma^k_{ji}$ in terms of the metric and its first order derivatives and can be used to write (20.1) as

$$\frac{\partial V}{\partial u^j} = \frac{\partial V^i}{\partial u^j}x_i + V^i\Gamma^l_{ij}x_l + \textit{normal parts}$$

The tangential part is thus

$$D_{u^j}V := (\frac{\partial V}{\partial u^j})_T = \left(\frac{\partial V^l}{\partial u^j} + V^i\Gamma^l_{ij}\right)x_l, \quad V = V^i x_i \tag{20.3}$$

This is called the covariant derivative of V with respect to u^j. Both V and $D_{u^j}V$ are tangent vector fields, with components

$$V^i \quad \text{and} \quad (D_{u^j}V)^l = \frac{\partial V^l}{\partial u^j} + V^i\Gamma^l_{ij}$$

20.3 Tangent vectors as derivatives

Next we introduce the modern view point on tangent vectors. Since every tangent vector defines a directional derivative, it has become customary to identify such first order differential operators with their direction vectors. In short, we think of

$$x_i = \frac{\partial x}{\partial u^i} \quad \text{and} \quad \frac{\partial}{\partial u^i}$$

as essentially the same objects. To see how this works in a point $x_0 \in M$ we use integral curves starting at x_0, that is, solutions of

$$\dot\gamma(t) = X(\gamma(t)), \quad \gamma(0) = x_0 \in M, \tag{20.1}$$

where X is a tangent vector field defined near x_0. The differential equation in (20.1) is called the *flow equation* for X. Using coordinates u, with $u = u_0$ corresponding to x_0, the expressions in (20.1) evaluate as

$$\gamma(t) = x(u(t)), \quad \dot\gamma(t) = \frac{\partial x}{\partial u^i}(u(t))\dot u^i(t) = \dot u^i(t)x_i, \quad X(\gamma(t)) = X^i(u(t))x_i,$$

so the system to be solved for $u = u(t)$ to obtain the integral curves is

$$\dot u^i = X^i(u), \quad u(0) = u_0. \tag{20.2}$$

The solution $u = u(t)$ exists locally and is unique. We have $\dot u^i(0) = X^i(u_0)$ and $X_0 := X(x_0) = \dot\gamma(0) = \dot u^i(0)x_i = X^i(u_0)x_i$. On scalar fields (functions) $\phi : M \to \mathbb{R}$, given in local coordinates as

$$\phi = \phi(u^1, \ldots, u^n),$$

the vector field X now acts through

$$\frac{d}{dt}|_{t=0}\phi(u(t)) = \frac{\partial\phi}{\partial u^i}(u_0)\dot u^i(0) = X_0^i\frac{\partial\phi}{\partial u^i}(u_0)$$

at ϕ in $u = u_0$, i.e. as the directional derivative

$$X_0^i\frac{\partial}{\partial u^i} \quad \text{corresponding to the direction vector} \quad X_0^i x_i$$

in $u = u_0$. The derivative only depends on the value of the vector field in x_0. Since the point $x_0 = x(u_0)$ was arbitrary we have

$$X = X^i\frac{\partial}{\partial u^i} \quad \text{corresponding to the tangent field} \quad X = X^i x_i = X^i\frac{\partial x}{\partial u^i}.$$

The two expressions above are merely different representations of the tangent vector field X (both in local coordinates):

The components

$$X^i\frac{\partial x^k}{\partial u^i}$$

of the *tangent field* X multiply

$$\frac{\partial\phi}{\partial x^k}$$

in the chain rule formula if ϕ is extended to a neighbourhood of M in $\mathbb{R}^d$. As *differential operator*

$$X = X^i\frac{\partial}{\partial u^i}$$

X acts on scalar fields like $\phi = \phi(u)$ and produces a scalar field $X\phi$, the derivative of ϕ in the direction of X. This directional derivative is denoted by

$$\nabla_X \phi = X\phi, \quad \text{replacing the notation} \quad \frac{\partial \phi}{\partial X}$$

in calculus texts. We already use the notation ∇_X customary for covariant differentiation. For reasons that should be clear, covariant differentiation of scalar fields is by definition the same as differentiation of scalar fields.

20.4 Commutators of tangent vector fields

If X and Y are scalar fields on M then the commutator of X and Y is defined as

$$[X, Y] = XY - YX$$

meaning that

$$\nabla_{[X,Y]}\phi = [X, Y]\phi = X(Y\phi) - Y(X\phi) = \nabla_X(\nabla_Y \phi) - \nabla_Y(\nabla_X \phi).$$

This commutator has a meaning by itself. If $\gamma(t)$ is the solution of (20.1), then the linearised flow equation transports the vector $Y(x_0)$ along $\gamma(t)$. Denoting the transported vector as $\xi(t)$, we may differentiate the difference of $\xi(t)$ and $Y(\gamma(t))$ with respect to t and evaluate the derivative in $t = 0$. This defines

$$(\mathcal{L}_X Y)(x_0) = \lim_{t \to 0} \frac{\xi(t) - Y(\gamma(t))}{t},$$

the Lie derivative of Y with respect to X in x_0.

In coordinates $\xi(t) = \xi^i(t) x_i$ with $\xi^i(t)$ is a solution of the linearizsation of (20.2) around $u(t)$,

$$\dot{\xi}^i = \underbrace{(\frac{\partial X^i}{\partial u^j})}_{\text{in } (u(t)} \xi^j(t), \quad \xi^j(0) = Y^i(u_0) \tag{20.1}$$

Writing

$$\xi(t) - Y(\gamma(t)) = \xi(t) - Y(x_0) - (Y(\gamma(t)) - Y(x_0))$$

you should verify that

$$(\mathcal{L}_X Y)(x_0) = (XY)(x_0) - (YX)(x_0)$$

so that

$$[X, Y] = \mathcal{L}_X Y \tag{20.2}$$

Note that $[X, Y]$ is bilinear over de scalar fields. Verify that

$$[X, Y]^j = X^k Y^j_k - Y^k X^j_k$$

and that the Jacobi identity

$$[[X, Y], Z] + [[Y, Z], X] + [[Z, X], Y] = 0 \tag{20.3}$$

holds.

20.5 Covariant differentiation of tangent vectors

Next we observe that

$$X = X^i \frac{\partial}{\partial u^i}$$

naturally acts covariantly on tangent fields V, if we replace

$$\frac{\partial}{\partial u^i} \quad \text{by} \quad D_{u^j},$$

as defined in (20.3) through

$$D_{u^j} V := \left(\frac{\partial V^l}{\partial u^j} + V^i \Gamma^l_{ij} \right) x_l \quad \text{for} \quad V = V^i x_i.$$

The result of this action is

$$X^j \left(\frac{\partial V^l}{\partial u^j} + V^i \Gamma^l_{ij} \right) x_l$$

and is denoted as

$$\nabla_X V = X^j \left(\frac{\partial V^l}{\partial u^j} + V^i \Gamma^l_{ij} \right) \frac{\partial}{\partial u^i} \tag{20.1}$$

in the modern notation for tangent vectors as differential operators.

The map

$$V \to \nabla_X V$$

is *not* linear over the scalar fields because

$$\nabla_X \phi V = X^j \left(\frac{\partial \phi V^l}{\partial u^j} + \phi V^i \Gamma^l_{ij} \right) x_l$$

$$= \phi X^j \left(\frac{\partial V^l}{\partial u^j} + V^i \Gamma^l_{ij} \right) x_l + X^j \frac{\partial \phi}{\partial u^j} V^l = \phi \nabla_X V + (\nabla_X \phi) V.$$

The latter term in this Leibniz rule destroys the tensor property of linearity over the scalar fields.

Convince yourself that in the non-abstract approach

$$\nabla_X V = X^j \left(\frac{\partial V^l}{\partial u^j} + V^i \Gamma^l_{ij} \right) x_l$$

is the tangential[5] *component of the derivative of* V *in the direction of* X and verify that

$$\nabla_X (V \cdot W) = \nabla_X V \cdot W + V \cdot \nabla_X W$$

if W is another tangent vector field on M.

20.6 Second fundamental form

The normal part of the derivative of V in the direction of X is denoted by $\mathrm{I\!I}(X, V)$, in which $\mathrm{I\!I}$ is called the *second fundamental form* of M. Verify that it is bilinear over the smooth fields on M. Since the normal part essentially comes from the mixed derivatives x_{ij}, the *second fundamental form must be symmetric*. Moreover, if N is a normal vector field on M and N, X, V are extended smoothly[6] to the ambient space $\mathbb{R}^d$ then

$$\bar{\nabla}_X (N \cdot Y) = \bar{\nabla}_X N \cdot Y + N \cdot \bar{\nabla}_X Y, \tag{20.1}$$

in which $\bar{\nabla}$ is the (standard covariant) derivative in $\mathbb{R}^d$. On M the left hand side of (20.1) is zero, and the second term $N \cdot \bar{\nabla}_X Y$ on the right hand side only sees the normal part of $\bar{\nabla}_X Y$ which is $\mathrm{I\!I}(X, Y)$. It follows that

$$\bar{\nabla}_X N \cdot Y = -N \cdot \mathrm{I\!I}(X, Y) \quad \text{on } M. \tag{20.2}$$

This is called Weingarten's relation. Note that in the codimension 1 case $d = n + 1$ we can choose a unit normal field N and define

$$h(X, Y) = N \cdot \mathrm{I\!I}(X, Y) = -\bar{\nabla}_X N \cdot Y = h_{ij} X^i Y^j \tag{20.3}$$

20.7 Curvature

The equality

$$\nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z = \nabla_{[X,Y]} Z + R(X, Y) Z \tag{20.1}$$

[5] to M

[6] This can be done, certainly locally, why?

defines $R(X,Y)Z$ for tangent vector fields X, Y, Z. You may verify that $R(X,Y)Z$ is multilinear in X, Y, Z over the scalar fields on M. In the case $M = \mathbb{R}^n = \mathbb{R}^d$ you will find that $R(X,Y)Z \equiv 0$. The standard way to write $R(X,Y)Z$ in local coordinates u is

$$(R(X,Y)Z)^\alpha = R^\alpha_{ijk} Z^i X^j Y^k. \tag{20.2}$$

So Z comes first[7] and then X and Y. Using (20.1) and writing

$$\Gamma^\alpha_{ij,k} = \frac{\partial \Gamma_{ij}}{\partial u^k}$$

you should verify that[8]

$$R^\alpha_{ijk} = \Gamma^\beta_{ik}\Gamma^\alpha_{\beta j} - \Gamma^\beta_{ij}\Gamma^\alpha_{\beta k} + \Gamma^\alpha_{ik,j} - \Gamma^\alpha_{ij,k} \tag{20.3}$$

and the zero ijk and jk cyclic sums

$$R^\alpha_{ijk} + R^\alpha_{kij} + R^\alpha_{jki} = 0 = R^\alpha_{ijk} + R^\alpha_{ikj} \tag{20.4}$$

If W is another tangent field then[9]

$$Rm(X,Y,Z,W) = R(X,Y)Z \cdot W = R^\alpha_{ijk} Z^i X^j Y^k g_{\alpha l} W^l = R_{lijk} W^l Z^i X^j Y^k, \tag{20.5}$$

which has the symmetries

$$Rm(X,Y,Z,W) + Rm(Y,Z,X,W) + Rm(Z,X,Y,W) = 0,$$

$$Rm(X,Y,Z,W) + Rm(Y,X,Z,W) = 0 = Rm(X,Y,Z,W) + Rm(X,Y,W,Z)$$

(the second one obtained from $Rm(X,Y,Z,Z) = 0$), implying

$$Rm(X,Y,Z,W) = Rm(Z,W,X,Z)$$

In the 2-dimensional case $n = 2$ the only possible nonzero entries of R_{ijk} are

$$R_{1212} = R_{2121} = -R_{1221} = -R_{2112}$$

In the codimension 1 case

$$R_{lijk} = h_{ik}h_{lj} - h_{ij}h_{lk}$$

[7] As if we would have prefered the notation $ZR(X,Y)$

[8] note the order ijk in the minus terms and the $j \leftrightarrow k$ relation with the plus terms

[9] $lijk$ = dead body, as if we would have prefered the notation $W \cdot ZR(X,Y)$

consists of all the 2×2 determinants you can get from the matrix h_{ij}. Note that similarly

$$(W \cdot X)(Y \cdot Z) - (W \cdot Y)(X \cdot Z) = \underbrace{(g_{ik}g_{lj} - g_{ij}g_{lk})}_{G_{lijk}} W^l Z^i X^j Y^k, \qquad (20.6)$$

in which G_{lijk} has the same symmetry properties as R_{lijk} (and depends only on G_{1212} if $n = 2$).

For submanifolds you can verify from the definitions that

$$Rm(X, Y, Z, W) = \mathrm{I\!I}(X, W)\mathrm{I\!I}(Y, Z) - \mathrm{I\!I}(X, Z)\mathrm{I\!I}(Y, W), \qquad (20.7)$$

which in the codimension 1 case (20.3) reduces to

$$Rm(X, Y, Z, W) = h(W, X)h(Y, Z) - h(W, Y)h(X, Z),$$

so

$$Rm(X, Y, Z, W) = \underbrace{(h_{ik}h_{lj} - h_{ij}h_{lk})}_{R_{lijk}} W^l Z^i X^j Y^k, \qquad (20.8)$$

Gauss computed this expression for R_{lijk} from $x_{ijk} = x_{ikj}$, see Chapter 10 in Schaum's Differential Geometry book by Martin Lipschutz. The Gauss curvature of a surface in $\mathrm{I\!R}^d$ is the scalar ratio between (20.7) and (20.6). In $\mathrm{I\!R}^3$ this is the scalar ratio between (20.8) and (20.6).

20.8 Geodesic curves

A smooth curve $\gamma(t) \in M$ may require several coordinate patches to describe it. For the moment we assume that it can be described by one coordinate patch. If

$$\gamma : [a, b] \ni t \to u(t) \to x(u(t)) \in M$$

is such a curve in M, then its velocity is given by

$$\dot{\gamma} = \frac{\partial x}{\partial u^1}\dot{u}^1 + \cdots + \frac{\partial x}{\partial u^n}\dot{u}^n = \sum_{i=1}^{n} \dot{u}^i \frac{\partial x}{\partial u^i} = \sum_{i=1}^{n} \dot{u}^i x_i.$$

Think of $\dot{\gamma}$ as a vector at the point $x = \gamma(t)$ in M. For every t this vector is tangent to M, and written as a linear combination of the tangent vectors obtained from the parameterisation:

$$x_1 = \frac{\partial x}{\partial u^1}, \ldots, x_n = \frac{\partial x}{\partial u^n}.$$

Its length l is given by

$$l = \int_a^b |\dot\gamma(t)|\,dt = \int_a^b \sqrt{\dot\gamma(t)\cdot\dot\gamma(t)}\,dt = \int_a^b \sqrt{x_i\dot u^i\cdot x_j\dot u^j}\,dt$$

$$= \int_a^b \sqrt{\dot u^i\dot u^j g_{ij}(u)}\,dt$$

We will work with another quantity, called the energy, which involves an L as in Chapter 1. Since I prefer to have u in L, my u's are the γ's in the book. My $\gamma(t)$ is what is $c(t)$ in the book. The energy is defined by

$$E = \frac{1}{2}\int_a^b |\dot\gamma(t)|^2\,dt = \frac{1}{2}\int_a^b \dot\gamma(t)\cdot\dot\gamma(t)\,dt = \frac{1}{2}\int_a^b x_i\dot u^i\cdot x_j\dot u^j\,dt$$

$$= \frac{1}{2}\int_a^b \dot u^i\dot u^j g_{ij}(u)\,dt = \int_a^b L(u(t),\dot u(t))\,dt,$$

in which

$$L = L(u,p) = \frac{1}{2}p^i p^j g_{ij}(u). \tag{20.1}$$

Playing with the estimate

$$\int_a^b |\dot\gamma(t)|\,dt = \int_a^b 1\,|\dot\gamma(t)|\,dt \le \sqrt{\int_a^b 1^2\,dt}\sqrt{\int_a^b |\dot\gamma(t)|^2\,dt}$$

and reparameterisation of γ to make $|\dot\gamma|$ constant you should easily conclude that minimizers of l are minimizers of E and vice versa if we keep $[a,b]$ fixed.

The Euler-Lagrange equations for E involve the derivatives of g_{ij} and come out as

$$\ddot u^i + \Gamma^i_{\alpha\beta}\dot u^\alpha\dot u^\beta = 0 \tag{20.2}$$

and are called the geodesic equations. Indeed,

$$\Gamma^i_{\alpha\beta} = \frac{1}{2}g^{ik}\left(g_{\alpha k,\beta} + g_{\beta k,\alpha} - g_{\alpha\beta,k}\right),$$

the symbols computed in (20.2). You should repeat this calculation without looking at the notes above. What is the conservation law for this system?

A nice example is a surface M which is described by a single set of coordinates $u\in\mathbb{R}^2$ with a metric

$$g_{ij}(u) = g(|u|)\delta_{ij} \tag{20.3}$$

in which $u \to g(|u|)$ is smooth and positive[10]. You can write the geodesic equations as in the book (2.1.27). In a special case the example is related to stereographic projection through

$$u^1 = \frac{x^1}{1-x^3}, \quad u^2 = \frac{x^2}{1-x^3},$$

which you may prefer as

$$u = \frac{x}{1-z}, \quad v = \frac{y}{1-z}$$

without indices.

- Verify that large circles on $x^2 + y^2 + z^2$ correspond to circles in the uv-plane. Hint: describe the large circles as $z = ax + by$ and avoid goniometric functions.
- The large circles not contained in this description are the vertical great circles which correspond to lines through the origin in the uv-plane. Assuming unit speed for both the vertical great circles and lines through the origin derive the formula for $g(|u|)$.

We return to (20.3) with general $g(|u|)$.

- Why are geodesics through the origin straight lines?
- Take a geodesic line parametrized by t such that $t = 0$ corresponds to $(0, 0)$ and that the speed in $(0, 0)$ is equal to 1. Use the conservation law to derive a first order equation for $R(t) = |u(t)|$ and solve it.
- Examine how long it takes for the geodesic curve to reach infinity. What is the condition on $g(|u|)$ to reach infinity in finite time? This should involve some integral with g. Do the same in dimension $n > 2$? Is there a difference?
- Can you cook up an example for which the geodesic cannot cross $|u| = 1$? Can you classify these examples?
- Incidentally, what is the Gauss curvature for metrics of the form (20.3) in $\mathbb{R}^2$?

[10] implying $0 = g'(0) = g'''(0) = g''''''(0) = \cdots$

20.9 The Jacobi equations

Consider the Lagrangian (20.1).

- Show that the Jacobi equations (20.4) for (20.1) are

$$\ddot{\eta}^i + 2\Gamma^i_{jk}\dot{u}^j\dot{\eta}^k + \Gamma^i_{jk,l}\dot{u}^j\dot{u}^k\eta^l = 0 \tag{20.4}$$

Both $\dot{u}^i(t)$ and $\eta^i(t)$ define vector fields along $\gamma(t) = x(u(t))$ in $M \in \mathbb{R}^d$ tangent to M through

$$\dot{\gamma}(t) = \dot{u}^i(t)x_i(u(t)), \quad \eta(t) = \eta^i(t)x_i(u(t))$$

The Jacobi equations are much more transparent if we work with the tangential parts D_tV of the time derivatives of such vector fields

$$V(\gamma(t)) = V^i(t)x_i(u(t))$$

- Derive that

$$D_tV = (D_tV)^j x_j \quad \text{with} \quad \dot{V}^j + V^\alpha\Gamma^j_{\alpha\beta}\dot{u}^\beta$$

- Derive that the geodesic equation (20.2) may be written as

$$D_t\dot{\gamma} = 0, \quad \dot{\gamma} = \dot{u}^i x_i$$

- Derive that (20.4) may be written as

$$(D_t^2\eta)^i + \dot{u}^\alpha R^i_{\alpha\beta k}\eta^\beta\dot{u}^k = 0, \quad \text{i.e.} \quad D_t^2\eta + R(\eta,\dot{\gamma})\dot{\gamma} = 0$$

21 Newton's method the hard way

Some time ago I was asked to give a talk on the work of Nash. I apologise for doing something else instead. On a family of theorems that bear his name and proofs Nash never wrote. In these notes I describe how Newton's method can be adapted in the case that the map

$$u \to u - f'(u)^{-1} f(u) \tag{1}$$

is not defined as a map from a Banach space X to itself. The resulting theorems are called HARD Implicit Function Theorems. My purpose here is to demystify the terminology and present a simple proof of convergence for a modification of Newton's method in such a case. Observe that a direct proof of the Inverse Function Theorem for a continuously differentiable function f amounts to solving the equation $f(u) = v$ for u given small v under the assumption that $f(0) = 0$, using the map

$$u \to u + f'(0)^{-1}(v - f(u)) \tag{2}$$

which is contractive if $f'(0)^{-1} : X \to X$ exists as a continuous linear map.

The proof of the Implicit Function Theorem for solving equations like $f(u, v) = 0$ in the form $u = u(v)$ if $f(0, 0) = 0$ and the partial derivative of f with respect to u is invertible in $(u, v) = (0, 0)$ is similar. To show that (2) produces a local solution $u = u(v)$ which is continuously differentiable the only regularity on f that has to be assumed is that $u \to f'(u)$ is continuous, as only $f'(u)$ is needed in the calcutions and estimates. Newton's method, which employs a suitable inverse of $f'(u)$ for all u in some (say the unit) ball B in X, relies an Taylor's theorem with a quadratic remainder and therefore the assumption that also $u \to f''(u)$ be continuous is required.

21.1 Newton's method: a convergence proof

I will modify the treatment in [KP][1] which begins with a somewhat alternative treatment of Newton's method in the standard case. So to warm up consider an equation of the form $f(u) = 0$ in which $f : B \to X$ is a twice continuously differentiable function defined on the open unit ball B in a Banach space X, with first and second order derivative satisfying bounds

$$||f'(u)|| \leq M_1 \quad \text{and} \quad ||f''(u)|| \leq M_2 \quad \forall u \in B. \tag{3}$$

The general case of Banach spaces is really not that different from the case in which $X = \mathbb{R}$, which you may think of in what follows below. Simply take $B = (-1, 1)$ and replace all norms by absolute values.

[1] Krantz & Parks, The Implicit Function Theorem, Birkhäuser 2003.

What we need is that Taylor's theorem with a second order remainder,

$$f(u_n) = \underbrace{f(u_{n-1}) + f'(u_{n-1})(u_n - u_{n-1})}_{\text{linear aproximation}} + Q_f(u_{n-1}, u_n), \tag{4}$$

in which

$$||Q_f(u_{n-1}, u_n)|| \leq \frac{M_2}{2}||u_n - u_{n-1}||^2, \tag{5}$$

applies to a sequence of iterates $u_n \in B$. For the standard Newton method one does not explicitly need the bound on $f'(u)$ in (3) which says that the linear map $f'(u) : X \to X$ satisfies

$$||f'(u)v|| \leq M_1||v|| \quad \forall u \in B \quad \forall u \in X, \tag{6}$$

but a similar bound

$$||L(u)|| \leq C \tag{7}$$

for maps $L(u)$, that act as right inverses of $f'(u)$ in the sense that

$$f'(u_{n-1})L(u_{n-1})f(u_{n-1}) = f(u_{n-1}), \tag{8}$$

is essential. Writing

$$p_n = ||u_n - u_{n-1}|| \text{ and } q_n = ||f(u_n)|| \tag{9}$$

the Newton scheme

$$u_n = u_{n-1} - L(u_{n-1})f(u_{n-1}) \quad (n \in \mathbb{N}), \tag{10}$$

starting with $u_0 = 0$, then defines $u_n \in B$ as long as

$$p_1 + p_2 + \cdots + p_n < 1, \tag{11}$$

and the inequalities

$$p_n \leq Cq_{n-1} \text{ and } q_n \leq \frac{1}{2}M_2 p_n^2 \tag{12}$$

are immediate from (4,5,10). Note that (10) kills the linear approximation in (4). The inequalities in (12) are complemented by

$$q_0 = ||f(0)|| \quad \text{and} \quad p_1 \leq Cq_0 = C||f(0)||. \tag{13}$$

21.2 The optimal result

Clearly (12) and (13) combine as

$$p_n \le \mu p_n^2 \quad \text{with} \quad \mu = \frac{1}{2}MC \quad \text{and} \quad p_1 \le C||f(0)||, \tag{14}$$

and the condition to be stated is which $\bar{P} = \bar{P}(\mu)$ guarantees that the implication

$$C||f(0)|| < \bar{P} \implies \sum_{n=1}^{\infty} p_n < 1 \tag{15}$$

holds. The larger $\bar{P}$ the stronger the statement in the sense that larger $||f(0)||$ are allowed to obtain a solution $u = \bar{u} \in B$ of $f(u) = 0$ via (10) with $u_0 = 0$. Note that with $C||f(0)|| \le \bar{P}$ the same conclusion will hold if only one of all the inequalites in the estimates below is strict, which will inevatibly be the case of course.

Obviously the smallest $\bar{P}$ we can get follows from replacing the three inequalities in (14) and (15) by inequalities. This leads to

$$p_n = \mu p_{n-1}^2 \ \text{ for } \ n \in \mathbb{N}; \quad p_1 = \bar{P}\,; \quad \sum_{n=1}^{\infty} p_n = 1. \tag{16}$$

Via $\xi_n = p_n$ and $\xi_n = \xi_{n-1}^2$ this is easily seen to be equivalent to

$$\mu = G(\mu\bar{P}) \quad \text{with} \quad G(\xi) = \xi + \xi^2 + \xi^4 + \xi^8 + \xi^{16} + \cdots \tag{17}$$

but this does not yield a simple formula for $\bar{P} = \bar{P}(\mu)$.

21.3 A suboptimal result

A rough estimate

$$G(\xi) < \xi + \xi^2 + \xi^3 + \xi^4 + \xi^5 + \cdots = \frac{\xi}{1-\xi} \tag{18}$$

leads to a simple but suboptimal formula:

$$\bar{P} = \frac{1}{1+\mu} \quad \text{or} \quad \mu = \frac{1}{\bar{P}} - 1. \tag{19}$$

21.4 Alternative proof of convergence

The alternative approach to (12) and (13) in [KP] is not to solve the corresponding system with equalities but to derive an estimate of the form

$$p_n \leq e^{-\gamma\lambda^n} \tag{20}$$

via induction starting from

$$p_1 \leq C||f(0)|| < \bar{P} = e^{-\gamma\lambda}, \tag{21}$$

with choices of γ and λ that guarantee both

$$\sum_{n=1}^{\infty} e^{-\gamma\lambda^n} \leq 1 \tag{22}$$

as well as that the induction step can be done via

$$p_{n-1} \leq e^{-\gamma\lambda^{n-1}} \implies p_n \leq \mu p_{n-1}^2 \leq \underbrace{\mu e^{-2\gamma\lambda^{n-1}} \leq e^{-\gamma\lambda^n}}_{\text{should hold for all } n\geq 1},$$

which is the case if

$$\ln\mu \leq \gamma\lambda^{n-1}(2-\lambda) \quad \forall n \geq 1.$$

21.5 The optimal alternative result

For a given μ this is equivalent to

$$\ln\mu \leq \gamma\lambda(2-\lambda) \quad \text{and} \quad \lambda \leq 2 \tag{23}$$

if we make the obvious restriction that γ and λ be positive. Conditions (21) and (23) suggest $\alpha = \gamma\lambda$ and λ as the more relevant parameter so we have to pick $\alpha > 0$ and $1 < \lambda \leq 2$ with

$$\ln\mu \leq \alpha(2-\lambda), \quad \sum_{n=0}^{\infty} e^{-\alpha\lambda^n} \leq 1 \quad \text{and} \quad \bar{P} = e^{-\alpha} \quad \text{maximal.} \tag{24}$$

For $\mu > 1$ the inequalities define a region in the first quadrant of the λ, α-plane bounded by the two curves given by

$$\ln\mu = \alpha(2-\lambda) \ \text{ and } \ \sum_{n=0}^{\infty} e^{-\alpha\lambda^n} = 1, \tag{25}$$

which intersect in one point.

This point defines the minimal value of $\alpha = -\ln \bar{P}$ via

$$1 = \sum_{n=0}^{\infty} e^{-\alpha\lambda^n} = \sum_{n=0}^{\infty} \bar{P}^{\lambda^n} = \sum_{n=0}^{\infty} \bar{P}^{(2+\frac{\ln \mu}{\ln \bar{P}})^n}$$

if $\mu > 1$. The curve defined by

$$1 = \sum_{n=0}^{\infty} \bar{P}^{(2+\frac{\ln \mu}{\ln \bar{P}})^n} \quad \text{and} \quad \mu \geq 1 \tag{26}$$

hits the curve defined by (17) in $\mu = 1$ and lies below (17) of course, but above (19) in view of

$$\mu = \frac{1}{\bar{P}} - 1 \implies \sum_{n=0}^{\infty} \bar{P}^{(2+\frac{\ln \mu}{\ln \bar{P}})^n} = \sum_{n=0}^{\infty} \bar{P}^{(1+\frac{\ln(1-\bar{P})}{\ln \bar{P}})^n} < \underbrace{\sum_{n=0}^{\infty} \bar{P}^{1+n\frac{\ln(1-\bar{P})}{\ln \bar{P}}}}_{\text{a geometric series}} = 1.$$

For $\mu \leq 0$ the optimal choice of $\bar{P}$ via (24) is given by

$$\sum_{n=0}^{\infty} \bar{P}^{2^n}.$$

21.6 A suboptimal alternative result

A more explicit formula is again obtained via a rough estimate

$$\sum_{n=1}^{\infty} e^{-\gamma\lambda^n} \leq \underbrace{\sum_{n=1}^{\infty} e^{-\gamma(1+n(\lambda-1))}}_{\text{a geometric series}} = \frac{e^{-\gamma\lambda}}{1 - e^{-\gamma(\lambda-1)}} = \frac{e^{-\alpha}}{1 - e^{\gamma} e^{-\alpha}} \tag{27}$$

and replacing (24) by

$$\ln \mu \leq \alpha(2 - \lambda), \quad \lambda \geq \frac{\alpha}{\ln(e^{\alpha} - 1)} \quad \text{and} \quad \bar{P} = e^{-\alpha} \quad \text{maximal.}$$

This leads to

$$\mu = e^{\alpha(2-\lambda)} = \frac{1}{\bar{P}^{2-\lambda}} = \frac{1}{\bar{P}^{2+\frac{\ln \bar{P}}{\ln(\frac{1}{\bar{P}}-1)}}} = \bar{P}^{\frac{\ln(\bar{P})-2\ln(1-\bar{P})}{\ln(1-\bar{P})-\ln(\bar{P})}}$$

so that

$$1 \leq \mu = \frac{1}{\bar{P}^{2+\frac{\ln \bar{P}}{\ln(\frac{1}{\bar{P}}-1)}}} < \frac{1}{\bar{P}} - 1 \tag{28}$$

defines another curve with

$$\bar{P} \leq \frac{3-\sqrt{5}}{2},$$

which is below the three curves above, but to leading coincides with them in the limit $\mu \to \infty$ and $\bar{P} \to 0$.

21.7 A lousy alternative result

The even rougher estimate used in [KP] via

$$\sum_{n=1}^{\infty} e^{-\gamma\lambda^n} \leq \sum_{n=1}^{\infty} e^{-n\gamma(\lambda-1)}$$

is to be avoided as at some point below the treatment of ill-behaved Newton's methods will show.

21.8 A much better suboptimal alternative result

Actually the first rough estimate above works better with α dan with γ, as I only noticed May 21. Directly in terms of γ and λ we have

$$\sum_{n=1}^{\infty} e^{-\gamma\lambda^n} = \sum_{n=1}^{\infty} e^{-\gamma\lambda\lambda^{n-1}} \leq \sum_{n=1}^{\infty} e^{-\gamma\lambda(1+(n-1)(\lambda-1))} = \frac{e^{-\gamma\lambda}}{1-e^{-\gamma\lambda(\lambda-1)}} \leq 1 \quad (29)$$

if

$$2-\lambda \leq \frac{\ln(e^{\gamma\lambda}-1)}{\gamma\lambda} = \frac{\ln(e^{\alpha}-1)}{\alpha},$$

so that we arrive at

$$\ln\mu \leq \alpha(2-\lambda), \quad \alpha(2-\lambda) \leq \ln(e^{\alpha}-1) \quad \text{and} \quad \bar{P} = e^{-\alpha} \quad \text{maximal.} \quad (30)$$

This is the optimal estimate using the Bernoulli type inequality

$$\lambda^n \geq 1 + n(\lambda - 1). \quad (31)$$

With equality in the final inequality in (29) we arrive at

$$\ln\mu \leq \ln(e^{\alpha}-1) = \ln(\frac{1}{\bar{P}} - 1),$$

which for $\mu > 1$ coincides with (19) and we can forget about the annoying (28) above. Note that factoring out another λ in the exponent in (29) will and cannot help to improve this result, which says that if $\mu > 1$ the bound

$$||f(0)|| \leq \frac{1}{C(\mu+1)}$$

suffices.

This bound may be compared with the bound in [KP], where all constants are named M, for unclear reasons $M > 2$ is assumed, and the $\frac{1}{2}$-coefficient in the Taylor-remainder term is omitted. Since $\mu = \frac{1}{2}CM$ our bounds looks similar to their bound $||f(0)|| \leq M^{-5}$. In the next section the comparison will be a true pain, as [KP] have a formulation in which again all constants are called M with apparently $M > 1$, and the bound on some norm of $f(0)$ (the wrong norm actually) involving M^{-307}. Comparing to the lectures notes of Schwartz from 60 years ago this is hardly an improvement as Schwartz had M^{-202} (also for the wrong norm).

22 Nash' modification of Newton's method

Now that we have seen several small variants of the method to obtain convergence for Newton's method, we consider the problem of solving $f(u) = 0$ in $B \subset X$ in the case that $f : B \to Z$ and $L(u) : Z \to Y$ with X, Y and Z *different* Banach spaces that we assume to belong to a family of spaces denoted by C^k, which we think of as function spaces. Here k denotes the number of possibly fractional derivatives that elements $u \in C^k$ have. Think of k for X, l for Z and m for Y. The goal is to have conditions that guarantee the existence of a solution to $f(u) = 0$ with k-norm smaller than 1, provided $f(0)$ has a norm bounded by some power of M, where M is a universal bound for all constants related to the derivatives of f.

Both [KP] and Schwartz require a very strong norm of $f(0)$ to be bounded, but the treatment below will show that a bound on the l-norm suffices. It should be noted that [KP] more or less copied from Schwartz with some additional details explained. Both formulate a statement for the case that $k > l$, but give a not completely correct proof for the case that $k = l > m$ (without mentioning the difference). The main additional assumption is a natural affine bound for $||L(u)f(u)||_{\bar{m}}$ in terms of $||u||_{\bar{k}}$, for $\bar{m}$ and $\bar{k}$ sufficiently large and $\bar{k} - \bar{m} = k - m$. The ratio

$$N = \frac{\bar{k} - k}{k - m} \tag{1}$$

measures the required higher regularity of the Newton map for the modified scheme described below to still do the job.

Below the norms $u \to ||u||_k$ on C^k are assumed to be monotone increasing in k and we assume that there are linear so-called smoothing operators $S(t)$ parametrized by $t \geq 1$ that satisfy

$$||S(t)u||_k \leq K_{kl} t^{k-l} ||u||_l \quad \text{and} \quad ||(I - S(t))u||_l \leq \frac{K^{kl}}{t^{k-l}} ||u||_k \tag{2}$$

for all $k > l$ in a sufficiently large range as needed in the particular implementation of the modified Newton method presented next. Thus $S(t)$ maps C^l to C^k, with an estimate for the ratio between the norms that grows worse as $S(t)$ approaches the identity I for $t \to \infty$, when I is considered as the embedding $I : C^k \to C^l$. It is convenient to write the norms of $S(t)$ and $I - S(t)$ with subscripts indicating the norms used for u, $S(t)u$ and $(I - S(t))u$. Thus (2) says that

$$||S(t)||_{kl} \leq K_{kl} t^{k-l} \quad \text{and} \quad ||(I - S(t))||_{lk} \leq \frac{K^{kl}}{t^{k-l}}. \tag{3}$$

Besides (3) we assume (now also) a bound M_1^{lk} on $||f'(u)||_{lk}$ and, as before, bounds M_2^{lk} on $||f''(u)||_{lk}$ and C_{ml} on $||L(u)||_{ml}$ for $||u||_k \leq 1$.

22.1 The modified scheme

The idea of Nash was to modify Newton's scheme into

$$u_n = u_{n-1} - S(t_{n-1})L(u_{n-1})f(u_{n-1}), \tag{4}$$

with a suitable choice of $t_n \to \infty$ as $n \to \infty$. In (4) the new factor $S_{n-1} = S(t_{n-1})$ maps $L(u_{n-1})f(u_{n-1})$ back to (the strict subset of smooth functions of) the original domain of f. This comes with a cost which is estimated using the norm of the smoothing operator S_{n-1} in the chain

$$u_{n-1} \in X = C^k \xrightarrow{f} Z = C^l \xrightarrow{L(u_{n-1})} Y = C^m \xrightarrow{S_{n-1}} u_n \in X = C^k.$$

Before we do so let's examine how (4) is modified when combined with (4). We have

$$f(u_n) = \underbrace{f(u_{n-1}) + f'(u_{n-1})(u_n - u_{n-1})}_{\text{vanishes with (10)}} + Q_f(u_{n-1}, u_n)$$

$$= \underbrace{f'(u_{n-1})(I - S_{n-1})L(u_{n-1})f(u_{n-1})}_{\text{because of (4)}} + Q_f(u_{n-1}, u_n),$$

so that, with

$$p_n = ||u_n - u_{n-1}||_k \quad \text{and} \quad q_n = ||f(u_n)||_l,$$

the estimate

$$q_n \leq \underbrace{M_1^{lk}||I - S_{n-1}||_{km}||L(u_{n-1})f(u_{n-1})||_m}_{\text{new error like term}} + \frac{1}{2}M_2^{lk}p_n^2 \tag{5}$$

holds.

22.2 The new error term

The third factor in the error like term in (5) will have to be controled using some assumption on the map

$$u \to L(u)f(u)$$

which was not needed in the case of (10) and that should guarantee that quadratic term in (5) will still allow us to establish a conclusion like (15). Clearly this is impossible if $m \leq k$ because we can only make $||I - S_n||_{km}$ small if $k < m$. Nash' solution was to replace m by a (much) larger $\bar{m}$ and assume an otherwise natural affine estimate of the form

$$||L(u)f(u)||_{\bar{m}} \leq A_{\bar{m}\bar{k}}(1 + ||u||_{\bar{k}}) \tag{6}$$

with

$$\bar{k} - \bar{m} = k - m,$$

which requires an additional estimate for

$$r_n = 1 + ||u_n||_{\bar{k}} \tag{7}$$

to be used in combination with

$$q_n \leq M_1^{lk} \underbrace{||I - S_{n-1}||_{k\bar{m}}}_{\text{controled by (3)}} r_{n-1} + \frac{1}{2} M_2^{lk} p_n^2 \tag{8}$$

and the estimate for p_n. Via (4) the latter now reads

$$p_n \leq ||S_{n-1}||_{km} C_{ml} q_{n-1} \tag{9}$$

because $||L(u_{n-1})f(u_{n-1})||_m \leq C_{ml} q_{n-1}$.

The additional estimate needed for r_n also follows from (4). In view of

$$||u_n - u_{n-1}||_{\bar{k}} \leq ||S_{n-1}||_{\bar{k}\bar{m}} ||L(u_{n-1}f(u_{n-1})||_{\bar{m}} \leq ||S_{n-1}||_{\bar{k}\bar{m}} A_{\bar{m}\bar{k}}(1 + ||u_{n-1}||_{\bar{k}})$$

we have

$$1 \leq r_n \leq 1 + A_{\bar{m}\bar{k}} \sum_{j=1}^{n} ||S_{j-1}||_{\bar{k}\bar{m}} r_{j-1}. \tag{10}$$

The "error"terms accumulate but can be kept under control as we shall see below.

The system of inequalities (9,8,10) and initial inequalities for q_0, $r_0 = 1$ and r_1 allows again estimates of the form (20), provided $\bar{k} - k = \bar{m} - m$

is sufficiently large in terms of (1). The idea is to get the first term in (8) controled by the right hand side of

$$p_n^2 \leq e^{-2\gamma\lambda^n}$$

in the induction argument, so that the norm $||S_n||_{km}$ in (9) can be chosen not to large so as still to have (20) with n if it already holds with $n-1$. To do so we need a control on $||S_{n-1}||_{km}$ of the same form and this is established by setting

$$t_{n-1} = e^{\beta\lambda^{n-1}} \tag{11}$$

with $\beta > 0$ to be chosen in terms of γ. Note that this gives λ^n in the exponents of the exponential bounds for S_n and $I - S_n$.

Here we choose to keep λ as a parameter in a range as large as possible, like we did in the analysis of (10). Clearly we can only complete the argument if we also specify a bound on r_n to be established in the course of the argument, and this bound has to be of the same form as the bound chosen for S_n. Thus we look for a proof that

$$p_n \leq e^{-\gamma\lambda^n} \quad \text{and} \quad r_n \leq e^{\delta\lambda^n} \tag{12}$$

with $\delta > 0$. We note that the proof presented in [KP] the choice $\delta = \gamma$ and $\lambda = \frac{3}{2}$ dates back to Schwartz's lecture notes. As we shall see below this is not quite the optimal choice.

22.3 The system of inequalities

With (11) we have the system of inequalities

$$p_n \leq K_{km} e^{(k-m)\beta\lambda^{n-1}} C_{ml} q_{n-1}; \tag{13}$$

$$q_n \leq M_1^{lk} K^{k\bar{m}} e^{(k-\bar{m})\beta\lambda^{n-1}} A_{\bar{m}\bar{k}} \underbrace{r_{n-1}}_{\leq e^{\delta\lambda^{n-1}}} + \frac{1}{2} M_2^{lk} \underbrace{p_n^2}_{\leq e^{-2\gamma\lambda^n}}; \tag{14}$$

$$1 \leq r_n \leq 1 + \underbrace{A_{\bar{m}\bar{k}} K_{\bar{k}\bar{m}}}_{\mu_3} \sum_{j=1}^{n} e^{(\bar{k}-\bar{m})\beta\lambda^{j-1}} \underbrace{r_{j-1}}_{\leq e^{\delta\lambda^{j-1}}}, \tag{15}$$

and we aim for a proof of (12) via induction, using the underbraced estimates in the three inequalities above as induction hypothesis. In (14) the estimate of the first term is controled by the estimate of the second term if

$$e^{(k-\bar{m})\beta\lambda^{n-1}} e^{\delta\lambda^{n-1}} \leq e^{-2\gamma\lambda^n},$$

requiring

$$(\bar{m} - k)\beta \geq \delta + 2\gamma\lambda, \tag{16}$$

which says that in the λ, β-plane we must be above a line that comes down as $\bar{m}$ is increased.

Combining the first two inequalities we arrive at

$$p_n \leq e^{(k-m)\beta\lambda^{n-1}}(\mu_1 e^{(k-\bar{m})\beta\lambda^{n-2}} r_{n-2} + \mu_2 p_{n-1}^2) \quad r_n \leq 1 + \mu_3 \sum_{j=0}^{n-1} e^{(\bar{k}-\bar{m})\beta\lambda^j} r_j, \tag{17}$$

the constants μ_{123} given by

$$\mu_1 = \underbrace{K_{km} C_{ml}}_{C} M_1^{lk} \underbrace{K^{k\bar{m}} A_{\bar{m}\bar{k}}}_{A}, \quad \mu_2 = \frac{1}{2} \underbrace{K_{km} C_{ml}}_{C} M_2^{lk}, \quad \mu_3 = \underbrace{K_{\bar{k}\bar{m}} A_{\bar{m}\bar{k}}}_{\bar{A}}. \tag{18}$$

22.4 Estimating the increments

Under the assumption that (16) holds, the induction hypotheses for p_{n-1} and r_{n-2} produce the desired inequality for p_n from (17) if

$$(\mu_1 + \mu_2) e^{(k-m)\beta\lambda^{n-1}} e^{-2\gamma\lambda^{n-1}} \leq e^{-\gamma\lambda^n}.$$

Thus we must have

$$\ln(\mu_1 + \mu_2) \leq -(k-m)\beta\lambda^{n-1} + 2\gamma\lambda^{n-1} - \gamma\lambda^n$$

for all $n \geq 2$. As in the case of the standard Newton scheme, this leads to

$$\ln(\mu_1 + \mu_2) \leq \lambda(\gamma(2-\lambda) - (k-m)\beta) \quad \text{with} \quad (k-m)\beta \leq \gamma(2-\lambda), \tag{19}$$

a sharp upper bound for β that we need to stay away from if we don't want to impose that $\mu_1 + \mu_2 \leq 1$.

As sufficient condition for

$$\sum_{n=1}^{\infty} p_n < 1$$

we can use the optimal condition found using Bernoulli's inequality, namely

$$\lambda\gamma(2-\lambda) \leq \ln(e^{\gamma\lambda} - 1). \tag{20}$$

22.5 Estimating the error terms

For the inductive construction of the upper bound for r_n we set

$$b = (\bar{k} - \bar{m})\beta = (k - m)\beta > 0 \tag{21}$$

and conclude from the inequality in (17) that (shifting the index)

$$r_n \leq 1 + \mu_3 \sum_{j=0}^{n-1} e^{b\lambda^j} r_j \leq 1 + \mu_3 \sum_{j=0}^{n-1} e^{b\lambda^j} e^{\delta\lambda^j}$$

in view of the induction assumption for (all) smaller n. Thus we need the inequality

$$1 + \mu_3 \sum_{j=0}^{n-1} e^{(b+\delta)\lambda^j} \leq e^{\delta\lambda^n} \tag{22}$$

for all $n \geq 2$. Recall that we start with $r_0 = 1 \leq e^{\delta}$ and

$$1 \leq r_1 \leq e^{\delta\lambda} \quad \text{(and also } p_1 \leq e^{\gamma\lambda} \text{ of course)} \tag{23}$$

via a smallness assumptions on q_0 still to be discussed.

Dividing by the right hand side, (22) is equivalent to

$$e^{-\delta\lambda^n} + \mu_3 (e^{(b+\delta-\delta\lambda)\lambda^{n-1}} + e^{-\delta\lambda^n} \sum_{j=0}^{n-2} e^{(b+\delta)\lambda^j}) \leq 1 \tag{24}$$

in which we have separated the probably dominant term with $j = n-1$ from the sum. Neglecting the sum in (24) a sufficient (and in any case necessary) condition for the induction step to work for all $n \geq 2$ would be that

$$\ln \mu_3 + (b + \delta - \delta\lambda)\lambda^{n-1} \leq 0 \quad \text{with} \quad b \leq \delta(\lambda - 1), \tag{25}$$

so that in particular we now need to impose two inequalities on b, namely

$$b < \delta(\lambda - 1) \quad \text{and} \quad b < \gamma(2 - \lambda), \tag{26}$$

the latter being the (strict) inequality from (19).

These two bounds severely restrict the bound in (16), which in terms of b becomes

$$\frac{\bar{m} - k}{k - m} b \geq \delta + 2\gamma\lambda, \tag{27}$$

and this does not really depend on how we turn the necessary condition (25) into a sufficient condition, which we do next, rewriting it as

$$e^{-\delta\lambda^n} + \mu_3(e^{(b+\delta-\delta\lambda)\lambda^{n-1}} + \sum_{j=0}^{n-2} e^{(b+\delta-\delta\lambda^{n-j})\lambda^j}) \le 1.$$

In view of (26) and using Bernoulli's inequality (31) the left hand side is smaller than

$$e^{-\delta\lambda^2} + \mu_3(e^{(b+\delta-\delta\lambda)\lambda} + \sum_{j=0}^{n-2} e^{(b+\delta-\delta\lambda^2)\lambda^j}) <$$

$$e^{-\delta\lambda^2} + \mu_3(e^{(b+\delta-\delta\lambda)\lambda} + \sum_{j=0}^{\infty} e^{(b+\delta-\delta\lambda^2)(1+j(\lambda-1))}) <$$

$$e^{-\delta\lambda^2} + \mu_3(e^{(b+\delta-\delta\lambda)\lambda} + \frac{e^{b+\delta-\delta\lambda^2}}{1-e^{(\lambda-1)(b+\delta-\delta\lambda^2)}}),$$

in which we used that $b+\delta-\delta\lambda^2 < b+\delta-\delta\lambda < 0$. Thus we arrive at

$$e^{-\delta\lambda^2} + \mu_3 e^{(b+\delta-\delta\lambda)\lambda}(1 + \frac{e^{-(b+\delta)(\lambda-1)}}{1-e^{(\lambda-1)(b+\delta-\delta\lambda^2)}}) \le 1 \tag{28}$$

Note that the first term on the right hand side of (31) is essential here. Without this first term the numerator, which is the first term ($j=0$) in the geometric series, would be 1 and we be stuck, as there would be no way to get a statement without an a priori bound on μ_3. We note that in [KP] the proof is without the 1 in (31) but an accidental mistake of computing the series with $j=1$ as the first term "allows" to conclude. Technically speaking that proof is incorrect[1].

The quickest way to finish is to estimate the sum of the geometric series by a fixed constant, rewriting it as

$$\frac{e^{-s}}{1-e^{s-S}} = \frac{e^S}{e^s(e^S-e^s)}$$

with

$$s = (b+\delta)(\lambda-1) \le \delta\lambda(\lambda-1) = s_0 < S = \delta\lambda^2(\lambda-1).$$

Provided

$$2e^{s_0} \le e^S \quad \text{or} \quad \ln 2 \le \delta\lambda(\lambda-1)^2,$$

[1] And it is not a proof of the theorem actually stated.

this expression is monotone decreasing in s on $[0, s_0]$ and thus

$$\frac{e^{-(b+\delta)(\lambda-1)}}{1-e^{(\lambda-1)(b+\delta-\delta\lambda^2)}} \leq \frac{1}{1-e^{-\delta(\lambda-1)\lambda^2}} \leq 2.$$

We conclude that

$$e^{-\delta\lambda^2} + 3\mu_3 e^{(b+\delta-\delta\lambda)\lambda} \leq 1 \quad \text{suffices if} \quad \ln 2 \leq \delta\lambda(\lambda-1)^2, \tag{29}$$

and the first inequality in (29) certainly holds if it holds with the first exponential replaced by the larger second exponential. Thus we arrive at

$$\ln(1+3\mu_3) \leq \lambda(\delta(\lambda-1)-b) \quad \text{and} \quad \ln 2 \leq \delta\lambda(\lambda-1)^2 \tag{30}$$

as the final condition needed.

22.6 Sufficient conditions for a convergence result

Summing up, with the condition on q_0 still to be imposed we arrive at

$$\lambda\gamma(2-\lambda) \leq \ln(e^{\gamma\lambda}-1), \tag{31}$$

$$\ln 2 \leq \delta\lambda(\lambda-1)^2, \tag{32}$$

$$(\bar{m}-k)\beta \geq \delta + 2\gamma\lambda, \tag{33}$$

$$(k-m)\beta < \gamma(2-\lambda) \quad \text{and} \quad (\bar{k}-\bar{m})\beta < \delta(\lambda-1) \tag{34}$$

as conditions on the parameters that we still have to choose.

The first inequality, (31), is to have the sum of the increments, and thereby the solution, bounded by 1 in the l-norm. Of course it can be replaced by just asking that

$$\sum_{n=1}^{\infty} e^{-\gamma\lambda^n} \leq 1.$$

The second, (32), was a technical condition to bound the sum of the geometric series in (28) by 2. The third, (33), allows to bound the error term in estimate (14) for q_n by the bound on p_n^2 that has to be established. It involves the choice of sufficiently large $\bar{m}$ and $\bar{k}$ with $\bar{k}-\bar{m} = k-m$.

The last two conditions are strict inequalities that have be chosen sufficiently strict depending on the constants related to f, to allow for an inductive proof of the desired estimates (12) for p_n and r_n. Thus, given μ_1, μ_2, μ_3, we need to choose $1 < \lambda < 2$ and $\gamma, \beta, \delta > 0$ such that

$$\lambda(\gamma(2-\lambda) - (m-k)\beta) \geq \ln(\mu_1 + \mu_2); \tag{35}$$

$$\lambda(\delta(\lambda-1) - (\bar{m}-\bar{k})\beta) \geq \ln(1+3\mu_3). \tag{36}$$

After a simultaneous rescaling of γ, β, δ, this is always possible once the first 5 conditions are satisfied. The inequalities in (34) being strict is essential for convergence of Nash' modified Newton scheme.

Of course we still have to formulate the necessary sufficient bound on $q_0 = ||f(0)||_l$, given the constants in (18) and the choice of parameters above. Recall that

$$\mu_1 + \mu_2 = \underbrace{K_{km}C_{ml}}_{C}(M_1^{lk}\underbrace{K^{k\bar{m}}A_{\bar{m}\bar{k}}}_{A} + \frac{1}{2}M_2^{lk}) = C(M_1 A + \frac{1}{2}M_2)$$

and

$$\mu_3 = K_{\bar{k}\bar{m}}A_{\bar{m}\bar{k}} = \bar{A},$$

with $C, M_1, M_2, A, \bar{A}$ constants related to f and the smoothing operators. From here on we drop the superscripts from the bounds M_1 and M_2 on the first and second derivative of $f : C^k \to C^l$.

22.7 Sufficient convergence condition on initial value

Finally we examine the initial inequalities we need. For p_1 we need, since $u_0 = 0$, that

$$p_1 = ||u_1||_k = ||S_0||_{km}||L(0)||_{ml}||f(0)||_l \leq e^{(k-m)\beta}\underbrace{K_{km}C_{ml}}_{C}||f(0)||_l \leq e^{-\gamma\lambda},$$

while via

$$||u_1||_{\bar{k}} \leq ||S(0)||_{\bar{k}m}||L(0)||_{ml}||f(0)||_l \leq K_{\bar{k}m}e^{(\bar{k}-m)\beta}C_{ml}||f(0)||_l \leq e^{\delta\lambda}$$

we need

$$1 + \underbrace{K_{\bar{k}m}C_{ml}}_{\bar{C}}e^{(\bar{k}-m)\beta}||f(0)||_l \leq e^{\delta\lambda}$$

for r_1. Thus

$$Cq_0 \leq e^{-(k-m)\beta}e^{-\gamma\lambda} \quad \text{and} \quad \bar{C}q_0 \leq e^{-(\bar{k}-m)\beta}(e^{\delta\lambda}-1) \tag{37}$$

are sufficient conditions on

$$q_0 = ||f(0||_l$$

to have a solution of $f(u) = 0$ with $||u||_k < 1$, once the parameters have been chosen according to Section 22.6 to make the induction steps work in the proof of the desired estimates (12) for p_n and r_n.

22.8 The optimal choice of parameters

At this point we compare (35) and (36) to (23). The strict inequalities in (34) are really strict in the sense that the gaps have to be taken sufficiently large large given the explicit constants related to f and $S(t)$. The other two inequalities are not strict. Recalling that $k - m = \bar{k} - \bar{m}$, the coefficient

$$\frac{\bar{m} - k}{k - m} = \frac{\bar{m} - m}{k - m} - 1 = \frac{\bar{m} - m}{\bar{k} - \bar{m}} - 1 = \frac{\bar{k} - k}{k - m} - 1 = N - 1 \tag{38}$$

has to be sufficiently large for the set of allowable b, as defined by (21), to be nonempty. Note that in Nash' strategy to get around the ill-posedness of Newton's method, (1) is the natural definition of N as the ratio of the required increase of smoothness by $\bar{k} - k$ to the loss of smoothness by $m - k$ in $u \to L(u)f(u)$.

The minimal largeness condition on N is obtained by taking the right hand sides of the inequalities in (34) equal to one another, so as to maximize the allowable upper bound for β. Thus we choose $1 < \lambda < 2$ such that

$$\gamma(2 - \lambda) = \delta(\lambda - 1) \quad \text{whence} \quad \lambda = \frac{2\gamma + \delta}{\gamma + \delta} \tag{39}$$

and (33,34) become

$$\frac{4\gamma^2 + 3\gamma\delta + \delta^2}{\gamma + \delta} \leq (N - 1)b < (N - 1)\frac{\gamma\delta}{\gamma + \delta} \tag{40}$$

for

$$b = (k - m)\beta = (\bar{k} - \bar{m})\beta$$

in terms of γ, δ, N, subject to (31,32) which reduce to

$$e^{\frac{2\gamma+\delta}{\gamma+\delta}\frac{\delta}{\gamma+\delta}} + 1 \leq e^{\gamma\frac{2\gamma+\delta}{\gamma+\delta}} \quad \text{and} \quad \ln 2 \leq \delta\,\frac{2\gamma + \delta}{\gamma + \delta}\left(\frac{\gamma}{\gamma + \delta}\right)^2. \tag{41}$$

In particular (40) requires

$$N > \frac{4\gamma}{\delta} + 4 + \frac{\delta}{\gamma} \geq 8, \tag{42}$$

the minimum 8 being realised by

$$\delta = 2\gamma. \tag{43}$$

The further choice of parameters depends on the constants which are as indicated in (18), at the end of Section 22.6 and in (37):

$$C = K_{km}C_{ml};\ \bar{C} = K_{\bar{k}m}C_{ml};\ A = K^{k\bar{m}}A_{\bar{m}\bar{k}};\ \bar{A} = K_{\bar{k}\bar{m}}A_{\bar{m}\bar{k}}; \tag{44}$$

$$\mu_1 + \mu_2 = C(M_1A + \frac{1}{2}M_2); \quad \mu_3 = \bar{A}. \tag{45}$$

We collect these constants in one single constant Θ as

$$\Theta = \frac{3}{4}\max(\ln C + \ln(M_1A + \frac{1}{2}M_2), \ln(1 + 3\bar{A})) \tag{46}$$

and, depending on N, the remaining parameters γ, b have to be chosen to control these constants via

$$\Theta \leq \frac{2\gamma}{3} - b \tag{47}$$

and

$$\frac{14\gamma}{3} \leq (N-1)b < \frac{2\gamma}{3}(N-1), \quad e^{\frac{8}{9}} + 1 \leq e^{\frac{4\gamma}{3}}, \quad \ln 2 \leq \frac{8\gamma}{27}, \tag{48}$$

which is (40,41) with $\delta = 2\gamma$. The last inequality now implies the one preceding it.

For the initial condition q_0 we arrive via (39) at

$$Cq_0 \leq e^{-\gamma\frac{2\gamma+\delta}{\gamma+\delta}}e^{-b} \quad \text{and} \quad \bar{C}q_0 \leq (e^{\delta\frac{2\gamma+\delta}{\gamma+\delta}} - 1)e^{-(\bar{k}-m)\beta} =$$

$$(e^{\delta\frac{2\gamma+\delta}{\gamma+\delta}} - 1)e^{-(\bar{k}-\bar{m})\beta}e^{-(\bar{m}-m)\beta} = (e^{\delta\frac{2\gamma+\delta}{\gamma+\delta}} - 1)e^{-(N+1)b},$$

so that with $\delta = 2\gamma$ the conditions on q_0 reduce to

$$Cq_0 \leq e^{-\frac{4\gamma}{3}}e^{-b} \quad \text{and} \quad \bar{C}q_0 \leq (e^{\frac{8\gamma}{3}} - 1)e^{-(N+1)b}. \tag{49}$$

Setting

$$\rho = \frac{2\gamma}{3}$$

we arrive at

$$\Theta \leq \rho - b, \quad 7\rho \leq (N-1)b, \quad \rho \geq \frac{9}{4}\ln 2,$$

$$\ln C + \ln q_0 \leq -2\rho - b, \quad \ln \bar{C} + \ln q_0 \leq \ln 80 - (N+1)b,$$

as sufficient conditions. Note that we have used the lower bound for ρ to relax the bound on $\bar{C}q_0$.

Choosing

$$N > 8 \quad \text{and} \quad \rho = \frac{N-1}{7}b$$

and using the last lower bound for ρ we arrive at

$$b \geq \max(\frac{63}{4}\frac{\ln 2}{N-1}, \frac{7\Theta}{N-8}) \quad \text{and} \quad q_0 \leq \min(\frac{1}{C}e^{-\frac{2N+5}{7N}b}, \frac{80}{\bar{C}}e^{-(N-1)b}) \quad (50)$$

as sufficient conditions, to be used as: given Θ choose $N > 8$ and $b = (k-m)\beta$ sufficiently large to make the condition on q_0 follow and thereby obtain a solution of $f(u) = 0$ with $||u||_k < 1$.

22.9 Continuity

Given the constants related to f and the smoothing operators we constructed a solution in the open unit k-ball, that is, with $||u||_k < 1$. We did not prove or state that the solution is unique, but it is well defined as the limit of an explicitly constructed sequence shown to be convergent if $||f(0)||_k$ is sufficiently small. The following issue relates to the continuity of the inverse function of f, if it were to exist, since we should naturally also ask for a condition $||f(0)||_k$ guaranteeing the constructed solution to have $||u||_k \leq \varepsilon$. This only changes the condition on the sum of the increments and leads to

$$\gamma\lambda(2-\lambda) \leq \ln(e^{\gamma\lambda} - \frac{1}{\varepsilon})$$

leading to

$$\Theta \leq \frac{2\gamma}{3} - b, \quad \frac{14\gamma}{3} \leq (N-1)b < \frac{2\gamma}{3}(N-1), \quad e^{\frac{8}{9}} + \frac{1}{\varepsilon} \leq e^{\frac{4\gamma}{3}}, \quad \ln 2 \leq \frac{8\gamma}{27},$$

in stead of (47,48). The conditions on γ rewrite as

$$\gamma \geq \max(\frac{3}{4}\ln(\frac{1}{\varepsilon} + e^{\frac{8}{9}}), 2^{\frac{27}{8}}) \sim \varepsilon^{-\frac{3}{4}}$$

as $\varepsilon \to 0$. This forces a larger choice of b and thereby via (49) a smaller (exponentially small in terms of ε in fact) bound on q_0 for the Nash scheme to converge within the ball of k-radius ε, as was to be expected of course. The fact that the limit u is a solution of $f(u) = 0$ is immediate from (14).

Note that for the standard Newton method the constructed solution of $f(u) = 0$ will have $||u|| < \varepsilon$ if we take equalities in (30) and replace the -1 by $-\frac{1}{\varepsilon}$. The upper bound $\bar{P}$ than has to be replaced by $\bar{P}_\varepsilon = \frac{\varepsilon}{1+\mu\varepsilon}$ and the condition on q_0 becomes $q_0 \leq C\bar{P}_\varepsilon$.

23 The Nash embedding theorem

The Schwartz's lecture notes contain a nice but nonconstructive argument to apply the above together with convexity arguments and the Hahn-Banach Theorem to prove that the n-dimensional torus with any nonstandard Riemannian metric embeds in some $\mathbb{R}^N$. To be explained here. See Chapter 20. Requires a deeper discussion of the smoothing operators used in the proof of Theroem 18.5 and the Fouriertransfrom.

24 Airy scary

I did the calculations below while reading Chapter 8 in Peter Olver's new PDE book with a little help of E.J. Hinch's nice little Cambridge Applied Math textbook on perturbation methods. Just goes to show how beautiful (also applied) complex analysis is.

The Airy function is defined by

$$\mathrm{Ai}(\xi) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i(\xi x + \frac{x^3}{3})}dx,$$

an integral barely convergent. The Airy function plays the same role in the theory for $u_t + u_{xxx} = 0$ as the Gaussian $e^{-\frac{1}{2}x^2}$ for $u_t = u_{xx}$. Both functions define the spatial profile of the fundamental solution.

Replacing $\xi \in \mathbb{R}$ by $\zeta \in \mathbb{C}$ the Airy function is a complex analytic function of

$$\zeta = \xi + i\eta = \rho e^{i\psi}.$$

Replacing also $x \in \mathbb{R}$ by

$$z = x + iy \in \mathbb{C}$$

one may deform the "real" contour C defined by $z = z(t) = t$ with $-\infty < t < \infty$ to another contour γ_ζ that connects two points at infinity. This can be done (without changing the outcome) as long as the integrals of

$$e^{i(\zeta z + \frac{z^3}{3})} = e^{\Phi(z;\zeta)}$$

over the connecting arcs $|z| = R$ between C and the new contour γ_ζ go to zero as $R \to \infty$.

To answer the question

$$\mathrm{Ai}(\rho)e^{i\psi} \sim \quad ? \quad \text{as} \quad \rho \to \infty \quad (\text{for } \psi \text{ fixed}),$$

one chooses the new contour to be one along which the absolute value of the integrand, $e^{\mathrm{Re}\,\Phi(z;\zeta)}$, is peaked and has fast decay as $|z| \to \infty$, and along which $\mathrm{Im}\,\Phi(z;\zeta) = \phi_\zeta$ is a ζ-dependent constant, so that

$$\mathrm{Ai}(\zeta) = \frac{1}{2\pi}\int_{\gamma_\zeta} e^{i(\zeta z + \frac{z^3}{3})}dz = \frac{1}{2\pi}\int_{\gamma_\zeta} \underbrace{e^{\mathrm{Re}\,\Phi(z;\zeta)}}_{\text{real, positive}}\, dz\, e^{i\phi_\zeta},$$

in which the integrand is real, although $dz = dx + idy$ will typically still make the integral complex. The factor $e^{i\phi_\zeta}$ contains the "stationary phase"

ϕ_ζ. If M_ζ is the maximum of $\operatorname{Re}\Phi(z;\zeta)$ along γ_ζ, realised in some $z = m_\zeta$, one may also factor out e^{M_ζ} and write

$$\mathrm{Ai}(\zeta) = \frac{e^{M_\zeta + i\phi_\zeta}}{2\pi} \underbrace{\int_{\gamma_\zeta} e^{-f_0(z;\zeta)} dz}_{\to\, ?\,\text{as}\,|\zeta|\to\infty},$$

in which $f_0(z;\zeta) \geq 0$ along γ_ζ. Typically $f_0(z;\zeta)$ has a unique global minimum zero along γ_ζ and $f_0(z;\zeta) \to +\infty$ as $|z| \to \infty$ along γ_ψ. Note though that the integrand is likely to be ill-behaved as $\rho = |\zeta| \to \infty$, also because the contour γ_ζ may disappear in the limit. The resolution of this latter complication may be prepared by scaling x *before* going to complex variables and making the optimal choice of γ_ζ.

Thus, returning to the definition of $\mathrm{Ai}(\zeta)$ one writes $\mathrm{Ai}(\rho e^{i\psi}) =$

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} \underbrace{e^{i(\rho e^{i\psi}x + \frac{x^3}{3})}}_{\text{scale}\, x = \rho^{\frac{1}{2}}u} dx = \frac{\rho^{\frac{1}{2}}}{2\pi}\int_{-\infty}^{\infty} e^{i\rho^{\frac{3}{2}}(e^{i\psi}u + \frac{u^3}{3})} du = \frac{\rho^{\frac{1}{2}}}{2\pi}\int_{-\infty}^{\infty} e^{\rho^{\frac{3}{2}}\Psi(u)} du,$$

in which you should now view u as $u = \operatorname{Re} w$ with $w = u + iv \in \mathbb{C}$. The effect of this scaling is that the level lines of $\operatorname{Im}\Psi(w)$ are independent of ρ.

One has

$$\Psi(w) = i(e^{i\psi}w + \frac{w^3}{3}) = f(u,v;\psi) + ig(u,v;\psi),$$

with

$$f(u,v;\psi) = -v\cos\psi - u\sin\psi + v(-u^2 + \frac{v^2}{3})$$

and

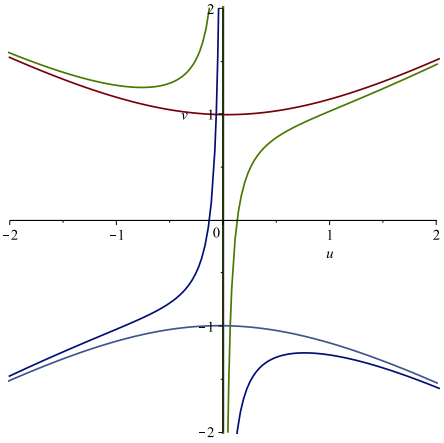
$$g(u,v;\psi) = u\cos\psi - v\sin\psi + u(\frac{u^2}{3} - v^2).$$

These harmonic functions have mutually perpendicular level curves. It is convenient to think of the level curves of the imaginary part $g(u,v;\psi)$ as orbits of

$$\dot{u} = \frac{du}{dt} = f_u = \frac{\partial f}{\partial u} = -\sin\psi - 2uv$$
$$\dot{v} = \frac{dv}{dt} = f_v = \frac{\partial f}{\partial v} = -\cos\psi - u^2 + v^2,$$

a system of ordinary differential equations for $u = u(t)$ and $v = v(t)$. In fact, Cauchy-Riemann gives

$$\frac{df}{dt} = f_u\dot{u} + f_v\dot{v} = f_u^2 + f_v^2 > 0\,, \quad \frac{dg}{dt} = g_u\dot{u} + g_v\dot{v} = g_uf_u + g_vf_v = 0.$$



Figuur 1: orbits for $\psi = \frac{\pi}{24}$

Thus the only orbits of interest as possible contours are the stable manifolds of saddle points, with the maximum of the real part f along the contour occurring in the saddle point.

This is illustrated for small positive ψ by Figure 1 which pictures the possibly relevant level curves of $g(u, v; \frac{\pi}{24})$. The red curve is the stable manifold of

$$m_\psi = (u_\psi, v_\psi) = (-\sin\frac{\psi}{2}, +\cos\frac{\psi}{2})$$

and asymptotes to $3v^2 = u^2$. In particular it has u ranging from $-\infty$ to $+\infty$, and may be written as the graph of a function $v = \varphi(u; \psi)$. The other stable manifold, that of

$$m_{\psi+2\pi} = (u_{\psi+2\pi}, v_{\psi+2\pi}) = (+\sin\frac{\psi}{2}, -\cos\frac{\psi}{2}),$$

the green curve on the right, fails this condition and has a vertical asymptote. The other branch of this level curve is the green curve on the left which is not a stable manifold of either two saddles. Neither of these two orbits is of direct use in relation to the Airy function, but this will change as ψ is taken larger. Note that although $\mathrm{Ai}(\rho e^{i\psi})$ is 2π-periodic in ψ, the parametrisation of the saddle point m_ψ is only 4π-periodic.

Deforming the contour as explained above,

$$\mathrm{Ai}(\rho e^{i\psi}) = \frac{\rho^{\frac{1}{2}}}{2\pi} \underbrace{\int_{-\infty}^{\infty} e^{\rho^{\frac{3}{2}}(f(u,\varphi(u;\psi);\psi)}(1 + i\varphi'(u;\psi))du}_{I(\rho,\psi)} e^{\rho^{\frac{3}{2}}(-\frac{2i}{3}(4\cos^2\frac{\psi}{2}-1)\sin\frac{\psi}{2})},$$

in which the phase factor has been made precise. It remains to examine the integral $I(\rho,\psi)$ in the limit $\rho\to\infty$. Clearly the most important information comes from the (second order) Taylor expansion

$$f = f(u,\varphi(u;\psi);\psi) = M_\psi - a_\psi^2(u-u_\psi)^2 + \cdots$$

with minor contributions coming from the higher order terms and the expansion of $\varphi'(u;\psi)$. Setting

$$u = u_\psi + p$$

one sees that to leading order the asymptotic expansion of the integral must be given by

$$I(\rho,\psi) \sim e^{M_\psi\rho^{\frac{3}{2}}}\int \underbrace{e^{-a_\psi^2\rho^{\frac{3}{2}}p^2}}_{\text{scale } s=a_\psi\rho^{\frac{3}{4}}p}\, dp \sim \frac{e^{M_\psi\rho^{\frac{3}{2}}}}{a_\psi\rho^{\frac{3}{4}}}\underbrace{\int e^{-s^2}ds}_{\sqrt{\pi}} + \cdots,$$

so that

$$\mathrm{Ai}(\rho e^{i\psi}) = \frac{1}{2\rho^{\frac{1}{4}}\sqrt{\pi}}\frac{e^{M_\psi\rho^{\frac{3}{2}}}}{a_\psi}\, e^{\rho^{\frac{3}{2}}(-\frac{2i}{3}(4\cos^2\frac{\psi}{2}-1)\sin\frac{\psi}{2})}(1+O(\rho^{-\frac{3}{2}}))$$

as $\rho\to\infty$. Notice the exponential decay combined with the increasingly rapid oscillations because of the phase factor.

At first sight you might expect an $O(\rho^{-\frac{3}{4}})$ error estimate but since the exponential function in the integrand expands as

$$e^{-s^2+b_3\frac{p^3}{\rho^{\frac{3}{4}}}+b_4\frac{p^4}{\rho^{\frac{6}{4}}}+b_5\frac{p^5}{\rho^{\frac{9}{4}}}+\cdots} = e^{-s^2}(1+(b_3\frac{p^3}{\rho^{\frac{3}{4}}}+\dots)+\frac{1}{2}(b_3\frac{p^3}{\rho^{\frac{3}{4}}}+\dots)^2+\cdots)$$

the higher order terms in the expansion of $\mathrm{Ai}(\rho e^{i\psi})$ involve the integrals

$$\int s^n e^{-s^2}ds \quad (n=3,4,\dots)$$

of which the odd ones vanish. Therefore a contribution of the first b_3-term appears only in combination with the first order term in the expansion of $\varphi'_\psi(u_\psi;\psi)$ (the second order term in the expansion of $\varphi_\psi(u_\psi;\psi)$). It is an exercise to make the expansion more precise.

One has

$$M_\psi = -\frac{2}{3}(4\cos^2\frac{\psi}{2}-3)\cos\frac{\psi}{2},$$

and by direct but tedious calculation the stable manifold is given by

$$v = \varphi_\psi(u) = \varphi_\psi(-s+p) = \frac{-sc + p\sqrt{1 - \frac{4}{3}sp + \frac{1}{3}p^2}}{-s+p} \quad (c = \cos\frac{\psi}{2},\ s = \sin\frac{\psi}{2}).$$

You should recognise a discriminant under the square root, which for the level curve going through the saddle has the property that it is everywhere positive, except in the saddle point m_ψ. Expansion gives

$$\varphi_\psi(-s+p) = c + \frac{-1+c}{s}p + \frac{1}{3}\frac{2c-1}{c+1}p^2 + \dots .$$

For $\psi = 0$ one has

$$v = \varphi_0(u) = 1 + \sqrt{1 + \frac{1}{3}u^2} = 1 + \frac{1}{6}u^2 - \frac{1}{72}u^4 + \cdots$$

and

$$f(u, \varphi_0(u)) = -2(1 + \frac{4}{9}u^2)\sqrt{1 + \frac{1}{3}u^2} = -\frac{2}{3} - u^2 - \frac{5}{36}u^4 + \cdots ,$$

so that

$$a_0 = 1,\ M_0 = -\frac{2}{3},\ A_0 = 0,$$

and

$$\mathrm{Ai}(\xi) \sim \frac{e^{-\frac{2}{3}\xi^{\frac{3}{2}}}}{2\sqrt{\pi}\xi^{\frac{1}{4}}}$$

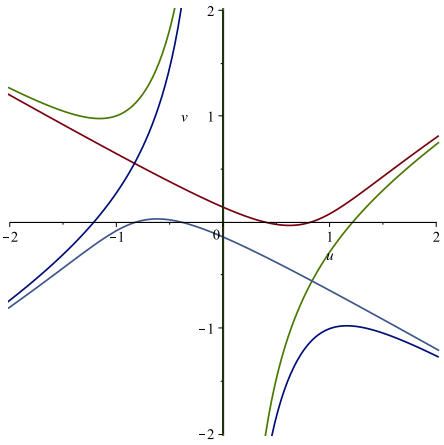
as $\xi \to +\infty$, give or take a mistake in the constants, without oscillations.

Increasing ψ there are changes as ψ crosses $\frac{\pi}{3}$ and $\frac{2\pi}{3}$. For all $0 \le \psi < \frac{2\pi}{3}$ it still holds that

$$\mathrm{Ai}(\rho e^{i\psi}) = \frac{\rho^{\frac{1}{2}}}{2\pi}\int_{-\infty}^{\infty} e^{\rho^{\frac{3}{2}}(f(u,\varphi(u;\psi);\psi)}(1 + i\varphi'(u;\psi))du\ e^{\rho^{\frac{3}{2}}(-\frac{2i}{3}(4\cos^2\frac{\psi}{2}-1)\sin\frac{\psi}{2})}.$$

Figure 2 shows the relevant orbits for $\psi = \frac{15\pi}{24}$, with the same stable manifold defining the contour, and the same asymptotics still valid, but with a different sign for M_ψ, as Figure 3 shows. The sign change occurs at $\frac{\pi}{3}$. Thus for $\frac{\pi}{3} < \psi < \frac{2\pi}{3}$ there is exponential growth of $\mathrm{Ai}(\rho e^{i\psi})$ as $\rho \to \infty$, while the nonzero phase factor accounts for increasingly rapid oscillations.

At $\psi = \frac{2\pi}{3}$, when the growth is maximal (and no oscillations, see Figure 8), the diagram (and the Maple automatic colour coding) changes. All orbits in Figure 4 are in the stable or unstable manifolds of the saddle points.



Figuur 2: orbits for $\psi = \frac{15\pi}{24}$

The appropriate contour now consists of 3 orbits: the 2 orbits in the stable manifold of $m_{\frac{2\pi}{3}}$ (one of which is in the unstable manifold of $m_{\frac{8\pi}{3}}$), and one orbit in the stable manifold of $m_{\frac{8\pi}{3}}$. Can you see which one? You should convince yourself that $M_{\frac{8\pi}{3}}$ only enters the asymptotics beyond any relevant order.

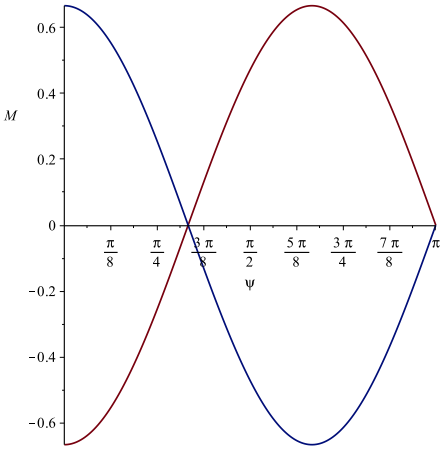
Only as ψ is increased to $\psi = \pi$ both M_π and $M_{3\pi}$ are on par: $M_\pi = M_{3\pi} = 0$. The phases are then $\phi_\pi = \frac{2}{3}$ and $\phi_{3\pi} = -\frac{2}{3}$, and the two stable manifolds are given by

$$v = \frac{(u+1)\sqrt{u(u-2)}}{u\sqrt{3}} \quad (u < 0)\,, \quad v = \frac{(u-1)\sqrt{u(u+2)}}{u\sqrt{3}} \quad (u > 0).$$

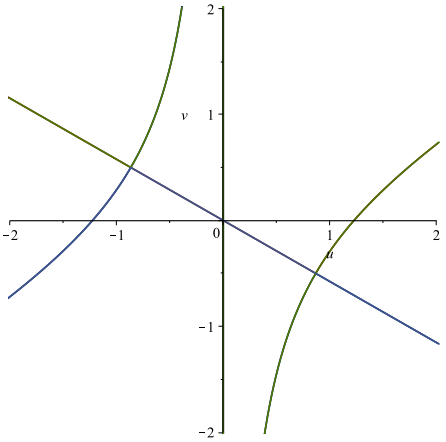
You can now compute the expansion using both contours, with u running from $-\infty$ to 0 for the first integral and from 0 to ∞ for the second. Note the symmetry in Figure 7.

Observe that for $\frac{2\pi}{3} < \psi \leq \pi$ the contours are different. In Figures 5 you see the red curve turning blue after the turning point, and as it escapes to infinity along the negative v-axis it is joined by the green curve which is the stable manifold of the other saddle point. As in the case that $\psi = \pi$, the appropriate contour consists in fact of two contours: the sum of the integrals along both stable manifolds defines $\mathrm{Ai}(\rho e^{i\psi})$. For $\psi < \pi$ the main contribution comes from the contour on the left. Solving a cubic equation this contour can be written as a graph $u = \varphi(v)$, but the main contribution can be computed as above, still writing $v = \varphi(u)$ near the saddle point.

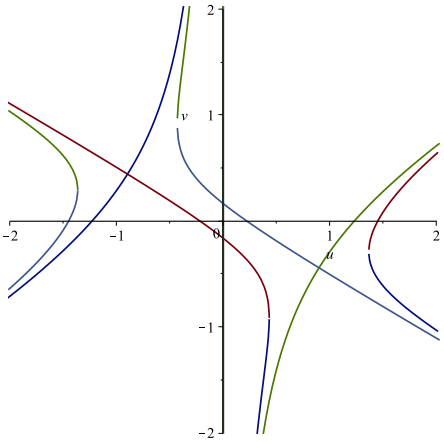
A similar program works for the solution of $u_t + \frac{1}{3}u_{xxx} = 0$ that starts



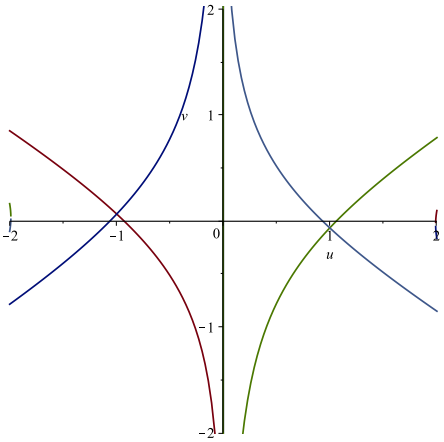
Figuur 3: M_ψ and $M_{\psi+2\pi}$



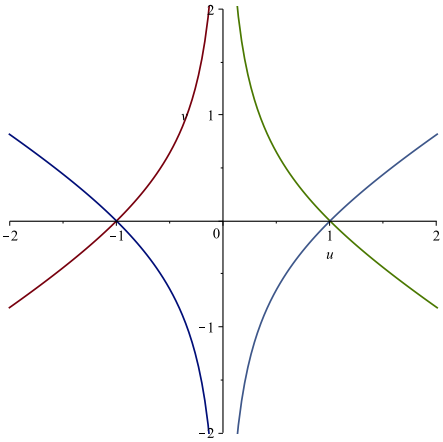
Figuur 4: orbits for $\psi = \frac{16\pi}{24} = \frac{2\pi}{3}$



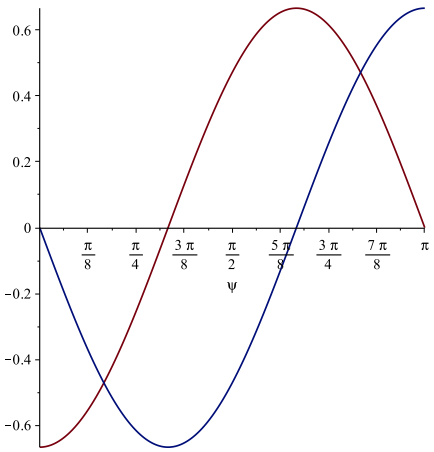
Figuur 5: orbits for $\psi = \frac{17\pi}{24}$



Figuur 6: orbits for $\psi = \frac{23\pi}{24}$



Figuur 7: orbits (stable manifolds: blue curves) for $\psi = \frac{24\pi}{24} = \pi$



Figuur 8: Amplitude M_ψ (red) and phase ϕ_ψ, changes at $\psi = 0, \frac{\pi}{3}, \frac{2\pi}{3}, \pi$.

from a "wave packet"

$$u_0(x) = e^{-\frac{x^2}{4a}} e^{ik_0 x},$$

along lines $x = ct + \xi$. One then has

$$u(t, ct+\xi) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{ik(c+\frac{1}{3}k^2)t + i\xi k - a(k-k_0)^2} dk = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{t\Phi} dk.$$

Replace k by $z = x + iy$ (not the same x of course) and write

$$\Phi = \Phi(z) = \Phi(z;t) = \Phi(z;t,\xi,k_0,a) = f + ig$$

with

$$f = -ty(c + x^2 - \frac{1}{3}y^2) + a(-(x-k_0)^2 + y^2) - \xi y$$

and

$$g = tx(c + \frac{1}{3}x^2 - y^2) - 2a(x-k_0)y + \xi x.$$

Then as before on may rewrite

$$u(t, ct+\xi) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\Phi(x;t)} dx = \frac{1}{\sqrt{2\pi}} \int_{\gamma} e^{\Phi(z;t)} dz$$

in which γ consists of orbits in stable manifolds of a suitable gradient flow of f, which is defined by

$$\dot{x} = \frac{dx}{d\tau} = \frac{1}{t}\frac{\partial f}{dx} = -2xy - \frac{2a}{t}(x - k_0),$$

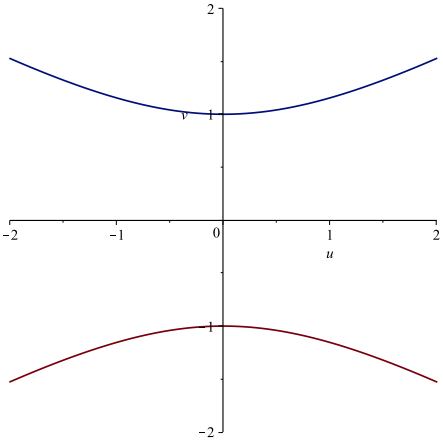
$$\dot{y} = \frac{dy}{d\tau} = \frac{1}{t}\frac{\partial f}{dy} = -c - x^2 + y^2 + \frac{2ay - \xi}{t}.$$

Unlike in the analysis of the Airy function integral, there is now no need to scale x and y, because in the limit $t \to \infty$ the diagram in the x, y-plane is well defined. For $c = 1$ it is the same as in Figure 7 and for $c = -1$ it coincides with Figure 9 (with u, v replaced by x, y). Unlike the u, v-diagram the x, y-diagram varies with the parameter under consideration, as the role of ρ is now played by t. One computes the relevant unstable manifold(s) directly from solving $g = \phi$, which is a quadratic equation in y, asking that the discriminant

$$D = \frac{4}{3}x^2(x^2 + 3c)t^2 + 4x(\xi x - \phi)t + 4a^2(x - k_0)^2$$

of this equation is positive except in the saddle point, thus first determining simultaneously the saddle point and the phase ϕ by solving

$$D = \frac{dD}{dx} = 0.$$



Figuur 9: u, v-plane for $\psi = 0$, the vertical axis also consist of orbits

The phase ϕ then drops out of

$$x\frac{dD}{dx} - D = t^2x^4 + (a^2 + \xi t + ct^2)x^2 - k_0^2a^2 = 0,$$

which determines the square of the positive solution $x = x_c > 0$ uniquely in terms of the parameters t, c, ξ, a, the y-coordinate $y = y_c$. The phase ϕ_c and the value M_c of f in the saddle point (x_c, y_c) are then given by

$$y_c(t) = \frac{a(k_0 - x_c)}{x_c t}, \quad \phi_c(t) = -\frac{2tx_c^3}{3} - \frac{2a^2k_0(x_c - k_0)}{x_c t},$$

and

$$M_c(t) = -a(x_c - k_0)^2 - \frac{a^3(x_c + 2k_0)(x_c - k_0)^2}{3x_c^3t^2}.$$

For the values of y, ϕ, M in the other saddle point replace x_c by $-x_c$. Note that $x_c = x_c(t)$ and likewise for y_c, ϕ_c, M_c (the other dependencies are also suppressed in the notation). Observe the different behaviours as $t \to \infty$ for $c < 0$ and $c > 0$.

At this point I found it convenient to continue the calculations for the stable manifold with x_c implicitly defined by the quartic $x\frac{dD}{dx} - D$ and all other quantities explicitly in terms of x_c. With

$$x = x_c + u, \quad y = y_c + v,$$

the real and imaginary parts of Φ rewrite as

$$f = M_c + F_c, \quad F = F_c(t) = -t(2x_cuv + v(u^2 - \frac{v^2}{3})) - \frac{ak_0}{x_c}(u^2 - v^2),$$

$$g = \phi_c + G_c, \quad G = G_c(t) = t(x_c(u^2 - v^2) + u(\frac{u^2}{3} - v^2)) - \frac{2ak_0uv}{x_c},$$

the latter defining the stable manifold as the graph $v = \varphi_c(u) = \varphi_c(u;t)$ obtained from solving $G = 0$ for v, the discriminant having the desired behaviour: positive except for $u = 0$. For $c > 0$ this gives a globally defined function and deforming contours as before it follows that

$$u(t, ct + \xi) = \frac{e^{M_c(t) + i\phi_c(t)}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{F_c(u,\varphi_c(u;t);t)}(1 + i\varphi_c'(u;t))du.$$

Note that c has disappeared completely from the formula's, except for the dependence through x_c.

The integral depends only on the formula's for F_c and G_c, φ_c being defined by solving $G = 0$, i.e.

$$(x_c + u)tv^2 + \frac{2ak_0u}{x_c}v - tu^2(x_c + \frac{u}{3}) = 0$$

and simplifying the discriminant using the quartic for x_c. This gives

$$v = \varphi_c(u) = \frac{u}{x_c(x_c + u)} \left(-\frac{ak_0}{t} + x_c R \right),$$

in which

$$R = \sqrt{\underbrace{c + 2x_c^2 + \frac{\xi}{t} + \frac{a^2}{t^2}}_{\text{positive}} + \frac{4x_c u}{3} + \frac{u^2}{3}}.$$

The derivative appears in the integral as

$$\varphi_c'(u) = -\frac{ak_0}{t(x_s + u)^2} + \frac{u(2x_c + u)}{3R(x_c + u)},$$

and the exponent in the integral rewrites as

$$F_c(u, \varphi_c(u;t);t) = -\frac{2}{9}\frac{tRu^2\,(3\,x_c + 2\,u)^2}{(x_c + u)^2} + \frac{2}{3}\frac{ak_0u^2\,(3\,x_c + 2\,u)}{(x_c + u)^2}$$

$$-\frac{2}{3}\frac{{k_0}^2a^2\,(Rx_ct - ak_0)\,u^2\,(3\,x_c + u)}{t^2{x_c}^3\,(x_c + u)^3}$$

Clearly these formula's suggest putting

$$u = x_c s, \quad k_0 = \frac{b}{a}, \quad t = b\tau, \quad \xi = b\eta$$

This scaling of u makes each term separate as far the integration variable and t are concerned, except for the R-terms. One now has to distinguish between $c < 0$ (the case discussed by Olver) when $u(t, ct + \xi)$ appears as the sum of 2 integrals involving the stable manifolds of both saddles, and $c > 0$, when $u(t, ct + \xi)$ appears as one single integral.

With x_c going to $\sqrt{-c}$ if $c < 0$, both integrals can be handled as in the Airy case, and the final expansion will depend on c. It may be handy to split the exponential in 3 separate exponentials before you proceed. From the c-dependence there should be a connection with the group velocity discussion by Olver, as we see below. On the other hand, when $c > 0$ (this case is not discussed by Olver) tx_c goes to a constant so $x_c \to 0$. Note that all 3 terms involve s^2, but with different signs. This is really an instructive example for understanding the method!

Now to back to WHY we did this analysis observe that the prefactor in the integral expression for

$$u(t, ct + \xi)$$

contains

$$e^{M_c}$$

which behaves very differently for $c < 0$ and $c > 0$.

For $c > 0$ it is the second term in the expression for $M_c(t)$ above that dominates and goes to infinity because tx_c goes to a constant, and this leading order term then goes to $-\infty$ linearly in t. Modulo the details of the analysis of the integral it follows that $u(t, ct + \xi) \to 0$ exponentially fast as $t \to \infty$.

On the other hand, if $c < 0$ then $x_c \to \sqrt{-c}$ and $M_c(t) \to -a(\sqrt{-c} - k_0)^2$ which is maximal and equal to zero for $c = -k_0^2$. Thus only for this value of c the solution is of order one along the line $x = ct + \xi$ as $t \to \infty$, with the more precise asymptotics following from a more detailed analysis of the integral, as in the Airy functions case, with contributions from both saddle points, and combining both phases and $\frac{2}{3}c^{\frac{3}{2}}t$ appearing in the imaginary part. Olver's point in the section about dispersion relations is that this *group velocity* $-c$ is 3 times larger as you would expect from looking at the single frequency solution with $a = 0$, and he did so by one single calculation starting from the dispersion relation. Read again what he did after the exam, and pay attention to the factor $\frac{1}{3}$ in the third order equation $u_t + \frac{1}{3}u_{xxx} = 0$ that I solved starting from a wave packet centered at $k = k_0$ rather than from a single wave with $k = k_0$.

1. This is an exercise about applying the Fourier transform to solve the equation $u_t + u_{xxx} = 0$ on the real line with initial data $u(0,x) = \delta(x)$, the Dirac δ-function, and to investigate the behaviour of the solution for $x \to -\infty$. The Fourier transform of a function $f = f(x)$ and the inverse transform are defined by

$$\hat{f}(k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{-ikx} dx, \quad f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \hat{f}(k) e^{ikx} dk,$$

respectively whenever f and $\hat{f}$ are sufficiently nice. The improper integrals are to be understood in the principal value sense

$$\int_{-\infty}^{\infty} = \lim_{R \to \infty} \int_{-R}^{R}$$

and are often easiest evaluated using complex integration over appropriate contours.

   (a) Explain why $\hat{\delta}(k) = \frac{1}{\sqrt{2\pi}}$.

   (b) Show using integration by parts that $\widehat{(f')}(k) = ik\hat{f}(k)$.

   (c) Let u be a smooth solution of $u_t + u_{xxx} = 0$ which decays to zero sufficiently fast as $|x| \to \infty$ to have $(\hat{u})_t = \widehat{(u_t)}$. Here $\hat{u} = \hat{u}(t,k)$ denotes the Fourier transform of the function $x \to u(x,t)$. Denote the initial value of u by u_0, that is, $u_0(x) = u(0,x)$. Show that

$$\hat{u}(t,k) = \widehat{u}_0(k) e^{ik^3 t}$$

   (d) Show that the inversion formula formally applied to the case that $\widehat{u}_0(k) = \hat{\delta}(k) = \frac{1}{\sqrt{2\pi}}$ defines a solution formula

$$u(t,x) = \frac{1}{(3t)^{\frac{1}{3}}} \mathrm{Ai}\left(\frac{x}{(3t)^{\frac{1}{3}}}\right)$$

   in which

$$\mathrm{Ai}(\xi) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i(\xi x + \frac{x^3}{3})} dx.$$

   (e) Use the methods above to determine the asymptotic behaviour of $\mathrm{Ai}(\xi)$ for $\xi \to -\infty$.

25 Hartman-Grobman stelling

Some material prepared for this very enjoyable event (see also Section 4.4):

`www.universiteitleiden.nl/agenda/2017/04/nationaal-wiskunde-symposium`

In [HM] hebben we uitgebreid gekeken naar de Methode van Newton voor het oplossen van vergelijkingen, met als eerste voorbeeld het snel benaderen van algebraïsche getallen, bijvoorbeeld $\sqrt{2}$, dat een vast punt is van de afbeelding

$$x \xrightarrow{F} F(x) = \frac{1}{2}(x + \frac{2}{x}),$$

een afbeelding die ongeveer 3000 jaar oud is, en later herontdekt is via

$$f(x) = x^2 - 2 \quad \text{en} \quad F(x) = x - f'(x)^{-1} f(x) = x - \frac{f(x)}{f'(x)}.$$

De mooie eigenschappen van het discrete dynamisch systeem gedefinieerd door

$$x_n = F(x_{n-1}) \quad (n \in \mathbb{N})$$

worden deels verklaard door het feit

$$F'(x) = \frac{f(x)f''(x)}{f'(x)^2}$$

gelijk is aan 0 in nulpunten van $f(x)$ en

$$F(x) = x \iff f(x) = 0$$

voor elke x met $f'(x) \neq 0$. Een curieus voorbeeldje in Hoofdstuk 6 van The Beauty of Fractals van Peitgen en Richter is

$$f(x) = \frac{x}{1-x} \quad \text{en} \quad F(x) = x^2,$$

en dat is er eentje uit een curieuze familie, bijvoorbeeld

$$f(x) = \frac{x}{(1-x)^{\frac{1}{7}}(1 + x + x^2 + x^3 + x^4 + x^5 + x^6 + x^7)^{\frac{1}{7}}} \quad \text{en} \quad F(x) = x^8,$$

maar dat terzijde. De afbeeldingen

$$x \to x^2 \quad \text{en} \quad x \to \frac{1}{2}(x + \frac{2}{x})$$

hebben vaste punten waar hun afgeleide 0 is en het is niet moeilijk jezelf ervan te overtuigen dat dit leidt tot snelle convergentie de rij x_n naar evenwicht, als je begint met x_0 in de buurt van een evenwicht.

Neem je zomaar een functie F om een dynamisch systeem te maken zoals hierboven, en is $F(0) = 0$, dan bepaalt de afgeleide $F'(0)$ in het algemeen of $x = 0$ een (lokaal) stabiel of onstabiel evenwicht is, zoals het voorbeeld

$$x \to \lambda x$$

met $\lambda \in \mathbb{R}$ bij inspectie meteen laat zien. Een voor de hand liggende vraag is dan of de dynamische systemen gedefinieerd door

$$x \to F(x) \quad \text{en} \quad \tilde{x} \to F'(0)\tilde{x}$$

niet eigenlijk hetzelfde zijn via een conjugatie:

$$\begin{array}{ccc} x & \stackrel{\phi}{\to} & \tilde{x} \\ F \downarrow & & \downarrow F'(0) \\ F(x) & \stackrel{\phi}{\to} & F'(0)\tilde{x} \end{array}$$

Dus is er een inverteerbare afbeelding ϕ waarmee

$$F'(0)\phi(x) = \phi(F(x))$$

voor x in een zo groot mogelijke buurt van $x = 0$? Als we $F(x)$ schrijven als

$$F(x) = \lambda x + a(x)$$

dan is de vraag dus of we gegeven $\lambda \in \mathbb{R}$ en $a : \mathbb{R} \to \mathbb{R}$ met $a'(0) = 0$ de functie $\phi : \mathbb{R} \to \mathbb{R}$ kunnen vinden zodanig dat

$$\lambda\phi(x) = \phi(\lambda x + a(x))$$

in de buurt van $x = 0$, en dit kunnen we proberen op te lossen door middel van

$$\phi_n(x) = \frac{\phi_{n-1}(\lambda x + a(x))}{\lambda} \quad \text{beginnend met} \quad \phi_0(x) = x.$$

Als we een a nemen met $a'(0) = 0$ en $a(x) = 0$ voor $|x|$ buiten een interval gedefinieerd door $|x| \leq \eta$ met η wellicht nog te kiezen, dan zien we dat de onbekende functie ϕ voor $|x| \geq \eta$ wel gegeven moet worden door $\phi(x) = x$. Hoewel? Is het duidelijk dat gegeven $\lambda \in \mathbb{R}$ uit

$$\lambda\phi(x) = \phi(\lambda x)$$

voor alle $x \in \mathbb{R}$ volgt dat $\phi(x) = x$ voor alle $x \in \mathbb{R}$? Niet meteen dus. Maar $\phi(x) = x$ doet het wel.

Terzijde, als ϕ differentieerbaar is volgt (als $\lambda \neq 0$) dat

$$\phi'(x) = \phi'(\lambda x)$$

voor alle $x \in \mathbb{R}$, en daarmee zijn heel veel ϕ' waarden gelijk aan elkaar, tenzij $|\lambda| = 1$. Als ϕ' continu is in 0 moet wel te bewijzen zijn dat $\phi'(x) = \phi'(0)$ voor alle x. En $\phi'(0) = 1$ ligt voor de hand als normaliserende voorwaarde.

Of we voor voor $a(x) \not\equiv 0$ zo'n differentieerbare ϕ wel maken is echter zeer de vraag. In het iteratieproces helpt de λ in de noemer wellicht als $|\lambda| > 1$ is. Weer terzijde, het voorbeeld met $a(x) = x^2$ laat zien dat zonder de aanname dat $a(x) \equiv 0$ voor $|x|$ groot er weinig hoop is, want we krijgen

$$\phi_1(x) = x + \frac{x^2}{\lambda},$$

$$\phi_2(x) = x + \left(\frac{1}{\lambda} + 1\right) x^2 + \frac{2x^3}{\lambda} + \frac{x^4}{\lambda^2},$$

$$\phi_3(x) = x + \left(\frac{1}{\lambda} + 1 + \lambda\right) x^2 + \left(\frac{2}{\lambda} + 2 + 2\,\lambda\right) x^3 + \left(\frac{1}{\lambda^2} + \frac{1}{\lambda} + 6 + \lambda\right) x^4$$
$$+ \left(\frac{6}{\lambda} + 4\right) x^5 + \left(\frac{2}{\lambda^2} + \frac{6}{\lambda}\right) x^6 + \frac{4x^7}{\lambda^2} + \frac{x^8}{\lambda^3},$$

$$\phi_4(x) = x + \left(\frac{1}{\lambda} + 1 + \lambda + \lambda^2\right) x^2 + \left(\frac{2}{\lambda} + 2 + 4\,\lambda + 2\,\lambda^2 + 2\,\lambda^3\right) x^3$$
$$+ \left(\frac{1}{\lambda^2} + \frac{1}{\lambda} + 7 + 7\,\lambda + 6\,\lambda^3 + \lambda^4 + 7\,\lambda^2\right) x^4 + \cdots + \frac{x^{16}}{\lambda^4},$$

enzovoorts.

De Stelling van Hartman Grobman gaat in het simpelste geval om de vraag of voor de afbeelding

$$(x, y) \xrightarrow{F} (\xi, \eta) = (\lambda x + a(x, y), \mu y + b(x, y))$$

het stelsel

$$\lambda\phi(x, y) = \phi(\lambda x + a(x, y), \mu y + b(x, y))$$
$$\mu\psi(x, y) = \psi(\lambda x + a(x, y), \mu y + b(x, y))$$

kunnen oplossen naar de functies ϕ, ψ onder de aanname dat

$$0 < |\mu| < 1 < |\lambda|,$$

teneinde de afbeelding F te conjugeren met de afbeelding

$$(\tilde{x}, \tilde{y}) \to (\tilde{\xi}, \tilde{\eta}) = (\lambda\tilde{x}, \mu\tilde{y}).$$

Dit zijn twee vergelijkingen, in de eerste is de onbekende de functie ϕ, in de tweede de functie ψ. Die voor ϕ lijkt op de vergelijking waarmee we begonnen en waarvoor de geschetste aanpak kans van slagen heeft als $|\lambda| > 1$. De aannamen op $a(x, y)$ en $b(x, y)$ zijn nu zoals die op $a(x)$ hierboven, dus

$$a_x(0, 0) = a_y(0, 0) = b_x(0, 0) = b_y(0, 0) = 0,$$

en $a(x, y)$ en $b(x, y)$ tenminste continu differentieerbaar. Met die conditie is het stelsel

$$\xi = \lambda x + a(x, y)$$

$$\eta = \mu y + b(x, y)$$

in de buurt van $(0, 0)$ op te lossen naar x, y in de vorm

$$x = \frac{1}{\lambda}\,\xi + \alpha(\xi, \eta)$$

$$y = \frac{1}{\mu}\,\eta + \beta(\xi, \eta)$$

met $\alpha(\xi, \eta)$ en $\beta(\xi, \eta)$ continu differentieerbaar in de buurt van $(0, 0)$ en

$$\alpha_\xi(0, 0) = \alpha_\eta(0, 0) = \beta_\xi(0, 0) = \beta_\eta(0, 0) = 0.$$

De eerste vergelijking houden we zoals die was, de tweede schrijven we in ξ, η. Beide vergelijkingen hebben dan dezelfde vorm:

$$\phi(x, y) = \frac{1}{\lambda}\phi(\lambda x + a(x, y), \mu y + b(x, y))$$

$$\psi(\xi, \eta) = \mu\psi(\frac{1}{\lambda}\xi + \alpha(\xi, \eta), \frac{1}{\mu}\eta + \beta(\xi, \eta))$$

Om deze vergelijkingen op te lossen moeten we dus eerst weten hoe we de eerdere vergelijking voor $\phi : \mathbb{R} \to \mathbb{R}$ oplossen.

Als $|a(x)| \leq \varepsilon|x|$ dan volgt

$$|\phi_1(x) - \phi_0(x)| = |\frac{1}{\lambda}(\lambda x + a(x)) - x| = |\frac{a(x)}{\lambda}| \leq \frac{\varepsilon}{\lambda}|x|,$$

en dan

$$|\phi_2(x) - \phi_1(x)| = |\frac{1}{\lambda}\phi_1(\lambda x + a(x)) - \frac{1}{\lambda}\phi_0(\lambda x + a(x))|$$

$$\leq \frac{\varepsilon}{|\lambda|^2}|\lambda x + a(x)| \leq \frac{\varepsilon(|\lambda| + \varepsilon)}{|\lambda|^2}|x|,$$

waarna

$$|\phi_3(x) - \phi_2(x)| = |\frac{1}{\lambda}\phi_2(\lambda x + a(x)) - \frac{1}{\lambda}\phi_1(\lambda x + a(x))|$$

$$\leq \frac{\varepsilon(|\lambda| + \varepsilon)}{|\lambda|^3}|\lambda x + a(x)| \leq \frac{\varepsilon(|\lambda| + \varepsilon)^2}{|\lambda|^3}|x|.$$

Zo wordt duidelijk dat

$$|\phi_n(x) - \phi_{n-1}(x)| \leq \frac{\varepsilon(\lambda + \varepsilon)^{n-1}}{|\lambda|^n}|x|,$$

niet genoeg om de rij $\phi_n(x)$ convergent te krijgen, maar als ook geldt dat $a(x) = 0$ voor $|x| \geq \eta$ dan kunnen we met $0 < \delta < 1$ de schattingen aanpassen als

$$|\phi_1(x) - \phi_0(x)| \leq \frac{\varepsilon}{|\lambda|}\eta^{1-\delta}|x|^\delta,$$

$$|\phi_2(x) - \phi_1(x)| \leq \frac{\varepsilon}{|\lambda|^2}\eta^{1-\delta}|\lambda x + a(x)|^\delta \leq \frac{\varepsilon}{|\lambda|^2}\eta^{1-\delta}(|\lambda| + \varepsilon)|x|^\delta,$$

en dan wordt duidelijk dat

$$|\phi_n(x) - \phi_{n-1}(x)| \leq \varepsilon\eta^{1-\delta}\frac{(|\lambda| + \varepsilon)^{\delta(n-1)}}{|\lambda|^n}|x|^\delta.$$

Uniforme convergentie volgt als

$$(|\lambda| + \varepsilon)^\delta < |\lambda|.$$

26 Welke fundamenten?

Deze oude inleiding was bedoeld voor een breed publiek. De eerstejaars wiskunde student kan voor de lol lezen wat ik hier schrijf. Ik begin met de verzameling $\mathbb{R}$ van de *reële getallen* en aftelbare sommen van die getallen. Als het onderstaande goed leesbaar is dan kun je rustig op weg met wat er verder komt in dit boek. Zo niet, dan zou het *groene* boekje met Ronald Meester[1] je wat op weg kunnen helpen. In dat boekje, dat vanaf nu [HM] heet, kwamen we vanuit getallenrepresentaties als

$$\frac{1}{3} = \frac{3}{10} + \frac{3}{100} + \frac{3}{1000} + \frac{3}{10000} + \frac{3}{100000} + \cdots = \sum_{n=1}^{\infty} \frac{3}{10^n} = 3\sum_{n=1}^{\infty} \frac{1}{10^n}$$

op natuurlijke wijze tot het inzicht dat ieder (reëel) getal van de vorm

$$k + \sum_{n=1}^{\infty} \frac{d_n}{10^n} \tag{1}$$

is. In (1) is $k \in \mathbb{Z}$, de verzameling van de *gehele getallen*. De decimalen zijn

$$d_n \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

waarbij n de verzameling $\mathbb{N}$ van de positieve[2] gehele getallen doorloopt, ook wel de *natuurlijke getallen* genoemd.

De verschillende notaties hierboven voor het rationale getal dan wel de breuk $\frac{1}{3}$ kunnen tot enige controverse leiden. De breuk $\frac{1}{3}$ heeft immers net als iedere andere breuk een teller en een noemer, in dit geval teller 1 en noemer 3. Evenzo heeft de breuk $\frac{2}{6}$ teller 2 en noemer 6. De breuken $\frac{1}{3}$ en $\frac{2}{6}$ zijn echter als rationale getallen gelijk aan elkaar. Mag je nu van het rationale getal $\frac{1}{3}$ zeggen dat zijn teller 1 en zijn noemer 3 is? Van mij wel, maar daar wordt soms anders over gedacht. Dus daarom hierbij de afspraak dat we stilzwijgend het rationale getal altijd als breuk met een minimale noemer[3] in $\mathbb{N}$ schrijven als we het over teller en noemer van het rationale getal hebben.

Ook de naam "reeks" voor de uitdrukking met het somteken Σ leidt tot controverses, alsmede het gebruik van het symbool ∞ boven op dat somteken. Wat het eerste betreft zou ik liever zoveel mogelijk over aftelbare sommen willen spreken, maar niet te vergeten dat de term "reeks" nu eenmaal door iedereen gebruikt wordt in zinsdelen als "de som van de reeks".

[1] vuuniversitypress.com/15-voor-auteurs/overige-content/108-wiskunde-in-je-vingers

[2] NB, 0 is niet positief, $\mathbb{N} = \{n \in \mathbb{Z} : n > 0\}$, $\mathbb{R}^+ = \{x \in \mathbb{R} : x > 0\}$.

[3] Ontbind teller en noemer in priemfactoren en streep gemeenschappelijke factoren weg.

Het gebruik van het symbool ∞ is wellicht te vermijden door

$$\sum_{n\in\mathbb{N}} \quad \text{in plaats van} \quad \sum_{n=1}^{\infty}$$

te schrijven, maar dan is de volgorde waarin de termen in de som bij elkaar op worden geteld niet meer zo eenduidig specificeerd als in de meer gebruikelijke notatie. Die wordt namelijk doorgaans uitgesproken als *de som van de termen in de reeks, waarbij n loopt vanaf het getal* 1 *tot (en niet tot en met) oneindig*[4]. In het derde hoofdstuk komen we hier nog op terug.

Tenslotte merken we op dat de schrijfwijze in (1) niet altijd uniek is omdat getallen van de vorm

$$k + \sum_{n=1}^{m} \frac{d_n}{10^n} \tag{2}$$

nu eenmaal twee representaties hebben, bijvoorbeeld

$$1 = \frac{9}{10} + \frac{9}{100} + \frac{9}{1000} + \frac{9}{10000} + \frac{9}{100000} + \cdots = \sum_{n=1}^{\infty} \frac{9}{10^n}, \tag{3}$$

wellicht het eerste voorbeeld van een zogenaamde meetkundige reeks dat ieder kind in het basisonderwijs hopelijk wel eens te zien krijgt.

Het simpelste voorbeeld van zo'n meetkundige reeks betreft de rij breuken

$$\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{32}, \frac{1}{64} \ldots,$$

met in de noemers de getallen uit de eerste rij getallen die ik ooit van mijn vader leerde, toen ik een jaar of 2^2 was. Als we die rij beschrijven met

$$a_n = \frac{1}{2^n}$$

met n de verzameling $\mathbb{N}$ doorlopend, dan is de bijbehorende som gelijk aan het getal 1. Nog altijd de mooiste som die er bestaat. Een eindeloze rij getallen die optellen tot 1. Wat wil je nog meer?

26.1 Academisch speelkwartier: kolomcijferen

We kiezen nu voor een wat basaler perspectief dan gebruikelijk om meer inzicht te krijgen in wat de reële getallen zijn. Deze inventariserende[5] subsectie

[4] Rekenen met ∞ doen wij hier niet.

[5] We gaan niet recht op een doel af nu.

kan overgeslagen worden bij eerste lezing[6], maar een opmerking van Jan Wiegerinck zette me aan het denken, ook in relatie tot Opgave 1.7 in [HM]. Hoe rekenen we met (eerste maar positieve) getallen in de vorm (1)? Een rekenvraag die we nu rekenend en redenerend willen beantwoorden zonder over verzamelingen van getallen te spreken.

Terzijde, helemaal consequent is de schrijfwijze in (1) niet. De k is duidelijk anders dan de rest van de termen in deze aftelbare som[7]. Als we ons beperken tot de positieve reële getallen, die we als verzameling[8] gezien aanduiden met $\mathbb{R}^+$, dan is het eleganter om het "Romeinse" perspectief van eenheden, tientallen, tienden, honderdtallen, hondersten, duizendtallen, duizendsten, etc alleen te combineren met de "Arabische" cijfers

$$0, 1, 2, 3, 4, 5, 6, 7, 8, 9.$$

Met

$$\text{tien} = 10, \quad \text{honderd} = 10 \times 10 = 100, \quad \text{duizend} = 10 \times 10 \times 10 = 1000,$$

enzovoorts, en op zijn kop

$$\text{een tiende} = \frac{1}{10}, \quad \text{een honderste} = \frac{1}{100}, \quad \text{een duizendste} = \frac{1}{1000},$$

und so weiter, is het niet heel raar om bijvoorbeeld

$$8 \times 1000 + 7 \times 100 + 6 \times 10 + 5 \times 1 + 4 \times \frac{1}{10} + 3 \times \frac{1}{100} + 2 \times \frac{1}{1000}$$

als corresponderend met een punt op een lijn van hier tot ginder te zien. Bij zo'n punt hoort een getal dat we noteren als

$$8765{,}4321 \tag{4}$$

in onze decimale notatie van vandaag de dag, met ook (hier 5 keer) de gewone eenheid 1, en een Nederlandse komma waarachter in dit geval vier cijfers staan.

Over getallen als (4) hoeven verder geen misverstanden te bestaan. Links van de komma tellen de decimalen van rechts af de 1-tallen, 10-tallen, 10×10-tallen, $10 \times 10 \times 10$-tallen, en rechts van de komma vanaf links de $\frac{1}{10}$-tallen, $\frac{1}{10}$-tallen, $\frac{1}{10} \times \frac{1}{10}$-tallen, $\frac{1}{10} \times \frac{1}{10} \times \frac{1}{10}$-tallen, $\frac{1}{10} \times \frac{1}{10} \times \frac{1}{10} \times \frac{1}{10}$-tallen. Aan beide kanten van de komma breekt het af, maar rechts is dat niet nodig. In

[6] En ook bij tweede lezing.

[7] Som van een eindeloze rij, zoals Marjolein Kool dat zo mooi noemt.

[8] Getallen in verzamelingen willen stoppen is een beetje een beroepsafwijking.

principe kunnen we er nog een rij cijfers achter plaatsen, en een groter getal maken, bijvoorbeeld

$$8765{,}432143214321 \quad \text{of} \quad 8765{,}432199999999999999999999999,$$

getallen die allebei groter zijn dan 8765,4321 en kleiner dan 8765,43211.

Omdat het in negen gelijke stukken verdelen van het lijnstuk tussen het beginpunt van diezelfde lijn van hier tot ginder, en het punt waar

$$0 \times 1000 + 0 \times 100 + 0 \times 10 + 1 \times 1 + 0 \times \frac{1}{10} + 0 \times \frac{1}{100} + 0 \times \frac{1}{1000} = 1$$

staat, negen lijnstukken geeft waarvan het eerste nog net niet loopt tot

$$0 \times 1000 + 0 \times 100 + 0 \times 10 + 1 \times 1 + 1 \times \frac{1}{10} + 1 \times \frac{1}{100} + 1 \times \frac{1}{1000} = 0{,}1111,$$

zien we dat er geen reden is waarom elk punt op de lijn een afbrekende getalrepresentatie zou moeten hebben. Wat heet, één negende correpondeert omherroepelijk met een representatie als in (4) waarbij er voor de komma alleen maar 0-en staan, en achter de komma alleen maar 1-en, zonder dat het rechts afbreekt[9].

Dat

$$9 \times 0{,}111111111\ldots = 9 \times 0{,}\underline{1} \quad \text{gelijk is aan} \quad 1,$$

is een conclusie die we willen trekken als resultaat van de herhaalde optelling

$$0{,}\underline{1} + 0{,}\underline{1} + 0{,}\underline{1} + 0{,}\underline{1} + 0{,}\underline{1} + 0{,}\underline{1} + 0{,}\underline{1} + 0{,}\underline{1} + 0{,}\underline{1} = 0{,}\underline{9} = 1.$$

Op dat optellen komen we zo terug, en op $0{,}\underline{9} = 1$ indirect ook. Bij een (met de nog te speciferen regels) decimaal geschreven getal $0{,}\underline{9}$ optellen verhoogt het gehele getal voor de komma met 1.

Verdelen we hetzelfde lijnstuk niet in negen maar in 99 gelijke stukken, dan zien we

$$0{,}010101010101010101010101\ldots \tag{5}$$

als getalrepresentatie voor één negenennegentigste verschijnen. De puntjes geven hier aan dat de decimale ontwikkeling niet eindigt. Tegenwoordig schrijven we

$$\frac{1}{9} = 0{,}\underline{1}, \quad \frac{1}{99} = 0{,}\underline{01}, \quad \frac{1}{999} = 0{,}\underline{001},$$

met links steeds een rationaal getal en rechts de decimale representatie van dat getal, dat we met liefde ook een breuk mogen noemen, een breuk met teller 1 en en een noemer met alleen maar 9-ens.

[9] En links eigenlijk ook niet, al schrijven we die nullen nooit op.

We zien in (5) dat de 0 als cijfer erg handig is, de 0 die correspondeert met nul vingers op de twee gebalde vuisten van je handen waar je geen ruzie mee wil krijgen. Het tellen zelf begint met 1, eindigt op de vingers bij tien = 10, en gaat daarna verder met 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376, 377, 378, 379, 380, 381, 382, 383, 384, 385, 386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399, en een kind kan al lerend voor het slapen gaan zien en begrijpen hoe dat zo altijd maar doorgaat als dromenland geen redding brengt. De eindeloze rij van de natuurlijke getallen, beginnend met 1, wordt zo ondubbelzinnig vastgelegd door de aftelling in het decimale stelsel.

Getallen uit die rij kunnen we optellen en vermenigvuldigen (in wezen herhaald optellen). Dat doen we cijferend met de getallen onder elkaar gezet, onhandig vanaf links of handig vanaf rechts per kolom. Die methodes[10] werken ook voor het rekenen met de positieve kommagetallen die we krijgen door voor de komma van het kommagetal het cijfer 0 of een natuurlijk getal te zetten, en achter de komma een natuurlijk getal opgevat als een rij cijfers, met voor dat getal al of niet nog een aantal nullen.

De natuurlijke getallen zelf zijn geen kommagetallen. Door getallen als 123456789 gelijk te zien aan een nepkommagetal 123456789,0 is dat snel ver-

[10] In het PO heeft de onhandige methode veelal de voorkeur gekregen.

holpen, maar dit wordt in de natuurkunde[11] terecht als een minder gelukkige en te mijden conventie gezien. Afbrekende kommagetallen kunnen wellicht beter vermeden worden, en de niet afbrekende kommagetallen zijn nu net de kommagetallen die we nog missen, en waar we zeker ook mee willen rekenen en cijferen. Dat *cijferen* moet dan wel vanaf links als je er even over nadenkt.

Tellen we bijvoorbeeld het getal dat gerepresenteerd wordt door (5) achtennegentig keer bij zichzelf op dan zien we dat 99 keer $\frac{1}{99}$ niet alleen gelijk is aan 1 maar ook aan

$$0{,}\underline{9} = 0{,}999999999999999999999999\ldots,$$

en deze decimale representatie kan de meestal toch onbereikbare eenheid 1 best vervangen, als we afspreken dat decimale representaties van positieve reële getallen naar links altijd, maar naar rechts nooit doorlopen met alleen maar 0-en. Een onhandige conventie wellicht, maar omdat het *cijferen* hier toch alleen maar onhandig vanaf links kan niet eens zo gek eigenlijk.

Bovendien is elk natuurlijk getal nu op een natuurlijke manier ook een echt kommagetal. Met $0{,}\underline{9}$, $1{,}\underline{9}$, $2{,}\underline{9}$, $3{,}\underline{9}$, $4{,}\underline{9}$, $5{,}\underline{9}$, $6{,}\underline{9}$, $7{,}\underline{9}$, $8{,}\underline{9}$, $9{,}\underline{9}$, $10{,}\underline{9}$, $11{,}\underline{9}$, $12{,}\underline{9}$, $13{,}\underline{9}$, $14{,}\underline{9}$, $15{,}\underline{9}$, $16{,}\underline{9}$, $17{,}\underline{9}$, $18{,}\underline{9}$, $19{,}\underline{9}$, $20{,}\underline{9}$, $21{,}\underline{9}$, $22{,}\underline{9}$, $23{,}\underline{9}$, $24{,}\underline{9}$, $25{,}\underline{9}$, $26{,}\underline{9}$, $27{,}\underline{9}$, $28{,}\underline{9}$, $29{,}\underline{9}$, $30{,}\underline{9}$, $31{,}\underline{9}$, $32{,}\underline{9}$, $33{,}\underline{9}$, $34{,}\underline{9}$, $35{,}\underline{9}$, $36{,}\underline{9}$, $37{,}\underline{9}$, $38{,}\underline{9}$, $39{,}\underline{9}$, $40{,}\underline{9}$, $41{,}\underline{9}$, $42{,}\underline{9}$, komen we ook op weg in de rekenkunde. Ook al bekt het niet zo lekker, we gaan rekenen met deze niet afbrekende kommagetallen zoals ze ons gegeven worden.

[11] Waar de laatste decimaal meestal met meetnauwkeurigheid te maken heeft.

26.1.1 Optellen

De vraag is nu of we met alle ons gegeven kommagetallen kunnen rekenen zoals je zou verwachten, en of rekenen dan (zoals tegenwoordig in het basisonderwijs) onhandig cijferen[12] kan worden, zoals bijvoorbeeld in

$$
\begin{array}{r}
0,9999999 \\
0,9999999 \\
\hline + \\
1,8000000 \\
0,1800000 \\
0,0180000 \\
0,0018000 \\
0,0001800 \\
0,0000180 \\
0,0000018 \\
\hline + \\
1,9999998
\end{array}
$$

Immers, met doorlopende negens gaat dit onhandig *cijferen* precies hetzelfde en duurt nauwelijks langer dan hierboven:

$$
\begin{array}{r}
0,99999\ldots \\
0,99999\ldots \\
\hline + \\
1,80000\ldots \\
0,18000\ldots \\
0,01800\ldots \\
0,00180\ldots \\
0,000180.. \\
0,000018.. \\
..... \\
\hline + \\
1,99999\ldots
\end{array}
\qquad (6)
$$

Gelukkig: $1 + 1 = 2!$[13] Weliswaar schendt de realistisch tussenstap hier wel de regel dat we rechts geen doorlopende 0-en mogen hebben, de uitkomst

[12] Onhandig cijferen wordt ook wel kolomrekenen genoemd.

[13] Lees: één en één is twee uitroepteken.

van de som is duidelijk: de 8 combineert steeds met de 1 op de volgende rij tot een 9. De 18 op elke rij is de som van 9 en 9. Op de eerste rij betreft het 0,9 + 0,9, op de tweede rij 0,09 + 0,09, op de derde rij 0,009 + 0,009, enzovoorts. Het is instructief[14] om zo'n sommetje als hierboven met twee andere doorlopende getallen met voor de komma alleen maar 0-en te doen. Dan vormen de twee cijfers op elke rij steeds een getal uit de rij

$$0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18,$$

en bij elk van deze getallen kan zowel van boven een getal uit de rij

$$0 = 00, 10, 20, 30, 40, 50, 60, 70, 80, 90$$

en daarna vanonder ook een getal uit de veel kortere rij $0, 1$ worden opgeteld. Op zijn hoogst krijgen we dus $90 + 18 + 1 = 109$.

Is de som groter dan 99 dan schuift er een 1 door naar links maar dat overhevelen blijft beperkt. Optellend per tweetal kolommen kan er een 1 naar links doorschuiven en een 1 van rechts binnenkomen. Die 1 kan van de 109 een 110 maken, maar ook die geeft nog steeds op zijn hoogst een 1 naar links door. Dat optellen van twee getallen onhandig cijferend per twee kolommen tegelijk vanaf links gaat dus altijd wel lukken.

Hoe zit het met drie getallen? We nemen weer de moeilijkste som van dat type, met de cijfers zo groot mogelijk, dus

$$\begin{array}{l} 0,99999\ldots \\ 0,99999\ldots \\ 0,99999\ldots \\ \hline\!\!\!

maar 0-en te doen. Dan vormen de twee cijfers op elke rij steeds een getal uit de rij

$$0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27,$$

en bij elk van deze getallen kan zowel een getal uit de rij

$$0 = 00, 10, 20, 30, 40, 50, 60, 70, 80, 90$$

en daarna ook een getal uit de veel kortere rij

$$0, 1, 2$$

worden opgeteld. Op zijn hoogst krijgen we nu $90 + 27 + 2 = 119$. Het overhevelen naar links blijft weer beperkt. Met enig werk gaan we hier wel inzien dat drie zulke *nul komma nog minstens wat getallen* altijd cijferend bij elkaar opgeteld kunnen worden en dat het niet uitmaakt[15] of we er eerst twee samen nemen en in welke volgorde we de getallen optellen, en dat in de som voor de komma ook een 1 of een 2 kan komen te staan.

Willen we positieve getallen met ook voor de komma decimalen hebben staan in de getallen die we bij elkaar optellen, dan doen we die apart. Bijvoorbeeld

$$99,\underline{9} + 88,\underline{8} + 77,\underline{7} = 99 + 88 + 77 + 0,\underline{9} + 0,\underline{8} + 0,\underline{7},$$

waarbij

$$\begin{array}{r} 99 \\ 88 \\ 77 \\ \text{—} + \\ \hline 264 \end{array}$$

nu ook (juist wel handig) van rechts af per kolom uitgecijferd[16] kan worden, en de som van de drie *nul komma nog minstens wat getallen* met de methode hierboven gelijk is aan $2,\underline{6}$. Alles bij elkaar vinden we zo dat

$$99,\underline{9} + 88,\underline{8} + 77,\underline{7} = 264 + 2,\underline{6} = 266,\underline{6},$$

al had dat vast handiger gekund.

[15] Lees: $a + b + c = (a + b) + c = c + (a + b)$ met alle gepermuteerde variaties.

[16] Ook kolomcijferen, maar wordt meestal mechanisch rekenen genoemd.

Is zo'n positief kommagetal p groter dan een ander positief kommagetal a, hetgeen betekent dat, na mogelijk een aantal gelijke decimalen van p en a, er een eerste decimaal is van p die groter is dan de overeenkomstige decimaal van a, dan kunnen we precies één (positieve) b vinden waarvoor geldt dat $p = a + b$. Het voorbeeldje

$$\begin{array}{l} 0{,}909090909090909090... \\ \\ 0{,}222222222222222222... \\ \hline\hline \\ 0{,}686868686868686868... \end{array} \quad -$$

kan van linksaf kolomcijferend worden aangepakt. In eerste instantie is de eerste decimaal achter de komma dan gelijk aan 7 maar bij de volgende decimaal moet er van links 1 geleend worden om $10 - 2 = 8$ te krijgen, waarmee de 7 een 6 wordt. Al spelend zie je wel hoe het in het algemeen gaat, en ook dat $p > a$ gelijkwaardig is met $p + w > a + w$ voor ieder willekeurig ander positief getal w.

Nog een optelvoorbeeldje om het af te leren:

$$\begin{array}{l} 0{,}12345... \\ 0{,}99999... \\ \\ \hline\hline \\ 1{,}00000... \\ 0{,}11000... \\ 0{,}01200... \\ 0{,}00130... \\ 0{,}000140.. \\ 0{,}000015.. \\ ..... \\ \hline\hline \\ 1{,}12345... \end{array} \quad + \qquad (8)$$

Bij een doorlopend kommagetal het getal $0{,}\underline{9}$ optellen laat uiteindelijk alle cijfers achter de komma ongemoeid en telt een 1 op bij het getal voor de komma. En zo hoort dat ook. Na wat oefenen lukt dat ook wel in één keer en is het wellicht verstandig om nu verder te gaan met Sectie 26.1.4.

26.1.2 Vermenigvuldigen?

Kunnen we ook vermenigvuldigen? Dit gemene[17] sommetje bijvoorbeeld?

$$
\begin{array}{l}
0,999999..... \\
0,999999..... \\
\hline\times \\
\\
?????? \\
\end{array}
$$

In de vorige subsectie is het gelukt om de som van deze twee getallen kolomcijferend vanaf links zondere hogere wiskunde uit te werken. Kan dat met het produkt ook? We laten ons niet afschrikken en schrijven het produkt cijferend uit, waarbij we het *cijferen* symmetrisch houden in beide factoren, net zoals in (6) en (7) de uitwerking van de som symmetrisch in de bijdragen van de aparte termen was.

$$
\begin{array}{l}
0,999999..... \\
0,999999..... \\
\hline\times \\
\\
0,810000..... \\
0,081000..... \\
0,081000..... \\
0,008100..... \\
0,008100..... \\
0,008100..... \\
0,000810..... \\
0,000810..... \\
0,000810..... \\
0,000810..... \\
0,00008 10... \\
0,00008 10... \\
0,00008 10... \\
0,00008 10... \\
0,00008 10... \\
..... \\
\hline\times \\
\\
???????
\end{array}
\qquad (9)
$$

[17] Denk nog niet meteen aan $0,\underline{6} \times 0,\underline{6}$.

Dat ziet er een stuk ingewikkelder uit dan (6). Misschien is het wel geen goed idee het produkt van twee kommagetallen zo in één keer te willen doen. In deze doorlopende som zien we tussen de horizontale strepen de termen staan die we krijgen als we het produkt van de eerste decimaal van de eerste factor met de eerste decimaal van de tweede factor nemen (één term), van de eerste met de tweede en de tweede met de eerste (twee termen), van de eerste met de derde, de tweede met de tweede en de derde met de eerste (drie termen), enzovoorts. Gelukkig zien we links steeds meer nullen waardoor het lijkt of het blokje 81 naar rechts opschuift.

Ieder zulk blokje is het produkt van twee decimalen op steeds twee andere posities, decimalen die we hier toevallig allemaal gelijk aan 9 genomen hebben om de som[18] zo moeilijk mogelijk te maken. Het is de positie van het blokje dat opschuift, en op het blokje staat steeds het produkt van twee cijfers. Dus dit zijn de blokjes die voor kunnen komen:

$$00, 01, 02, 03, 04, 05, 06, 07, 08, 09, 10, 12, 14, 15, 16, 18, 20, 24,$$

$$25, 27, 28, 30, 32, 35, 36, 40, 42, 45, 48, 49, 54, 56, 63, 64, 72, 81.$$

Met alle blokjes gelijk aan 81 gaat het in (9) om de som van de getallen in het schema dat begint met

$$\begin{array}{cccccc}
 & & & \frac{81}{10^2} & & \\
 & & \frac{81}{10^3} & & \frac{81}{10^3} & \\
 & \frac{81}{10^4} & & \frac{81}{10^4} & & \frac{81}{10^4} \\
\frac{81}{10^5} & & \frac{81}{10^5} & & \frac{81}{10^5} & & \frac{81}{10^5} \\
 & \frac{81}{10^6} & & \frac{81}{10^6} & & \frac{81}{10^6} & & \frac{81}{10^6} & & \frac{81}{10^6} \\
\frac{81}{10^7} & & \frac{81}{10^7} & & \frac{81}{10^7} & & \frac{81}{10^7} & & \frac{81}{10^7} & & \frac{81}{10^7}
\end{array} \tag{10}$$

en naar beneden breder en breder doorloopt. Let wel, de volgorde waarin we cijferend optellen in (9) komt overeen met per regel optellen in (10) en leidt in de somnotatie tot

$$81 \times \sum_{n=1}^{\infty} \frac{n}{10^{n+1}} \tag{11}$$

[18] Het betreft $1 \times 1 = 1$, maar dat terzijde.

als maximale uitkomst (vast wel gelijk[19] aan 1) van een produkt van twee *nul komma (minstens) nog wat getallen*.

En met drie zulke getallen gaat het om maximaal

$$\begin{array}{cccccccccc} & & & & \frac{729}{10^3} & & & & & \\ & & & \frac{729}{10^4} & \frac{729}{10^4} & \frac{729}{10^4} & & & & \\ & & \frac{729}{10^5} & \frac{729}{10^5} & \frac{729}{10^5} & \frac{729}{10^5} & \frac{729}{10^5} & \frac{729}{10^5} & & \\ \frac{729}{10^6} & \frac{729}{10^6} & \frac{729}{10^6} & \frac{729}{10^6} & \frac{729}{10^6} & \frac{729}{10^6} & \frac{729}{10^6} & \frac{729}{10^6} & \frac{729}{10^6} & \frac{729}{10^6} \end{array} \tag{12}$$

enzovoorts, met $3, 6, 10, 15, 21, \ldots$ termen op elke regel, en maximaal

$$729 \times \sum_{n=1}^{\infty} \frac{n(n+1)}{2} 10^{n+2}$$

als maximale[20] uitkomst (vast wel gelijk aan 1) van een produkt van drie *nul komma (minstens) nog wat getallen*.

Het wordt er niet eenvoudiger op. We kijken nog een keer naar (9) waarmee we begonnen zijn. Het aantal niet-nullen is in de kolommen rechts van de komma achtereenvolgens $1, 3, 5, 7, 9, \ldots$, en

in kolom	1	2	3	4	5	6	7	8	...
zien we	1	3	5	7	9	11	13	15	...

niet-nullen. Per kolom gaan we bij het optellen dus onvermijdelijk over de 9 heen, en daarbij blijft het niet als we doorcijferen naar rechts, met een ruwe schatting

in kolom	1	2	3	4	5	6	7	8	...
maximaal	1×9	3×9	5×9	7×9	9	11×9	13×9	15×9	...

voor de kolomsommen, hetgeen leidt tot de vraag of

$$9 \times (\frac{1}{10} + \frac{3}{10^2} + \frac{5}{10^3} + \frac{7}{10^3} + \frac{9}{10^4} + \cdots) = 9 \times \sum_{n=1}^{\infty} \frac{2n-1}{10^n}$$

[19] Equivalent met

$$\sum_{n=1}^{\infty} \frac{n}{10^n} = \frac{10}{81}, \quad \text{wat zou} \quad \sum_{n=1}^{\infty} \frac{n^2}{10^n} \quad \text{zijn? Zie verder.}$$

[20] Het betreft immers $1 \times 1 \times 1 = 1$.

een decimaal ontwikkelbaar getal definieert waar *elke* eindige som van termen in (9) niet boven kan komen, een vraag vergelijkbaar met de minder ruw afgeleide vraag over (11). Maar het moge duidelijk zijn dat we opnieuw afdwalen van de basisschoolstof waar het hier toch om zou moeten gaan[21].

Het *cijferen* geeft wellicht meer begrip. Onhandig kolomcijferend zien we in (9) kolomsommen

$$8, 17, 26, 35, 44, 53, 62, 71, 80, 89, 98, 107, 116, 125, 134, 143, 152, 161, 170, 179,$$

enzovoorts verschijnen. Cijferend optellen geeft dat met weglating van de nul komma

$$\begin{array}{cccccccccccccccccc}
8 & 7 & 6 & 5 & 4 & 3 & 2 & 1 & 0 & 9 & 8 & 7 & 6 & 5 & 4 & 3 & 2 & 1 & 0 & 9 \\
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 8 & 9 & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 7 & 8 \\
 & & & & & & & & & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1
\end{array}$$

en alles loopt niet alleen naar rechts maar ook naar beneden door.

Opnieuw optellen per kolom geeft

$$\begin{array}{cccccccccccccccccc}
9 & 9 & 9 & 9 & 9 & 9 & 9 & 9 & 8 & 9 & 9 & 9 & 9 & 9 & 9 & 9 & 9 & 9 & 8 & 8 \\
 & & & & & & & & 1 & & & & & & & & & & 1 & 1
\end{array}$$

Enzovoorts. Zo te zien krijgen we op iedere plek inderdaad uiteindelijk een negen, maar alles loopt nog steeds (rechts) naar beneden door, al past het niet meer op de pagina.

De vraag is hoe we uitgaande van dit voorbeeld zien dat er voor twee willekeurige getallen zo altijd een decimale ontwikkeling van het produkt ontstaat, waarmee dan het produkt ondubbelzinnig vast ligt, en ook of er in het geval van het "maximale" voorbeeld alleen maar negens uitkomen. En ook voor produkten van drie getallen natuurlijk, met dezelfde overwegingen als bij optellen[22]. Als we dat wiskundig precies willen maken hebben we nodig dat volgordes en eerst samen nemen niet uit moet maken bij het optellen in doorlopende schema's beginnend als (12), als de breedte maar niet te snel toeneemt. Dat idee verkennen we in de volgende subsectie, waarin we opnieuw afdwalen van het cijferen.

26.1.3 Andere aftelbare sommen?

Een analysevraag om te stellen lijkt: voor welke rijen $a_1, a_2, a_3, \ldots$ gehele nietnegatieve getallen correspondeert een aftelbare maar niet eindige som

$$\sum_{n=1}^{\infty} \frac{a_n}{10^n} \tag{13}$$

[21] Voor een analysecursus zijn dit vragen om te onthouden!

[22] Lees: $a \times b \times c = (a \times b) \times c = c \times (a \times b)$, weer met alle gepermuteerde variaties.

ondubbelzinnig met een getal

$$\sum_{n=1}^{\infty} \frac{d_n}{10^n}$$

waarin alle d_n een cijfer zijn, i.e. $0, 1, 2, 3, 4, 5, 6, 7, 8$ of 9? Het liefst beantwoorden we die vraag zonder over andere uitdrukkingen dan die van de vorm (13) te praten.

Een noodzakelijke voorwaarde is dat de eindige sommen

$$A_1 = \frac{a_1}{10}, \quad A_2 = \frac{a_1}{10} + \frac{a_2}{10^2}, \quad A_3 = \frac{a_1}{10} + \frac{a_2}{10^2} + \frac{a_3}{10^3}, \quad \ldots \tag{14}$$

allemaal kleiner dan $1 = 0{,}\underline{9}$ zijn. Als $0{,}\underline{9}$ zo'n (strikte) bovengrens is dan is wellicht $0{,}8\underline{9}$ dat ook. Of niet. Kies het minimale cijfer d_1 waarvoor $0{,}d_1\underline{9}$ zo'n bovengrens is. Kies vervolgens het minimale cijfer d_2 waarvoor $0{,}d_1d_2\underline{9}$ een bovengrens is, enzovoorts. Dit proces definieert ondubbelzinnig een getal

$$0 = d_1d_2d_3\cdots = \sum_{n=1}^{\infty} \frac{d_n}{10^n}$$

dat kleiner is dan alle bovengrenzen $0{,}d_1\underline{9}$, $0{,}d_1d_2\underline{9}$, $0{,}d_1d_2d_3\underline{9}$, $\ldots$, en voor de bijbehorende

$$D_1 = \frac{d_1}{10}, \quad D_2 = \frac{d_1}{10} + \frac{d_2}{10^2}, \quad D_3 = \frac{d_1}{10} + \frac{d_2}{10^2} + \frac{d_3}{10^3}, \quad \ldots,$$

geldt dat

$$D_1 + \frac{1}{10}, \quad D_2 + \frac{1}{10^2}, \quad D_3 + \frac{1}{10^3}, \quad \ldots$$

bovengrenzen zijn. We kunnen niet uitsluiten dat na verloop van tijd alle d_n nul zijn, maar ze zijn zeker niet allemaal nul.

Kan het zo zijn dat de A_n-tjes niet boven de D_1 uitkomen? Wel, in dat geval zijn alle $A_n < D_1$ (want met $A_n = D_1$ komt een volgende A_n boven D_1), voor alle $n = 1, 2, 3, \ldots$, en was d_1 kennelijk niet minimaal gekozen om alle A_n onder $0{,}d_1\underline{9}$ te hebben. Dus A_n komt wel boven D_1 en blijft dan groter dan D_1. Hetzelfde geldt met het zelfde argument voor D_2, D_3, etcetera.

Als n_1 de eerste n is waarvoor $A_n > D_1$, n_2 de eerste n waarvoor $A_n > D_2$, n_3 de eerste n waarvoor $A_n > D_3$, etcetera, dan is n_2 minstens n_1, n_3 minstens n_2, n_4 minstens n_3, enzovoorts. We concluderen dat voor iedere k geldt dat

$$D_k < A_n < D_k + \frac{1}{10^k} \tag{15}$$

voor alle n vanaf $n = n_k$, en dat zou moeten betekenen dat

$$A_n \to D = 0{,}d_1d_2d_3d_4\ldots, \tag{16}$$

een nog niet precies gemaakte uitspraak voor een rij breuken A_n, breuken met noemers machten van 10 en $A_n \leq A_{n+1}$, met strikte ongelijkheid voor niet per se alle maar wel willekeurig grote n.

Elke A_n heeft decimalen genummerd door $j = 1, 2, 3, 4, \ldots$. De eerste decimaal kan niet kleiner worden met toenemende n. Dat betekent dat vanaf zekere $n = m_1$ de eerste decimaal van A_n niet meer verandert en gelijk is aan een vast cijfer α_1. Daarna geldt hetzelfde voor de tweede decimaal die vanaf zekere $n = m_2$ (waarbij we m_2 minstens gelijk aan m_1 kunnen nemen) niet meer verandert en gelijk is aan een vast cijfer α_2, enzovoorts.

Deze eigenschap moet voor de niet-dalende rij $A, A_2, A_3, \ldots$ toch wel de enige zinvolle definitie van

$$A_n \to 0{,}\alpha_1\alpha_2\alpha_3\alpha_4\ldots$$

zijn. Graag zouden[23] we nu uit (15) concluderen dat

$$0{,}d_1d_2d_3d_4\cdots = 0{,}\alpha_1\alpha_2\alpha_3\alpha_4\ldots,$$

waarbij we opmerken dat de ontwikkeling in het rechterlid bij constructie niet af kan breken maar de ontwikkeling in het linkerlid wel. Het kan dus gebeuren dat de eerste zoveel α_n en d_n hetzelfde zijn, daarna één keer $\alpha_n + 1 = d_n$, en vervolgens alle $d_n = 0$ en alle $\alpha_n = 9$. Hoe het ook zij, de uitdrukking in (13) definieert dus ondubbelzinnig een *nul komma minstens nog wat getal*, mits we weten dat alle eindige sommen in (14) kleiner zijn dan $0{,}\underline{9}$. Maar wie voor (10) en (12) meteen ziet dat dat inderdaad zo is mag het zeggen. We zijn er dus nog niet uit wat betreft produkten van positieve kommagetallen.

26.1.4 Een cijfer keer een kommagetal

Terug naar het cijferen. We houden ons nog even aan de afspraak dat positieve kommagetallen de getallen zijn met een na de komma doorlopende rij cijfers waarin niet-nullen blijven voorkomen hoe ver je ook gaat in de decimale ontwikkeling. Zo'n positief kommagetallen heeft voor de komma een natuurlijke getal of een 0 staan. Het produkt van twee zulke getallen moet wel de som van vier bijdragen zijn: wat je krijgt van voor de komma keer voor de komma, van voor de komma keer achter de komma, van achter de komma keer voor de komma, en van achter de komma keer achter de komma.

De laatste lijkt het moeilijkst. Als we die kunnen dan kunnen we daarna ook alle produkten van positieve kommagetallen door eerst de komma's naar links te schuiven en in het antwoord de komma naar rechts te schuiven. Twee

[23] Nog even nagaan dit dus.

keer naar rechts eigenlijk, om beide verschuivingen naar links goed te maken. Helaas zijn we hierboven nog niet bevredigend uit produkten van zulke *nul komma nog wat getallen* gekomen.

De eerste van de vier bijdragen is het makkelijkst, hoe het daarmee zit is basisschoolstof. De volgende twee bijdragen zijn wat lastiger. Met een 1-cijferig natuurlijk getal 1, 2, 3, 4, 5, 6, 7, 8 of 9 is de moeilijkste $9 \times 0{,}\underline{9}$. Net zo moeilijk is $0{,}9 \times 0{,}\underline{9}$:

$$
\begin{array}{l}
0,999999..... \\
0,9 \\
\text{———————} \times \\
0,810000..... \\
0,081000..... \\
0,008100..... \\
0,000810..... \\
0,0000810... \\
..... \\
\text{———————} \times \\
0,899999.....
\end{array}
$$

Daarna zijn produkten van cijfers met kommagetallen geen probleem meer. Met twee cijfers tegelijk in elke stap geeft een cijfers keer een blokje van twee maximaal $9 \times 99 = 891$. Cijferend per blokjes van twee vanaf links schuift er dus steeds maximaal een 8 naar links door. Bij het eerste blokje komt die gewoon voor het blokje te staan. Van het tweede blokje schuift er maximaal een 8 door naar links waarmee het blokje dat daar maximaal voor staat op zijn hoogst $91 + 9 = 99$ wordt. Enzovoorts. Het is weer instructief om een paar voorbeeldjes te doen en in één keer het antwoord op te schrijven op basis van de decimalen die je hebt in je voorbeeld.

26.1.5 Produkten van kommagetallen

Als het bovenstaande eenmaal in in één keer lukt als

$$
\begin{array}{l}
0,999999..... \\
0,9 \\
\text{———————} \times \\
0,899999.....
\end{array}
$$

dan kan daarna

$$
\begin{array}{l}
0,999999..... \\
0,999999..... \\
\text{————————} \times \\
\\
0,899999..... \\
0,089999..... \\
0,008999..... \\
0,000899..... \\
0,000089..... \\
.....
\end{array}
\tag{17}
$$

ook, en vervolgens kunnen we dan van boven af de kommagetallen term voor term optellen met wat we kolomcijferend geleerd hebben in sommetjes als (8).

De eerste stap is

$$
\begin{array}{l}
0,899999..... \\
0,089999..... \\
\text{————————} + \\
\\
0,899999..... \\
0,009999..... \\
0,080000..... \\
\text{————————} + \\
\\
0,909999..... \\
0,080000..... \\
\\
\text{————————} + \\
\\
0,989999.....
\end{array}
\tag{18}
$$

In (18) hebben we de tweede rij negens afgesplitst. Opgeteld bij het kommagetal erboven verhogen die de 89 tot 90, en met de 8 eronder maken ze van de 89 een 98, waarbij de decimalen achter de 89 ongewijzigd blijven. Het resultaat is de som van de eerste twee kommagetallen in (17), waarbij in dit voorbeeld de 8 eentje opgeschoven is naar rechts.

Zo gaat dat verder. Nu we met (17) zijn gevorderd tot

$$
\begin{array}{l}
0,999999..... \\
0,999999..... \\
\text{————————} \times \\
\\
0,989999..... \\
0,008999..... \\
0,000899..... \\
0,000089..... \\
.....
\end{array}
\tag{19}
$$

zien we dat het patroon zich herhaalt in

$$
\begin{array}{l}
0,989999..... \\
0,008999..... \\
\text{————————} + \\
\\
0,998999.....
\end{array}
$$

met als resultaat de som van de eerste drie kommagetallen in (17). De 8 is weer eentje opgeschoven en dat gaat zo door. In de volgende stap zien we

$$
\begin{array}{l}
0,999999..... \\
0,999999..... \\
\text{————————} \times \\
\\
0,998999..... \\
0,000899..... \\
0,000089..... \\
.....
\end{array}
\tag{20}
$$

met nu boven de drie nullen na de komma in (20) alleen het derde cijfer dat nog zal veranderen bij verder cijferen. Zo vinden we al cijferend dat

$$0,\underline{9} \times 0,\underline{9} = 0,\underline{9},$$

hetgeen zoveel wil zeggen dat $1 \times 1 = 1$.

Is ieder tweetal kommagetallen zo cijferend met elkaar te vermenigvuldigen? Merk op dat een staartstuk in de ontwikkeling van de tweede factor steeds maximaal uit een rij negens bestaat en zo het cijfer in het antwoord op de positie waarna dat staartstuk begint maximaal met 1 verhoogt.

Om nog verder uit te werken dit alles, maar niet hier. Het idee is wel duidelijk nu. Zonder hier nu meteen Turing aan te roepen is het aardig om deze sectie te besluiten met de opmerking dat je in gedachten een machientje zou kunnen maken dat als input de doorlopende kommagetallen krijgt die als het ware van de ene kant cijfer voor cijfer naar binnen schuiven, en dan vervolgens aan de andere kant als output de som of produkt cijfer voor cijfer als doorlopend kommagetal uitspuugt, en het machientje daarmee tot het einde der tijden doorgaat.

26.2 Kleinste bovengrenzen

Net als de aftelbare som in (1) met $k \geq 0$ is (3) een mooi voorbeeld van

$$\sum_{n=0}^{\infty} a_n \tag{21}$$

met $a_n \geq 0$ voor alle $n \in \mathbb{N}_0 = \mathbb{N} \cup \{0\}$. Als de partiële sommen

$$S_N = \sum_{n=0}^{N} a_n$$

begrensd zijn dan is de *kleinste bovengrens* van de aftelbare vereniging

$$\cup_{N \in \mathbb{N}_0} \{S_N\} = \{S_0, S_1, S_2, \dots\}$$

per definitie de som van de *reeks* in (21), notatie

$$S = \sum_{n=0}^{\infty} a_n.$$

In het geval van (1) is $S_N \leq k+1$ voor alle $N \in \mathbb{N}_0$ en kan deze uitspraak dus als *tautologie* gezien worden: het reële getal S is de limiet van zijn decimale ontwikkeling, een ontwikkeling waarin de decimalen d_n uit de cijfers 0 tot en met 9 gekozen worden.

Dat het überhaupt mogelijk is dat er uit een som met oneindig veel termen als (21) een eindig getal kan komen is zo vanuit (1) vanzelfsprekend, ook al dacht ene Zeno daar destijds anders over. Mooie voorbeelden waarbij er uit de som geen eindig getal komt zijn

$$S = \sum_{n=0}^{\infty} 1 \quad \text{met} \quad S_N = N, \quad \text{en} \quad S = \sum_{n=0}^{\infty} \frac{1}{n}. \tag{22}$$

Geen van deze twee definieert een $S \in \mathbb{R}^+$.

Waarom eigenlijk niet? Wel, de eerste S zou een kleinste bovengrens in $\mathbb{R}$ voor de verzameling $\mathbb{N}$ zijn. Maar dan is $S - \frac{1}{2}$ geen bovengrens voor $\mathbb{N}$. En dus is er een $N \in \mathbb{N}$ met $N > S - \frac{1}{2}$ en volgt dat $N + 1 > S + \frac{1}{2}$. Maar $N + 1 \in \mathbb{N}$ dus is S geen bovengrens voor $\mathbb{N}$, een tegenspraak[24]. Gelukkig maar, want het zou wel heel gek zijn als $\mathbb{N}$ wel begrensd is in $\mathbb{R}$. Komt meteen te pas bij het tweede voorbeeld in (22), waarover we opmerken dat

$$\underbrace{1 + \frac{1}{2} + \underbrace{\frac{1}{3} + \frac{1}{4}}_{>\frac{1}{2}}}_{>1} + \underbrace{\underbrace{\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}}_{>\frac{1}{2}} + \underbrace{\frac{1}{9} + \frac{1}{10} + \frac{1}{11} + \frac{1}{12} + \frac{1}{13} + \frac{1}{14} + \frac{1}{15} + \frac{1}{16}}_{>\frac{1}{2}}}_{>1},$$

enzovoorts, en zo komen de bijbehorende S_N boven elke $n \in \mathbb{N}$. Ook niet begrensd dus. Maar de som

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \cdots$$

heeft wel een uitkomst[25], althans indien opgevat als

$$\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n},$$

al zijn noch de positieve termen

$$1 + \frac{1}{3} + \frac{1}{5} + \frac{1}{7} + \cdots,$$

noch de negatieve termen

$$-\frac{1}{2} - \frac{1}{4} - \frac{1}{6} - \cdots$$

op te tellen tot een eindige som. Sterker, gegeven een $S \in \mathbb{R}$ kun je de positieve en negatieve termen verweven[26] tot een rij a_n op zo'n manier dat

$$S = a_1 + a_2 + a_3 + a_4 + \cdots,$$

een goede reden om zoveel mogelijk alleen maar over reeksen zoals in Sectie 26.3 te spreken.

[24] Overtuigd?

[25] Ik meen $\ln 2$.

[26] Kies positieve termen om boven S te komen, dan negatieve om onder S, dan ...

26.3 Absoluut convergente reeksen

Als we hadden leren rekenen met $d_n \in \{-5,-4,-3,-2,-1,0,1,2,3,4\}$ en de tafels tot en met vijf, dan was (1) een voorbeeld geweest van (21) zonder de a priori informatie dat $a_n \geq 0$ maar *wel* met de eigenschap dat

$$\sum_{n=0}^{\infty} |a_n| < \infty, \tag{23}$$

omdat

$$\sum_{n=1}^{\infty} |\frac{d_n}{10^n}| \leq \sum_{n=0}^{\infty} \frac{5}{10^n} = \frac{5}{10} + \frac{5}{100} + \frac{5}{1000} + \frac{5}{10000} + \frac{5}{100000} + \cdots = \frac{5}{9}.$$

Ook nu geldt dat $S_N \to S$ voor een unieke $S \in \mathbb{R}$, dus

$$S = \sum_{n=0}^{\infty} a_n, \tag{24}$$

en hernummeren van de som verandert niets aan die uitkomst. Reeksen van de vorm (21) waarvoor (23) geldt heten *absoluut convergent* en zijn *onvoorwaardelijk convergent*: de volgorde van sommeren maakt niet uit voor de waarde S van de som en bovendien geldt dat

$$|S| = |\sum_{n=0}^{\infty} a_n| \leq \sum_{n=0}^{\infty} |a_n|. \tag{25}$$

Wat betreft het bewijs van (24) gegeven (23), de invariantie onder hernummeren en de aftelbare 3-hoeksongelijkheid (25): dat bewijs maakt gebruik van het feit dat in de reële getallen *Cauchyrijen*, dat zijn rijen waarvoor geldt dat

$$x_n - x_m \to 0 \quad \text{als} \quad m,n \to \infty,$$

een unieke limiet $\bar{x}$ hebben, een limiet $\bar{x}$ die bestaat als dan inderdaad het *enige* reële getal waarvoor

$$x_n \to \bar{x} \quad \text{als} \quad n \to \infty.$$

Zulke rijen heten *convergent*.

Uit Hoofdstuk 10 van [HM] of Hoofdstuk 8 van het *Basisboek Wiskunde* is de lezer wellicht al bekend met de wiskundige definitie van het begrip (limiet van een) convergente rij, waarin alleen [27] de "voor alle $p > 0$" nog door een

[27] Didactisch aardig in het basisboek is het gebruik van grote P naast kleine p.

"voor alle $\varepsilon > 0$" moet worden vervangen om tot het gebruikelijke jargon te komen, en later eventueel door $\forall \varepsilon > 0$. Wel is het in de analyse straks *praktischer* om met

$$|x_n - \bar{x}| \leq \varepsilon$$

te werken.

Wat ook elegant en praktisch is in het Basisboek Wiskunde is de zorgvuldige manier waarop gesproken wordt over *de rij waarvan het n-de element gelijk is aan* x_n, en het aan de lezer wordt overgelaten zich daarbij te realiseren dat n de getallen $1, 2, 3, 4, \ldots$ doorloopt, of een andere steeds met stap 1 oplopende rij gehele getallen. Wij zullen de notatie in het Basisboek Wiskunde afkorten tot simpelweg *de (door n genummerde) rij* x_n, vaak de rij reële getallen x_n. Evenzo spreken we over de rij rationale getallen q_n of de rij $q_n \in \mathbb{Q}$. De laatste notatie wordt hieronder nog gebruikt.

De ε-definities van uitspraken als hierboven komen in deze cursus aan de orde op het moment dat dat nodig is. Want ze zijn nodig, bijvoorbeeld om precies te maken dat sommen als (24) bestaan du moment dat je met één $M \in \mathbb{R}^+$ een schatting

$$\sum_{n=0}^{N} |a_n| \leq M$$

hebt voor alle partiële sommen *tegelijk*, en daaruit afleidt dat de door N genummerde rij S_N een Cauchyrij is. We merken hierbij op dat het in zogenaamde genormeerde ruimten dan om twee equivalente uitspraken gaat, uitspraken waarin noch de limiet $\bar{x}$ van de rij, noch de som S van de reeks waar het om gaat expliciet voorkomen:

absoluut convergente reeksen convergent $\iff$ Cauchyrijen convergent

Je kunt dus weten of $\bar{x}$ en S in $\mathbb{R}$ bestaan zonder ze eerst te hebben bepaald.

26.4 Verzamelingen in de praktijk

Voor sommige wiskundigen van de meer zuivere inclinatie zijn de uitspraken hierboven niet los te zien van een precieze maar voor de analyse zelf niet altijd even verhelderende wiskundige constructie van de reële getallen. Maar interessant zijn die constructies natuurlijk wel, en je moet ergens beginnen als je de wiskunde per se axiomatisch en wiskundig streng wil opzetten[28], vanuit wat men de leer van verzamelingen noemt.

Deze verzamelingenleer is iets waarover Paul Halmos in zijn mooie boekje Naive Set Theory [29] schreef: alle wiskundigen vinden dat je er wat van

[28] Een vriendje van Einstein heeft helaas laten zien dat dat nooit bevredigend zal lukken.
[29] Vertaald ooit als Prisma pocket verkrijgbaar.

gezien moet hebben, maar ze zijn het oneens over *wat* precies. Je kunt verzamelingenleer bijvoorbeeld bij het begin beginnen met het axioma dat de lege verzameling[30] bestaat.

Dat doen wij hier niet. Maar mocht je dat wel doen dan komen toch op enig moment ook de axioma's voor de verzameling van de natuurlijke getallen $\mathbb{N}$ voorbij, natuurlijke getallen die iedereen die op zijn vingers heeft leren tellen allang kent. En tellen begint natuurlijk bij 1[31], al is het handig om de verzameling

$$\mathbb{N}_0 = \mathbb{N} \cup \{0\} = \{0, 1, 2, 3, \dots\}$$

in te voeren, hier in een zuiver wiskundig gezien af te keuren maar wel zo begrijpelijke notatie met onfatsoenlijke stippeltjes, waarin we het hierboven al gebruikte verenigingssymbool $\cup$ weer terugzien[32].

Het is wel goed om één van die axioma's voor $\mathbb{N}$ te relateren aan de wiskundige praktijk van alledag. Want hoe bewijs je bijvoorbeeld dat voor iedere $N \in \mathbb{N} = \{1, 2, 3, \dots\}$ geldt dat de uitspraak

$$(P_N) \qquad 1^2 + 2^2 + \cdots + N^2 = \frac{N^3}{3} + \frac{N^2}{2} + \frac{N}{6}$$

waar is, *zonder* voor elke $N \in \mathbb{N}$ *apart* de uitspraak (P_N) te moeten controleren?

Binnen de zuivere wiskunde hoort daar een verhaal bij waarin voor al de puntjes hierboven eigenlijk geen plaats is. Dat verhaal eindigt met het principe van volledige inductie[33], dat er op neerkomt dat[34] als je voor $N = 1$ de uitspraak controleert, en je vervolgens laat zien dat de *implicatie*

$$(P_N) \implies (P_{N+1}) \tag{26}$$

geldt voor alle N waarvoor je hem nodig hebt, namelijk om via herhaald toepassen van de inductiestap (26) tot

$$(P_1) \implies (P_2) \implies (P_3) \implies (P_4) \implies (P_5) \implies (P_6) \implies (P_7) \implies \cdots$$

te komen, zover als je maar wil, de uitspraak inderdaad geldt voor alle $N \in \mathbb{N}$. De implicatie (26) moet daartoe voor alle $N \in \mathbb{N}$ worden aangetoond om beginnend met de juistheid voor $N = 1$ de keten hierboven zonder die stippeltjes in één keer af te maken.

[30] In LaTeχ: $\emptyset$. Op het schoolbord liever $\not\emptyset$.
[31] Tellend is $\mathbb{N}$ met nul een beetje flauwe kul, op 0^0 komen we nog terug.
[32] Gaat er dus eigenlijk om hoe je al die getallen zonder stippels tussen accolades vangt.
[33] Naamgeving volledig intimiderend, vriendelijker is: dominoprincipe.
[34] Nu komt een lange zin.

Dit soort oefeningen kunnen elders gedaan worden. Zie bijvoorbeeld in [HM] Sectie 6.1 en voetnoot 16. Relevant uit die sectie voor deze cursus zijn bewijzen voor rekenpartijtjes als in (P_N) hierboven, waarin niet alleen N maar ook $3 = p \in \mathbb{N}$ een parameter is, en de inductiestap van de vorm

$$(P_1) \wedge (P_2) \wedge \cdots \wedge (P_N) \implies (P_{N+1}) \tag{27}$$

is[35]. Zie verder ook Hoofdstuk IV [BE] .

Wat we hier precies met $(A) \implies (B)$ bedoelen moge duidelijk zijn: als de uitspraak (A) waar is dan is ook de uitspraak (B) waar. Hetgeen in onze wiskundige redenaties equivalent is met: als uitspraak (B) niet waar is dan kan uitspraak (A) ook niet waar zijn. Deze logica kan geformaliseerd worden met waarheidstabellen vol nullen en enen opgeleukt met bijzonder fraaie algebra, maar wat dat betreft laten we hier liever de Boole de Boole[36].

Du moment dat er over het bestaan van[37] $\mathbb{N}$ geen twijfel meer is, worden in de verzamelingsleer, bijvoorbeeld zoals in Hoofdstuk V en VI van [BE], $\mathbb{Z}$, $\mathbb{Q}$ en uiteindelijk $\mathbb{R}$ *wiskundig netjes* geconstrueerd. De constructie van $\mathbb{R}$ is in [BE] gebaseerd op de gedachte dat iedere manier om $\mathbb{Q}$ in twee stukken te knippen overeen zou moeten komen met een reëel getal, waarbij de rationale getallen dan wel met de schaar te maken krijgen en de overige getallen niet[38].

Het is instructief om de constructies van $\mathbb{Z}$ en $\mathbb{Q}$ uit $\mathbb{N}$ met elkaar te vergelijken zoals dat gebeurt in [BE]. Die van $\mathbb{Z}$ is inderdaad tamelijk kunstmatig. Die van $\mathbb{Q}$ is echter heel natuurlijk en gebaseerd op hoe je eigenlijk altijd al met de rationale getallen rekende, namelijk als breuken. Breuken met een teller en een noemer. Bijvoorbeeld

$$\frac{14}{333} = \frac{42}{999} = 0.\underline{042}$$

met een streep die aangeeft dat de decimale ontwikkeling van de breuk zich herhaalt. Anders dan gesuggereerd in de in

`http://www.few.vu.nl/~jhulshof/TAL.pdf`

besproken TAL-boekjes van het Freudenthal Instituut doe je echter het rekenen met rationale getallen bij voorkeur niet met zulke decimale ontwikkelingen, maar juist wel met de niet unieke representatie van rationale getallen als quotiënten van gehele getallen, dus in de vorm

$$q = \frac{t}{d}$$

[35] De $\wedge$ staat voor "en", dat is logisch. Denken aan dominosteentjes is nu lastiger.
[36] https://www.youtube.com/watch?v=DOzqUyW7jog
[37] Eventueel via Peano's axioma's.
[38] Want ze bestaan op dat moment nog niet.

met teller t en noemer d in $\mathbb{Z}$, de d niet gelijk aan 0, waarbij je moet afspreken dat

$$\frac{t_1}{d_1} = \frac{t_2}{d_2} \quad \text{als} \quad t_1 d_2 = t_2 d_1.$$

26.5 Equivalentierelaties

Wiskundigen noemen zo'n afspraak een equivalentierelatie. We komen nu in relatie tot $\mathbb{R}$ meer over dit belangrijke begrip te spreken, ook voor wie van $\mathbb{R}$ graag een inzichtelijke constructie wil zien. Een constructie waarvan de details overigens niet thuis horen in of voorafgaand aan een eerste vak Analyse. Ik meen dat ik zelf de constructie van $\mathbb{R}$ voor het eerst zag bij een college over de integraal van Lebesgue van Jan van de Craats in het vierde semester van wat toen de kandidaatsstudie wiskunde in Leiden was.

De onderliggende maattheorie voor dat vak over die andere integraal begint met de vraag wat de *oppervlakte* $|A|$ is van een willekeurige deelverzameling A van $\mathbb{R}^2$, en komt onvermijdelijk tot twee constateringen. Vroeger of later zijn dat respectievelijk

(i) het komt voor dat $A \subset B$ en $|A| = |B|$;

(ii) het zou kunnen voorkomen dat A eindige oppervlakte $|A|$ heeft maar opgeknipt kan worden in aftelbaar veel stukjes die allemaal dezelfde maat zouden moeten hebben[39],

en daar moet je mee omgaan. Leuk is dat (ii) ons dan later[40] weer terugvoert naar het boekje van Halmos. In een vroeger stadium doet (i) ons echter al het dringende verzoek om A en B in zekere[41] zin als hetzelfde te zien, en bijvoorbeeld ook hetzelfde als een C met $C \subset A$ en $|C| = |A|$, waarbij C geen deelverzameling van B hoeft te zijn of omgekeerd. Hoe formuleer je dan rechtstreeks dat B en C equivalent zijn?

Anders van aard is het gebruik van equivalentierelaties bij een inzichtelijke constructie van $\mathbb{R}$, waarbij je denkt aan reële getallen als denkbeeldige limieten van Cauchyrijtjes rationale getallen, zoals bijvoorbeeld de hierboven besproken decimale ontwikkelingen, maar dan moet je wel een goede afspraak maken over wat het betekent dat twee zulke Cauchijrijtjes hetzelfde reële getal (zouden moeten) definiëren. Denk bijvoorbeeld aan binaire benaderingen met alleen maar nullen en enen, of aan benaderingen met kettingbreuken, allebei erg fraai of juist minder[42] fraai, omdat ze afstand nemen van de vingers waarin onze wiskunde zit. Kortom, een belangrijke vraag is hoe je van

[39] Waarom is dat een paradox?

[40] Maar niet hier.

[41] Lees: in maatheoretische zin.

[42] Over gebrek aan smaak valt niet te twisten.

twee Cauchyrijen rationale getallen q_n en r_n zegt dat ze hetzelfde reële getal definiëren[43].

Als je er even over nadenkt is het logisch dat dit een definitie zou kunnen zijn:

$$q_n \sim r_n \quad \Longleftrightarrow \quad q_n - r_n \to 0 \quad \text{als} \quad n \to \infty$$

Deze tweezijdige equivalentiepijl definieert een *equivalentierelatie* op de verzameling van alle rijen rationale getallen. We walsen nu wellicht even over wat belangrijke details heen, maar een equivalentierelatie is niets anders dan een relatie met formeel dezelfde eigenschappen als de gelijkheidsrelatie voor elementen van een willekeurige verzameling A. Voor alle $a, b, c \in A$ geldt

$$a = a,$$

$$a = b \implies b = a,$$

$$a = b \wedge b = c \implies a = c$$

De relatie[44] gedefinieerd door het $=$ teken heet daarom reflexief, symmetrisch en transitief, en ook $\sim$ is zo'n equivalentierelatie, op de verzameling van alle rijen rationale getallen in dit geval. En die equivalentierelatie doet het!

Wat doet $\sim$ dan? De equivalentierelatie $\sim$ deelt de verzameling van alle rijen rationale getallen in. Waarin? In equivalentieklassen natuurlijk. Iedere rij $r_n \in \mathbb{Q}$ definieert een equivalentieklasse

$$[r_n] = \{q_n \in \mathbb{Q} : \, q_n \sim r_n\} \tag{28}$$

waar die rij zelf in zit, en een reëel getal is *per definitie* de *equivalentieklasse* van een Cauchyrij $r_n \in \mathbb{Q}$.

So much for the construction of the real numbers en we zullen het $\sim$ tekentje nu weer in laten leveren, omdat we dat symbool toch liever gebruiken als

$$x_n \sim y_n \quad \Longleftrightarrow \quad \frac{x_n}{y_n} \to 1 \quad \text{als} \quad n \to \infty$$

voor een andere en in de praktijk vaker gebruikte equivalentierelatie[45] op de verzameling van alle reële rijen. Een voorbeeld is

$$n! \sim (\frac{n}{e})^n \sqrt{2\pi n},$$

uitvoerig besproken[46] in [HM].

[43] In je hoofd of op de getallenlijn.

[44] Ook dat woord heeft een *wiskundige* definitie natuurlijk.

[45] Wel een goede vraag hierboven is: wat is de beste representant?

[46] Zie echter ook Sectie 5.8.

26.6 Analyse in en van wat?

Of er nog andere verzamelingen zoals deze $\mathbb{R}$ zijn is niet een standaardvraag om hier te stellen. Wel belangrijk voor een eerste vak over analyse is dat de *rationale getallen* $\mathbb{Q}$, dat zijn de getallen die ontstaan als quotiënten van getallen in $\mathbb{Z}$ en getallen in $\mathbb{N}$, de "Cauchy eigenschap" niet hebben. Dat is de reden waarom we de analyse in $\mathbb{R}$ doen, het unieke geordende getallenlichaam waarin (alle) Cauchyrijen en absoluut convergente reeksen convergent zijn.

In het Basisboek Wiskunde worden deze getallen besproken in Hoofdstuk 24, en we gebruiken vrijwel dezelfde notaties, met de accolades ook. Lees ook Hoofdstuk 25 nog even door, we nemen de daar gebruikte input-output voorstelling voor functies[47] hier graag over als

$$x \xrightarrow{f} f(x) \quad \text{en} \quad D_f \xrightarrow{f} \mathbb{R}$$

met D_f het domein van f. Het bereik en de grafiek[48] van f zijn

$$B_f = \{f(x) : x \in D_f\} \quad \text{en} \quad G_f = \{(x, y) : x \in D_f,\, y = f(x)\}.$$

Soms zullen we liever over functies $f : \mathbb{R} \to \mathbb{R}$ spreken die op een bepaalde deelverzameling van $\mathbb{R}$ een bepaalde eigenschap hebben. Het *domein* D_f is dan de verzameling bestaande uit alle $x \in \mathbb{R}$ waarvoor $f(x)$ gedefinieerd is. Is het domein van f niet heel $\mathbb{R}$, dan kun je natuurlijk altijd $f(x)$ voor x-waarden buiten het domein een waarde geven die je toevallig goed uitkomt, nul bijvoorbeeld[49].

In deze cursus behandelen we ondermeer de analyse die de calculus onderbouwt voor functies $f : I \to \mathbb{R}$ met $I \subset \mathbb{R}$ een interval. Vaak, met $a, b \in \mathbb{R}$, is I daarbij een gesloten begrensd interval

$$I = [a, b] = \{x \in \mathbb{R} : a \leq x \leq b\},$$

of een open begrensd interval

$$I = (a, b) = \{x \in \mathbb{R} : a < x < b\}.$$

We beginnen met integraalrekening, eerst voor monotone functies, zonder over limieten te spreken, en daarna voor uniform continue functies $f : [a, b] \to \mathbb{R}$, waarbij we voor het eerst het limietbegrip tegenkomen en nodig hebben.

Voor zulke functies wordt

$$\int_a^b f(x)\, dx$$

via benaderende sommen gedefinieerd in relatie tot wat de oppervlakte van het gebied ingesloten door $x = a$, $x = b$, $y = 0$ en $y = f(x)$ in het x, y-vlak moet zijn in het geval dat f een positieve functie is. Je zou kunnen zeggen dat dit de eerste *probleemstelling* is in dit boek, geformuleerd in drie punten als:

Teken plaatje!

hoe definieer je de oppervlakte van niet meteen arbitraire verzamelingen;

en hoe reken je die vervolgens uit?

wat kun je vervolgens leren van de oplossing?

Dat laatste doe je dan wellicht zonder meteen een nieuw probleem te willen formuleren. Spelen met de verworven inzichten zonder een concreet doel op zich.

Bijvoorbeeld: met een variabele bovengrens in de integraal ontdekken we de opzet van de differentiaalrekening met behulp van lineaire benaderingen. Die werken we later uit voor *machtreeksen*

$$P(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3 + \cdots ,$$

waarmee we een grote klasse van standaardfuncties tot onze beschikking krijgen, waarvoor "mag dat" vragen kort maar krachtig met "ja natuurlijk" te beantwoorden zijn. Binnen die klasse is de analyse namelijk ondergeschikt aan de algebra, en zodra je die algebra goed begrijpt, voor x^7 of zo, ben je wel klaar en daarmee dient een andere probleemstelling zich aan:

Hoe zit het met al die andere functies?

Als we buiten de klasse van machtreeksen treden verandert alles en moet er gewerkt worden. Dat werk halen we nu naar voren, waar we dat in [HM] zo lang mogelijk uitstelden.

De *middelwaardestelling* blijkt het belangrijkste hulpmiddel om ogenschijnlijk evidente uitspraken ook werkelijk te bewijzen. De uitspraak van die stelling is dat differentiequotiënten als

$$\frac{F(b) - F(a)}{b - a},$$

de richtingcoëfficiënt van

$$\text{de lijn door } (x, y) = (a, F(a) \text{ en } (x, y) = (b, F(b))$$

[47] Van het Latijnse fungor (deponens: een passieve vorm met actieve betekenis).
[48] Vaak slordig: de grafiek $y = f(x)$ in het x, y-vlak.
[49] Zoals wel eens voorgesteld in relatie tot $x \to \frac{1}{x}$ en het rekenonderwijs.

in het x, y-vlak, zelf doorgaans gelijk zijn aan de richtingscoëfficiënt van de raaklijn aan de grafiek van F in een punt met x-waarde tussen a en b, lees: aan de met de differentiaalrekening gedefinieerde

$$\text{afgeleide van } F'(x) \text{ van } F(x) \text{ in een tussenpunt.}$$

Is $F'(x)$ overal tussen a en b gelijk aan nul, dan is $F(x)$ kennelijk constant, aldus de NIET-TRIVIALE STELLING in [HM]. Die STELLING is niet zozeer de oplossing van een probleem, maar formuleert juist iets dat je zeker wil weten bij het oplossen van (bijvoorbeeld) differentiaalvergelijkingen. Het bewijs van de STELLING maakt essentieel gebruik van een fundamentele stelling over het bestaan van convergente deelrijen, die, triviaal[50] of niet, toch maar een apart hoofdstuk krijgt, waarin een wat minder bekende stelling die ik ken via Han Peters wordt geformuleerd.

Vergelijkingen oplossen is een belangrijke tak van niet alleen maar recreatieve *sport*[51] in de wiskunde. In de context van vergelijkingen van de vorm $F(x, y) = 0$, waarbij F een functie is van twee variabelen, introduceren we daarom ook meteen maar het begrip impliciete functie, met als speciaal geval het al behandelde begrip inverse functie. Het bewijs van de impliciete functiestelling draaien we binnenste buiten in een aparte sectie, gevolgd door twee secties waarin weer met verworven inzichten wordt gespeeld en een basis wordt gelegd voor alles dat later komt. Na een zijstapje over de methode van Newton wordt de basale theorie in Hoofdstuk 4 afgesloten met differentiaalrekening voor integralen met parameters, en partieel integreren en een stelling over Taylorbenaderingen met polynomen.

Daarna nemen we in Hoofdstuk 5 de tijd voor voorbeelden en meer voorbeelden, en herhalen de rekenregels nog een keer in de kale context van functies van één variabele zonder er functies van x en y bij te halen, niet alleen voor wie dat stuk in Hoofdstuk 4 heeft overgeslagen. We gaan uitvoerig in op de natuurlijke logaritme ln als inverse van de exponentiële functie exp en introduceren in die context ook zogenaamde asymptotische formules, waarvan de formule van Stirling[52] voor $n!$ als $n \to \infty$ een mooi voorbeeld[53] is.

In het tweede deel, dat begint met Hoofdstuk 6, kunnen we de meeste van de in het eerste deel geformuleerde definities, stellingen en bewijzen uit de differentiaalrekening voor functies van $\mathbb{R}$ naar $\mathbb{R}$ vrijwel letterlijk overnemen. Alleen de notaties hoeven nog te worden uitgepakt. We beginnen daartoe met $\mathbb{C}$, de verzameling van de complexe getallen, en een in ons Leidse wat vergeten maar wel zo snel bewijs van de hoofdstelling van de algebra. Daarna komen

[50] Denk ook aan valsspelen met meetwaarden.

[51] Geen sport zonder *techniek*.

[52] De voorbeeldformule met $\sim$ een paar pagina's terug, uit te spreken als "twiddles".

[53] En buitengewoon relevant voor probleemstellingen in de natuurkunde.

functies van $\mathbb{C}$ naar $\mathbb{C}$ en afbeeldingen en functies met meerdere variabelen. Lineaire functies beschrijven we dan in matrixnotatie, en matrixrekening behandelen we daartoe zo kort door de bocht als hier mogelijk en voor het uitpakken voldoende is.

De kettingregel is een belangrijk voorbeeld en we laten zien hoe die regel op verschillende manieren wordt gebruikt, ook in de door fysici gebruikte manipulaties met afhankelijke en onafhankelijke grootheden bij het transformeren en oplossen van partiële differentiaalvergelijkingen. Integraalrekening in het vlak wordt nog wat kort behandeld, zowel in rechthoekige als in de uitvoerig besproken poolcoördinaten.

Nieuw is daarna de opzet van complexe functietheorie met lijnintegralen over alleen maar lijnstukjes en meteen de belangrijke hoofdstellingen, eerst zonder kromme poespas. Daarna bekijken we onderzoekend wat voor kromme krommen we na limietovergangen krijgen, en hoe we daarlangs kunnen integreren. De aanpak is zo precies tegenovergesteld aan de die van Conway, wiens fraaie opzet met equivalentieklassen van rectificeerbare krommen hier niet realiseerbaar is. Naast, voor of na de kromme aanpak, verkennen we de toepassingen van de hoofdstellingen bij het uitbreiden van de definitie van $f(z)$ met $z \in \mathbb{C}$ naar $f(A)$, eerst voor $A : \mathbb{R}^2 \to \mathbb{R}^2$ een lineaire afbeelding gegeven door een matrix, en daarna algemener, voor A van een (uiteindelijke complexe) Banachruimte X naar zichzelf. Een eerste kennismaking met Banachalgebra's[54] ligt hier voor de hand.

Gewone differentiaalvergelijkingen, al aan de orde geweest in de context van machtreeksen, motiveren het opnemen van Hoofdstuk 10, waarin we ook de calculus voor functies op en naar Banachruimten introduceren, met als belangrijkste voorbeeld $X = C([a, b])$, de ruimte van de continue $\mathbb{R}$-waardige functies op een interval $[a, b]$, waarin we de bijbehorende integraalvergelijkingen formuleren en oplossen.

De impliciete functiestelling uit Hoofdstuk 4 kan dan weer worden overgeschreven. In Hoofdstuk 6 doen we dat al in één moeite door in combinatie de multiplicatorenmethode van Lagrange[55] voor stationaire punten van gewone functies van meer variabelen onder randvoorwaarden. Essentieel hier is het inzicht dat de oplossingsverzameling van een stelsel van bijvoorbeeld 3 vergelijkingen in $\mathbb{R}^{5=2+3}$, lokaal te schrijven is als de grafiek van een functie van $x \in \mathbb{R}^2$ naar $y \in \mathbb{R}^3$, tenzij er te veel nullen in de relevante berekeningen voorkomen.

De term onderdompeling wordt hier nog niet geintroduceerd[56]. De meer

[54] Door mijn medestudenten destijds ook wel Bananachalgebra's genoemd.

[55] De eerste stelling die ik ooit zelf aan anderen uitlegde, maar nu heel anders.

[56] Zie www.encyclo.nl/begrip/Submersie èn www.encyclo.nl/begrip/Immersie.

abstracte formulering van de methode van Lagrange in Hoofdstuk 11 is opgenomen for amusement. Ook wat pittiger is de behandeling van tweede orde afgeleiden die we pas in abstracte setting in meer detail doen. Het hoofdresultaat is het Lemma van Morse, Stelling 11.6, waarin met een coördinatentransformatie een functie waarvan de tweede afgeleide continu is, in de buurt van een stationair punt puur kwadratisch gemaakt wordt. Denk aan

$$F(x,y) = ax^2 + bxy + cy^2 + \cdots$$

en een transformatie die de puntjes wegwerkt als de discriminant niet gelijk is aan 0.

Zo'n transformatie is van de vorm

$$\begin{pmatrix} \xi \\ \eta \end{pmatrix} = A(x,y) \begin{pmatrix} x \\ y \end{pmatrix},$$

met $A(x,y)$ een van x en y afhankelijke matrix met

$$A(0,0) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

die maakt dat

$$F(x,y) = a\xi^2 + b\xi\eta + c\eta^2.$$

We laten zien hoe $A = A(x,y)$ gevonden kan worden als oplossing van een kwadratische matrixvergelijking voor A die met worteltrekken kan worden opgelost.

We besluiten met Hoofdstuk 12, eigenlijk een appendix, waarin we het gebruikelijke jargon met metrieken, omgevingen en open en gesloten verzamelingen samenvatten voor wie zich minder gelukkig voelt met informele uitdrukkingen als in de buurt van, zoals ook Adams die gebruikt in zijn nu vrijwel overal gebruikte calculus boek, dat door de theoretisch hoofdstukken in dit boek stevig wordt onderbouwd.

27 Welke horizon?

Kijk eens naar dit filmpje:

`https://www.youtube.com/watch?v=0l9hrFHNw3Q#t=133`

De formule van Pythagoras wordt toegepast op een rechthoekige driehoek met schuine zijde $R + l$, lange zijde R en korte zijde z. Ik heb van de r een R gemaakt, omdat R GROOT is ten opzichte van z en l. En z is weer groot ten opzichte van l. Dat schrijven we zo

$$R >> z >> l,$$

met dubbele $>>$ symbolen (die vaak een minder precies omschreven betekenis hebben dan de enkele symbolen, maar dat terzijde). Deze orde van grootte beschouwingen, die essentieel zijn voor de modelvorming en uitwerking blijven achterwege in het filmpje, net als de uitwerking zelf. De uiteindelijke vergelijking is

$$R^2 + z^2 = (R + l)^2,$$

met de lengte z van de zichtlijn als de onbekende. De straal R van de aarde en de ooghoogte l van het poppetje zijn als gegeven of op te zoeken te beschouwen.

Zoals gezegd, de uitwerking blijft achterwege. Wel wordt een samenvatting gegeven: eerst de straal R van de aarde opzoeken en dan wat algebraïsche manipulaties. Dat moet natuurlijk precies andersom, eerst de formule uitwerken en daarna de getallen invoeren als dat zin heeft of nodig is. Als we rechts het kwadraat uitwerken dan krijgen we

$$R^2 + z^2 = R^2 + 2Rl + l^2,$$

en gaat het verder met eenvoudige letteralgebra. We strepen links en rechts de gigantisch grote R^2 weg en houden een formule over die niet direct in verband gebracht kan worden met een meetkundige figuur, maar daarom niet minder is:

$$z^2 = 2Rl + l^2.$$

Is verder rekenen nu nodig? Ik zou zeggen nauwelijks. Rechts staan twee termen waarvan de eerste VEEL groter is dan de tweede, dus het kan niet anders dan dat bij benadering z de wortel uit $2Rl$ is. NU PAS is het moment daar om getallen in te vullen. Uit mijn hoofd is R ongeveer (en meer dan) 6 duizend kilometer en voor Jos is l ongeveer 2 meter. De dominante term rechts is dus ongeveer 2 keer 6 duizend kilometer keer 2 meter, en dat is 24

miljoen meter kwadraat, waarbij het vast handig is om van die 24 maar 25 te maken. Dan hebben 3 kwadraten op een rijtje in dit product en zien we dat z ongeveer 5 kilometer is. Nu, dat is te lopen.

In mijn tweeslagmodel kijken we nog even terug en wat preciezer. Die eenheden zijn misschien wat verwarrend maar ook in de laatste vergelijking zijn alle termen het product van twee lengten. De grootste daarvan is R en die kun je op 1 zetten door als fysische eenheid van lengte de EU te nemen, de zogenaamde Earth Unit. Als je dat meteen al gedaan had dan was de vergelijking voor z dus $z^2 = 2l + l^2$ geweest, met z en l in EU. Een beetje gevaarlijke formule die makkelijk een eigen leven gaat leiden waarbij de eenheden vergeten worden. Zelf geef ik er in dit voorbeeld nu de voorkeur aan om pas te schalen bij het uitwerken van de vergelijking voor z, maar dat is misschien een kwestie van smaak.

Omdat alles net zo positief als dit commentaar zelve kunnen we links en rechts de wortel nemen. Factorizerend komen we dan uit op

$$z = \sqrt{2Rl + l^2} = \sqrt{2Rl}\sqrt{1 + \frac{l}{2R}}.$$

De laatste stap is essentieel. Hier pas wordt echt duidelijk hoe volstrekt verwaarloosbaar de *dimensieloze* tweede term onder de tweede wortel is. Als we links en rechts door R delen zien we dat

$$\frac{z}{R} = \sqrt{\frac{2l}{R}}\sqrt{1 + \frac{l}{2R}},$$

een formule waarin zowel z als l met R geschaald zijn. Noemen we de geschaalde grootheden ζ en λ, dan is

$$\zeta = \sqrt{2\lambda}\sqrt{1 + \frac{1}{2}\lambda}$$

De vervolgvraag die betrekking op het (hoeveel) groter zijn van z bij grotere l heeft weinig te maken met logisch redeneren en wiskundige denkactiviteiten maar des te meer met het verontachtzaamde rekenen en de algebra van de formules. For all practical purposes is de tweede wortel niet te onderscheiden van 1. Maken we l (en dus λ) vier keer zo klein dan zien we dat papa Jos maar twee keer zo ver kijkt als de dreumes die naast hem staat. Dat hadden we natuurlijk al gezien aan de eerste benadering waarin z als ongeveer $\sqrt{2Rl}$ tevoorschijn kwam.

Hier ligt natuurlijk een leuk opstapje voor echte calculus met reeksontwikkelingen. Het belang bijvoorbeeld van het binomium van Newton

$$(1 + x)^\alpha = 1 + \alpha x + \frac{\alpha(\alpha - 1)}{2}x^2 + \frac{\alpha(\alpha - 1)(\alpha - 2)}{3 \cdot 2}x^3 + \cdots$$

met α niet per se geheel en positief en $|x| < 1$. Voor $\alpha = \frac{1}{2}$ geeft afkappen na twee termen een methode om de wortel uit 1 plus een klein beetje de bepalen die ten behoeve van de benadering van $\sqrt{2}$ al 2000 jaar B.C. werd gebruikt getuige[1] kleitablet YBC7289: de ontwikkeling

$$\zeta = 2^{\frac{1}{2}}\lambda^{\frac{1}{2}} + 2^{-\frac{3}{2}}\lambda^{\frac{3}{2}} + \cdots$$

is met de hand goed te doen en laat bij invullen van getallen zien hoe klein de fout is in de eerste benadering hierboven.

[1] Zo leerde ik uit het boek van Eves over de geschiedenis van de wiskunde.