OTTO DIEKA VAN

OVER NIET-LINEAIRE INTERNATIONALE INTERNATIONALE ONDERNEMINGEN

UvA

OVER NIET-LINEAIRE INTEGRAALVERGELIJKINGEN EN MATHEMATISCHE EPIDEMIOLOGIE

TW

OVER NIET-LINEAIRE INTEGRAALVERGELIJKINGEN EN MATHEMATISCHE EPIDEMIOLOGIE

ACADEMISCH PROEFSCHRIFT

TER VERKRIJGING VAN DE GRAAD VAN
DOCTOR IN DE WISKUNDE EN NATUURWETENSCHAPPEN
AAN DE UNIVERSITEIT VAN AMSTERDAM
OP GEZAG VAN DE RECTOR MAGNIFICUS
DR. G. DEN BOEF
HOOGLERAAR IN DE FACULTEIT
DER WISKUNDE EN NATUURWETENSCHAPPEN
IN HET OPENBAAR TE VERDEDIGEN
IN DE AULA DER UNIVERSITEIT
(TIJDELIJK IN DE LUTHERSE KERK, INGANG SINGEL 411, HOEK SPUI)
OP WOENSDAG 18 OKTOBER 1978 DES NAMIDDAGS TE 13.30 UUR

DOOR

ODO DIEKMANN

GEBOREN TE DIEPENVEEN

1978
MATHEMATISCH CENTRUM, AMSTERDAM

BIBLIOTHEEK MATHEMATISCH CENTRUM
AMSTERDAM

1703.399

PROMOTOR : PROF. DR. H.A. LAUWERIER

COPROMOTOR: PROF. DR. IR. L.A. PELETIER

COREFERENT: PROF. DR. N. CHAFEE

Aan mijn ouders

INHOUD 2

VOORWOORD 22

INLEIDING 1

1. Hoe een probleem ontstond 2
2. Volterra integraalvergelijkingen en dynamische systemen 13
3. Verspreiding in ruimte en tijd 21
4. Model en realiteit 27

VIER ARTIKELEN

[A] Limiting behaviour in an epidemic model 39
[B] Thresholds and travelling waves for the geographical spread of infection 53
[C] On the bounded solutions of a nonlinear convolution equation 85
[D] Run for your life. A note on the asymptotic speed of propagation of an epidemic. 113

SAMENVATTING 135

VOORWOORD

Dit proefschrift bestaat uit de vier artikelen

[A] *Limiting behaviour in an epidemic model, Nonl. Anal.* Th. Meth. Appl. 1 (1977) 459-470.

[B] *Thresholds and travelling waves for the geographical spread of infection,* J. Math. Biol. 1978.

[C] met H.G. KAPER, *On the bounded solutions of a nonlinear convolution equation,* Nonl. Anal. Th. Meth. Appl. 1978.

[D] *Run for your life. A note on the asymptotic speed of propagation of an epidemic,* voor publikatie aangeboden.

voorafgegaan door een inleiding. De artikelen zijn geschreven in het Engels en in een onpersoonlijke wetenschappelijke stijl. Ik heb geprobeerd om in de inleiding, die geschreven is in het Nederlands, op een informele en persoonlijke manier een aantal achtergronden van de artikelen te belichten.

Als ik, achterom kijkend, de weg overzie die leidde naar dit proefschrift, dan blijkt dat mijn promotor, Prof.dr. H.A. Lauwerier, op beslissende momenten een grote invloed heeft gehad. Door zijn colleges aan de Universiteit van Amsterdam (en door die van Prof.dr. E.M. de Jager) werd mijn belangstelling voor toegepaste wiskunde niet alleen gewekt, maar zelfs zo aangewakkerd dat ik besloot om "om te zwaaien" van theoretische natuurkunde naar wiskunde. Via hem werd ik eerst kandidaatsassistent en later wetenschappelijk medewerker op het Mathematisch Centrum. En bovendien was hij het die mij aanraadde om te gaan werken aan een bepaalde integraalvergelijking uit de mathematische epidemiologie.

Elke keer dat ik dreigde de moed te verliezen, omdat ik al te lang met mijn handen in het haar had gezeten, stond mijn copromotor, Prof.dr.ir. L.A. Peletier, klaar om mij verder te helpen met suggesties en adviezen. De vele gesprekken die ik met hem gevoerd heb zijn van grote invloed geweest op de ontwikkeling van mijn belangstelling en op mijn stijl van werken. Zijn kritische kanttekeningen bij eerste versies van de artikelen waren steeds opbouwend en op de inhoudelijke kern gericht.

Ik ben blij dat Prof.dr. N. Chafee erin heeft toegestemd om coreferent te zijn.

Aan de wieg van dit proefschrift staat Drs. J.A.J. Metz, want hij is

de geestelijke vader van het probleem waar het allemaal mee begon. Bovendien is zijn bezielende visie vaak mijn leidraad geweest bij verkenningen in het raakgebied van wiskunde en biologie.

Van samenwerken gaat een sterke stimulerende kracht uit. Dat althans was mijn ervaring tijdens de samenwerking met Dr. H.G. Kaper.

De direktie van het Mathematisch Centrum bood mij de gelegenheid om dit onderzoek uit te voeren. De kameraadschappelijke samenwerking met de leden van de afdeling Toegepaste Wiskunde heeft het werken op het MC voor mij steeds tot een plezier gemaakt.

De werkgroepen Biomathematica en Niet-lineaire Analyse hebben zowel als inspiratiebron alsook als klankbord gefungeerd. Bij de leden van deze werkgroepen vond ik vaak de zo onontbeerlijke belangstelling voor mijn werk.

Het moeilijke typewerk is voortreffelijk uitgevoerd door Mevr. R.W.T. Riechelmann-Huis en haar medewerksters. Het druk- en bindwerk werd verzorgd door D. Zwarst en zijn medewerkers, en R.T. Baanders ontwierp de omslag en hij maakte de tekeningen in de inleiding.

Alle bovengenoemde personen zeg ik bij deze hartelijk dank voor de wijze waarop zij aan de totstandkoming van dit proefschrift hebben bijgedragen.

INLEIDING

1. Hoe een probleem ontstond
2. Volterra integraalvergelijkingen en dynamische systemen
3. Verspreiding in ruimte en tijd
4. Model en realiteit

1. HOE EEN PROBLEEM ONTSTOND

Bijna elk deelgebied van wetenschappelijke activiteit kent zijn "klassieken", een aantal oudere artikelen dat steeds opnieuw als inspiratiebron voor verder onderzoek blijkt te functioneren. Soms kan het zelfs gewoonte zijn om naar zo'n artikel te verwijzen, terwijl toch velen die dat doen het betreffende artikel nooit zelf gelezen hebben.

Op het vakgebied van de wiskundige epidemiológie behoort het artikel *A contribution to the mathematical theory of epidemics* [28] tot deze categorie. Het werd in 1927 in de Proceedings of the Royal Society (London) gepubliceerd. De auteurs zijn twee Schotten, W.O. KERMACK, een Professor in de Biochemie in Aberdeen, en A.G. McKENDRICK die op dat moment werkzaam was in het Laboratory of the Royal College of Physicians in Edinburgh, maar die daarvoor lange tijd Luitenant-Kolonel bij de medische dienst van het leger in India was geweest. KENDALL [27] zegt dan ook:

> "*Mathematicians may be blamed for subsequently carrying the game too far, but its highly respectable medical origin should not be overlooked*".

In hun artikel bouwen Kermack en McKendrick een deterministisch model voor de verspreiding (als een functie van de tijd) van een besmettelijke ziekte in een populatie van mensen, dieren of planten. Dit model blijkt te leiden tot een niet-lineaire Volterra integraalvergelijking, die zij vervolgens analyseren. Al doende geven zij steeds de medisch-biologische interpretatie van hun resultaten.

In het volgende zal ik eerst het model toelichten en de integraalvergelijking afleiden. Daarna zal ik mijn versie geven van hun belangrijkste resultaat: "the celebrated threshold theorem" (KENDALL [25]).

Voor een gedetailleerde en kritische beschrijving van de aannamen die aan het model ten grondslag liggen verwijs ik naar REDDINGIUS [42] en METZ [36]. Hier volsta ik met een informele beschrijving.

Beschouw een populatie die verdeeld is in twee deelgroepen S en I. We zullen de dichtheden van de individuen die tot een bepaalde groep behoren aangeven met dezelfde letters. We beschouwen deze dichtheden als functies van de tijd en wel als "gladde" functies (een aanname die duidelijk onrealistisch is als we een kleine populatie van bijvoorbeeld tien individuen beschouwen).

De groep S (susceptibles) bestaat uit degenen die vatbaar zijn voor een

bepaalde besmettelijke ziekte en de groep I (infectives) bestaat uit degenen die besmet werden. De groep I geven we een interne structuur door de groepsleeftijd bij te houden. Dat wil zeggen dat we de individuen van groep I onderscheiden op grond van de tijd die er verstreken is sinds zij besmet werden (tot I gingen behoren). We geven met $i(t,\tau)$ de dichtheid op tijdstip t aan van leden van I die groepsleeftijd τ hebben, zodat

$$I(t) = \int_0^\infty i(t,\tau)d\tau.$$

We nemen nu aan dat:

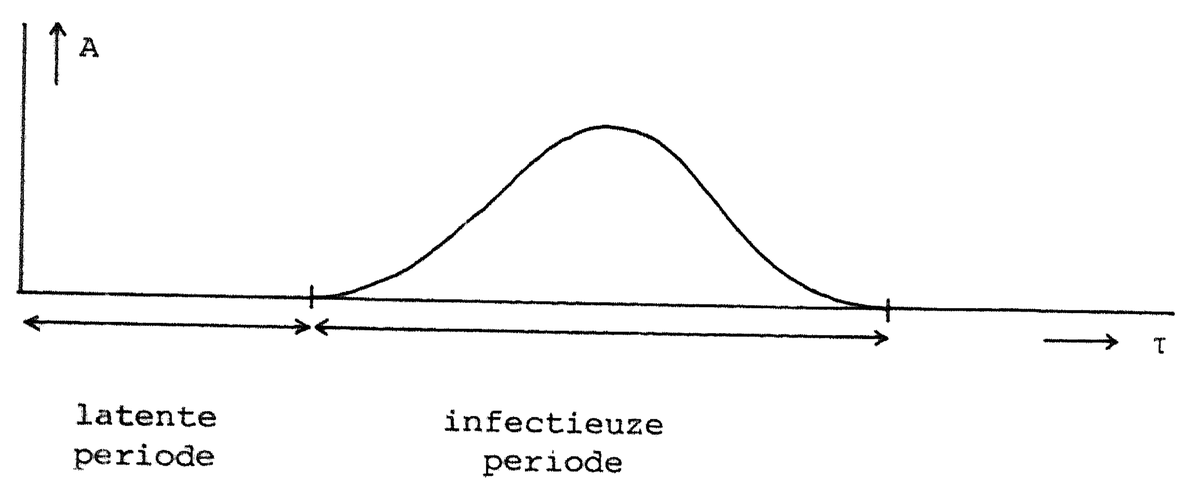
(i) de ziekte leidt tot permanente immuniteit, zodat de overgang van I naar S niet voorkomt;

(ii) alle veranderingen in de dichtheden zijn een gevolg van besmetting ($S(t) + I(t)$ = constant, de populatie is demografisch gesloten; we verwaarlozen veranderingen ten gevolge van geboorten, migratie enz.);

(iii) er is een niet-negatieve functie A gegeven zó dat

$$\dot{S}(t) = -S(t) \int_0^\infty i(t,\tau)A(\tau)d\tau. \tag{1.1}$$

Met betrekking tot (iii) merken we op dat de interaktie tussen de groepen S en I kennelijk op dezelfde wijze beschreven wordt als bij de zogeheten "law of mass action" voor chemische reakties, met dien verstande dat de factor I vervangen is door een gewogen gemiddelde over de leeftijds-gestructureerde dichtheid. De functie A is karakteristiek voor de ziekte (zie het voorbeeld in Figuur 1).



Figuur 1.

De interpretatie die men aan immuun geeft kan variëren van hersteld en niet meer vatbaar, tot in quarantaine tot dood.

Op grond van deze aannamen geldt

$$\frac{\partial i}{\partial t} + \frac{\partial i}{\partial \tau} = 0,$$

zodat

$$i(t,\tau) = i(t-\tau,0)$$

en

$$\dot{S}(t) = -\dot{I}(t) = -i(t,0).$$

Om het model te voltooien nemen we aan dat al deze relaties gelden voor $t > 0$ en dat we op $t = 0$ alle informatie bezitten die nodig is om op een goed gesteld probleem uit te komen. Het komt er op neer dat we stellen

$$S(0) = S_0$$

$$i(0,\tau) = i_0(\tau), \quad 0 \le \tau < \infty,$$

waarbij S_0 een gegeven getal en i_0 een gegeven functie is. We herschrijven nu (1.1):

$$\begin{aligned}\dot{S}(t) &= -S(t)\left\{\int_0^t i(t-\tau,0)A(\tau)d\tau + \int_t^\infty i(0,\tau-t)A(\tau)d\tau\right\} \\ &= S(t)\left\{\int_0^t \dot{S}(t-\tau)A(\tau)d\tau - \int_0^\infty i_0(\tau)A(t+\tau)d\tau\right\}.\end{aligned}$$

Delen we links en rechts door $S(t)$ en integreren we vervolgens van 0 tot t dan is het resultaat

$$\ln\frac{S(t)}{S_0} = \int_0^t (S(t-\tau) - S_0)A(\tau)d\tau - f(t),$$

waarbij per definitie

$$f(t) = \int_0^t \int_0^\infty i_0(\tau)A(s+\tau)d\tau\, ds.$$

De functie f is dus monotoon niet-dalend. Na invoering van de nieuwe afhankelijke variabele

$$u(t) = -\ln \frac{S(t)}{S_0}$$

kunnen we deze vergelijking in de standaardvorm van een niet-lineaire Volterra integraalvergelijking schrijven

$$(1.2) \qquad u(t) = S_0 \int_0^t g(u(t-\tau))A(\tau)d\tau + f(t),$$

waarbij

$$g(u) = 1 - e^{-u}.$$

We zijn er dus in geslaagd om, uitgaande van aannamen voor de interaktie tussen de verschillende groepen, te komen tot een vergelijking in slechts één afhankelijke variabele u, die een maat is voor de dichtheid van de vatbaren. Dit was mogelijk omdat we het aantal besmettingsgevallen gelijk konden stellen aan de afname van de dichtheid van de vatbaren. Wiskundig bezien hoeven we kennelijk van de besmetten geen aparte boekhouding bij te houden.

Het is een koud kunstje om existentie en uniciteit van een oplossing van (1.2) te bewijzen onder vrij zwakke voorwaarden voor A en f. De niet-negativiteit van A en f zorgt er voor dat ook de oplossing deze eigenschap heeft, en bovendien wordt de monotonie van f door de oplossing overgenomen (deze eigenschappen zijn ook noodzakelijk vanwege de biologische interpretatie; zo is de monotonie een weerspiegeling van onze aannamen dat de populatie demografisch gesloten is en dat de ziekte permanente immuniteit induceert). Nemen we tenslotte aan dat

$$\gamma := \int_0^\infty A(\tau)d\tau < \infty$$

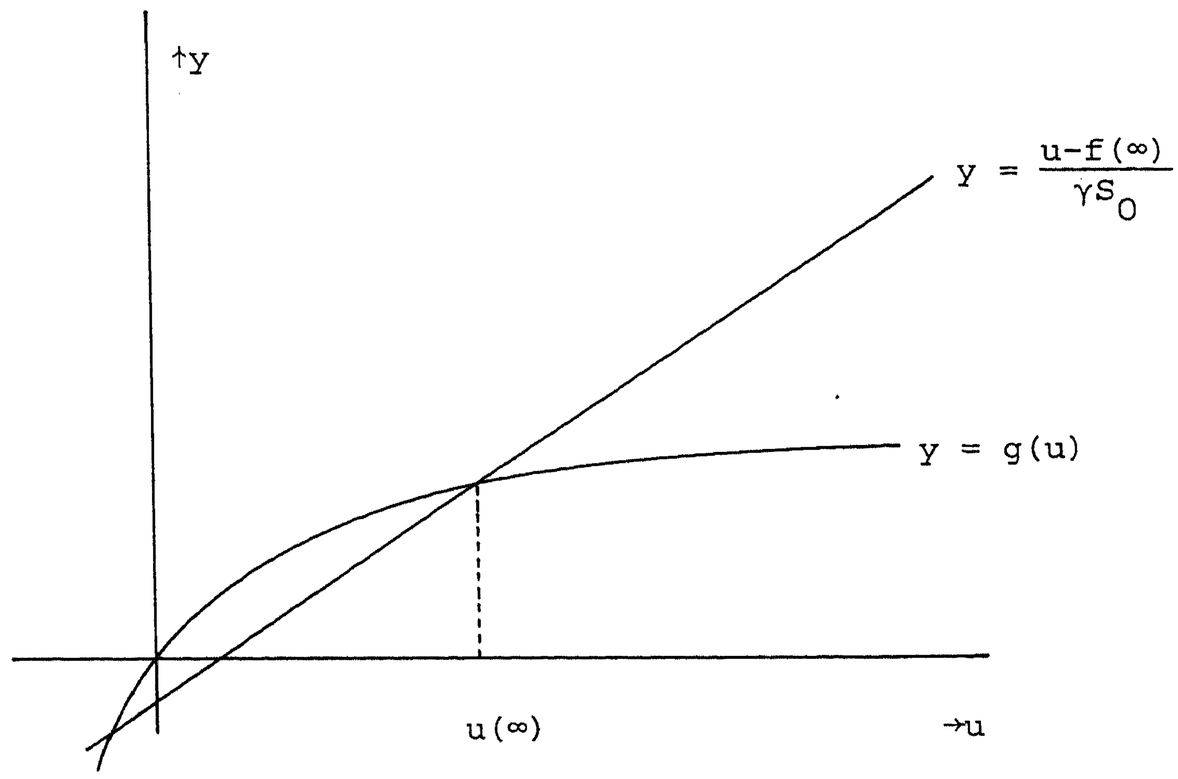
en

$$f(\infty) < \infty$$

dan kan bewezen worden dat de oplossing begrensd is en dus dat $u(\infty)$ bestaat, en voldoet aan de scalaire vergelijking

$$(1.3) \qquad u(\infty) = \gamma S_0 \, g(u(\infty)) + f(\infty).$$

Het is de bedoeling dat het in Figuur 2 getekende plaatje de lezer er van overtuigt dat de vergelijking (1.3) voor iedere positieve $f(\infty)$ een



Figuur 2.

eenduidig bepaalde positieve oplossing $u(\infty)$ bezit. Uit het feit dat $u(\infty) < \infty$ trekken Kermack en McKendrick de volgende conclusie:

> *"An epidemic, in general, comes to an end, before the susceptible population has been exhausted".*

Vervolgens vragen zij zich af wat het gevolg is van de introductie van een zeer gering aantal besmetten in een tot dan toe onbesmette populatie. In wiskundige termen komt dit neer op de vraag: wat gebeurt er met $u(\infty)$ als we $f(\infty)$ naar nul laten naderen?

Indien $f(\infty)$ naar nul nadert dan schuift de rechte lijn $y = (\gamma S_0)^{-1}(u-f(\infty))$ in Figuur 2 evenwijdig aan zichzelf omhoog tot hij door de oorsprong gaat. We zien hieruit dat $u(\infty)$ monotoon stijgend is als functie van $f(\infty)$. Bovendien blijkt direkt dat de verhouding tussen de richtingscoëfficiënt $(\gamma S_0)^{-1}$ van de lijn en de richtingscoëfficiënt 1 van de raaklijn aan $y = g(u)$ in de oorsprong, van cruciaal belang is voor het al dan niet nul zijn van $\lim_{f(\infty)\downarrow 0} u(\infty)$. Indien $\gamma S_0 > 1$ dan heeft de scalaire vergelijking

$$y = \gamma S_0\, g(y)$$

een positieve wortel, die we p zullen noemen, en er geldt

$$\lim_{f(\infty)\downarrow 0} u(\infty) = p.$$

(Merk op dat in [A] deze wortel x_∞ wordt genoemd in plaats van p.) Als daarentegen $\gamma S_0 \leq 1$ dan geldt

$$\lim_{f(\infty)\downarrow 0} u(\infty) = 0$$

en als $\gamma S_0 < 1$ dan zelfs

$$u(\infty) \leq (1 - \gamma S_0)^{-1} f(\infty).$$

We zien hieruit dat het kwalitatieve gedrag van de oplossing sterk afhankelijk is van de waarde van de parameter γS_0. Men zegt wel dat 1 een drempelwaarde (threshold value) voor deze parameter is, en de bovenstaande opmerkingen vormen een wiskundige formulering van het Threshold Theorem. Een meer biologisch getinte formulering is de volgende: Als $\gamma S_0 > 1$ dan geldt dat er voor de fractie van de vatbare populatie die aan de ziekte ontsnapt een bovengrens is, die onafhankelijk van de hoeveelheid geïntroduceerde besmetten gekozen kan worden (namelijk $\frac{S(\infty)}{S_0} < e^{-p}$). Als $\gamma S_0 \leq 1$ dan bestaat er niet een dergelijke bovengrens kleiner dan één.

Er is een eenvoudige biologische interpretatie van de parameter γS_0 die dit resultaat onmiddellijk duidelijk maakt. Men kan namelijk γS_0 opvatten als het aantal besmettingsgevallen dat het gevolg is van de introduktie van één vers besmet individu in de populatie (zie ook HOPPENSTEADT [23]). Is dit getal groter dan één dan ontstaat er een lawine, is het evenwel kleiner dan één dan loopt het met een sisser af.

In dit verband is het ook aantrekkelijk om het Threshold Theorem te formuleren in termen van stabiliteit. De afbeelding $f(\infty) \longmapsto u(\infty)$ is continu in nul dan en slechts dan als $\gamma S_0 \leq 1$ en men zou dus ook kunnen zeggen dat de toestand waarin elk spoor van de ziekte in de populatie ontbreekt, stabiel is dan en slechts dan als $\gamma S_0 \leq 1$. Echter, een snel doorbladeren van R.K. MILLER's boek [37] zal de lezer overtuigen dat deze uitspraak niet direkt aansluit bij een eenvoudige standaard definitie van stabiliteit in de context van Volterra integraalvergelijkingen. Hierop kom ik in de volgende paragraaf nog terug.

In het voorafgaande heb ik een vrije vertaling gegeven van een belangrijk deel van het artikel van Kermack en McKendrick. In het artikel worden voorts nog benaderingen gegeven voor allerlei grootheden onder veronderstellingen dat sommige andere grootheden erg klein of juist erg groot zijn, er

wordt een ingewikkelder model bekeken en ér wordt verder ruime aandacht besteed aan het speciale geval dat

$$A(t) = \alpha e^{-\frac{\alpha}{\gamma}t},$$

waarbij α een willekeurige positieve parameter is. Het bijzondere van dit zogeheten "constant rates" geval is, dat het model dan blijkt te leiden tot een stelsel gewone differentiaalvergelijkingen. Schrijven we namelijk (1.1) als

$$\dot{S}(t) = -\alpha S(t)Y(t), \tag{1.4}$$

waarbij per definitie

$$Y(t) = -\frac{1}{\alpha}\left\{\int_0^t \dot{S}(\tau)A(t-\tau)d\tau - \int_0^\infty i_0(\tau)A(t+\tau)d\tau\right\},$$

dan volgt dat Y voldoet aan

$$\dot{Y}(t) = \alpha S(t)Y(t) - \frac{\alpha}{\gamma}Y(t). \tag{1.5}$$

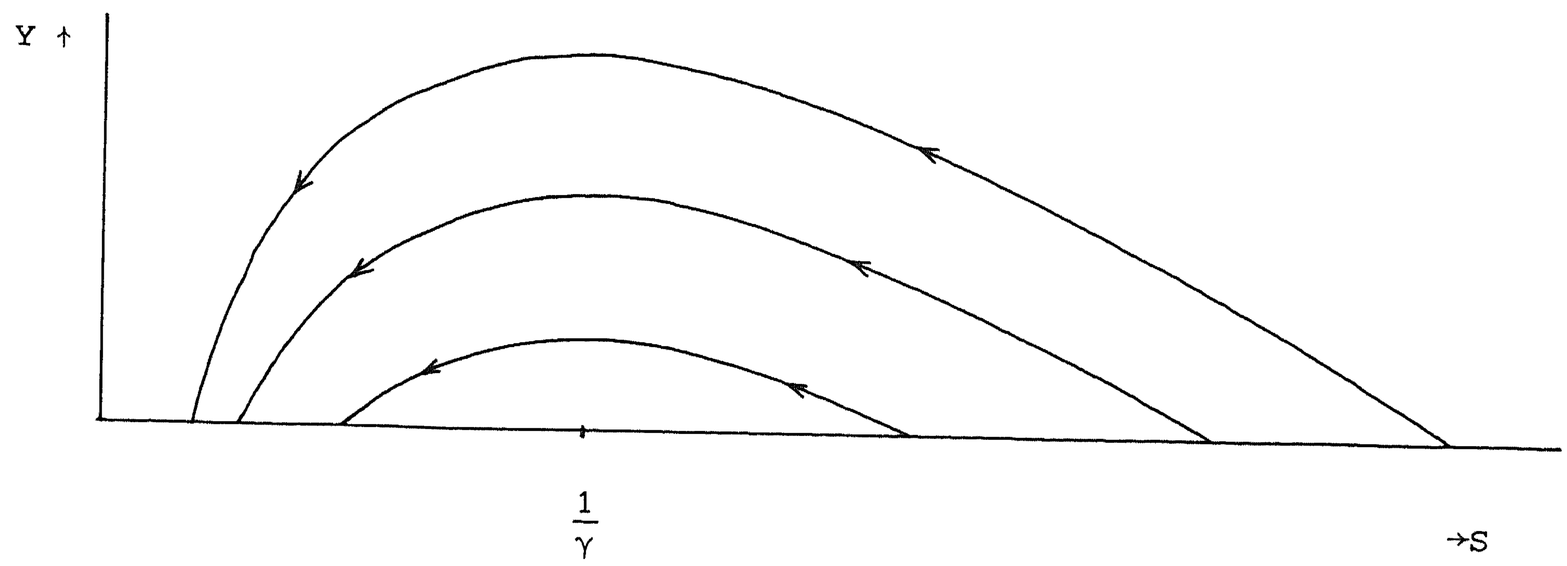
Omgekeerd kunnen we ook het stelsel (1.4) - (1.5), met daaraan beginvoorwaarden voor S en Y toegevoegd, in de vorm (1.2) brengen. Het stelsel laat de volgende alternatieve interpretatie toe: Y is de dichtheid van de besmetten en iedere besmette heeft een constante infectiviteit α, maar zij worden "verwijderd" met een snelheid die evenredig is aan hun eigen dichtheid met een evenredigheidsconstante α/γ.

Uit (1.4) en (1.5) volgt rechtstreeks dat het stelsel een eerste integraal bezit:

$$Y(t) + S(t) - \frac{1}{\gamma}\ln S(t) = \text{constant}. \tag{1.6}$$

In Figuur 3 zijn enige banen in het (S,Y)-vlak geschetst, en het Threshold Theorem kan uit deze figuur onmiddellijk afgelezen worden. Een stabiliteitsinterpretatie is nu ook direkt duidelijk: het stelsel heeft de S-as als een één-parameterschaar van evenwichtstoestanden en voor $S > \gamma^{-1}$ is er instabiliteit terwijl er voor $S < \gamma^{-1}$ stabiliteit is (maar niet asymptotisch!).

Op dit punt wil ik de aandacht vestigen op een raadselachtig feit. Er heerst een wijd en zijd verspreid misverstand dat het stelsel (1.4) - (1.5) <u>het</u> Kermack en McKendrick model zou zijn zoals zij dat in 1927 behandelen (zie bijvoorbeeld ARONSON [1], BAILEY [5], BRAUN [10], KENDALL [25],[27], MOLLISON [39] en HOPPENSTEADT [24]). Waarschijnlijk is het zo dat vele van



Figuur 3.

de latere auteurs hun kennis baseren op de artikelen van KENDALL [25],[27] zonder nog het oorspronkelijke werk te bestuderen. Kendall zelf wekt echter sterk de indruk dat hij het oorspronkelijke artikel wel gelezen heeft. In dat geval zou hij dus om de een of andere reden het algemene model en de Volterra integraalvergelijking voor zijn lezers verzwegen hebben!

Tot zover deze opsomming van oude resultaten gestoken in het nieuwe jasje dat speciaal op maat gesneden werd om de nu volgende vraagstelling te verduidelijken.

In het academisch jaar 1973-1974 bracht J.A.J. Metz (Instituut voor Theoretische Biologie, Leiden) tijdens één van de bijeenkomsten van de Werkgroep Biomathematica van het Mathematisch Centrum het volgende probleem naar voren.

Beschouw de oplossing van (1.2) voor een vaste waarde van t als functie van de "aandrijvende" term f, die we voor het gemak met $f(\infty)$ blijven meten, dan geldt

$$\lim_{f(\infty)\downarrow 0} u(t) = 0$$

op grond van de continue afhankelijkheid van data voor dit soort vergelijkingen (het bewijs is eenvoudig te geven met behulp van een lineaire afschatter voor g en een expliciete voorstelling voor de oplossing van lineaire Volterra vergelijkingen). Indien $\gamma S_0 > 1$ dan geldt dus dat de limieten $t \to \infty$ en $f(\infty) \downarrow 0$ niet verwisseld mogen worden:

$$p = \lim_{f(\infty)\downarrow 0} \lim_{t\to\infty} u(t) > \lim_{t\to\infty} \lim_{f(\infty)\downarrow 0} u(t) = 0.$$

Met andere woorden, er is hier sprake van niet-uniforme convergentie. Is het mogelijk een niet-triviale limiet te vinden die de limieten 0 en p aan elkaar plakt?

Op dit probleem was Metz gestuit bij zijn pogingen om een stochastisch model voor de beginfase van de epidemie (immers, dan zijn er slechts weinig besmetten zodat een deterministische beschrijving aanvechtbaar is), vast te plakken aan een deterministisch model voor de latere fasen (zie [36]). Hij had ook al een vermoeden (dat later juist is gebleken) hoe zo'n limiet geconstrueerd zou kunnen worden.

Het komt zelden voor dat een (theoretisch) bioloog een probleem heeft dat een dergelijke preciese wiskundige formulering toelaat, en er werd dan ook door J. Grasman, de singuliere storingsdeskundige van de werkgroep, enthousiast op gereageerd. Zijn aanpak kwam neer op de constructie van formele reeksontwikkelingen (in de parameter $f(\infty)$) voor de oplossing in verschillende ($f(\infty)$-afhankelijke) gebieden van de t-as (zie GRASMAN & MATKOWSKY [18] voor de uiteindelijke resultaten). Op 17 januari 1974 hield hij een werkbespreking over dit onderzoek en bij deze gelegenheid raakte Professor Lauwerier geïnteresseerd. In een ongepubliceerd manuscript gaf Lauwerier vervolgens zijn visie op het probleem, waarbij hij vooral gebruik maakte van Laplace transformatie (zie ook [30]).

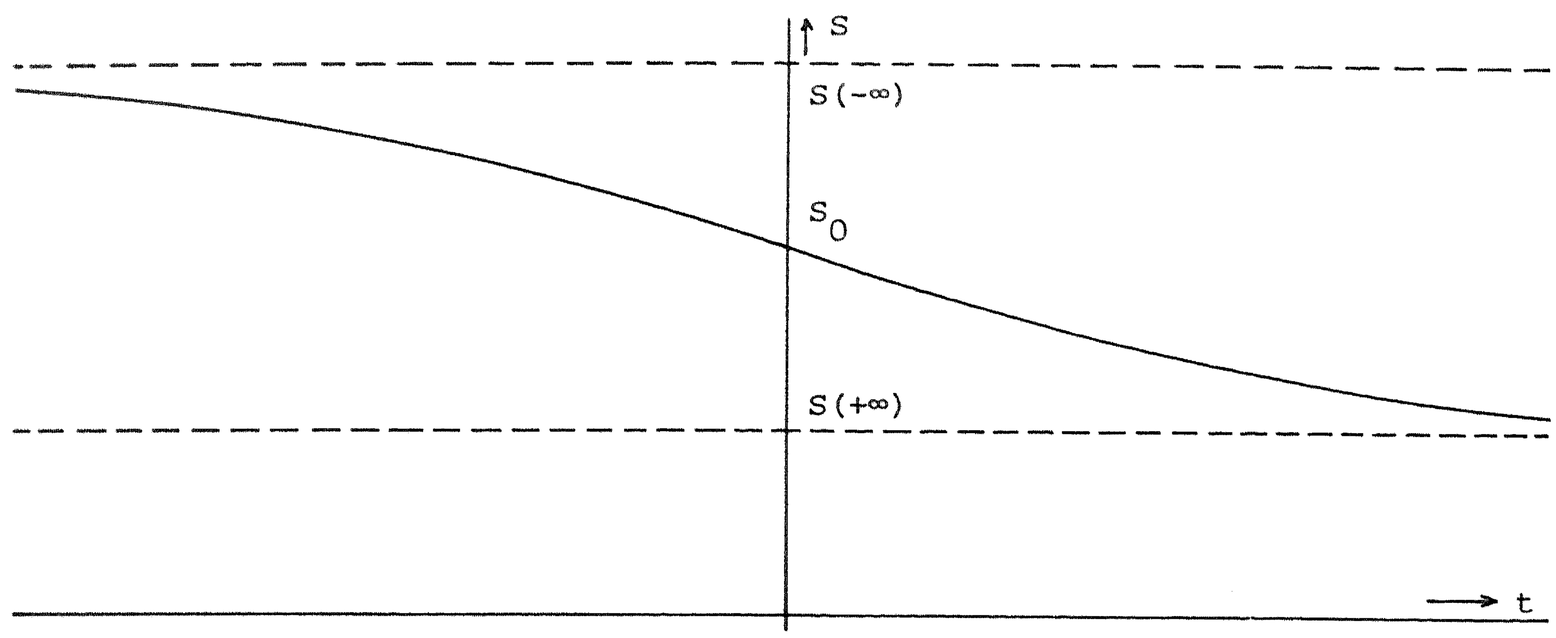
Zowel Grasman als Lauwerier waren bij hun analyse via een limietovergang gestuit op de homogene vergelijking

$$(1.7) \qquad w(t) = S_0 \int_0^\infty g(w(t-\tau))A(\tau)d\tau, \qquad -\infty < t < \infty.$$

Lauwerier bewees (met behulp van monotone iteratie) dat deze vergelijking een monotoon stijgende oplossing heeft met $w(-\infty) = 0$ en $w(\infty) = p$ en hij sprak het vermoeden uit dat deze uniek was modulo translatie (merk op dat de vergelijking translatie-invariant is). In september 1974 suggereerde Lauwerier mij om te proberen een bewijs te vinden van de juistheid van dit vermoeden. Het artikel [A] bevat de resultaten van het onderzoek dat uit deze suggestie voortkwam. In het nu volgende zal ik proberen om op een informele manier duidelijk te maken hoe men mijns inziens tegen het probleem aan moet kijken.

Voor een goed begrip is het verhelderend om een speciaal geval te bekijken. Met een baan van het stelsel (1.4) - (1.5) in het (S,Y)-vlak, zoals getekend in Figuur 3, correspondeert een oplossing die gedefinieerd is voor $-\infty < t < \infty$. Indien we een vaste waarde kiezen voor de eerste integraal (1.6) dan is daarmee deze oplossing eenduidig bepaald op translatie van de tijds-

variabele na. We kunnen de oplossing dan dus eenduidig vastleggen door de waarde S_0 van S voor t = 0 te kiezen en we brengen dit in de notatie tot uiting door te schrijven $S = S(t;S_0)$.



Figuur 4.

Veranderingen in S_0 komen overeen met verschuivingen in t. In formulevorm wordt dit uitgedrukt door de semigroep eigenschap

$$S(t;S(t^*,S_0)) = S(t+t^*;S_0). \tag{1.8}$$

In dit voorbeeld vinden we de niet-uniforme convergentie op de volgende manier terug: voor iedere eindige t geldt

$$\lim_{S_0 \uparrow S(-\infty)} S(t;S_0) = S(-\infty),$$

terwijl voor iedere vaste $S_0 \in (S(-\infty),S(+\infty))$

$$\lim_{t\to\infty} S(t;S_0) = S(\infty).$$

De limieten $S_0 \uparrow S(-\infty)$ en $t \to \infty$ zijn dus niet verwisselbaar. Maar we zien in dit voorbeeld heel precies wat er aan de hand is: als $S_0 \uparrow S(-\infty)$ dan schuift de kromme naar rechts om als het ware in het oneindige te verdwijnen. Bovendien is het duidelijk dat we de niet-triviale limiet die $S(-\infty)$ aan $S(\infty)$ vastplakt kunnen vinden door gewoon net zo hard terug te schuiven.

In dit voorbeeld is alles triviaal omdat het in feite steeds om dezelfde functie gaat en omdat (1.8) ons precies vertelt hoe het schuiven in z'n werk gaat. We zullen nu onze aanname dat de waarde van de eerste integraal

een vast getal is laten vallen. Als we de beginvoorwaarden die we aan (1.4) - (1.5) toevoegen op de een of andere manier naar één van de instabiele stationaire punten op de S-as laten naderen, dan zal de baan in het (S,Y)-vlak naderen naar de baan die uit dat punt vertrekt (de instabiele variëteit). Als een functie van t krijgen we in eerste instantie als limieten alleen het begin- en het eindpunt van die baan (niet-uniforme convergentie), maar door de limiet te nemen na uitvoering van een handig gekozen, van de beginvoorwaarden afhankelijke, verschuiving van de tijdsas, krijgen we ook het ontbrekende tussenstuk. De uniciteitsuitspraak die al in deze beweringen ligt opgesloten is in feite een uitspraak over de dimensie van de instabiele variëteit (namelijk dat deze één is).

We zien dus dat de analyse van het probleem voor stelsels gewone differentiaalvergelijkingen eenvoudig is omdat we gebruik kunnen maken van het feit dat deze vergelijkingen een dynamisch systeem definiëren en omdat we stellingen kennen over het bestaan en de dimensie van invariante variëteiten in de omgeving van een evenwichtspunt. Voor Volterra integraalvergelijkingen is de situatie ingewikkelder. Ten eerste is de wijze waarop deze vergelijkingen een dynamisch systeem definiëren essentieel gecompliceerder en ten tweede is er voor deze vergelijkingen geen standaard theorie die analoge resultaten over invariante variëteiten omvat. In zekere zin zou men kunnen zeggen dat deze variëteiten corresponderen met oplossingen van (1.7), en men kan dan ook zeggen dat in [A] deze problemen voor dit speciale geval overwonnen werden op een ad hoc manier. In de volgende paragraaf zal ik een eerste aanzet geven tot een veel fundamentelere aanpak van de problematiek.

2. VOLTERRA INTEGRAALVERGELIJKINGEN EN DYNAMISCHE SYSTEMEN

In de populatiedynamica is het eerder regel dan uitzondering dat de toekomst mede bepaald wordt door het recente verleden. Er zijn dan ook veel modellen die leiden tot functionaal-differentiaalvergelijkingen (van "retarded" type), Volterra integraalvergelijkingen of Volterra integrodifferentiaalvergelijkingen (zie bijvoorbeeld MAY [33], HOPPENSTEADT [24], MAYNARD SMITH [34],[35] en CUSHING [13]). Met name Volterra zelf heeft veel van dit soort modellen opgesteld en geanalyseerd en men kan zonder overdrijving zeggen dat zijn interesse in populatiedynamica één van de drijfveren is geweest bij zijn studie van de later naar hem genoemde vergelijkingen (zie bijvoorbeeld [44]). Iemand die vanuit deze achtergrond belangstelling heeft gekregen voor dit soort vergelijkingen en die een indruk wil krijgen van de bestaande theorie, zal waarschijnlijk met twee boeken uit de bibliotheek komen, namelijk met *Theory of functional differential equations* [20] van J.K. HALE en met *Nonlinear Volterra integral equations* [37] van R.K. MILLER. Als hij of zij ze dan achtereenvolgens doorbladert zal hij of zij tot zijn of haar verbazing ontdekken dat er betrekkelijk weinig overeenkomst is.

In Hale's boek wordt een kwalitatieve theorie ontwikkeld die een grote mate van overeenkomst vertoont met die van gewone differentiaalvergelijkingen: variatie-van-constanten formule, exponentiële schattingen, zadelpunteigenschap (invariante variëteiten) enz.. Het belangrijkste verschil met gewone differentiaalvergelijkingen is dat de onderliggende toestandsruimte oneindig dimensionaal is, waardoor het nodig wordt gebruik te maken van functionaalanalyse.

In Miller's boek wordt weliswaar een schat van informatie geboden maar de structuur is veel minder doorzichtig. Een vergelijkbare kwalitatieve theorie ontbreekt.

Ik denk dat dit verschil gedeeltelijk terug te voeren is op de wijze waarop men met de vergelijkingen een dynamisch systeem associeert. Bovendien moet het volgens mij mogelijk zijn om voor een ruime klasse van Volterra integraalvergelijkingen dit op zo'n manier te doen dat men vervolgens op de standaard manier een kwalitatieve theorie op kan bouwen.

Om deze bewering toe te lichten wil ik het allereenvoudigste voorbeeld bekijken, namelijk één enkele lineaire vergelijking van convolutie type. Er zijn twee alternatieve manieren om zo'n vergelijking voor het voetlicht te brengen, namelijk

$$(2.1) \qquad x(t) = \int_0^t x(\tau)B(t-\tau)d\tau + f(t)$$

en

$$(2.2) \qquad x(t) = \int_{t-b}^t x(\tau)B(t-\tau)d\tau.$$

Ik hoop dat de discussie van het epidemiemodel de lezer een idee heeft gegeven van de manier waarop dit soort vergelijkingen ontstaat bij de beschrijving van biologische verschijnselen waarbij men rekening moet houden met de leeftijdsstruktuur van een populatie (zie HOPPENSTEADT [24] en GURTIN & MacCAMY [19] voor andere voorbeelden).

De standaard schrijfwijze is (2.1). Hierbij beschrijft de gegeven functie f een combinatie van de direkte invloed van het verleden en van de invloed van mogelijke *externe* storingen. Ons uitgangspunt is dat existentie en eenduidigheid van een (locale) oplossing van (2.1) verzekerd is (voor ons doel heeft het op dit moment geen zin om hierover precieser te zijn).

Bij de schrijfwijze (2.2) wordt benadrukt dat $x(t)$ berekend kan worden zodra we $x(\tau)$ kennen voor $t-b \le \tau \le t$. We nemen dus aan, zoals ook al in (2.2) expliciet tot uitdrukking wordt gebracht, dat de gegeven kern B een drager heeft die bevat is in $[0,b]$, met andere woorden dat slechts een eindig deel van het verleden de toekomst rechtstreeks beïnvloedt. De vergelijking (2.2) is niet volledig, dat wil zeggen, door (2.2) wordt geen eenduidige oplossing bepaald. Om in dit mankement te voorzien kunnen we een beginconditie toevoegen

$$(2.3) \qquad x(t) = \phi(t) \quad \text{voor} \quad -b \le t \le 0.$$

Met andere woorden, we veronderstellen ofwel dat we x kennen op een interval ter lengte b ofwel, in een experimentele situatie, dat we x kunnen kiezen (voorschrijven) in zo'n interval. Met deze beginconditie kunnen we (2.2) in de vorm (2.1) brengen met in dat geval

$$(2.4) \qquad f(t) = \begin{cases} \displaystyle\int_{t-b}^{0} \phi(\tau)B(t-\tau)d\tau & \text{voor } t \le b \\ 0 & \text{voor } t \ge b. \end{cases}$$

Ook dan is dus existentie en eenduidigheid van een oplossing verzekerd. Bij deze visie op de vergelijking is het duidelijk dat f alleen de direkte invloed van het stukje verleden met $-b \le t \le 0$ beschrijft. Bovendien is het

van belang op te merken dat de vergelijking (2.2) autonoom, dat wil zeggen, translatie-invariant is (vandaar ook dat het geen enkele beperking van de algemeenheid is, dat we in (2.3) het segment ter lengte b kozen dat eindigt in 0).

In principe kan men ook zoeken naar oplossingen van (2.2) die gedefinieerd zijn voor $-\infty < t < \infty$. Gezien de lineairiteit en de translatie-invariantie ligt het voor de hand om te proberen of misschien exponentiële functies aan de vergelijking kunnen voldoen. Substitutie van $x(t) = \exp \lambda t$ leidt tot de karakteristieke vergelijking

$$(2.5) \qquad \int_0^b e^{-\lambda\tau} B(\tau)d\tau = 1$$

en we concluderen dan ook dat wortels van (2.5) oplossingen van (2.2) genereren (zie ook [A,§3]).

Wat willen we bereiken? We willen op de een of andere manier met de vergelijking een sterk continue semigroep van begrensde lineaire operatoren op een Banach ruimte associëren zó dat het puntspectrum van de infinitesimale generator van die semigroep samenvalt met de wortels van de karakteristieke vergelijking (2.5). (Vergelijk de situatie bij een stelsel gewone differentiaalvergelijkingen $\frac{dx}{dt} = Ax$, $x(0) = x_0$; daar is de matrix A de infinitesimale generator van de semigroep $T(t)$ gedefinieerd door $T(t)x_0 = x(t)$ en de karakteristieke vergelijking is de eigenwaardenvergelijking $Ax = \lambda x$.) Voor de definities en een zeer leesbare inleiding tot de abstracte theorie zij verwezen naar LADAS & LAKSHMIKANTHAM [29]. Dat dit een aanpak is die veel vruchten kan afwerpen blijkt overduidelijk uit de toepassing op de functionaal-differentiaalvergelijkingen (zie HALE [20]) en op semilineaire parabolische differentiaalvergelijkingen (zie FRIEDMAN [17] maar vooral HENRY [22]).

Met het voorbeeld van de constructie voor functionaal-differentiaalvergelijkingen voor ogen proberen we het eerst op de volgende manier. We gaan uit van de formulering (2.2) - (2.3) en we kiezen als toestandsruimte de ruimte van continue functies op $[-b,0]$, voorzien van de supremum norm, die we aangeven met $C([-b,0])$. We definiëren de afbeeldingen $T(t)$ door

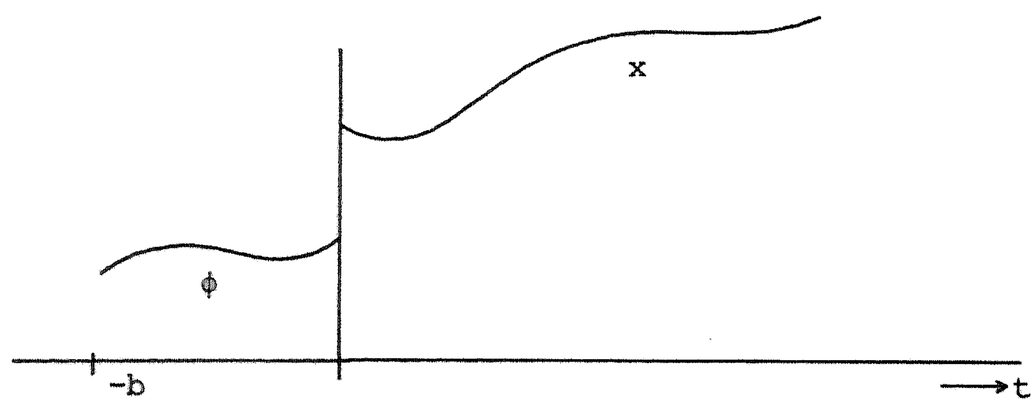
$$(2.6) \qquad (T(t)\phi)(\tau) = x(t+\tau), \quad -b \le \tau \le 0,$$

(dus $T(t)\phi$ is het stukje van de oplossing met argument tussen $t-b$ en t, maar dan teruggeschoven naar $[-b,0]$). Deze definitie is voor de hand liggend en verdedigbaar op grond van de interpretatie dat $T(t)\phi$ de

informatie bevat die nodig en voldoende is om de toekomst vanaf t te bepalen. Deze aanpak lijkt dus op het eerste gezicht een schot in de roos. Maar zodra we nagaan of T(t) wel aan de voorwaarden voldoet, stuiten we op moeilijkheden. Immers, de eerste vraag die we moeten stellen is: is T(t) een afbeelding van C([-b,0]) in zichzelf? En het antwoord is: nee, want in het algemeen is

$$\phi(0) \neq \lim_{t\downarrow 0} x(t) = \int_{-b}^{0} \phi(\tau)B(-\tau)d\tau$$

(in het algemeen is de beginvoorwaarde niet compatibel met de dynamica van het model!),



Figuur 5

zodat voor $0 < t < b$ kan gelden dat $T(t)\phi \notin C([-b,0])$. Waarom gaat hier iets mis, terwijl bij functionaal-differentiaalvergelijkingen een volkomen analoge aanpak geen enkel probleem oplevert? Beschouw daartoe even de vergelijking die uit (2.2) ontstaat door het linkerlid te vervangen door $\frac{dx}{dt}(t)$, en stel dat de beginfunctie ϕ differentieerbaar is in nul. Dan ontstaat er misschien een sprong in de afgeleide, maar dat levert geen enkel probleem op omdat we in de toestandsruimte C([-b,0]) werken!

Van dit "sprong-discontinuïteit"-probleem werd ik mij bewust bij mijn pogingen om vat te krijgen op de vergelijking (1.7). In de hierboven beschreven vorm werd het probleem echter geformuleerd door J.K. Hale tijdens een gesprek dat ik met hem had naar aanleiding van het artikel [A]. In dit gesprek stelde hij dat het de moeite waard was om te proberen dit struikelblok uit de weg te ruimen en bovendien had hij hiervoor een bepaalde strategie op het oog, namelijk om (2.6) te beschouwen op de variëteit in C([-b,0]) gedefinieerd door

$$\phi(0) = \int_{-b}^{0} \phi(\tau)B(-\tau)d\tau,$$

omdat immers voor $t > b$ door $T(t)$ de hele ruimte $C([-b,0])$ op deze variëteit wordt afgebeeld (vergelijk ook HALE & MARTINEZ-AMORES [21], waar een soortgelijke aanpak wordt gevolgd bij de analyse van functionaal-differentiaalvergelijkingen van "neutraal" type). In principe werkt dit inderdaad en men krijgt een infinitesimale generator met alleen maar puntspectrum dat samenvalt met de wortels van (2.5). Onze motivatie voor deze hele constructie is echter dat we vervolgens via de variatie-van-constanten formule niet-lineaire vergelijkingen willen analyseren. Dan krijgen we te maken met het probleem dat we iedere vergelijking op een *eigen* variëteit beschouwen en dit maakt het toepassen van storingstheorie tot allesbehalve een routine procedure. Verder dan tot deze constatering ben ik nooit gekomen.

Laten we nu even terugkeren naar de formulering (2.1). Vanuit vooral een gerichtheid op industriële toepassingen is er een uitgebreide, zeer algemene, theorie opgebouwd over de continuïteit van de afbeelding $f \longmapsto x$, waarbij dan f en x behoren tot ruimten van functies op $\mathbb{R}_+$ die van een geschikte topologie zijn voorzien (meestal Fréchet of Banach ruimten). Dit geeft dan vaak meteen enige informatie over het asymptotisch gedrag van de oplossing als $t \to \infty$ en men spreekt in deze context wel over input-output stabiliteit (zie naast MILLER [37] bijvoorbeeld ook CORDUNEANU [12] en WILLEMS [46]). Daarnaast is er een verwante, ook weer zeer algemene, topologisch dynamische aanpak van MILLER & SELL [38] (zie ook MILLER [37, Hoofdstuk III]), die vergelijkbaar is met die voor niet-autonome gewone differentiaalvergelijkingen (zie SELL [43]). Het idee is daarbij om eerst formeel na te gaan hoe de vergelijking transformeert onder translatie van de tijd en om op grond daarvan aannamen te doen en afbeeldingen te definieren. In ons eenvoudige voorbeeld werkt dat als volgt. Met de notatie $x_s(t) = x(t+s)$ schrijft men de identiteit

$$x(t+s) = \int_{0}^{t+s} x(\tau)B(t+s-\tau)d\tau + f(t+s)$$

als

$$x_s(t) = \int_{0}^{t} x_s(\tau)B(t-\tau)d\tau + f(t+s) + \int_{0}^{s} x(\tau)B(t+s-\tau)d\tau.$$

Hieruit blijkt dat men na translatie over een afstand s de vergelijking weer in dezelfde vorm kan schrijven door f te vervangen door $T(s)f$ gedefinieerd

door

$$(2.7)\qquad (T(s)f)(t) = f(t+s) + \int_0^s x(\tau)B(t+s-\tau)d\tau.$$

In feite is er nu nog een heleboel vrijheid in de keuze van een (translatie-invariante) ruimte van functies X waartoe we f laten behoren en waarop we T(s) laten werken. Maar laten we voorlopig formeel verder gaan en de infinitesimale generator A uitrekenen:

$$(Af)(t) = \lim_{s\downarrow 0} \frac{1}{s}[(T(s)f)(t) - f(t)] = \dot{f}(t) + f(0)B(t),$$

waarbij $\mathcal{D}(A)$, het definitiegebied van A, bestaat uit de verzameling van functies waarvoor de uitdrukking rechts bestaat en tot de ruimte X behoort waarin we werken. De eigenwaardenvergelijking $Af = \lambda f$ is dus de differentiaalvergelijking

$$\dot{f}(t) + f(0)B(t) = \lambda f(t),$$

met als oplossing

$$(2.8)\qquad f(t) = f(0)e^{\lambda t}\left(1 - \int_0^t e^{-\lambda\tau}B(\tau)d\tau\right).$$

Het puntspectrum wordt nu in feite bepaald door de eis dat de in (2.8) gedefinieerde functie tot X moet behoren. Stel bijvoorbeeld dat X de ruimte van continue functies ψ op $\mathbb{R}_+$ met $\lim_{t\to\infty} \psi(t) = 0$ is, voorzien van de supremumnorm, dan behoort het hele linkerhalfvlak Re $\lambda < 0$ tot het puntspectrum van A en dit, voorwaar, is een bedroevend resultaat! Vandaar misschien dat de mensen die werken in ruimten van op $\mathbb{R}_+$ gedefinieerde functies nooit, bij mijn beste weten, met deze infinitesimale generator op de proppen zijn gekomen.

Na deze lange uitweiding over allerlei doodlopende wegen ben ik eindelijk gekomen bij een lichtpuntje. Als we nagaan hoe de formulering (2.2) - (2.3), die het *autonome* karakter van de dynamica benadrukt, past in de formulering (2.1), dan is (2.4) de sleutelformule. We kunnen (2.4) als volgt lezen: als we f interpreteren als een beschrijving van de direkte invloed van het verleden tussen -b en 0, dan is noodzakelijkerwijs f identiek nul voor $t \geq b$. Deze opmerking maakt het natuurlijk om te definiëren

$$(2.9)\qquad X = \{f \in C(\mathbb{R}_+) \mid f(t) = 0 \text{ voor } t \geq b\}$$

en zie, nu blijkt wel degelijk de eis dat f in (2.8) tot X behoort te leiden tot de karakteristieke vergelijking (2.5). Bovendien geldt dat A een zuiver puntspectrum heeft.

Hoe komt het nu dat we van de sprong-discontinuïteit geen last meer hebben? Als we (2.7) uitschrijven onder gebruikmaking van (2.3), (2.4) en de notatie $x_s(t) = x(t+s)$ dan krijgen we

$$(T(s)f)(t) = \begin{cases} \displaystyle\int_{t-b}^{0} x_s(\tau)B(t-\tau)d\tau & \text{voor } t \le b, \\ 0 & \text{voor } t \ge b. \end{cases}$$

De functie $x_s(\tau)$ heeft mogelijk een sprong-discontinuïteit voor $\tau = -s$, maar deze wordt door de convolutie met B gladgestreken! Dit is immers ook precies de reden dat de sprong in $t = 0$ de continuïteit van de oplossing voor $t > 0$ niet in gevaar brengt.

Het idee om X te definiëren als in (2.9) is ontstaan tijdens het schrijven van deze inleiding. Het was namelijk mijn bedoeling om de haken en ogen van een fundamentele benadering toe te lichten ter verontschuldiging van de ad hoc methode in [A]. Al denkende over de juiste formuleringen kwam toen dit nieuwe idee naar voren en ik hoop nu dat het mogelijk zal zijn om vanuit deze benadering te komen tot een bevredigende kwalitatieve theorie. Als een illustratie van het belang van een dergelijke theorie wil ik nog de mogelijkheid noemen om via een Hopf-bifurcatiestelling het bestaan en de stabiliteit van periodieke oplossingen van het niet-lineaire analogon van (2.2) te bewijzen. De uitwerking zal, ondanks het feit dat het grotendeels gaat om het bewandelen van reeds gebaande paden, nog wel de nodige tijd en inspanning kosten. Ik hoop te zijner tijd deze vage beweringen door tastbare resultaten te kunnen vervangen.

Tot slot van deze paragraaf geef ik een samenvatting van de belangrijkste punten van mijn betoog.

Bij de beschrijving van dynamische verschijnselen in populaties met een leeftijdsstruktuur ontstaan Volterra integraalvergelijkingen van convolutie type. Het autonome karakter van de dynamica wordt het beste tot uiting gebracht in een formulering als (2.2) - (2.3), maar helaas blijkt een daarbij aansluitende semigroep constructie via (2.6) op een technisch probleem stuk te lopen.

Voor de formulering als in (2.1) bestaat een uitgebreide theorie die vooral gericht is op industriële toepassingen, waarbij f gezien wordt als

een externe input en x als de output. Het is dan natuurlijk om f te laten behoren tot een klasse van functies die op $\mathbb{R}_+$ gedefinieerd en bijvoorbeeld begrensd is, zonder verder veel beperkingen op te leggen. Het blijkt nu dat onze biologische invalshoek het natuurlijk maakt om juist te eisen dat f identiek nul is voor $t \geq b$. Dit brengt ons dan bij de semigroep van operatoren, gedefinieerd in (2.7), werkend op de ruimte X, gedefinieerd in (2.9). Het lijkt mogelijk en zinvol om vanuit deze constructie een kwalitatieve stabiliteitstheorie voor lineaire en niet-lineaire vergelijkingen op te bouwen die overeenkomst vertoont met de theorie van functionaal-differentiaalvergelijkingen van "retarded" type.

3. VERSPREIDING IN RUIMTE EN TIJD

De oplossing van het in paragraaf 1 geschetste probleem en het schrijven van [A] hield mij bezig tot in het voorjaar van 1976. In datzelfde voorjaar werkte ik intensief aan het Colloquium Niet-lineaire Diffusieproblemen, dat de afdeling Toegepaste Wiskunde van het Mathematisch Centrum organiseerde (zie [15]). Het was dus zeer voor de hand liggend dat ik mij vervolgens af ging vragen waar het epidemiemodel toe zou leiden als men de aanname laat vallen dat de bevolking zich homogeen mengt (hetgeen de rechtvaardiging was voor het veronachtzamen van plaatsafhankelijkheid). In eerste instantie was het mijn bedoeling om op diffusievergelijkingen uit te komen, maar nadat Metz mij gewezen had op BAILEY [5], KENDALL [26], [27] en MOLLISON [39],[40], ging ik mij concentreren op de vergelijking

$$(3.1) \qquad u(t,x) = S_0 \int_0^t A(t-\tau) \int_{\mathbb{R}^n} g(u(\tau,\xi))V(x-\xi)\,d\xi\,d\tau + f(t,x)$$

en op een iets algemenere variant. Het essentiële bij de afleiding van deze vergelijking (zie [B]) is dat de infectiviteit een gewogen gemiddelde is over de leeftijds- en plaatsgestructureerde dichtheid. Het zijn niet de individuen die bewegen, het is de ziekte (men denke bijvoorbeeld aan plantenziekten; zie MOLLISON [41] voor achtergrondinformatie).

Een eerste vraag, die zeer voor de hand ligt, is: vertoont deze vergelijking een drempelverschijnsel zoals we dat in paragraaf 1 tegenkwamen? Het antwoord is: ja en zelfs zeer sterk, want als de parameter boven de drempelwaarde ligt dan leidt *elke* verstoring, zelfs als die in een heel klein gebiedje geconcentreerd is, uiteindelijk *overal* tot grote gevolgen (men zegt dan dat de vergelijking het "hair-trigger effect" vertoont). Dit wordt in [B,§5] bewezen voor $n = 1$ en $n = 2$ door gebruik te maken van bekende resultaten voor lineaire convolutie-ongelijkheden (zie ook [14]) en in [D] wordt het bewezen voor willekeurige n met behulp van de constructie van geraffineerde vergelijkingsfuncties.

Een tweede vraag, die ook in de literatuur steeds opnieuw naar voren komt, is: zijn er lopende golfoplossingen? Voordat men deze vraag kan beantwoorden is eerst enige bezinning nodig. Een lopende (vlakke) golfoplossing is een oplossing van de vorm $u(t,x) = w(x\cdot\nu + ct)$, waarbij de vaste eenheidsvector ν de richting aangeeft en het vaste getal c de snelheid. Als we bij vaste t, x laten variëren, dan varieert het argument van w van $-\infty$ tot $+\infty$. We kunnen dus ook bij vaste x, t laten variëren van $-\infty$ tot $+\infty$,

zodat de oplossing gedefinieerd is voor alle t. Het gaat dus niet om een beginwaardeprobleem en de vergelijking moet het verschijnsel voor alle t beschrijven. (Voor gewone differentiaalvergelijkingen maakt dit niets uit, maar voor Volterra integraalvergelijkingen wel degelijk.) Het komt er op neer dat we in plaats van naar (3.1) nu moeten kijken naar de homogene vergelijking

$$(3.2) \qquad u(t,x) = S_0 \int_{-\infty}^{t} A(t-\tau) \int_{\mathbb{R}^n} g(u(\tau,\xi))V(x-\xi)\,d\xi\,d\tau .$$

Substitutie van de vooronderstelling dat we kunnen schrijven $u(t,x) = w(x\cdot\nu + ct)$ leidt tot een convolutievergelijking voor w

$$(3.3) \qquad w(y) = S_0 \int_{-\infty}^{\infty} g(w(\eta))V_c(y-\eta)\,d\eta, \qquad -\infty < y < \infty,$$

waarbij de kern V_c onafhankelijk van ν is als V een radiële functie is (zie [B,§6] en [D,§2] voor de details).

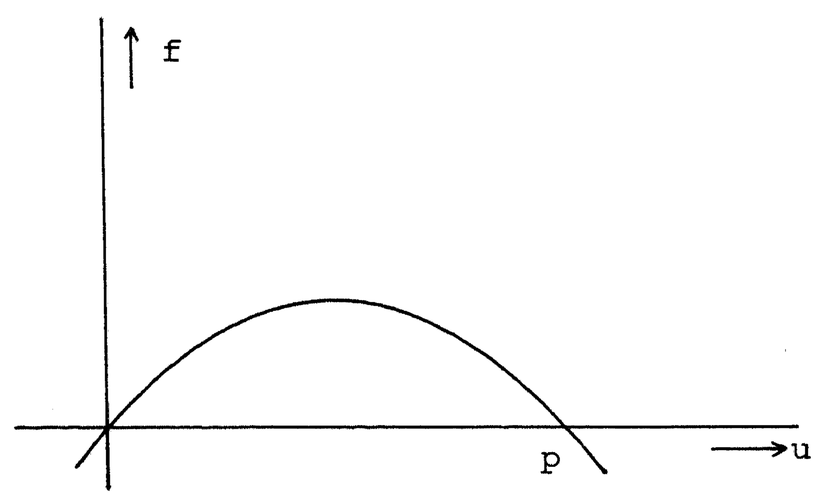
Om de situatie enigszins toe te lichten wil ik een kort uitstapje maken naar een niet-lineair diffusieprobleem. Beschouw de vergelijking

$$\frac{\partial u}{\partial t} = \Delta u + f(u),$$

dan leidt substitutie van $u(t,x) = w(x\cdot\nu + ct)$ tot de tweede-orde gewone differentiaalvergelijking

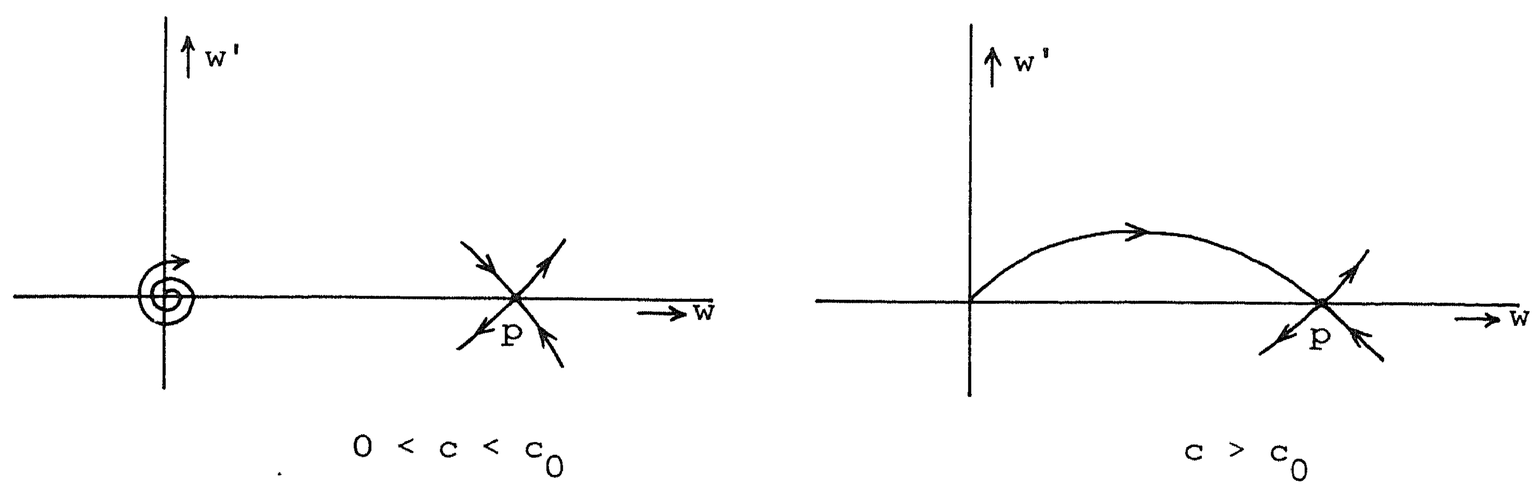
$$(3.4) \qquad w'' - cw' + f(w) = 0.$$

Stel dat f "er uit ziet" als in Figuur 6.



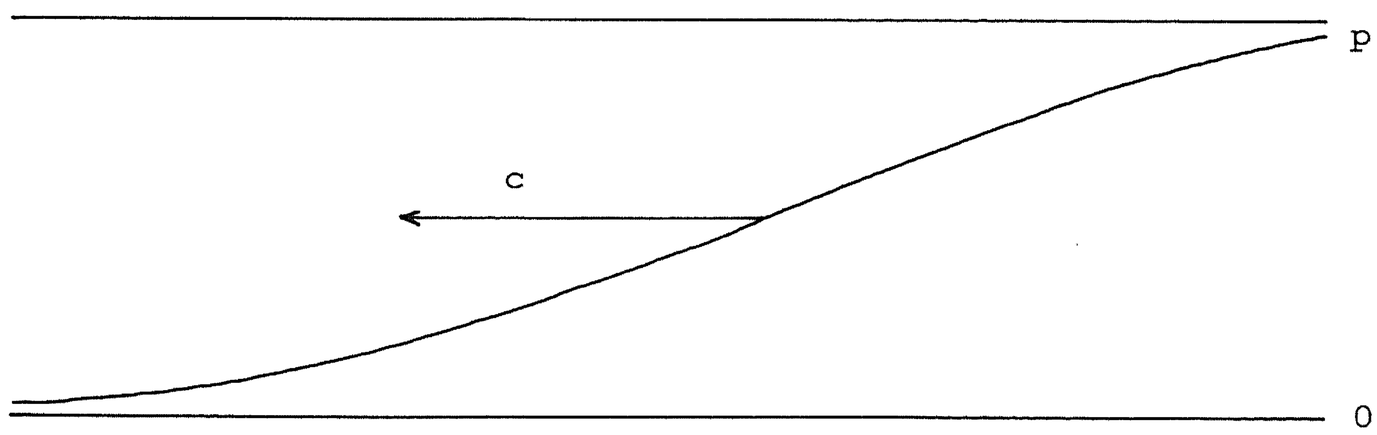
Figuur 6.

Dan zijn er dus voor iedere c twee constante oplossingen van (3.4), namelijk 0 en p. Stel nu dat we ons afvragen of er ook oplossingen zijn, die tussen 0 en p inliggen, dan kunnen we het antwoord op die vraag vinden door een fasevlak-analyse te maken (zie [15, Hoofdstuk IV]) voor een gedetailleerde behandeling van dit probleem).



Figuur 7.

Zo'n analyse begint met het onderzoek naar de aard van de singuliere punten (0,0) en (p,0). Dat wil zeggen dat men de vergelijking lineairiseert om die punten en dat men nagaat waar de wortels van de karakteristieke vergelijking (een tweede-graads polynoom) liggen in het complexe vlak. Zo blijkt dat er een getal c_0 bestaat met de eigenschap dat voor $0 < c < c_0$, (0,0) een instabiel spiraalpunt is, zodat er voor die waarden van c onmogelijk zo'n oplossing met $0 < w(x) < p$ kan bestaan. Voor $c > c_0$ is (0,0) een instabiele knoop en voor alle $c > 0$ is (p,0) een zadelpunt. Onder bepaalde aannamen voor het globale gedrag van f kan men nu bewijzen dat voor $c > c_0$ de stabiele variëteit van (p,0) uit (0,0) afkomstig is (zie Figuur 7), zodat er dan dus wel een oplossing bestaat van het type waar we in geïnteresseerd zijn. Bovendien blijkt dat die oplossing



Figuur 8.

een vorm heeft als getekend in Figuur 8 en dat er modulo translatie precies één zo'n oplossing is (als $w(x)$ aan (3.4) voldoet dan voldoet ook $w_h(x) = w(x+h)$ aan (3.4), omdat de vergelijking translatie-invariant is; maar al deze oplossingen corresponderen met dezelfde baan in het fasevlak en de uniciteit volgt uit onze kennis van het gedrag van banen in de buurt van het zadelpunt $(p,0)$).

We kunnen nu de resultaten van [B,§6] en [C] kort en bondig samenvatten door te zeggen dat (3.3) precies hetzelfde beeld te zien geeft als (3.4).

In (3.3) staat een één-parameter familie van (translatie-invariante) convolutievergelijkingen op de hele reële rechte. Voor iedere c heeft de vergelijking de constante oplossingen 0 en p. We zijn geïnteresseerd in een karakterisering van de waarden van c waarvoor er ook oplossingen zijn met $0 < w(x) < p$. Het is duidelijk dat we nu niet onze toevlucht kunnen nemen tot fasevlak-technieken, maar desondanks wordt bij de analyse een hoofdrol gespeeld door de karakteristieke vergelijking die ontstaat uit de om 0 gelineairiseerde vergelijking. Voor een uitgebreide discussie van deze vergelijking verwijs ik naar [B,§6].

Bij mijn onderzoek van (3.3) was ik gevorderd tot en met een analyse van deze karakteristieke vergelijking, op grond waarvan ik een sterk vermoeden had wat de resultaten zouden gaan worden, toen het artikel van ATKINSON & REUTER [4] verscheen. In dit artikel wordt het stelsel vergelijkingen behandelt dat men krijgt door, uitgaande van (1.4) - (1.5), plaatsafhankelijkheid in het model op te nemen. Voor dit speciale geval leverde hun resultaten een bevestiging van mijn vermoedens. Bovendien bevatte het artikel een bewijs van het bestaan van oplossingen voor $c > c_0$ dat vrij gemakkelijk te generaliseren viel.

Dat existentiebewijs is gebaseerd op het volgende recept. Vat het rechterlid van (3.3) op als een (formele) integraaloperator T werkend op w, dan is T monotoon. Construeer twee functies ϕ en ψ zó dat $0 \le \phi \le \psi \le p$ en $T\phi \ge \phi$, $T\psi \le \psi$. De monotonie van T impliceert nu dat er een dekpunt w van T bestaat met $\phi \le w \le \psi$. De enige moeilijkheid is in feite de constructie van ϕ en ψ en deze is gebaseerd op de informatie die de karakteristieke functie verschaft (zie [B,§6]).

Op dat moment was het voor mij een open probleem om te bewijzen dat er geen oplossingen zijn voor $0 < c < c_0$ (het bewijs van Atkinson en Reuter bevatte elementen die niet gegeneraliseerd konden worden) en dat er voor $c > c_0$ modulo translatie slechts één oplossing is (hetgeen ook door Atkinson

en Reuter als open probleem werd genoemd). Ik stelde H.G. Kaper, die een half jaar als gastmedewerker op het Mathematisch Centrum kwam werken, voor om deze problemen gezamenlijk te lijf te gaan en dat heeft uiteindelijk geresulteerd in [C]. De aanpak in [C], die zonder meer klassiek te noemen is, is gebaseerd op het gebruik van twee fundamentele werktuigen: Pitt's versie van Wiener's algemene Tauberstelling en Laplace transformatie. Voor de volledigheid merk ik hierbij op dat we, voordat we met deze aanpak succes hadden, ook pogingen hebben gedaan om ons doel langs functionaalanalytische weg te bereiken met behulp van Fourier transformatie van distributies en pogingen om de Wiener-Hopf methode toe te passen.

Wat het niet-existentie deel betreft is er een alternatieve aanpak mogelijk. Het recept is als volgt. Construeer een functie ψ met compacte drager zó dat $T(\delta\psi) \geq \delta\psi$ voor δ positief en voldoende klein en zó dat $\liminf_{n\to\infty} T^{(n)}(\delta\psi) \geq p$. Zij dan w een oplossing van (3.3) die niet identiek 0 is, dan is het mogelijk δ zo klein te kiezen dat $w \geq \delta\psi$. Omdat T monotoon is geldt dan $w \geq T^{(n)}(\delta\psi)$ voor alle n en dus $w \geq p$. De lezer kan in [D] alle voor dit recept benodigde ingrediënten vinden (zie ook WEINBERGER [45]).

In [C] wordt uniciteit modulo translatie bewezen onder de nogal beperkende voorwaarde dat g op het interval [0,p] een Lipschitz constante heeft die kleiner of gelijk aan g'(0) is. Met een iets gewijzigde bewijsvoering, waarbij meer gebruik wordt gemaakt van de eigenschappen van g(w) voor w nabij p, is het mogelijk om deze voorwaarde geheel te laten vallen.

We hebben tijdens de excursie naar het niet-lineaire diffusieprobleem gezien dat daar een minimale golfsnelheid c_0 bestaat. In twee invloedrijke artikelen [2],[3] hebben ARONSON & WEINBERGER aangetoond dat oplossingen van het beginwaardeprobleem zich in een zekere zin asymptotisch met die snelheid c_0 voortplanten. Aan het eind van [B] opperde ik de mogelijkheid dat voor (3.1) wel eens hetzelfde zou kunnen gelden. Kort daarna kreeg ik preprints van [1] en [45] onder ogen. In [1] bewijst ARONSON dat voor het stelsel, dat ook Atkinson en Reuter geanalyseerd hebben, c_0 inderdaad de asymptotische voortplantingssnelheid is, en in [45] bewijst WEINBERGER hetzelfde voor de integro-differentievergelijking

$$u_{m+1}(x) = \int_{\mathbb{R}^n} g(u_m(\xi))k(x-\xi)\,d\xi$$

(hierbij is m een discrete tijdsvariabele, de generatie, en u is de relatieve dichtheid van een bepaald gen in de populatie; de kern k beschrijft

het effect van migratie). Nu kon ik de kunst van het construeren van bepaalde vergelijkingsfuncties voor convolutievergelijkingen van de "meesters" afkijken en alles wat ik verder nog moest doen was het vinden van een hanteerbaar vergelijkingsprincipe voor (3.1). De resultaten van dit onderzoek vormen de inhoud van [D].

"Physical models are as different from the world as a geographical map is from the surface of the earth"

L. Brillouin

4. MODEL EN REALITEIT

De werkwijze in het voorafgaande en in de verschillende artikelen kan men als volgt omschrijven. Het uitgangspunt is een deterministisch model van de verspreiding van een besmettelijke ziekte. Zo'n model bestaat uit een aantal aannamen die wiskundig geformuleerd kunnen worden. Na enige manipulatie ontstaat er uit de aannamen één of andere vergelijking. De biologische betekenis van de diverse grootheden leidt op een natuurlijke manier tot een aantal preciese wiskundige vragen over het kwalitatieve gedrag van de oplossing(en) van deze vergelijking en de antwoorden op deze vragen laten een duidelijke biologische interpretatie toe.

Men noemt dit *toegepaste* wiskunde omdat de wiskundige problemen gemotiveerd en toegelicht worden aan de hand van een model van een (natuurwetenschappelijk) verschijnsel. Men moet dus ook deze terminologie interpreteren als: vanuit de wiskunde bezien is er hier sprake van een toepassing. Een voor de hand liggende vraag is dan: is het ook toegepaste of toepasbare wetenschap?

De kans is groot dat een praktisch ingestelde bioloog of medicus meewarig zijn hoofd schudt als hij dit proefschrift onder ogen krijgt. Hij zal zich verbazen over de naïviteit waarmee het verschijnsel "verspreiding van een besmettelijke ziekte" benaderd wordt. En hij zal zich afvragen hoe mensen er toe komen om te gaan werken vanuit een paar onrealistische algemene aannamen, terwijl het toch empirisch volkomen duidelijk is dat er een grote diversiteit van ziekten is, ieder met eigen belangrijke en onderling sterk verschillende kenmerken!

In deze paragraaf wil ik vooral op deze laatste vraag nader ingaan. Ik wil proberen om, veelal aan de hand van citaten, iets duidelijk te maken van de ideeën die leven onder sommige theoretisch biologen die zich met dit soort modelvorming en modelanalyse bezighouden. Wat zijn hun uitgangspunten en hun doestellingen en welke resultaten verwachten zij van hun werk?

Laten we beginnen met terug te keren tot KERMACK & McKENDRICK. Zij openen hun artikel [28] als volgt:

> "*One of the most striking features in the study of epidemics is the difficulty of finding a causal factor which appears to be adequate to account for the magnitude of the frequent epidemics of disease which visit almost every population. It was with a view to obtaining more insight regarding the effects of the various factors which govern the spread of contagious epidemics that the present investigation was undertaken*"

en even verder zeggen zij

> "*In the course of time the epidemic may come to an end. One of the most important problems in epidemiology is to ascertain whether this termination occurs only when no susceptible individuals are left, or whether the interplay of the various factors of infectivity, recovery and mortality, may result in termination, whilst many susceptible individuals are still present in the unaffected population*".

We hebben in de eerste paragraaf gezien dat zij in het raamwerk van een vrij eenvoudig model een precies antwoord op de laatste vraag kunnen geven en dat zij er bovendien in slagen om de hele probleemstelling terug te brengen tot een analyse van de scalaire vergelijking (1.3). Op dat moment is er een eenduidig verband gelegd tussen een aantal aannamen en een eenvoudig plaatje, Figuur 2. Wat daartussenin is gebeurd is nu in feite niet meer belangrijk en men mag het vergeten. De parameters γ, S_0 en $f(\infty)$ en de variabele $u(\infty)$ hebben een duidelijke biologische betekenis binnen het model en men kan kwalitatieve conclusies in één oogopslag uit Figuur 2 aflezen. De belangrijkste conclusie is dan

> "*In general a threshold density of population is found to exist, which depends upon the infectivity, recovery and death rates peculiar to the epidemic. No epidemic can occur if the population density is below this threshold value*".

Los van de vraag hoe goed het model in staat is om de verspreiding van sommige "echte" ziekten te beschrijven is er nu iets bereikt. Immers, er is een duidelijk beeld gegeven van de manier waarop dit hypothetische niet-lineaire interaktiemechanisme zich manifesteert en het is juist via zo'n eenvoudig en volledig te analyseren voorbeeld dat men een "gevoel" kan opbouwen voor de "kwalitatieve werking" van een interaktiemechanisme.

In deze benadering is het misschien beter om te spreken over een proefmodel in plaats van over een model. In feite gaat het om een experiment: er wordt onderzocht wat het kwalitatieve beeld is dat hoort bij een aantal hypothesen. Dat dit idee veelvuldig voorkomt in de mathematische biologie moge blijken uit de volgende aanhaling van MAYNARD SMITH [35, p.2]

"*In all actual ecosystems, migration, territorial behaviour, climatic fluctuations and intraspecific genetic variation are all relevant. But we will not for this reason put all these features into all our models. Instead we adopt the method of the experimental scientist, which is to vary one factor at a time, and to do so in a system which is otherwise as simple as possible*".

Om vat te krijgen op een complex verschijnsel probeert men verschillende aspecten er uit te lichten om deze dan eerst afzonderlijk grondig te onderzoeken in de context van een geïdealiseerd model. In een later stadium kunnen dan de resultaten van allerlei van dit soort onderzoekingen gecombineerd worden tot een totaalbeeld en men hoopt dat men uit de vergelijking van dit totaalbeeld met de werkelijkheid conclusies kan trekken. Hierbij merkt MAYNARD SMITH [34, p.92] op

"*It is characteristic of mathematical ideas in biology that they are most suggestive when they are contradicted by experiment*"

en KENDALL [27] zegt iets soortgelijks

"*To know the consequences may be to reject the theory, but this is one way in which truth prevails*".

Het is duidelijk dat de weg naar het einddoel lang is en dat verschillende mensen tegelijkertijd aan het werk kunnen zijn zonder dat zij oog hebben voor elkaar.

Een variant van het los van elkaar bestuderen van verschillende aspecten is de stap voor stap methode. Hierbij begint men met het veronachtzamen van zoveel mogelijk factoren om er daarna geleidelijk weer meer in het model mee te nemen. Een goed voorbeeld hiervan is Kendall's introductie van plaatsafhankelijkheid in het speciale Kermack en McKendrick model. Op dat moment is de dynamica zonder plaatsafhankelijkheid volledig bekend en begrepen en men vraagt zich af: wat blijft er hetzelfde en wat voor nieuwe verschijnselen komen erbij? Het is deze vraagstelling die ten grondslag ligt aan het grootste deel van dit proefschrift.

Uit dit voorbeeld blijkt ook dat opbouwende kritiek op de aannamen van een model vruchtbaar kan werken. In feite dient het model juist gezien te worden als een kapstok voor begripsvorming en niet als een star dogma. METZ [36] zegt hierover

"*The reason for listing some set of assumptions is that one seeks to answer questions about the real world phenomena which one tries to model. However, those questions are usually inherently vague, and have to be made more precise within the framework of the model.*

> *Actually, the most important reason for building a model at all is that it leads to new conceptualizations and so to asking novel questions*"

en MAY [33, p.11] merkt op

> "*Such general models, even though they do not correspond in detail to any single real community, aim to provide a conceptual framework for the discussion of broad classes of phenomena. Such framework can serve a useful purpose in indicating key areas or relevant questions for the field and laboratory ecologist, or simply in sharpening discussion of contentious issues*".

In het bovenstaande heb ik al even aangestipt dat men via een gedetailleerde analyse van een zo eenvoudig mogelijk probleem een "gevoel" kan opbouwen voor de essentiële kenmerken van een veel grotere klasse van problemen. Voorbeelden van deze aanpak vindt men volop in de natuurwetenschappen: de harmonische oscillator, het perfekte kristal, het Ising model. Een noodzakelijke voorwaarde voor het begrijpen van iets is een gevoel van vertrouwdheid ermee en het is juist bij het nemen van deze psychologische hindernis dat het prototype probleem als polsstok fungeert. Evenzo blijkt de zeggingskracht van een numerieke analyse van een ingewikkeld probleem vele malen groter te worden indien men de resultaten via een "net zo iets als" formulering in verband kan brengen met de kwalitatieve analyse van een prototype probleem (met name bij de bestudering van prikkelgeleiding in zenuwen heeft dit een rol gespeeld; zie [15] en de daarin vermelde bronnen).

Op het grensvlak van natuurwetenschap en wiskunde werkt een prototype probleem als een tweesnijdend zwaard. Immers, ook binnen de wiskunde ontwikkelt een theorie zich vaak in het voetspoor van de volledige uitwerking van een modelprobleem (bijvoorbeeld de van der Pol vergelijking, of de Laplace vergelijking) en nadat de theorie eenmaal ontwikkeld is fungeert zo'n probleem nog vaak als een didactisch erg belangrijk illustratief voorbeeld.

Waarom raakte de Volterra integraalvergelijking van Kermack en McKendrick in het vergeetboek? Ik denk omdat dit soort vergelijkingen toen (en ook nu nog vaak) als "vreemd" ervaren werd. Het is juist door de uitwerking van specifieke vergelijkingen die uit een natuurwetenschappelijk model ontstaan dat men vertrouwd raakt met de essentiële kenmerken van het type vergelijking en dat men er spelenderwijs toe komt om een algemene theorie op te bouwen.

Zover ik weet was KENDALL de eerste die een model formuleerde waarbij plaatsafhankelijkheid beschreven wordt via een convolutie-integraal met

positieve kern (zie [26]). Toen hij enige tijd later het bestaan van lopende golfoplossingen wilde onderzoeken, ging hij echter over op een "diffusiebenadering" (zie [27] en ook HOPPENSTEADT [24]). Waarom? Achteraf kan men concluderen dat hij het zich daarmee niet makkelijker maakte (ik ben zelfs geneigd om te zeggen moeilijker), maar op dat moment was er kennelijk een aantrekkingskracht van het vertrouwde, de diffusievergelijkingen, en een afstotende kracht van het onbekende, de convolutievergelijkingen.

Al met al komt het er op neer dat men hoopt en verwacht dat modellen zoals dat van Kermack, McKendrick en Kendall vooral zullen fungeren als gangmakers (in een zeer lezenswaardige verhandeling [16, Hoofdstuk 5] over methodologische aspecten van modelbouw in de biologie, noemt DOUCET dit de heuristische rol van een model). Enerzijds dragen de resultaten van een uitwerking bij tot een begripsmatige verheldering, anderzijds stimuleert de uitwerking de ontwikkeling van methoden en technieken die bruikbaar zijn bij een analyse van verwante meer realistische modellen.

De bovenstaande opmerkingen en citaten zijn vooral gericht op een verduidelijking van het theoretisch wetenschappelijk kader tegen de achtergrond waarvan men een deel van dit proefschrift zou kunnen zien. Het gaat hoe dan ook om theorie, met in eerste instantie als voornaamste doel een vergroting van het begrip van verschijnselen die men waarneemt. De vraag blijft welke invloed deze theorie heeft of kan hebben op de praktische werkelijkheid. Op deze vraag moet ik het antwoord schuldig blijven omdat ik daar weinig of niets van afweet. Ik denk trouwens dat een precies antwoord op die vraag onmogelijk is, omdat veel van die invloed indirekt is. Het is via een lange keten van elkaar beïnvloedende mensen dat er verband bestaat tussen theorie en praktijk en dat maakt het moeilijk om een goed beeld van dat verband te krijgen.

Maar laat ik aan de hand van een voorbeeld aantonen dat er wel degelijk vanuit een veel meer op de praktijk gerichte instelling met soortgelijke modellen gewerkt wordt. In [47] beschrijven ZADOKS & KAMPMEIJER een simulatiestudie van een model dat in essentie hetzelfde is als dat van Kermack, McKendrick en Kendall. Hun belangstelling komt voort uit het feit dat bepaalde graansoorten voorkomen in verschillende variëteiten die een verschillende resistentie hebben voor bepaalde ziekten. Het is hun doel om na te gaan welke menging van zaden de hoogste opbrengst geeft.

Natuurlijk ligt vaak aan het gebruik van modellen in de epidemiologie de wens ten grondslag om een bepaalde ziekte te bestrijden (zie [7],[11] en [32]) en soms gaat dit samen met het toepassen van (optimale) controle

theorie (zie ook [6]). In 1760 hield DANIËL BERNOULLI een hartstochtelijk pleidooi (gepubliceerd in 1766, zie [8]) voor het toepassen op grote schaal van inenting tegen pokken. Zijn argumentatie is gebaseerd op wiskundige berekeningen, maar af en toe onderstreept hij zijn betoog door er op te wijzen dat het maatschappelijk nut van een individu leeftijdsafhankelijk is. Dat doet vermakelijk aan omdat het lang geleden is en omdat het in feite losse opmerkingen zijn.

Toch is het onontkoombaar dat bij een formulering van beslissingsproblemen andere dan strikt natuurwetenschappelijke factoren een rol gaan spelen. Zo beschouwen bijvoorbeeld LONGINI, ACKERMAN & ELVEBACK [31] een model waarbij de bevolking verdeeld is in bepaalde leeftijdsgroepen die verschillen in een aantal kenmerken zoals de vatbaarheid, de frequentie van onderlinge contacten en de kans om aan de ziekte (griep in dit geval) te sterven. Zij dringen er op aan om te komen tot het vaststellen van de doelstellingen van een immunisatiecampagne, omdat

> "*Public-health authorities must decide which segments of the population should be immunized with limited supplies of vaccine*".

Bij het vaststellen van die doelstellingen speelt het dan een rol welke negatieve waarde men toekent aan enerzijds de sterfte (uitgedrukt in het aantal verwachte levensjaren dat verloren gaat) en anderzijds de materiële schade (uitgedrukt in geld). Heeft men eenmaal de doelstellingen gekozen dan kan een optimale vaccinatiestrategie uitgerekend worden.

Op zichzelf zie ik de redelijkheid van een dergelijk betoog in: als er dan een beslissing genomen moet worden dan is het maar beter om je precies rekenschap te geven van de gronden waarop je dat doet! Maar anderzijds stuit een dergelijk voorbeeld van technocratisch denken mij gevoelsmatig tegen de borst. Stap voor stap is de technologisch-wetenschappelijke ontwikkeling gebaseerd op een rationale benadering van de problemen waarvoor men zich gesteld ziet. Maar de oplossing van ieder probleem roept een nieuw probleem in het leven en het hangt af van de manier waarop men dit in een winst- en verliesrekening onderbrengt of men blij is met de totale ontwikkeling of niet (zie [9]).

LITERATUUR

[1] ARONSON, D.G., *The asymptotic speed of propagation of a simple epidemic*, in: Nonlinear Diffusion, W.E. Fitzgibbon & H.F. Walker, eds., Research Notes in Mathematics 14 (Pitman Publishing Co., London, 1977) 1-23.

[2] ARONSON, D.G. & H.F. WEINBERGER, *Nonlinear diffusion in population genetics, combustion, and nerve pulse propagation*, in: Partial Differential Equations and Related Topics, J.A. Goldstein, ed., Lecture Notes in Math. 446 (Springer, Berlin, 1975) 5-49.

[3] ARONSON, D.G. & H.F. WEINBERGER, *Multidimensional nonlinear diffusion arising in population genetics*, te verschijnen in Advances in Math.

[4] ATKINSON, C. & G.E.H. REUTER, *Deterministic epidemic waves*, Math. Proc. Camb. Phil. Soc. 80 (1976) 315-330.

[5] BAILEY, N.T.J., *The Mathematical Theory of Epidemics* (Griffin, London, 1957).

[6] BANKS, H.T., *Modeling and Control in the Biomedical Sciences*, Lecture Notes in Biomath. 6 (Springer, Berlin, 1975).

[7] BERGER, J. & W. BÜHLER & R. REPGES & P. TAUTU, eds., *Mathematical Models in Medicine*, Lecture Notes in Biomath. 11 (Springer, Berlin, 1976).

[8] BERNOULLI, D., *Essai d'une nouvelle analyse de la mortalité causée par la petite Vérole & des avantages de l'Inoculation pour la prévenir*, Mémoires de mathématiques et de physique tirés des registres de l'Académie Royale des Sciences, de l'Année M.DCCLX, Imprimer (avec un introduction) M.DCCLXVI.

[9] BOERS, C., *De anti-technologiebeweging: de discussie over de technologie*, Intermediair 14e jaargang 27 (1978) 51-59.

[10] BRAUN, M., *Differential Equations and Their Applications*, Applied Math. Sciences 15 (Springer, Berlin, 1978).

[11] CHARNES, A. & W.R. LYNN, eds., *Mathematical Analysis of Decision Problems in Ecology*, Lecture Notes in Biomath. 5 (Springer, Berlin, 1975).

[12] CORDUNEANU, C., *Integral Equations and Stability of Feedback Systems* (Academic Press, New York, 1973).

[13] CUSHING, J.M., *Integrodifferential Equations and Delay Models in Population Dynamics*, Lecture Notes in Biomath. 20 (Springer, Berlin, 1977).

[14] DIEKMANN, O., *On a nonlinear integral equation arising in mathematical epidemiology,* te verschijnen in: W. Eckhaus & E.M. de Jager, eds., Proceedings of the Third Scheveningen Conference on Differential Equations.

[15] DIEKMANN, O. & N.M. TEMME, *Nonlinear Diffusion Problems,* MC Syllabus 28 (Math. Centrum, Amsterdam, 1976).

[16] DOUCET, P.G., *Studies on L-Systems,* Proefschrift (Utrecht, 1976).

[17] FRIEDMAN, A., *Partial Differential Equations* (Holt, Rinehart and Winston, New York, 1969).

[18] GRASMAN, J. & B.J. MATKOWSKY, *Evolution of an epidemic with a small number initial infectives,* TW Report 162 (Math. Centrum, Amsterdam, 1976).

[19] GURTIN, M.E. & R.C. MacCAMY, *Nonlinear age-dependent population dynamics,* Arch. Rat. Mech. Anal. 54 (1974) 281-300.

[20] HALE, J.K., *Theory of Functional Differential Equations,* Applied Math. Sciences 3 (Springer, Berlin, 1977).

[21] HALE, J.K. & P. MARTINEZ-AMORES, *Stability in neutral equations,* Nonl. Anal. Th. Meth. Appl. 1 (1977) 161-175.

[22] HENRY, D., *Geometric Theory of Semilinear Parabolic Equations,* te verschijnen.

[23] HOPPENSTEADT, F., *Thresholds for deterministic epidemics,* in: Mathematical Problems in Biology, Victoria Conference 1973, P. van den Driessche, ed., Lecture Notes in Biomath. 2 (Springer, Berlin, 1974) 96-101.

[24] HOPPENSTEADT, F., *Mathematical Theories of Populations: Demographics, Genetics and Epidemics,* SIAM Regional Conference Series in Applied Math. 20 (SIAM, Philadelphia, 1975).

[25] KENDALL, D.G., *Deterministic and stochastic epidemics in closed populations,* in: Proc. Third Berkeley Symposium, 4, J. Neyman, ed., (Univ. of California Press, 1956) 149-165.

[26] KENDALL, D.G., *Discussion of "Measles periodicity and community size"* by M.S. Bartlett, J. Roy. Statist. Soc. A 120 (1957) 64-67.

[27] KENDALL, D.G., *Mathematical models of the spread of infection,* in: Mathematics and Computer Science in Biology and Medicine (Medical Research Council, London, 1965) 213-224.

[28] KERMACK, W.O. & A.G. McKENDRICK, *A contribution to the mathematical theory of epidemics,* Proc. Roy. Soc. A 115 (1927) 700-721.

[29] LADAS, G.E. & V. LAKSHMIKANTHAM, *Differential Equations in Abstract Spaces* (Academic Press, New York, 1972).

[30] LAUWERIER, H.A., *Mathematische modellen voor epidemische processen*, Notitie uitgereikt tijdens het twaalfde Nederlands Mathematisch Congres (Amsterdam, 1976).

[31] LONGINI, Jr., I.M. & E. ACKERMAN & L.R. ELVEBACK, *An optimization model for Influenza A epidemics*, Math. Biosc. 38 (1978) 141-157.

[32] LUDWIG, D. & K.L. COOKE, eds., Proceedings of a SIMS Conference on Epidemiology (SIAM, Philadelphia, 1975).

[33] MAY, R.M., *Stability and Complexity in Model Ecosystems* (Princeton Univ. Press, 1973).

[34] MAYNARD SMITH, J., *Mathematical Ideas in Biology* (Cambridge Univ. Press, 1968).

[35] MAYNARD SMITH, J., *Models in Ecology* (Cambridge Univ. Press, 1974).

[36] METZ, J.A.J., *The epidemic in a closed population with all susceptibles equally vulnerable; some results for large susceptible populations and small initial infections*, Acta Biotheoretica 27 (1978) 75-123.

[37] MILLER, R.K., *Nonlinear Volterra Integral Equations* (Benjamin, Menlo Park, 1971).

[38] MILLER, R.K. & G.R. SELL, *Volterra Integral Equations and Topological Dynamics*, Memoir of the AMS 102 (AMS, Providence, 1970).

[39] MOLLISON, D., *The rate of spatial propagation of simple epidemics*, in: L.M. le Cam, J. Neyman & E.L. Scott, eds., Proc. Sixth Berkeley Symposium, III (Univ. of California Press, 1972) 579-614.

[40] MOLLISON, D., *Possible velocities for a simple epidemic*, Advances in Applied Prob. 4 (1972) 233-257.

[41] MOLLISON, D., *Spatial contact models for ecological and epidemic spread*, J. Roy. Statist. Soc. B 39 (1977) 283-326.

[42] REDDINGIUS, J., *Notes on the mathematical theory of epidemics*, Acta Biotheoretica 20 (1971) 125-157.

[43] SELL, G.R., *Topological dynamics and ordinary differential equations* (Van Nostrand Reinhold, London, 1971).

[44] VOLTERRA, V., *Leçons sur la Théorie Mathématique de la Lutte pour la Vie* (Gauthier-Villars, Paris, 1931).

[45] WEINBERGER, H.F., *Asymptotic behavior of a model in population genetics*, in: J.M. Chadam, ed., Nonlinear Partial Differential Equations and Applications, Lecture Notes in Math. 648 (Springer, Berlin, 1978).

[46] WILLEMS, J.C., *The Analysis of Feedback Systems*, Research Monograph 62 (M.I.T. Press, Cambridge, 1971).

[47] ZADOKS, J.C. & P. KAMPMEIJER, *Epimul, a simulator of foci and epidemics in mixtures of resistant and susceptible plants, mosaics and multilines* (Simulation Monograph, Centre for Agriculture Publishing and Documentation, Wegeningen, 1977).

Nonlinear Analysis, Theory, Methods & Applications. Vol. 1 No. 5, pp. 459–470. Pergamon Press, 1977. Printed in Great Britain

LIMITING BEHAVIOUR IN AN EPIDEMIC MODEL

O. DIEKMANN

Mathematisch Centrum, 2e Boerhaavestraat 49, Amsterdam, Holland

(*Received* 29 *October* 1976)

Key words: epidemic model of Kermack and McKendrick, threshold phenomenon, nonlinear Volterra integral equations, limiting behaviour as the forcing function tends to zero while $x = 0$ is a unstable steady state, linear and nonlinear Volterra equations on the whole line

1. INTRODUCTION TO THE BIOLOGICAL ASPECTS

IN 1927 Kermack and McKendrick [1] proposed a fairly general deterministic model for the spread (as a function of time) of an infectious disease in a closed population. The population is divided into a class of susceptibles S and a continuum of classes of infectives I_τ, $\tau \geqslant 0$, where τ is the length of time that passed away since exposure. The transfer of individuals from S to I_0 is assumed to be proportional to the number of susceptibles and to the 'total infectivity', to which every member of I_τ contributes a certain amount depending on τ. The disease is supposed to induce permanent immunity so there is no transfer to class S.

As an idealization of reality we suppose $S(t)$, the number of susceptibles at time t, to be real (and not necessarily integral) and moreover continuously differentiable with respect to t. Then the equation describing the dynamics of the epidemic is

$$\dot{S}(t) = S(t)\left\{\int_0^t \dot{S}(\tau)A(t-\tau)\,d\tau - h(t)\right\}. \tag{1.1}$$

Here $A(t)$ denotes the infectivity of an individual which has been infected at $t = 0$ and so $A(t)$ is nonnegative. The history up to $t = 0$ is described by the nonnegative function $h(t)$, which we assume to be known. Integration of equation (1.1) yields

$$\ln\frac{S(t)}{S(0)} = \int_0^t S(\tau)A(t-\tau)\,d\tau - S(0)\int_0^t A(\tau)\,d\tau - \int_0^t h(\tau)\,d\tau. \tag{1.2}$$

By the biological interpretation we expect the solution to be monotone nonincreasing and bounded from below (at least by zero). It is easily verified that this is in fact the case and hence the limit $S(\infty)$ exists. Assuming both

$$\gamma = \int_0^\infty A(\tau)\,d\tau < \infty \quad \text{and} \quad H(\infty) = \int_0^\infty h(\tau)\,d\tau < \infty$$

(for instance one might argue that for most diseases A and h have compact support) one deduces for $S(\infty)$ the equation

$$\ln\frac{S(\infty)}{S(0)} = \gamma(S(\infty) - S(0)) - H(\infty), \tag{1.3}$$

which has, as a graphical argument shows, a unique solution satisfying $0 < S(\infty) < S(0)$. If either $\gamma = \infty$ or $H(\infty) = \infty$ one has $S(\infty) = 0$ and equation (1.3) does only hold formally.

In their paper Kermack and McKendrick investigate the dependence of the limit $S(\infty)$ on the parameters $S(0)$ (which is approximately the total population size), γ (the total infectivity due to one infected individual during the course of his illness) and $H(\infty)$ (the total infectivity due to the history up to $t = 0$). In particular they discuss the following *threshold phenomenon* (also see [2, 3] and the references there) : if $\gamma S(0) \leqslant 1$ then $S(0) - S(\infty)$ is small if $H(\infty)$ is small, whereas if $\gamma S(0) > 1$ then $S(0) - S(\infty)$ is relatively large independently of how small we choose $H(\infty)$. This again can easily be verified by plotting a picture. In loose terms one can say that an epidemic 'occurs' if and only if $\gamma S(0) > 1$.

In this paper the dependence of the solution (and not merely the final size) on the function $h(t)$ is discussed in much more detail. We prove that if $\gamma S(0) > 1$ and if we let $H(\infty)$ tend to zero then the solution tends in a well-defined translation sense to a nontrivial limit which can be interpreted as an epidemic starting at $t = -\infty$. This result was conjectured by J. A. J. Metz [15]. In the special case that the epidemic is described by an autonomous ordinary differential equation a related point of view was taken up by Kendall [4].

2. LIMITING BEHAVIOUR OF SOLUTIONS OF A VOLTERRA INTEGRAL EQUATION

Consider the nonlinear Volterra integral equation

$$x(t) = \int_0^t g(x(\tau))A(t - \tau)\,d\tau + f(t;\varepsilon), \tag{2.1}$$

where g, A and f satisfy

H_g: g is uniformly Lipschitz continuous, monotone increasing and bounded on $[0, \infty)$; $g(0) = 0$.
H_A: A is piecewise continuous and nonnegative on $[0, \infty)$; $\int_0^\infty A(\tau)\,d\tau = \gamma < \infty$.
H_f: f is uniformly continuous, nonnegative and bounded on $[0, \infty) \times [0, \varepsilon_0]$; f is monotone nondecreasing as a function of t and monotone decreasing as a function of ε; $f(t; 0) \equiv 0$.

Since the equation is of convolution type it is called a nonlinear renewal equation as well. The parameter ε is introduced for notational convenience. The relation of (2.1) with (1.2) is easily seen by putting $x(t) = \ln S(0) - \ln S(t)$ and by making the special choice

$$g(x) = S(0)(1 - e^{-x}). \tag{2.2}$$

The existence of a solution $x(t;\varepsilon)$ of (2.1) is readily shown (for instance by monotone iteration) and uniqueness follows from a contraction argument on a small enough time interval. Furthermore $x(t;\varepsilon)$ is uniformly continuous, nonnegative and monotone nondecreasing as a function of t. The assumptions imply that $x(t;\varepsilon)$ is bounded from above by any root $\bar{x}(\varepsilon)$ of the scalar equation

$$x = \gamma g(x) + f(\infty;\varepsilon), \tag{2.3}$$

which satisfies the additional condition $\bar{x}(\varepsilon) > x(0;\varepsilon) = f(0;\varepsilon)$. For suppose $x(t;\varepsilon) < \bar{x}(\varepsilon)$ for all $t \in [0, t_0)$ then

$$x(t_0;\varepsilon) = \int_0^{t_0} g(x(t_0 - \tau))A(\tau)\,d\tau + f(t_0;\varepsilon) < \gamma g(\bar{x}(\varepsilon)) + f(\infty;\varepsilon) = \bar{x}(\varepsilon)$$

and hence $x(t;\varepsilon)$ cannot become equal to $\bar{x}(\varepsilon)$ in finite time. On account of H_g equation (2.3) has at least one positive root and every positive root satisfies the additional condition. So $x(t;\varepsilon)$,

being monotone nondecreasing and bounded from above, approaches a finite limit $x(\infty;\varepsilon)$ as t tends to infinity. Using the fact that $A \in L^1[0,\infty)$ one can show that the limit has to satisfy (2.3) and hence $x(\infty;\varepsilon)$ is the smallest positive root of (2.3).

Now we shall study the dependence of $x(t;\varepsilon)$ on ε (this kind of problem has been investigated by Brauer [5] under different hypotheses). For all finite t we have $\lim_{\varepsilon\downarrow 0} x(t;\varepsilon) = 0$ by continuous dependence on the forcing function. The next step is to investigate $x(\infty;\varepsilon)$. Since $f(\infty;\varepsilon) > 0$ and $x(\infty;\varepsilon)$ is the *smallest* positive root of (2.3), we know that $x < \gamma g(x) + f(\infty;\varepsilon)$ in a left-hand neighbourhood of $x(\infty;\varepsilon)$ and so in general $x > \gamma g(x) + f(\infty;\varepsilon)$ in a right-hand neighbourhood of $x(\infty;\varepsilon)$. If, however, for some ε, $\varepsilon = \varepsilon_1$ say, the straight line $y = x$ and the graph of $y = \gamma g(x) + f(\infty;\varepsilon)$ do not cross at $x(\infty;\varepsilon)$, then they do cross for all $\varepsilon < \varepsilon_1$ with $|\varepsilon - \varepsilon_1|$ sufficiently small. It follows that $\lim_{\varepsilon\downarrow 0} x(\infty;\varepsilon) = x_\infty$, where x_∞ is defined as the smallest nonnegative solution of the scalar equation

$$x = \gamma g(x), \tag{2.4}$$

such that $x > \gamma g(x)$ in a right-hand neighbourhood of x_∞.

Suppose $x_\infty > 0$, then we have nonuniform convergence: the limits $\varepsilon \downarrow 0$ and $t \to \infty$ are not interchangeable (this can be interpreted as instability of $x = 0$ as well) and the problem is to find a nontrivial limit that matches the limits 0 and x_∞. We shall show that such a limit can be constructed by considering properly chosen translates of $x(t;\varepsilon)$ (see [6] for an approach using singular perturbation methods). Choose x_0 with $0 < x_0 < x_\infty$ and define $\bar{t}(\varepsilon)$ by

$$x(\bar{t}(\varepsilon);\varepsilon) = x_0 \quad \text{and} \quad x(t;\varepsilon) < x_0 \quad \text{for all} \quad t < \bar{t}(\varepsilon),$$

and subsequently $x_\varepsilon(t)$ by

$$x_\varepsilon(t) = x(t + \bar{t}(\varepsilon);\varepsilon), \quad t \geqslant -\bar{t}(\varepsilon),\ \varepsilon > 0. \tag{2.5}$$

So we performed an ε-dependent shift of the time scale such that $x_\varepsilon(0)$ has an ε-independent value. Note that $\bar{t}(\varepsilon) \to \infty$ as $\varepsilon \downarrow 0$. It turns out to be convenient to define $x_\varepsilon(t)$ for $t < -\bar{t}(\varepsilon)$ too. This is done by setting

$$x_\varepsilon(t) = x(0;\varepsilon), \quad t < -\bar{t}(\varepsilon). \tag{2.6}$$

Let $C(\mathbf{R})$ be the Banach space of real-valued bounded continuous functions on $\mathbf{R}$, with the norm given by

$$\|x\| = \sup_{-\infty < t < \infty} |x(t)|.$$

Then clearly $x_\varepsilon \in C(\mathbf{R})$ for all $\varepsilon > 0$. Let, as ε tends to zero, X denote the limit set of $\{x_\varepsilon\}$ in $C(\mathbf{R})$, i.e.

$$X = \{x \in C(\mathbf{R}) \mid \text{there exists a sequence } \{\varepsilon_n\} \text{ with } \varepsilon_n \downarrow 0 \text{ as } n \to \infty \text{ such that } \lim_{n\to\infty} \|x_{\varepsilon_n} - x\| = 0\}. \tag{2.7}$$

We now state the main result of this section.

THEOREM 2.1. Suppose $x < \gamma g(x)$ for $0 < x < x_\infty$, then X is nonempty and every $x \in X$ has the following properties:

(i) $x(0) = x_0$,
(ii) $x(t)$ is monotone nondecreasing,

(iii) x satisfies the limit equation

$$x(t) = \int_{-\infty}^{t} g(x(\tau))\, A(t-\tau)\, d\tau, \quad -\infty < t < \infty, \tag{2.8}$$

(iv) $\lim_{t \to -\infty} x(t) = 0$ and $\lim_{t \to +\infty} x(t) = x_\infty$.

The proof consists of several steps. Let, as $\varepsilon \downarrow 0$, X_c denote the limit set of $\{x_\varepsilon\}$ with respect to the topology of uniform convergence on compact subsets of $\mathbf{R}$.

LEMMA 2.2. X_c is nonempty.

Proof. We intend to apply the Arzela–Ascoli theorem to the family of functions $\{x_\varepsilon\}$. We know already that $\{x_\varepsilon\}$ is uniformly bounded, so it remains to verify the equicontinuity. Let $z_\varepsilon(t)$ be defined by

$$z_\varepsilon(t) = \int_{-\bar{t}(\varepsilon)}^{t} g(x_\varepsilon(\tau))\, A(t-\tau)\, d\tau, \quad t \geqslant -\bar{t}(\varepsilon), \tag{2.9}$$

then x_ε satisfies

$$x_\varepsilon(t) = z_\varepsilon(t) + f(t + \bar{t}(\varepsilon); \varepsilon), \quad t \geqslant -\bar{t}(\varepsilon),$$

The estimate

$$|z_\varepsilon(t_1) - z_\varepsilon(t_2)| \leqslant g(\infty) \left\{ \int_0^\infty |A(t_1 - t_2 + \tau) - A(\tau)|\, d\tau + \int_0^{t_1 - t_2} A(\tau)\, d\tau \right\}$$

(with $t_1 \geqslant t_2 \geqslant -\bar{t}(\varepsilon)$) shows the equicontinuity of the first term on the right-hand side, whereas the same result for the second term is a consequence of the hypothesis H_f. Now the result follows from application of the Arzela–Ascoli theorem on an expanding sequence of compact intervals together with a diagonalization argument.

LEMMA 2.3. Every $x \in X_c$ has the properties (i)–(iv).

Proof. Let $x \in X_c$ then clearly x satisfies (i), (ii) and $0 \leqslant x(t) \leqslant x_\infty$. Put

$$z(t) = \int_{-\infty}^{t} g(x(\tau))\, A(t-\tau)\, d\tau$$

and let $\{\varepsilon_n\}$ be a sequence such that, as $n \to \infty$, $x_{\varepsilon_n}(t) \to x(t)$ uniformly on compact sets. We show that, for each t, $|z_{\varepsilon_n}(t) - z(t)| \to 0$ as $n \to \infty$. Let $\eta > 0$ be arbitrary, then there exists a number $T = T(\eta)$ such that $\int_T^\infty A(\tau)\, d\tau < (4g(\infty))^{-1}\eta$. For a given t there exists a number $N = N(\eta)$ such that for $n > N$, $|g(x(\tau)) - g(x_{\varepsilon_n}(\tau))| < (2\gamma)^{-1}\eta$ uniformly on $[t-T, t]$. So for $n > N$

$$\begin{aligned} |z_{\varepsilon_n}(t) - z(t)| &\leqslant \int_{-\infty}^{t-T} (g(x(\tau)) + g(x_{\varepsilon_n}(\tau)))\, A(t-\tau)\, d\tau + \int_{t-T}^{t} |g(x(\tau)) - g(x_{\varepsilon_n}(\tau))|\, A(t-\tau)\, d\tau \\ &\leqslant 2\, g(\infty) \int_T^\infty A(\tau)\, d\tau + \gamma \sup_{\tau \in [t-T, t]} |g(x(\tau)) - g(x_{\varepsilon_n}(\tau))| \\ &< \tfrac{1}{2}\eta + \tfrac{1}{2}\eta = \eta. \end{aligned}$$

Hence we may take limits in the equation and (iii) is obtained. Finally, the monotonicity and (iii) imply (iv).

LEMMA 2.4. Let $\{y_n\} \subset C(\mathbf{R})$ be a sequence of monotone nondecreasing functions which as $n \to \infty$ converge to $y \in C(\mathbf{R})$ uniformly on compact subsets of $\mathbf{R}$. Moreover, suppose that $y_n(-\infty) \to y(-\infty)$ and $y_n(+\infty) \to y(+\infty)$ as $n \to \infty$. Then $\lim_{n \to \infty} \|y_n - y\| = 0$.

Proof. First of all we note that $y(t)$ is monotone nondecreasing. Let $\eta > 0$ be arbitrary. Let $k_1 = k_1(\eta)$ and $k_2 = k_2(\eta)$ be such that

$$y(k_1) \leqslant y(-\infty) + \tfrac{1}{2}\eta, \quad y(k_2) \geqslant y(+\infty) - \tfrac{1}{2}\eta.$$

Let $N_1 = N_1(\eta)$ be such that

$$|y_n(t) - y(t)| < \tfrac{1}{2}\eta \text{ for all } t \in [k_1, k_2] \text{ and for all } n > N_1.$$

Let $N_2 = N_2(\eta)$ be such that

$$|y_n(-\infty) - y(-\infty)| < \tfrac{1}{2}\eta \quad \text{and} \quad |y_n(+\infty) - y(+\infty)| < \tfrac{1}{2}\eta$$

for all $n > N_2$. Then for $n > \max\{N_1, N_2\}$ we have

$$|y_n(t) - y(t)| < \eta \quad \text{for all} \quad t \in \mathbf{R}.$$

Namely for $t \in [k_1, k_2]$ by the definition of N_1, whereas for $t < k_1$

$$y(-\infty) \leqslant y(t) \leqslant y(k_1) \quad \text{and} \quad y_n(-\infty) \leqslant y_n(t) \leqslant y_n(k_1)$$

and so

$$|y_n(t) - y(t)| \leqslant \max\{|y(-\infty) - y_n(k_1)|, |y_n(-\infty) - y(k_1)|\}.$$

Furthermore

$$|y(-\infty) - y_n(k_1)| \leqslant |y(-\infty) - y(k_1)| + |y(k_1) - y_n(k_1)| < \eta$$

and

$$|y_n(-\infty) - y(k_1)| \leqslant |y_n(-\infty) - y(-\infty)| + |y(-\infty) - y(k_1)| < \eta$$

and therefore

$$|y_n(t) - y(t)| < \eta.$$

For $t > k_2$ the reasoning is the same.

Proof of Theorem 2.1. Lemma 2.3 and Lemma 2.4 (notably property iv) imply that if $x \in X_c$ then $x \in X$ and so by Lemma 2.2 we may conclude indeed that X is nonempty. The properties of the elements of X are established in Lemma 2.3.

Remark 2.5. In the exceptional case that $x^* = \gamma g(x^*)$ for some $x^* \in (0, x_\infty)$, Lemma 2.2 holds equally well and limits do have the properties (i)–(iii). But then property (iv) is lacking and so we will no longer obtain convergence in $C(\mathbf{R})$. Moreover the limit set will depend strongly on the position of x_0 relative to x^*.

Theorem 2.1 bears some analogy with well-known theorems on the asymptotic behaviour as $t \to \infty$ of bounded solutions of Volterra integral equations (for instance see [7, section III.7, or 8]). Then it is shown that elements of the ω-limit set satisfy an equation like (2.8) (in general an inhomogeneous version). Our result concerns a special case of the asymptotic behaviour as the forcing function tends to zero while $x = 0$ is a unstable steady state.

One would like to obtain the much stronger result that $\lim_{\varepsilon \downarrow 0} \| x_\varepsilon - x \| = 0$ for some $x \in C(\mathbf{R})$. This can be achieved by showing that the properties (i)–(iv) define x uniquely (which implies that X consists of precisely one function). So we are led to study equation (2.8). In Section 4 it is shown that uniqueness is guaranteed if g fulfills an additional condition. The proof requires some knowledge of solutions of the linearized equation and therefore in Section 3 we gather together some results for the linear analogue of (2.8).

3. POSITIVE SOLUTIONS OF THE LINEAR EQUATION

Applications in population dynamics, such as the epidemic model discussed in Section 1, make it interesting to study the homogeneous linear renewal equation on the whole line

$$x(t) = \int_{-\infty}^{t} x(\tau)\, A(t - \tau)\, \mathrm{d}\tau, \quad -\infty < t < \infty, \tag{3.1}$$

with the kernel A satisfying H_A. Since x then stands for some population size one is particularly interested in solutions which are nonnegative and $O(1)$ for $t \to -\infty$.

This kind of equation on the whole line occurs in probability theory too, but there the starting-point is somewhat different. Then usually the integration is from $-\infty$ to $+\infty$ (i.e. A is not necessarily concentrated on $[0, \infty)$), $\int_{-\infty}^{\infty} A(\tau)\, \mathrm{d}\tau = 1$ and first of all one is interested in solutions which are bounded on $\mathbf{R}$. An important result in this area is that under these hypotheses constants are the only bounded solutions of the homogeneous equation (see [9, and 10, Chap. XI]).

By a solution of (3.1) we mean a continuous function $x(t)$ such that the integral on the right-hand side converges absolutely and the equation (3.1) is satisfied. Since the equation is linear and autonomous (in the sense that if $x(t)$ is a solution then so is every translate $x(t + h)$) exponential functions are good candidates for being solutions. Substitution shows that indeed $\exp(st)$ is a solution, provided that

$$L(s) = 1, \tag{3.2}$$

where L is defined by

$$L(s) = \int_0^{\infty} \mathrm{e}^{-s\tau} A(\tau)\, \mathrm{d}\tau. \tag{3.3}$$

Taking s as a complex variable it follows that L is defined and analytic in a half-plane (in some biological applications A will have compact support and then L is in fact an entire function). If s_0 is a root of (3.2) of multiplicity $k > 1$ then again by substitution one verifies that $t^l \exp(s_0 t)$ is a solution of (3.1) for $l = 1, \ldots, k - 1$. The question arises whether or not all solutions of (3.1) are generated in this manner, i.e. can every solution of (3.1) be written as a linear combination of these elementary solutions? The obvious way to answer this question is to use the Fourier transform. In order to apply Fourier transformation one needs to specify the function class in which solutions are to be found and in particular a growth condition for $|t| \to \infty$ will simplify the analysis considerably. However, from the structure of (3.1), which reflects the fact that the future is determined

by the history, it is clear that we do not need to *impose* a growth bound on $x(t)$ for $t \to +\infty$, but that instead an *a priori* estimate can be found. This idea is worked out in the following lemma.

LEMMA 3.1. Let $x(t)$ be a solution of (3.1) for which $\int_{-\infty}^{0} x(\tau)\, A(t-\tau)\, d\tau$ is bounded for $t \in [0, \infty)$. Let $\beta \geqslant 0$ be such that $L(\beta) < 1$. Then there exists a positive constant $k(\beta)$ such that

$$|x(t)| < k(\beta)\, e^{\beta t}, \quad \text{for} \quad t \geqslant 0.$$

Proof. For $t \geqslant 0$ we can write (3.1) in the form

$$x(t) = \int_0^t x(\tau)\, A(t-\tau)\, dt + \int_{-\infty}^{0} x(\tau)\, A(t-\tau)\, d\tau.$$

Choose any $T > 0$ and multiply the above identity by $\exp(-\beta t)$. Then some manipulations yield the estimate

$$\sup_{0 \leqslant t \leqslant T} |x(t)\, e^{-\beta t}| \leqslant L(\beta) \sup_{0 \leqslant t \leqslant T} |x(t)\, e^{-\beta t}| + \sup_{0 \leqslant t \leqslant T} \left|e^{-\beta t} \int_{-\infty}^{0} x(\tau)\, A(t-\tau)\, d\tau\right|.$$

Since $L(\beta) < 1$, $\beta \geqslant 0$ and $|\int_{-\infty}^{0} x(\tau)\, A(t-\tau)\, d\tau| \leqslant c$ for some $c > 0$, we may write as well

$$\sup_{0 \leqslant t \leqslant T} |x(t)\, e^{-\beta t}| \leqslant \frac{c}{1 - L(\beta)}.$$

The right-hand side does not depend on T and therefore

$$|x(t)| \leqslant \frac{c}{1 - L(\beta)}\, e^{\beta t}, \quad 0 \leqslant t < \infty.$$

Remark 3.2. If A has compact support then the condition $\int_{-\infty}^{0} x(\tau)\, A(t-\tau)\, d\tau$ is bounded on $[0, \infty)$ is fulfilled for every solution. In the general case a sufficient requirement is for instance that $x(t) = 0\,(\exp(\alpha t))$ for $t \to -\infty$ with $\alpha \leqslant 0$ and $L(\alpha) < \infty$. This follows from the estimates

$$\left|\int_{-\infty}^{0} x(\tau)\, A(t-\tau)\, d\tau\right| \leqslant K \int_{-\infty}^{0} e^{\alpha\tau}\, A(t-\tau)\, d\tau = K e^{\alpha t} \int_t^{\infty} e^{-\alpha\tau}\, A(\tau)\, d\tau \leqslant K L(\alpha).$$

THEOREM 3.3. (Titchmarsh). Let $x(t)$ be a solution of (3.1) and suppose that for some $\alpha \leqslant 0$ we have $x(t) = 0\,(\exp(\alpha t))$ for $t \to -\infty$ and $L(\alpha - \varepsilon) < \infty$ for some $\varepsilon > 0$. Then $x(t)$ is a linear combination of the finitely many elementary solutions generated by the roots of (3.2) satisfying $\operatorname{Re} s \geqslant \alpha$.

Proof. By Lemma 3.1 and Remark 3.2 we know that $x(t)$ satisfies an exponential bound for $t \to +\infty$ as well, say $x(t) = 0\,(\exp(\beta t))$, where $\beta \geqslant 0$ is such that $L(\beta) < 1$. We are now in a position to apply a result of Titchmarsh, notably Theorem 146 of [11], which says that x indeed can be written as a linear combination of the elementary solutions generated by the roots of (3.2) satisfying $\alpha \leqslant \operatorname{Re} s \leqslant \beta$. The proof of that theorem uses Fourier transforms in the complex domain in a way similar to the Wiener–Hopf method. It remains to show that there are only finitely many such roots and that there are no roots with $\operatorname{Re} s > \beta$. For this we split (3.2) by considering the real and the imaginary part:

$$h_1(a, b) = \int_0^{\infty} e^{-at} \cos bt A(t)\, dt = 1, \tag{3.4a}$$

$$h_2(a, b) = \int_0^\infty e^{-at} \sin bt A(t)\, dt = 0, \tag{3.4b}$$

where $s = a + ib$.

Since $h_1(a, 0)$ is a monotone decreasing function of a and since $|h_1(a, b)| \leqslant h_1(a, 0)$ there are no roots with $\alpha > \beta$. The Lemma of Riemann–Lebesgue (see [11], p. 11) implies that $h_1(a, b) \to 0$ as $|b| \to \infty$ and this can be shown to hold uniformly with respect to $a \in [\alpha, \beta]$. Hence the roots in the strip $\alpha \leqslant a \leqslant \beta$ lie in fact in a rectangle and then the analyticity of $L(s)$ implies that there are only finitely many.

So far we did not completely exploit the nonnegativity of A. The fact that $h_1(a, 0)$ is a monotone decreasing function of a implies that (3.2) has at most one real root σ which moreover is simple (the constant σ is sometimes called the Malthusian parameter). Since $|h_1(a, b)| < h_1(a, 0)$ if $b \neq 0$, all other roots satisfy $\operatorname{Re} s < \sigma$, and so they generate oscillating solutions which are, as $t \to -\infty$, asymptotically dominant over $\exp(\sigma t)$. This observation immediately leads to the following result.

COROLLARY 3.4. Let $x(t)$ be a *nonnegative* solution of (3.1) such that the conditions of Theorem 3.3 are satisfied. If the real root σ of (3.2) exists and satisfies $\sigma \geqslant \alpha$, then $x(t) = c \exp(\sigma t)$ for some $c \geqslant 0$; else $x(t) \equiv 0$.

If $\gamma \geqslant 1$ then $\sigma \geqslant 0$ exists whereas if $\gamma < 1$ then $\sigma < 0$ may or may not exist depending on the behaviour of $A(t)$ as $t \to \infty$ (if for instance $\gamma < 1$ and $A(t) \sim t^{-2}$ as $t \to \infty$ then σ does not exist). If A has compact support then σ exists in all cases.

We conclude this section with a remark concerning the inhomogeneous equation

$$x(t) = \int_{-\infty}^t x(\tau) A(t - \tau)\, d\tau + f(t), \quad -\infty < t < \infty. \tag{3.5}$$

By substituting the right-hand side for x at the right-hand side and repeating this process indefinitely one can construct the particular solution

$$\tilde{x}(t) = f(t) + \sum_{n=1}^\infty \int_{-\infty}^t f(\tau) A^{n*}(t - \tau)\, d\tau, \tag{3.6}$$

provided the terms make sense and the sum converges. Here $A^{1*}(t) = A(t)$ and

$$A^{(n+1)*}(t) = \int_0^t A^{n*}(\tau) A(t - \tau)\, d\tau, \quad n \geqslant 1.$$

An easy criterion is the following: if $|f(t)| < \psi(t)$ and $\int_0^\infty \psi(t - \tau) A(\tau)\, d\tau < c\psi(t)$ where $c < 1$ then the sum in (3.6) converges absolutely (especially exponential functions are suited to take for ψ). All other solutions of (3.5) are then found by adding to (3.6) an arbitrary solution of the homogeneous equation. However, we emphasize that as well (3.5) may not have a solution at all, for instance because $\int_{-\infty}^t f(\tau) A(t - \tau)\, d\tau$ diverges. For a comprehensive treatment of equations like (3.5) we refer to [9]. In [12] inhomogeneous nonlinear equations are analysed under different hypotheses.

4. UNIQUENESS BY GEOMETRICAL ARGUMENTS

In the foregoing section it was found that the linear equation (3.1) has in general many solutions and so we expect this to be the case for the nonlinear equation

$$x(t) = \int_{-\infty}^{t} g(x(\tau))\, A(t-\tau)\, \mathrm{d}\tau, \quad -\infty < t < \infty, \tag{4.1}$$

with A and g satisfying H_A and H_g respectively, as well. Since the analysis in Section 3 led to the conclusion that by imposing some further, biologically significant, conditions one particular solution of (3.1) is selected, we wonder whether the same is true for (4.1).

Suppose $x < \gamma g(x)$ on $(0, x_\infty)$ then we know by Theorem 2.1 that a monotone nondecreasing solution satisfying

$$0 < x(t) < x_\infty \tag{4.2}$$

exists and the question arose whether or not these qualifications imply uniqueness (apart from translation). As a first remark we observe that if two solutions coincide for $-\infty < t \leqslant t_0$ then they do so for all t.

It turns out that we can make some progress by imposing an additional condition on the function g. For ease of formulation we first introduce the notion of sublinearity (*cf.* [13]).

Definition. The function g is called *sublinear* on $[0, \bar{x}]$ if $g(\alpha x) \geqslant \alpha g(x)$ for every $x \in [0, \bar{x}]$ and every $\alpha \in [0, 1]$.

This property can be geometrically formulated by saying that for every $x \in (0, \bar{x}]$ the graph of g on the interval $(0, x)$ lies above the straight line through 0 and $g(x)$. Note that every concave function g with $g(0) \geqslant 0$ is sublinear.

THEOREM 4.1. Let $x_1(t)$ and $x_2(t)$ satisfy (4.1), (4.2) and $x_1(0) = x_2(0)$. In addition to H_A, H_g assume that g is sublinear on $[0, x_\infty]$. Then either $x_1(t) \equiv x_2(t)$ or $x_1(t)/x_2(t)$ does assume neither a minimum nor a maximum on $(-\infty, 0]$.

Proof. First of all we note that a maximum of $x_1(t)/x_2(t)$ is a minimum of $x_2(t)/x_1(t)$ and so it is sufficient to consider the minimum only. Let

$$\alpha = \inf_{-\infty < t \leqslant 0} \frac{x_1(t)}{x_2(t)},$$

then $0 \leqslant \alpha \leqslant 1$, and so

$$\begin{aligned} x_1(t) &= \int_{-\infty}^{t} g(x_1(\tau))\, A(t-\tau)\, \mathrm{d}\tau \geqslant \int_{-\infty}^{t} g(\alpha x_2(\tau))\, A(t-\tau)\, \mathrm{d}\tau \\ &\geqslant \alpha \int_{-\infty}^{t} g(x_2(\tau))\, A(t-\tau)\, \mathrm{d}\tau = \alpha x_2(t). \end{aligned}$$

Let $V = \{t \leqslant 0 \mid x_1(t) = \alpha x_2(t)\}$ and suppose $t_0 \in V$. Then the above inequalities are in fact equalities for $t = t_0$. Concerning the first inequality this yields as a necessary condition that $t_0 - \tau \in V$ for all τ in the support of A. For some $\tau_1 \geqslant 0$ and some $\delta > 0$ the support of A contains the interval

$[\tau_1, \tau_1 + \delta]$. Hence we may conclude that $[t_0 - \tau_1 - \delta; t_0 - \tau_1] \subset V$. Repeating the same reasoning n times we deduce $[t_0 - n\tau_1 - n\delta, t_0 - n\tau_1] \subset V$. As soon as $n\delta > \tau_1$ two successive intervals are overlapping, and we arrive at the conclusion that V contains an interval of the form $(-\infty, t_1]$. Now suppose $\alpha < 1$. Then $x_1(t) > x_2(t)$ for $-\infty < t \leqslant t_1$ and consequently $x_1(t) > x_2(t)$ for all t. This clearly contradicts the assumption $x_1(0) = x_2(0)$ and so V must be empty if $\alpha < 1$. Finally, if $\alpha = 1$ then $0 \in V$ and so $x_1(t) = x_2(t)$ for $-\infty < t \leqslant t_1$. This implies $x_1(t) \equiv x_2(t)$.

Unfortunately we cannot conclude uniqueness from the result of Theorem 4.1 because of the noncompactness of $(-\infty, 0]$. However, if we can show that x_1/x_2 approaches a limit as t tends to $-\infty$ then the possibility that x_1/x_2 assumes neither a maximum nor a minimum on $(-\infty, 0]$ is excluded, and we have attained our end. It is here that linearization plays a role. As a side-step we observe that some of the arguments in the proof of Theorem 4.1 can be used to show that any monotone nondecreasing solution is in fact monotone increasing.

THEOREM 4.2. Let $x(t)$ be a monotone nondecreasing solution of (4.1) satisfying (4.2). In addition to H_A, H_g suppose that $L(-\varepsilon) < \infty$ fo some $\varepsilon > 0$, that $g'(0)$ exists and that $\gamma g'(0) > 1$. Then $\lim_{t \to -\infty} x(t)\exp(-\sigma t) = c$ for some positive constant c, where $\sigma > 0$ is defined by

$$g'(0) \int_0^\infty e^{-\sigma\tau} A(\tau)\, d\tau = 1.$$

Proof. Let $\{x_h \mid -\infty < h \leqslant 0\}$ be the family of functions of t, $-\infty < t \leqslant 0$, defined by

$$x_h(t) = \frac{x(t+h)}{x(h)}.$$

As a first step we analyse the limit set of $\{x_h\}$ as $h \to -\infty$. This is done in exactly the same way as the limit set of $\{x_\varepsilon\}$ was analysed in Theorem 2.1, so we omit the details. The monotonicity implies the uniform boundedness and by making use of the equation one verifies the equicontinuity. So the limit set with respect to the topology of uniform convergence on compact subsets of $(-\infty, 0]$ is nonempty. Let y be an element of this set, then $y(t)$ satisfies the linearized equation

$$y(t) = g'(0) \int_{-\infty}^{t} y(\tau) A(t - \tau)\, d\tau.$$

Moreover $y(t) > 0$, $y(0) = 1$ and thus by the results of Section 3, $y(t) = \exp(\sigma t)$. Note that $x \leqslant \gamma g(x)$ on $(0, x_\infty)$ implies $\gamma g'(0) \geqslant 1$ and consequently $\sigma \geqslant 0$; the assumption $\gamma g'(0) > 1$ guarantees that $\sigma > 0$. Hence, using the monotonicity as in Lemma 2.4, we deduce

$$\lim_{h \to -\infty} \frac{x(t+h)}{x(h)} = e^{\sigma t}, \text{ uniformly on } (-\infty, 0].$$

Putting $x(t) = \exp(\sigma t + z(t))$ we obtain that $z(t+h) - z(h) \to 0$ as $h \to -\infty$ uniformly on $(-\infty, 0]$. So for every $\varepsilon > 0$ there exists a $T(\varepsilon)$ such that for $t_1, t_2 < T(\varepsilon)$, $|z(t_1) - z(t_2)| < \varepsilon$. We draw the conclusion that $z(t)$ approaches a finite limit as t tends to $-\infty$.

Before we combine Theorems 4.1 and 4.2 we make some observations. In case g is sublinear $g(x)/x$ is monotone nonincreasing and consequently $g'(0)$ exists. In Theorem 4.1 we assumed g to be

sublinear on $[0, x_\infty]$ but, since in the proof the solutions are considered for $-\infty < t \leqslant 0$ only, the same result holds if g is assumed to be sublinear in a right-hand neighbourhood of $x = 0$ only. Note that this will be the case if for instance $g''(0) < 0$.

COROLLARY 4.3. In addition to H_A, H_g let g be sublinear on $[0, \bar{x}]$ for some $\bar{x} > 0$, $L(-\varepsilon) < \infty$ for some $\varepsilon > 0$, and $\gamma g'(0) > 1$, then there is (apart from translation) one and only one monotone nondecreasing solution of (4.1) satisfying (4.2). This solution is in fact monotone increasing.

Proof. The existence was established in Theorem 2.1. Let $x_1(t)$ and $x_2(t)$ be two such solutions, then by Theorem 4.2 $x_1(t) \sim c_1 \exp(\sigma t)$ and $x_2(t) \sim c_2 \exp(\sigma t)$ for $t \to -\infty$, with $c_1, c_2 > 0$. Therefore $\lim_{t \to -\infty} x_1(t)/x_2(t) = c_1/c_2$. Finally, by Theorem 4.1, $x_1(t) \equiv x_2(t)$ if $x_1(0) = x_2(0)$.

5. INTERPRETATION OF THE RESULTS

In Section 4 we found some conditions on g which guarantee that the limit set X, defined in (2.7), consists of precisely one function. As a consequence we have $\lim_{\varepsilon \downarrow 0} \| x_\varepsilon - x \| = 0$ in that case, or, in other words, $x(t; \varepsilon)$ tends in a well-defined translation sense to a well-defined nontrivial limit. Note that the limit does not depend either on the choice of x_0 inbetween 0 and x_∞ or on the particular $f(t; \varepsilon)$ satisfying H_f.

For the special function $g(x)$ defined in (2.2) the conditions are fulfilled and so our results apply to the epidemic problem. In biological terms we can say that as we let the influence of the prescribed history tend to zero then the epidemic curve approaches a characteristic shape. This shape can be found by computing a certain nontrivial solution, defined for $-\infty < t < \infty$, of the limit equation

$$\dot{S}(t) = S(t) \int_{-\infty}^{t} \dot{S}(\tau)\, A(t - \tau)\, \mathrm{d}\tau$$

describing the interaction process underlying the model. In a sense the dynamics of the epidemic are much better described by the limit equation since no arbitrary choice concerning an initial time $t = 0$ and the history up to $t = 0$ has to be made.

Though uniqueness is lacking for the limit equation it was shown in this paper that it does define a *biological* solution uniquely. Let $\bar{S}$ denote the total population size, then this biological solution is given by $S(t) \equiv \bar{S}$ (no epidemic) if $\gamma \bar{S} \leqslant 1$, whereas if $\gamma \bar{S} > 1$ it is nontrivial (an epidemic 'occurs'). So it is demonstrated again that the parameter $\gamma \bar{S}$ has a threshold value 1.

6. REMARKS ON GENERALIZATIONS

The hypotheses H_g, H_A and H_f for (2.1) were motivated by the epidemic model. The monotonicity of the solution leads to a straightforward calculation of the limit as $t \to \infty$, and therefore the nonuniform convergence, as the forcing function f tends to zero, could easily be detected.

In the study of nonlinear Volterra integral equations much attention has been devoted to the boundedness of solutions and the question whether all bounded solutions tend to limits and how these limits can be characterized (see [14] and the references given there). As soon as for a given class of forcing functions it is known that the solution approaches a certain limit (for instance in [14] a population growth model is discussed and conditions are given such that every solution tends to the same limit as $t \to \infty$) the analysis of this paper can be repeated and the analogue of Theorem 2.1 is easily obtained. In general it is difficult to analyse the limit set any further.

Geometrical conditions in the spirit of Section 4 seem to be far from necessary, but to the author no other method is known. If monotonicity is not part of the qualifications the situation is much more awkward.

Acknowledgement—We would like to thank J. A. J. Metz for stating the conjecture which led to this investigation, L. A. Peletier for many very stimulating discussions and J. W. de Roever for clarifying conversations on the subject of Fourier transformation.

REFERENCES

1. KERMACK W. O. & MCKENDRICK A. G., A contribution to the mathematical theory of epidemics, *Proc. R. Soc.* **A115**, 700–721 (1927).
2. HOPPENSTEADT F., Mathematical Theories of Populations: Demographics, Genetics and Epidemics, *SIAM Regional Conference Series in Applied Mathematics*, Vol. 20 (1975).
3. REDDINGIUS J., Notes on the mathematical theory of epidemics, *Acta Biotheor.* **20**, 125–157 (1971).
4. KENDALL D. G., Deterministic and stochastic epidemics in closed populations, *Proc. Third Berkeley Symposium* (Edited by J. NEYMAN) **4**, pp. 149–165. University of California Press (1956).
5. BRAUER F., Perturbations of the nonlinear renewal equation, *Adv. Math.* **22**, 32–51 (1976).
6. GRASMAN J. & MATKOWSKY B. J., Evolution of an epidemic with a small number of initial infectives, *Math. Centre Report* TW 162 (1976).
7. MILLER R. K., *Nonlinear Volterra Integral Equations*. Benjamin, Menlo Park (1971).
8. LEVIN J. J. & SHEA D. F., On the asymptotic behaviour of the bounded solutions of some integral equations, *J. math. Analysis Applic.* **37**, 42–82, 288–326, 537–575 (1972).
9. KARLIN S., On the renewal equation, *Pacif. J. Math.* **5**, 229–257 (1955).
10. FELLER B., *An Introduction to Probability Theory and Its Applications*, Vol. II. Wiley, New York (1966).
11. TITCHMARSH E. C., *Introduction to the Theory of Fourier Integrals*. Clarendon Press, Oxford (1937).
12. LEITMAN M. J. & MIZEL V. J., Hereditary laws and nonlinear integral equations on the line, *Adv. Math.* **22**, 220–267 (1976).
13. AMANN H., On the number of solutions of nonlinear equations in ordered Banach Spaces, *J. funct. Analysis* **11**, 346–384 (1972).
14. BRAUER F., On a nonlinear integral equation for population growth problems, *SIAM J. math. Analysis* **6**, 312–317 (1975).
15. METZ, J. A. J., The epidemic in a closed population, to appear in *Acta Biotheor.*

Thresholds and travelling waves for the geographical spread of infection

by

O. Diekmann

ABSTRACT

A nonlinear integral equation of mixed Volterra-Fredholm type describing the spatio-temporal development of an epidemic is derived and analysed. Particular attention is paid to the hair-trigger effect and to the travelling wave problem.

1. INTRODUCTION

An important approach in mathematical population dynamics is to investigate which qualitative phenomena one can understand on the basis of some simple assumptions concerning an interaction mechanism. Kendall expressed this once as follows [12]: "I expect the usable results to be not numbers but a list of possible modes of behaviour". In this spirit we study in the present paper some qualitative aspects of the development of an epidemic in space and time.

In 1927, KERMACK and McKENDRICK [13] proposed a deterministic model for the evolution in time of a contageous disease in a closed population. The model leads to a nonlinear Volterra integral equation which can be analysed in great detail (see [13], [14] and [4]). Among other things, the analysis reveals that even in the long run a nontrivial fraction of the susceptible population escapes from getting the disease. This fraction is found from a root of a nonlinear scalar equation; some important conclusions can be drawn from the dependence of the relevant root on the various parameters.

A special case of the Kermack and McKendrick model leads to a system of ordinary differential equations. KENDALL [11] has introduced a space-dependent analogue of this system by taking for the infectivity at each point a weighted spatial average of the density of infectives. In this case, the final fraction of susceptibles is a function of position; it is found from a solution of a nonlinear integral equation. On the basis of this equation Kendall made it plausible that, for certain values of the parameters, the ultimate effect of the introduction of an arbitrary small amount of infectivity at some point would be large at every point of the spatial domain. The system exhibits therefore what ARONSON and WEINBERGER [2] call the hair-trigger effect. In this paper we consider the space-dependent analogue of the general Kermack and McKendrick model and we prove the hair-trigger effect for one - and two-dimensional spatial domains.

A topic that has received considerable attention is the travelling wave problem. As far as we know, all work done so far takes the Kendall system as a starting point. MOLLISON [15] proves the existence of travelling waves for a special weight function which enables him to reduce the problem to a system of ordinary differential equations and then to use invariant manifold

theory. KENDALL [12] arrives at a system of ordinary differential equations by replacing the spatial average by a diffusion approximation (see also [8] and [10]). It seems that ATKINSON and REUTER [3] were the first to prove existence and nonexistence of travelling waves by analyzing the integro-differential system itself. In this paper we deal with the travelling wave problem for the general Kermack and McKendrick model. This problem leads us to an integral equation which can be treated in the style of Atkinson and Reuter.

As a general introduction to the field of mathematical epidemiology we refer to the lecture notes of HOPPENSTEADT [10]. As far as mathematics is concerned, the main subject of this paper is the study of nonlinear convolution equations on the real line. This work was begun in [4] and will be continued in [5].

2. THE MODEL

Consider a population living in a habitat Ω (a closed subset of $\mathbb{R}^n$) and susceptible to some contageous disease. Our first aim is to give a heuristic derivation of an equation for the evolution of an epidemic among the population on the basis of some simple assumptions. The only dynamical phenomenon that we are going to consider is infection. So, according to the custom to study but one interaction mechanism at a time, we simply ignore the effect of changes due to birth, migration, etcetera. We assume that space dependence manifests itself through the possibility that an infectious individual at position x_1 can infect a susceptible individual at position x_2. In this connection one may think of such phenomena as forest-fires, the wind borne dissemination of spores that cause plant-diseases, or contagion among animals (or people) walking around by day and returning to their home for the night.

Let $S(t,x)$ and $I(t,x)$ denote the density of *susceptibles* and *infectives*, respectively, at time t and position x. Let $i(t,\tau,x)d\tau$ be the density of infectives which were infected some time between $t-\tau$ and $t-\tau-d\tau$. Then,

$$I(t,x) = \int_0^\infty i(t,\tau,x)d\tau. \tag{2.1}$$

We suppose that the population size is large, so we can consider the variables S, I and i as continuous, or even continuously differentiable, real-valued functions of their arguments.

Let the *infectivity* $B = B(t,x)$ be defined as the rate at which susceptibles become infective. The basic assumptions in this paper are that

(i) the disease induces permanent immunity, so the transition from I to S does not occur,

(ii) a nonnegative function $A = A(\tau,x,\xi)$ is given such that

$$B(t,x) = \int_0^\infty \int_\Omega i(t,\tau,\xi)A(\tau,x,\xi)d\xi d\tau. \tag{2.2}$$

Thus, $A(\tau,x,\xi)$ describes the infectivity at x due to one infective of "age of illness" τ at ξ. Many characteristics of both the disease and the habitat are in fact incorporated in A.

The above definitions and assumptions lead to the following system of dynamical equations:

$$\frac{\partial S}{\partial t}(t,x) = -S(t,x)B(t,x), \tag{2.3}$$

$$\frac{\partial S}{\partial t}(t,x) = -i(t,0,x), \tag{2.4}$$

$$i(t,\tau,x) = i(t-\tau,0,x). \tag{2.5}$$

Elimination of i from (2.2) by means of (2.4) and (2.5), followed by substitution of the result into (2.3) yields an equation for S alone,

$$\frac{\partial S}{\partial t}(t,x) = S(t,x)\int_0^\infty \int_\Omega \frac{\partial S}{\partial t}(t-\tau,\xi)A(\tau,x,\xi)d\xi d\tau. \tag{2.6}$$

A solution S of (2.6) has to be defined at least for $-\infty < t \le T$ for some finite T. In section 6, equation (2.6) will be taken as the starting point for an analysis of travelling wave solutions.

An appropriate *initial value problem* is obtained by prescribing

$$i(0,\tau,x) = i_0(\tau,x) \text{ and } S(0,x) = S_0(x)$$

and by restricting the validity of the dynamical equations to $t > 0$. Then, instead of (2.6), we obtain

$$\frac{\partial S}{\partial t}(t,x) = S(t,x)\left\{\int_0^t\int_\Omega \frac{\partial S}{\partial t}(t-\tau,\xi)A(\tau,x,\xi)d\xi d\tau - h(t,x)\right\}, \tag{2.7}$$

where

$$h(t,x) = \int_0^\infty\int_\Omega i_0(\tau,\xi)A(t+\tau,x,\xi)d\xi d\tau. \tag{2.8}$$

Finally, assuming $S_0(x) > 0$ for every $x \in \Omega$, we integrate equation (2.7) once with respect to t. If we interchange the order of the integrations, we can write the resulting equation in the form

$$u(t,x) = \int_0^t\int_\Omega g(u(t-\tau,\xi))S_0(\xi)A(\tau,x,\xi)d\xi d\tau + f(t,x), \tag{2.9}$$

where

$$u(t,x) = -\ln\frac{S(t,x)}{S_0(x)}, \tag{2.10}$$

$$g(y) = 1 - e^{-y}, \tag{2.11}$$

$$f(t,x) = \int_0^t h(\tau,x)d\tau. \tag{2.12}$$

In the next section we shall discuss the problems of existence, uniqueness and asymptotic behaviour of solutions of equation (2.9). To some extent the specific form of the forcing function f will be irrelevant; all we shall need is the fact that f is a given nonnegative function which is monotonic nondecreasing with respect to t.

To conclude this section we make some remarks concerning the biological aspects of the model and we discuss some special cases.

REMARK 2.1. If Ω is bounded, the influence of the boundary has to be incorporated in the function A. How this is done depends on the underlying mech-

anism. For instance, if Ω is a cornfield and the wind is acting as a driving force for the spread of spores, then the boundary $\partial\Omega$ has no active influence on the spread and all spores falling outside Ω are simply without any effect. However, if Ω is a town, then individuals living near the boundary are attracted by the centre and, consequently, A will include anisotropic boundary effects. Likewise, internal inhomogeneities of Ω (mountains, rivers) can give rise to anisotropy in A.

REMARK 2.2. Another feature of A is its coupling of the age of illness τ and the space variables x and ξ. Again, this coupling varies with the type of the underlying mechanism. An important special case arises if we assume that there is no coupling at all,

$$A(\tau,x,\xi) = H(\tau)W(x,\xi) \quad , \tag{2.13}$$

where we can normalize so that

$$\int_{\Omega} W(x,\xi)dx = 1 \quad . \tag{2.14}$$

The situation can be simplified even further if we assume that the function W depends on the distance only,

$$W(x,\xi) = V(x-\xi) \quad , \tag{2.15}$$

where V is a radial function.

REMARK 2.3. Two special cases of equation (2.7) have received much attention in the literature. Both cases have in common that $\Omega = \mathbb{R}^n$, that (2.13) - (2.15) hold, and that $i_0(\tau,x) = j_0(x)\delta(\tau)$ (introduction of fresh infectives only; δ is the Dirac delta function), with $S_0(x) + j_0(x) = 1$ (constant population density). With these assumptions (2.7) can be written as

$$\frac{\partial S}{\partial t}(t,x) = S(t,x)\int_{\mathbb{R}^n}\left\{\int_0^t \frac{\partial S}{\partial t}(t-\tau,\xi)H(\tau)d\tau - H(t)j_0(\xi)\right\}V(x-\xi)d\xi.$$

Two special cases are:

a) *a simple epidemic* ([3], [15], [16]): $H(t) \equiv a$.

Then

$$\frac{\partial S}{\partial t}(t,x) = aS(t,x)\{\bar{S}(t,x) - 1\} \quad ,$$

where $\bar{S}$ is a weighted spatial average of S,

$$\bar{S}(t,x) = \int_{\mathbb{R}^n} S(t,\xi)V(x-\xi)d\xi.$$

b) *a general epidemic* ([3], [8], [10], [11], [12]): $H(t) = ae^{-\mu t}$.

Define

$$Y(t,x) = -\int_0^t \frac{\partial S}{\partial t}(\tau,x)e^{-\mu(t-\tau)}d\tau + j_0(x)e^{-\mu t}.$$

Then Y and S satisfy the system of equations

$$\frac{\partial Y}{\partial t}(t,x) = -\frac{\partial S}{\partial t}(t,x) - \mu Y(t,x) \quad ,$$

$$\frac{\partial S}{\partial t}(t,x) = -aS(t,x)\bar{Y}(t,x) \quad ,$$

where $\bar{Y}$ is the weighted spatial average of Y, as defined above. This system admits another interpretation: Y is the density of the infectives, every infective has constant infectivity a, and the infectives are removed with rate μ.

3. THE INITIAL VALUE PROBLEM

The nonlinear integral equation

$$u(t,x) = \int_0^t \int_\Omega g(u(t-\tau,\xi))S_0(\xi)A(\tau,x,\xi)d\xi d\tau + f(t,x), \tag{3.1}$$

is of Volterra type with respect to t and of Fredholm type with respect to x. In making the assumptions concerning A, g, S_0 and f we shall be guided by the biological interpretation and by the desire to let the analysis proceed along standard lines. We do not aim at generality.

Let $BC(\Omega)$ denote the Banach space of the bounded continuous functions on Ω equipped with the supremum norm. Then a convenient framework for the study of (3.1) is provided by the Banach space $C_T = C([0,T];BC(\Omega))$ of continuous functions on $[0,T]$ with values in $BC(\Omega)$, equipped with the norm

$$\|f\|_{C_T} = \sup_{0\le t\le T} \|f[t]\|_{BC(\Omega)} \quad .$$

When looking at u as an element of C_T we shall write $u[t](x)$ instead of $u(t,x)$. With this convention we can write (3.1) as

$$(3.2) \qquad u[t] = Qu[t] + f[t] \quad ,$$

where Q is defined by

$$(3.3) \qquad Qu[t](x) = \int_0^t \int_\Omega g(u[t-\tau](\xi))S_0(\xi)A(\tau,x,\xi)d\xi d\tau .$$

As will be proved in Lemma 3.1 below, the following assumptions guarantee that Q is a mapping from C_T into C_T.

A_{S_0} : $S_0 \in L_\infty(\Omega)$; S_0 is nonnegative.

A_g : $g:\mathbb{R} \to \mathbb{R}$ is continuous ; $g(0) = 0$.

A_A^1 : $A(\cdot,\cdot,\cdot)$ is defined and nonnegative on $[0,\infty)\times\Omega\times\Omega$; for every $x \in \Omega$ and every $T > 0$, $A(\cdot,x,\cdot) \in L_1([0,T]\times\Omega)$.

A_A^2 : Let $\eta(t,x) = \int_0^t \int_\Omega A(\tau,x,\xi)d\xi d\tau$ and let $T > 0$ be arbitrary, then the family of functions on $[0,T]$, $\{\eta(\cdot,x) \mid x \in \Omega\}$ is uniformly bounded and equicontinuous.

A_A^3 : For every $T > 0$ and every $\varepsilon > 0$ there exists $\delta = \delta(\varepsilon,T) > 0$ such that if $x_1, x_2 \in \Omega$ and $|x_1 - x_2| < \delta$ then

$$\int_0^T \int_\Omega |A(\tau,x_1,\xi) - A(\tau,x_2,\xi)| d\xi d\tau < \varepsilon.$$

For notational convenience we define

$$\bar{S}_0 = \sup_{x\in\Omega} S_0(x) \quad .$$

<u>LEMMA</u> 3.1. *For every* $T > 0$ *and every* $u \in C_T$, $Qu \in C_T$.

<u>PROOF</u>. Let $T > 0$ and $u \in C_T$ be arbitrary and put $k = \|u\|_{C_T}$. First we prove that for fixed $t \in [0,T]$, $Qu[t] \in BC(\Omega)$. The boundedness follows from

$$|Qu[t](x)| \le \sup_{|y|\le k} |g(y)|\cdot\bar{S}_0\cdot \int_0^T \int_\Omega A(\tau,x,\xi)d\xi d\tau \quad ,$$

and the uniform boundedness assumption in A_A^2. The continuity with respect to x follows from

$$|Qu[t](x_1) - Qu[t](x_2)| \le \sup_{|y|\le k} |g(y)|\cdot\bar{S}_0\cdot \int_0^T \int_\Omega |A(\tau,x_1,\xi) - A(\tau,x_2,\xi)| d\xi d\tau$$

and A_A^3.

Next we show that the mapping $t \mapsto Qu[t]$ from $\mathbb{R}$ to $BC(\Omega)$ is continuous. Suppose $0 \le t_2 \le t_1 \le T$, then

$$\sup_{x\in\Omega} | Qu[t_1](x) - Qu[t_2](x) | \le$$

$$\bar{S}_0 \sup_{x\in\Omega} \int_0^T \int_\Omega A(\tau,x,\xi)d\xi d\tau\cdot \sup_{\substack{\xi\in\Omega \\ \tau\in[0,t_2]}} |g(u[t_1 - t_2 + \tau](\xi)) - g(u[\tau](\xi))| +$$

$$\sup_{|y|\le k} |g(y)|\cdot\bar{S}_0\cdot\sup_{x\in\Omega} \int_{t_2}^{t_1} \int_\Omega A(\tau,x,\xi)d\xi d\tau.$$

If $t_1 - t_2 \downarrow 0$ then both terms tend to zero. For the first term this follows from A_A^2 and the uniform continuity of $t \mapsto u[t]$ on $[0,T]$ and $y \mapsto g(y)$ on $|y| \le k$, and for the second term this follows from the equicontinuity assumption in A_A^2. $\square$

So if we assume that f satisfies

A_f: $\qquad f: [0,\infty) \to BC(\Omega)$ is continuous,

then equation (3.2) makes sense in C_T for arbitrary $T > 0$.

<u>THEOREM</u> 3.2. *(local existence and uniqueness).*
Suppose g *is locally Lipschitz continuous, then there exists a* $T > 0$ *such that (3.2) has a unique solution* u *in* C_T *and the mapping* $f \mapsto u$ *is continuous from* C_T *into* C_T.

<u>PROOF</u>. Consider the closed subset of C_T

$$X_{f,T} = \{h \in C_T \mid \|h - f\|_{C_T} \le 1\}$$

and let the mapping $R_f: C_T \to C_T$ be defined by

$$R_f u = Qu + f.$$

Choose $T_0 > 0$ and let ℓ be the Lipschitz constant of g corresponding to the set $|y| \le 1 + \|f\|_{C_{T_0}}$. Since $\|f\|_{C_T}$ is a nondecreasing function of T we have for all $T < T_0$ and $u \in X_{f,T}$,

$$\|u\|_{C_T} \le 1 + \|f\|_{C_{T_0}} .$$

Let $u,v \in X_{f,T}$ then

$$\|Qu\|_{C_T} \le \ell \bar{S}_0 \sup_{x\in\Omega} \int_0^T \int_\Omega A(\tau,x,\xi)d\xi d\tau. \ (1 + \|f\|_{C_{T_0}}),$$

and

$$\|R_f u - R_f v\|_{C_T} \le \ell \bar{S}_0 \sup_{x\in\Omega} \int_0^T \int_\Omega A(\tau,x,\xi)d\xi d\tau . \ \|u - v\|_{C_T} .$$

So if

$$\sup_{x\in\Omega} \int_0^T \int_\Omega A(\tau,x,\xi)d\xi d\tau < (\ell\bar{S}_0(1 + \|f\|_{C_{T_0}}))^{-1} \quad ,$$

then R_f is a contraction mapping of $X_{f,T}$ into itself. From A_A^2 we conclude that we can choose indeed $T \in (0,T_0]$ such that the inequality holds and then the first half of the theorem follows from the Banach contraction mapping theorem. This existence proof is constructive: if $u_0 = f$ and $u_{n+1} = Qu_n + f$, then u_n converges as $n \to \infty$ in C_T to the solution u.

In order to prove the continuity of the mapping $f \mapsto u$ one needs to modify the foregoing proof just a little to show that for a given $f \in C_{T_0}$ there exist a neighbourhood Y of f in C_{T_0}, a $T \in (0,T_0]$ and a closed subset $X_{f,T}$ such that $R_{\tilde{f}}$ is a contraction mapping of $X_{f,T}$ into $X_{f,T}$ uniformly for $\tilde{f} \in Y$ (we leave the details of this modification to the reader). Clearly the mapping $f \mapsto R_f u$ is continuous on C_T for each $u \in C_T$ and, consequently, the continuous dependence of the fixed point on f follows from [9, Theorem 0.3.2]. We remark that one can also prove the continuous dependence by making use of Gronwall's inequality [9, Corollary I.6.6] and a slightly different assumption concerning A. □

Once we know a solution of (3.2) in C_{T_1} we can write

$$\tilde{u}[t] = Q\tilde{u}[t] + \tilde{f}[t] \quad ,$$

with

$$\tilde{u}[t] = u[T_1 + t] \quad ,$$

$$\tilde{f}[t](x) = f[T_1 + t](x) + \int_0^{T_1} \int_\Omega g(u[\tau](\xi))S_0(\xi)A(T_1 + t-\tau,x,\xi)d\xi d\tau .$$

Application of Theorem 3.2 then yields a continuation of u in C_{T_2}, $T_2 > T_1$, and by repetition of the argument one finds a maximally defined solution (compare [9, section I.2]).

THEOREM 3.3. *(global existence and uniqueness).*
Suppose g *is uniformly Lipschitz continuous then (3.2) has a unique continuous solution* $u: [0,\infty) \to BC(\Omega)$.

PROOF. In the continuation procedure one can choose $T_{n+1} - T_n$ independent of n. □

We now turn our attention to some further properties of the solution that arise as a consequence of the structure of the forcing function f and the properties of g in the biological model. On $BC(\Omega)$ we introduce a partial ordering by

$$\phi \geq \psi \text{ if and only if } \phi(x) \geq \psi(x) \text{ for all } x \in \Omega.$$

THEOREM 3.4. *(positivity and monotonicity).*

(a) Suppose $g(y) > 0$ *for* $y > 0$ *and* $f[t] \geq 0$ *for all* $t \geq 0$, *then* $u[t] \geq 0$ *on the domain of definition of* u.

(b) Suppose, in addition, g *is monotone nondecreasing and* $f[t+h] \geq f[t]$ *for all* $h \geq 0$, *then* $u[t+h] \geq u[t]$ *for all* $h \geq 0$ *and* $t \geq 0$ *such that* $t + h$ *is in the domain of definition of* u.

PROOF. Both (a) and (b) follow from the construction of the solution (see the proof of Theorem 3.2). □

If f is given by (2.12) and g by (2.11) then they have the properties that are assumed in Theorem 3.4. The conclusion is biologically evident, since we assumed that the disease induces permanent immunity.

Next we discuss the boundedness and the asymptotic behaviour as $t \to \infty$, of a globally defined solution. If g, f and η (see A_A^2) are bounded from above, then a simple estimate shows that u is bounded from above. In combination with Theorem 3.4(a), (b) this yields the convergence of $u[t](x)$ to a finite limit as $t \to \infty$, pointwise in x. By imposing further conditions on A and f we can strengthen the convergence and we can deduce an equation for the limit.

THEOREM 3.5. *(stabilization).*
In addition to the assumptions of Theorem 3.4(a), (b), suppose that g *is bounded and uniformly Lipschitz continuous on* $[0,\infty)$, *that the subset*

$\{f[t] \mid t \geq 0\}$ *of* $BC(\Omega)$ *is uniformly bounded and equicontinuous and that* A *satisfies*

A_A^4: *For each* $x \in \Omega$

$$\int_0^t A(\tau,x,\cdot)d\tau \to \int_0^\infty A(\tau,x,\cdot)d\tau \text{ in } L_1(\Omega) \text{ as } t \to \infty,$$

and for some $C > 0$

$$\sup_{x\in\Omega} \int_\Omega \int_0^\infty A(\tau,x,\xi)d\tau d\xi < C.$$

A_A^5: *For each* $\varepsilon > 0$ *there exists* $\delta = \delta(\varepsilon) > 0$ *such that if* $x_1,x_2 \in \Omega$ *and* $|x_1 - x_2| < \delta$ *then*

$$\int_\Omega \int_0^\infty |A(\tau,x_1,\xi) - A(\tau,x_2,\xi)|\, d\tau d\xi < \varepsilon.$$

Then the solution u *of (3.2) is defined on* $[0,\infty)$ *and there exists* $u[\infty] \in BC(\Omega)$ *such that, as* $t \to \infty$, $u[t] \to u[\infty]$ *in* $BC(\Omega)$ *if* Ω *is compact, and uniformly on compact subsets of* Ω *if* Ω *is not compact. Moreover* $u[\infty]$ *satisfies the limit equation*

$$u[\infty](x) = \int_\Omega g(u[\infty](\xi))S_0(\xi) \int_0^\infty A(\tau,x,\xi)d\tau d\xi + f[\infty](x). \tag{3.4}$$

<u>PROOF</u>. The global existence follows from Theorem 3.4(a) and the uniform Lipschitz continuity of g on $[0,\infty)$. We have noticed before that $u[t](x)$ converges pointwise in x. The estimate

$$|u[t](x_1) - u[t](x_2)| \leq \sup_{y\geq 0} g(y)\cdot\bar{S}_0\cdot \int_\Omega \int_0^\infty |A(\tau,x_1,\xi) - A(\tau,x_2,\xi)|d\tau d\xi +$$
$$+ |f[t](x_1) - f[t](x_2)| \quad ,$$

shows that the subset $\{u[t] \mid t \geq 0\}$ of $BC(\Omega)$ is equicontinuous. Hence the Arzela-Ascoli theorem can be applied to strengthen the convergence as stated in the theorem. We leave it to the reader to write down the straightforward

but rather tedious estimates that show that one may take limits in the equation in order to get (3.4). □

In biological terms $u[\infty]$ is the fraction of the susceptible population that escapes from getting the disease. Therefore we call $u[\infty]$ the *final size* and (3.4) the *final size equation*.

The assumptions conserning A are rather technical. However, in the special case that (2.13) - (2.15) hold they simplify considerably: if $\Omega = \mathbb{R}^n$, $V \in L_1(\mathbb{R}^n)$ and $H \in L_1([0,T])$ for all $T > 0$, then $A_A^1 - A_A^3$ are satisfied (recall that translation is continuous in L_1), and if $H \in L_1([0,\infty))$ then A_A^4 and A_A^5 are satisfied as well.

4. THE FINAL SIZE EQUATION

The analysis in [4] of the initial value problem for the space-independent model proceeds along more or less the same lines as followed in section 3. There the final size equation is a scalar equation

$$x(\infty) = \gamma s_0 g(x(\infty)) + f(\infty) \quad , \tag{4.1}$$

which can be analysed, for instance graphically. With g given by (2.11) one finds that (4.1) has a unique positive solution $x(\infty)$ for each positive $f(\infty)$. The dependence of $x(\infty)$ on the parameters s_0 and γ is best illustrated by looking at

$$\underline{x} = \inf_{f(\infty)>0} x(\infty).$$

It follows that $\underline{x}$ satisfies the homogeneous equation

$$y = \gamma s_0 g(y) \tag{4.2}$$

and $\underline{x}$ is positive if and only if $\gamma s_0 > 1$. This is the so-called *threshold phenomenon* which has received considerable attention (see [13], [4]).

The investigation of the final size equation (3.4) for the space-dependent model is not so simple. However, it is still possible to show that there

is an analogous threshold phenomenon. The analysis requires one extra step: with (3.4) we shall associate a scalar equation (which is essentially the same as in the space-independent case) and then we shall show that a positive solution of the scalar equation serves pointwise as a lower bound for any bounded, nonnegative solution of (3.4).

Put

$$s_0 = \inf_{x\in\Omega} S_0(x)$$

and

$$\gamma = \inf_{x\in\Omega} \int_\Omega \int_0^\infty A(\tau,x,\xi)d\tau d\xi$$

Thus, γ is the infimum of the total infectivity at x due to a homogeneously distributed unit density of infectives during the whole course of the disease. The scalar equation we have in mind is (4.2). With

$$\underline{u} = \inf_{x\in\Omega} u[\infty](x) \quad ,$$

we obtain from (3.4) and the monotonicity of g

$$\underline{u} \geq s_0\gamma g(\underline{u}) \tag{4.3}$$

and this makes it tempting to conjecture that $\underline{u} \geq p$, where p is defined as the positive solution of (4.2) if $\gamma s_0 > 1$ and $p = 0$ if $\gamma s_0 \leq 1$. This is sometimes called Kendall's Pandemic Threshold Theorem (see [11]).

<u>THEOREM</u> 4.1. *Assume that*

(i) Ω *is compact and connected,*

(ii) $s_0\gamma g(y) > y$ *for* $0 < y < p$,

(iii) for each $x \in \Omega$ *there exists* $\delta = \delta(x)$ *such that the set* $\{\xi \mid |x-\xi| \leq \delta\} \cap \Omega$ *is contained in the support of*

$$\int_0^\infty A(\tau,x,\cdot)d\tau.$$

If $f[\infty] \in BC(\Omega)$ *is nonnegative, but not identically zero, then every nonnegative solution* $u[\infty] \in BC(\Omega)$ *of (3.4) satisfies* $u[\infty] \geq p$.

PROOF. From (ii) and (4.3) we obtain that either $\underline{u} = 0$ or $\underline{u} \geq p$. Suppose $\underline{u} = 0$. Since $u[\infty]$ achieves its infimum, there exists $x_0 \in \Omega$ such that $u[\infty](x_0) = 0$. In (3.4) both terms are nonnegative, so (iii) implies that $u[\infty](x) = 0$ for $x \in \{\xi \mid |x_0 - \xi| \leq \delta(x_0)\} \cap \Omega$. Repeating the same argument one finds a maximal subset X of Ω where $u[\infty](x) = 0$. Necessarily $\partial X \subset \partial\Omega$ and hence $X = \Omega$. But with $u[\infty] \equiv 0$, (3.4) cannot be satisfied, and our assumption $\underline{u} = 0$ must be false. □

So the conjecture is seen to be true for compact Ω. The conditions, that Ω is connected and that A satisfies (iii) of Theorem 4.1, express that Ω should not consist of parts which are isolated with respect to infection.

Mathematically, the problem is much more interesting if Ω is not compact. Restricting our attention to the special case $\Omega = \mathbb{R}^n$, and assuming that (2.13) - (2.15) hold, we obtain the following inequality from (3.4)

$$u[\infty](x) \geq s_0\gamma \int_{\mathbb{R}^n} g(u[\infty](\xi))V(x-\xi)d\xi.$$

This leads to the investigation of the bounded solutions of the equation

$$w(x) = \int_{\mathbb{R}^n} g(w(\xi))V(x-\xi)d\xi + h(x),$$

with $h(x) \geq 0$. (We have incorporated the constant $s_0\gamma$ in the function g.) The results in the next section will imply the correctness of the conjecture for $n = 1$ or $n = 2$.

It is worthwhile to keep in mind that the lower bound p is independent of the function $f[t]$ as long as f is not identically zero. Hence, if $s_0\gamma > 1$, then equation (3.1) manifests the hair-trigger effect: no matter how little infectivity is introduced in an arbitrarily small subset of Ω, eventually there will be a large effect at every point.

5. A NONLINEAR CONVOLUTION EQUATION

Consider the nonlinear convolution equation

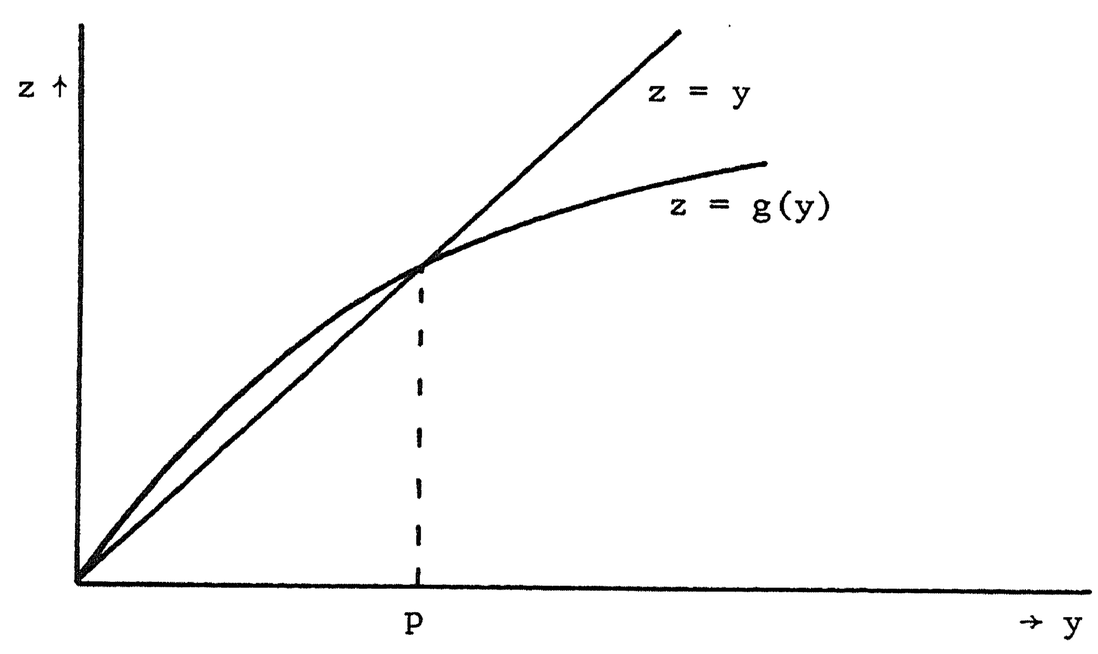
$$(5.1) \qquad w(x) = \int_{\mathbb{R}^n} g(w(\xi))V(x-\xi)d\xi + h(x) \ ,$$

for the unknown function $w\colon \mathbb{R}^n \to \mathbb{R}$, where $n = 1$ or $n = 2$. The functions g, V and h satisfy the following hypotheses:

H_g: $g\colon \mathbb{R} \to \mathbb{R}$ is continuous and monotone nondecreasing; $g(0) = 0$ and the equation

$$(5.2) \qquad y = g(y)$$

has one positive root p. Moreover, $g(y) > y$ for $0 < y < p$ and $g(y) < y$ for $y > p$.



H_V: $V \in L_1(\mathbb{R}^n)$, $\int_{\mathbb{R}^n} V(x)dx = 1$; $\int_{\mathbb{R}^n} |x|^n V(x)dx < \infty$; ($|\cdot|$ denotes the Euclidean norm)

V is a nonnegative radial function.

H_h: $h\colon \mathbb{R}^n \to \mathbb{R}$ is continuous, nonnegative and not identically zero.

Under only slightly more restrictive assumptions the existence of a bounded nonnegative solution of (5.1) can be established by means of a monotone iteration process. But we have a different aim in view, viz. to prove that p is a lower bound for any such solution.

THEOREM 5.1. *Let* $w: \mathbb{R}^n \to \mathbb{R}$ *be a continuous nonnegative solution of (5.1). Then* $w(x) \geq p$ *for every* $x \in \mathbb{R}^n$.

PROOF. Let v be defined by $v(x) = \min\{w(x),p\}$. Then (5.1) implies

$$v(x) \geq \int_{\mathbb{R}^n} v(\xi)V(x-\xi)d\xi. \tag{5.3}$$

It is known that in fact the inequality has to be an equality. A proof of this result can be found in ESSÉN [6, Theorem 3.1] or in FELLER [7, sections VI.10 and XVIII.7]. The basic idea of that proof is to show that the existence of a bounded function v satisfying (5.3) with strict inequality in a set of positive measure would imply the convergence of the infinite sum

$$\sum_{k=1}^{\infty} \int_{|x|\leq h} V^{k*}(x)dx$$

for some $h > 0$ (here V^{k*} denotes the $(k-1)$-times iterated convolution of V with itself), whereas the vanishing of the expectation(s) of V (recall that V is a radial function) implies the divergence of the same infinite sum for each $h > 0$. The latter result is proved by means of Fourier analysis. Note that here the dimension n and the assumptions concerning the existence of moments of V are important.

The only bounded continuous solutions of the equation

$$v(x) = \int_{\mathbb{R}^n} v(\xi)V(x-\xi)d\xi \quad ,$$

are constants (see [6, Theorem 2.1] or [7, section XI.2]). Thus, $v(x) \equiv c$. The case $c = 0$ can be excluded, because then $w(x) \equiv 0$ and, hence, (5.1) cannot be satisfied. If $0 < c < p$, then the inequality (5.3) is strict. So $c = p$ and the theorem is proved. □

The following result will be needed in section 6.

THEOREM 5.2. *Let* $w: \mathbb{R}^n \to \mathbb{R}$ *be a bounded nonnegative solution of the homogeneous equation*

$$(5.4) \qquad w(x) = \int_{\mathbb{R}^n} g(w(\xi))V(x-\xi)d\xi \ .$$

Then either $w(x) \equiv 0$ *or* $w(x) \equiv p$.

PROOF. As in the proof of Theorem 5.1 it follows that either $w(x) \equiv 0$ or $w(x) \geq p$ for every $x \in \mathbb{R}^n$. Let

$$\beta = \sup_{x \in \mathbb{R}^n} w(x) \quad ,$$

then (5.4) implies $w(x) \leq g(\beta)$. Suppose $\beta > p$, then

$$w(x) \leq g(\beta) < \beta + \frac{1}{2}(g(\beta) - \beta).$$

Hence $\beta - w(x)$ is bounded away from zero which is in contradiction with the definition of β. □

One can also consider the convolution equation (5.1) with a Borel probability measure instead of a L_1-kernel. The results of this section remain true if it is assumed that the support of the measure is non-arithmetic (or non-lattice, i.e., the support should not lie in any cyclic subgroup of $\mathbb{R}^n$; see [6], [7]).

The hypothesis concerning g was chosen to cover the case $g(y) = s_0\gamma(1-\exp(-y))$ and to avoid any multiplicity complication. If, on the other hand, the graph of $z = g(y)$ winds itself around the straight line $z = y$, then the homogeneous equation (5.4) may have many more bounded solutions. However, the nonexistence of nontrivial solutions satisfying certain bounds can still be shown along the same lines.

Thus far we have assumed that the population density exceeds the threshold on all of $\mathbb{R}^n$. Suppose, on the contrary, that the threshold is exceeded

only on a half-space, then the question arises whether the introduction of infectives on the other half-space will give rise to a severe epidemic. The following result answers this question positively.

THEOREM 5.3. *Let* $w: \mathbb{R}^n \to \mathbb{R}$ *be a bounded nonnegative solution of*

$$w(x) = \int_{\mathbb{R}^n} g(w(\xi))q(\xi)V(x-\xi)d\xi + h(x). \tag{5.5}$$

In addition to H_g, H_V *and* H_h *assume that:*

(i) *The support of* V *contains an interval if* $n = 1$ *and a ring or a circle if* $n = 2$,

(ii) *there exist* $\delta > 0$, $\ell > 1$ *such that* $g(x) \geq \ell x$ *for* $0 \leq x \leq \delta$,

(iii) $q \in L_\infty(\mathbb{R}^n)$ *and* q *is strictly positive.*

Then

(a) *if* $n = 1$, $\liminf_{x\to\infty} q(x) > 1$ *implies* $\liminf_{x\to\infty} w(x) \geq p$.

(b) *if* $n = 2$, *the existence of a sector* Λ *such that* $q(x) > 1$ *asymptotically for* $|x| \to \infty$ *in* Λ *implies that* $w(x) \geq p$ *asymptotically for* $|x| \to \infty$ *in every sector* Λ' *that is contained in* Λ.

PROOF. Since the arguments are exactly the same in both cases, we shall only give the proof of the result in case $n = 1$. We split the proof into several steps. Some of the arguments in step 6 are inspired by the discussion of Kendall [11].

STEP 1. The minimal nonnegative solution of (5.5) arises by monotone iteration with $h(x)$ as a starting function. Therefore it is sufficient to prove the result for a function h which has a compact support (for the minimal solution for arbitrary h is pointwise greater than or equal to the minimal solution corresponding to a compact supported $\tilde{h}$ which satisfies $\tilde{h} \leq h$). So let us assume that h has compact support.

STEP 2. u is uniformly continuous. For the first term on the right-hand side of (5.5) the uniform continuity follows from the continuity of translation in L_1 and for the second term from our assumption that h has compact support.

STEP 3. Let the limit set Q be defined by

$Q = \{\zeta |$ there exists a sequence $\{x_n\}$ with $x_n \to \infty$ and $w(x_n) \to \zeta$ as $n \to \infty\}$.

If $\zeta \in Q$, then either $\zeta = 0$ or $\zeta \geq p$.
This can be seen as follows. Suppose $w(x_n) \to \zeta$ and $x_n \to \infty$ as $n \to \infty$. By the Arzela-Ascoli theorem, the sequence of translated functions $w(x + x_n)$ has a subsequence which converges uniformly on compact subsets to a bounded, non-negative function $v_\zeta(x)$ say. Let $\bar{x}$ be such that $q(x) \geq 1$ for $x \geq \bar{x}$. From

$$w(x + x_n) \geq \int_{-\infty}^{x+x_n-\bar{x}} g(w(x + x_n - \xi))V(\xi)d\xi$$

we obtain upon passing to the limit

$$v_\zeta(x) \geq \int_{-\infty}^{\infty} g(v_\zeta(x - \xi))V(\xi)d\xi.$$

Hence either $v_\zeta(x) \equiv 0$ and $\zeta = 0$ or $v_\zeta(x) \geq p$ for every $x \in \mathbb{R}^n$ and $\zeta \geq p$ (Theorem 5.1).

STEP 4. Q is connected (intermediate value argument). So if $0 \in Q$ then $Q = \{0\}$ or, in other words, $\lim_{x\to\infty} w(x) = 0$.

STEP 5. w is strictly positive.
Indeed, suppose $w(x_0) = 0$ for some x_0, then from (5.5) it follows that $w(x) = 0$ for all x such that $x - x_0$ is contained in the support of V and, by assumption, this set contains an interval. By repetition of the argument we can cover the whole line and hence $w(x_0) = 0$ implies $w(x) \equiv 0$, which is impossible in view of (5.5).

STEP 6. It is impossible that $\lim_{x\to\infty} w(x) = 0$.
Suppose $\lim_{x\to\infty} w(x) = 0$. Then there exists $\tilde{x} = \tilde{x}(\delta)$ such that $w(x) \leq \delta$ for all $x \geq \tilde{x}$ and consequently $g(w(x)) \geq \ell w(x)$ for $x \geq \tilde{x}$. For every $d > 0$ (5.5) implies

$$w(x) \geq \int_{-d}^{d} g(w(x - \xi))V(\xi)d\xi, \quad \text{if } x \geq \bar{x} + d,$$

so if $x \geq a+d$, where $a = \max\{\tilde{x},\bar{x}\}$, then

$$w(x) \geq \ell \int_{-d}^{d} w(x-\xi)V(\xi)d\xi \ ,$$

and by iteration, if $x \geq a+md$, then

$$w(x) \geq \ell^m \int_{-md}^{md} w(x-\xi)V_d^{m*}(\xi)d\xi. \tag{5.6}$$

Here V_d^{m*} is the $(m-1)$-times iterated convolution of the truncated kernel V with itself (the symmetry of the kernel is conserved under convolution with itself!).

Since w is positive we can define for $n > (w(0))^{-1}$,

$$x_n = \sup\{x \mid w(\xi) \geq n^{-1} \text{ for } 0 \leq \xi \leq x\}.$$

It follows that $w(x_n) = n^{-1}$, $w(x) \geq w(x_n)$ for $0 \leq x \leq x_n$ and furthermore $x_n \to \infty$ for $n \to \infty$. At this moment we make a particular choice for the parameters d, m and n:

d so large that $\theta > 1$, where $\theta = \ell \int_{-d}^{d} V(\xi)d\xi$,

m so large that $\theta^m > 2$,

n so large that $x_n \geq a+md$.

Then we obtain from (5.6)

$$w(x_n) \geq \ell^m \int_0^{md} w(x_n-\xi)V_d^{m*}(\xi)d\xi \geq \frac{1}{2}\theta^m w(x_n) > w(x_n),$$

which is impossible.

Finally, combination of the results stated in the steps 3, 4 and 6 yields that if $\zeta \in Q$ then $\zeta \geq p$ and consequently $\liminf_{x\to\infty} w(x) \geq p$. □

In conclusion of this section we remark that one can use the same kind of arguments as have been used in the proof of Theorem 5.3 in the analysis of the final size equation for non-compact Ω other than $\mathbb{R}^n$, for instance for

half-lines or half-planes.

6. TRAVELLING WAVES

Now that we have shown that solutions of the initial value problem stabilize, we may ask how they approach the final distribution. Instead of studying this question in general, we concentrate on a special case and a particular set of solutions.

Let $\Omega = \mathbb{R}$ and $A(\tau,x,\xi) = \gamma H(\tau)V(x-\xi)$. Throughout this section we shall assume that

(i) $H \in L_1([0,\infty))$; $\int_0^\infty H(\tau)d\tau = 1$; H is nonnegative;

(ii) $V \in L_1(\mathbb{R})$; $\int_{-\infty}^{\infty} V(x)dx = 1$; V is nonnegative; $V(-x) = V(x)$;

if $W_\lambda(x) = e^{-\lambda x}V(x)$, then $W_\lambda \in L_1(\mathbb{R})$ for λ in a vertical strip (of positive width) of the complex plane.

We take the time-translation invariant equation (2.6) as our starting point. Let $S(-\infty,x) = S$ be a given constant (so we assume constant population density). We integrate equation (2.6) once with respect to t. Thus we obtain an equation of the form

$$u(t,x) = \int_0^\infty H(\tau) \int_{-\infty}^{\infty} g(u(t-\tau,\xi))V(x-\xi)d\xi d\tau \tag{6.1}$$

for

$$u(t,x) = -\ell n \frac{S(t,x)}{S} ,$$

where

$$g(y) = \gamma S(1-e^{-y}). \tag{6.2}$$

A *travelling wave solution* of (6.1) is a solution of the form $u(t,x) = w(x+ct)$. One may visualize a travelling wave solution as a function of x which, as time increases, is propagated (to the left if $c > 0$ and to the right if $c < 0$) with constant velocity c without any alteration in shape.

Since a travelling wave w depends on a linear combination of the independent variables x and t, one expects that it is possible to deduce an equation for w in one independent variable. Indeed, this is easily achieved by substitution and rearranging:

$$(6.3) \qquad w(\xi) = \int_{-\infty}^{\infty} g(w(\eta))V_c(\xi-\eta)d\eta, \quad \xi = x+ct,$$

where

$$(6.4) \qquad V_c(\xi) = \int_0^{\infty} H(\tau)V(\xi-c\tau)d\tau, \quad -\infty < \xi < \infty.$$

Incidentally we observe that a simlar equation arises in the study of plane wave solutions in higher dimensions.

For every c, (6.3) is a homogeneous nonlinear convolution equation with a nonnegative kernel. The constant solutions of (6.3) are the roots of

$$(6.5) \qquad y = g(y)$$

and we assume that (6.5) has, in addition to the root $y=0$, one positive root $y = p$ (in view of (6.2) this amounts to the assumption $\gamma S > 1$).

If $c = 0$ (equilibrium solutions) the kernel V_c is symmetric and we may apply Theorem 5.2 to conclude that $w \equiv 0$ and $w \equiv p$ are the only bounded nonnegative solutions of (6.3). The question arises whether or not there are values of c for which (6.3) has a solution satisfying $0 < w(\xi) < p$ and how these values of c can be characterized. From the symmetry of V it is immediately clear that, if $w(\xi)$ satisfies (6.3) for $c = \bar{c}$, then $w(-\xi)$ satisfies (6.3) for $c = -\bar{c}$, so we can confine ourselves to the case $c > 0$. In order to get a first impression we note that

$$\int_{-\infty}^{\infty} V_c(\xi)d\xi = 1$$

and

$$\int_{-\infty}^{\infty} \xi V_c(\xi)d\xi = c \int_0^{\infty} \tau H(\tau)d\tau,$$

if the integral at the right-hand side converges. Or, in words: the total mass of V_c is constant, but the distribution of mass becomes more and more lopsided as c increases. It turns out that (6.3) may have a nontrivial solution if V_c is lopsided enough.

Consider the linearized (about $w \equiv 0$) equation

$$(6.6) \qquad v(\xi) = g'(0) \int_{-\infty}^{\infty} v(\eta)V_c(\xi - \eta)d\eta$$

and the corresponding characteristic equation

$$(6.7) \qquad L_c(\lambda) = 1,$$

where

$$(6.8) \qquad L_c(\lambda) = g'(0) \int_{-\infty}^{\infty} e^{-\lambda\xi}V_c(\xi)d\xi = g'(0) \int_0^{\infty} e^{-\lambda c\tau}H(\tau)d\tau . \int_{-\infty}^{\infty} e^{-\lambda\xi}V(\xi)d\xi .$$

If λ_0 is a root of equation (6.7) of multiplicity $k \geq 1$, then $\xi^m \exp(\lambda_0\xi)$ is a solution of (6.6) for $m = 0,1,\ldots,k-1$ (one can verify this by substitution) and essentially all solutions of (6.6) are generated in this manner (see TITCHMARSH [17, Theorem 146] and [4]). The observation that positive solutions of (6.6) correspond to real roots of (6.7) now motivates us to study L_c as a function of a real variable and to list some of its properties:

(i) $L_c(\lambda)$ is defined in a right-hand neighbourhood of $\lambda = 0$;

(ii) $L_c(0) = g'(0) > 1$ (by assumption);

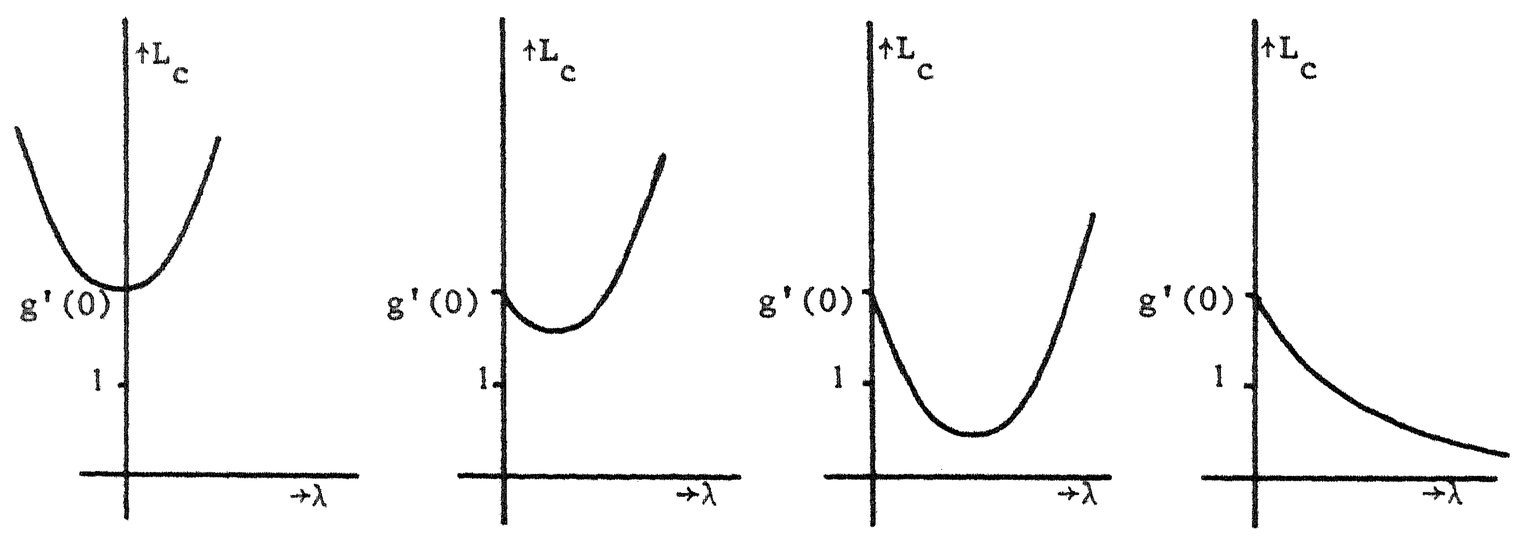
(iii) $\dfrac{dL_c}{d\lambda}(0+) = -cg'(0) \displaystyle\int_0^{\infty} \tau H(\tau)d\tau < 0$ (possibly $-\infty$);

(iv) $\dfrac{d^2L_c}{d\lambda^2}(\lambda) = g'(0) \displaystyle\int_{-\infty}^{\infty} \xi^2 e^{-\lambda\xi}V_c(\xi)d\xi > 0$, i.e., L_c is a convex function.

Next we describe the dependence on c. If $c = 0$, $L_c(\lambda)$ has a minimum for $\lambda = 0$. From (iii) and (iv) it follows that for $c > 0$ a minimum can occur only for positive λ. For every $\lambda > 0$, $L_c(\lambda)$ is a monotonically decreasing function of c, and we can achieve that $L_c(\lambda) < 1$ by choosing c sufficiently large. Consequently, the set

$$\{c \mid \text{there exists } \lambda > 0 \text{ such that } L_c(\lambda) < 1\}$$

consists of a half-line, (c_0,∞) say.



$c = 0$ $\qquad$ $0 < c < c_0$ $\qquad$ $c > c_0$ $\qquad$ or $\qquad$ $c > c_0$

We have provided enough motivation to state the following theorem.

<u>THEOREM</u> 6.1. *Let* $g\colon \mathbb{R} \to \mathbb{R}$ *be continuous and monotone nondecreasing and suppose* $g(0) = 0$, $g(y) > y$ *for* $0 < y < p$ *and* $g(p) = p$. *Assume, furthermore, that* $g'(0)$ *exists and that, for some* $k > 0$,

$$g'(0)y - ky^2 \le g(y) \le g'(0)y, \quad \text{for } 0 \le y \le p.$$

Let $K \in L_1(\mathbb{R})$ *be nonnegative and let* $\int_{-\infty}^{\infty} K(x)dx = 1$. *Suppose there exists* $\lambda_0 > 0$ *for which* $L(\lambda_0) < 1$, *where*

$$L(\lambda) = g'(0) \int_{-\infty}^{\infty} e^{-\lambda x} K(x)dx.$$

(Note that the convergence of the integral for $\lambda = \lambda_0$ *is part of the assumption.)*
Then the equation

$$(6.9) \qquad w(x) = \int_{-\infty}^{\infty} g(w(\eta))K(x-\eta)d\eta$$

has a monotone nondecreasing solution w *which has the following properties,*

$$0 < w(x) < p, \lim_{x\to-\infty} w(x) = 0, \lim_{x\to+\infty} w(x) = p.$$

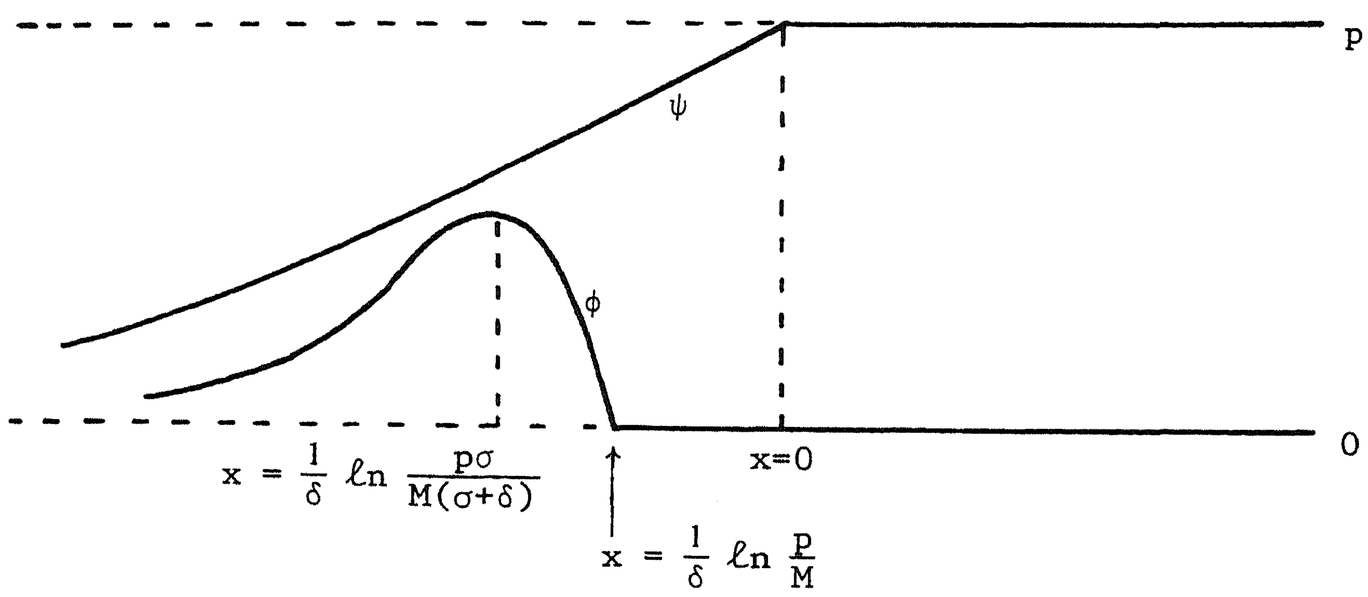
REMARK. The subjoined proof is essentially the same as the proof of a related result in [3], but it is different in a number of details.

PROOF. The convergence of the defining integral for $L(\lambda)$ for $\lambda = 0$ and $\lambda = \lambda_0$ implies the convergence for $0 < \lambda < \lambda_0$. Since $L(\lambda)$ is continuous and $L(0) > 1$, there exists a $\sigma \in (0,\lambda_0)$ such that $L(\sigma) = 1$ and, since L is convex, $L(\lambda) < 1$ for $\sigma < \lambda \le \lambda_0$. We introduce the following functions

$$\psi(x) = \min\{pe^{\sigma x}, p\} \quad ,$$

$$\phi(x) = \max\{pe^{\sigma x} - Me^{(\sigma+\delta)x}, 0\} \quad ,$$

with $M,\delta > 0$ still at our disposal.



A first requirement we want to meet is that $\phi(x) < \psi(x)$ for all $x \in \mathbb{R}$. If the maximum of $p\exp(\sigma x) - M\exp(\sigma x + \delta x)$ is achieved for negative x then this will certainly be the case and thus we find as a sufficient condition

$$(6.10) \qquad M > \frac{\sigma p}{\sigma + \delta} \quad .$$

Defining a formal integral-operator T by the right-hand side of (6.9), we now try to choose M and δ so that $T\psi \le \psi$ and $T\phi \ge \phi$. The former inequality follows directly from the assumptions and from the equality $L(\sigma) = 1$, so we concentrate on the latter.

The properties of g imply the estimate

$$(T\phi)(x) \ge g'(0) \int_{-\infty}^{\infty} \phi(x-\eta)K(\eta)d\eta - k \int_{-\infty}^{\infty} \phi^2(x-\eta)K(\eta)d\eta.$$

By choosing $0 < \delta \le \sigma$ we can achieve that

$$\phi^2(x) \le \psi^2(x) \le p^2 \, e^{(\sigma+\delta)x} \quad ,$$

and hence

$$(T\phi)(x) \ge p \, e^{\sigma x} - ML(\sigma+\delta)e^{(\sigma+\delta)x} - \frac{kp^2}{g'(0)} L(\sigma+\delta)e^{(\sigma+\delta)x} \quad .$$

Consequently,

$$(T\phi)(x) \ge p \, e^{\sigma x} - Me^{(\sigma+\delta)x} \quad ,$$

provided

$$(6.11) \qquad ML(\sigma+\delta) + \frac{kp^2}{g'(0)} L(\sigma+\delta) \le M \quad .$$

Choose δ such that $0 < \delta \le \min\{\sigma, \lambda_0 - \sigma\}$, then $L(\sigma+\delta) < 1$ and we can satisfy (6.10) and (6.11) by choosing M sufficiently large that

$$M \ge \max\left\{ \frac{kp^2}{g'(0)} (1 - L(\sigma+\delta))^{-1}, \ \frac{\sigma p}{\sigma+\delta} \right\} \quad .$$

Then indeed $T\phi \geq \phi$ (note that $T\phi \geq 0$ is trivial).

The operator T is monotone, i.e., $\chi_1 \leq \chi_2$ implies $T\chi_1 \leq T\chi_2$. Starting with ψ and iterating with T we obtain a monotone decreasing sequence $\{\psi_n = T^n\psi \mid n = 0,1,\ldots\}$ which is bounded from below by ϕ and, hence, pointwise convergent. As before, the Arzela-Ascoli theorem can be applied to show that the convergence is in fact uniform on compact subsets, which implies that the limit, say w, satisfies the equation (6.9). Since ψ is monotone nondecreasing, the same is true for $T^n\psi$ and, hence, for w. From the properties of ψ and ϕ we obtain $\lim_{x\to-\infty} w(x) = 0$. The monotonicity and the boundedness of w imply the existence of $\lim_{x\to+\infty} w(x)$ and, finally, a translation argument shows that this limit is a constant solution of (6.9) and, thus, equal to p. □

<u>COROLLARY</u> 6.2. *Let* g *be given by (6.2) with* $\gamma S > 1$. *Then there exists a* $c_0 > 0$ *such that (6.1) has a travelling wave solution for every* $c > c_0$. *The constant* c_0 *is given by* $c_0 = \inf\{c \mid \text{there exists } \lambda > 0 \text{ such that } L_c(\lambda) < 1\}$.

In [5], the nonlinear convolution equation (6.9) is studied and Tauberian methods are used to prove a nonexistence result. Under some reasonable conditions on H and V this yields a complementary nonexistence result for travelling waves with $0 < c < c_0$. The question whether or not there is a travelling wave solution for $c = c_0$ appears to be more delicate.

The results of [5] suggest that, under some additional conditions, travelling waves for fixed c are unique modulo translation. A special case deserves attention: if the support of V is compact and contained in $[-a,a]$, and the support of H is contained in $[b,\infty)$ with $b > 0$, then for $c \geq ab^{-1}$ the support of V_c is contained in $[0,\infty)$ and the uniqueness of the travelling wave follows from [4, section 4].

An obvious continuation of the above analysis would be to study the stability of the travelling waves, that is, their relation to solutions of the initial value problem (3.1). A related approach, with a somewhat different flavour, would be to investigate the relation between c_0 and the asymptotic speed of propagation of infectivity starting from a compactly supported initial function. Here, ARONSON and WEINBERGER [1], [2] could be a source of inspiration. Some work on the rate of propagation of infectivity has already been done by MOLLISON [16].

ACKNOWLEDGEMENT

The author is grateful to H.G. Kaper and L.A. Peletier. Their comments on a preliminary draft of the paper have led to many improvements!

REFERENCES

[1] ARONSON, D.G. & H.F. WEINBERGER, *Nonlinear diffusion in population genetics, combustion, and nerve pulse propagation,* in: J.A. Goldstein, ed., Partial Differential Equations and Related Topics, Lecture Notes in Math. 446 (Springer, Berlin, 1975) pp. 5-49.

[2] ARONSON, D.G. & H.F. WEINBERGER, *Multidimensional nonlinear diffusion arising in population genetics,* to appear in Advances in Math..

[3] ATKINSON, C. & G.E.H. REUTER, *Deterministic epidemic waves,* Math. Proc. Camb. Phil. Soc. 80 (1976) pp. 315-330.

[4] DIEKMANN, O., *Limiting behaviour in an epidemic model,* to appear in J. Nonl. Anal.-Theory Meth. Appl..

[5] DIEKMANN, O. & H.G. KAPER, *On the bounded solutions of a nonlinear convolution equation,* in preparation.

[6] ESSÉN, M., *Studies on a convolution inequality,* Ark. Mat. 5 (1963) pp. 113-152.

[7] FELLER, W., *An Introduction to Probability Theory and Its Applications,* Vol. II (Wiley, New York, 1966).

[8] HADELER, K.P. & F. ROTHE, *Travelling fronts in nonlinear diffusion equations,* J. Math. Biol. 2 (1975) pp. 251-263.

[9] HALE, J.K., *Ordinary Differential Equations,* (Wiley, New York, 1969).

[10] HOPPENSTEADT, F., *Mathematical Theories of Populations: Demographics, Genetics and Epidemics,* SIAM Regional Conference Series in Applied Mathematics, Vol. 20 (SIAM, Philadelphia, 1975).

[11] KENDALL, D.G., *Discussion of "Measles periodicity and community size" by M.S. Bartlett*, J. Roy. Statist. Soc. A 120 (1957) pp. 64-67.

[12] KENDALL, D.G., *Mathematical models of the spread of infection*, in Mathematics and Computer Science in Biology and Medicine (Medical Research Council, London, 1965) pp. 213-224.

[13] KERMACK, W.O. & A.G. McKENDRICK, *A contribution to the mathematical theory of epidemics*, Proc. Roy. Soc. A 115 (1927) pp. 700-721.

[14] METZ, J.A.J., *The epidemic in a closed population with all susceptibles equally vulnerable; some results for large susceptible populations and small initial infections*, to appear in Acta Biotheoretica.

[15] MOLLISON, D., *Possible velocities for a simple epidemic*, Advances in Applied Prob. 4 (1972) pp. 233-257.

[16] MOLLISON, D., *The rate of spatial propagation of simple epidemics*, in L.M. le Cam, J. Neyman & E.L. Scott, eds., Proc. Sixth Berkeley Symposium on Mathematical Statistics and Probability, Vol. III (Univ. of California Press, Berkeley, 1972) pp. 579-614.

[17] TITCHMARSH, E.C., *Introduction to the Theory of Fourier Integrals*, (Clarendon Press, Oxford, 1937).

On the bounded solutions of a nonlinear convolution equation

by

O. Diekmann & H.G. Kaper*

ABSTRACT

This investigation is concerned with the nonlinear convolution equation $u(x) - (g\circ u)*k(x) = 0$ on the real line $\mathbb{R}$. The kernel k is nonnegative and integrable on $\mathbb{R}$, with $\int_{\mathbb{R}} k(x)dx = 1$; the function g is real-valued and continuous on $\mathbb{R}$, $g(0) = 0$, and there exists a $p > 0$ such that $g(x) > x$ for $x \in (0,p)$ and $g(p) = p$. Sufficient conditions are given for the non-existence of bounded nontrivial solutions. Implications for the solution of the inhomogeneous equation $u(x) - (g\circ u)*k(x) = f(x)$, $x \in \mathbb{R}$, are discussed. Finally, uniqueness (modulo translation) is shown to hold. The results are applied to a problem of mathematical epidemiology.

---

* Applied Mathematics Division, Argonne National Laboratory, Argonne, IL 60439, USA. Work performed under the auspices of the U.S. Department of Energy.

1. INTRODUCTION

Certain deterministic models for the spatial spread of an epidemic or an advantageous gene among a population along a line can be analyzed in terms of the nonlinear convolution equation

$$(1.1) \qquad u(x) = (g\circ u)*k(x) \qquad x \in \mathbb{R},$$

where $g\circ u$ is the composite of g and u, $g\circ u(x) = g(u(x))$, and $*$ denotes convolution, $\phi*\psi(x) = \int_{\mathbb{R}}\phi(x-y)\psi(y)dy$. In the epidemic model, g typically has the form $g(x) = \alpha(1-e^{-x})$, α some constant ($\alpha>1$), cf. DIEKMANN [6], while in the genetic model g is given by $g(x) = [\alpha x^2+\beta x(1-x)]/[\alpha x^2+2\beta x(1-x)+ +\gamma(1-x)^2]$, α, β, and γ positive constants, cf. WEINBERGER [16]. In both cases, k is a nonnegative function, normalized such that $\int_{\mathbb{R}}k(x)dx = 1$.

We consider eq. (1.1) under the following hypotheses:

(H_g^1) g: $\mathbb{R} \to \mathbb{R}$ is continuous; $g(0) = 0$; there exists a $p > 0$ such that $g(x) > x$ for $0 < x < p$ and $g(p) = p$.

(H_k^1) k: $\mathbb{R} \to \mathbb{R}$ is nonnegative; $k \in L_1(\mathbb{R})$ with $\int_{\mathbb{R}}k(x)dx = 1$; $\int_{\mathbb{R}}|x|k(x)dx < \infty$.

Further hypotheses on g and k will be introduced as needed.

Equation (1.1) has at least two constant solutions, viz., u_0: $x \mapsto 0$ and u_p: $x \mapsto p$. These we call trivial solutions. By a nontrivial solution of (1.1) we mean a Lebesgue measurable function u: $\mathbb{R} \to \mathbb{R}$ which satisfies (1.1) and the inequalities $0 \le u(x) \le p$ for almost all $x \in \mathbb{R}$ and which is not a trivial solution. We are interested in nontrivial solutions of (1.1) and in the solution of the corresponding inhomogeneous equation

$$(1.2) \qquad u(x) = (g\circ u) * k(x) + f(x) \qquad x \in \mathbb{R},$$

where f is a given nonnegative function.

In section 2 we establish various general properties of the solutions of equation (1.1). In Sections 3 and 4 we prove some non-existence results, i.e., we establish conditions on g and k under which one can prove that the equation (1.1) does not have a nontrivial solution. These results are

complementary to some earlier results of WEINBERGER [16] and of one of the authors [6] concerning the existence of nontrivial solutions of (1.1). As we will show, the existence or non-existence of such solutions hinges upon the distribution of the mass of the kernel k. Formally, this is determined by the existence or non-existence of real zeros of the function $\lambda \mapsto 1 - g'(0)K(\lambda)(\lambda \in \mathbb{R})$, where K is the two-sided Laplace transform of k. The problem has some Tauberian aspects and part of our arguments are based on Pitt's form of Wiener's general Tauberian theorem. This approach is related to the use of Tauberian methods in the study of linear convolution equations by KARLIN [9] and ESSÉN [7], cf. [14, Chapter 9].

In Section 5 we study the inhomogeneous equation (1.2). In Section 6 we turn our attention to the case when a nontrivial solution of (1.1) exists and investigate the question of its uniqueness (modulo translation, as eq. (1.1) is translation invariant). Here, our method of proof forces us to impose further conditions on g and k.

In the final Section 7 we apply our results to a problem in mathematical epidemiology, viz., the traveling wave problem for the Kermack-McKendrick-Kendall model for the spatial spread of an epidemic, cf. [1], [2], [3], [4], [6] [10], [11], [12].

NOTATION. $\mathbb{N}_+$ denotes the set of all positive integers; $\mathbb{R}_+ = \{x \in \mathbb{R} : x \geq 0\}$ and $\mathbb{R}_- = \{x \in \mathbb{R} : x \leq 0\}$. The letter C is used to denote a generic positive constant.

2. PRELIMINARIES

Starting from k we form a sequence of functions $\{k^{n*} : n \in \mathbb{N}_+\}$. We interpret k^{1*} as k, and define for $n \in \mathbb{N}_+$, $k^{(n+1)*} := k^{n*} * k$. Thus, $k^{n*} \in L_1(\mathbb{R})$ and $\int_{\mathbb{R}} k^{n*}(x)dx = 1$ for all n. Following FELLER [8, Sections V.2, V.4], we call x a point of increase of k if $\int_I k(y)dy > 0$ for every open interval I containing x, and we define $\sum := \bigcup_{n=1}^{\infty} \sum_n$, where $\sum_n$ denotes the set formed by the points of increase of k^{n*}. We say that k is concentrated on a set I if $\int_{\mathbb{R}\setminus I} k(x)dx = 0$. The following result will be used several times.

LEMMA 2.1. *If* k *is concentrated on* $\mathbb{R}_+$ ($\mathbb{R}_-$), *then there exists an* $a \geq 0$ *such that* $\sum$ *contains the interval* $[a,\infty)$ ($(-\infty,-a]$). *If* k *is not concentrated on either* $\mathbb{R}_+$ *or* $\mathbb{R}_-$, *then* $\sum = \mathbb{R}$.

PROOF. For an arbitrary real number $b > 0$ we define $\ell(x) := \min\{k(x),b\}$. Then $\ell \in L_\infty(\mathbb{R}) \cap L_1(\mathbb{R})$ and, furthermore,

$$k^{2*}(x) \geq \ell^{2*}(x) \geq \int_0^x \ell(x-y)\ell(y)dy =: L(x).$$

We claim that $L(x)$ is uniformly continuous. Indeed, $|L(x_1)-L(x_2)| \leq b^2|x_1-x_2| + b\int_{\mathbb{R}}|\ell(y)-\ell_{x_1-x_2}(y)|dy$, where $\ell_x(y) := \ell(y-x)$, and the translation map $x \mapsto \ell_x$ from $\mathbb{R}$ into $L_1(\mathbb{R})$ is uniformly continuous, see [13, Section 1.1.5].

Suppose $\int_{\mathbb{R}+} k(x)dx > 0$. Then L cannot be identically zero on $\mathbb{R}_+$ and, hence, there exist $x_1 \geq 0$ and $\delta > 0$ such that $k^{2*}(x) \geq L(x) > 0$ for $x \in (x_1,x_1+\delta)$. From this we deduce that $k^{4*}(x) = (k^{2*} * k^{2*})(x) > 0$ for $x \in (2x_1,2x_1+2\delta)$ and, by induction, $k^{2n*}(x) > 0$ for $x \in (nx_1,nx_1+n\delta)$, $n \in \mathbb{N}_+$. For sufficiently large $n(n>\delta^{-1}x_1)$, successive intervals become overlapping and, consequently, $\sum$ contains an interval $[a,\infty)$ for some $a \geq 0$. A similar argument shows that if $\int_{\mathbb{R}_-} k(x)dx > 0$, then $\sum$ contains an interval $(-\infty,-a]$ for some $a \geq 0$.

Suppose now that both $\int_{\mathbb{R}+} k(x)dx > 0$ and $\int_{\mathbb{R}_-} k(x)dx > 0$. Let $x_0 \in \mathbb{R}$ be arbitrary. For sufficiently large y we know that there exist positive integers m and n such that $k^{n*}(x) > 0$ for x in a neighborhood of $\frac{1}{2}x_0 + y$ and $k^{m*}(x) > 0$ for x in a neighborhood of $\frac{1}{2}x_0 - y$. Consequently, $k^{(n+m)*}(x) > 0$ for x in a neighborhood of x_0 and $x_0 \in \sum$. □

The next lemma contains some basic information concerning nontrivial solutions of (1.1).

LEMMA 2.2. *Any nontrivial solution* u *of* (1.1) *is uniformly continuous, strictly positive, and such that* $\inf\{u(x): x \in \mathbb{R}\} = 0$.

PROOF. Let u be a nontrivial solution of (1.1). The uniform continuity of u follows from the continuity of translation in $L_1(\mathbb{R})$ (cf. the proof of Lemma 2.1).

Next, we observe that $g\circ u(x) \geq u(x)$ for all $x \in \mathbb{R}$, so $u(x) \geq u*k(x)$ and, upon iteration, $u(x) \geq u*k^{n*}(x)$ for $n \in \mathbb{N}_+$. Hence, if $u(x_0) = 0$ for some $x_0 \in \mathbb{R}$, then $u*k^{n*}(x_0) = 0$ for $n \in \mathbb{N}_+$, which implies $u(x) = 0$ for all x such that $x_0 - x \in \sum$. On account of Lemma 2.1 this implies that $u(x) = 0$ for all $x \in \mathbb{R}$ if k is not concentrated on either $\mathbb{R}_+$ or $\mathbb{R}_-$. If, on the other hand, k is concentrated on a half-axis, $\mathbb{R}_+$ say, then Lemma 2.1 implies that $u(x) = 0$ for all $x \in (-\infty, x_0-a]$ for some $a \in \mathbb{R}_+$ and, consequently, $u(x) = 0$ for all $x \in \mathbb{R}$, since the equation (1.1) has to be satisfied. In all cases we thus arrive at the conclusion that $u(x) = 0$ for all $x \in \mathbb{R}$, contrary to the assumption that u is nontrivial.

Let $\alpha := \inf\{u(x): x \in \mathbb{R}\}$ and suppose $\alpha > 0$. Then $g\circ u(x) \geq \min\{g(y): \alpha \leq y \leq p\}$. Since g is continuous, there exists a $\beta \in [\alpha,p]$ such that $g(\beta) = \min\{g(y): \alpha \leq y \leq p\}$. So $u(x) = (g\circ u)*k(x) \geq g(\beta) = \alpha + (g(\beta)-\alpha) \geq \alpha + (g(\beta)-\beta) = \alpha + \varepsilon$, where $\varepsilon := g(\beta) - \beta > 0$. Hence, $\inf\{u(x): x \in \mathbb{R}\} \geq \alpha + \varepsilon$, which is in contradiction with the definition of α. We conclude that $\alpha = 0$. □

3. THE FIRST NON-EXISTENCE RESULT

In this section we give a non-existence result which can be proved by elementary arguments. Let L be defined by

$$L := \limsup_{x\downarrow 0} \frac{g(x)}{x} . \tag{3.1}$$

Note that $L \geq 1$; L may be infinite. If g is differentiable at the origin with derivative value $g'(0)$, then $L = g'(0)$.

<u>LEMMA 3.1</u>. *Suppose* g *is monotone nondecreasing on* $[0,p]$. *If* $\int_{\mathbb{R}_\pm} k(x)dx > L^{-1}$, *then any nontrivial solution* u *of* (1.1) *is such that* $\liminf_{x\to\pm\infty} u(x) > 0$.

<u>PROOF</u>. We prove the lemma only for the upper sign. Let u be a nontrivial solution of (1.1) and suppose that $\liminf_{x\to+\infty} u(x) = 0$. Choose a monotonically decreasing sequence $\{y_n: n \in \mathbb{N}_+\}$ such that $y_n \downarrow 0$ and $g(y_n)/y_n \to L$ as $n \to \infty$. Define x_n for $n = n_0, n_0+1,\ldots,$ where n_0 is such that $u(0) > y_{n_0}$, by $x_n := \sup\{x \in \mathbb{R}_+ : u(y) \geq y_n \text{ for all } y \in [0,x]\}$. Then $u(x_n) = y_n$ and

$x_n \to \infty$ as $n \to \infty$. Furthermore, as g is monotone nondecreasing,

$$u(x_n) \geq \int_0^{x_n} g\circ u(x_n-y)k(y)dy \geq g\circ u(x_n) \int_0^{x_n} k(y)dy,$$

so $(g(y_n)/y_n)\int_0^{x_n} k(y)dy \leq 1$ for all n. Letting $n \to \infty$ we obtain the inequality $L \int_{\mathbb{R}_+} k(y)dy \leq 1$, a contradiction. We conclude that $\liminf_{x\to\infty} u(x) > 0$. □

The following theorem is an immediate consequence of Lemmas 2.2 and 3.1.

THEOREM 3.2. *Suppose* g *is monotone nondecreasing on* [0,p]. *If* $\int_{\mathbb{R}_+} k(x)dx > L^{-1}$ *and* $\int_{\mathbb{R}_-} k(x)dx > L^{-1}$, *then there is no nontrivial solution of* (1.1).

PROOF. Let u be a nontrivial solution of (1.1). Then either $\liminf_{x\to+\infty} u(x) = 0$ or $\liminf_{x\to-\infty} u(x) = 0$, by Lemma 2.2. However, from Lemma 3.1 we infer that $\liminf_{x\to+\infty} u(x) > 0$ and $\liminf_{x\to-\infty} u(x) > 0$, and we arrive at a contradiction. We conclude that there is no nontrivial solution of (1.1).

REMARK 3.3. Up to this point we have not made use of the hypothesis that $\int_{\mathbb{R}} |x|k(x)dx < \infty$, so all foregoing results remain valid if this hypothesis is actually violated.

4. ANOTHER NON-EXISTENCE RESULT

In the present section we establish a non-existence result whose proof is rather involved. It is based on various reformulations of (1.1) as an inhomogeneous linear convolution equation. First we rewrite (1.1) in the form

$$u(x) - u*k(x) = \phi(x) \qquad x \in \mathbb{R}, \tag{4.1}$$

where

$$\phi(x) := (g\circ u-u)*k(x). \tag{4.2}$$

The function ϕ is nonnegative on $\mathbb{R}$ and we will exploit this fact in the investigation of solutions of (4.1), cf. ESSÉN [7].

Let m denote the first moment of the kernel k,

$$(4.3) \qquad m = \int_{\mathbb{R}} xk(x)dx.$$

The first theorem requires barely more than a reference to the literature.

THEOREM 4.1. *If* $m = 0$ *then there is no nontrivial solution of* (1.1).

PROOF. It is known that the condition $m \neq 0$ is necessary and sufficient for the existence of a bounded continuous function u such that $u(x) - u*k(x) \geq 0$, with strict inequality in some point, cf. [7, Theorem 3.1]. □

Having established this non-existence result for $m = 0$, we may henceforth assume $m \neq 0$.

Let

$$(4.4) \qquad n(x) := \begin{cases} \int_x^\infty k(y)dy & x \geq 0, \\ -\int_{-\infty}^x k(y)dy & x < 0. \end{cases}$$

Note that $n(x) = H(x) - \int_{-\infty}^x k(y)dy$ and, consequently, $n' = \delta - k$ in the sense of distributions (here H denotes the Heaviside step function and δ the Dirac distribution). The condition $\int_{\mathbb{R}} |x|k(x)dx < \infty$ implies that $n \in L_1(\mathbb{R})$ with $\|n\|_1 = \int_{\mathbb{R}} |x|k(x)dx$, cf. [8, Section V.6.]. The Fourier transform $\hat{n}$ of n, which is a continuous function, is related to the Fourier transform $\hat{k}$ of k via the identity $\hat{n}(\lambda) = (1-\hat{k}(\lambda))/(i\lambda)$ for $\lambda \neq 0$, while $\hat{n}(0) = m$. If $m \neq 0$, then $\hat{n}(\lambda) \neq 0$ for all $\lambda \in \mathbb{R}$.

LEMMA 4.2. *Suppose* $m > 0$. *If* u *is any nontrivial solution of* (1.1), *then* $\lim_{x\to-\infty} u(x)$ *and* $\lim_{x\to\infty} u(x)$ *exist and* $u(-\infty) = 0$, $u(\infty) = p$.

PROOF. Let u be a nontrivial solution of (1.1). Then u satisfies (4.1) with ϕ given by (4.2). We integrate both members of (4.1) from a to x,

$$\int_a^x u(\xi)d\xi - \int_a^x \int_{-\infty}^{\infty} u(\xi-\eta)k(\eta)d\eta d\xi = \int_a^x \phi(\xi)d\xi.$$

Interchanging the order of the integrals in the second term and integrating by parts we obtain the identity

$$\int_a^x \int_{-\infty}^{\infty} u(\xi-\eta)k(\eta)d\eta d\xi = \int_a^x u(\xi)d\xi - \int_{-\infty}^{\infty} n(x-\xi)u(\xi)d\xi + \int_{-\infty}^{\infty} n(a-\xi)u(\xi)d\xi,$$

where n is defined by (4.4). Hence, u satisfies

$$(4.5) \qquad u*n(x) - u*n(a) = \int_a^x \phi(\xi)d\xi \qquad x \in \mathbb{R}.$$

Since u is bounded and $n \in L_1(\mathbb{R})$, $u*n$ is bounded. Also, $\phi(x)$ is nonnegative for all $x \in \mathbb{R}$, so the integral $\int_a^x \phi(\xi)d\xi$, besides being bounded, depends monotonically on x. It follows that $\lim_{x\to\infty} \int_a^x \phi(\xi)d\xi$ exists and hence, from (4.5), that $\lim u*n(x)$ exists. In fact,

$$u*n(\infty) = u*n(a) + \int_a^{\infty} \phi(\xi)d\xi.$$

Now, n belongs to the Wiener class W (i.e., $n \in L_1(\mathbb{R})$ and $\{\lambda \in \mathbb{R} : \hat{n}(\lambda) = 0\} = \emptyset$), and u is bounded and uniformly continuous. It follows from Pitt's form of Wiener's fundamental Tauberian theorem, cf. [14, Section 9.7] or [17, Section V.10], that $\lim_{x\to\infty} u(x)$ exists and is given by

$$u(\infty) = m^{-1}\left(u*n(a) + \int_a^{\infty} \phi(x)dx\right).$$

Similarly, one shows that $\lim_{x\to-\infty} u(x)$ exists and is given by

$$u(-\infty) = m^{-1}\left(u*n(a) - \int_{-\infty}^{a} \phi(x)dx\right).$$

A simple argument shows that both $u(\infty)$ and $u(-\infty)$ satisfy the scalar equation $y = g(y)$, so $\{u(\infty),u(-\infty)\} \subset \{0,p\}$. Since $u(\infty)-u(-\infty) = \int_{-\infty}^{\infty} \phi(x)dx > 0$, we conclude that $u(-\infty) = 0$ and $u(\infty) = p$. □

REMARK 4.3. The case $m < 0$ can be reduced to the case $m > 0$ by a change of variable $\tau: x \mapsto -x$.

The next step towards a non-existence result for $m \neq 0$ consists of an analysis of the rate of convergence of a nontrivial solution towards its limiting value 0 as $x \to -\infty$. The following lemma is needed in the proof of Lemma 4.5.

LEMMA 4.4. *Let* $f: \mathbb{R} \to \mathbb{R}$ *be nonnegative,* $f \in L_1(\mathbb{R})$. *Define* $f^{(0)}(x) := f(x)$ *and, recursively,* $f^{(k)}(x) := \int_{-\infty}^{x} f^{(k-1)}(y)dy$ *for those values of* $k \in \mathbb{N}_+$ *for which* $f^{(k-1)} \in L_1(\mathbb{R}_-)$. *Then, for any* $r \in \mathbb{R}$,

$$\int_{-\infty}^{r} (r-x)^k f(x)dx = k! \int_{-\infty}^{r} f^{(k)}(x)dx,$$

in the sense that if one side converges so does the other.

PROOF. The identity given in the lemma follows upon induction from the identity

$$\int_{-\infty}^{r} (r-x)^k f(x)dx = k \int_{-\infty}^{r} (r-x)^{k-1} f^{(1)}(x)dx.$$

We omit the proof of the latter identity, as it is similar to the proof of Lemma 1 in [8, Section V.6]. □

Let ℓ be defined by

$$\ell := \liminf_{x \downarrow 0} \frac{g(x)}{x} . \tag{4.6}$$

Note that $1 \le \ell \le L$, cf. (3.1); ℓ may be infinite. If g is differentiable at the origin with derivative value $g'(0)$, then $\ell = L = g'(0)$.

LEMMA 4.5. *Let* u *be a nontrivial solution of* (1.1) *such that* $\lim_{x \to -\infty} u(x) = 0$. *If* $\ell > 1$, *then there exists a* $\delta > 0$ *such that the integral* $\int_{-\infty}^{\infty} e^{-\lambda x} u(x)dx$ *converges for* $\lambda \in S_\delta$, *where* $S_\delta := \{\lambda \in \mathbb{C}: 0 < \mathrm{Re}\lambda < \delta\}$.

PROOF. It suffices to prove the convergence of the integral $\int_{-\infty}^{r} e^{-\lambda x}u(x)dx$ for some $r < \infty$ and λ real, $0 < \lambda < \delta$.

Take $\alpha \in (1,\ell)$ and let c be such that $\int_{-c}^{c} k(x)dx = (\alpha+1)/(2\alpha)$. Define a new function k_c on $\mathbb{R}$ by putting $k_c(x) := k(x)$ if $|x| \le c$, $k_c(x) := 0$ if $|x| > c$. If u is a nontrivial solution of (1.1), then $u(x) = (g\circ u)*k_c(x) + \psi_1(x)$ for all $x \in \mathbb{R}$, with $\psi_1(x) := (g\circ u)*(k-k_c)(x) \ge 0$. We rewrite the latter identity in the form

$$(4.7) \qquad u(x) - \frac{2\alpha}{\alpha+1}\, u * k_c(x) = \frac{\alpha-1}{\alpha+1}\, u(x) + \psi(x) \qquad x \in \mathbb{R},$$

where $\psi(x) := \frac{2}{\alpha+1}[\psi_1(x) + \psi_2(x)]$, with $\psi_2(x) := (g\circ u-\alpha u)*k_c(x)$. Let

$$n_c(x) := \begin{cases} \dfrac{2\alpha}{\alpha+1}\displaystyle\int_x^{\infty} k_c(y)dy & x \ge 0, \\[2ex] -\dfrac{2\alpha}{\alpha+1}\displaystyle\int_{-\infty}^{x} k_c(y)dy & x < 0. \end{cases}$$

Then $\mathrm{supp}(n_c) \subset [-c,c]$. We integrate both members of (4.7) from x to some $r < \infty$. Thus,

$$\int_x^r u(\xi)d\xi - \frac{2\alpha}{\alpha+1}\int_x^r\int_{-\infty}^{\infty} u(\xi-\eta)k_c(\eta)d\eta d\xi = \frac{\alpha-1}{\alpha+1}\int_x^r u(\xi)d\xi + \int_x^r \psi(\xi)d\xi.$$

Interchanging the order of the integrals in the second term and integrating by parts we obtain the identity

$$(4.8) \qquad u*n_c(r) - u*n_c(x) = \frac{\alpha-1}{\alpha+1}\int_x^r u(\xi)d\xi + \int_x^r \psi(\xi)d\xi,$$

cf. (4.5). Because $\lim_{x\to-\infty} u(x) = 0$, there exists $x_\alpha \in \mathbb{R}$ such that $g\circ u(x) \ge \alpha u(x)$ for all $x \in (-\infty,x_\alpha]$ and, hence, $\psi_2(x) \ge 0$ for all $x \in (-\infty,x_\alpha-c]$. The same is then true of $\psi(x)$ and we conclude, as in the proof of Lemma 4.2, that $\lim_{x\to-\infty}\int_x^r u(\xi)d\xi$ and $\lim_{x\to-\infty}\int_x^r\psi(\xi)d\xi$ exist. Letting x tend to $-\infty$ we obtain the identity

$$\int_{-\infty}^{r} u(x)dx = \frac{\alpha+1}{\alpha-1}\left[u*n_c(r) - \int_{-\infty}^{r}\psi(x)dx\right].$$

If r is chosen such that $r \le x_\alpha - c$ it follows that

$$\int_{-\infty}^{r} u(x)dx \le \frac{\alpha+1}{\alpha-1} \max\{u(x): x \in [r-c,r+c]\} \int_{\mathbb{R}} |n_c(x)|dx.$$

Next, we put $u^{(0)}(x) := u(x)$ and define $u^{(k)}(x) := \int_{-\infty}^{x} u^{(k-1)}(\xi)d\xi$, $x \in \mathbb{R}$, for those values of $k \in \mathbb{N}_+$ for which the integral exists. The functions $\psi^{(k)}$ are defined similarly. Integration of (4.7) from $-\infty$ to x yields an equation for $u^{(1)}$ similar to (4.7), with ψ replaced by $\psi^{(1)}$, from which one deduces the relation

$$u^{(1)}*n_c(r) - u^{(1)}*n_c(x) = \frac{\alpha-1}{\alpha+1} \int_{x}^{r} u^{(1)}(\xi)d\xi + \int_{x}^{r} \psi^{(1)}(\xi)d\xi,$$

cf. (4.8). The integrability of u on $(-\infty,r)$ and the continuity of u imply the boundedness of $u^{(1)}$ on $(-\infty,r+c]$. As before, it follows that $\lim_{x\to-\infty} \int_x^r u^{(1)}(\xi)d\xi$ and $\lim_{x\to-\infty} \psi^{(1)}(\xi)d\xi$ exist,

$$\int_{-\infty}^{r} u^{(1)}(x)dx = \frac{\alpha+1}{\alpha-1} \left[u^{(1)}*n_c(r) - \int_{-\infty}^{r} \psi^{(1)}(x)dx\right]$$

and, therefore,

$$\int_{-\infty}^{r} u^{(1)}(x)dx \le \frac{\alpha+1}{\alpha-1} \max\{u^{(1)}(x): x \in [r-c,r+c]\} \int_{\mathbb{R}} |n_c(x)|dx.$$

Continuing in this manner we obtain the sequence of inequalities,

$$\int_{-\infty}^{r} u^{(k)}(x)dx \le \frac{\alpha+1}{\alpha-1} \max\{u^{(k)}(x): x \in [r-c,r+c]\} \int_{\mathbb{R}} |n_c(x)|dx,$$

for $k \in \mathbb{N}_+$. Now, for each $k \in \mathbb{N}_+$, $u^{(k)}$ is a monotonically increasing function, so

$$\begin{aligned} \max\{u^{(k)}(x): x \in [r-c,r+c]\} &= u^{(k)}(r+c) \\ &\le \int_{-\infty}^{r} u^{(k-1)}(x)dx + cu^{(k-1)}(r+c) \\ &\le \delta^{-1}u^{(k-1)}(r+c), \end{aligned}$$

where $\delta := [(\alpha+1)(\alpha-1)^{-1} \int_{\mathbb{R}} |n_c(x)|dx+c]^{-1}$. Hence, by induction,

$$\max\{u^{(k)}(x): x \in [r-c,r+c]\} \le \delta^{-k}\max\{u(x): x \in [r-c,r+c]\} \le p\delta^{-k}.$$

Thus, there exists a positive constant C such that

$$\int_{-\infty}^{r} u^{(k)}(x)dx \le C\delta^{-k} \qquad k \in \mathbb{N}_+ .$$

Next, we apply Lemma 4.4 to the function u and conclude that each of the integrals $\int_{-\infty}^{r} (r-x)^k u(x)dx$ converges. Moreover,

$$\frac{1}{k!}\int_{-\infty}^{r} (r-x)^k u(x)dx \le C\delta^{-k}.$$

This result, in turn, implies the convergence of the infinite sum $\sum_{k=0}^{\infty} (k!)^{-1}\lambda^k \int_{-\infty}^{r} (r-x)^k u(x)dx$ and therefore of the integral $\int_{-\infty}^{r} e^{\lambda(r-x)}u(x)dx$, at least for $\lambda < \delta$. □

LEMMA 4.6. *Let* u *be a nontrivial solution of* (1.1). *If the integral* $\int_{-\infty}^{\infty} e^{-\lambda x}u(x)dx$ *converges in the strip* S_δ (cf. Lemma 4.5) *then so does the integral* $\int_{-\infty}^{\infty} e^{-\lambda x}k(x)dx$.

PROOF. It suffices to prove the convergence of the integral $\int_{-\infty}^{\infty} e^{-\lambda x}k(x)dx$ for λ real, $0 < \lambda < \delta$.

Since u is a nontrivial solution of (1.1) we have $u*k(x) \le u(x)$ for all $x \in \mathbb{R}$, so $\int_{-\infty}^{\infty} e^{-\lambda x}u*k(x)dx \le \int_{-\infty}^{\infty} e^{-\lambda x}u(x)dx < \infty$ for any $\lambda \in (0,\delta)$. On the other hand, $\int_{-\infty}^{\infty} e^{-\lambda x}u*k(x)dx = (\int_{-\infty}^{\infty} e^{-\lambda x}k(x)dx)(\int_{-\infty}^{\infty} e^{-\lambda x}u(x)dx)$, in the sense that if the integral on the left hand side exists then so do the integrals on the right hand side. Hence, $\int_{-\infty}^{\infty} e^{-\lambda x}k(x)dx$ exists for $\lambda \in (0,\delta)$ and satisfies the inequality $\int_{-\infty}^{\infty} e^{-\lambda x}k(x)dx \le 1$ there. □

We remark that Lemmas 4.5 and 4.6 together imply that, if $\ell > 1$ then a necessary condition for the existence of nontrivial solutions of (1.1) is that $\int_{-\infty}^{\infty} e^{-\lambda x}k(x)dx$ converges in S_δ for some $\delta > 0$.

The above results lead us to define a real number Λ_k,

$$(4.9)\qquad \Lambda_k := \sup\left\{\lambda \in \mathbb{R} : \int_{\mathbb{R}} e^{-\lambda x}k(x)\,dx < \infty\right\}$$

and to associate with each nontrivial solution u of (1.1) a number Λ_u thus,

$$(4.10)\qquad \Lambda_u := \sup\left\{\lambda \in \mathbb{R} : \int_{\mathbb{R}} e^{-\lambda x}u(x)\,dx < \infty\right\} .$$

It follows from Lemma 4.6 that $\Lambda_u \le \Lambda_k$. Furthermore, let

$$K(\lambda) := \hat{k}(-i\lambda) = \int_{\mathbb{R}} e^{-\lambda x}k(x)\,dx,$$

$$U(\lambda) := \hat{u}(-i\lambda) = \int_{\mathbb{R}} e^{-\lambda x}u(x)\,dx,$$

whenever the defining integrals converge. It is a standard result from Laplace transform theory that if $\Lambda_u > 0$ then U is analytic in the strip $S_u := \{\lambda \in \mathbb{C}: 0 < \mathrm{Re}\lambda < \Lambda_u\}$, cf. [17, Section VI.4]. The following result is a consequence of the positivity of u(x).

LEMMA 4.7. *If* u *is a nontrivial solution of* (1.1) *and* $0 < \Lambda_u < \infty$, *then* $U(\lambda)$ *is singular at* $\lambda = \Lambda_u$.

PROOF. Cf. [17, Section II.5]. □

At this point we make two further hypotheses concerning the functions g and k:

(H_g^2) the function g is differentiable at 0 with derivative value g'(0) and $g(x) = g'(0)x + O(x^{1+\varepsilon})$ as $x \downarrow 0$ for some $\varepsilon > 0$;

(H_k^2) the kernel k is exponentially small at $-\infty$, i.e., there exists a $\delta > 0$ such that $e^{-\lambda x}k(x)$ is bounded for $\lambda \in (0,\delta)$.

Under these additional hypotheses we can prove the following result.

LEMMA 4.8. *Let* u *be a nontrivial solution of* (1.1) *with* $\lim_{x\to-\infty} u(x) = 0$. *If* $0 < \Lambda_u < \Lambda_k$, *then* $1 - g'(0)K(\Lambda_u) = 0$.

PROOF. We rewrite (1.1) as a linear inhomogeneous integral equation,

$$u(x) - g'(0)u*k(x) = r(x) \qquad x \in \mathbb{R},$$

where $r(x) := (g\circ u-g'(0)u)*k(x)$. A two-sided Laplace transform yields the equation

$$(4.11) \qquad (1-g'(0)K(\lambda))U(\lambda) = R(\lambda),$$

which is valid for all $\lambda \in S_{\Lambda_u}$. In fact, the right abscissa of convergence of R lies to the right of Λ_u. This can be shown in the following manner. Because of the hypothesis (H_k^2), $u(x)e^{-\lambda x} = (g\circ u)*k(x)e^{-\lambda x} \le C\int_{\mathbb{R}} g\circ u(y)e^{-\lambda y}dy$. The integral is bounded for $\mathrm{Re}\lambda$ sufficiently small, since $g(x) < Cx$ for x near zero $(x>0)$. Hence, using the hypothesis (H_g^2) we obtain for $\lambda \in (0,\Lambda_u)$ and ν sufficiently small $(\nu>0)$,

$$\begin{aligned}
&\left|\int_{-\infty}^{\infty} e^{-(\lambda+\nu)x}(g\circ u-g'(0)u)*k(x)dx\right| \\
&= K(\lambda+\nu)\int_{-\infty}^{\infty} e^{-(\lambda+\nu)x}|(g\circ u-g'(0)u)(x)|dx \\
&\le C\,K(\lambda+\nu)\int_{-\infty}^{\infty} e^{-\lambda x}u(x)(e^{-(\nu/\varepsilon)x}u(x))^{\varepsilon}dx \\
&\le C\,K(\lambda+\nu)U(\lambda)(\sup\{e^{-(\nu/\varepsilon)x}u(x): x \in \mathbb{R}\})^{\varepsilon}.
\end{aligned}$$

This result shows that the abscissa of convergence of R lies to the right of Λ_u. Consequently, $R(\lambda)$ is analytic in the neighborhood of $\lambda = \Lambda_u$. Since $U(\lambda)$ is singular at $\lambda = \Lambda_u$ (Lemma 4.7) we conclude from (4.11) that $1 - g'(0)K(\lambda) = 0$ at $\lambda = \Lambda_u$. □

The preceding lemmas lead to the following non-existence theorem.

<u>THEOREM 4.9</u>. *Suppose* $m > 0$, $g'(0) > 1$ *and* $1 - g'(0)K(\lambda) \neq 0$ *for all* $\lambda \in [0,\Lambda_k)$. *If there exists a* $\lambda < \Lambda_k$ *such that* $\int_{-\infty}^{0} e^{-\lambda x}k(x)dx > (g'(0))^{-1}$, *then there is no nontrivial solution of* (1.1).

PROOF. Suppose (1.1) has a nontrivial solution u. Then $\lim_{x\to-\infty} u(x) = 0$ by Lemma 4.2, and $\Lambda_u > 0$ by Lemma 4.5 and the definition (4.10) of Λ_u. For any $\alpha \in (1,g'(0))$ there exists an x_α such that $g\circ u(x) \geq \alpha u(x)$ for all $x \leq x_\alpha$. Take any $c > 0$. Then $u(x) \geq \alpha \int_{-c}^{\infty} u(x-y)k(y)dy$ whenever $x \leq x_\alpha - c$ and, consequently,

$$u(x)e^{-\lambda x} \geq \alpha \int_{-c}^{\infty} e^{-\lambda(x-y)}u(x-y)e^{-\lambda y}k(y)dy$$

for real λ. Now, $\liminf_{x\to-\infty} u(x)e^{-\lambda x} = 0$ for any $\lambda \in [0,\Lambda_u)$, so if we take a monotonically decreasing sequence $\{y_n : n \in \mathbb{N}_+\}$ such that $y_n \downarrow 0$ as $n \to \infty$, and define x_n for $n = n_0, n_0+1, \ldots$, where n_0 is such that $u(0) > y_{n_0}$, by $x_n := \inf\{x \in \mathbb{R}_- : u(y)e^{-\lambda y} \geq y_n \text{ for all } y \in [x,0]\}$, then $u(x_n)e^{-\lambda x_n} = y_n$ and $x_n \to -\infty$ as $n \to \infty$. Now $[-c,0] \subset [x_n,0]$ for n sufficiently large, so $u(x_n-y)e^{-\lambda(x_n-y)} \geq y_n$ for all $y \in [-c,0]$ and

$$\begin{aligned} y_n = u(x_n)e^{-\lambda x_n} &\geq \alpha \int_{-c}^{\infty} e^{-\lambda(x_n-y)}u(x_n-y)e^{-\lambda y}k(y)dy \\ &\geq \alpha \int_{-c}^{0} e^{-\lambda(x_n-y)}u(x_n-y)e^{-\lambda y}k(y)dy \\ &\geq \alpha y_n \int_{-c}^{0} e^{-\lambda y}k(y)dy. \end{aligned}$$

Thus, we obtain the inequality

$$\alpha \int_{-c}^{0} e^{-\lambda y}k(y)dy \leq 1, \tag{4.12}$$

which is valid for any $\lambda \in [0,\Lambda_u)$. If Λ_u were strictly less than Λ_k, then the function $\lambda \mapsto 1 - g'(0)K(\lambda)$ would have a zero at Λ_u by Lemma 4.8, i.e., inside the interval $[0,\Lambda_k)$, contrary to the hypothesis of the theorem. Hence, $\Lambda_u = \Lambda_k$ and the inequality (4.12) holds for all $\lambda \in [0,\Lambda_k)$. Since α and c were chosen arbitrarily subject to the constraints $1 < \alpha < g'(0)$ and $c > 0$ we conclude that the assumption of the existence of a nontrivial solution of (1.1) leads to the validity of the inequality $g'(0)\int_{-\infty}^{0} e^{-\lambda y}k(y)dy \leq 1$ for all $\lambda \in [0,\Lambda_k)$. This proves the theorem. □

If $\Lambda_k = \infty$, or if $K(\lambda) \to \infty$ for $\lambda \uparrow \Lambda_k'$, then there certainly exists a $\lambda < \Lambda_k$ such that $\int_{-\infty}^{0} e^{-\lambda x}k(x)dx > (g'(0))^{-1}$. If, on the other hand, $K(\lambda)$ approaches a finite limit as $\lambda \uparrow \Lambda_k$, then Theorem 4.9 is not necessarily applicable. In that case, one can conclude from a Tauberian theorem for Laplace transforms, cf. [17, Section V.4], that the function $u_\lambda: x \mapsto u(x)e^{-\lambda x}$ belongs to $L_1(\mathbb{R})$ for $\lambda = \Lambda_k$. This implies that, under the hypothesis that $k_\lambda: x \mapsto k(x)e^{-\lambda x}$ is bounded for $\lambda = \Lambda_k$, u_λ is bounded for $\lambda = \Lambda_k$. Then one can repeat the analysis of the present section, starting from the identiy $u_\lambda(x) = \int_{\mathbb{R}} e^{-\lambda y}(g\circ u)(y)k_\lambda(x-y)dy$ for $\lambda = \Lambda_k$, to obtain a non-existence result. We do not elaborate this idea any further.

REMARK 4.10. The condition that k has a finite first absolute moment ($\int_{\mathbb{R}}|x|k(x)dx < \infty$) has been used only once, in Lemma 4.2, to prove that any nontrivial solution of (1.1) has limits as $x \to \pm\infty$. The proofs of Lemmas 4.5 through 4.8 and Theorem 4.9 proceeded without an explicit reference to this condition. It remains an open question whether, if k does not have a finite first absolute moment, there exist nontrivial solutions of (1.1) which do not have limits as $x \to \pm\infty$.

5. THE INHOMOGENEOUS EQUATION

As the greater part of the proofs in the foregoing section were based on inequalities, rather than equalities, the same or similar arguments can be used in the analysis of the inhomogeneous equation

$$u(x) = (g\circ u)*k(x)+f(x) \qquad x \in \mathbb{R}. \tag{5.1}$$

In fact, let f satisfy the following hypothesis:

(H_f) $f: \mathbb{R} \to \mathbb{R}$ is uniformly continuous, nonnegative and not equal to the zero function.

We then have the following result.

THEOREM 5.1. *Suppose* $g(x) \geq p$ *for* $x \geq p$. *The same hypotheses on* g *and* k *which lead to the non-existence results of Theorem* 4.1 *and Theorem* 4.9 *guarantee that any bounded nonnegative solution* u *of* (5.1) *necessarily*

satisfies the inequality $u(x) \geq p$ *for all* $x \in \mathbb{R}$.

PROOF. Let u satisfy (5.1). We introduce the function w by the definition $w(x) := \min\{u(x),p\}$, $x \in \mathbb{R}$. Then $g \circ u(x) \geq w(x)$ for all $x \in \mathbb{R}$, so $u(x) \geq (g \circ u) * k(x) \geq w * k(x)$. But $w(x) \leq p$ for all $x \in \mathbb{R}$, so $w * k(x) \leq p \int_{\mathbb{R}} k(x)dx = p$. Combining these two results we see that w satisfies the inequality $w * k(x) \leq \min\{u(x),p\} = w(x)$ for all $x \in \mathbb{R}$. Under the hypotheses of Theorem 4.1, this inequality leads to the conclusion that w is constant, $w(x) = c$ say. The case $0 \leq c < p$ is excluded, because then $u(x) = c$ and (5.1) cannot be satisfied. Hence, $w(x) = p$ and $u(x) \geq p$ for all $x \in \mathbb{R}$. Under the hypotheses of Theorem 4.9 it follows that $w(-\infty) := \lim_{x \to -\infty} w(x)$ and $w(\infty) := \lim_{x \to \infty} w(x)$ exist and are such that $w(\infty) - w(-\infty) \geq 0$ and $\{w(-\infty), w(\infty)\} \subset \{0,p\}$, cf. Lemma 4.2. One shows as in Lemma 2.2 that $w(x) > 0$ for all $x \in \mathbb{R}$, and that either $w(x) = p$ for all $x \in \mathbb{R}$ or $\inf\{w(x): x \in \mathbb{R}\} = 0$. It follows that either $w(x) = p$ for all $x \in \mathbb{R}$ or $w(-\infty) = 0$. In the latter case, the chain of arguments leading from Lemma 4.4 to Theorem 4.9 can be repeated mutatis mutandis. Under the conditions of Theorem 4.9 the assumption that there exists a nontrivial w such that $\lim_{x \to -\infty} w(x) = 0$ leads to a contradiction, so necessarily $w(x) = p$, i.e., $u(x) \geq p$, for all $x \in \mathbb{R}$. □

REMARK 5.2. A close examination of the proof of Theorem 5.1 shows that the condition (H_f) can be weakened somewhat if there exists a monotonically nondecreasing function g_0 on $[0,p]$ such that $g(x) \geq g_0(x) \geq x$ for all $x \in [0,p]$. In fact, let $f: \mathbb{R} \to \mathbb{R}$ be such that $f(x) \geq f_0(x)$ for all $x \in \mathbb{R}$, where $f_0: \mathbb{R} \to \mathbb{R}$ satisfies (H_f). Suppose u is a solution of (5.1). Consider the equation

$$(5.2) \qquad v(x) = (g_0 \circ v) * k(x) + f_0(x) \qquad x \in \mathbb{R}.$$

The minimal solution of (5.2) can be constructed via the method of successive iterations, $v^{(0)}(x) := f_0(x)$, $v^{(n)}(x) := (g_0 \circ v^{(n-1)}) * k(x) + f_0(x)$ $(n \in \mathbb{N}_+)$. Any of these iterates is majorized by the solution u of (5.1). The monotonicity of g_0 implies that the sequence $\{v^{(n)}(x): n \in \mathbb{N}_+\}$ is monotonically nondecreasing for each $x \in \mathbb{R}$, so the sequence converges to a limit function, v say, which satisfies the equation (5.2) and the inequality

BIBLIOTHEEK MATHEMATISCH CENTRUM
AMSTERDAM

$v(x) \le u(x)$ for all $x \in \mathbb{R}$. Suppose now that the function g_0 is such that any bounded nonnegative solution v of (5.2) necessarily satisfies the inequality $v(x) \ge p$ for all x, according to the theorem, then the solution u of (5.1) must also satisfy the inequality $u(x) \ge p$ for all $x \in \mathbb{R}$.

6. ASYMPTOTIC BEHAVIOR AND UNIQUENESS CRITERIA

The key assumption leading to the non-existence result of Theorem 4.9 is that the function $\lambda \mapsto 1 - g'(0)K(\lambda)$ does not vanish in the interval $[0,\Lambda_k)$. In the present section we will drop this assumption, i.e., we assume that the function $\lambda \mapsto 1 - g'(0)K(\lambda)$ has a real positive zero. The nonnegativity of k implies that K is a convex function on $[0,\Lambda_k)$, so $\lambda \mapsto 1 - g'(0)K(\lambda)$ has at most two real positive zeros. Let σ denote the smallest positive zero; we assume that it is simple.

In Section 4 we established various conditions which guarantee that any nontrivial solution of (1.1) has the property that $\lim_{x\to-\infty} u(x) = 0$. We also analyzed the rate of convergence of $u(x)$ to its limiting value 0 in terms of the Laplace transform of u. Our first objective in the present section is to establish conditions which guarantee that any nontrivial solution of (1.1) has the much stronger property that $u(x) \sim e^{\sigma x}$ as $x \to -\infty$. We will then use this knowledge of the asymptotic behavior to establish a uniqueness result for the nontrivial solutions of (1.1).

The following lemma is a modified version of Ikehara's Theorem and will be needed in the proof of Theorem 6.2.

<u>LEMMA 6.1</u>. *If the real-valued function ϕ is nonnegative and nonincreasing on $\mathbb{R}_+$, and there exists a $\tau > 0$ such that the integral $f(\lambda) = \int_0^\infty e^{-\lambda x}\phi(x)dx$ converges for $\mathrm{Re}\lambda > -\tau$, and if, furthermore, for some constant A and some function $g\colon \mathbb{R} \to \mathbb{R}$,*

$$\lim_{x\downarrow-\tau} [f(x+iy) - A(\tau+x+iy)^{-1}] = g(y),$$

uniformly on compact subsets of $\mathbb{R}$, then

$$\lim_{x\to\infty} \phi(x)e^{\tau x} = A.$$

PROOF. Analogous to the proof of Ikehara's Theorem, see e.g. [17, Section V.17]. □

THEOREM 6.2. *Suppose* $m > 0$, $g'(0) > 1$ *and* $g(x) \le g'(0)x$ *for* $x \in [0,p]$. *If* u *is a monotone nondecreasing nontrivial solution of* (1.1), *then there exists a positive constant* A *such that* $\lim_{x\to-\infty} u(x)e^{-\sigma x} = A$.

PROOF. Let u be a monotone nondecreasing nontrivial solution of (1.1). It follows from Lemma 4.5 and the definition (4.10) of Λ_u that $U(\lambda)$ is defined for $\lambda \in [0,\Lambda_u)$ for some $\Lambda_u > 0$. Moreover, $[1 - g'(0)K(\lambda)]U(\lambda) = R(\lambda)$, cf. (4.11). The continuity of g, together with the properties $g(0) = 0$, $g'(0) > 1$, and $g(0) = p$, imply that the inequality $\le$ in the condition $g(x) \le g'(0)x$ for $x \in [0,p]$ in the statement of the theorem, is a strict inequality, $g(x) < g'(0)x$, at least on some subinterval of $[0,p]$. Hence, $g\circ u(x) < g'(0)u(x)$ on a set of positive measure on $\mathbb{R}$ and, consequently, the function $R: \lambda \mapsto \int_{\mathbb{R}} e^{-\lambda x}(g\circ u-g'(0)u)*k(x)dx$ does not vanish at $\lambda = \sigma$. Thus,

$$U(\lambda) = \frac{R(\lambda)}{1-g'(0)K(\lambda)} \sim \frac{R(\sigma)}{g'(0)K'(\sigma)}(\sigma-\lambda)^{-1} \quad \text{as} \quad \lambda \uparrow \sigma.$$

The function R is regular in a neighborhood of the line $\mathrm{Re}\lambda = \sigma$. The zeros of the function $\lambda \mapsto 1 - g'(0)K(\lambda)$, which lie in a vertical strip left-adjacent to the line $\mathrm{Re}\lambda = \sigma$ lie, in fact, in a rectangle according to the Lemma of Riemann-Lebesgue [15, Section 1.8]. The analyticity of K then implies that there are only finitely many inside this rectangle. Hence, if we choose the width of the strip sufficiently small, then there are no other zeros of $\lambda \mapsto 1 - g'(0)K(\lambda)$ in this strip, besides the real zero $\lambda = \sigma$, and we can apply Lemma 6.1 (with the obvious modifications) to obtain the statement of the theorem. □

The use of an Ikehara-type lemma makes the proof of the asymptotic behavior rather simple, but has the disadvantage that the conclusion holds only for monotone solutions. Our next objective is to give an alternate proof of the asymptotic behavior, under slightly different assumptions on k, without presupposing that the solution be monotone.

THEOREM 6.3. *Suppose* $m > 0$, $g'(0) > 1$, *and* $g(x) \le g'(0)x$ *for* $x \in [0,p]$. *Let* $\varepsilon > 0$ *be such that* $g(x) = g'(0)x + O(x^{1+\varepsilon})$, cf. (H_g^2). *Suppose that* $K(-\delta) < \infty$ *for some* $\delta > 0$, *and that there exists a* $\beta \in (\sigma,\sigma(1+\varepsilon))$ *such that*

(i) $g'(0)K(\beta) < 1$, *and*
(ii) $k(x)e^{-(\beta/(1+\varepsilon))x}$ *is bounded.*

If u *is a nontrivial solution of* (1.1), *then there exists a positive constant* A *such that* $\lim_{x\to-\infty} u(x)e^{-\sigma x} = A$.

PROOF. In this proof we use for any given function $\phi: \mathbb{R} \to \mathbb{R}$ the symbol ϕ_λ to denote the mapping $x \mapsto \phi(x)e^{-\lambda x}$, $x \in \mathbb{R}$. Let u be a nontrivial solution of (1.1). We know that $u_{\beta/(1+\varepsilon)} \in L_1(\mathbb{R})$ and $k_{\beta/(1+\varepsilon)} \in L_\infty(\mathbb{R})$. Hence, $u_{\beta/(1+\varepsilon)}(x) \le g'(0)u_{\beta/(1+\varepsilon)}*k_{\beta/(1+\varepsilon)}(x) \le C$, or, in other words, $u(x) \le Ce^{-(\beta/(1+\varepsilon))x}$ for all $x \in \mathbb{R}$. We rewrite (1.1) as an inhomogeneous linear convolution equation,

$$(6.1) \qquad u(x) = g'(0)u*k(x) + r(x) \qquad x \in \mathbb{R},$$

where $r(x) := (g\circ u - g'(0)u)*k(x)$. There exists a positive constant C such that $0 \ge r(x) \ge -Cu^{1+\varepsilon}*k(x) \ge -Ce^{-\beta x}$ for all $x \in \mathbb{R}$. Multiplying both sides of (6.1) with $e^{-\beta x}$ we obtain

$$(6.2) \qquad u_\beta(x) = g'(0)u_\beta*k_\beta(x) + r_\beta(x) \qquad x \in \mathbb{R}.$$

The function r_β is nonpositive and bounded. Consider the iterative scheme $v^{(0)}(x) := r_\beta(x)$, $v^{(n)}(x) := g'(0)v^{(n-1)}*k_\beta(x) + r_\beta(x)$, $n \in \mathbb{N}_+$. For each x, the sequence $\{v^{(n)}(x): n \in \mathbb{N}_+\}$ is monotone nonincreasing. Moreover, $v^{(n)}(x) \ge \inf\{r_\beta(x): x \in \mathbb{R}\}\,(1+\alpha+\alpha^2+\ldots+\alpha^n)$, where $\alpha = g'(0)K(\beta) < 1$. So the sequence $\{v^{(n)}: n \in \mathbb{N}_+\}$ being bounded and monotone, converges to a limit function, v say, which satisfies the equation $v(x) = g'(0)v*k_\beta(x) + r_\beta(x)$ for all $x \in \mathbb{R}$. The same equation is satisfied by the function u_β, cf. (6.2). Since v is nonpositive and u_β is positive, the difference $u_\beta - v$ is not the zero function. Let the function w be defined by $w_\beta(x) := u_\beta(x) - v(x)$. Then w satisfies the homogeneous linear convolution equation

$$(6.3) \qquad w(x) = g'(0)w*k(x) \qquad x \in \mathbb{R},$$

and the estimates $0 \le w(x) \le C(1+e^{\beta x})$ for all $x \in \mathbb{R}$. We claim that this implies that $w(x) = Ae^{\sigma x}$ for some $A > 0$.

In order to support this claim we first collect some information concerning the roots of the characteristic equation $g'(0)K(\lambda) = 1$ in the strip $S_\beta = \{\lambda \in \mathbb{C}: 0 \le \mathrm{Re}\lambda \le \beta\}$. In this strip, σ is the only real root; all other roots occur in complex conjugate pairs. The nonnegativity of k implies that the real part of any root other than σ is strictly less than σ. As in the proof of Theorem 6.2 it follows that there are only finitely many roots in S_β.

The function w, being a solution of (6.3) which satisfies the estimates $0 \le w(x) \le C(1+e^{\beta x})$, is a linear combination of exponential functions, possibly multiplied by polynomial functions, where the exponents are found as the roots of the characteristic equation in S_β, cf. [15, Section 11.2]. (It is in this step that the assumption $K(-\delta) < \infty$ is used.) Next, observe that any contribution from a pair of complex conjugate roots (with real parts less than σ) leads to an oscillatory behavior of w which is asymptotically dominant over $e^{\sigma x}$ as $x \to -\infty$. Hence, if there were a contribution to w from any complex conjugate pair of roots of the characteristic equation in the strip S_β, w(x) would certainly become negative for some x sufficiently large negative, which is impossible because w(x) is nonnegative, as we have seen. The conclusion is, therefore, that the only nonzero coefficient in the finite expansion for w is the one corresponding to σ. In other words, $w(x) = Ae^{\sigma x}$ for some $A > 0$, as claimed.

Finally, $u(x) = w(x) + v(x)e^{\beta x}$ and so $\lim_{x\to-\infty} u(x)e^{-\sigma x} = \lim_{x\to-\infty} w(x)e^{-\sigma x} = A$. □

Our final results concern the uniqueness of nontrivial solutions of (1.1). We observe that the homogeneous equation (1.1) is invariant under translation, so uniqueness is to be understood in the sense of uniqueness modulo translation.

THEOREM 6.4. *Suppose that the conditions of Theorem* 6.3 *are satisfied and that, in addition,* g *is such that* $|g(x)-g(y)| \le g'(0)|x-y|$ *for all pairs*

$x,y \in [0,p]$. *Then there is at most one nontrivial solution (modulo translation) of* (1.1).

PROOF. Let u_1^* and u_2 be two nontrivial solutions of (1.1). There exist positive constants C^* and C such that $\lim_{x\to-\infty} u_1^*(x)e^{-\sigma x} = C^*$ and $\lim_{x\to-\infty} u_2(x)e^{-\sigma x} = C$, according to Theorem 6.3. Define u_1 by putting $u_1(x) := u_1^*(x-\tau)$ for all $x \in \mathbb{R}$, where $\tau = \sigma^{-1}\ell n(C^*/C)$. Then u_1 is also a nontrivial solution of (1.1) and $\lim_{x\to-\infty} u_1(x)e^{-\sigma x} = C$. Thus, if we consider the function v defined by $v(x) := (u_1-u_2)(x)e^{-\sigma x}$ for all $x \in \mathbb{R}$, then v is continuous on $\mathbb{R}$, $\lim_{x\to\pm\infty} v(x) =: v(\pm\infty)$ exist, and $v(-\infty) = v(\infty) = 0$. Also, if $z := \sup\{|v(x)|: x \in \mathbb{R}\}$, then $z = v(x_0)$ for some $x_0 \in \mathbb{R}$. Now, $v(x) = (u_1-u_2)(x)e^{-\sigma x} = (g\circ u_1-g\circ u_2)*k(x)e^{-\sigma x}$, so since the Lipschitz constant of g on $[0,p]$ is at most equal to $g'(0)$ we have the inequality $|v(x)| \le g'(0)|v|*k_\sigma(x)$, where $k_\sigma(x) := k(x)e^{-\sigma x}$, which is valid for all $x \in \mathbb{R}$. From this point on, the arguments parallel some of the arguments used in the proof of Lemma 2.2, cf. also FELLER [8, Section XI.2]. Upon induction we obtain the estimate

$$(6.4) \qquad |v(x)| \le (g'(0))^n|v|*k_\sigma^{n*}(x) \qquad x \in \mathbb{R},$$

for any $n \in \mathbb{N}_+$. Now, the expression in the right member of this inequality is at most equal to $z(g'(0))^n \int_{\mathbb{R}} k_\sigma^{n*}(x)dx$ which, in turn, is equal to $z(g'(0) \int_{\mathbb{R}} k_\sigma(x)dx)^n = z(g'(0)K(\sigma))^n = z$. Hence, at $x = x_0$, the inequality (6.4) must be an equality, $|v(x_0)| = (g'(0))^n|v|*k_\sigma^{n*}(x_0)$ for all $n \in \mathbb{N}_+$. But this is only possible if $v(x_0-y) = v(x_0)$ for all $y \in \Sigma$. (We recall that $\Sigma = \cup_{n=1}^\infty \Sigma_n$, Σ_n being the set consisting of the points of increase of k^{n*}.) It follows that $v(x) = v(x_0)$ for all $x \in \mathbb{R}$ if k is not concentrated on either $\mathbb{R}_+$ or $\mathbb{R}_-$, cf. Lemma 2.1. If k is concentrated on $\mathbb{R}_+$, say, then it follows that $v(x) = v(x_0)$ for all $x \in (-\infty,x_0-a]$ for some $a \in \mathbb{R}_+$, but as v satisfies the inequality (6.4) one may conclude again that $v(x) = v(x_0)$ for all $x \in \mathbb{R}$. Hence, v is a constant function on $\mathbb{R}$. Since its limiting values at $\pm\infty$ are zero, it follows that v is the zero function, i.e., $u_1(x) = u_2(x)$ for all $x \in \mathbb{R}$. □

The proof of the following uniqueness theorem is analogous to the

proof of Theorem 6.4.

THEOREM 6.5. *Suppose that the conditions of Theorem* 6.2 *are satisfied and that, in addition,* g *is such that* $|g(x)-g(y)| \le g'(0)|x-y|$ *for all pairs* $x,y \in [0,p]$. *Then there is at most one monotone nondecreasing nontrivial solution (modulo translation) of*(1.1).

REMARK 6.6. If g is monotone nondecreasing and sublinear (i.e., $g(\alpha x) \ge \alpha g(x)$ for $\alpha \in [0,1]$, $x \in [0,p]$), then the Lipschitz constant of g on $[0,p]$ is at most $g'(0)$. The former properties were used by one of the authors in [5] to prove a uniqueness result. The fact that one can use the less restrictive condition involving the Lipschitz constant was observed by BARBOUR [3] for a function g which was explicitly given by an expression of the form $g(x) = \alpha(1-e^{-x})$.

REMARK 6.7. We observe that a combination of the existence result for monotone solutions (cf. [6], [16]) and the uniqueness result of Theorem 6.4 leads, in a very indirect way, to the conclusion that monotone functions g lead to monotone nontrivial solutions of (1.1) only. It would be of interest to have a direct proof of this fact.

7. APPLICATION TO A DETERMINISTIC MODEL FROM MATHEMATICAL EPIDEMIOLOGY

The spatio-temporal development of an epidemic among a closed population, with habitat $\mathbb{R}^n$, can be described by an integral equation of the form

$$(7.1) \qquad u(t,x) = \int_0^\infty H(\tau) \int_{\mathbb{R}^n} g\circ u(t-\tau,y)V(x-y)dy d\tau$$

$-\infty < t < \infty$, $x \in \mathbb{R}^n$, cf. [6]. Here u is a measure for the density of susceptible individuals. The equation is based on various assumptions, the most important being

(i) the members of the population can be categorized as either susceptible to or infected by the disease;

(ii) the infectivity of an infected individual as a function of time

elapsed since exposure and position relative to the individual's own position is given by $H: \mathbb{R}_+ \to \mathbb{R}$ and $V: \mathbb{R}^n \to \mathbb{R}$, respectively;

(iii) the disease induces permanent immunity, so an individual can pass from the class of susceptibles to the class of infectives, but not vice versa.

Both H and V are nonnegative, and normalized to have integral one. The assumption that the habitat $\mathbb{R}^n$ is homogeneous and isotropic is reflected in the fact that V is a radial function. The nonlinearity of the equation comes about through the function g which has the form $g(x) = \alpha(1-e^{-x})$. The parameter α has a threshold value 1, cf. [1], [6]; here, we restrict our attention to the more interesting case $\alpha > 1$. In the so-called traveling (plane) wave problem one looks for solutions of (7.1) of the form $u(t,x) = w(x\cdot\nu+ct)$, where ν is a fixed unit vector. If one chooses a basis in $\mathbb{R}^n$ such that $\nu = (1,0,\ldots,0)$, then the function w must satisfy the following convolution equation on the line,

$$(7.2) \qquad w(x) = (g\circ w)*V_c(x) \qquad x \in \mathbb{R},$$

where $V_c(x) := \int_0^\infty H(\tau)\tilde{V}(x-c\tau)d\tau$, $x \in \mathbb{R}$, with $\tilde{V}(x_1) := \int_{\mathbb{R}^{n-1}} V(x_1,x_2,\ldots,x_n)dx_2\ldots dx_n$, $x_1 \in \mathbb{R}$. If $c = 0$ the function V_c is symmetric. One can then apply Theorem 4.1, from which it follows that (7.1) does not admit any standing wave solution. Next, we consider the case $c \neq 0$. Because of the symmetry of V we can restrict our attention to the case $c > 0$.

As c increases, the mass of V_c shifts to the right. This can be seen by inspecting the qualitative behavior of the characteristic function L_c, associated with the linearized equation, as a function of the wave velocity c. By definition,

$$L_c(\lambda) := g'(0) \int_{\mathbb{R}} e^{-\lambda y}V_c(y)dy$$

$$= g'(0) \int_{\mathbb{R}_+} e^{-\lambda c\tau}H(\tau)d\tau \int_{\mathbb{R}} e^{-\lambda y}\tilde{V}(y)dy.$$

The nonnegativity of H and V implies that, for c fixed, L_c is a convex function of λ ($\lambda \in \mathbb{R}$) and, for λ fixed ($\lambda > 0$), it is a monotonically decreasing function of c. Furthermore, $L_c'(0) = g'(0) = \alpha > 1$. Hence, the number $c_0 := \inf\{c > 0$: there exists $\lambda > 0$ such that $L_c(\lambda) = 1\}$ is well-defined if H and V satisfy the appropriate hypotheses [6].

Various authors have given constructive proofs for the existence of (monotone) nontrivial solution of (7.2) in the case $c \geq c_0$, cf. [2], [4], [6], [16]. From the results of our investigation it follows that there exists exactly one such solution (modulo translation) at each speed $c > c_0$, at least if $\tilde{V}_\lambda : x \mapsto \tilde{V}(x)e^{-\lambda x}$ is bounded for the value of λ specified in Theorem 6.4. (If, for instance, V has compact support then this condition is certainly fulfilled.) A similar conclusion has been reached at by BARBOUR [3] who used probabilistic arguments.

Furthermore, it follows from the results of our investigation that no traveling waves exist with speed $c < c_0$. The same result has been obtained for the special case $H(t) = ae^{-\mu t}$ (a,μ nonnegative constants) by ATKINSON and REUTER [2] and by ARONSON [1], and for a discrete time model in population genetics by WEINBERGER [16]. Both Aronson and Weinberger based their proofs on the construction of a subsolution. This approach made it actually possible for them to conclude that c_0 is the asymptotic speed of propagation of disturbances from a rest state of the associated initial value problem (see also MOLLISON [12]).

In our opinion, the main advantage of the approach presented in this paper is that the same chain of purely analytical arguments leads to the nonexistence, as well as the uniqueness results.

ACKNOWLEDGMENT

Part of this investigation was carried out while the second author was on foreign assignment to the Mathematisch Centrum, Amsterdam (Neth.). The final draft of this paper was written while the first author was visiting the Applied Mathematics Division at Argonne National Laboratory. Both authors wish to thank their respective host institutions for their hospitality and financial support.

REFERENCES

[1] ARONSON, D.G., *The asymptotic speed of propagation of a simple epidemic*, to appear in: Nonlinear Diffusion, W.E. Fitzgibbon & H.F. Walker, eds., Research Notes in Mathematics, Pitman Publishing Co., 1977.

[2] ATKINSON, C. & G.E.H. REUTER, *Deterministic epidemic waves*, Math. Proc. Camb. Phil. Soc. 80 (1976) 315-330.

[3] BARBOUR, A.D., *The uniqueness of Atkinson and Reuter's epidemic waves*, Math. Proc. Camb. Phil. Soc. 82 (1977) 127-130.

[4] BROWN, K.J. & J. CARR, *Deterministic epidemic waves of critical velocity*, Math. Proc. Camb. Phil. Soc. 81 (1977) 431-435.

[5] DIEKMANN, O., *Limiting behaviour in an epidemic model*, Nonlinear Analysis, Theory, Methods & Applications 1 (1977) 459-470.

[6] DIEKMANN, O., *Thresholds and travelling waves for the geographical spread of infection*, Mathematical Centre Report TW 166, Amsterdam, 1977.

[7] ESSÉN, M., *Studies on a convolution inequality*, Ark. Mat. 5 (1963) 113-152.

[8] FELLER, W., *An Introduction to Probability Theory and Its Applications*, Vol. II, Wiley, New York, 1966.

[9] KARLIN, S., *On the renewal equation*, Pacific J. Math. 5 (1955) 229-257.

[10] KENDALL, D.G., *Mathematical models of the spread of infection*, in: Mathematics and Computer Science in Biology and Medicine, Medical Research Council, London, 1965, 213-224.

[11] KERMACK, W.O. & A.G. McKENDRICK, *A contribution to the mathematical theory of epidemics*, Proc. Roy. Soc. A 115 (1927) 700-721.

[12] MOLLISON, D., *The rate of spatial propagation of simple epidemics*, in: Proc. Sixth Berkeley Symposium on Mathematical Statistics and Probability, Vol. III, L.M. le Cam, J. Neyman & E.L. Scott, eds., Univ. of California Press, Berkeley, 1972, 579-614.

[13] RUDIN, W., *Fourier Analysis on Groups*, Interscience, New York, 1962.

[14] RUDIN, W., *Functional Analysis*, McGraw-Hill, New York, 1973.

[15] TITCHMARSCH, E.C., *Introduction to the Theory of Fourier Integrals*, Clarendon Press, Oxford, 1937.

[16] WEINBERGER, H.F., *Asymptotic behavior of a model in population genetics*, to appear in: Indiana University Seminar in Applied Mathematics, J. Chadam, ed., Springer Lecture Notes.

[17] WIDDER, D.V., *The Laplace Transform*, University Press, Princeton, 1946.

Run for your life. A note on the asymptotic speed of propagation of an epidemic

by

O. Diekmann

ABSTRACT

We study the large-time behaviour of the solution of a nonlinear integral equation of mixed Volterra-Fredholm type describing the spatio-temporal development of an epidemic. For this model it is known that there exists a minimal wave speed c_0 (i.e., travelling wave solutions with speed c exist if $|c| > c_0$ and do not exist if $|c| < c_0$). In this paper we show that c_0 is the asymptotic speed of propagation (i.e., for any c_1, c_2 with $0 < c_1 < c_0 < c_2$ the solution tends to zero uniformly in the region $|x| \geq c_2 t$, whereas it is bounded away from zero uniformly in the region $|x| \leq c_1 t$ for t sufficiently large).

1. INTRODUCTION

In this paper, which may be regarded as a sequel to [4], we shall investigate the large-time behaviour of the solution u of the nonlinear integral equation of mixed Volterra-Fredholm type

$$(1.1)\qquad u(t,x) = \int_0^t A(t-\tau)\int_{\mathbb{R}^n} g(u(\tau,\xi))V(x-\xi)d\xi d\tau + f(t,x), \quad t \geq 0,\ x \in \mathbb{R}^n .$$

In [4] it was shown how an equation of this form arises as a model of the spatio-temporal development of an epidemic and, among other things, the existence, uniqueness and nonnegativity of a solution was established under some suitable assumptions concerning the given functions A, V, g and f.

Equation (1.1) corresponds to an initial value problem (the history up to $t = 0$ is prescribed; in fact it is incorporated in the function f). On the other hand, if one wants to describe an epidemic which has been evolving from the beginning of time then one arrives at the time-translation invariant homogeneous equation

$$(1.2)\qquad u(t,x) = \int_0^\infty A(t-\tau)\int_{\mathbb{R}^n} g(u(\tau,\xi))V(x-\xi)d\xi d\tau, \quad -\infty < t < \infty,\ x \in \mathbb{R}^n .$$

When looking for travelling (plane) wave solutions (i.e., solutions of the form $u(t,x) = w(x\cdot\nu+ct)$, where ν is a fixed unit vector) one has to consider equation (1.2). The investigations in [4] and [5] have revealed that there exists $c_0 > 0$ such that (1.2) has a (for each ν unique modulo translation) travelling wave solution with speed c if $|c| > c_0$ and no such solution if $|c| < c_0$.

With this knowledge available several questions concerning the asymptotic behaviour (as $t \to \infty$) of solutions of (1.1) immediately present themselves. For instance, one might try to characterize those functions f for which the solution of (1.1) converges to a travelling wave of given speed c in some appropriate sense. Or one can investigate whether for a large class of functions f the solution develops into some structure of travelling waves. These questions have been successfuly studied for reaction-diffusion equations (see [6] for a survey of ideas, results and open

problems) and one can acquire a lot of intuition by studying this theory. However, at present it seems that in order to deal with the *asymptotic form* of solutions of (1.1) one has to overcome some hard technical problems.

In two papers ([2],[3]) on Fisher's equation from population genetics, Aronson and Weinberger have introduced the concept of the *asymptotic speed* of propagation of disturbances from the rest state. Roughly speaking $c^* > 0$ is called the asymptotic speed if for any c_1, c_2 with $0 < c_1 < c^* < c_2$, the solution tends to zero uniformly in the region $|x| \geq c_2 t$, whereas it is bounded away from zero uniformly in the region $|x| \leq c_1 t$ for t sufficiently large (see Theorems 1 and 2 for a precise formulation). Any one who has read the papers of Aronson and Weinberger and the opening remarks above will immediately conjecture that for equation (1.1) c_0 is the asymptotic speed of propagation. It is the object of this paper to prove that this is indeed the case.

The organization of the paper is as follows. In section 2 we formulate the hypotheses concerning the functions A, V, g and f and we discuss in some more detail the known results for the equations (1.1) and (1.2). In section 3 we formulate our results in two theorems and we prove one of these, while the second one is proved in section 4 by means of a sequence of lemmas. Finally, in section 5 we review the epidemic model and we interpret our results biologically.

At this place we would like to acknowledge our thanks to Professors D.G. Aronson and H.F. Weinberger for the inspiration that came from their published works and for the stimulation received during a short visit.

NOTATION.

$\mathbb{R}_+ = \{y \in \mathbb{R} \mid y \geq 0\}$.

$BC(\mathbb{R}^n)$: the space of the bounded continuous functions on $\mathbb{R}^n$ equipped with the supremum norm.

$C(\mathbb{R}_+; BC(\mathbb{R}^n))$: the set of functions mapping $\mathbb{R}_+$ continuously into $BC(\mathbb{R}^n)$.

$B_R = \{x \in \mathbb{R}^n \mid |x| \leq R\}$.

supp ϕ: the support of the function ϕ.

$\phi \succ \psi$: we use this notation if ϕ and ψ are continuous functions defined on $\mathbb{R}^n$ and such that $\phi(x) \geq \psi(x)$ with strict inequality for $x \in \operatorname{supp} \psi$.

2. THE CHARACTERISTIC EQUATION AND THE MINIMAL WAVE SPEED c_0

We consider the equation (1.1) under the following hypotheses:

H_A: $A: \mathbb{R}_+ \to \mathbb{R}$ is nonnegative, $A \in L_1(\mathbb{R}_+)$ and $\int_0^\infty A(\tau)d\tau = 1$.

H_V: $V: \mathbb{R}^n \to \mathbb{R}$ is nonnegative, $V \in L_1(\mathbb{R}^n)$ and $\int_{\mathbb{R}^n} V(x)dx = 1$;
V is a radial function;
there exists $\delta > 0$ such that $\int_{\mathbb{R}^n} V(x)e^{\lambda x_1}dx < \infty$ for $\lambda \in (-\delta,\delta)$.

H_g: $g: \mathbb{R} \to \mathbb{R}$ is continuously differentiable;
$g(0) = 0$, $g'(0) > 1$ and there exists $C > 0$ such that $0 < g'(x) < C$ for $x \geq 0$; there exists $p > 0$ such that $g(x) > x$ for $0 < x < p$ and $g(p) = p$.

H_f: $f: \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}$ is nonnegative but not identically zero;
$f \in C(\mathbb{R}_+;BC(\mathbb{R}^n))$.

In [4] it was shown that the existence and the uniqueness of a solution $u \in C(\mathbb{R}_+;BC(\mathbb{R}^n))$ is guaranteed under less restrictive assumptions and that u is nonnegative.

If we look for a travelling wave solution $u(t,x) = w(x.\nu+ct)$ of equation (1.2), then w has to satisfy the following nonlinear convolution equation on the line

$$w(y) = \int_{-\infty}^{\infty} g(w(\eta))\tilde{V}_c(y-\eta)d\eta, \quad -\infty < y < \infty, \tag{2.1}$$

where

$$\tilde{V}_c(y) := \int_0^\infty \tilde{V}(y+c\tau)A(\tau)d\tau, \tag{2.2}$$

$$\tilde{V}(y) := \int_{\mathbb{R}^{n-1}} V(y,x_2,\ldots,x_n)dx_2\ldots dx_n. \tag{2.3}$$

The equation does not depend on ν since V is a radial function and for the same reason we can restrict our attention to $c \geq 0$. For any c, equation

(2.1) has the constant solutions $w(y) = 0'$ and $w(y) = p$, but the point is whether there exist nontrivial solutions with $0 < w(y) < p$.

In our analysis an important role is played by the characteristic equation

$$(2.4) \qquad L_c(\lambda) = 1,$$

where

$$(2.5) \qquad L_c(\lambda) := g'(0) \int_{-\infty}^{\infty} \tilde{V}_c(y) e^{\lambda y} dy$$

or, equivalently,

$$(2.6) \qquad L_c(\lambda) = g'(0) \int_0^{\infty} e^{-\lambda c \tau} A(\tau) d\tau \int_{\mathbb{R}^n} V(x) e^{\lambda x_1} dx.$$

This characteristic equation arises by linearization of (2.1) followed by substitution of an exponential function.

For any $c \geq 0$, the function L_c is defined at least on the set

$$(2.7) \qquad S := \{\lambda \geq 0 \mid \int_{\mathbb{R}^n} V(x) e^{\lambda x_1} dx < \infty\}.$$

L_c is a convex function of λ (see (2.5) and notice that $\tilde{V}_c$ is nonnegative). For $c = 0$, L_c achieves its infimum for $\lambda = 0$ and consequently $L_0(\lambda) \geq L_0(0) = g'(0) > 1$. By continuity this implies that $L_c(\lambda) > 1$ for $\lambda \in S$ and c sufficiently small. From (2.6) it follows that for fixed $\lambda > 0$, $L_c(\lambda)$ is a monotone decreasing function of c which tends to zero as $c \to \infty$. So the definition

$$(2.8) \qquad c_0 := \inf\{c > 0 \mid L_c(\lambda) = 1 \quad \text{for some } \lambda \in S\}$$

makes sense and $0 < c_0 < \infty$. Moreover, if $c > c_0$ then the set $\{\lambda \in S \mid L_c(\lambda) < 1\}$ is nonempty.

It has been shown that (2.1) has a (unique modulo translation) nondecreasing solution w with $w(-\infty) = 0$ and $w(\infty) = p$ if $c > c_0$ (or even $c=c_0$),

whereas no nontrivial solution exists if $0 \le c < c_0$ (see [4] or [10] for the existence proof and [5] for the uniqueness and the non-existence results; we point out that the proofs require some extra technical assumptions concerning A, V and g, so one is advised to consult these references for a precise formulation of the results).

3. c_0 IS THE ASYMPTOTIC SPEED OF PROPAGATION

The assertion that c_0 is the asymptotic speed of propagation for equation (1.1) naturally splits into two parts. Our first theorem deals with the part that admits a straightforward proof.

THEOREM 1. *Suppose that*

(i) $\sup\{f(t,x) \mid t \in \mathbb{R}_+, x \in \mathbb{R}^n\} = C < \infty$;

(ii) *there exists* $R > 0$ *such that* $\operatorname{supp} f(t,.) \subset B_R$ *for all* $t \in \mathbb{R}_+$;

(iii) $g(x) \le g'(0)x$ *for all* $x \ge 0$.

Then for any $c > c_0$: $\lim_{t\to\infty} \sup\{u(t,x) \mid |x| \ge ct\} = 0$.

PROOF. If we write (1.1) symbolically as $u = Qu + f$, then u is the limit of the nondecreasing sequence u_n defined by $u_0 = f$, $u_{n+1} = Qu_n + f$ (see [4]). Let $c_1 > c_0$ be arbitrary and choose $c_2 \in (c_0,c_1)$. Because of the assumption (iii) we know that

$$u_1(t,x)e^{\lambda(x_1-c_2t)} \le$$

$$\le g'(0)\int_0^t A(\tau)e^{-\lambda c_2\tau}\int_{\mathbb{R}^n} f(t-\tau,\xi)e^{\lambda(\xi_1-c_2(t-\tau))} V(x-\xi)e^{\lambda(x_1-\xi_1)}d\xi d\tau$$

$$+ f(t,x)e^{\lambda(x_1-c_2t)} \le Ce^{\lambda R}(1+L_{c_2}(\lambda)),$$

and by induction we find

$$u_n(t,x)e^{\lambda(x_1-c_2t)} \le Ce^{\lambda R}(1+L_{c_2}(\lambda)+\ldots+(L_{c_2}(\lambda))^n). \tag{3.1}$$

Since $c_2 > c_0$ we can choose $\lambda > 0$ such that $L_{c_2}(\lambda) < 1$. For this choice

of λ the right-hand side of (3.1) is bounded from above uniformly in n and we obtain

$$u(t,x) \le \frac{Ce^{\lambda R}}{1-L_{c_2}(\lambda)} \; e^{\lambda(c_2 t - x_1)} .$$

If ν is any unit vector, we can rotate the coordinate axes in such a way that with respect to the new basis $\nu = (1,0,\ldots,0)$. Since V is a radial function the estimates given above are not affected by this rotation, and we conclude that we can replace x_1 by $x.\nu$ for any unit vector ν. The choice $\nu = x|x|^{-1}$ leads to

$$u(t,x) \le \frac{Ce^{\lambda R}}{1-L_{c_2}(\lambda)} \; e^{\lambda(c_2 t - |x|)} .$$

Consequently

$$\sup\{u(t,x) \mid |x| \ge c_1 t\} \le \frac{Ce^{\lambda R}}{1-L_{c_2}(\lambda)} \; e^{\lambda(c_2 - c_1)t}$$

and since $\lambda > 0$ and $c_1 > c_2$ this proves the theorem. □

<u>REMARK</u>. Actually the result of Theorem 1 is true under less restrictive conditions on f. For instance, suppose that there exists $\lambda_0 > 0$ such that $L_{c_0}(\lambda_0) = 1$, then $L_c(\lambda_0) < 1$ for all $c > c_0$, and one can verify that the proof of Theorem 1 still works if we assume that

$$\sup\{f(t,x)e^{\lambda_0(|x|-c_0 t)} \mid t \in \mathbb{R}_+ , x \in \mathbb{R}^n\} < \infty.$$

The second and last theorem establishes that the solution of (1.1) approaches or passes over p on the set $|x| \le ct$ where c is any number between 0 and c_0.

<u>THEOREM 2</u>. *For any* $c \in (0,c_0)$: $\liminf_{t\to\infty} \min\{u(t,x) \mid |x| \le ct\} \ge p$.

4. PROVING THEOREM 2 BIT BY BIT

Since the proof of Theorem 2 is rather involved we shall split it into several steps which are formulated as lemmas. The proof is based upon a comparison principle and the construction of a suitable subsolution. In the construction of this subsolution we mimic ARONSON and WEINBERGER ([1], [2], [3] and [4]) and some of our proofs are merely adaptations of their analysis to the present context.

For any $T > 0$ we define a mapping E_T by

$$(4.1) \qquad E_T[\phi](t,x) := \int_0^T A(\tau) \int_{\mathbb{R}^n} g(\phi(t-\tau,\xi))V(x-\xi)d\xi \, d\tau.$$

So E_T maps a function defined on $\mathbb{R}_+ \times \mathbb{R}^n$ onto a function defined on $[T,\infty) \times \mathbb{R}^n$. Roughly speaking E_T is a sort of time-T map for the dynamical system defined by (1.1).

<u>LEMMA 1</u>. (*Comparison Principle*).
Suppose that

$$E_T[\psi](t,.) \succ \psi(t,.) \qquad \text{for all } t \geq T,$$

where ψ: $\mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}$ *is a nonnegative continuous function such that*

(i) *for any* $t_1 > 0$ *there exists* $R = R(t_1) < \infty$ *such that for any* $t \in [0,t_1]$, $\operatorname{supp} \psi(t,.) \subset B_R$;

(ii) *if* $\{(t_n,x_n)\}_{n=1}^{\infty} \subset \mathbb{R}_+ \times \mathbb{R}^n$ *is a sequence for which* $x_n \in \operatorname{supp} \psi(t_n,.)$ *and* $\lim_{n\to\infty} (t_n,x_n) = (\bar{t},\bar{x})$, *then necessarily* $\bar{x} \in \operatorname{supp} \psi(\bar{t},.)$.

If there exists $t_0 \geq 0$ *such that*

$$u(t_0+t,.) \succ \psi(t,.) \qquad \text{for } 0 \leq t \leq T,$$

then this relation holds for all $t \geq 0$.

<u>PROOF</u>. Let $\bar{t} := \sup\{t \geq T \mid u(t_0+t,.) \succ \psi(t,.)\}$ and suppose that $\bar{t} < \infty$. Since u is nonnegative it follows that there exists a sequence

$\{(t_n,x_n)\}_{n=1}^{\infty} \subset \mathbb{R}_+ \times \mathbb{R}^n$ such that: (a) $x_n \in \operatorname{supp} \psi(t_n,.)$; (b) $u(t_0+t_n,x_n) \le \psi(t_n,x_n)$; (c) $t_n \downarrow \bar{t}$ as $n \to \infty$. From (i) we deduce that the sequence $\{x_n\}_{n=1}^{\infty}$ is contained in a compact subset of $\mathbb{R}^n$ and hence it contains a convergent subsequence. Subsequently (ii) and (b) imply that there exists $\bar{x} \in \operatorname{supp} \psi(\bar{t},.)$ such that $u(t_0+\bar{t},\bar{x}) \le \psi(\bar{t},\bar{x})$.

On the other hand, since $\bar{t} \ge T$ and $t_0 \ge 0$, the definition of $\bar{t}$ also implies that

$$\begin{aligned} u(t_0+\bar{t},\bar{x}) &\ge \int_0^T A(\tau) \int_{\mathbb{R}^n} g(u(t_0+\bar{t}-\tau,\xi))V(\bar{x}-\xi)d\xi d\tau \\ &\ge \int_0^T A(\tau) \int_{\mathbb{R}^n} g(\psi(\bar{t}-\tau,\xi))V(\bar{x}-\xi)d\xi d\tau = E_T[\psi](\bar{t},\bar{x}) \\ &> \psi(\bar{t},\bar{x}). \end{aligned}$$

So we obtain a contradiction and our assumption $\bar{t} < \infty$ must be false. □

Such a function ψ for which for some $T > 0, E_T[\psi](t,.) \succ \psi(t,.)$ for all $t \ge T$ we will call a *subsolution*.

Next our efforts are directed at the construction of a subsolution ψ with the property that $\psi(t,.)$ is bounded away from zero uniformly on the set $|x| \le ct$. This construction is relatively easy if A and V have compact support and if $g(x) \ge g'(0)x$. For the general case we need to cut off the kernels and we use the inequality $g(x) \ge hx$ for $h < g'(0)$ and x sufficiently small. So firstly we have to show that this can be done without changing the characteristic function too much.

Let the function $K_c = K_c(h,T,R,\lambda)$ be defined by

$$K_c(h,T,R,\lambda) := h \int_0^T e^{-\lambda c\tau} A(\tau)d\tau \int_{B_R} V(x)e^{\lambda x_1}dx, \tag{4.2}$$

or, equivalently,

$$K_c(h,T,R,\lambda) = h \int_{-\infty}^{\infty} e^{\lambda x_1} \Phi_c(R,T,x_1)dx_1, \tag{4.3}$$

where

$$(4.5)\qquad \Phi_c(R,T,x_1) := \int_0^T \tilde{V}^R(x_1+c\tau)A(\tau)d\tau,$$

$$(4.6)\qquad \tilde{V}^R(x_1) := \begin{cases} \tilde{V}(x_1) & \text{for } |x_1| \le R, \\ 0 & \text{for } |x_1| > R . \end{cases}$$

Then we have the following result.

LEMMA 2. *Suppose that* $c \in [0,c_0)$. *There exist numbers* $h \in (0,g'(0))$, $T > 0$ *and* $R > 0$ *such that*

$$K_c(h,T,R,\lambda) > 1 \quad \textit{for all } \lambda \in \mathbb{R} .$$

PROOF. Since $K_c(h,T,R,-\lambda) \ge K_c(h,T,R,\lambda)$ for $\lambda \ge 0$, it suffices to prove the inequality for $\lambda \ge 0$. We split the proof into two steps.

STEP 1. We claim that there exist $\lambda_0 > 0$, $h_0 \in (0,g'(0))$, $T_0 > 0$ and $R_0 > 0$ such that $K_c(h,T,R,\lambda) > 1$ for all $\lambda \ge \lambda_0$, $h \ge h_0$, $T \ge T_0$ and $R \ge R_0$. Indeed, choose R_0 and T_0 such that $\int_0^\infty \Phi_c(R_0,T_0,x_1)dx_1 > 0$ (this can be done since $\int_0^\infty \tilde{V}_c(x_1)dx_1 > 0$ and since Φ_c is a monotone nondecreasing function of R and T converging to $\tilde{V}_c$ pointwise), then $\lim_{\lambda\to\infty} K_c(h,T_0,R_0,\lambda) = \infty$ for all $h > 0$. So we can choose h_0 and λ_0 such that $K_c(h_0,T_0,R_0,\lambda) > 1$ for all $\lambda \ge \lambda_0$ and this proves the claim since K_c is a monotone function of h,T and R.

STEP 2. Suppose that the assertion is not true, then there exist sequences $\{h_n\}$, $\{T_n\}$, $\{R_n\}$ and $\{\lambda_n\}$ such that $h_n \uparrow g'(0)$, $T_n \uparrow \infty$, $R_n \uparrow \infty$ and $\lambda_n \ge 0$ and such that $K_c(h_n,T_n,R_n,\lambda_n) \le 1$. Since $\{\lambda_n\} \subset [0,\lambda_0]$ we can choose a subsequence $\{\lambda_{n_k}\}$ which converges to a limit, say $\bar{\lambda}$. By Fatou's Lemma we have

$$L_c(\bar{\lambda}) \le \liminf_{k\to\infty} K_c(h_{n_k},T_{n_k},R_{n_k},\lambda_{n_k}) \le 1,$$

which is impossible. □

The function $K_c(h,T,R,.)$ is the (two-sided) Laplace transform of the

function $\Phi_c(R,T,.)$ which is defined on $\mathbb{R}$. Our next step is to give explicit solutions of a linear one-dimensional convolution-inequality if it is known that the Laplace transform of the kernel is bounded from below by 1. As candidates for solutions we take the members of the two-parameter family of functions

$$(4.7) \qquad q(y;\alpha,\beta) := \begin{cases} e^{-\alpha y}\sin\beta y & \text{for } 0 \le y \le \frac{\pi}{\beta} \\ 0 & \text{for } y \in \mathbb{R}\setminus[0,\frac{\pi}{\beta}]. \end{cases}$$

LEMMA 3. (ARONSON and WEINBERGER [1],[10]).
Let $k \in L_1(\mathbb{R})$ *be a nonnegative function with compact support such that*

$$L(\lambda) := \int_{-\infty}^{\infty} e^{\lambda y}k(y)dy > 1 \quad \textit{for all } \lambda \in \mathbb{R}.$$

Then there exist a positive number β_0, *a continuous function* $\tilde{\alpha} = \tilde{\alpha}(\beta)$ *and a positive function* $\Delta = \Delta(\beta)$ *defined on* $[0,\beta_0]$ *such that for any* $\beta \in [0,\beta_0]$ *and for any* $\delta\in[0,\Delta(\beta))$

$$\phi * k \succ \phi_\delta,$$

where $\phi(y) := q(y;\tilde{\alpha}(\beta),\beta)$ *and* $\phi_\delta(y) := \phi(y-\delta)$.

PROOF. We split the proof into five steps.

STEP 1. Since k is nonnegative we know that L is convex and that $L(\lambda) \to \infty$ as $|\lambda| \to \infty$. So L achieves its infimum, say for $\lambda = \mu$. Then

$$\int_{-\infty}^{\infty} ye^{\mu y}k(y)dy = 0.$$

STEP 2. We define a function $H = H(\alpha,\beta)$ by

$$H(\alpha,\beta) := \beta^{-1}\int_{-\infty}^{\infty} e^{\alpha y}\sin\beta y\, k(y)dy \quad \text{for } \beta \neq 0,$$

$$H(\alpha,0) := \lim_{\beta\to 0} H(\alpha,\beta) = \int_{-\infty}^{\infty} ye^{\alpha y}k(y)dy.$$

Then $H(\mu,0) = 0$ and

$$\frac{\partial H}{\partial \alpha}(\mu,0) = \int_{-\infty}^{\infty} y^2 e^{\mu y} k(y)dy > 0,$$

so the implicit function theorem implies that there exists $\beta_1 > 0$ and a continuous function $\tilde{\alpha} = \tilde{\alpha}(\beta)$ with $\tilde{\alpha}(0) = \mu$ such that for $0 \le \beta \le \beta_1$, $H(\tilde{\alpha}(\beta),\beta) = 0$. Hence for $0 < \beta \le \beta_1$

$$\int_{-\infty}^{\infty} e^{\tilde{\alpha}(\beta)y} \sin\beta y \; k(y)dy = 0.$$

<u>STEP 3.</u> Since $\int_{-\infty}^{\infty} e^{\alpha y}\cos\beta y \; k(y)dy > 1$ for $\alpha = \mu$ and $\beta = 0$ there exists $\beta_2 > 0$ such that this inequality holds for $\alpha = \tilde{\alpha}(\beta)$ and $0 \le \beta \le \beta_2$.

<u>STEP 4.</u>

$$\begin{aligned}
\phi * k(y) &= \int_0^{\pi/\beta} e^{-\tilde{\alpha}(\beta)\eta} \sin\beta\eta \; k(y-\eta)d\eta \\
&\overset{(1)}{\ge} \int_{-\infty}^{\infty} e^{-\tilde{\alpha}(\beta)\eta} \sin\beta y \; k(y-\eta)d\eta \\
&= e^{-\tilde{\alpha}(\beta)y} \sin\beta y \int_{-\infty}^{\infty} e^{\tilde{\alpha}(\beta)\eta} \cos\beta\eta \; k(\eta)d\eta \\
&\quad - e^{-\tilde{\alpha}(\beta)y} \cos\beta y \int_{-\infty}^{\infty} e^{\tilde{\alpha}(\beta)\eta} \sin\beta\eta \; k(\eta)d\eta \\
&\overset{(2)}{\ge} e^{-\tilde{\alpha}(\beta)y} \sin\beta y.
\end{aligned}$$

(1): We restrict our attention to $y \in [0, \frac{\pi}{\beta}]$ (since anyhow $\phi * k(y) \ge 0$). In order to make this inequality valid we have to take care that for $\eta \in \mathbb{R} \setminus [0, \frac{\pi}{\beta}]$ either, $\sin\beta\eta \le 0$ or $k(y-\eta) = 0$. Let $B > 0$ be such that $\operatorname{supp} k \subset [-B,B]$. If $y \in [0, \frac{\pi}{\beta}]$ and $|y-\eta| \le B$ then $\eta \in [-B, \frac{\pi}{\beta} + B]$ and this interval is contained in $[-\frac{\pi}{\beta}, \frac{2\pi}{\beta}]$ provided $\beta \le \frac{\pi}{B}$.

If either $y = 0$ or $y = \frac{\pi}{\beta}$, then the part that is added to the integral yields a strictly negative contribution (note that both

$\int_0^B k(y)dy > 0$ and $\int_{-B}^0 k(y)dy > 0$) and hence the inequality is strict for those values of y.

(2): Since $\int_{-\infty}^{\infty} e^{\tilde{\alpha}(\beta)\eta} \cos\beta\eta\, k(\eta)d\eta > 1$ for $\beta \le \beta_2$, this inequality is strict for $y \in (0, \frac{\pi}{\beta})$.

STEP 5. At this point we know that $\phi * k \succ \phi$ for $\beta \in [0,\beta_0]$ where $\beta_0 = \min\{\beta_1, \beta_2, \frac{\pi}{B}\}$. From continuity considerations it follows that also $\phi * k \succ \phi_\delta$ if δ is sufficiently small, notably if $\delta \in [0,\Delta(\beta))$, where

$$\Delta(\beta) := \inf\{\sup\{\varepsilon > 0 \,|\, \phi * k(y) > \phi(y-\varepsilon)\} \,|\, 0 \le y \le \tfrac{\pi}{\beta}\}. \quad \square$$

Starting from q we form a three-parameter family of nonincreasing functions r as follows,

$$(4.8) \qquad r(y;\alpha,\beta,\gamma) := \max_{\eta \ge -\gamma} q(y+\eta;\alpha,\beta)$$

or, equivalently,

$$(4.9) \qquad r(y;\alpha,\beta,\gamma) = \begin{cases} M & \text{for } y \le \gamma + \rho \\ q(y-\gamma;\alpha,\beta) & \text{for } \gamma + \rho \le y \le \gamma + \frac{\pi}{\beta} \\ 0 & \text{for } y \ge \gamma + \frac{\pi}{\beta} \end{cases}$$

where $M = M(\alpha,\beta) := \max\{q(y;\alpha,\beta) \mid 0 \le y \le \frac{\pi}{\beta}\}$ and $\rho = \rho(\alpha,\beta)$ is the value for which the maximum is achieved. The following lemma completes the construction of subsolutions with the desired property.

LEMMA 4. (ARONSON and WEINBERGER [3],[10]).
Let $c \in (0,c_0)$ *be given. There exist numbers* $T > 0$, $\beta > 0$, $\alpha \in \mathbb{R}$, $D > 0$ *and* $\sigma_0 > 0$ *such that for any* $\sigma \in (0,\sigma_0)$ *and for any* $t \ge T$

$$E_T[\sigma\psi](t,.) \succ \sigma\psi(t,.),$$

where $\psi(t,x) := r(|x|;\alpha,\beta,D+ct)$.

PROOF. Choose $h \in (0,g'(0))$, $T > 0$ and $R > 0$ such that $K_c(h,T,R,\lambda) > 1$

for all $\lambda \in \mathbb{R}$ (see Lemma 2). According to Lemma 3 we can choose $\beta > 0$, $\alpha = \tilde{\alpha}(\beta)$ and $\delta \in (0,\Delta(\beta))$ such that for $\delta \le x_1 \le \frac{\pi}{\beta} + \delta$

$$(4.10) \qquad h \int_{-\infty}^{\infty} q(x_1-\xi_1)\Phi_c(R,T,\xi_1)d\xi_1 > q(x_1-\delta).$$

(Here and in the following we suppress the dependence on α and β in the notation.) Let σ_h be the smallest positive root of the equation $g(y) = hy$. Then $g(y) > hy$ for $0 < y < \sigma_h$. Choose $\sigma_0 \in (0,\sigma_h M^{-1})$, where M is the maximum of q. With the exception of D we now have chosen all the parameters and it remains to verify the conclusion of the lemma. Let $\sigma \in (0,\sigma_0)$ and $t \ge T$.

$$E_T[\sigma\psi](t,x) = \int_0^T A(\tau) \int_{\mathbb{R}^n} g(\sigma\psi(t-\tau,x-\xi))V(\xi)d\xi d\tau$$

$$\ge \int_0^T A(\tau) \int_{B_R} g(\sigma\psi(t-\tau,x-\xi))V(\xi)d\xi d\tau.$$

We distinguish two cases.

(i) $|x| \le D + \rho + c(t-T) - R$.

If $|\xi| \le R$ and $\tau \in [0,T]$ then

$$|x-\xi| \le D + \rho + c(t-T) \le D + \rho + c(t-\tau)$$

and consequently

$$E_T[\sigma\psi](t,x) = g(\sigma M) \int_0^T A(\tau)d\tau \int_{B_R} V(\xi)d\xi > \sigma M K_c(h,T,R,0) > \sigma M.$$

(ii) $D + \rho + c(t-T) - R \le |x| \le \frac{\pi}{\beta} + D + ct$.

If $|\xi| \le R$ and $t \ge T$ then

$$\begin{aligned} |x-\xi| &= (|x|^2-2x.\xi+|\xi|^2)^{\frac{1}{2}} \\ &\le |x| - \frac{x.\xi}{|x|} + \frac{|\xi|^2}{2|x|} \\ &\le |x| - \frac{x.\xi}{|x|} + \frac{R^2}{2(D+\rho-R)} \\ &\le |x| - \frac{x.\xi}{|x|} + \delta \end{aligned}$$

if we choose $D \geq \frac{R^2}{2\delta} - \rho + R$.

Since ψ is a nonincreasing function of $|x|$ this implies

$$E_T[\sigma\psi](t,x) \geq \sigma h \int_0^T A(\tau) \int_{B_R} \max_{\eta \geq -D-c(t-\tau)} q\left(|x| - \frac{x.\xi}{|x|} + \delta + \eta\right) V(\xi)\, d\xi d\tau$$

$$= \sigma h \max_{\eta \geq -D-ct} \int_{-\infty}^{\infty} q(|x|+\delta+\eta-\xi_1)\Phi_c(R,T,\xi_1)\, d\xi_1$$

$$> \sigma \max_{\eta \geq -D-ct} q(|x|+\eta) = \sigma\psi(t,x).$$

Here we have used the fact that V is a radial function and the inequality (4.10).

Finally, combination of (i) and (ii) yields that

$$E_T[\sigma\psi](t,.) \succ \sigma\psi(t,.). \quad \square$$

Now that we have a comparison principle and suitable subsolutions $\sigma\psi$, it remains to show that the solution of (1.1) is bounded from below by $\sigma\psi$ on a sufficiently large time interval if σ is chosen sufficiently small. In this connection the next result is useful.

<u>LEMMA 5</u>. *For any* $T > 0$ *and any* $R > 0$ *there exists* $t_0 = t_0(T,R)$ *such that* $u(t,x) > 0$ *for* $(t,x) \in [t_0, t_0+T] \times B_R$.

<u>PROOF</u>. If we define the function w by $w(t,x) := \min\{u(t,x),p\}$ then clearly

$$u(t,x) \geq \int_0^t A(\tau) \int_{\mathbb{R}^n} w(t-\tau,\xi)V(x-\xi)\, d\xi d\tau.$$

From $w(t,x) \leq p$, $\int_0^\infty A(\tau)d\tau = 1$ and $\int_{\mathbb{R}^n} V(x)dx = 1$ we deduce that also

$$p \geq \int_0^t A(\tau) \int_{\mathbb{R}^n} w(t-\tau,\xi)V(x-\xi)\, d\xi d\tau.$$

Hence

$$w(t,x) \geq \int_0^t A(\tau) \int_{\mathbb{R}^n} w(t-\tau,\xi)V(x-\xi)\, d\xi d\tau$$

and by iteration

$$(4.11)\qquad w(t,x) \geq \int_0^t A^{m*}(\tau) \int_{\mathbb{R}^n} w(t-\tau,\xi)V^{m*}(x-\xi)d\xi d\tau$$

for all $m \geq 1$, where A^{m*} and V^{m*} denote the $(m-1)$-times iterated convolutions of, respectively, A and V with itself.
One can show that V^{m*} is positive on a ball of radius R_m and that A^{m*} is positive on an interval $[a_m,b_m]$ where both $R_m \to \infty$ and $b_m - a_m \to \infty$ as $m \to \infty$ (see [5, Lemma 2.1]; these properties are easily verified if V is positive on some ball and if A is positive on some interval).

Let $\bar{t}$ be such that $f(\bar{t},\bar{x}) > 0$ for some $\bar{x} \in \mathbb{R}^n$, then $u(\bar{t},\bar{x}) \geq f(\bar{t},\bar{x}) > 0$ and hence also $w(\bar{t},\bar{x}) > 0$. So (4.11) shows that $u(t,x) \geq w(t,x) > 0$ for $t \in [\bar{t}+a_m,\bar{t}+b_m]$ and x in a ball of radius R_m centered at $\bar{x}$. Now for any $T > 0$ and any $R > 0$ this set contains the set $[\bar{t}+a_m,\bar{t}+a_m+T] \times B_R$ if m is sufficiently large and from this observation the conclusion of the lemma follows. □

REMARK. Actually one can show that there is a finite speed of propagation of the boundary of supp u(t,.) if the support of A is bounded away from zero and if V has bounded support. This has been pointed out recently by H. Thieme (Asymptotic speed for the spread of populations, preprint).

Although our subsolutions are bounded away from zero on the set $|x| \leq ct$, they do not grow to p on such a set. The idea is now to use estimates on $|x| \leq ct$ and more detailed information about g to get better estimates on a smaller set. Our last lemma is intended to show that we can come arbitrarily close to p in this manner.

LEMMA 6. *Let a sequence* $\{N_n\}$ *of real numbers be defined by* $N_0 = a > 0$,

$$N_{n+1} = \left(\int_0^t A(\tau)d\tau \int_{B_R} V(x)dx\right)g(N_n).$$

For any $\varepsilon > 0$ *there exist positive numbers* $\bar{t}(\varepsilon)$, $\bar{R}(\varepsilon)$ *and* $\bar{n}(\varepsilon)$ *such that for any* $t \geq \bar{t}(\varepsilon)$, $R \geq \bar{R}(\varepsilon)$ *and* $n \geq \bar{n}(\varepsilon)$

$$N_n > p - \varepsilon.$$

PROOF. Let $\varepsilon > 0$ be arbitrary. Since $g(x) > x$ for $0 < x < p$ and $g'(0) > 1$, we know that $\sup\{x^{-1}g(x) \mid 0 < x \le p - \varepsilon\} > 1$. Hence we can choose $\alpha(\varepsilon) < 1$ such that $\alpha(\varepsilon)g(x) > x$ for $0 < x \le p - \varepsilon$. Let the sequence $\{M_n\}$ be defined by $M_0 = a$, $M_{n+1} = \alpha(\varepsilon)g(M_n)$. We observe that

(i) if $0 < M_n \le p - \varepsilon$ then $M_{n+1} = \alpha(\varepsilon)g(M_n) > M_n$,

(ii) if $M_n > p - \varepsilon$ then $M_{n+1} \ge \alpha(\varepsilon)g(p-\varepsilon) > p - \varepsilon$.

Suppose that $M_n \le p - \varepsilon$ for all n, then (i) shows that M_n converges to a limit $M \le p - \varepsilon$. But then necessarily $M = \alpha(\varepsilon)g(M)$ which is impossible for $M \le p - \varepsilon$. So there exists $\bar{n}(\varepsilon)$ such that $M_{\bar{n}(\varepsilon)} > p - \varepsilon$ and subsequently (ii) implies that $M_n > p - \varepsilon$ for all $n \ge \bar{n}(\varepsilon)$.

Choose $\bar{t}(\varepsilon)$ and $\bar{R}(\varepsilon)$ such that

$$\int_0^{\bar{t}(\varepsilon)} A(\tau)d\tau \int_{B_{\bar{R}(\varepsilon)}} V(x)dx \ge \alpha(\varepsilon).$$

For any $t \ge \bar{t}(\varepsilon)$ and $R \ge \bar{R}(\varepsilon)$ we have $N_1 \ge \alpha(\varepsilon)g(a) = M_1$ and by induction $N_n \ge M_n$. Hence $N_n \ge p - \varepsilon$ for $n \ge \bar{n}(\varepsilon)$. □

Finally, we gather together the pieces in order to give the

PROOF OF THEOREM 2. Let $c_1 \in (0,c_0)$ be arbitrary and choose $c_2 \in (c_1,c_0)$. Let $T > 0$, $\beta > 0$, $\alpha \in \mathbb{R}$, $D > 0$ and $\sigma_0 > 0$ be such that for any $\sigma \in (0,\sigma_0)$ and for any $t \ge T$

$$E_T[\sigma\psi](t,.) \succ \sigma\psi(t,.),$$

where $\psi(t,x) := r(|x|;\alpha,\beta,D+c_2t)$, (see Lemma 4). Let t_0 be such that $u(t,x) > 0$ for $t \in [t_0,t_0+T]$ and $|x| \le D + c_2T + \frac{\pi}{\beta}$ (cf. Lemma 5). Then we can choose $\sigma_1 \in (0,\sigma_0)$ such that $u(t_0+t,.) \succ \sigma_1\psi(t,.)$ for $0 \le t \le T$, and we infer from the comparison principle Lemma 1 that this relation holds for all $t \ge 0$ (note that ψ has all the required properties). Hence $u(t_0+t,x) \ge \sigma_1 M$ for $t \ge 0$, $|x| \le \rho + D + c_2t$.

Next we use the inequality

$$u(t_0+t,x) \ge \int_0^t A(\tau) \int_{B_R} g(u(t_0+t-\tau,x-\xi))V(\xi)d\xi d\tau$$

to conclude by induction that

$$u(t_0+t,x) \geq N_n \quad \text{for} \quad t \geq 0,\ |x| \leq \rho + D + c_2 t - nR,$$

where N_n is defined as in Lemma 6 with $N_0 = a = \sigma_1 M$. So for any $\varepsilon > 0$ we can find $\bar{t}(\varepsilon)$, $\bar{R}(\varepsilon)$ and $\bar{n}(\varepsilon)$ such that

$$u(t,x) \geq p - \varepsilon \text{ for } t \geq t_0 + \bar{t}(\varepsilon),\ |x| \leq \rho + D - c_2 t_0 - \bar{n}(\varepsilon)\bar{R}(\varepsilon) + c_2 t.$$

Finally, since $c_2 > c_1$ this implies that

$$u(t,x) \geq p - \varepsilon \text{ for } |x| \leq c_1 t$$

$$\text{if } t \geq \max\left\{t_0+\bar{t}(\varepsilon),\ \frac{\bar{n}(\varepsilon)\bar{R}(\varepsilon)+c_2 t_0-\rho-D}{c_2-c_1}\right\}. \quad \square$$

5. THE EPIDEMIC MODEL OF KERMACK, McKENDRICK AND KENDALL

In our previous paper [4] we have formulated a model for the geographical spread of a contagious disease. This model is based on a combination of ideas of KERMACK & McKENDRICK [9] and KENDALL [7], [8] and the main assumptions are:

(i) the members of the population can be categorized as either susceptible to or infected by the disease;

(ii) the infectivity of an infected individual as a function of time elapsed since exposure and position relative to the individual's own position is given by $H: \mathbb{R}_+ \to \mathbb{R}$ and $W: \mathbb{R}^n \to \mathbb{R}$, respectively;

(iii) the disease induces permanent immunity, so an individual can pass from the class of susceptibles to the class of infectives, but not vice versa.

If S denotes the density of the susceptibles and u is defined by

$$u(t,x) := -\ln \frac{S(t,x)}{S(0,x)}, \tag{5.1}$$

then the model leads to the following equation for u

$$(5.2)\qquad u(t,x) = \int_0^t H(\tau) \int_{\mathbb{R}^n} \tilde{g}(u(t-\tau,\xi))S(0,\xi)W(x-\xi)d\xi d\tau + f(t,x),$$

where

$$(5.3)\qquad \tilde{g}(y) = 1 - e^{-y},$$

and f is a given nonnegative function describing the history up to $t = 0$ (see [4] for the details). If $S(0,\xi)$ is constant, say S_0, then clearly (5.2) is of the form (1.1) with $g(y) := \alpha\tilde{g}(y)$ and

$$(5.4)\qquad \alpha := S_0 \int_0^\infty H(\tau)d\tau \int_{\mathbb{R}^n} W(x)dx.$$

For this function g the hypothesis H_g is satisfied if and only if $\alpha > 1$. As we will show, the parameter α has a threshold value 1, i.e., the qualitative behaviour is very different in the two cases $\alpha < 1$ and $\alpha > 1$.

By using $g(y) \le \alpha y$ in the same way as in the proof of Theorem 1 it follows that u is bounded from above by constant. $\sup\{f(t,x) \mid t \in \mathbb{R}_+, x \in \mathbb{R}^n\}$ if $\alpha < 1$. One can interpret this result as stability of the rest state $u \equiv 0$. Moreover, Theorem 1 holds for the case $\alpha < 1$ with the definition $c_0 = 0$.

In contrast with this, Theorem 2 shows that for $\alpha > 1$, equation (5.2) exhibits the hair-trigger effect, i.e., *every* nontrivial nonnegative perturbation of the rest state $u \equiv 0$ has *everywhere* eventually a large effect. In [4] we proved this result for the case of space dimension $n = 1$ or $n = 2$, by using the structure of the forcing function f in the epidemic model and some known results about convolution inequalities.

Moreover, Theorems 1 and 2 show that $c_0 = c_0(\alpha)$ is the asymptotic speed for equation (5.2) if $\alpha > 1$. In biological terms this means that if an epidemic draws near then an observer moving with a fixed speed c will consider the epidemic as severe if and only if c is less than c_0. Although the model does not describe the moving of individuals, one can, by way of speaking, also say that somebody who tries to escape the epidemic will improve his changes considerably if and only if he runs away with a speed

that exceeds c_0. For a special case of the model it was shown before by ARONSON [1] that c_0 is the asymptotic speed.

Thus far our discussion was limited to the case that $S(0,\xi)$ is assumed to be constant. A close examination of the proofs of the Theorems 1 and 2 reveals that they are based on inequalities, rather than on equalities. If $S(0,\xi)$ is not constant then one can easily obtain results of the type of Theorem 1 by using an upper bound for $S(0,\xi)$ and results of the type of Theorem 2 by using a lower bound. We do not elaborate this idea any further.

REFERENCES

[1] ARONSON, D.G., *The asymptotic speed of propagation of a simple epidemic*, in: Nonlinear Diffusion, W.E. Fitzgibbon & H.F. Walker, eds., Research Notes in Mathematics (Pitman Publishing Co., London, 1977).

[2] ARONSON, D.G. & H.F. WEINBERGER, *Nonlinear diffusion in population genetics, combustion, and nerve pulse propagation*, in: Partial Differential Equations and Related Topics, J.A. Goldstein, ed., Lecture Notes in Math. 446 (Springer, Berlin, 1975) 5-49.

[3] ARONSON, D.G. & H.F. WEINBERGER, *Multidimensional nonlinear diffusion arising in population genetics*, to appear in Advances in Math.

[4] DIEKMANN, O., *Thresholds and travelling waves for the geographical spread of infection*, to appear in J. Math. Biol.

[5] DIEKMANN, O. & H.G. KAPER, *On the bounded solutions of a nonlinear convolution equation*, to appear in Nonlinear Analysis, Theory, Methods & Applications.

[6] FIFE, P.C., *Asymptotic states for equations of reaction and diffusion*, to appear in Bulletin AMS.

[7] KENDALL, D.G., *Discussion of "Measles periodicity and community size" by M.S. Bartlett*, J. Roy. Statist. Soc. A 120 (1957) 64-67.

[8] KENDALL, D.G., *Mathematical models of the spread of infection*, in: Mathematics and Computer Science in Biology and Medicine (Medical Research Council, London, 1965) 213-224.

[9] KERMACK, W.O. & A.G. McKENDRICK, *A contribution to the mathematical theory of epidemics*, Proc. Roy. Soc. A 115 (1927) 700-721.

[10] WEINBERGER, H.F., *Asymptotic behavior of a model in population genetics*, to appear in: Indiana University Seminar in Applied Mathematics, J. Chadam, ed., Springer Lecture Notes in Math.

SAMENVATTING

Dit proefschrift handelt over de afleiding en de kwalitatieve analyse van verscheidene niet-lineaire integraalvergelijkingen die de verspreiding beschrijven van een besmettelijke ziekte in een populatie. Het proefschrift bestaat uit een inleiding en uit vier artikelen ([A] t/m [D]), die in wetenschappelijke tijdschriften verschenen zijn of eventueel nog zullen verschijnen.

Het uitgangspunt is steeds een model dat gebaseerd is op ideeën van Kermack, McKendrick en Kendall. Bij dit model hoort een integraalvergelijking van gemengd Volterra-Fredholm type. Analyse van deze vergelijking brengt aan het licht dat een willekeurig kleine locale infectiebron een overal optredende uitbraak van de ziekte veroorzaakt, indien de populatiedichtheid een bepaalde kritieke waarde te boven gaat (het zogenaamde "hair-trigger effect"; [B] en [D]).

Er wordt voorts aandacht besteed aan lopende golfoplossingen. Deze blijken te moeten voldoen aan een niet-lineaire convolutievergelijking op de hele reële rechte. Bewezen wordt dat een getal c_0 bestaat zó dat er voor iedere c groter dan c_0 een lopende golfoplossing bestaat met snelheid c [B]. In samenwerking met H.G. Kaper wordt een aantal algemene resultaten afgeleid betreffende asymptotisch gedrag, uniciteit modulo translatie en niet-existentie van begrensde oplossingen van convolutievergelijkingen [C]. Toepassing van deze resultaten op het epidemiemodel leidt tot de conclusie dat er geen lopende golfoplossingen bestaan met snelheid c kleiner dan c_0 en dat er voor c groter dan c_0 modulo translatie precies één zo'n oplossing bestaat. Uit een analyse van de Volterra-Fredholm vergelijking blijkt vervolgens dat de ziekte zich asymptotisch verspreidt met de minimale golfsnelheid c_0 [D].

Voor het speciale geval dat plaatsafhankelijkheid veronachtzaamd wordt, wordt het effect van de introductie van slechts heel weinig besmetten in een tot dan toe onbesmette populatie nader onderzocht. Wiskundig bezien komt dit neer op de karakterisering van het gedrag van oplossingen van een Volterra integraalvergelijking in de omgeving van een instabiel evenwichtspunt ([A] en de inleiding).

In de inleiding wordt een overzicht gegeven van de resultaten, er wordt aandacht besteed aan de manier waarop men met Volterra integraalvergelijkingen een dynamisch systeem kan associëren en tot slot wordt de theoretisch biologische motivatie voor dit soort onderzoek gekenschetst.

nse 45 G 99,

92 A 15

45 D 05

45 B 05

45 M 05

45 M 99

SOK

ONTVANGEN 22 AUG. 1978

STELLINGEN

BIJ HET PROEFSCHRIFT

OVER NIET-LINEAIRE INTEGRAALVERGELIJKINGEN EN MATHEMATISCHE EPIDEMIOLOGIE

VAN

ODO DIEKMANN

18 OKTOBER 1978

I

Zij de functie $z = z(w)$ gedefinieerd door de vergelijking

$$z - \ln(1+z) = \frac{1}{2} w^2$$

en door de conditie dat $z(0) = 0$ en $z(w) > 0$ voor $w > 0$.
Dan geldt voor de coëfficiënten d_k in de machtreeksontwikkeling

$$z(w) = \sum_{k=1}^{\infty} d_k w^k$$

de asymptotische benadering

$$d_k = \left(\frac{-1}{2\sqrt{\pi}}\right)^k \frac{2\sqrt{2}}{k\sqrt{k}} \left\{\alpha_k + \left(\frac{3}{4}\alpha_k + \frac{\pi}{3}\beta_k\right)\frac{1}{k} + O\left(\frac{1}{k^2}\right)\right\}, \quad k \to \infty,$$

waarbij

$$\alpha_k := \cos\left(\frac{k+1}{4}\pi\right) \quad \text{en} \quad \beta_k := \cos\left(\frac{k-1}{4}\pi\right)$$

G.N. WATSON, *Theorems stated by Ramanujan* (V): *Approximations connected with* e^x, Proc. Lond. Math. Soc.(2) 29 (1929) 293-308.

O. DIEKMANN, *Asymptotic expansion of certain numbers related to the gamma function*, MC Rapport TN 80, 1975.

II

In de oplossing van sommige lineaire elliptische singuliere storingsproblemen treedt bij variatie van de coëfficiënten van de vergelijking een overgang op van gewone - via parabolische - naar vrije grenslaag. Door de methode van strekking van variabelen ook toe te passen op de coëfficiënten kan men aantonen dat deze overgang soepel verloopt. Bovendien worden zodoende verborgen grenslagen aan het licht gebracht.

O. DIEKMANN, *Some aspects of non-uniform convergence in an elliptic singular perturbation problem*, J. Eng. Math. 9 (1975) 227-233.

III

Zij $y(x;\varepsilon)$ de oplossing van het tweepuntsrandwaardeprobleem

$$\begin{cases} \varepsilon x y'' + (g(x)-y)y' = 0 \\ y(0) = 0, \quad y(1) = 1 \end{cases}$$

waarbij de gegeven continu differentieerbare functie g voldoet aan $g(0) = 0$ en $g(1) \geq 1$.
Dan geldt uniform voor $x \in [0,1]$ dat

$$\lim_{\varepsilon \downarrow 0} y(x;\varepsilon) = y_0(x)$$

waarbij

$$y_0(x) := \min\{\max\{\min\{g(y) \mid x \leq y \leq 1\}, 0\}, 1\}$$

O. DIEKMANN, D. HILHORST & L.A. PELETIER, *A singular boundary value problem arising in a pre-breakdown gas discharge*, MC Rapport TW 174, 1978.

IV

Het probleem

$$\begin{cases} \frac{1}{r} \frac{d}{dr}\left[r \frac{dV}{dr}\right] = -\frac{1}{r} \int_0^{\min\{r,\frac{1}{2}\}} \frac{s\,ds}{\sqrt{V(s)-V(r)+1}} \\ \frac{dV}{dr}(0) = 0, \quad V(1) = 0 \end{cases}$$

heeft één en slechts één oplossing.

O. DIEKMANN, N.M. TEMME & J.G. VERWER, *Berekening van een ruimteladings-effect*, MC Rapport TN 85, 1976.

V

Het artikel *Qualitative Theory of the FitzHugh-Nagumo Equations* (Adv. in Math. 27 (1978) 12-44) van J. RAUCH en J. SMOLLER, bevat een aantal onjuiste beweringen, onder andere

then $R(\varepsilon_0)$ *is the desired rectangle.*	p.25 r.-8
This R *does the trick.*	p.35 r.-1
For $t \geq T$, U_0 *satisfies the* FN *equations* (4.9), (4.10) *on* $-\infty < x < \infty$.	p.38 r.-5

Desondanks is het artikel de moeite van het lezen meer dan waard.

VI

Zij $f:[0,1] \to [0,1]$ gedefinieerd door

$$f(x) = \begin{cases} x + \frac{1}{2} & \text{voor } 0 \leq x \leq \frac{1}{2} \\ 2 - 2x & \text{voor } \frac{1}{2} \leq x \leq 1 \end{cases} .$$

Voor ieder natuurlijk getal k bestaat er in [0,1] een periodieke oplossing van (kleinste) periode k van de differentievergelijking

$$x_{n+1} = f(x_n) .$$

T-Y LI & J.A. YORKE, *Period three implies chaos*, Amer. Math. Monthly 82 (1975) 985-992.

VII

Het artikel *Catastrophe Theory as Applied to the Social and Biological Sciences: A Critique* (Synthese 37 (1978) 117-216) van H.J. SUSSMANN en R.S. ZAHLER laat op overtuigende wijze zien dat het sprookje van de Keizer zonder kleren niet aan actuele waarde heeft ingeboet.

VIII

De opmerking "*Het lijkt wel alsof een jaar steeds korter wordt*" vindt zijn oorsprong in het feit dat ieder mens de tijd meet met zijn eigen leeftijd als eenheid.

IX

Van alle tot verslaving leidende verdovende middelen kent televisie de meeste gebruikers.