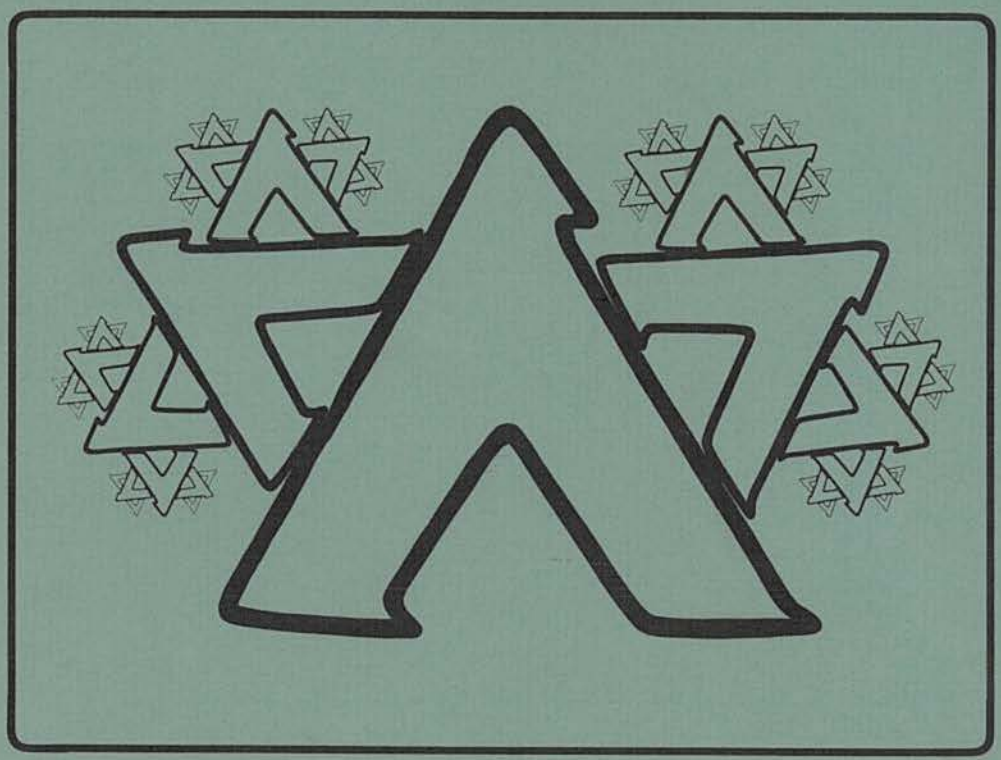
Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

TW
IN BEELD



BIJ HET AFSCHEID VAN
PROF.DR. H.A. LAUWERIER

AMSTERDAM, DECEMBER 1988

Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

The CWI is a research institute of the Stichting Mathematisch Centrum, which was founded on February 11, 1946, as a nonprofit institution aiming at the promotion of mathematics, computer science, and their applications. It is sponsored by the Dutch Government through the Netherlands Organization for the Advancement of Research (NWO).

Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

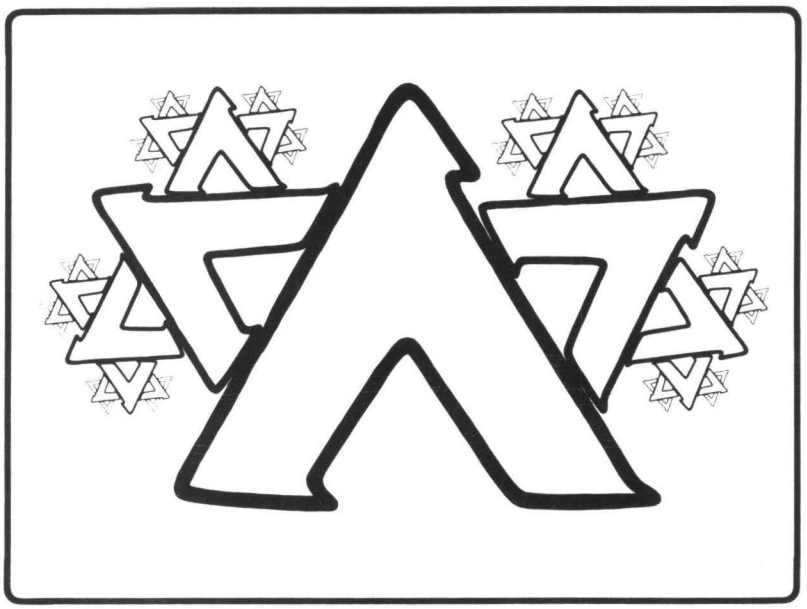
TW in Beeld

bij het afscheid van

Prof.dr. H.A. Lauwerier

onder redactie van

N.M. Temme



ISBN 90 6196 362 1

Hoe dat precies werkt is niet zo eenvoudig te zeggen, daarvoor schiet onze omgangstaal te kort. Maar nu komt de wiskunde te hulp. Wiskunde is zelf een taal, de taal waarin de verschijnselen van de natuur nauwkeurig en ondubbelzinnig beschreven kunnen worden.

HANS LAUWERIER

Symmetrie (1988)

Beste Hans,

Het eerste kontakt dat ik op het Mathematisch Centrum (zo heette het Centrum destijds nog) in de Boerhaavestraat met je had vond plaats, toen ik als student je colleges Asymptotiek en Methoden van de Mathematische Fysica volgde. Het eerste college sprak me erg aan en niet alleen mij. Ook alle toenmalige leden van de afdeling Toegepaste Wiskunde werden, voornamelijk op de eerste rij gezeten, van je rijke ervaringen op de hoogte gesteld. In de pauze trok het grote gezelschap zich dan terug om, zoals ik dacht, diepgaande discussies met je over de geboden leerstof te beginnen. Later merkte ik dat het er tussen de bedrijven door gewoon gezellig aan toe ging.
De colleges worden al lang niet meer op het Centrum gegeven, maar je streven om af en toe eens voor een onderonsje bij elkaar te komen is altijd blijven bestaan, waarbij het aanstekelijke enthousiasme waarmee je over je recente onderzoek of leeservaringen vertelde vaak aanleiding gaf tot een nieuwe en interessante kijk op de wiskunde.

Ongeveer gelijktijdig met je afscheid verdwijnt op het CWI door reorganisatie van diverse afdelingen de naam Toegepaste Wiskunde. De nieuwe afdeling Analyse, Algebra en Meetkunde, die uit de afdelingen Toegepaste Wiskunde en Zuivere Wiskunde is ontstaan, moet nog wennen aan zijn identiteit. Het is dan ook een goede gelegenheid om even (voor één keertje nog?) achterom te kijken om te zien hoe het ook weer was. Maar vooral om te constateren dat er tijdens je vertrek als chef van de afdeling TW een actieve groep onderzoekers bezig was, met een brede scala aan onderzoeksonderwerpen. Deze ter ere van jou geschreven en hierbij aangeboden artikelen geven een beeld van TW ten tijde van je afscheid. In sommige artikelen (over biomathematica en asymptotiek) zijn je eigen onderzoeksgebieden duidelijk terug te vinden, in andere (over tomografie en mathematische morfologie) vind je nieuwe onderwerpen, die je met veel interesse de afdeling hebt zien binnenkomen, maar waarin je zelf niet actief bent geweest. In een komend nummer van de CWI Quarterly, dat geheel aan je afscheid zal zijn gewijd, zullen artikelen uit deze bundel voor een groter publiek beschikbaar komen.

De ongedwongen sfeer, de vrijheid van handelen voor de onderzoekers op je afdeling en je eigen levendige belangstelling en activiteiten hebben een klimaat geschapen waarin wetenschappelijk onderzoek zeer wel mogelijk is geweest. De onderhavige bundel wordt je daarom door ons, medewerkers van je voormalige afdeling TW, met dankbaarheid aangeboden.

Nico Temme
Afdeling AM

CWI, 9 december 1988

INHOUDSOPGAVE

On sums of remainders and almost perfect numbers

H.A.J.M. Schellinx
3e Oosterparkstraat 172-I
1092 EE Amsterdam

J.A.P. Heesterbeek
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam
The Netherlands

This note contains some preliminary results of current research on the sum of remainders $S(n)$ of a natural number n on division by all $m<n$. This function is related with the problems of perfection in numbers. We state a conjecture concerning $S(n)$ which, if true, would imply that the powers of two are the only almost perfect numbers. The main part of this note discusses some elementary properties of $S(n)$ and a related function $P(n):=S(n)-S(n-1)$.

For every problem you can't
solve, there exists an easier
problem you also can't solve.
Find it.
(P. Erdös)
"Nothing works!"
(Catweazle)

1. Introduction

Thanks to Euclid it is known that for any two natural numbers n,m $(m\neq 0)$ there exist unique natural numbers k,r with $n=m\cdot k+r$, $0\leqslant r<m$. Let us write $\mathrm{mod}(n,m)=r$ for the remainder on division by m. We now define the following number theoretic function,

$$S(n) := \sum_{i<n} \mathrm{mod}(n,i) \qquad n\in\mathbb{N}, i\in\mathbb{N}_{>0}.$$

This function was probably first mentioned by Lucas in his Théorie des Nombres (1891) (see also Dickson, 1919, p. 312). His result on $S(n)$, which

appears here as our theorem 1, gives an immediate link of $S(n)$ to perfect (and related) numbers (see also CROSS, 1974).

In this note we will look at some properties of $S(n)$ and the derived function $P(n):=S(n)-S(n-1)$. This is mainly done in section 4, where $S(n)$ and $P(n)$ are decomposed into subsums whose calculation shows strikingly regular patterns. The way these patterns are generated will be derived. We shall only use elementary arithmetical operations. In section 2 we derive a limit relation between $S(n)$ and $d(n)$, the number of divisors of n. In section 3 we discuss a conjecture that has implications for the almost-perfect-number problem.

2. ESTIMATING $S(n)$

For convenience we introduce $\rho(n,i):=i\cdot\lfloor\frac{n}{i}\rfloor$ (with $\lfloor x\rfloor$ being the largest integer $\leqslant x$). It is easy to see that

$$n = \rho(n,i)+\mathrm{mod}(n,i). \tag{1}$$

Summing both sides of (1) over all $i\leqslant n$ gives, after some rearranging,

$$S(n) = n^2-\sum_{i=1}^{n}\rho(n,i). \tag{2}$$

As the last term is positive, we have $S(n)<n^2$ for all n. A lower upper bound is obtained after summing (1) over the non-divisors of n,

$$S(n) = n^2-n\cdot d(n)-\sum_{i\nmid n}\rho(n,i)$$

where $d(n):=\#\{k\leqslant n:k|n\}$. This gives $S(n)\leqslant n^2-n\cdot d(n)$ with equality only for $n\leqslant 2$. As a bound this is not very useful because $d(n)$ is a function even more erratic than $S(n)$. However we can derive a limit-relation between $d(n)$ and $S(n)$. For this we need a few well-known results. The first is by Catalan (DICKSON, 1919)

$$\sigma(k) = \sum_{i=1}^{k}(\rho(k,i)-\rho(k-1,i)) \tag{3}$$

where $\sigma(k):=\Sigma_{i|k}i$. Combining (3) with (2) gives (Cesàro)

$$S(n) = n^2-\sum_{k=1}^{n}\sigma(k). \tag{4}$$

Using a result from HARDY & WRIGHT (1960, theorem 324) we can rewrite this as

$$S(n) = n^2-\tfrac{1}{12}\pi^2n^2+O(n\log n).$$

For large n, this gives an estimate for $S(n)$ in terms of n: $S(n)\approx(1-\frac{1}{12}\pi^2)\cdot n^2$. Since $d(n)/n\to 0$ as $n\to\infty$ (Hardy & Wright, 1960), we have

PROPOSITION 1. *Let* $n \in \mathbb{N}$, *then*

$$\lim_{n\to\infty} \frac{n^2 - n\cdot d(n)}{S(n)} = \frac{1}{1-\frac{1}{12}\pi^2}. \quad \square$$

We will not pursue this matter further here but concentrate on describing $S(n)$.

3. A CONJECTURE ON EQUALITY I.

An obvious question to ask is: Can one characterize pairs (n,m) of natural numbers with $S(n)=S(m)$? Concerning this we state the following

CONJECTURE 1. Let $n,m \in \mathbb{N}$, $n \neq m$, then

$$S(n) = S(m) \Leftrightarrow n = 2^k,\ m = 2^k - 1 \text{ for some } k \geqslant 0.$$

We will show that the implication from right to left holds, in fact we will show something stronger. The proof rests on the following theorem:

THEOREM 1. (Lucas) *Let* $n \in \mathbb{N}_{>0}$, *then*

$$S(n) - S(n-1) = 2n - 1 - \sigma(n).$$

The theorem follows immediately after subtracting $S(n-1)$ from $S(n)$, both written as in equation (4). If we take n to be a prime-power in theorem 1 we get

$$S(p^n) - S(p^n - 1) = \frac{p-2}{p-1}(p^n - 1).$$

The implication "$\Leftarrow$" of conjecture 1 now follows from the observation that for $n > 0$: $S(p^n) = S(p^n - 1) \Leftrightarrow p = 2$.

As stated in the introduction, theorem 1 connects $S(n)$ with almost perfect, perfect and quasiperfect numbers. Recall that a number n is called perfect if $\sigma(n) = 2n$, almost perfect if $\sigma(n) = 2n - 1$, and quasiperfect if $\sigma(n) = 2n + 1$. If we put $P(n) := S(n) - S(n-1)$ then: $P(n) = 0 \Leftrightarrow n$ almost perfect; $P(n) = -1 \Leftrightarrow n$ perfect; $P(n) = -2 \Leftrightarrow n$ quasiperfect.

It is known at least since Euler that all even perfect numbers are of the form $2^{n-1}(2^n - 1)$ with $2^n - 1$ a (Mersenne) prime. It is not known whether there are any odd perfect numbers. There are no known quasiperfect numbers and the only known almost perfect numbers are the powers of two (GUY, 1981).

If our conjecture 1 is true it would follow that the powers of two are indeed the only almost perfect numbers. This relation to the somewhat esoteric problem of 'perfection of numbers' makes the existence of an elementary proof of the left-to-right implication very unlikely. Still, further study of the functions $S(n)$ and $P(n)$ might be helpful, and proves to be rewarding in its own right.

4. DESCRIBING $S(n)$ & $P(n)$.

Taking some fixed $n \in \mathbb{N}$ and calculating $\mathrm{mod}(n, 1), \mathrm{mod}(n, 2), \ldots, \mathrm{mod}(n,n)$, we notice that $S(n)$ may be considered as being composed of subsums, each of which is the contribution to $S(n)$ of a set $\{k, k+1, \ldots, k+l\}$ of numbers $\leqslant n$, with the property that $\exists m \in \mathbb{N}$ such that for all $0 \leqslant i < l$, $\mathrm{mod}(n, k+i) - \mathrm{mod}(n, k+i+1) = m$. In fact, for each $m \in \mathbb{N}$ there is a unique set with this property (which will be almost always empty), and sets belonging to different m's are disjoint. So for each $m \in \mathbb{N}$ we find a subsum, say $\theta_m(n)$ (which is almost always zero), and $S(n) = \Sigma_{m=0}^{\infty} \theta_m(n)$. To make this more precise let us define, for $n, m \in \mathbb{N}$,

$$G_m(n) := \{k \in \mathbb{N} : \frac{m}{m+1} \cdot n \leqslant k < \frac{m+1}{m+2} \cdot n\}. \tag{5}$$

Furthermore let $\|G_m(n)\|$ denote the number of elements in $G_m(n)$ and let $g_m(n) := \Sigma_{i=0}^{m} \|G_i(n)\|$. We refer to table 1 for an example of the sets $G_m(n)$ for $n = 80$.

LEMMA 1.

$$k \in G_m(n) \Leftrightarrow \mathrm{mod}(n, n-k) = k + m \cdot (k-n).$$

PROOF: As $n = n - k + k$ we have $n \equiv k \equiv k + m(k-n)$ $(\mathrm{mod}(n-k))$ for all $m \in \mathbb{Z}$.

Therefore

$$\begin{aligned} & \mathrm{mod}(n, n-k) = k + m \cdot (k-n) \\ \Leftrightarrow\ & 0 \leqslant k + m \cdot (k-n) < n - k \\ \Leftrightarrow\ & \frac{m}{m+1} \cdot n \leqslant k < \frac{m+1}{m+2} \cdot n. \end{aligned}$$ □

If we put $g_{-1}(n) := 0$ for all n, the following lemma is obvious from the definitions given above.

LEMMA 2.

a) $g_m(n) = \lceil \frac{m+1}{m+2} \cdot n \rceil$,

b) $\|G_m(n)\| = g_m(n) - g_{m-1}(n) = \lceil \frac{m+1}{m+2} \cdot n \rceil - \lceil \frac{m}{m+1} \cdot n \rceil$. □

(Here $\lceil x \rceil$ denotes the smallest integer $\geqslant x$.)

Let s be the smallest integer satisfying $\frac{s+1}{s+2} \cdot n > n - 2$ for all $n \in \mathbb{N}_{\geqslant 2}$. A simple calculation shows that $s = \lfloor \frac{n-2}{2} \rfloor$. Then we have partitioned the natural numbers smaller than $n-1$ in $s+1$ pairwise disjoint, possibly empty, sets $G_m(n)$, $0 \leqslant m \leqslant s$. So $G_m(n) = \{x_1, \ldots, x_t\}$, if $G_m(n) \neq \varnothing$, with $x_1, \ldots, x_t$ consecutive natural numbers. $G_m(n) = \varnothing$ for all $m > s$, except for $m = n-1$, in

G,	$n-k$,	$\mathrm{mod}(n,k)$	G,	$n-k$,	$\mathrm{mod}(n,k)$	G,	$n-k$,	$\mathrm{mod}(n,k)$
$G_0(n)$	0	0	$G_2(n)$	54	2	$G_5(n)$	67	2
	1	1		55	5		68	8
	2	2		56	8	$G_6(n)$	69	3
	3	3		57	11	$G_7(n)$	70	0
	·	·		58	14		71	8
	·	·		59	17	$G_9(n)$	72	0
	39	39	$G_3(n)$	60	0	$G_{10}(n)$	73	3
$G_1(n)$	40	0		61	4	$G_{12}(n)$	74	2
	41	2		62	8	$G_{15}(n)$	75	0
	42	4		63	12	$G_{19}(n)$	76	0
	·	·	$G_4(n)$	64	0	$G_{25}(n)$	77	2
	·	·		65	5	$G_{39}(n)$	78	0
	53	26		66	10	$G_{79}(n)$	79	0

TABLE 1. Explicit calculation of $S(n)$ for $n=80$, with contents of the sets $G_m(80)$.

which case $G_m(n)=\{n-1\}$. If we define

$$\theta_m(n) := \sum_{i=1}^{t} \mathrm{mod}(n, n-x_i) \tag{6}$$

then it is clear that $\theta_m(n)$ is a subsum of $S(n)$, as mentioned at the beginning of this section, and

$$S(n) = \sum_{m=0}^{s} \theta_m(n) = \sum_{m=0}^{\infty} \theta_m(n) \tag{7}$$

PROPOSITION 2. *For all* $m,n \in \mathbb{N}$

$$\theta_m(n) = \tfrac{1}{2}\cdot\|G_m(n)\|\cdot\{(m+1)(g_m(n)+g_{m-1}(n)-1)-2mn\}. \tag{8}$$

PROOF: For $n=0$ and $n=1$ this is easily verified by direct computation. For all other n the following counting argument proves (8). It follows from (5) and lemma 2a that $g_{m-1}(n)$ is the smallest element of $G_m(n)$. Then $G_m(n)=\{x_1,\ldots,x_t\}$ is given by

$$G_m(n) = \{g_{m-1}(n),\ g_{m-1}(n)+1,\ \ldots,\ g_{m-1}(n)+\|G_m(n)\|-1\} \tag{9}$$

because of the definition of $\|G_m(n)\|$.
Let us put $\|G_m(n)\|-1=k$ for convenience. Combining (9) with (6) and lemma 1 leads to

$$\begin{aligned}\theta_m(n) &= \sum_{j=0}^{k}(g_{m-1}(n)+j+m\cdot(g_{m-1}(n)+j-n)),\\ &= \tfrac{1}{2}(k+1)((m+1)(k+1+2g_{m-1}(n)-1)-2mn).\end{aligned}$$

Putting k back to its original meaning we arrive at

$$\theta_m(n) = \tfrac{1}{2}\cdot\|G_m(n)\|\cdot\{(m+1)(\|G_m(n)\|+2g_{m-1}(n)-1)-2mn\}$$

Combining this with lemma 2 gives (8). □

If we put $\Delta\theta_m(n) := \theta_m(n)-\theta_m(n-1)$, $n>0$ we get from (7) and the definition of $P(n)$

$$P(n) = \sum_{m=0}^{s} \Delta\theta_m(n) = \sum_{m=0}^{\infty} \Delta\theta_m(n). \tag{10}$$

Calculating $\theta_m(n)$ and $\Delta\theta_m(n)$ brings out a striking regularity as shown in tables 2 and 3. We will now derive these patterns and an expression for $\Delta\theta_m(n)$, proposition 5, from (8). In order to do so, we need a few rules for the manipulation of expressions involving the functions $\lceil\cdot\rceil$ and $\lfloor\cdot\rfloor$, the easy proofs of which we leave to the reader.

$\theta_1(n)$	$\theta_2(n)$	$\theta_3(n)$	$\theta_4(n)$
0 0 0	0 0 0 0	0 0 0 0 0	0 0 0 0 0 0
0 0 1	0 0 0 1	0 0 0 0 1	0 0 0 0 0 1
0 1 2	0 0 1 2	0 0 0 1 2	0 0 0 0 1 2
1 2 4	0 1 2 3	0 0 1 2 3	0 0 0 1 2 3
2 4 6	1 2 3 5	0 1 2 3 4	0 0 1 2 3 4
4 6 9	2 3 5 7	1 2 3 4 6	0 1 2 3 4 5
6 9 12	3 5 7 9	2 3 4 6 8	1 2 3 4 5 7
9 12 16	5 7 9 12	3 4 6 8 10	2 3 4 5 7 9
12 16 20	7 9 12 15	4 6 8 10 12	3 4 5 7 9 11
16 20 25	9 12 15 18	6 8 10 12 15	4 5 7 9 11 13
20 25 30	12 15 18 22	8 10 12 15 18	5 7 9 11 13 15
25 30 36	15 18 22 26	10 12 15 18 21	7 9 11 13 15 18
30 36 42	18 22 26 30	12 15 18 21 24	9 11 13 15 18 21
36 42 49	22 26 30 35	15 18 21 24 28	11 13 15 18 21 24
42 49 56	26 30 35 40	18 21 24 28 32	13 15 18 21 24 27
. . .	. . . .	. . . . .	. . . . . .
. . .	. . . .	. . . . .	. . . . . .

TABLE 2. $\theta_m(n)$ for $m=1$ to 4 with $0\leqslant n<15\cdot m$. The entries are arranged in sets of $m+2$, to be read from left to right.

LEMMA 3.

For all $n,m\in\mathbb{N}$, $q\in\mathbb{Q}_{\geqslant 0}$:

a) $\lceil\frac{n}{m}\rceil = \lfloor\frac{n}{m}\rfloor+\lceil\frac{\mathrm{mod}(n,m)}{m}\rceil$, $m\neq 0$,

b) $\lceil n+q\rceil = n+\lceil q\rceil$; $\lfloor n+q\rfloor = n+\lfloor q\rfloor$,

c) $n-\lceil\frac{mn}{m+1}\rceil = \lfloor\frac{n}{m+1}\rfloor$,

d) $(m+1)\lceil\frac{mn}{m+1}\rceil = mn+\mathrm{mod}(n,m+1)$. □

Let us define

$$\gamma_m(n) := \lfloor\frac{n}{m+2}\rfloor+\mathrm{mod}(n,m+2). \tag{11}$$

Thus $\gamma_m(n)$ is the sum of the number of times n is divisible by $m+2$ plus the remainder of n on division by $m+2$. This function will play a major role in what follows.

$\Delta\theta_1(n)$	$\Delta\theta_2(n)$	$\Delta\theta_3(n)$	$\Delta\theta_4(n)$
0 0 0	0 0 0 0	0 0 0 0 0	0 0 0 0 0 0
0 1 -1	0 0 1 -1	0 0 0 1 -1	0 0 0 0 1 -1
1 1 -1	0 1 1 -2	0 0 1 1 -2	0 0 0 1 1 -2
1 2 -2	1 1 1 -2	0 1 1 1 -3	0 0 1 1 1 -3
2 2 -2	1 1 2 -3	1 1 1 1 -3	0 1 1 1 1 -4
2 3 -3	1 2 2 -4	1 1 1 2 -4	1 1 1 1 1 -4
3 3 -3	2 2 2 -4	1 1 2 2 -5	1 1 1 1 2 -5
3 4 -4	2 2 3 -5	1 2 2 2 -6	1 1 1 2 2 -6
4 4 -4	2 3 3 -6	2 2 2 2 -6	1 1 2 2 2 -7
4 5 -5	3 3 3 -6	2 2 2 3 -7	1 2 2 2 2 -8
5 5 -5	3 3 4 -7	2 2 3 3 -8	2 2 2 2 2 -8
5 6 -6	3 4 4 -8	2 3 3 3 -9	2 2 2 2 3 -9
6 6 -6	4 4 4 -8	3 3 3 3 -9	2 2 2 3 3 -10
6 7 -7	4 4 5 -9	3 3 3 4 -10	2 2 3 3 3 -11
7 7 -7	4 5 5 -10	3 3 4 4 -11	2 3 3 3 3 -12
7 8 -8	5 5 5 -10	3 4 4 4 -12	3 3 3 3 3 -12
. . .	. . . .	. . . . .	. . . . .
. . .	. . . .	. . . . .	. . . . .

TABLE 3. $\Delta\theta_m(n)$ for $m=1$ to 4 with $1\leqslant n\leqslant 15\cdot m$. The entries are arranged in sets of $m+2$ to be read from left to right.

PROPOSITION 3. *For all* $m,n\in\mathbb{N}$:

a) $\|G_m(n)\| = \lfloor\frac{\gamma_m(n)}{m+1}\rfloor$,

b) $(m+1)\{\lceil\frac{m+1}{m+2}\cdot n\rceil+\lceil\frac{m}{m+1}\cdot n\rceil-1\}-2mn =$
$(m+1)\{\lfloor\frac{\gamma_m(n)}{m+1}\rfloor-1\}+2\mathrm{mod}(\gamma_m(n),m+1)$.

PROOF: Choose any $m,n\in\mathbb{N}$, and write $n=k\cdot(m+2)+r$ with $r,k\in\mathbb{N}$, $0\leqslant r<m+2$. Then $k=\lfloor\frac{n}{m+2}\rfloor$, $r=\mathrm{mod}(n,m+2)$ and we have $\lfloor\frac{\gamma_m(n)}{m+1}\rfloor=\lfloor\frac{k+r}{m+1}\rfloor$.

Substituting $k\cdot(m+2)+r$ for n and applying lemma 2b, 3 b&c we obtain:

$$\|G_m(n)\| = (m+1)k+r-\lceil\frac{mk(m+2)+mr}{m+1}\rceil. \tag{12}$$

An application of lemma 3 a&b gives:

$$\lceil\frac{mk(m+2)+mr}{m+1}\rceil = mk+\lceil\frac{m(k+r)}{m+1}\rceil. \tag{13}$$

Combining (12) and (13) and applying lemma 3c we obtain $\|G_m(n)\|=\lfloor\frac{k+r}{m+1}\rfloor$, which completes the proof of a). Statement b) can be proved in a similar way, e.g. using lemma 3d. □

By rewriting (8) using proposition 2 we get

$$\theta_m(n) = \tfrac{1}{2}\lfloor\frac{\gamma_m(n)}{m+1}\rfloor\cdot\{(m+1)(\lfloor\frac{\gamma_m(n)}{m+1}\rfloor-1)+2\mathrm{mod}(\gamma_m(n),m+1)\}.$$

Now define for each $m\in\mathbb{N}$ a mapping $f_m:\mathbb{N}\to\mathbb{N}$ by $f_m(n)=\frac{1}{2}k(n+r-(m+1))$, if $n=k\cdot(m+1)+r$, $k,r\in\mathbb{N}$, $0\leqslant r<m+1$. Explicitly, $f_m(n)=\frac{1}{2}\cdot\lfloor\frac{n}{m+1}\rfloor\cdot(n+\mathrm{mod}(n,m+1)-(m+1))$. The following is now obvious:

PROPOSITION 4. *For all* m, $n\in\mathbb{N}$,

$$\theta_m(n) = f_m(\gamma_m(n)). \quad \square$$

As it is an easy calculation to show that, for each m, f_m is strictly increasing on $\{n\in\mathbb{N}: n\geqslant m+1\}$ and zero on $\{n\in\mathbb{N}:n\leqslant m+1\}$, we also have:

COROLLARY 1. *For all* n, n_1, $m\in\mathbb{N}$,
a) $\theta_m(n)=0\Leftrightarrow\gamma_m(n)\leqslant m+1$,
b) If $\theta_m(n)>0$ then: $\theta_m(n)=\theta_m(n_1)\Leftrightarrow\gamma_m(n)=\gamma_m(n_1)$. □

The mappings f_m in fact define $\mathbb{N}\times(m+2)$-matrices the first 15 rows of which are shown in tables 2 and 3 for $1\leqslant m\leqslant 4$.
We associate with $(\theta_m(n))_{n\in\mathbb{N}}$ a matrix $\mathfrak{A}=(A_{ij}^m)$ $(i\in\mathbb{N}, 0\leqslant j<m+2)$ with entries $A_{ij}^m=f_m(i+j)=\theta(i\cdot(m+2)+j,m)$. So, the entries in $\mathfrak{A}$, read from left to right, constitute precisely the sequence $(\theta_m(n))_{n\in\mathbb{N}}$.
From $\mathfrak{A}$ we obtain for each $m\in\mathbb{N}$ a second $\mathbb{N}\times(m+2)$-matrix $\mathfrak{B}=(B_{ij}^m)$ $(i\in\mathbb{N}, 0\leqslant j<m+2)$ by defining:

$$B_{ij}^m = \begin{cases} A_{i,j+1}^m - A_{i,j}^m & \text{if } j<m+1, \\ A_{i+1,0}^m - A_{i,m+1}^m & \text{if } j=m+1. \end{cases}$$

It is clear by construction, that the entries in $\mathfrak{B}$, read from left to right, constitute precisely the sequence $(\Delta\theta_m(n))_{n\in\mathbb{N}_{>0}}$, with $i=\lfloor\frac{n-1}{m+2}\rfloor$,

$j = \mathrm{mod}(n-1, m+2)$.
By a straightforward calculation one obtains from the definition of f_m and A_{ij}^m:

$$B_{ij}^m = \begin{cases} \lfloor \frac{i+j}{m+1} \rfloor & \text{if } j < m+1, \\ -i + \lfloor \frac{i}{m+1} \rfloor & \text{if } j = m+1. \end{cases}$$

This immediately proves (using proposition 3a)

PROPOSITION 5. *For all* $m \in \mathbb{N}$, $n \in \mathbb{N}_{>0}$.

$$\Delta\theta_m(n) = -\lfloor \frac{\beta}{m+1} \rfloor \cdot (\alpha+1) + \|G_m(n-1)\|$$

where $\alpha = \lfloor \frac{n-1}{m+2} \rfloor$, $\beta = \mathrm{mod}(n-1, m+2)$. □

5. A CONJECTURE ON EQUALITY II.

In section 4 we derived the blue print for a recursive machinery generating $S(n)$ and $P(n)$, using elementary arithmetic. As far as conjecture 1 in section 3 is concerned however, we already mentioned that the probability of finding an elementary proof for the right-to-left implication is very small. Therefore the results of section 4 are unlikely to bring us any nearer to a possible solution. All we have is a bone-ache.

REFERENCES

J.T. CROSS (1974). A note on almost perfect numbers. *Math. Mag. 47:* 230-231.

L.E. DICKSON (1919). *History of the theory numbers,* vol. I. Carnegie Institution, Washington.

R.K. GUY (1981). *Unsolved problems in number theory.* Springer Verlag, New York.

G.H. HARDY, E.M. WRIGHT (1960). *Introduction to the theory of numbers.* Oxford University Press.

E. LUCAS (1891). *Theórie de Nombres.* Tome Premier. Gauthier-Villars & Fils, Paris.

Products of random matrices or 'why do biennials live longer than two years?'

J.B.T.M. Roerdink
Centre for Mathematics and Computer Science
Kruislaan 413, 1098 SJ Amsterdam

We study a population model with two age classes which describes the growth of biennial plants in a randomly fluctuating environment. A fraction of the oldest age class delays its flowering each year. Using the theory of random matrix products we show that delay of flowering increases the growth rate of the population. We investigate the dependence of the optimal flowering fraction on the model parameters.

1. Introduction

Strictly biennial plants lead a vegetative existence in the first year, flower in the second year and then die. In natural populations of 'biennials' often a large fraction of the plants delays its flowering past the second year: the non-reproductive period of a lineage may span 2-5 years. This is surprising because classical life-history theory predicts that a biennial has to produce four times as many seeds as a perennial and twice as many seeds as an annual plant to attain the same rate of increase. This conclusion was reached by assuming a deterministic (nonchanging) environment.

This picture changes completely when the population grows in a randomly fluctuating environment. Simulation studies by Klinkhamer and de Jong [1] indicated that some delay of flowering is profitable if the per capita reproductive success varies strongly over the years. These authors used the following discrete-time model pertaining to the case of density-independent population growth:

$$\begin{pmatrix} N_{1,t+1} \\ N_{2,t+1} \end{pmatrix} = \begin{pmatrix} 0 & f\phi_t \\ s & (1-f)s \end{pmatrix} \begin{pmatrix} N_{1,t} \\ N_{2,t} \end{pmatrix} \quad t = 0,1,2,\cdots, \tag{1.1}$$

where t is a time just before flowering, and

$N_{1,t}$ the number of individuals which are one year old at time t;
$N_{2,t}$ the number of individuals older than one year at time t;
s survival rate of individuals older than one year, $0 \leqslant s \leqslant 1$;
f fraction of individuals older than one year that flowers in a given year, $0 \leqslant f \leqslant 1$;
ϕ_t fluctuating number of offspring per flowering plant after one year, $\phi_t \geqslant 0$.

We assume that $\{\phi_t\}$ is a sequence of independent, identically-distributed (i.i.d.) random variables, where each ϕ_t has a two-parameter gamma distribution with density given by

$$g(\phi) = \{k^a/\Gamma(a)\}\phi^{a-1}e^{-k\phi}, \quad a>0, \quad k>0, \tag{1.2}$$

with $\Gamma(a)$ the gamma function. Mean and variance of this distribution are

$$\bar{\phi} := \mathbb{E}(\phi) = a/k; \quad \bar{\bar{\phi}} := \mathrm{Var}(\phi) = a/k^2, \tag{1.3}$$

where $\mathbb{E}$ denotes the mathematical expectation operator.

The long-run properties of the solution of (1.1) depend on the behaviour of the random matrix products $M_t M_{t-1} \cdots M_0$, where M_t is the two-by-two matrix in Eq. (1.1). Below we give an informal presentation of the results which have been obtained for this model in [2] by using the theory of products of random matrices (for the latter see e.g. [3,4]). In particular we will discusss the existence of an optimal flowering fraction, f_{opt}, for our model and discuss how the value of f_{opt} depends on the model parameters $s, \bar{\phi}$ and $\bar{\bar{\phi}}$. We do not go into the question as to how the delayed flowering is brought about in natural populations. For a more extensive discussion of this point and other biologically relevant questions we refer to [5].

2. Average behaviour is not typical behaviour

The aim of this paragraph is to define the concept of 'optimal flowering fraction' and to bring out the difference between deterministic and stochastic environments with regard to the proper definition of this concept.

First we will look at the *deterministic* case. So let us assume that ϕ_t in (1.1) has a constant value $\bar{\phi}$ (which will later be identified with the average value of ϕ_t in the stochastic case). We define the optimal flowering fraction f_{opt} in this case as the value of f for which the total population $N_t := N_{1,t} + N_{2,t}$ has a maximal growth rate μ_0, defined by

$$\mu_0 := \lim_{t\to\infty} \frac{1}{t} \ln N_t. \tag{2.1}$$

It is easy to see that μ_0 is equal to $\ln\lambda_0$, where λ_0 is the maximal eigenvalue of the matrix M_t with $\phi_t = \bar{\phi}$, which is given by

$$\lambda_0(f) = \tfrac{1}{2}[s(1-f) + \{s^2(1-f)^2 + 4f\bar{\phi}s\}^{1/2}]. \tag{2.2}$$

If we assume that without delayed flowering the population is *increasing*, i.e. $\lambda_0(1) = \sqrt{\bar{\phi}s} > 1$, then delayed flowering is not profitable since λ_0 (or μ_0) is an increasing function of f for $\bar{\phi}s > 1$ (in fact for $\bar{\phi} > s$). As an example we have plotted μ_0 as a function of f for the parameter values $s = 0.9$, $\bar{\phi} = 2$ in Fig. 1 (broken line).

If we now turn to the case of a *random* environment our first difficulty is the choice of the optimality or 'fitness' criterion to be used in defining an optimal flowering fraction. There are at least two intuitively plausible ways to generalize the maximization of (2.1). First, we may maximize the growth rate of the

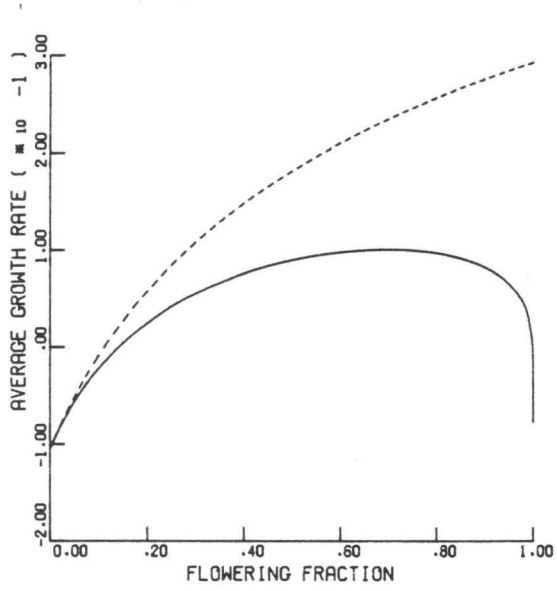


FIGURE 1. Geometric growth rate γ (solid line) and arithmetic growth rate μ vs. the flowering fraction; $s=0.9$; $\bar{\phi}=2$, $\bar{\bar{\phi}}=5$.

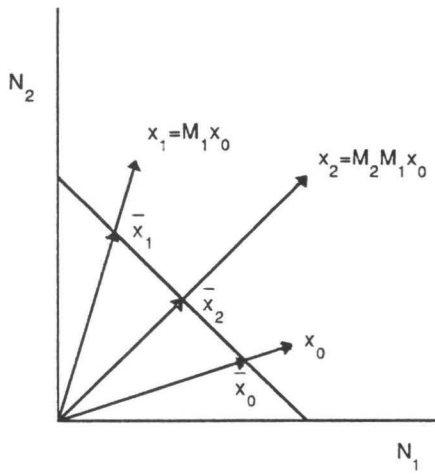


FIGURE 2. Iterates of the population vector x_t and the age-structure $\bar{x}_t$ under the action of (1.1).

average population, defined as

$$\mu = \lim_{t\to\infty} \frac{1}{t} \ln \mathbb{E}(N_t). \tag{2.3}$$

In fact, $\mu=\mu_0=\ln\lambda_0$ with λ_0 given by (2.2), where now $\bar{\phi}$ is the average of ϕ_t. A second possibility is to consider the average growth rate (also called the upper Lyapunov exponent) defined by

$$\gamma = \lim_{t\to\infty} \frac{1}{t} \mathbb{E}(\ln N_t). \tag{2.4}$$

In general the arithmetic growth rate μ will be strictly larger than the geometric growth rate γ.

The theory of random matrix products tells us that a typical realization is characterized by γ, not by μ. More precisely, under some mild conditions (which are satisfied in our case),

$$\lim_{t\to\infty} \frac{1}{t} \ln N_t = \gamma \tag{2.5}$$

for almost all sequences of matrices $\{M_t\}$ and any initial value $N_0\neq 0$. Thus, although eventually each population growing according to (1.1) attains a growth rate γ, there are at each given time t, no matter how large, always populations which have been growing at a much larger rate, due to a rare succession of years with favourable environmental conditions. It is because of these exceptional realizations that the average (2.3) differs from the typical growth rate γ. Also, it can be shown that a maximal geometric growth rate is the biologically appropriate 'fitness criterion': if in a population with geometric growth rate γ_0 a mutant occurs with growth rate γ_1, then the mutant will invade and eventually outnumber the original population if $\gamma_1>\gamma_0$.

The analytical computation of γ is in general not a trivial matter. In fact there are only a few models for which γ has been explicitly determined, our model (1.1) being one of them. The calculation will be briefly sketched in the next paragraph.

3. The upper Lyapunov exponent γ

First we introduce the following notation.

$$x_t = (N_{1,t}, N_{2,t}), \tag{3.1}$$

$$|x_t| = N_t = N_{1,t}+N_{2,t}, \tag{3.2}$$

$$\bar{x}_t = x_t/|x_t|. \tag{3.3}$$

In biological terms: x_t is the population vector, $|x_t|$ the total population and $\bar{x}_t$ the 'age-structure'. So $\bar{x}_t$ is defined on the simplex

$$C = \{x\in\mathbb{R}^2: x_1\geq 0,\ x_2\geq 0,\ x_1+x_2=1\} \tag{3.4}$$

Every time the iteration (1.1) is applied we get a new population vector x_t and corresponding age-structure $\bar{x}_t$, see Fig. 2. The sequence $\{\bar{x}_t\}$ constitutes a

Markov chain on the simplex C and the distribution of age-structures evolves towards a unique stationary probability measure $\nu(\bar{x})$, which is invariant with respect to the common distribution $\mu(M)$ of the matrices $\{M_t\}$.

In terms of the given measure μ and the (as yet unknown) measure ν the expression for the growth rate γ is,

$$\gamma = \int\int \ln|M\bar{x}|d\mu(M)d\nu(\bar{x}). \tag{3.5}$$

The problem is therefore to determine the stationary age-structure distribution ν. It is convenient to introduce the scaled variable

$$\tau_t = (1-f)^{-1}\{N_{1,t}/N_{2,t}\}, \tag{3.6}$$

so that (1.1) can be written as

$$\tau_{t+1} = \eta\phi_t/(1+\tau_t) \tag{3.7}$$

where

$$\eta = f\{(1-f)^2 s\}^{-1}. \tag{3.8}$$

For the invariant measure expressed in the variable τ we write $\tilde{\nu}(\tau)$. The corresponding density (which exists) is denote by $h(\tau)$. By using the invariance of $h(\tau)$ under the action of (3.7) one can derive the integral equation

$$h(\tau) = \int_0^\infty h(\tau')g\left[\frac{(\tau'+1)\tau}{\eta}\right]\left[\frac{\tau'+1}{\eta}\right]d\tau' \tag{3.9}$$

where $g(\cdot)$ it the gamma distribution (1.2). The solution is

$$h(\tau) = K^{-1}\tau^{a-1}(1+\tau)^{-a}e^{-z\tau} \tag{3.10}$$

where K is a normalization constant.

By using this result in (3.5) and performing some simplifications in the resulting formula one ends up with the following expression for $\gamma = \gamma(f)$,

$$\gamma(f) = \ln\{s(1-f)\} + \frac{\int_0^\infty \{\ln(1+\tau)\}\tau^{a-1}(1+\tau)^{-a}e^{-z\tau}d\tau}{\int_0^\infty \tau^{a-1}(1+\tau)^{-a}e^{-z\tau}d\tau}, \tag{3.11}$$

where

$$z = ks(1-f)^2/f. \tag{3.12}$$

The limiting values at the endpoints of the interval [0,1] are,

$$\gamma(0) = \ln s, \quad \gamma(1) = \tfrac{1}{2}\{\ln(s/k)+\psi(a)\} \tag{3.13}$$

where $\psi(a) = \frac{d}{da}\Gamma(a)$ is the digamma function. As a final remark we note that the second term in the r.h.s. of (3.11) can be expressed in terms of Kummer functions and derivatives thereof [2].

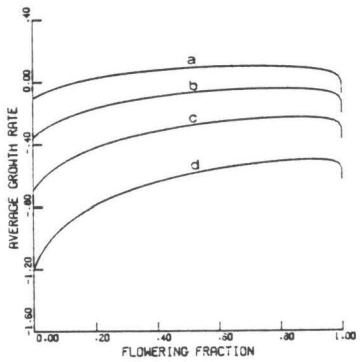


FIGURE 3. Geometric growth rate vs. the flowering fraction ($\bar{\phi}=2$, $\bar{\bar{\phi}}=5$): (a) $s=0.9$; (b) $s=0.7$; (c) $s=0.5$; (d) $s=0.3$.

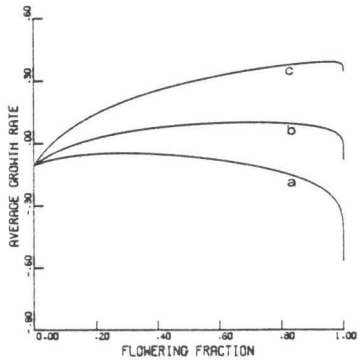


FIGURE 4. Geometric growth rate vs. the flowering fraction ($s=0.9$, $\bar{\bar{\phi}}=5$); (a) $\bar{\phi}=1$; (b) $\bar{\phi}=2$; (c) $\bar{\phi}=3$.

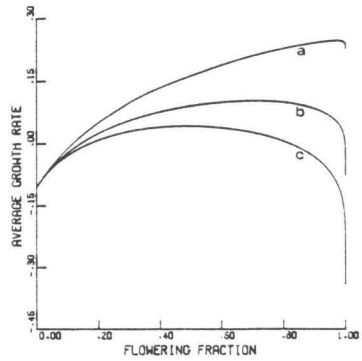


FIGURE 5. Geometric growth rate vs. the flowering fraction ($s=0.9$, $\bar{\phi}=2$); (a) $\bar{\bar{\phi}}=1$; (b) $\bar{\bar{\phi}}=5$; (c) $\bar{\bar{\phi}}=8$;

4. Maximizing the fitness

The biologically interesting question is now to see for which values of the flowering fraction f the geometric growth rate γ attains a maximum. If such a maximum occurs for a value $f_{opt}<1$ then delayed flowering is profitable.

First we calculate γ as a function of the flowering fraction f for the parameter values $s=0.9$, $\bar{\phi}=2$, $\bar{\bar{\phi}}=5$ (where $\bar{\phi}$, $\bar{\bar{\phi}}$ are the mean and variance of the offspring number ϕ_t), see the solid curve in Fig. 1. The result is obtained by numerical evaluation of the integrals occurring in (3.11). We note the following:

(i) γ attains a maximum for a value $f_{opt}<1$: delayed flowering is profitable ('spreading of the risk')

(ii) $\gamma<\mu$ for $0<f\leqslant 1$. In fact $\mu(1)>0$, $\gamma(1)<0$, so if $f=1$ the population will almost surely go extinct while the average population increases indefinitely!

(iii) a very steep decrease of γ near $f=1$ is observed (in fact [2]: $\frac{d\gamma}{df}\Big/_{f=1}=-\infty$): even a tiny amount of delayed flowering is advantageous. The precise amount is not very crucial, as can be seen from the small variation of γ in the region $0.5\lesssim f\lesssim 0.9$.

Of course the value of f_{opt} will depend on the parameters of the model, i.e. on s, a and k or, equivalently, on $s, \bar{\phi}$ and $\bar{\bar{\phi}}$. In Fig. 3 we have plotted γ as a function of f for the values $s=0.3, 0.5, 0.7, 0.9$ with $\bar{\phi}=2$, $\bar{\bar{\phi}}=5$. The optimal fraction f_{opt} increases as s decreases. This is to be expected since a smaller survival probability has to be balanced by a larger amount of flowering.

In Figs. 4 and 5 we show curves of γ versus f for various values of $\bar{\phi}$ and $\bar{\bar{\phi}}$ respectively, keeping the other parameters fixed. One observes that f_{opt} decreases as the average offspring number $\bar{\phi}$ decreases or its variance $\bar{\bar{\phi}}$ increases. Hence if the average reproductive success is low or varies strongly there is a strong environmental pressure on the population to delay its flowering.

Field studies of biennials suggest that the model (1.1) is certainly inadequate in many respects, see e.g. [5]. For example, no density dependence is assumed and there is no advantage of delayed flowering through increased probability of survival and increased seed production. Nevertheless, the prediction of the simple model (1.1) that f_{opt} is considerably smaller than unity for a wide range of parameters is in agreement with the fact that only very few species are strictly biennial.

References

[1] P.G.L. Klinkhamer, T.J. de Jong (1983). *Is it Profitable for Biennials to Live Longer than Two Years?* Ecol. Modell. *20*, 223-232.

[2] J.B.T.M. Roerdink (1988). *The Biennial Life Strategy in a Random Environment,* J. Math. Biol. *26*, 199-215, and Supplement (to appear).

[3] J.E. Cohen, H. Kesten, C.M. Newman (eds.) (1986). *Random Matrices*

and their Applications, A.M.S., Providence (Contemporary Mathematics, Vol. 50).

[4] P. BOUGEROL, J. LACROIX (1985). Products of random matrices with application to Schrödinger operators, Birkhäuser, Boston (Progress in Probability and Statistics, Vol. 8) Part A.

[5] T.J. DE JONG, P.G.L. KLINKHAMER, S.A.H. GERITZ and E. VAN DER MEIJDEN, *Why biennials delay flowering: an optimization model and field data on Cirsium vulgare and Cynoglossum officinale,* Acta Botanica Neerlandica, submitted.

Asymptotic Expansion of a Special Integral

N.M. Temme

Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

The integral

$$I = \int_a^b \left[\frac{x+a}{x-a}\right]^{2ai} \left[\frac{b-x}{b+x}\right]^{2bi} \frac{dx}{x}$$

is considered for large positive values of a and b; the parameters tend to infinity in such a manner that the quotient $b/a = c$ is a constant greater than 1. In a recent paper, Mahler showed that the integral tends more rapidly to 0 than any finite negative power of a and he gives an upper bound of the integral. As Mahler admitted, his results do not imply estimates of the form $I = \mathcal{O}(e^{-a})$. Our results give $I = \mathcal{O}(e^{-2\pi a})$. Mahler's technique is based on integration by parts. Here we use a different technique, based on complex variables, and we construct the leading term and the first terms in the asymptotic expansion.

1. Mahler's approach

Quite recently the physicist J. Lekner of the Victoria University of Willington, New Zealand presented K. Mahler of the Australian National Univeristy of the following integral

$$I = \int_a^b \left[\frac{x+a}{x-a}\right]^{2ai} \left[\frac{b-x}{b+x}\right]^{2bi} \frac{dx}{x}. \tag{1.1}$$

Lekner was interested in the behaviour of I for large values of the parameters a, b. He used this integral for describing the Rayleigh approximation for a reflection amplitude in the theory of electromagnetic and particle waves. See formula 6.64 in [2]. Mahler [3] proved the following two theorems.

Theorem 1. *Assume that the two positive parameters a and b tend tp $+\infty$ in such a manner that the quptient $b/a = c$ remains equal to a constant $c > 1$. Then the integral I in (1.1) tends more rapidly to 0 than any finite, negative, power of a*

In other words, the integral I is asymptotically equal to zero with respect to the scale $\{a^{-n}\}$, as $a \to \infty$. That is,

$$I \sim 0 \quad \{a^{-n}\}, \text{ as } a \to \infty.$$

For this notation we refer to Lauwerier [1].

Another theorem appeared in the second part of [3]:

THEOREM 2. *Denote by $A(a)$ a monotone, increasing, positive-valued continuous function of a which tends arbitrarily slowly to $+\infty$ as a tends to $+\infty$; further let $\Gamma>0$ be an arbitrarily large positive constant. Then for all sufficiently large positive a the integral (1.1) satisfies the inequality*

$$|I| < \exp\left[\frac{-a\Gamma}{A(a)}\right] \tag{1.2}$$

Mahler suggests to take for A the m-times iterated logarithm

$$\ln_m(a) = \ln(\ln(...(\ln a)...)), \ (m \ log's),$$

where m is any positive integer. Observe that in (1.2) $\Gamma/A(a)$ tends to 0. As Mahler admits, his result does not quite imply the estimate $|I| = \mathcal{O}(e^{-a})$, and he concludes his paper with the words *'and I do not know whether it is true'*. From our results it follows that $|I| = \mathcal{O}(e^{-2\pi a})$, and the purpose of the paper is to give a complete description of the asymptotic behaviour of I.

Mahler proved his result by considering

$$I = \int_1^c F(u)^{2ai}\frac{du}{u}, \quad F(u) = \frac{u+1}{u-1}\left[\frac{c-u}{c+u}\right]^c. \tag{1.3}$$

This representation easily follows by using $b = ac$ and introducing a new variable of integration u by writing $x = au$. Then Mahler used integration by parts to prove the theorems.

2. COMPLEX VARIABLE APPROACH

As remarked earlier, the above theorems give only a partial result, since no information is given on the leading term of the asymptotic estimate and of the terms in the asymptotic expansion. It does not seem possible that one can obtain these leading terms by using only real integration variables. Therefore we replace (1.1) by a loop integral in the complex plane, which gives the required information. But first we transform (1.1) into an integral on the interval $[0,\infty)$. It is easier to handle then the various branch-points of the integrand, since one of them is sent to ∞ by this transformation.

Let us write $t = (x-a)/(b-x)$. Then we obtain

$$I = A\int_0^\infty \frac{t^{-2ai}(t+t_1)^{2ai}dt}{(t+t_2)^{2bi}(1+t)(1+ct)}, \tag{2.1}$$

where

$$t_1 = \frac{2}{c+1}, \quad t_2 = \frac{c+1}{2c} \tag{2.2}$$

and

$$A = (2c)^{-2bi}(c+1)^{2ai}(c-1)^{2bi-2ai+1}. \tag{2.3}$$

Recall that $c>1$. Hence we have the inequalities

$$0<\frac{1}{c}<\frac{2}{c+1}<\frac{c+1}{2c}<1.$$

This shows that the five singular points $0,-1/c,-t_1,-t_2,-1$ of the above integral satisfy $-1<-t_2<-t_1<-1/c<0$. Since c is assumed to be fixed, no confluence of singularities can happen. Otherwise the problem would be much more difficult.

Before choosing a proper loop integral based on (2.1), we compute the stationary point(s) of the integrand.

We write

$$\frac{t^{-2ai}(t+t_1)^{2ai}}{(t+t_2)^{2bi}}=e^{-2ai\phi(t)},$$

introducing the function

$$\phi(t)=\ln t-\ln(t+t_1)+c\ln(t+t_2). \tag{2.4}$$

For $t>0$ we assume real values of the logarithms. It is straightforward to verify that

$$\phi'(t)=\frac{c(t+1/c)^2}{t(t+t_1)(t+t_2)}. \tag{2.5}$$

Hence ϕ has a (double) stationary point at $-1/c$. It follows that we can expand

$$\phi(t)=\phi(-1/c)+\frac{1}{6}\phi'''(-1/c)(t+1/c)^3+O(t+1/c)^4. \tag{2.6}$$

A few computations give

$$\phi'''(-1/c)=-\frac{4c^4(c+1)}{(c-1)^2}. \tag{2.7}$$

We observe that (2.1) has a double stationary point outside the interval of integration, and that this point coincides with a single pole of the integrand. From an asymptotic point of view, this combination of phenomena is not just trivial. However, in Lauwerier's book the theory needed to handle this problem is presented for an analogue case.

3. A LOOP INTEGRAL

The final preparatory step is to introduce a suitable loop integral of which the path of integration can be shifted to the stationary point at $-1/c$. We introduce

$$J=\int_{-\delta-i\infty}^{-\delta+i\infty}\frac{(-t)^{-2ai}(t+t_1)^{2ai}dt}{(t+t_2)^{2bi}(1+t)(1+ct)}, \tag{3.1}$$

where δ is a positive number satisfying $0<\delta<1/c$. It is not difficult to verify that the integral is convergent at ∞. The phase of the complex parameter t is,

initially, between $\pi/2$ and $3\pi/2$. The minus-sign in $(-t)$ is interpreted as $e^{-\pi i}$ (this choice is irrelevant, but it brings a nice symmetry in the relation between I and J). Hence we assume that for negative values of t the phase of $-t$ equals zero. The branches of the remaining many-valued functions are chosen in the normal way: we assume that the phases of $t+t_1$ and $t+t_2$ are zero for positive values of t.

Our procedure is as follows. First we show, by modifying the vertical path of integration, that J equals I (up to a simple function of a and b). On the other hand, we can shift the vertical path to the left until it meets the real negative axis at $-1/c$; that is, we let $\delta \to 1/c$. Then we apply the method of stationary phase from asymptotics.

To recover I from the complex integral, we bend the vertical path around the interval $[0,\infty)$. At the upper side of this interval, where $\arg t = 0$, we have

$$(-t)^{-2ai} = (e^{-\pi i}|t|)^{-2\pi a}|t|^{-2ai}.$$

At the lower side, where $\arg t = 2\pi i$, we have

$$(-t)^{-2ai} = (e^{-\pi i}|t|e^{2\pi i})^{-2ai} = e^{2\pi a}|t|^{-2ai}.$$

The integration near the origin gives no problems. So we arrive at the result

$$J = -2\sinh(2\pi a)I/A, \tag{3.2}$$

where A is given in (2.3).

4. Asymptotic expansion

We slightly change the phase function introduced in (2.4) by writing

$$\phi(t) = \ln(e^{-\pi i}t) - \ln(t+t_1) + c\ln(t+t_2). \tag{4.1}$$

The formulas (2.5), (2.6) and (2.7) also hold for this new ϕ. We have

$$J = \int_{\mathfrak{L}} e^{-2ia\phi(t)} \frac{dt}{(1+t)(1+ct)},$$

where $\mathfrak{L}$ is the above introduced vertical, now with $\delta = 1/c$ and with a small semi-circle at the right of the pole at $-1/c$. We introduce the transformation of variables (see (2.6))

$$w^3 = \frac{\phi(t)-\phi(-1/c)}{\phi'''(-1/c)/6}, \tag{4.2}$$

and we choose the branch that satisfies $w \sim t+1/c$ in a neighborhood of the stationary point $-1/c$.

On using (4.1), (4.2) in (3.1), we obtain

$$J = B\int e^{\frac{1}{3}i\mu a w^3} f(w)\frac{dw}{w}, \tag{4.3}$$

where

$$B = e^{-2ai\phi(-1/c)} = (c-1)/A, \tag{4.4}$$

$$f(w) = \frac{w}{(1+t)(1+ct)}\frac{dt}{dw}, \tag{4.5}$$

$$\mu = -\phi'''(-1/c). \tag{4.6}$$

Since μ is positive (see (2.7)), the 'best' path in (4.3) is the steepest descent path defined by the rays

$$\arg w = -\pi/2, \ \arg w = \pi/6. \tag{4.7}$$

Locally, the same holds for the t-plane near $t = -1/c$. In order to avoid the pole, the path in the w-plane has a small circular arc near the origin. The integration runs from $-i\infty$ to $\infty\exp(\pi i/6)$, and the pole at the origin is at the left hand side of the contour.*)

We substitute the MacLaurin series

$$f(w) = \sum_{k=0}^{\infty} c_k w^k \tag{4.8}$$

in (4.3), and we interchange summation and integration. The result is the asymptotic expansion

$$J \sim B\sum_{k=0}^{\infty} c_k F_k, \quad F_k = \int e^{\frac{1}{3}i\mu a w^3} w^{k-1} dw, \tag{4.9}$$

as $a \to \infty$. To compute F_k we use the path described by (4.7). F_0 needs some special care. We write

$$F_0 = [\int_{-i\infty}^{-ir} + \int_{r\exp(i\pi/6)}^{\infty\exp(i\pi/6)}] e^{\frac{1}{3}i\mu a w^3} \frac{dw}{w} + i\int_{-\pi/2}^{\pi/6} e^{\frac{1}{3}i\mu a r^3} e^{3i\theta} d\theta$$

for any positive number r. The first two integrals cancel. The third one assumes in the limit $r \to 0$ the value $2\pi i/3$, which is $1/3$ of the residue of the pole. Hence $F_0 = 2\pi i/3$. The remaining integrals follow straightforwardly:

$$F_k = [e^{i\pi k/6} - e^{-i\pi k/2}] \int_0^{\infty} e^{-\frac{1}{3}\mu a w^3} w^{k-1} dw$$

$$= \frac{2i}{3} e^{-i\pi k/6} \sin(k\pi/3)\Gamma(k/3)(\mu a/3)^{-k/3}, \quad k = 1,2,3,\ldots.$$

Observe that this result can also be interpreted for $k = 0$. Combining (3.2), (4.4) and (4.9) we obtain the final result

$$I \sim -\frac{c-1}{2\sinh(2\pi a)} \sum_{k=0}^{\infty} c_k F_k, \quad \text{as } a \to \infty. \tag{4.10}$$

The dominant term in this expansion reads very simple. We have from (4.5) by

*) We do not prove that the function f admits such a contour; to do so, we should examine the mapping (4.2) more globally. For the construction of the asymptotic expansion we only need a local analysis around the origin.

using l'Hôpital's rule

$$c_0 = f(0) = \frac{dt}{dw}|_{w=0} \frac{c}{c-1} \lim_{w\to 0} \frac{w}{1+ct} = \frac{1}{c-1}.$$

This gives

$$I \sim -\frac{\pi i}{3\sinh 2\pi a}, \quad \text{as } a\to\infty.$$

We conclude this section by giving the first few coefficients c_k. They are obtained by using (4.2), (4.5), and (4.8). Thus we obtain

$$c_0 = \frac{1}{c-1}, \quad c_1 = 0, \quad c_2 = -\frac{2c^2(c^2+1)}{5(c-1)^3}, \quad c_3 = \frac{4c^3(c^2+1)}{5(c-1)^4},$$
$$c_4 = -\frac{2c^4(3c^4+17c^2+24)}{35(c-1)^5}, \quad c_5 = \frac{8c^5(3c^4+3c^2+10)}{35(c-1)^6}.$$

5. A Fourier integral

When we take in (1.3)

$$t = -\ln F(u) = \ln(u-1) - \ln(u+1) + c\ln(c+u) - c\ln(c-u) \quad (5.1)$$

as a new variable of integration, we obtain the Fourier integral

$$I = \int_{-\infty}^{\infty} e^{-2ait} g(t)dt, \quad g(t) = \frac{1}{u}\frac{dt}{du} = \frac{(u^2-1)(c^2-u^2)}{2u^3(c^2-1)}. \quad (5.2)$$

By considering the mapping $u\mapsto t$ in more detail, we see that it is one-to-one on $[1,c]$, and that, consequently, g is a C^∞-function on $\mathbb{R}$. Moreover, g is exponentially small at $\pm\infty$. That is,

$$g(t) = O(e^{t/c}), \text{ as } t\to-\infty, \quad g(t) = O(e^{-t}), \text{ as } t\to+\infty.$$

This easily follows from (5.1) and (5.2). By using these properties, Theorem 1 can be proved immediately.

The function g is singular at $u = 0$, that is, at $t = -i\pi$. Hence, we can shift the contour of integration in (5.2) downwards to this point, and we can expand the function g at this singularity. Observe that the exponential function in (5.2) assumes the value $\exp(-2\pi a)$ at this point. This dominant factor also occurs in (4.10), and we expect that (5.2) can be used to obtain the same or a similar expansion.

References

[1] H.A. Lauwerier, *Asymptotic analysis*, Mathematical Centre Tract 54, Amsterdam, 1974.

[2] J. Lekner, *Theory of reflection of electromagnetic and particle waves*, Martinus Nijhoff, Dordrecht, 1987.

[3] K. Mahler, *On a special integral, I, II*. Research Reports 37, 38; Department of Mathematics, The Australian National University, 1986.

A Radon transform on circles through the origin in $\mathbb{R}^2$.

M. Zwaan

We invert a special kind of Radon transform that maps a function to its integrals over circles through the origin.

1. Introduction

The Radon transform is an integral transform, named after J. Radon, that maps a function $f:\mathbb{R}^n \to \mathbb{R}$ on the set of its integrals over the hyperplanes of $\mathbb{R}^n$. If the hyperplanes of $\mathbb{R}^n$ are parametrized by a unit normal $\theta \in S^{n-1}$, the unit sphere in $\mathbb{R}^n$, and the distance to the origin $|p|$, then we denote the integral of f over one such plane by $Rf(p,\theta)$ and the map $f \to Rf$ is the Radon transform.

This transform has many practical applications in engineering and medicine ([3], [5], [7]), but it is also of theoretical interest, with applications to partial differential equations, integral geometry and Lie groups ([1], [2]). In a more abstract setting the Radon transform is defined as the transform that maps a function f defined on a differentiable manifold, to its integrals over certain submanifolds.

As a special case of this generalized Radon transform we want to study the Radon transform, denoted by Qf, which maps a function $f:\mathbb{R}^2 \to \mathbb{R}$ onto its integrals over circles through the origin. This transform has been studied by John ([7]) and Cormack and Quinto ([5]) for functions $f:\mathbb{R}^n \to \mathbb{R}$. Cormack and Quinto give an inversion formula for this transform by expressing the function f in terms of a (orthonormal) basis of spherical harmonics (in $L_2(S^{n-1})$), $f(p\theta) = \Sigma_l f_l(p)\, Y_l(\theta)$ $(p \in \mathbb{R}, \theta \in S^{n-1})$ and by giving an inversion formula for each component $f_l(p)$. In the case $n=2$ Cormack [6] gives an explicit alternative derivation of an inversion formula for the transform $f \to Qf$.

Here we give an inversion formula for functions $f:\mathbb{R}^2 \to \mathbb{R}$, by relating the transform Qf to Rf by transforming lines onto circles and by using the known inversion formula for Rf.

2. Definitions

Define the line $L_{p,\phi} \subset \mathbb{R}^2$ by

$$L_{p,\phi} := \left\{ x \in \mathbb{R}^2 | x \cdot \theta - p = 0; \theta = \begin{bmatrix} \cos\phi \\ \sin\phi \end{bmatrix} \right\}. \tag{2.1}$$

Let $f: \mathbb{R}^2 \to \mathbb{R}$ be a continuous function with compact support. The Radon transform of f is

$$Rf(p,\phi) := \int_{L_{p,\phi}} f(x)dx. \tag{2.2}$$

notice that $L_{p,\phi} = L_{-p,\phi+\pi}$. In fact, we can represent-each line $L_{p,\phi}$ as a point on the cylinder, $(p,\theta) \in \mathbb{R} \times S^1$, with $\theta := [\cos\phi, \sin\phi]^T$; we can identify the points (p,θ) and $(-p,-\theta)$. This interpretation gives us the relation

$$Rf(p,\phi) = Rf(-p,\phi+\pi). \tag{2.3}$$

Let $C_{p',\phi'}$ be the circle through the origin, (see fig. 1)

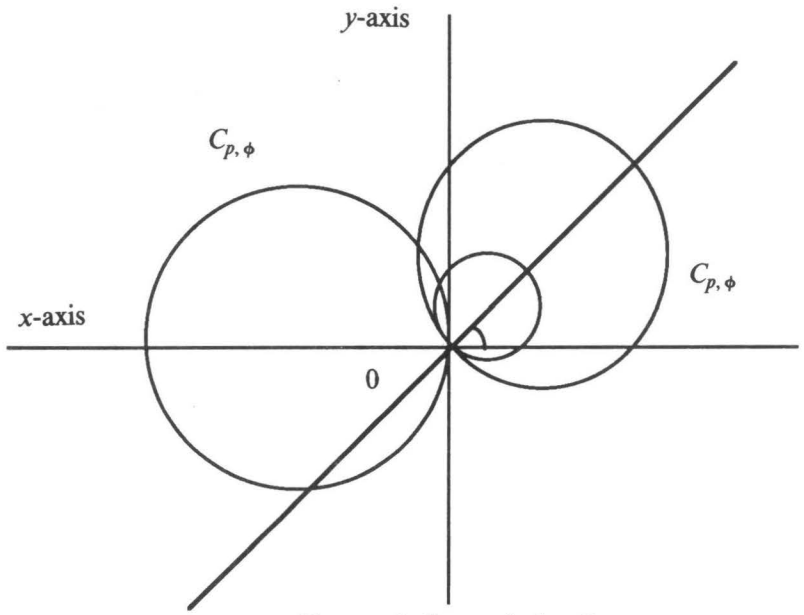


FIGURE 1. Some circles $C_{p',\phi'}$

$$C_{p',\phi'} := \{(x,y) \in \mathbb{R}^2 | (x - \frac{p'}{2}\cos\phi')^2 + (y - \frac{p'}{2}\sin\phi')^2 = \frac{(p')^2}{4}\}.$$

Define the Radon transform $f \to Qf$ as

$$Qf(p',\phi') := \int_{C_{p',\phi'}} f(x)dx. \tag{2.4}$$

Again we have $C_{p',\phi'} = C_{-p',\phi'+\pi}$, so

$$Qf(p',\phi') = Qf(-p',\phi' + \pi). \tag{2.5}$$

For the transform $f \to Rf$, there is an inversion formula ([3] p. 21, or [4], [8])

$$g(x) = \frac{1}{4\pi^2} \int_0^{2\pi} \int_{-\infty}^{\infty} \frac{\frac{\partial}{\partial p} Rg(p,\phi)}{\theta \cdot x - p} dp d\phi, \tag{2.6}$$

valid for continuous functions with compact support.

3. Relation between Q and R

Suppose we can find a function $\Phi: \mathbb{R}^2 \to \mathbb{R}^2$, which maps the line $L_{p,\phi}$ onto the circle $C_{p',\phi'}$, which is invertible and continuously differentiable. Then we have

$$\begin{aligned} Qf(p',\phi') &= \int_{C_{p',\phi'}} f(x)\, dx = \int_{\Phi(L_{p,\phi})} f(x)\, dx = \\ &= \int_{L_{p,\phi}} J\Phi(x)\, f(\Phi(x))\, dx = \\ &= R[\, J\Phi(\cdot)\, f(\Phi(\cdot))\,](p,\phi). \end{aligned} \tag{3.1}$$

Where $J\Phi(x)$ is the Jacobian of Φ in x.

If we use the inversion formula for the transform Rf, (§ 2), then we find an inversion formula for Qf. Now, take $g(x) := J\Phi(x)\, f(\Phi(x))$, in formula (2.6), then we obtain.

$$f(\Phi(x)) = \frac{1}{4\pi^2} \left[J\Phi_{|x} \right]^{-1} \int_0^{2\pi} \int_{-\infty}^{\infty} \frac{\frac{\partial}{\partial p} Qf(p',\phi')}{x \cdot \theta - p} dp d\phi, \tag{3.2}$$

substituting $y = \Phi(x)$, we obtain

$$f(y) = \frac{1}{4\pi^2} \left[J\Phi_{|\Phi^{-1}(y)} \right]^{-1} \int_0^{2\pi} \int_{-\infty}^{\infty} \frac{\frac{\partial}{\partial p} Qf(p',\phi')}{\Phi^{-1}(y) \cdot \theta - p} dp d\phi. \tag{3.3}$$

Here p' and ϕ' are functions of p and ϕ respectively.

4. A formula for Φ

We want a function $\Phi: \mathbb{R}^2 \to \mathbb{R}^2$ that maps the family of lines $L_{p,\phi}$, $p \in (-\infty, \infty)$, $\phi \in [0, 2\pi)$ onto the family of circles $C_{p',\phi'}$, $p' \in (-\infty, \infty)$ $\phi' \in [0, 2\pi)$. A known type of transformation, that maps lines onto circles and vice versa, is the Möbius transformation. Consider the mapping $M: \mathbf{C}_\infty \to \mathbf{C}_\infty$,

$$M(z) = \frac{1}{z}.$$

($\mathbf{C}_\infty$ is the extended complex plane $\mathbf{C} \cup \{\infty\}$.) M maps a line, not through the

origin, onto a circle through the origin and a line through the origin onto a line through the origin (i.e. a circle with radius ∞). M is a conformal mapping, so it preserves angles. Let S be the point of intersection of the line $L_{p,\phi}$ and of the line L, going through the origin, orthogonal to $L_{p,\phi}$ (fig. 2). The line through the origin at $M(S)$ intersects the circle $M(L_{p,\phi})$ orthogonally in the point $M(S)$, because the line L and $L_{p,\phi}$ are orthogonal in S. So the segment $[O,M(S)]$ determines the circle $M(L_{p,\phi})$, in the sense that we can compute p' and ϕ' if we know $M(S)$. (see fig. 2). The coordinates of S follow directly from the definition: $S=(p\cos\phi, p\sin\phi)$. Write S as a complex number, $S=pe^{i\phi}$, then $M(S)=\dfrac{e^{-i\phi}}{p}$. So $\phi'=-\phi$ and $p'=1/p$. (fig. 2)

If we write M in real coordinates, we obtain the desired mapping $\Phi:\mathbb{R}^2\to\mathbb{R}^2$,

$$\Phi(x_1,x_2) = \left[\frac{x_1}{x_1^2+x_2^2}, \frac{-x_2^2}{x_1^2+x_2^2}\right]^T. \tag{4.1}$$

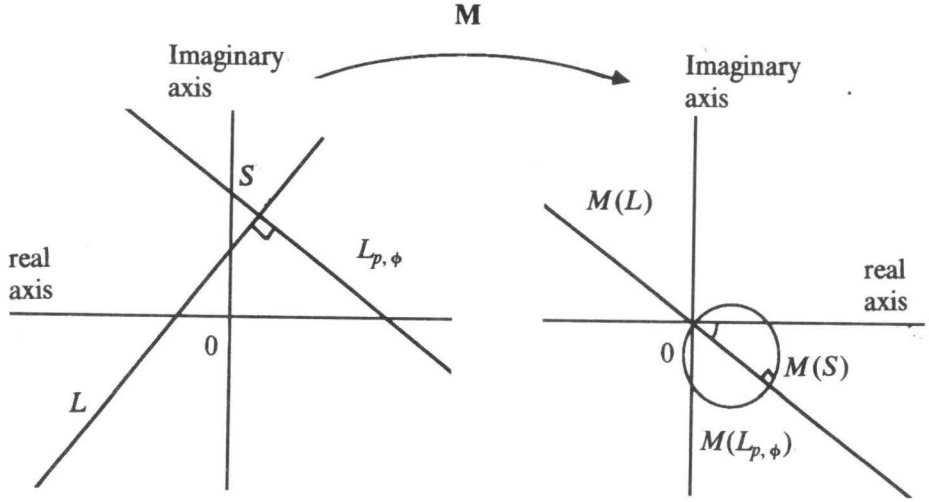


FIGURE 2 $L_{p,\phi}$ and L are orthogonal

The Jacobian is

$$J\Phi(x) = \frac{1}{\|x\|^2}. \tag{4.2}$$

With the relations $\phi'=-\phi$ and $p'=1/p$, we have that the family of lines $\{L_{p,\phi}\}_{p\in\mathbb{R},\phi\in[0,2\pi]}$ is mapped onto the family of circles $\{C_{p',\phi'}\}_{p'\in\mathbb{R},\phi'\in[0,2\pi]}$.

5. AN INVERSION FORMULA FOR Q

We have formula (3.3)

$$f(y) = \frac{1}{4\pi^2 J\Phi_{|\theta\cdot\Phi^{-1}(y)}} \int_0^{2\pi}\int_{-\infty}^{\infty} \frac{\dfrac{\partial}{\partial p}Qf(p',\phi')}{\theta\cdot\Phi^{-1}(y)-p}\,dp\,d\phi$$

Using the transformation of variables

$$p' = \frac{1}{p} \text{ and } \phi' = -\phi, \text{ so } dp = -\frac{1}{p'^2}dp'$$

and $\frac{\partial}{\partial p}Qf(p',\phi') = -p'^2\frac{\partial}{\partial p'}Qf(p',\phi')$ we obtain

$$f(y) = \frac{1}{4\pi^2 J\Phi_{|_{\Phi^{-1}(y)}}} \int_0^{2\pi}\int_{-\infty}^{\infty} \frac{\frac{\partial}{\partial p'}Qf(p',\phi')}{1/p' - \theta'\cdot\Phi^{-1}(y)}dp'd\phi'. \tag{5.1}$$

Using (4.1), (4.2) and $\Phi^{-1}(y)=\Phi(y)$. We have

$$f(y_1,y_2) = \frac{1}{4\pi^2}\int_0^{2\pi}\int_{-\infty}^{\infty} \frac{\frac{\partial}{\partial p'}Qf(p',\phi')}{(1/p')(y_1^2+y_2^2)-(y_1\cos\phi'+y_2\sin\phi')}dp'd\phi' \tag{5.2}$$

or with an obvious change of notation

$$f(y) = \frac{1}{4\pi^2}\int_{S^1}\int_{\mathbf{R}} \frac{\frac{\partial}{\partial p'}Qf(p',\theta')}{(1/p')\|y\|^2 - y\cdot\theta'}dp'(d\theta');\ \theta' = \begin{bmatrix}\cos\phi'\\ \sin\phi'\end{bmatrix}. \tag{5.3}$$

6. The domain of Q

In § 3 we inverted the Radon transform $f\to Rf$ of a function g of the form

$$g(x) = J\Phi(x)\, f(\Phi(x)).$$

If g is a continuous function with compact support, then the Radon inversion formula holds. So we want to find a region G such that g is continuous on G with support in G. Define the annulus around m by $A(m,r_1,r_2) := B(m,r_2)\setminus\overline{B(m,r_1)}$, $r_1<r_2$. We know that $\Phi[A(0,\epsilon,1)] = A(0,1,1/\epsilon)$, because $M[A(0,\epsilon,1)]=A(0,1,1/\epsilon)$. So if f has support in $A(0,\epsilon,1)$, then g has support in $A(0,\epsilon,1/\epsilon)$, and vice versa. Thus if f is continuous with support in $G := A(0,1,R)$, then g is continuous with support in $A(0,1/R,1)$ and the Radon inversion formula can be applied.

7. The Ludwig-Helgason conditions

It is not only important to have an inversion formula for a Radon transform, but also to know some of its properties. For example, for $f\to Rf$ there are the so-called Ludwig-Helgason conditions for a function h to satisfy $h(p,\theta)=Rg(p,\theta)$ for some g and $(p,\theta)\in\mathbf{R}\times S^1$. In this section we formulate analogous conditions for the transform $f\to Qf$.

Recall the symmetry property for the Radon transform $f\to Rf$.

$$Rf(p,\theta) = Rf(-p,-\theta). \tag{7.1}$$

This property together with the homogeneity property

$$\int_{-\infty}^{\infty} p^k Rg(p,\theta)dp = \pi_k(\theta), \quad \theta \in S^1, \tag{7.2}$$

are called the Ludwig-Helgason conditions, where $\pi_k(\theta)$ is a homogeneous polynomial of degree k in θ. The conditions (7.1) and (7.2) are properties of functions h in the range of R,

$$\text{Im } R := \{h : \mathbf{R} \times S^1 \to \mathbf{R} \mid \exists g : \mathbf{R}^2 \to \mathbf{R}, \quad Rg = h\}.$$

These conditions are of practical importance in e.g. computerized tomography, where the measurements $h(p,\theta) = Rg(p,\theta)$, which may be corrupted by noise, or incomplete, should satisfy (7.1) and (7.2). For $f \to Qf$ we can find analogous formulas, for continuous functions with support in G. We already have the symmetry property (2.5), with a change of notation,

$$Qf(p',\theta') = Qf(-p',-\theta'). \tag{7.3}$$

By substituting $g(x) := J\Phi(x)\, f(\Phi(x))$ in formula (7.2) we obtain

$$\int_{-\infty}^{\infty} p^k Qf(p',\theta')dp = \pi_k(\theta). \tag{7.4}$$

With $p = 1/p'$,

$$\int_{-\infty}^{\infty} \left(\frac{1}{p'}\right)^{k+2} Qf(p',\theta')dp' = \pi_k(\theta'), \tag{7.5}$$

and $\pi_k(\theta')$ is a homogeneous polynomial of degree k. Formula (7.5) is the analogue of (7.2).

8. Summary and Remarks

We have obtained an inversion formula for the transform $f \to Qf$ (formula 5.3),

$$f(y) = \frac{1}{4\pi^2} \int_{S^1} \int_{\mathbf{R}} \frac{\frac{\partial}{\partial p'} Qf(p',\theta')}{(1/p')\|y\|^2 - y \cdot \theta'} \, dp'(d\theta'),$$

if f is a continuous function with support in the annulus $G := A(0,1,R)$. We found the Ludwig-Helgason conditions for this transform, (7.3) and (7.5)

$$Qf(p',\theta') = Qf(-p',-\theta'),$$

$$\int_{-\infty}^{\infty} \frac{1}{(p')^{k+2}} Qf(p',\theta')dp' = \pi_k(\theta'),$$

where $\pi_k(\theta')$ is a homogeneous polynomial of degree k. The inversion formula, obtained here, only holds for functions $f : \mathbf{R}^2 \to \mathbf{R}$. For the higher dimensional case an inversion formula is derived by Cormack and Quinto [5]. They consider the components $f_l(p)$ of a function $f \in C^\infty(\mathbf{R}^n)$, given by $f_l(p) := \int_{S^{n-1}} f(p\omega) Y_l(\omega) dS(\omega)$. Here $p \in \mathbf{R}, \omega \in S^{n-1}$ and Y_l is a normalized spherical harmonic. Cormack and Quinto obtain an inversion formula for each

component $f_l(p)$.

For the special case $n=2$, CORMACK ([6]) finds an inversion formula for integral transforms of a function over a whole class of families of curves in the plane the so called β-curves,

$$p^\beta \cos\{\beta(\theta-\phi)\} = r^\beta;\ \beta>0\ \ |\theta-\phi|<\frac{\pi}{2\beta}. \tag{8.1}$$

If we fix $\beta=1$, then (8.1) describes a family of circles through the origin, as defined here in § 2.

The cases $\beta=½$ and $\beta=2$ correspond to a family of cardioids and a family of one-branched lemniscates respectively. Cormack's inversion formula is $(m=1/\beta)$

$$f(p,\phi) = \frac{1}{2\pi^2 t} \int_0^{2\pi} \int_0^{\infty} \frac{\frac{\partial}{\partial p'} Qf(p',\phi')\ U_{m-1}(t/p')}{T_m(t/p') - \cos(\phi-\phi')} dp'\, d\phi'.$$

Here $T_k(x)$ and $U_k(x)$ are the Tschebycheff polynomials of the first and the second kind, respectively. If $\beta=1$ $(m=1)$, then $(U_0(x)\equiv 1,\ T_1(x)=x)$

$$f(p,\phi) = \frac{1}{2\pi^2 t} \int_0^{2\pi} \int_0^{\infty} \frac{\frac{\partial}{\partial p'} Qf(p',\phi')}{t/p' - \cos(\phi-\phi')} dp' d\phi'.$$

Substituting $t = \|y\|$ and $y = \begin{bmatrix} \|y\| \cos\phi \\ \|y\| \sin\phi \end{bmatrix}$ we obtain formula (5.3). So the result obtained here is a special case of the formula in [6], found by Cormack.

REFERENCES

[1] I.M. GEL'FAND, M.I. GRAEV, N.YA. VILENKIN (1966). *Generalized functions vol 5:* Integral Geometry and Representation Theory. Academic Press, New York.

[2] S. HELGASON (1980). *The Radon transform.* Birkhäuser, Boston.

[3] F. NATTERER (1986). *The Mathematics of Computerized Tomography* John Wiley & Sons, Chichester, New York.

[4] Y. NIEVERGELT. Elementary inversion of Radon's Transform. *SIAM Review vol 28 no. 1,* March 1986.

[5] A.M. CORMACK and E.T. QUINTO. A Radon transform on spheres through the origin in $\mathbb{R}^n$ and applications to the Darboux equation. *Transactions of the A.M.S. vol 260 no. 2,* August 1980, pp. 575-581.

[6] A.M. CORMACK. Radon's problem, Old and New. *SIAM, AMS proceedings, vol. 14,* Inverse problems April 1983, pp. 33-43.

[7] F. JOHN. *Plane Waves and Spherical means with applications to P.D.E.* Interscience, New York, 1955.

[8] N.M. TEMME. The Radon transform; First steps. *C.W.I. Newsletter, No. 15,* june 1987.

A short introduction to exit problems

H. Roozen
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

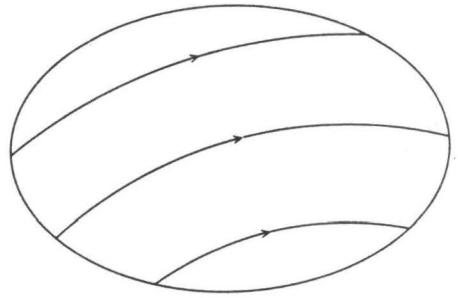
Many phenomena that occur in nature and technology exhibit a stochastic behaviour. When the stochastic element is relevant, it has to be included in the modeling of such phenomena. We shall discuss models with a deterministic component and a small stochastic component. The short term behaviour of these models is determined mainly by the deterministic component, while the long term behaviour is influenced considerably by the stochastic component. For the description of the long term behaviour, deterministic stability concepts (stable, neutral equilibrium) are inadequate and have to be replaced by stochastic stability concepts (the expected exit time from a region containing such a deterministic equilibrium). We will consider these stochastic models from an asymptotic (in the limit for small noise) point of view.

1. Introduction

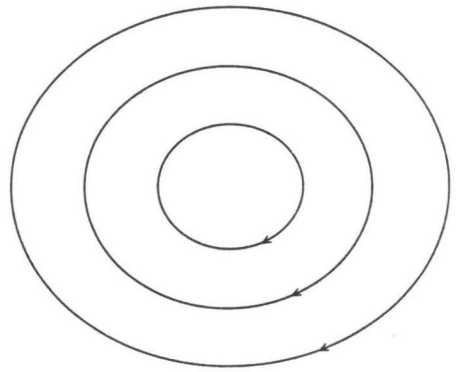
In this contribution we will study some aspects of stochastic dynamical systems that have a deterministic part (referred to as 'the deterministic system') and a small stochastic part consisting of Gaussian white noise (referred to as 'stochastic fluctuations').

In some of these systems, the dynamical characteristics of interest are dominated by the deterministic system, while the stochastic fluctuations are only of secondary importance, in the sense that the omission of the stochastic fluctuations does not essentially alter these characteristics. This is demonstrated, for example, by a 'diffusion *with* the flow', see figure 1a. Starting at a point in a bounded domain D, the trajectories of the stochastic dynamical system leave the domain D with probability close to one in the same time as the deterministic trajectory through that point. The probability density function defined on the boundary ∂D, describing the point of exit from D of the stochastic dynamical system, is concentrated near the deterministic exit point. Stochastic systems of this type will not be considered here.

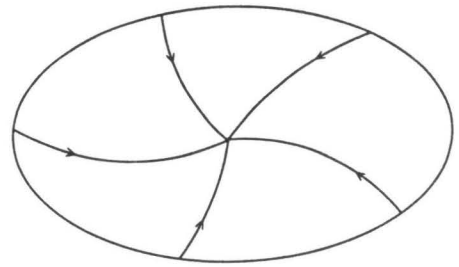
In other stochastic dynamical systems, the stochastic fluctuations, though small compared to the deterministic system, are of such importance to the dynamical characteristics of interest, that these are essentially changed on omission of the stochastic fluctuations. One such example is a 'diffusion *across* the flow', as depicted in figure 1b. The deterministic system consists of a center point, surrounded by closed trajectories. Consider a domain D, enclosed by one of these trajectories. In the deterministic system no exit from D can occur, since we follow ceaselessly the closed trajectory through the starting point. In contrast with this fact, in the stochastic dynamical



a



b



c

FIGURE 1. Illustration of diffusion (a) with, (b) across and (c) against the flow.

system, i.e. in the deterministic system perturbed by stochastic fluctuations, exit will occur with probability one in finite time. Another such example is a

'diffusion *against* the flow', depicted in figure 1c. A bounded domain D is entered at its boundary ∂D by deterministic trajectories that converge to an asymptotically stable limit point contained in D. In this deterministic system, if we start at some point in D, we approach the limit point along the trajectory through the starting point. Again, the deterministic system does not allow exit from D, but when this system is perturbed by stochastic fluctuations, exit will happen in finite time with probability one. Although more complicated systems exist that exhibit a similar behaviour, such as attracting limit cycles or strange attractors, etc., notably in higher dimensional domains, we will confine ourselves to systems of the two simple types described here, in particular to systems of the last type.

We will concentrate on a few statistical characteristics related to the problem of exit from a domain, like the expectation value of the time of first exit (which provides a measure for the stability of the stochastic system) and the distribution of exit points over the boundary of the domain.

2. The equations

In this section we shall given the formulation of the forward and the backward (Fokker-Planck or Kolmogorov) [19,55] equations, which form the starting point of our analysis. We consider a stochastic dynamical system that has been defined on the n-dimensional domain D in the state space. Let $v(x,t)dx$ denote the probability that the system is in the infinitesimal n-dimensional subregion $(x,x+dx)\in D$ on time t. This function satisfies the forward equation

$$\frac{\partial v}{\partial t} = M_\epsilon v, \quad x \in D, \tag{1}$$

where M_ϵ is the differential operator:

$$M_\epsilon v \equiv -\sum_{i=1}^{n} \frac{\partial}{\partial x_i}(b_i(x)v) + \frac{\epsilon}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2}{\partial x_i \partial x_j}(a_{ij}(x)v). \tag{2}$$

Equation (1) has to be supplemented with the relevant initial and boundary conditions. The first term on the right side of (2) represents the deterministic part of the dynamical system, b is called the deterministic or drift vector. The second term on the right side represents the stochastic fluctuations. The symmetric positive (semi-)definite matrix a is known as the diffusion matrix. The parameter ϵ, $0<\epsilon\ll 1$, indicates that the stochastic fluctuations are small relative to the deterministic part. When $\epsilon=0$ the stochastic fluctuations are absent and equation (1) reduces to the Liouville equation. Then, if the initial position is deterministic, the initial probability density is a delta function, say $\delta(x-x_0)$, and the solution of the Liouville equation corresponds to the solution of the system of ordinary differential equations

$$\frac{dx_i}{dt} = b_i(x), \quad i=1,2,...,n \tag{3a}$$

with initial conditions

$$x(0) = x_0. \tag{3b}$$

This system is defined as the deterministic system corresponding to the stochastic dynamical system.

In order to determine the distribution of exit points over ∂D, as well as the expected time of first exit from D, we will use backward equations. Let $p(x,y)\,dS_y$ be the probability of exit at $dS_y \in \partial D$, given that we started at $x \in D$ on time $t=0$, i.e. p is the exit density. We define the function $u_s(x)$ as follows:

$$u_s(x) = \int_{\partial D} f(y)p(x,y)\,dS_y, \tag{4}$$

where f is a function on ∂D that can be chosen arbitrarily. With f defined as the indicator function

$$f = \begin{cases} 1 & \text{on } \partial_1 D, \text{ where } \partial_1 D \subseteq \partial D, \\ 0 & \text{on } \partial_0 D = \partial D \setminus \partial_1 D, \end{cases} \tag{5}$$

$u_s(x)$ is the probability of exit at $\partial_1 D$, given that we started at $x \in D$ on time $t=0$. The function u_s is the solution of the stationary backward equation

$$L_\epsilon u_s = 0, \quad x \in D, \tag{6a}$$

subject to the boundary condition

$$u_s = f(x), \quad x \in \partial D, \tag{6b}$$

where L_ϵ is the differential operator

$$L_\epsilon u \equiv \sum_{i=1}^{n} b_i(x)\frac{\partial u}{\partial x_i} + \frac{\epsilon}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}(x)\frac{\partial^2 u}{\partial x_i \partial x_j}, \tag{7}$$

and a and b are the same functions as above.

We will consider the time-dependent backward equation

$$\frac{\partial u}{\partial t} = L_\epsilon u \tag{8a}$$

as well. With the boundary condition

$$u = f(x), \quad x \in \partial D, \tag{8b}$$

where f is the indicator function (5), and the initial condition

$$u(x,0) = 0, \quad x \in D, \tag{8c}$$

$u(x,t)$ is the probability that exit occurs at $\partial_1 D$ on the time interval $(0,t]$, given that we started at $x \in D$ on time $t=0$.

Let $T(x)$ be the expected time of first exit from D, given that we started at x on time $t=0$:

$$T(x) = \inf\{t | x(t) \in \partial D, \ x(0) = x \in D\}. \tag{9}$$

The function T is the solution of the boundary value problem:

$$L_\epsilon T = -1, \quad x \in D, \tag{10a}$$

$$T = 0, \quad x \in \partial D. \tag{10b}$$

Equation (10a) is known as the Dynkin equation.

The reader interested in the details of the equations and the corresponding conditions that we have given here, and related material, is referred to the literature [19,55]. In later sections we shall be concerned with the asymptotic solution of (1), (6), (8) and (10) for small ϵ.

The backward and forward differential operators L_ϵ and M_ϵ defined above are formal adjoints, which means that the following relation holds [48]:

$$\int_D \int (vL_\epsilon u - uM_\epsilon v)\, dx = \int_{\partial D} P \cdot \xi \; dS_x, \tag{11}$$

where P is the vector with components

$$P_i = \sum_{j=1}^{n} \frac{\epsilon}{2} \left[a_{ij} v \frac{\partial u}{\partial x_j} - u \frac{\partial (a_{ij} v)}{\partial x_j} \right] + b_i uv, \quad i = 1,2,\ldots,n \tag{12}$$

and ξ denotes the outward normal on ∂D.

3. The boundary

In the study of exit problems the behaviour of the stochastic system at and near the boundary of the domain deserves special attention, since the domain is left via the boundary. For a given stochastic system we must verify whether the boundary can actually be reached from the interior domain.

In many practical situations the type of the boundary is determined by the drift vector and the diffusion matrix. For one-dimensional stochastic systems there is a classification of such boundaries originating from Feller [16]. In a semi-group approach to adjoint forward and backward equations he distinguished the regular, exit, entrance and natural boundaries. In table 1 we have repeated schematically the boundary classification as it has been described in [54]. The type of boundary depends on the integrability at the boundary point of some of the following integrals:

$$\begin{aligned} I_1(x) &= \exp\left[-\frac{4}{\epsilon} \int^x b(t)/a(t)\, dt \right], \\ I_2(x) &= \frac{2}{\epsilon a(x)} \exp\left[\frac{4}{\epsilon} \int^x b(t)/a(t)\, dt \right], \\ I_3(x) &= I_1(x) \int^x I_2(t)\, dt, \\ I_4(x) &= I_2(x) \int^x I_1(x)\, dt. \end{aligned} \tag{1}$$

If there are sample paths that hit the boundary in finite time, the boundary is attainable, otherwise it is unattainable. From table 1 we infer that it makes sense to talk of an exit problem only if at least one of the boundaries of the domain (an interval) is a regular or exit boundary (which are the only cases

that permit the boundary to be reached in finite time from the interior domain). When we have a regular boundary we should speak of the problem of *first* exit from a domain, since in that case the exited domain can be re-entered and subsequently re-exited. For higher dimensional stochastic systems a similar classification has never been published.

integrable				type of boundary	boundary is attainable	interior is attainable
I_1	I_2	I_3	I_4			
yes	yes			regular	yes	yes
yes	no	yes		exit	yes	no
yes	no	no		natural attracting	no	no
no	yes		yes	entrance	no	yes
no	yes		no	natural repelling	no	no
no	no					

TABLE 1. Boundary classification for one-dimensional stochastic systems. Five boundary types are defined according to the integrability of some of the integrals I_1 to I_4. The last two columns indicate whether the boundary is attainable from the interior domain and whether the interior domain is attainable from the boundary.

In other situations we dispose of a stochastic system, defined by a drift vector and a diffusion matrix on a domain D, and we want to erect a boundary of a desired type at any place in D, thereby restricting the domain to a subdomain D' of D. Examples of such boundaries are absorbing and reflecting boundaries [19]. On reaching an absorbing boundary from the interior domain D', the system is taken apart (is absorbed) so that this domain cannot be entered again, comparable with an exit boundary. A reflecting boundary does not let through probability, so that no exit is possible at this boundary. With respect to the solution of the forward equation an absorbing boundary implies the boundary condition $v(x,t)=0$, where $x\in\partial D'$, and a reflecting boundary implies the condition $\xi\cdot J(x,t)=0$, where $x\in\partial D'$, ξ is the outward normal on $\partial D'$ and J is the probability current, i.e. the vector with components

$$J_i(x,t) = b_i(x)v - \frac{\epsilon}{2}\sum_{j=1}^{n}\frac{\partial}{\partial x_j}(a_{ij}(x)v), \quad i=1,2,\ldots,n. \tag{2}$$

Absorbing and reflecting boundaries can be set up in domains of any dimension.

4. AN EXAMPLE OF DIFFUSION ACROSS THE FLOW

In this section we will treat a simple example of diffusion across the flow. Consider an oscillator with a small damping, that is subjected to a stochastic forcing. The damping effect is introduced here since it leads to realistic model, without essential complications for the analysis that follows. As a consequence

of the stochastic effects, the energy of the oscillator can reach a critical level after some time. This critical level is chosen as the energy at which the oscillator breaks down. We will derive an expression for the expected time needed to reach the critical energy level, which is a measure for the stochastic stability of the oscillator. In non-dimensional form, the stochastic differential equation [1] for this problem is [53]:

$$\ddot{x} + \epsilon\alpha\dot{x} + x = \sqrt{\epsilon}q(x)\xi, \tag{1}$$

where x is the deviation from the equilibrium position, the dot denotes differentiation with respect to the time t, ϵ is a small positive parameter, and $\epsilon\alpha$ is a nonnegative $O(\epsilon)$ damping constant. The right side of (1) represents a Gaussian white noise process with intensity $\epsilon q^2(x)$. The function q is approximated by the first two terms of its Taylor expansion around $x=0$:

$$q(x)\approx\beta_0 + \beta_1 x. \tag{2}$$

The second order differential equation (1) can be written as the system of first order equations:

$$\begin{aligned} \frac{dx}{dt} &= \dot{x}, \\ \frac{d\dot{x}}{dt} &= -(\epsilon\alpha\dot{x}+x) + \sqrt{\epsilon}(\beta_0 + \beta_1 x)\xi. \end{aligned} \tag{3}$$

The undisturbed ($\epsilon=0$) system (1) is an undamped oscillator, whose dynamics are described by closed trajectories around the origin in the $(x,\dot{x})$-phase space. Each trajectory corresponds to an enery level. The energy is larger for orbits farther away from the origin. The effect of a nonzero ϵ is that the trajectories tend to spiral slightly inwards to approach the origin as a consequence of damping if $\alpha\neq0$ and contain stochastic fluctuations in the $\dot{x}$-direction. The backward equation corresponding to (3) reads [19]:

$$\frac{\partial u}{\partial t} = \dot{x}\frac{\partial u}{\partial x} - (\epsilon\alpha\dot{x}+x)\frac{\partial u}{\partial \dot{x}} + \frac{\epsilon}{2}(\beta_0+\beta_1 x)^2\frac{\partial^2 u}{\partial \dot{x}^2}, \tag{4}$$

with u defined as in section 2. This equation is studied asymptotically for small ϵ and on the time scale of $O(\epsilon^{-1})$. With

$$t = \tilde{t}/\epsilon, \quad u = u^0 + \epsilon u^1 + \cdots, \tag{5}$$

and the transformations $(x,\dot{x})\rightarrow(r,\theta)$ defined by

$$x = \sqrt{2}r\cos\theta, \quad \dot{x} = \sqrt{2}r\sin\theta, \tag{6}$$

we obtain to leading order in ϵ:

$$\frac{\partial u^0}{\partial \theta} = 0, \tag{7}$$

implying that u^0 is a function of r and $\tilde{t}$ only. The variable r is the square root of the dimensionless energy of the undisturbed ($\epsilon=0$) system. The

nondimensionalization process can be carried out such that the critical energy corresponds to $r^2=1$, thus $r\in[0,1]$. To the next order in ϵ we obtain an equation in terms of u^0 and u^1. Terms with u^1 vanish by integration of this equation with respect to θ from 0 to 2π and the additional assumption that u^1 is periodic in θ with period 2π. The resulting equation for u^0 reads:

$$\frac{\partial u^0}{\partial \tilde{t}} = (\frac{a_0}{r} + a_2 r)\frac{\partial u^0}{\partial r} + (a_0 + a_1 r^2)\frac{\partial^2 u^0}{\partial r^2}, \tag{8a}$$

with

$$a_0 = \beta_0^2/8, \quad a_1 = \beta_1^2/16, \quad \alpha_2 = 3\beta_1^2/16 - \alpha/2. \tag{8b}$$

The description to this order in ϵ includes the effects of damping and stochastic fluctuations. If, as a consequence of the latter effect the critical energy $r^2=1$ is reached in finite time with probability one, starting from $r\in[0,1]$, the oscillator is said to be stochastically unstable. In that case, the stability of the oscillator is measured by the expected time of exit from the unit interval at 1.

In the present discussion we only consider the case $|\beta_0|, |\beta_1| \gg O(\epsilon^{1/2})$, so that a_0 and a_1 do not vanish in the asymptotics leading to equation (8a) and thus appear in this equation indeed. The boundary $r=0$ is then an entrance boundary and at $r=1$ we adopt an absorbing boundary in order to model the breakdown of the oscillator at the critical energy. Thus exit from the unit interval can take place only at $r=1$. Let $u_s(r)$ be the probability of exit at $r=1$, given that we started at r on time $t=0$. The leading order term $u_s^0(r)$ in the expansion of $u_s(r)$ in powers of ϵ is obtained by solving the stationary equation (8a) with boundary condition $u_s^0(1)=1$. The only relevant solution (i.e. giving values $u_s^0(r)\in[0,1]$) is $u_s^0(r)\equiv 1$. There is no freedom to specify an arbitrary boundary condition at $r=0$. We conclude that if we start somewhere on the interval $[0,1]$, exit at $r=1$ will occur with probability one, so that the oscillator is stochastically unstable. Next we consider the expected exit time $T(r)$, starting from a point r. Similar to the time scaling in (5) we put $T=\tilde{T}/\epsilon$ and similar to the expansion of u in (5) we put $\tilde{T}=\tilde{T}^0+\epsilon\tilde{T}^1+\cdots$, so that $T=\tilde{T}^0/\epsilon+\tilde{T}^1+\cdots$. An approximation for T is found by solving the Dynkin equation

$$-1 = (\frac{a_0}{r} + a_2 r)\frac{\partial \tilde{T}^0}{\partial r} + (a_0 + a_1 r^2)\frac{\partial^2 \tilde{T}^0}{\partial r^2}, \tag{9a}$$

with the boundary conditions:

$$\tilde{T}^0(0) \text{ is finite}, \tag{9b}$$

$$\tilde{T}^0(1) = 0. \tag{9c}$$

For $a_2 \neq a_1$ we find:

$$T(r) \sim \frac{1}{\epsilon(a_1 - a_2)} \int_r^1 \left[\left(\frac{a_1}{a_0} s^2 + 1 \right)^{-\frac{a_2}{2a_1} + \frac{1}{2}} - 1 \right] \frac{1}{s} ds. \tag{10}$$

If $a_2 = a_1$, this is substituted into equation (9a). Solving the corresponding boundary value problem we find:

$$T(r) \sim \frac{1}{2\epsilon a_1} \int_r^1 \frac{1}{s} \log\left[\frac{a_1}{a_0} s^2 + 1\right] ds. \tag{11}$$

The reader is asked to take notice of the order of magnitude of the results (10) and (11) in order to compare this with results to be derived later for diffusion against the flow systems. The cases that either β_0 or β_1 are of order $O(\epsilon^{1/2})$ must be treated separately. In the case $\beta_0 = O(\epsilon^{1/2})$ it can be shown that if the damping is larger than a certain value the oscillator is stochastically stable on the time scale under consideration. This means that on this time scale the probability of exit is less than one, in contrast with the result above. A more detailed description of the exit problem for oscillators as described here can be found in [53]. The stochastic stability of oscillators with a different type of damping (as cubic damping) or noise (red, dichotomic, etc.) and with a forcing described by a potential function has been treated in [14]. The asymptotics that we have used in this example to arrive at equation (8) are well established and are known under the names of averaging technique [3,31,35,50,56] and adiabatic elimination of fast variables [19].

5. Diffusion against the flow

In this section we will discuss the exit problem for systems that are of diffusion against the flow type. First we will treat a one-dimensional system, then a multidimensional potential system that can be treated with essentially the same means, and we will conclude with more general multidimensional systems.

5.1. A one-dimensional system

Consider the stochastic system defined on $[\alpha,\beta]$, where $\alpha<0$ and $\beta>0$, with drift coefficient $b(x)$ satisfying

$$b(x)\begin{cases} > 0, & x\in[\alpha,0), \\ = 0, & x = 0, \\ < 0, & x\in(0,\beta], \end{cases} \tag{1a}$$

so that $x=0$ is an attractor, and diffusion coefficient $\epsilon a(x)/2$, $0<\epsilon\ll 1$, with

$$a(x)>0, \quad x\in[\alpha,\beta]. \tag{1b}$$

For this system we will find the functions u_s and T defined in section 2, asymptotically for small ϵ. The boundary value problem for u_s reads:

$$L_\epsilon u_s \equiv \frac{\epsilon}{2} a(x) \frac{d^2 u_s}{dx^2} + b(x) \frac{du_s}{dx} = 0, \tag{2a}$$

$$u_s(\alpha) = c_\alpha, \quad u_s(\beta) = c_\beta, \tag{2b}$$

where c_α and c_β are given constants. The reduced equation (2a), i.e. equation (2a) with $\epsilon=0$, is solved by any constant c_0. This solution is valid way from α

and β but not near these points since the boundary conditions (2b) cannot be satisfied. We assume that the functions a and b have the Taylor series expansions

$$\begin{aligned} a(x) &= a(\alpha) + a'(\alpha)(x-\alpha) + \cdots \quad \text{near } x = \alpha, \\ b(x) &= b(\alpha) + b'(\alpha)(x-\alpha) + \cdots \quad \text{near } x = \alpha, \\ a(x) &= a(\beta) + a'(\beta)(x-\beta) + \cdots \quad \text{near } x = \beta, \\ b(x) &= b(\beta) + b'(\beta)(x-\beta) + \cdots \quad \text{near } x = \beta. \end{aligned} \tag{3}$$

As an abbreviation we use the notation $\tilde{b}(x)=2b(x)/a(x)$. it follows from (1) that $\tilde{b}(\alpha)>0$ and $\tilde{b}(\beta)<0$. A boundary layer analysis near $x=\alpha$ and $x=\beta$ shows the presence of $O(\epsilon)$ boundary layers near these points. An asymptotic expression for u_s to leading order in ϵ that is uniformly valid on $[\alpha,\beta]$ is given by

$$\begin{aligned} u_s(x) \sim c_0 &+ (c_\alpha - c_0)\exp[-\tilde{b}(\alpha)(x-\alpha)/\epsilon] \\ &+ (c_\beta - c_0)\exp[-\tilde{b}(\beta)(x-\beta)/\epsilon]. \end{aligned} \tag{4}$$

Note that the constant c_0 is left undetermined by the given asymptotics. To find c_0 we utilize a variational formulation of the boundary value problem (2), following [23], see also [61]. After multiplication by the factor

$$g(x) = \exp\left[\int_0^x \frac{2b(s)-\epsilon a'(s)}{\epsilon a(s)}\,ds\right], \tag{5}$$

equation (2a) can be written as the Euler equation

$$\frac{dF_{u'_s}}{dx} - F_{u_s} = 0, \tag{6}$$

with $F=\frac{\epsilon}{4}(u'_s)^2 ag$. Consequently, the solution of (2) corresponds to an extremal of the functional

$$J[u_s] = \int_\alpha^\beta \frac{\epsilon}{4}(u'_s)^2 ag\,dx, \tag{7}$$

with respect to functions u_s satisfying the boundary conditions (2b), see [5,9,46,47]. The expression (4) for u_s is substituted into the integral in (7), and this integral is evaluated asymptotically for small ϵ by the method of Laplace [2,4]. The constant c_0 is determined by the requirement that the corresponding function u_s is an extremal of the functional J thus obtained, that is, by

$$\frac{dJ}{dc_0} = 0. \tag{8}$$

In addition to (1) we shall henceforth assume that

$$\tilde{b}'(x)<0, \quad x\in[\alpha,\ \beta]. \tag{9}$$

Carrying out the above procedure we then find that the largest contributions to

the integral in (7) are from the neighbourhoods of α and β, and c_0 is given by

$$c_0 = \frac{c_\alpha \tilde{b}(\alpha)\exp[-I(\alpha)/\epsilon] - c_\beta \tilde{b}(\beta)\exp[-I(\beta)/\epsilon]}{\tilde{b}(\alpha)\exp[-I(\alpha)/\epsilon] - \tilde{b}(\beta)\exp[-I(\beta)/\epsilon]}, \tag{10}$$

where $I(x)$ is defined as:

$$I(x) = -\int_0^x \tilde{b}(s)\,ds, \quad (>0 \text{ for } x \neq 0). \tag{11}$$

The result (10) simplifies to:

$$c_0 = \begin{cases} c_\alpha, & \text{if } I(\alpha) < I(\beta), \\ c_\beta, & \text{if } I(\beta) < I(\alpha), \\ \dfrac{c_\alpha \tilde{b}(\alpha) - c_\beta \tilde{b}(\beta)}{\tilde{b}(\alpha) - \tilde{b}(\beta)}, & \text{if } I(\alpha) = I(\beta). \end{cases} \tag{12}$$

Thus, in the limit $\epsilon \to 0$, if we start outside $O(\epsilon)$-neighbourhoods of the boundaries α and β, exit will occur with probability one at the boundary with the smallest value of I. If $I(\alpha) = I(\beta)$ and if we start outside $O(\epsilon)$-neighbourhoods of the boundaries, the probabilities of exit at α and β are constants with values between zero and one, depending on $\tilde{b}(\alpha)$ and $\tilde{b}(\beta)$. The asymptotic result that we have derived above is found alternatively by the evaluation for small ϵ of the exact solution of the boundary value problem (2).

Next we derive an expression for the expected time T of exit from the interval $[\alpha, \beta]$. The function T satisfies the inhomogeneous equation

$$\frac{\epsilon}{2} a(x) \frac{d^2 T}{dx^2} + b(x) \frac{dT}{dx} = -1, \tag{13a}$$

with the condtions

$$T(\alpha) = 0, \quad T(\beta) = 0. \tag{13b}$$

The approach to this boundary value problem is largely the same as above, the only additional difficulty is the appearance of the inhomogenous term in (13a). We anticipate that T is of the form

$$T(x) = c_0(\epsilon)\tau(x), \tag{14a}$$

where c_0 is a constant with respect to x that depends on ϵ in the following way

$$1/c_0(\epsilon) = o(\epsilon), \tag{14b}$$

asymptotically for small ϵ. Expression (14a) is substituted into (13) and the corresponding boundary value problem is asymptotically solved to obtain τ. For T we find:

$$T(x) \sim c_0(\epsilon) \left\{ 1 - \exp[-\tilde{b}(\alpha)(x-\alpha)/\epsilon] - \exp[-\tilde{b}(\beta)(x-\beta)/\epsilon] \right\}, \tag{15}$$

to leading order in ϵ uniformly on $[\alpha,\beta]$. The unknown constant c_0 is determined again from a variational principle. Equation (13a) is multiplied by the factor g defined in (5). The solution of the boundary value problem (13) then corresponds to an extremal of the functional

$$J[T] = \int_{\alpha}^{\beta} \left[\frac{\epsilon}{4}(T')^2 a - T \right] g \, dx \,, \tag{16}$$

with respect to functions $T(x)$ satisfying the boundary conditions (13b). This functional is evaluated by substitution of (15) into (16) and application of the method of Laplace. The major contributions to the integral in (16) are from neighbourhoods of α and β and from a neighbourhood of $x=0$. Putting (8) it is found that

$$c_0 = \frac{\dfrac{4}{a(0)} \sqrt{\dfrac{2\pi\epsilon}{-\tilde{b}'(0)}}}{\tilde{b}(\alpha)\exp[-I(\alpha)/\epsilon] - \tilde{b}(\beta)\exp[-I(\beta)/\epsilon]}. \tag{17}$$

This result simplifies to:

$$c_0 = 4\sqrt{\frac{\pi\epsilon}{-b'(0)a(0)}} \cdot \begin{cases} \dfrac{1}{\tilde{b}(\alpha)}\exp[I(\alpha)/\epsilon], & \text{if } I(\alpha)<I(\beta), \\ \dfrac{1}{-\tilde{b}(\beta)}\exp[I(\beta)/\epsilon], & \text{if } I(\beta)<I(\alpha), \\ \dfrac{1}{\tilde{b}(\alpha)-\tilde{b}\beta)}\exp[I(\beta)/\epsilon], & \text{if } I(\alpha) = I(\beta). \end{cases} \tag{18}$$

Thus, in the limit $\epsilon \to 0$, if we start outside an $O(\epsilon)$-neighbourhood of the boundaries α and β, the expected exit time equals one of the constants given in (18), depending on the magnitude of $I(\alpha)$ and $I(\beta)$. Note that in the first order asymptotics to u_s and T, the position of the starting point is of importance only if we start in $O(\epsilon)$-neighbourhoods of α and β.

Other asymptotic approaches to the type of problem we encountered in this subsection can be found in de GROEN [24], who used an eigenfunction expansion method, in JIANG FURU [26], who used the two-scale method, and in MATKOWSKY and SCHUSS [41], whose method will be explained further on. A biologically relevant model in which at one of the boundaries of the domain both the drift and diffusion coefficients vanish, linearly with the distance to this boundary, has been treated in [25,59,60].

5.2. Potential systems

The method to determine c_0 in the previous section was based on the fact that with the factor g defined in (5.1:5) the nonself-adjoint backward differential operator L_ϵ turned into a self-adjoint operator, so that consequently variational formulations of the boundary value problems of exit became feasible. In this subsection we shall see that for multidimensional stochastic systems a

similar factor g exists only for a class of so-called potential systems. Results for these systems will be derived.

We consider an n-dimensional stochastic system with a domain D that contains a deterministic point attractor and with a boundary ∂D at which the deterministic trajectories enter D. First we study the asymptotic solution of the boundary value problem (2:6). The reduced equation (2:6a) is solved by a constant c_0. It can be shown [19,55] that this solution is valid outside an $O(\epsilon)$-neighbourhood of ∂D. We assume that ∂D is smooth. For points $x \in D$ near ∂D, we introduce $n-1$ new coordinates along ∂D, and the new coordinate $\rho = |x - x'|$, where x' is the projection of x on ∂D. Using the stretching transformation

$$z = \epsilon\rho \tag{1}$$

we then obtain from (2:6a) the boundary layer equation

$$\tfrac{1}{2}\bar{a}(x')\frac{\partial^2 u_s}{\partial z^2} + \bar{b}(x')\frac{\partial u_s}{\partial z} = 0, \tag{2a}$$

with $\bar{a}$ and $\bar{b}$ defined as

$$\bar{a}(x') = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}(x')\xi_i\xi_j, \quad \bar{b}(x') = -\sum_{i=1}^{n} b_i(x')\xi_i, \tag{2b}$$

where ξ denotes the outward normal on ∂D. Equation (2a) is solved with the conditions (2:6b) and $\lim_{z\to\infty} u_s = c_0$. In the original variable x we find:

$$u_s(x) \sim c_0 + (f(x') - c_0)\exp[-\tilde{b}(x')|x - x'|/\epsilon], \tag{3}$$

uniformly on D, where $\tilde{b}(x') = 2\bar{b}(x')/\bar{a}(x')$. We intent to determine the unknown constant c_0 from a variational principle again. The backward operator L_ϵ defined in (2:7) is nonself-adjoint in general. A factor $g(x)$ is sought such that gL_ϵ is self-adjoint. This requirement leads to the following expression:

$$\epsilon\frac{\partial \log g}{\partial x_i} = \sum_{j=1}^{n} a_{ij}^{-1}\left[2b_j - \epsilon\sum_{k=1}^{n}\frac{\partial a_{jk}}{\partial x_k}\right] \equiv V_i, \quad 1,2,\ldots,n \tag{4}$$

where a_{ij}^{-1} denotes the inverse of the diffusion matrix (we assume this matrix is invertible). A function g satisfying (4) exists only if the vector field V is irrotational, that is, can be described by a potential function ϕ as follows:

$$V_i = -\frac{\partial \phi}{\partial x_i}. \tag{5}$$

Stochastic systems for which (5) holds are called potential systems. The remaining analysis in this subsection will be restricted to such systems. In order for the vector field V to be irrotational independent of the value of ϵ, we assume in addition that

$$\phi = \phi_0 + \epsilon\phi_1. \tag{6}$$

From (4),(5) and (6) it follows for g that

$$g(x) = \exp[-\phi_0(x)/\epsilon + \phi_1(x)], \tag{7a}$$

with:

$$\phi_0(x) = -\int_{x_0}^{x} \sum_{i=1}^{n} \sum_{j=1}^{n} 2a_{ij}^{-1} b_j \, dx_i, \quad \phi_1(x) = -\int_{x_0}^{x} \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}^{-1} \sum_{k=1}^{n} \frac{\partial a_{jk}}{\partial x_k} dx_i. \tag{7b}$$

The integrals in (7b) are functions of x that are independent of the path of integration. The integrals equal zero at the point x_0, which is chosen to coincide with the position of the deterministic equilibrium. Using the relationship (4) with the matrix a brought to the left side, we find that equation (2:6a) multiplied by g can be written as the Euler equation

$$-F_u + \sum_{i=1}^{n} \frac{\partial}{\partial x_i} F_{u_{x_i}} = 0, \tag{8a}$$

with F equal to

$$F = \frac{\epsilon}{4} g \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} u_{x_i} u_{x_j}. \tag{8b}$$

In these expressions we suppressed the subscript s of u for the reason of clarity. Thus, the solution of the boundary value problem (2:6) corresponds to an extremal of the functional

$$J[u_s] = \int_D F \, dx, \tag{9}$$

taken over functions u_s satisfying the boundary condition (2:6b). Expression (3) for u_s is substituted into the integral in (9), which subsequently is evaluated for small ϵ by the method of Laplace. To be definite we assume that the drift vector and the diffusion matrix are such that the major contributions to this integral come from the boundary ∂D. Setting (5.1:8) we then find:

$$c_0 = \frac{\int_{\partial D} f(y) b(y) \cdot \xi(y) \exp[-\phi_0(y)/\epsilon + \phi_1(y)] \, dS_y}{\int_{\partial D} b(y) \cdot \xi(y) \exp[-\phi_0(y)/\epsilon + \phi_1(y)] \, dS_y}. \tag{10}$$

Using the definition (2:4) of u_s we write:

$$\lim_{\epsilon \to 0} \int_{\partial D} f(y) \left[p(x,y) - \frac{b(y) \cdot \xi(y) \exp[-\phi_0(y)/\epsilon + \phi_1(y)]}{\int_{\partial D} b(y) \cdot \xi(y) \exp[-\phi_0(y)/\epsilon + \phi_1(y)] \, dS_y} \right] dS_y = 0. \tag{11}$$

This result indicates that for small ϵ the exit density p is independent of the starting point x, given $|x - x'| \gg O(\epsilon)$, and that this density is sharply peaked near the boundary point(s) with minimal potential ϕ_0. In typical situations, there is a unique y^* such that

$$\phi_0(y) > \phi_0(y^*), \quad \text{for } y \neq y^* \; y, \; y^* \in \partial D, \tag{12}$$

see figure 2. Then (11) implies that in the limit $\epsilon \to 0$ the exit density becomes:

$$p(x,y) = \delta(y - y^*), \tag{13}$$

that is, exit occurs with probability one at y^*. For cases that the minimum of ϕ_0 on ∂D is attained on a set larger than one point, the reader is referred to the literature [41].

An asymptotic expression for the expected time of exit from a region, for systems of the potential type considered above, can be derived as in subsection 5.1. This is left as an exercise for the reader.

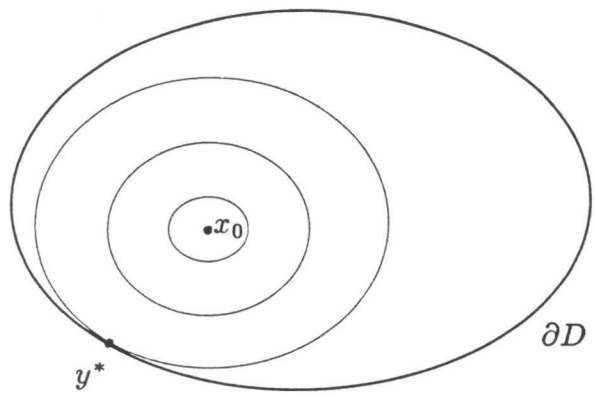


FIGURE 2. Contours on which ϕ_0 attains a constant value. This value is higher for contours farther away from x_0. The lowest value of ϕ_0 on ∂D is attained at y^*.

5.3. More general multidimensional systems

As we have seen in section 5.2, the method to determine c_0 described in section 5.1 for one-dimensional stochastic systems is applicable to multidimensional systems only if they are of a particular potential type. In the present section we give a brief outline of the approach to more general multidimensional systems, due to MATKOWSKY and SCHUSS [41].

We take over the discussion of section 5.2 until the determination of the constant c_0. The idea is to determine this constant by the employment of the relationship (2:11) between the backward operator (2:7) and the adjoint forward operator (2:2). To this aim, we first construct a solution of the stationary forward equation (2:1). This is done by means of the WKB-method [37], which assumes that this solution is of the form:

$$v(x) = w(x) \exp[-Q(x)/\epsilon], \tag{1a}$$

for small ϵ, where

$$Q(x_0) = 0, \quad w(x_0) = 1. \tag{1b}$$

The condition on w is a normalization. Substitution of (1a) into (2:1) with the left side set equal to zero yields to leading order in ϵ eikonal equation

$$\sum_{i=1}^{n} b_i \frac{\partial Q}{\partial x_i} + \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{a_{ij}}{2} \frac{\partial Q}{\partial x_i} \frac{\partial Q}{\partial x_j} = 0, \tag{2}$$

and to the next order in ϵ the transport equation

$$\sum_{i=1}^{n} \left[\sum_{j=1}^{n} a_{ij} \frac{\partial Q}{\partial x_j} + b_i \right] \frac{\partial w}{\partial x_i} + \sum_{i=1}^{n} \left[\sum_{j=1}^{n} \left\{ \frac{a_{ij}}{2} \frac{\partial^2 Q}{\partial x_i \partial x_j} + \frac{\partial a_{ij}}{\partial x_i} \frac{\partial Q}{\partial x_j} \right\} + \frac{\partial b_i}{\partial x_i} \right] w = 0. \tag{3}$$

The functions Q and w are solutions of equations (2) and (3). The relation (2:11) is evaluated with the function v defined by (1),(2),(3) and the expression (5.2:3) for u_s. In the limit $\epsilon \to 0$ we obtain:

$$c_0 = \frac{\int_{\partial D} f(y) b(y) \cdot \xi(y)\, w(y)\, \exp[-Q(y)/\epsilon]\, dS_y}{\int_{\partial D} b(y) \cdot \xi(y)\, w(y)\, \exp[-Q(y)/\epsilon]\, dS_y}. \tag{4}$$

Following the argument of the previous subsection, we find that for small ϵ the exit density p is peaked near the boundary point(s) with the lowest value of Q. Thus, the role played by the potential ϕ_0 in section 5.2 is taken over here by the function Q. The potential ϕ_0 was expressed explicitly in terms of the drift vector and the diffusion matrix by (5.2:7b). Except in some special cases, no such explicit expression exists for Q. In practice this function is obtained through numerical integration of the eikonal equation by the ray method [37]. Such an integration scheme may include the transport equation as well in order to determine w. The method described in the present section is powerful in the sense that it can be applied to a large class of problems in arbitrary dimension. However, the asymptotics to the stationary forward equation (2:1) are not (yet) supported by a solid mathematical background. The asymptotic method described above is similar to an asymptotic method used frequently in geometrical optics and diffraction theory. For the latter method a more or less extensive literature exists, see for example the publications of KELLER and coworkers [7,30] and LUDWIG [36], up to the more recent work of DUISTERMAAT [13], MASLOV [39], MASLOV and FEDORIUK [40], etc.. For the former method, i.e. the asymptotic method to the partial differential equations related to exit problems, the literature is limited, see for example COHEN and LEWIS [8], LUDWIG [37] and a more recent paper of BRANNAN [6].

For the expected exit time the following formula has been derived [41]:

$$T(x) \sim \frac{\sqrt{2\pi\epsilon}\, \exp[Q(y^*)/\epsilon]}{H_1^{1/2}(x_0) \left[b \cdot \xi\, w\, H_2^{-1/2} \right](y^*)} \{1 - \exp[-\tilde{b}(x')|x - x'|/\epsilon]\}, \tag{5a}$$

in which

$$H_1(x_0) = \det\left\{\frac{\partial^2 Q}{\partial x_i \partial x_j}(x_0)\right\}_{i,j=1,2,\ldots,n}, \tag{5b}$$

$$H_2(y^*) = \det\left\{\frac{\partial^2 Q}{\partial y_i \partial y_j}(y^*)\right\}_{i,j=1,2,\ldots,n-1}, \tag{5c}$$

where x_0 is the deterministic equilibrium point and y^* is the unique (by assumption, for other cases see the literature) point on ∂D with the lowest value of Q.

Now that we have obtained expressions for the expected exit time for a diffusion across the flow in section 4 and for diffusions against the flow in section 5, it is interesting to compare them in their dependence on the small parameter ϵ. For the former type of diffusion this dependence algebraic, while for the latter it is exponential. Thus, these results express the quantitative difference in stochastic stability between systems of each type, where the diffusions against the flow are the more stable ones (conform intuition I hope).

In the stochastic systems under consideration the deterministic flow was directed inward at the boundary of the domain. Other systems, in which the deterministic flow at the boundary coincides with the boundary, have been analysed in [42] and [43]. In the first paper there are no critical points of the deterministic system located on this boundary, whereas in the second paper there are.

In the present paper we studied exit problems using formal asymptotic methods. The same subject has been studied by VENTCEL and FREIDLIN [17,58], FRIEDMAN [18] and others from a probabilistic point of view. Rigorous mathematical methods have been used by DAY [11,12], EVANS and ISHII [15], KAMIN [27,28] and others.

The stochastic systems that we considered had a continuous domain. In chemistry, physics, biology and other areas one meets processes with a discontinuous domain, for example birth or birth-death processes. For these processes, asymptotic methods that resemble the method described in this subsection have been presented in [32,33,34,44].

6. SOME APPLICATIONS

Exit models have a wide variety of applications. We shall mention only a few of them below. There are applications in population genetics, see for example CROW and KIMURA [10], who describe the change in gene frequency of biological populations by means of a stochastic diffusion model. Exit from a domain here corresponds to the fixation of a gene. See also MARUYAMA [38] and GILLESPIE [20]. Another application in biology is the description of the dynamics of stochastic populations. In such applications, exit corresponds to extinction of a species. Examples can be found in GOEL and RICHTER-DYN [21], LUDWIG [37], MAY [45], NISBET and GURNEY [49], ROOZEN [51,52], ROUGHGARDEN [54]. Other applications are in mechanics and reliability theory. Many mechanical systems near equilibrium behave essentially like the diffusion across the flow model or the diffusion against the flow model that have been

studied in this paper. The stochastic domain can be chosen as the domain in which the system is known to function properly. Exit corresponds to a break down of the system. The expected exit time is a measure for the reliability of the system. See for example GRASMAN [22], KATZ and SCHUSS [29], ROOZEN [53]. For an application of an exit model to the dynamics of the atmospheric circulation, see DE SWART and GRASMAN [57]. The expected exit times predict the lifetimes of alternative circulation types. Other applications of exit models can be found in the literature.

REFERENCES

1. L. ARNOLD (1971), *Stochastic differential equations: Theory and applications.* J. Wiley.
2. C.M. BENDER and S.A. ORSZAG (1978), *Advanced mathematical methods for scientists and engineers.* Mcgraw-Hill Book Company.
3. A. BENSOUSSAN, J.L. LIONS, G. PAPANICOLAOU (1978), *Asymptotic analysis for periodic structures.* Studies in mathematics and its applications, Vol. 5.
4. N. BLEISTEIN and R.A. HANDELSMAN (1975), *Asymptotic expansions of integrals.* Holt, Rinehart and Winston.
5. O. BOLZA (1904), *Lectures on the calculus of variations.* Chelsea publishing company.
6. J.R. BRANNAN (1987), *Parabolic ray theory with boundaries.* Dept. of Math. Sc., Clemson University, Clemson, South Carolina.
7. R.N. BUCHAL and J.B. KELLER (1960), *Boundary layer problems in diffraction theory.* Comm. Pure Appl. Math. Vol. 13.
8. J.K. COHEN and R.M. LEWIS (1967), *A ray method for the asymptotic solution of the diffusion equation.* J. Inst. Math. Appl., 3.
9. R. COURANT and D. HILBERT (1953), *Methods of mathematical physics,* Vol. I. Interscience Publishers Inc.
10. J.F. CROW and M. KIMURA (1970), *An introduction to population genetics theory.* Harper and Row.
11. M.V. DAY (1982), *Exponential leveling for stochastically perturbed dynamical systems.* SIAM J. Math. Anal. Vol. 13, No. 4.
12. M.V. DAY (1986), *Recent progress on the small parameter exit problem.* Dept. of Math. Virginia Polyt. Inst. and St. Univ.
13. J.J. DUISTERMAAT (1974), *Oscillatory integrals, Lagrange immersions and unfolding of singularities.* Comm. Pure. Appl. Math., Vol. 27.
14. M.M. DYGAS, B.J. MATKOWSKY, Z. SCHUSS (1987), *Stochastic stability of nonlinear oscillators.* Technical Report No. 8615, Appl. Math. Northwestern University, Evanston, Illinois.
15. L.C. EVANS, H. ISHII (1985), *A PDE approach to some asymptotic problems concerning random differential equations with small noise intensities.* Ann. Inst. Henri Poincaré - Analyse non linéaire - Vol. 2, No. 1.
16. W. FELLER (1952), *The parabolic differential equations and the associated semigroups of transformations.* Ann. of Math. Vol. 55, No. 3.
17. M.I. FREIDLIN and A.D. WENTZELL (1984), *Random perturbations of*

dynamical systems. Springer-Verlag.

18. A. FRIEDMAN (1976), *Stochastic differential equations and applications, Vol. 2.* Academic Press.
19. C.W. GARDINER (1983), *Handbook of stochastic methods for physics, chemistry and the natural sciences.* Springer-Verlag.
20. J.H. GILLESPIE (1985), *The interaction of genetic drift and mutation with selection in a fluctuating environment.* Theor. Pop. Biol. 27.
21. N.S. GOEL and N. RICHTER-DYN (1974), *Stochastic models in biology.* Academic Press.
22. J. GRASMAN (1988), *Lifetime distributions and stochastic dynamical systems.* Proc. Int. Symp. Probab. and Bayesian Statistics, Innsbruck 1986, Plenum Press.
23. J. GRASMAN and B.J. MATKOWSKY (1977), *A variational approach to singularly perturbed boundary value problems for ordinary and partial differential equations with turning points.* SIAM J. Appl. Math. Vol. 32, No. 3.
24. P.P.N. DE GROEN (1980), *The singularly perturbed turning point problem: a spectral approach.* MRC Techn. Summ. Rep. #2140, Univ. of Wisconsin-Madison.
25. F.B. HANSON and C. TIER (1981), *An asymptotic solution of the first passage problem for singular diffusion in population biology.* SIAM J. Appl. Math. Vol. 40, No. 1.
26. JIANG FURU (1980), *On the boundary value problems for a class of ordinary differential equations with turning points.* Applied Mathematics and Mechanics (Engl. Ed., Vol. I, No. 2).
27. S. KAMIN (1978), *Elliptic perturbation of a first order operator with a singular point of attracting type.* Indiana Univ. Math. J., Vol. 27, No. 6.
28. S. KAMIN (1979), *On elliptic singular perturbation problems with turning points.* SIAM J. Math. Anal. Vol. 10, No. 3.
29. A. KATZ and Z. SCHUSS (1985), *Reliability of elastic structures driven by random loads.* SIAM J. Appl. Math. Vol. 45, No. 3.
30. J.B. KELLER (1958), A geometrical theory of diffraction. *Calculus of variations and its applications,* Mcgraw-Hill.
31. R.Z. KHAS'MINSKII (1966), *A limit theorem for the solutions of differential equations with random right-hand sides.* Theory of probability and its applications, Vol. XI, Number 3.
32. C. KNESSL, M. MANGEL, B.J. MATKOWSKY, Z. SCHUSS, C. TIER (1984), *Solution of Kramers-Moyal equations for problems in chemical physics.* J. Chem. Phys. 81 (3).
33. C. KNESSL, B.J. MATKOWSKY, Z. SCHUSS, C. TIER (1985), *An asymptotic theory of large deviations for Markov jump processes.* SIAM J. Appl. Math. Vol. 46, No. 6.
34. C. KNESSL, B.J. MATKOWSKY, Z. SCHUSS, C. TIER (1986), *A singular perturbation approach to first passage times for Markov jump processes.* J. Stat. Phys., Vol. 42, Nos. 1/2.
35. H.A. KRAMERS (1940), *Brownian motion in a field of forces and the diffusion model of chemical reaction.* Physica 7.

36. D. LUDWIG (1966), *Uniform asymptotic expansions at a caustic.* Comm. Pure Appl. Math., Vol. 19.
37. D. LUDWIG (1975), *Persistence of dynamical systems under random perturbations.* SIAM Rev. Vol. 17, No. 4.
38. T. MARUYAMA (1977), *Stochastic problems in population genetics.* Lecture Notes in Biomathematics. Springer-Verlag.
39. V.P. MASLOW (1972), *Théorie des perturbations et méthodes asymptotiques.* Dunod, Gauthier-Villars.
40. V.P. MASLOW and M.V. FEDORIUK (1981), *Semi-classical approximation in quantum mechanics.* D. Reidel.
41. B.J. MATKOWSKY and Z. SCHUSS (1977), *The exit problem for randomly perturbed dynamical systems.* SIAM J. Appl. Math. Vol. 33, No. 2.
42. B.J. MATKOWSKY and Z. SCHUSS (1982), *Diffusion across characteristic boundaries.* SIAM J. Appl. Math. Vol. 42, No. 4.
43. B.J. MATKOWSKY, Z. SCHUSS, and C. TIER (1983), *Diffusion across characteristic boundaries with critical points.* SIAM J. Appl. Math. Vol. 43, No. 4.
44. B.J. MATKOWSKY, Z. SCHUSS, C. KNESSL, C.TIER, M. MANGEL (1984), *Asymptotic solution of the Kramers-Moyal equation and first passage times for Markov jump processes.* Phys. Rev. A, Vol. 29, No. 6.
45. R.M. MAY (1974), *Stability and complexity in model ecosystems.* Princeton Univ. Press, Sec. ed.
46. S.G. MICHLIN (1962), *Variationsmethoden der mathematischen Physik.* Akad.-Verlag.
47. S.G. MIKHLIN (1965), *The problem of the minimum of a quadratic functional.* Holden-Day Inc.
48. S.G. MIKHLIN (1967), *Linear equations of mathematical physics.* Holt, Rinehart and Winston.
49. R.M. NISBET and W.S.C. GURNEY (1982), *Modelling fluctuating populations.* Wiley-Interscience.
50. G.C. PAPANICOLAOU (1977), *Introduction to the asymptotic analysis of stochastic equations.* Modern modelling of continuum phenomena, R.C. DiPrima ed., Lectures in Applied Mathematics, Vol. 16, American Mathematical Society.
51. H. ROOZEN (1987), *Equilibrium and extinction in stochastic population dynamics.* Bull. Math. Biol., Vol. 49, No. 6.
52. H. ROOZEN (1987), *An asymptotic solution to a two-dimensional exit problem arising in population dynamics.* Accepted for publication in SIAM J. Appl. Math.
53. H. ROOZEN (1988), *Stochastic stability of the loaded stiff rod.* notition.
54. J. ROUGHGARDEN (1979), *Theory of population genetics and evolutionary ecology: an introduction.* Macmillian.
55. Z. SCHUSS (1980), *Theory and applications of stochastic differential equations.* Wiley.
56. R.L. STRATONOVICH (1963), *Topics in the theory of random noise, Vol. I.* Gordon and Breach.

57. H.E. DE SWART and J. GRASMAN (1987), *Effect of stochastic perturbations on a low-order spectral model of the atmospheric circulation.* Tellus, 39A.
58. A.D. VENTCEL and M.I. FREIDLIN (1970), *On small perturbations of dynamical systems.* Uspehi Mat. Nauk 25; Russian Math. Surveys 25.
59. A.M. WAZWAZ and F.B. HANSON (1986), *Matched uniform approximations for a singular boundary point and an interior turning point.* SIAM J. Appl. Math. Vol. 46, No. 6.
60. A.M. WAZWAZ and F.B. HANSON (1986), *Singular boundary resonance with turning point resonance.* SIAM J. Appl. Math. Vol. 46, No. 6.
61. M. WILLIAMS (1982), *Asymptotic exit time distributions.* SIAM J. Appl. Math. Vol. 42, No. 1.

Iteration of Morphological Transformations

H.J.A.M. Heijmans
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

Morphological transformations play an important role in the morphological analysis of images. Certain features of images can be revealed by putting them through a cleverly chosen sequence of morphological transformations. This paper deals with some aspects of morphological transformations, in particular with the question under what conditions iteration of such mappings yields openings or closings.

1980 Mathematics Subject Classification: 68U10
Keywords & Phrases: image analysis, mathematical morphology, morphological transformation, dilation, erosion, hit-or-miss transformation, opening, closing, domain of invariance, $\downarrow$ -continuous mapping, iteration of morphological transformations.

1. Introduction

Mathematical morphology is a particular branch in image processing founded by Matheron [3] and Serra [6, 7], which is concerned with the study of probabilistic models for images, the investigation of image transformations and functionals, and the development of specific algorithms. It derives its tools from algebra, topology, integral geometry, and stochastic geometry. In this paper our attention will be focussed on binary (= black and white) images, although an extension to grey-level functions is straightforward. Moreover we will be only concerned with morphological image transformations. These are transformations which map the space of all subsets of $\mathbb{R}^d$(continuous case) or $\mathbb{Z}^d$ (dicrete case) into itself and which are invariant under translations. Or in mathematical language: a morphological transformation is a transformation ψ which maps the space $P(E)$ of all subsets of E (where $E=\mathbb{R}^d$ or $\mathbb{Z}^d$) into itself and satisfies

$$\psi(X_h)=(\psi(X))_h, \text{ for } X\subseteq E \text{ and } h\in E. \tag{1.1}$$

Here $X_h=\{x+h|x\in X\}$ is the translate of X along h. We point out that there does not exist an unambiguous definition of a morphological transformation in the literature: for our purpose the above definition will do. Morphological transformations are used to detect certain features of an object, or, as Serra puts it in the introduction to [7]: *"In order to compare bodies, to recognize them, and to uncover their genesis, or to follow their evolution in time – in brief, to reduce them to their essentials – mathematical morphology classifies them into groups of more or less similar entities by putting them through sequences of set*

transformations." A common feature of morphological transformations (we will see some typical examples in the forthcoming sections) which, from a mathematical point of view, makes them very interesting, is their nonlinear and irreversible nature. Furthermore, they are often *locally defined*: to determine if h lies in $\psi(X)$ one doesn't have to know the entire image X but only $X \cap M_h$, the part of X within some bounded mask M positioned at h. To be precise

$$h \in \psi(X) \text{ if and only if } h \in \psi(X \cap M_h). \tag{1.2}$$

Of course, M depends on ψ and by iteration of ψ the size of the mask may increase, eventually becoming unbounded.

The remainder of this section will be used to introduce some further terminology. Let ψ: $P(E) \to P(E)$, where $E = \mathbb{R}^d$ or $\mathbb{Z}^d$, be an arbitrary mapping. Then ψ is called

- *increasing* if $X \subseteq Y$ implies $\psi(X) \subseteq \psi(Y)$
- *extensive* if $X \subseteq \psi(X)$, for every $X \subseteq E$
- *anti-extensive* if $\psi(X) \subseteq X$, for every $X \subseteq E$
- *idempotent* if $\psi^2(X) = \psi(X)$, for every $X \subseteq E$.

Here $\psi^2 = \psi \circ \psi$. We denote by *id* the identity mapping given by $id\,(X) = X$. If ψ_1, ψ_2 are two mappings then we write $\psi_1 \leqslant \psi_2$ if $\psi_1(X) \subseteq \psi_2(X)$ for every $X \subseteq E$. Thus a mapping is extensive if $\psi \geqslant id$. If ψ_i is a mapping for any i in the index set I, then $\cap_{i \in I} \psi_i$ is the mapping given by

$$(\cap_{i \in I}\, \psi_i)(X) = \cap_{i \in I}\, \psi_i(X),\ X \subseteq E.$$

$\cup_{i \in I}\, \psi_i$ is defined similarly. The *dual* (or *complementary*) mapping of ψ is defined as

$$\psi^*(X) = (\psi(X^c))^c,\ X \subseteq E.$$

Here X^c stands for the complement of X.

Finally we devote some words to the discrete space $E = \mathbb{Z}^2$ subdivided by the square grid. In this case an image X can be represented as a (possibly infinite) collection of pixels. Every pixel has 4 horizontal and vertical neighbours, the so-called 4-neighbours. In addition it has 4 diagonal neighbours which together with the 4-neighbours are called the 8-neighbours. An object $X \subseteq \mathbb{Z}^2$ is called 4-connected if for every pair $h, k \in X$ there exists a sequence $h = h_0, h_1, h_2, \ldots, h_m = k$ in X such that h_{i-1} and h_i $(i = 1, \ldots, m)$ are 4-neighbours. Similarly, 8-connectedness is defined. One can still think of other neighbourhood relations. Serra and co-workers [6] have chosen to work on the hexagonal grid where every pixel has six neighbours. The main advantage of the hexagonal grid over the square grid is that it has more rotational symmetry. This is reflected in the resulting algorithms which are simpler than in the square case.

2. Dilation, erosion, and matheron's theorem

In classical signal analysis an important role is played by linear operations such as convolution and Fourier transformation. The characteristic feature of mathematical morphology is that the object space is not a vector space but has a lattice structure [1,7]. It is therefore not too surprising that the two basic morphological transformations, dilation and erosion, are in a very specific manner compatible with the ordering structure of this space: dilation since it commutes with unions, and erosion since it commutes with intersections. The mathematical definitions are as follows. Let $A \subseteq E$. The *dilation* of the image X by A is defined as

$$X \oplus A = \cup_{h \in A} X_h. \tag{2.1}$$

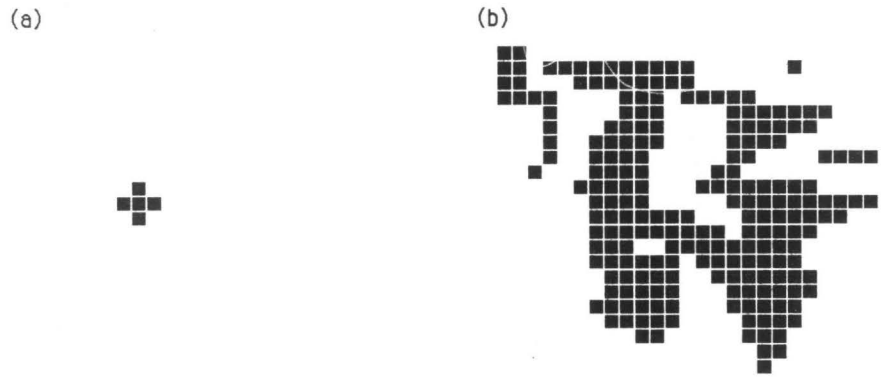
The *erosion* of X by A is given by

$$X \ominus A = \cap_{h \in A} X_{-h}. \tag{2.2}$$

The operations $\oplus$ and $\ominus$ have first ocurred in the context of *integral geometry* [3] where they were called Minkowski addition and Minkowski subtraction, respectively. We warn the reader that our definition of erosion is slightly different from that in [3, 6, 7]. A is called the *structuring element* and is usally assumed to be compact (*if* $E = \mathbb{R}^d$) or finite (*if* $E = \mathbb{Z}^d$). In the discrete case $E = \mathbb{Z}^2$ the most obvious choices for A are the set consisting of one central pixel and its four horizontal and vertical neighbours (hereafter called CROSS), and the set which contains in addition its four diagonal neighbours (hereafter called SQUARE). If $0 \in A$ then a natural though unnecessary restriction is

$$X \ominus A \subseteq X \subseteq X \oplus A.$$

The action of dilation and erosion is depicted in Figure 1. Throughout this paper we use the following convention in illustrations. Points belonging both to the initial image X and its transform $\psi(X)$ are denoted by ■, points belonging to $\psi(X)$ but not to X are denoted by ◘ and points in X which do not belong to $\psi(X)$ are denoted by □.



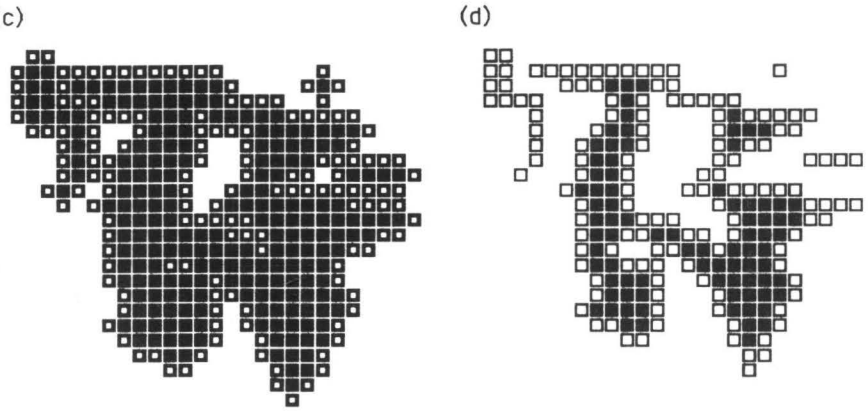


FIGURE 1: Dilation and erosion of X by A. (a) Structuring element A. (b) Orginal image X. (c) Dilated set $X \oplus A$. (d) Eroded set $X \ominus A$.

Let

$$\overset{\vee}{A} = \{-h | h \in A\}$$

be the reflected structuring element. The following algebraic relations hold.

$$(X \oplus A)^c = X^c \ominus \overset{\vee}{A} \tag{2.3}$$

$$(X \oplus A) \oplus B = X \oplus (A \oplus B) \tag{2.4a}$$

$$(X \ominus A) \ominus B = X \ominus (A \oplus B) \tag{2.4b}$$

$$X_h \oplus A = (X \oplus A)_h \tag{2.5a}$$

$$X_h \ominus A = (X \ominus A)_h \tag{2.5b}$$

$$(\cup_{i \in I} X_i) \oplus A = \cup_{i \in I} (X_i \oplus A) \tag{2.6a}$$

$$(\cap_{i \in I} X_i) \ominus A = \cap_{i \in I} (X_i \ominus A). \tag{2.6b}$$

Here $X_i \subseteq E$ for every i in the (finite or infinite) index set I. The duality relation (2.3) states that dilation of an object yields the same result as the erosion of its background. The main implication of this relation is that properties of dilations and erosions always occur in pairs: to every property of dilations there corresponds a dual property of erosions and vice versa. Yet there exists a second duality relation between dilations and erosion which is only based on the (partial) ordering relation $\subseteq$ and which for that reason can also be used in more general situations:

$$X \oplus A \subseteq Y \text{ if and only if } X \subseteq Y \ominus A. \tag{2.7}$$

In a slightly different setting, this relation is also known as *Galois connection* or *adjunction*: see [1]. From (2.7) it follows immediately that

$$(X \ominus A) \oplus A \subseteq X \subseteq (X \oplus A) \ominus A, \tag{2.8}$$

where, in general, the inclusions may be strict. Both dilation and erosion belong to the class of increasing and translation-invariant transformations. It is easily seen that by taking intersections and unions of dilations and erosions, respectively, we stay within this class. But the converse also holds: every increasing translation-invariant mapping on $\mathrm{P}(E)$ can be decomposed as an intersection of dilations or, equivalently, as a union of erosions. This is essentially the content of Matheron's theorem which we state below. Before we do this we introduce the notion of a kernel. Let $\psi: \mathrm{P}(E) \to \mathrm{P}(E)$ be an arbitrary mapping. Then its kernel $\mathcal{V}[\psi]$ is defined as

$$\mathcal{V}[\psi] = \{A \subseteq E \mid 0 \in \psi(A)\}. \tag{2.9}$$

By $\mathcal{V}^*$ we denote the kernel of the dual mapping i.e. $\mathcal{V}^* = \mathcal{V}[\psi^*]$.

THEOREM 2.1 (Matheron's theorem).
Let $\psi: \mathrm{P}(E) \to \mathrm{P}(E)$ be an increasing translation-invariant mapping with kernel $\mathcal{V}$. Then

$$\psi(X) = \cup_{A \in \mathcal{V}} (X \ominus A) = \cap_{A \in \mathcal{V}} (X \oplus \check{A})$$

We give an illustration of this theorem by applying it to the so-called *median filter*. Let $A \subseteq \mathbb{Z}^d$ be a finite structuring element containing an odd number of points, say $2p-1$. For an image X we define $\mu_A(X)$ by

$$h \in \mu_A(X) \text{ if and only if } \#(X \cap A_h) \geqslant \mathrm{p}.$$

Here $\#(Y)$ denotes the number of points of Y. In Figure 2 we have depicted the action of the median filter μ_A in the 2-dimensional case with CROSS as structuring element. This particular example is called the 4-median filter.

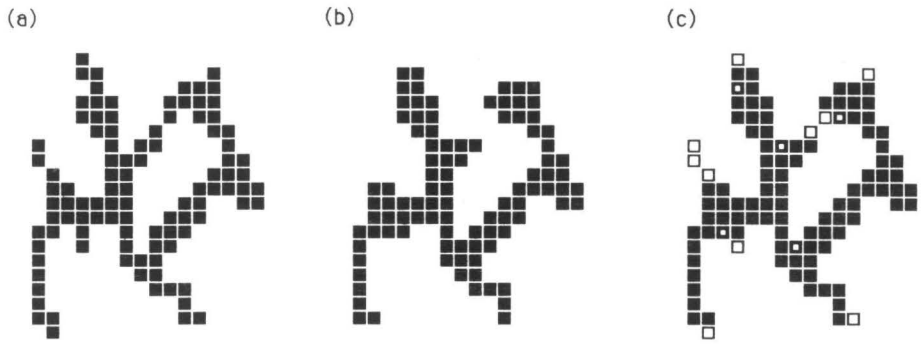


FIGURE 2: The 4-median filter. (a) Orginal image X. (b) 4-median of X. (c) Comparison of the original and the transformed image.

Before we decompose μ_A as a union of erosions we state a trivial but, from a practical point of view, very important result.

PROPOSITION 2.2. *Let ψ be an increasing translation-invariant transformation with kernel $\mathcal{V}$. If $\mathcal{V}_0 \subseteq \mathcal{V}$ is such that for every $A \in \mathcal{V}$ there exists an $A_0 \in \mathcal{V}_0$ such that $A_0 \subseteq A$, then*

$$\psi(X) = \cup_{A_0 \in \mathcal{V}_0} (X \ominus A_0)$$

If the underlying space is discrete and if the transformation ψ is locally defined (c.f. Section 1) then $\mathcal{V}_0$ can be chosen finite. This is the content of the next theorem.

THEOREM 2.3. *Let $E = \mathbb{Z}^d$. If ψ is an increasing translation-invariant transformation which is locally defined, then there exists a finite subset $\mathcal{V}_0$ of $\mathcal{V}[\psi]$ such that*

$$\psi(X) = \cup_{A_0 \in \mathcal{V}_0} (X \ominus A_0).$$

PROOF. Let $\mathcal{V}$ be the kernel of ψ, i.e. $A \in \mathcal{V}$ if and only if $0 \in \psi(A)$. Since ψ is locally defined there exists a bounded (hence finite) mask $M \subseteq \mathbb{Z}^d$ such that

$$0 \in \psi(A) \text{ if and only if } 0 \in \psi(A \cap M).$$

Define

$$\mathcal{V}_0 = \{A \cap M | A \in \mathcal{V}\}.$$

Then $\mathcal{V}_0 \subseteq \mathcal{V}$, and for every $A \in \mathcal{V}$ there exists an $A_0 \in \mathcal{V}_0$ such that $A_0 \subseteq A$, namely $A_0 = A \cap M$. Since $\mathcal{V}_0 \subseteq \mathrm{P}(M)$ and M is finite, also $\mathcal{V}_0$ is finite. □

In the case of μ_A we have

$$\mathcal{V}[\mu_A] = \{B \subseteq \mathbb{Z}^d | \#(A \cap B) \geqslant p\},$$

and we may choose

$$\mathcal{V}_0 = \{B \subseteq \mathbb{Z}^d | B \subseteq A \text{ and } \#(B) = p\}.$$

Thus, by a well-known combinatorial result, $\#\mathcal{V}_0 = \begin{bmatrix} 2p-1 \\ p \end{bmatrix}$. In the example of Figure 2, we have $p = 3$ and hence $\#(\mathcal{V}_0) = 10$. The number $\#(\mathcal{V}_0)$ increases exponentially with p, and eventually there will be more economic ways to perform μ_A.

3. ITERATION OF TRANSFORMATIONS WHICH ARE NOT INCREASING

Although a substantial part of mathematical morphology is concerned with increasing transformations, many important algorithms use transformations which are not increasing. We mention $X \to X^c$ as an obvious but also important member of this class. Adding this prototype example to the increasing

transformations of the previous section and taking unions, intersections and compositions, we obtain a large family of morphological transformations which are not necessarily increasing. A member of this family which deserves extra attention is the so-called *hit-or-miss transformation*. Let A,B be two non-intersecting structuring elements. We define

$$X \circledast (A,B) = \{h \in E | A_h \subseteq X \text{ and } B_h \subseteq X^c\}. \tag{3.1}$$

One can easily give a description of this mapping in terms of erosions and dilations by using the following (geometric) characterization of erosion:

$$X \ominus A = \{h \in E | A_h \subseteq X\}. \tag{3.2}$$

Combining (3.1), (3.2) and (2.3) we get

$$X \circledast (A,B) = (X \ominus A) \cap (X^c \ominus B) = (X \ominus A) \cap (X \overset{\vee}{\oplus} B)^c$$
$$= (X \ominus A) \setminus (X \overset{\vee}{\oplus} B).$$

Putting $T = (A,B)$ we may also write $X \circledast T$ instead of $X \circledast (A,B)$. The hit-or-miss transformation is well suited to locate within an object points with certain (local) geometric properties, e.g. isolated points, border points, or corner points. In Figure 3 below, a hit-or-miss transformation is used to locate all lower left corner points of an object: note that the structure of (A,B) reflects the structure one is looking for within the object.

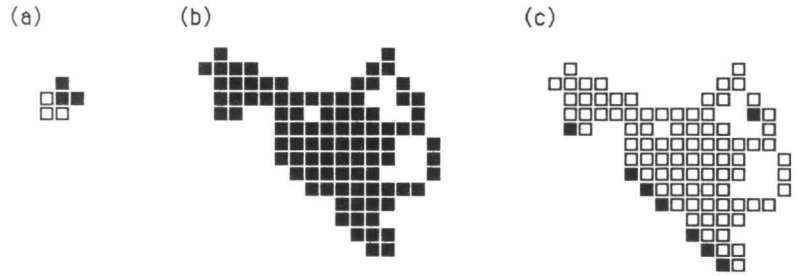


FIGURE 3: Hit-or-miss transformation. (a) Structuring element (A,B); the points of A are denoted by ■, the points of B by □. (b) Original image. (c) The hit-or-miss transformation $X \circledast (A,B)$ consists of all lower left corner points of X.

As we already mentioned, the hit-or-miss transformation plays a prominent role in quite a number of morphological image processing algorithms. Below, we shall describe in some detail two such algorithms: the (finite) convex hull and the skeleton. In both examples (and many others not mentioned here) the basic idea is to add or delete points from an object depending on the state of their neighbouring pixels. This is also the key feature of the so-called *cellular automata*: the state of a cell (pixel) at time $t+1$ is determined by its own state and those of its neigbours at time t: see [4]. Probably the best known example is Conway's life game.

We introduce two other transformations which are not increasing, namely the *thickening* and the *thinning*. The thickening and the thinning of an image X with the structuring element $T=(A,B)$ are respectively defined as

$$X \odot T = X \cup (X \circledast T) \tag{3.3}$$

$$X \oslash T = X \setminus (X \circledast T). \tag{3.4}$$

Since $X \circledast T \subseteq X$ if $0 \in A$ and $X \circledast T = \varnothing$ if $A \cap B \neq \varnothing$, the thickening operator yields a non-trivial result only if $A \cap B = \varnothing$ and $0 \notin A$. Similarly, the thinning is a non-trivial operation if $A \cap B = \varnothing$ and $0 \notin B$. It is not difficult to check that thickening and thinning are dual operations:

$$X^c \odot (A,B) = (X \oslash (B,A))^c. \tag{3.5}$$

The examples which we give below involve a sequential application of the thickening and thinning operation. So let us first make some general remarks about iteration of morphological transformations. Let $\psi: \mathrm{P}(E) \to \mathrm{P}(E)$ be a translation-invariant (not necessarily increasing) transformation. Suppose furthermore that ψ is anti-extensive, i.e. $\psi(X) \subseteq X$ for all $X \subseteq E$. Then, for every non-negative integer k, $\psi^{k+1} \leqslant \psi^k$ and we define ψ^∞ by

$$\psi^\infty(X) = \bigcap_{k \geqslant 1} \psi^k(X). \tag{3.6}$$

Then ψ^∞ is translation-invariant and anti-extensive. Futhermore, ψ^∞ is increasing if ψ is. It is rather easy to show that ψ^∞ is idempotent if and only if $\psi \circ \psi^\infty = \psi^\infty$. In Section 4 we present an example of a morphological transformation ψ which is translation-invariant, increasing, and anti-extensive, but for which ψ^∞ is *not* idempotent.

For extensive mappings ψ we define ψ_∞ by

$$\psi_\infty(X) = \bigcup_{k \geqslant 1} \psi^k(X). \tag{3.7}$$

Convex hull

Convexity of an object is a global property. Local information does not suffice to decide about the convexity of an object. So how can morphological transformations, which essentially only require local knowledge of an image (at least in the practical cases where the structuring elements are bounded), be used to construct the convex hull of an object? We consider the case $E = \mathbb{Z}^2$ and define the convex hull $CH(X)$ as the intersection of all discrete halfplanes which contain X: by a discrete halfplane we mean a set of points $(x,y) \in \mathbb{Z}^2$ satisfying the requirement $ax + by \leqslant c$, where $a,b,c \in \mathbb{Z}$. A set $X \subseteq \mathbb{Z}^2$ is called convex if $X = CH(X)$. For a thorough exposition on discrete convexity we refer to [5].

Obviously the mapping $X \to CH(X)$ is a translation-invariant and increasing mapping, and application of Matheron's theorem shows that $CH(X)$ can be obtained as a union of erosions. So far, so good. Unfortunately, the kernel $\mathcal{V}$ of CH contains infinitely many structuring elements and cannot be reduced to a finite set $\mathcal{V}_0$ as in Proposition 2.2: in fact, this would contradict the global

character of convexity. Therefore the representation of CH as a union of erosions is not of any practical value.

If, in the definition of the discrete convex hull we only admit those halfplanes for which $|a|,|b| \leq 1$ (that are halfplanes bounded by lines whose angle with the positive x-axis is an integer multiple of 45°) then the situation changes drastically. We call the resulting notion 45-convexity and denote the 45-convex hull of X by $CH_{45}(X)$. There exists an algorithm which is essentially an iteration of thickenings and which yields the 45-convex hull if the initial object is 4-connected. Let $T_1,...,T_8$ be as depicted in Figure 4a and define

$$\psi(X) = (....((X \odot T_1) \odot T_2) \odot ..... \odot T_8).$$

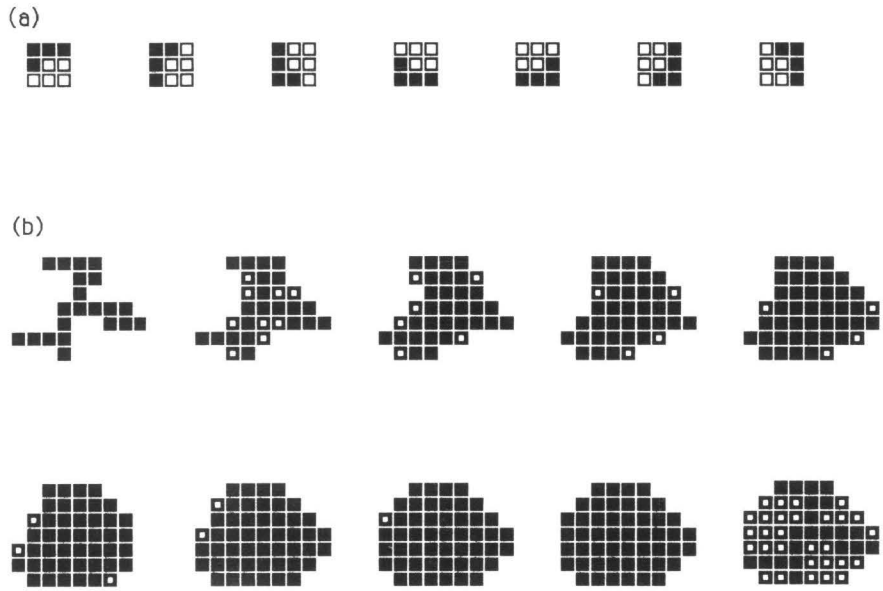


FIGURE 4: (a) Structuring elements for the computation of the 45-convex hull; T_2,T_3,T_4, etc. are the 45°-, 90°-, 135°-, etc. rotations of T_1. (b) The 45-convex hull can be computed by iteration of $\psi(X)=(...((X \odot T_1) \odot T_2) \odot ... \odot T_8)$, with $T_1,T_2,...,T_8$ as depicted in (a). The first object is the initial image X, the second object is $\psi(X)$, etc. The one but last object shows the final result $\psi_\infty(X)=CH_{45}(X)$, which is reached after 7 iterations. The last object compares the original image to its 45-convex hull.

Then ψ is an extensive, translation-invariant mapping and $\psi_\infty(X) = CH_{45}(X)$ if X is 4-connected. We shall not prove this result, but rather illustrate it by means of an example: see Figure 4b above.

Skeleton

Our second example handles the computation of the skeleton of a discrete object. Our use of words ("the " skeleton) might suggest that this notion is well-defined, but unfortunately this is far beyond the truth, both in the continuous and the discrete case. We refer the reader to chapter XI of [6] and chapters 11-13 of [7] for an interesting discussion on both the theoretical and practical aspects of the skeleton. A possible definition of the skeleton in the continuous case $E = \mathbb{R}^d$ goes as follows: The skeleton of an object X is the set of all points x such that the maximal ball B centered at x and contained in X, intersects the boundary of X in two or more points. One may think of the skeleton as a set which is a union of arcs and which has the same homotopy as the set X.

Algorithms for the computations of "the" skeleton (at least, something which looks like it) in the discrete case are sometimes called *homotopic thinnings*. As an example we mention the sequential thinning by the structuring elements $T_1, T_2, ..., T_8$ of Figure 5a. Formally, this algorithm, which is originally due to LEVIALDI [2], transforms X into $\psi^\infty(X)$: here $\psi(X) = (...((X \circ T_1) \circ T_2) \circ ... \circ T_8)$. Again, a figure can explain more than thousand words.

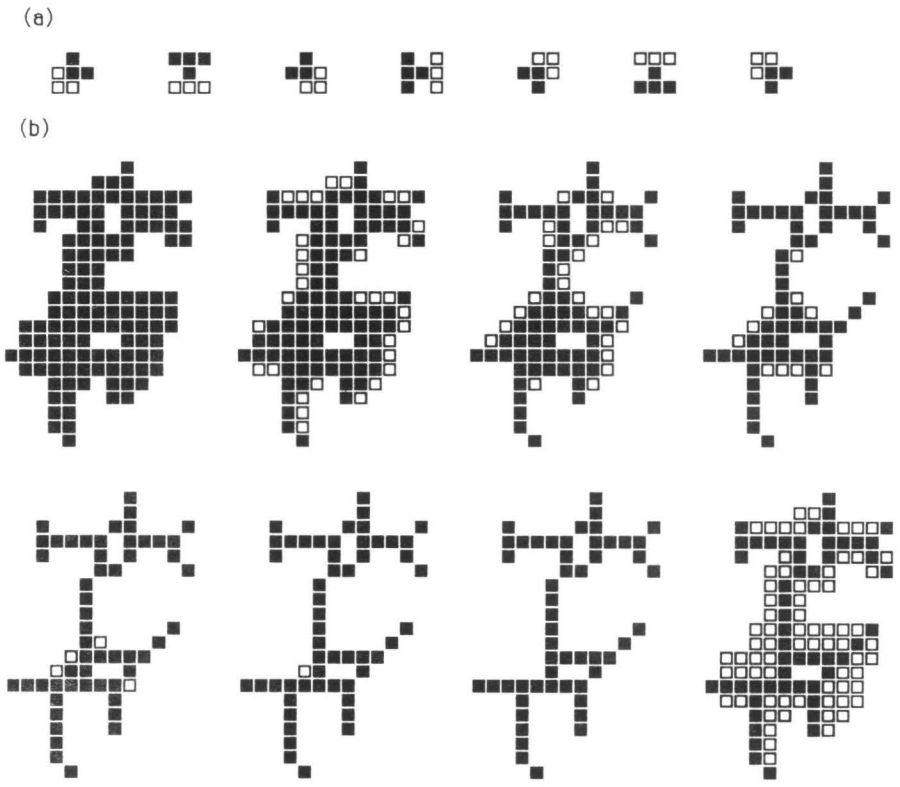


FIGURE 5: (a) Structuring elements for the computation of the homotopic thinning; T_3, T_5, T_7 are the 90°-, 180°-, 270°-, rotations of T_1. The same applies to T_2, T_4, T_6, T_8. (b) Homotopic thinning obtained by iteration of the mapping $\psi(X)=(...((X \bigcirc T_1) \bigcirc T_2) \bigcirc ... \bigcirc T_8)$, with $T_1, T_2, ..., T_8$ as depicted in (a) (Levialdi's algorithm). The first object is the initial image, the second $\psi(X)$, etc. The one but last object is $\psi^\infty(X)$, the homotopic thinning of X, and is reached after 5 iterations. The last object compares the homotopic thinning $\psi^\infty(X)$ to the initial image X.

4. OPENING, CLOSING AND ITERATION OF INCREASING TRANSFORMATIONS

In Section 2 we saw that dilation and erosion are irreversible operations and that

$$(X \ominus A) \oplus A \subseteq X \subseteq (X \oplus A) \ominus A$$

where in general the inclusions are strict. The operations

$$X^A = (X \oplus A) \ominus A \tag{4.1}$$

$$X_A = (X \ominus A) \oplus A \tag{4.2}$$

are called the *closing* resp. *opening* of X by A. Like dilation and erosion, closing and opening are translation-invariant increasing transformations, and they are related by the duality relation

$$(X^c)_A = (\check{X^A})^c. \tag{4.3}$$

Furthermore, one can easily show that

$$X_A = \cup \{A_h | h \in E, A_h \subseteq X\}, \tag{4.4}$$

or in words, the opening of X by A consists of all translates of A which are contained in X. Depending on the structure of A, one can think of the opening as an operation which deletes small isolated particles and removes thin recesses from an object. On the other hand, the action of the closing is best understood by interpreting it as the opening of the complement (background) with the reflected structuring element: it fills small holes and narrow coves in an object. In Figure 6 the closing and opening of the object of Figure 1 is performed: again we have chosen CROSS as structuring element.

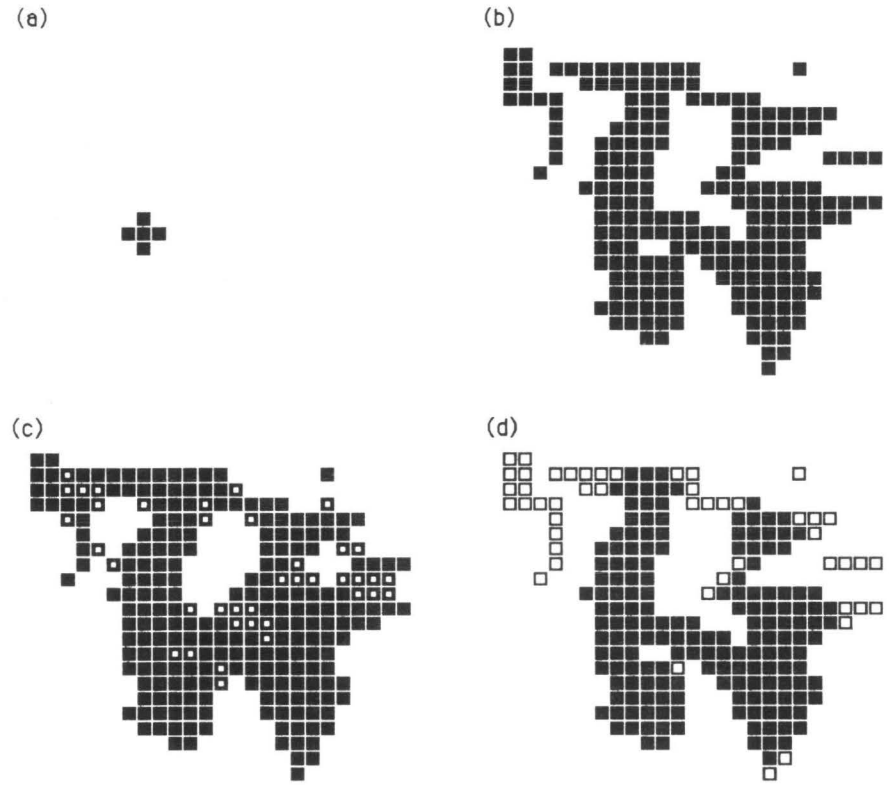


FIGURE 6: Closing and opening of X by A. (a) Structuring element A. (b) Orginal image X. (c) Closing X^A. (d) Opening X_A.

It follows from the geometric interpretation that a repeated application of the closing (or opening) has no further effect: both are idempotent transformations. Or in mathematical terms:

$$(X^A)^A = X^A, (X_A)_A = X_A.$$

DEFINITION. A mapping ψ:P(E)→P(E) is called a closing (opening) if ψ is translation-invariant, increasing, idempotent, and extensive (anti- extensive).

As much as posible we shall denote a closing by ϕ and an opening by α. In the exposition below we shall merely be concerned with openings, but it should be clear that by duality all that is going to be said about openings has its counterpart for closings: if α is an opening then α^* is a closing and vice versa.

The opening given by (4.2) is a very special one: we call it a *structural opening* to indicate that it uses one structuring element A. Below we shall prove that the class of structural openings forms a basis for the overall class of openings: every opening can be obtained as a union of structural openings. To do this we need some auxiliary results.

PROPOSITION 4.1. *Let α_i be an opening for every i in the index set I. Then $\alpha = \cup_{i\in I}\alpha_i$ is an opening as well.*

PROOF. Translation-invariance, increasingness, and anti-extensivity of α are trivial. We only prove the idempotence. Since $\alpha \leqslant id$ we immediately get that $\alpha^2 \leqslant \alpha$. On the other hand $\alpha^2 \geqslant \alpha_i\alpha_i = \alpha_i^2 = \alpha_i$, hence $\alpha^2 \geqslant \cup_{i\in I}\alpha_i = \alpha$. □

The action of a particular opening is best understood by studying its *domain of invariance*. Let ψ: P(E)→ P(E) be an arbitrary mapping. The domain of invariance (= set of fixed points) of ψ is defined as

$$\mathrm{Inv}(\psi) = \{X \subseteq E | \psi(X) = X\}.$$

If α is an opening, then Inv(α) is closed under translations and arbitrary unions (in particular, $\varnothing \in$ Inv(α)). We say that Inv(α) is *invariant under dilations*, since $A \in$ Inv(α) and $B \subseteq E$ arbitrary implies that $A \oplus B \in$ Inv(α). Our next result shows among others, that every opening is uniquely characterized by its domain of invariance.

PROPOSITION 4.2. *Let α_1, α_2 be openings. Then $\alpha_1 \geqslant \alpha_2$ if and only if* Inv(α_2) ⊆ Inv(α_1), *and in that case $\alpha_1\alpha_2 = \alpha_2\alpha_1 = \alpha_2$. In particular, $\alpha_1 = \alpha_2$ if and only if* Inv(α_1) = Inv(α_2).

The proof is straightforward and we omit it.

If $\mathbf{A} \subseteq$ P(E) is invariant under dilations, then

$$\alpha_{\mathbf{A}}(X) = \cup\{A \in \mathbf{A} | A \subseteq X\} \tag{4.5}$$

defines an opening with Inv($\alpha_{\mathbf{A}}$) = $\mathbf{A}$. In particular, $\alpha_{\mathbf{A}}(X) \in \mathbf{A}$ for every $X \subseteq E$. From (4.5) it follows immediately that $X_A \subseteq \alpha_{\mathbf{A}}(X)$, for $A \in \mathbf{A}, X \subseteq E$, where for $A = \alpha_{\mathbf{A}}(X)$ equality holds. This shows that

$$\alpha_{\mathbf{A}}(X) = \cup_{A\in\mathbf{A}} X_A. \tag{4.6}$$

THEOREM 4.3. *Every opening α can be written as the union of structural openings in the following way*

$$\alpha(X) = \cup_{A\in \mathrm{Inv}(\alpha)} X_A. \tag{4.7}$$

PROOF. Defining $\mathbf{A}$ = Inv(α) we get that $\mathbf{A}$ = Inv(α) = Inv($\alpha_{\mathbf{A}}$), hence $\alpha = \alpha_{\mathbf{A}}$ by Proposition 4.2. Now the result follows immediately from (4.6). □

Since in most applications, Inv(α) is very large, the practical use of (4.7) is rather limited. Fortunately, it is often possible to reduce the number of structuring elements in (4.7) considerably.

PROPOSITION 4.4. *Let α be an opening and let $\mathbf{A}_0 \subseteq P(E)$ be such that $\text{Inv}(\alpha) = \{A \oplus B | A \in \mathbf{A}_0, B \subseteq E\}$, i.e., $\text{Inv}(\alpha)$ is the smallest dilation-invariant family in $P(E)$ which contains $\mathbf{A}_0$. Then*

$$\alpha(X) = \cup_{A \in \mathbf{A}_0} X_A. \tag{4.8}$$

The proof is left to the reader. Note that, because of Proposition 4.1, (4.8) defines an opening for every collection $\mathbf{A}_0$ of structuring elements.

EXAMPLE. The mapping $X \to \text{int}(X)$, where $\text{int}(X)$ denotes the interior of X, defines an opening on $P(\mathbb{R}^d)$. For $\mathbf{A}_0$ we can choose the family of open balls B_r with radius $r > 0$. Then

$$\text{int}(X) = \cup_{r>0} X_{B_r}, \; X \subseteq \mathbb{R}^d.$$

So far, the results stated in this section are due to MATHERON [3]. We now show how to construct openings from arbitrary increasing translation-invariant mappings. We refer to Section 5.7 of [7] for some related results. To our knowledge the results given below are new. Let ψ be an increasing translation-invariant mapping which is anti-extensive. Then $\text{Inv}(\psi)$ is a dilation-invariant subset of $P(E)$. If α is the opening "generated" by $\text{Inv}(\psi)$, i.e. α is the opening with $\text{Inv}(\alpha) = \text{Inv}(\psi)$, then $\alpha \leqslant \psi$. Note that $\alpha = \psi$ if and only if ψ is an opening. Under some extra assumption on ψ, α can be obtained by iteration of ψ. Or, to put it differently : iteration of ψ yields an idempotent mapping. We refer to Theorem 4.5 below for a precise statement. We recall that

$$\psi^\infty = \bigcap_{n \geqslant 1} \psi^n.$$

In order for ψ^∞ to be idempotent it is necessary and sufficient that $\psi \circ \psi^\infty = \psi^\infty$, i.e., $\psi(\cap_{n \geqslant 1} \psi^n(X)) = \cap_{n \geqslant 1} \psi^n(X)$. The following counterexample illustrates that this identily does not hold in general.

Counterexample.
Let $E = \mathbb{Z}$ and let the structuring element A be given by $A = \{..., -7, -5, -3, -1, 2\}$. Define $\psi(X) = (X \oplus A) \cap X$. Then ψ is an increasing, translation-invariant, anti-extensive mapping on $P(\mathbb{Z})$. However, ψ^∞ is not idempotent as we show now. Let $X = \{0, 1, 3, 5, 7, ...\}$. Then (see Figure 7) $\psi(X) = \{0, 3, 5, 7, ...\}$, $\psi^2(X) = \{0, 5, 7, ...\}$, etc. Hence $\psi^\infty(X) = \{0\}$. But $\psi(\psi^\infty(X)) = \psi(\{0\}) = \varnothing$.

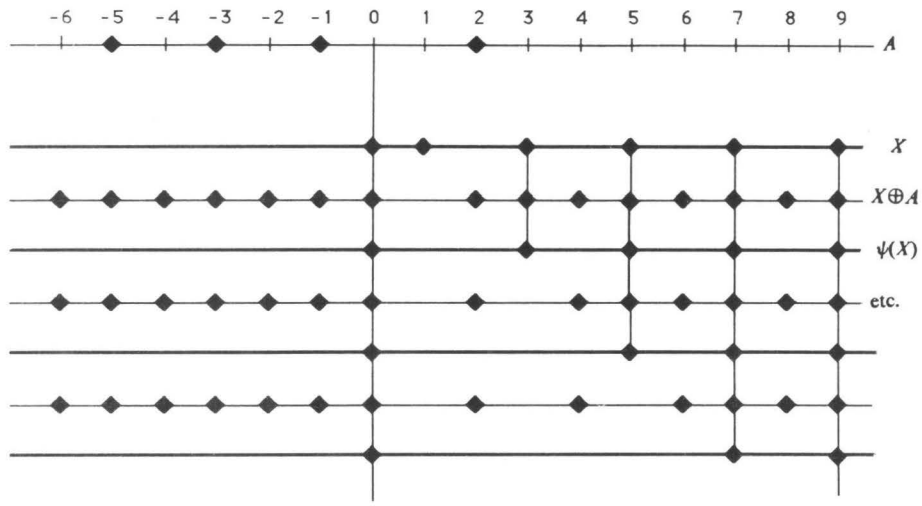


FIGURE 7: $\psi(X)=(X\oplus A)\cap X$, where $A=\{...,-5,-3,-1,2\}$. If $X=\{0,1,3,5,..\}$ then $\psi^n(X)=\{0,2n+1,2n+3,...,\}$. Hence $\psi^\infty(X)=\{0\}$. But $\psi\circ\psi^\infty(X)=\varnothing\neq\psi^\infty(X)$.

In order for ψ^∞ to be idempotent we have to impose an extra condition on ψ. Let $X_n\subseteq E$, $n\in\mathbb{N}$, and $X\subseteq E$. By $X_n\downarrow X$ we mean that

$$...\subseteq X_{n+1}\subseteq X_n\subseteq X_{n-1}\subseteq...\subseteq X_1,$$

and that

$$\cap_{n\geq 1}X_n = X,$$

i.e., X_n is a non-increasing sequence in $\mathrm{P}(E)$ which "converges" to X. The mapping $\psi:\mathrm{P}(E)\to\mathrm{P}(E)$ is called $\downarrow$-*continuous* if $X_n\downarrow X$ implies that $\psi(X_n)\downarrow\psi(X)$ for every non-increasing sequence $\{X_n\}$.

THEOREM 4.5. *Let ψ be an increasing, translation-invariant, anti-extentive mapping which is $\downarrow$-continuous. Then ψ^∞ is an opening (in particular, ψ^∞ is idempotent) and* $\mathrm{Inv}(\psi^\infty) = \mathrm{Inv}(\psi)$.

PROOF. For every $X\subseteq E$ we have by definition that $\psi^n(X)\downarrow\psi^\infty(X)$, so by the $\downarrow$-continuity of ψ we get that $\psi^{n+1}(X)\downarrow\psi(\psi^\infty(X))$. Since the limits must be the same we find that $\psi(X) = \psi(\psi^\infty(X))$. The relation $\mathrm{Inv}(\psi) = \mathrm{Inv}(\psi^\infty)$ follows immediately. □

Since erosions commute with intersections (see (2.6b)) they automatically belong to the class of $\downarrow$-continuous mappings. But this class is actually much larger.

PROPOSITION 4.6. *Compositions, arbitrary intersections, and finite unions of* $\downarrow$*-continuous mappings are* $\downarrow$*-continuous. In particular, if* A *is finite, then the dilation* $X \to X \oplus A$ *is* $\downarrow$*-continuous.*

PROOF. We only prove that the union of a finite number of $\downarrow$-continuous mappings is $\downarrow$-continuous. The other statements are almost trivial.

Let ψ_i be $\downarrow$-continuous for $i = 1,...,p$, and define $\psi = \cup_{i=1}^{p} \psi_i$. We show that ψ is $\downarrow$-continuous. Let $X_n \downarrow X$. Since $X \subseteq X_n$ we have $\psi_i(X) \subseteq \psi_i(X_n)$ $(i = 1,...,p)$ and so $\psi(X) \subseteq \psi(X_n)$. This proves that $\psi(X) \subseteq \cap_{n \geqslant 1} \psi(X_n)$.

To prove the other inclusion, assume that $y \in \cap_{n \geqslant 1} \psi(X_n)$. Thus $y \in \psi(X_n) = \cup_{i=1}^{p} \psi_i(X_n)$, for every $n \geqslant 1$. So there must be some index i $(1 \leqslant i \leqslant p)$ and an infinite subsequence X_{n_k} $(k \geqslant 1)$ such that $y \in \psi_i(X_{n_k})$. But, since $\psi_i(X_{n_k})$ is decreasing, $y \in \psi_i(X_n)$ for all n. Now it follows from the $\downarrow$-continuity of ψ_i that $y \in \cap_{n \geqslant 1} \psi_i(X_n) = \psi_i(X) \subseteq \psi(X)$, and the result is proved. □

Combining these result one ends up with a large class of $\downarrow$ -continuous mappings, in particular if the underlying space E is discrete. The following result is an immediate consequence of Theorem 2.3.

THEOREM 4.7. *Every increasing translation-invariant transformation on* $P(\mathbb{Z}^d)$ *which is locally defined, is* $\downarrow$*-continuous.*

We conclude this section with two examples.

The median opening.
The median filter μ_A of Section 2, with $\#(A)$ finite and odd is a $\downarrow$-continuous mapping since it can be written as a finite union of erosions. So it follows from Theorem 4.5 that $\cap_{n \geqslant 1} (\mu_A \cap id)^n$ is an opening, the median opening. A good understanding of this operation can be achieved by finding its domain of invariance. In the example below (see Figure 8), where $E = \mathbb{Z}^2$ and A is the CROSS, the domain of invariance consists of all objects in which every pixel has at least two 4-neighbours.

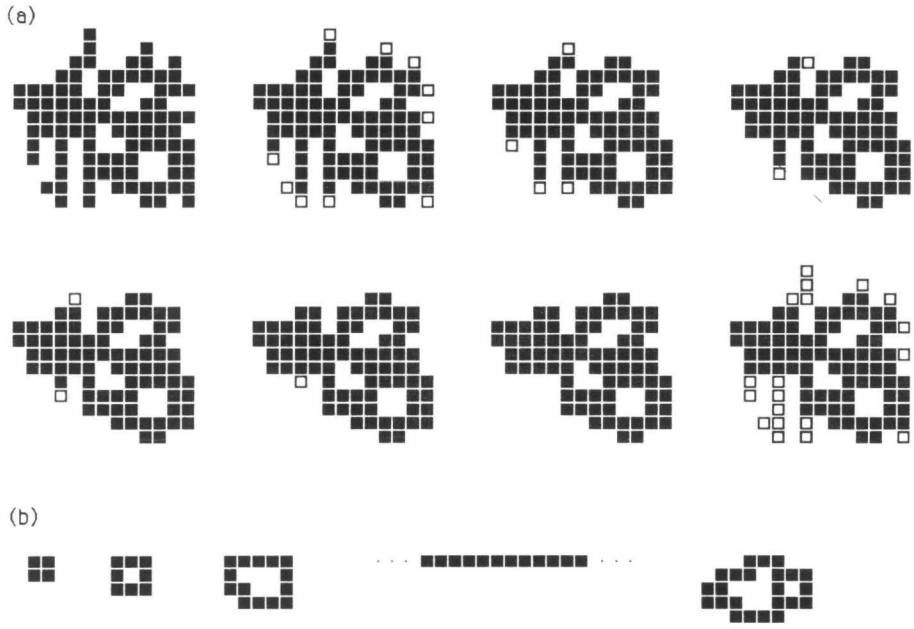


FIGURE 8 a: The median opening $(\mu_A \cap \mathrm{id})^\infty$ where $A = \mathrm{CROSS}$. (a) The first object is the orginal image X, the second object is $\mu_A(X) \cap X$, etc. The one but last object is the median opening of X by A. The last object compares X with its median opening. (b) Some examples of objects which lie in the invariance domain of the median opening.

Closing by iteration.

By duality, all the result mentioned above for openings carry over to closings: if ψ is an increasing, translation-invariant, extensive mapping which is $\uparrow$-continuous, then $\psi_\infty = \cap_{n \geq 1} \psi^n$ is a closing with $\mathrm{Inv}(\psi_\infty) = \mathrm{Inv}(\psi)$. As an example we consider the mapping $\psi(X) = (X \oplus A) \ominus B$ where $B \subseteq A$ and B is finite. Then ψ is extensive since

$$\psi(X) \supset (X \oplus A) \ominus A = X^A \supset X. \qquad (\star)$$

Moreover, ψ is $\uparrow$-continuous and we may therefore conclude that ψ_∞ is a closing with invariant elements $\mathrm{Inv}(\psi)$. It follows from (*) that $X \in \mathrm{Inv}(\psi_\infty)$ implies $X^A = X$. In Figure 9 below we have chosen SQUARE and CROSS for A and B respectively. Then X is invariant under ψ_∞ if and only if X consists of disjoint rectangles which lie "far enough apart". If desired the reader may give a precise description of $\mathrm{Inv}(\psi_\infty)$.

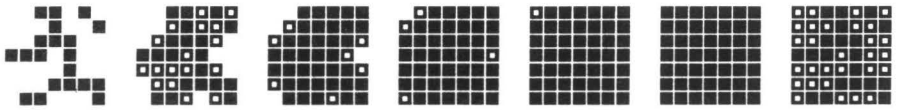


FIGURE 9: Iteration of the mapping $\psi(X)=(X\oplus A)\ominus B$ with $A=$SQUARE and $B=$CROSS. The final result $\psi_\infty(X)$ consists of a family of rectangles which surround the orginal image. The first object is X, the second $\psi(X)$, etc. The one but last object is $\psi_\infty(X)$ reached after 4 iterations. The last object compares X and $\psi_\infty(X)$.

REFERENCES

[1] H.J.A.M. HEIJMANS, C. RONSE (1988). *The algebraic basis of mathematical morphology. Part I: Dilations and erosions,* CWI Report AM-R8807, 1988.

[2] S. LEVIALDI (1971). *Parallel pattern processing,* IEEE Trans. Syst. Man. Cybern SMC-1, 292-296.

[3] G. MATHERON (1975). *Random Sets and Integral Geometry,* J. Wiley & Sons, New York.

[4] K. PRESTON JR., M.J.B. DUFF (1984). *Modern Cellular Automata: Theory and Applications,* Plenum Press, New York and London.

[5] C. RONSE (1985). *Definitions of convexity and convex hulls in digital images,* Bull. Soc. Math. Belg. Ser. B-37, 71-85.

[6] J. SERRA (1982). *Image Analysis and Mathematical Morphology,* Academic Press, London.

[7] J. SERRA, ed. (1988). *Image Analysis and Mathematical Morphology. Volume 2: Theoretical Advances,* Academic Press, London.

Exploring linear chain trickery for physiologically structured populations

Odo Diekmann & Hans Metz
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands
&
Institute of Theoretical Biology
University of Leiden
Groenhovenstraat 5, 2311 BT Leiden, The Netherlands

In this paper we address the following question:
when can one represent the input-output mapping of an infinite dimensional dynamical system corresponding to a physiologically structured population model by means of a finite dimensional system of ordinary differential equations? We concentrate in particular on sufficient conditions in terms of the individual growth and death rates.

Note: this paper is dedicated to H.A. Lauwerier on the occasion of his 65th birthday and in appreciation of all the methods, techniques and tricks we learned from him.

1. Prelude: Simplifying Daphnia (on paper)

Suppose we want to model a population of ectothermic invertebrates, e.g. the water flea Daphnia magna. Experimentally it appears that reproduction depends on the size of the individual animals and this observation motivated Kooijman & Metz (1984) to introduce a size structured model. In Metz & Diekmann (1986; I.3) the assumptions underlying the model are described in some detail, so here we restrict ourselves to a mathematical formulation:

$$\begin{cases} \dfrac{\partial}{\partial t}n(t,l)+\dfrac{\partial}{\partial l}(v(s,l)n(t,l)) = -\mu(s,l)n(t,l) \\ v(s,l_b)n(t,l_b) = \displaystyle\int_{l_b}^{l_{\max}} \beta(s,l)n(t,l)dl \end{cases} \tag{1.1}$$

Here l denotes length and s substrate (more precisely: concentration of algae). The individual growth, death and reproduction rates are denoted by, respectively, v, μ and β. The density n describes the number of Daphnia as well as their distribution with respect to length. All individuals are born with length l_b and $l_{\max}$ is the maximal attainable length under abundant food conditions.

To describe some experiments one should consider s as a given function of

time but to describe others one has to specify the dynamics of s as well. In the latter case we take

$$\frac{ds}{dt} = h(s) - \int_{l_b}^{l_{max}} \gamma(s,l)n(t,l)dl \tag{1.2}$$

where h corresponds to the rate of change of the algae concentration in the absence of daphnids and γ is the per capita consumption rate. Under appropriate assumptions on the ingredients ν, μ, β, γ and h, (1.1) and (1.2) together generate an infinite dimensional nonlinear dynamical system.

Since daphnids are filter feeders it is reasonable to assume that the consumption rate γ is proportional to the surface area which in turn is proportional to l^2. So we put

$$\gamma(s,l) = f(s)l^2. \tag{1.3}$$

If a constant fraction of the ingested energy is allotted to reproduction we may put

$$\beta(s,l) = \alpha f(s)l^2 \tag{1.4}$$

(at this point we deliberately ignore the experimental fact that daphnids don't reproduce if they are still too small; see METZ & DIEKMANN (1986, I.3) for a formulation which does take into account a juvenile period characterized by $l < l_j$). If the rest of the ingested energy is allotted to individual growth and maintenance and if maintenance is proportional to weight, which in turn is proportional to l^3, we may take

$$\frac{d}{dt}l^3 = 3\delta f(s)l^2 - 3\epsilon l^3$$

and therefore

$$\nu(s,l) = \frac{d}{dt}l = \delta f(s) - \epsilon l. \tag{1.5}$$

Finally we take

$$\mu(s,l) = \mu, \quad \text{a constant.} \tag{1.6}$$

We now introduce

$$N_i(t) = \int_{l_b}^{l_{max}} l^i n(t,l)dl, \; i = 0,1,2 \tag{1.7}$$

and find, using (1.1)-(1.7) and some straightforward integrations (by parts), that (N,s) satisfies the *closed* system of ordinary differential equations

$$\begin{aligned}
\frac{dN_0}{dt} &= \alpha f(s)N_2 - \mu N_0 \\
\frac{dN_1}{dt} &= l_b \alpha f(s)N_2 - \delta f(s)N_0 - (\mu+\epsilon)N_1 \\
\frac{dN_2}{dt} &= l_b^2 \alpha f(s)N_2 + 2\delta f(s)N_1 - (\mu+2\epsilon)N_2 \\
\frac{ds}{dt} &= h(s) - f(s)N_2.
\end{aligned} \tag{1.8}$$

The powerful qualitative theory of finite dimensional dynamical systems can now be used to study (1.8) and likewise one can choose from a multitude of well established schemes to study (1.8) numerically (DE ROOS (1988) exploits this fact to investigate the accuracy of the escalator boxcar train, a method to solve the partial differential equation (1.1) numerically).

Of course neglecting the juvenile period has consequences, the main difference between the present model and the full one being that the latter not only allows the occurrence of predator prey oscillations due to the lag in recovery of the food population, but in addition oscillations related to the development lag (see METZ, DE ROOS & VAN DEN BOSCH (1988) or DE ROOS, DIEKMANN & METZ (preprint)).

2. INTRODUCTION: GENERAL MODELS AND TRACTABLE SPECIAL CASES

Every painter needs technical tools to materialize the views he has in mind. Likewise, to practise the art of modelling natural phenomena one needs methods, techniques and tricks. This paper is about a coherent collection of tricks which reduce certain infinite dimensional dynamical systems to finite dimensional ones. In particular the paper concentrates on the applicability of such tricks in the context of physiologically structured population models (METZ & DIEKMANN, 1986).

Before embarking upon the mathematical aspects we want to make some remarks about modelling in general in order to bring the underlying ideas in perspective. The structured population methodology provides first of all a framework for strategic modelling. By this we mean the building of master models which, except in those parts representing the specific mechanistic feature about which we wish to obtain information and understanding, are as general as possible. Within such a framework we can try to prove general theorems and develop general concepts.

If we are lucky the specific mechanism under investigation, by its very nature, allows some sort of model reduction (an example is provided by the use of an equivalent 'age' (:= time elapsed since infection) representation in the Kermack-McKendrick epidemic model, METZ & DIEKMANN, 1986, section IV.4.1) and completely general assertions can rigorously be shown to be valid. Usually, however, it is very difficult, if not impossible, to obtain such strong results. In that case we may start considering simplifications chosen for their mathematical convenience (these we may call tactic models with a strategic

objective as opposed to tactic models with a concrete objective, viz. prediction and/or testing). Here our hope is set on robustness: we expect that the results derived in the context of the simplified models are more or less typical for (some subclass of) the full master model. This hope seems especially justified when the results allow a mechanistic interpretation (an example is provided by the various criteria specifying the qualitative properties of the bifurcation diagram for the cannibalism model of VAN DEN BOSCH, DE ROOS & GABRIEL, to appear.)

As a warning note we recall from section 1 that DE ROOS, DIEKMANN & METZ, preprint, report a certain type of oscillations for the Daphnia model with a juvenile period taken into account for which there is no counterpart in the special case described by (1.8).

Linear chain trickery provides one particular means for carrying out mathematical simplifications. The name arose in the context of delay equations, see MACDONALD, 1978.

3. AN ABSTRACT FORMULATION OF PHYSIOLOGICALLY STRUCTURED POPULATION MODELS

Let the individuals of a population be characterized by finitely many variables, which together we call the *i*-state. So the set of feasible *i*-states Ω is a nice subset of $\mathbb{R}^n$, for some n. At the individual level a model amounts to a specification of (i) the rate of *i*-state change, (ii) the death rate, (iii) the birth rate and in particular how (i), (ii) and (iii) depend on the *i*-state and the prevailing environmental conditions. The latter are described by a, possibly even infinite dimensional, variable E. In the case of the birth rate we have to specify the (distribution of the) state at birth as well.

Once we have a model at the individual level we can immediately derive balance laws doing the necessary bookkeeping. These balance laws generate the time evolution at the population level. There are two types of balance laws, related to each other by duality. The Kolmogorov backward equation is concerned with the clan mean of a continuous function on Ω. The Kolmogorov forward equation describes infinitesimal changes in the measure which assigns to every measurable subset of Ω the number of individuals which have at that instant an *i*-state which belongs to the particular subset. This measure is called the *p*-state (*p* for population) and the space $M(\Omega)$ of regular Borel measures on Ω is called the *p*-state space. Frequently (but not always) we can restrict our attention to densities and formulate the Kolmogorov forward equation for $L_1(\Omega)$.

We can use duality since for E a given function of time, by assumption, the equations are linear. Some environmental variables, like food, are in turn influenced by the population, e.g. through consumption. Thus nonlinear problems come into being by allowing feedbacks through the environment.

Formally at least the Kolmogorov forward equation can be written as

$$\frac{dn}{dt} = A(E)^* n. \tag{3.1}$$

Quite in general we have a decomposition

$$A(E) = A_0(E)+B(E) \tag{3.2}$$

where A_0 describes i-state changes and deaths and B represents births (the importance of this decomposition derives from the fact that for $B = 0$ we can write down explicit solutions by integration along characteristics, i.e. following cohorts). The mathematical theory to provide a rigorous justification and interpretation is still in its infancy. First steps towards a general framework have been made in CLEMENT et al, 1987, 1988, preprints (see DIEKMANN, preprint, for a survey), but much work remains to be done. In the present paper we simply ignore all problems related to existence and uniqueness of solutions and to the precise interpretation of the abstract differential equation. We shall restrict ourselves to formal manipulations.

4. AN ABSTRACT FORMULATION OF LINEAR CHAIN TRICKERY

Let $P: M(\Omega) \to \mathbb{R}^k$ be a mapping such that

$$PA(E)^* = M(E)P \tag{4.1}$$

for some family $M(E)$ of $k \times k$ matrices. Define

$$N(t) = Pn(t) \tag{4.2}$$

then (3.1) implies that N satisfies the ode

$$\frac{dN}{dt} = M(E)N. \tag{4.3}$$

Remark. If we assume that

$$Px^* = <\Phi, x^*> \tag{4.4}$$

for some k vector Φ with components $\phi_i \in X$ (note that this is an assumption since X need not be reflexive) then we can reformulate (4.1) as

$$A(E)\Phi = M(E)\Phi \tag{4.5}$$

provided $\Phi \in D(A(E))$ for all E. Actually $\bigcap_E D(A(E))$ may be empty. However, within the context of dual semigroups one can extend $A(E)$ to an operator $A(E)^{\odot *}$ which has its range in a larger space $X^{\odot *}$ and therefore has larger domain as well (see CLEMENT et al, 1987, 1988, preprints or DIEKMANN, preprint). One can then replace (4.5) by

$$A(E)^{\odot *}\Phi = M(E)\Phi. \tag{4.6}$$

In the following we shall restrict our attention to mappings P of the form (4.4) and we shall not go into the distinction between (4.5) and (4.6) (in fact we shall omit the precise definition of domains of unbounded operators).

Next assume that E is finite dimensional and that feedback through the

environment is incorporated by allowing the equations for E to depend on $C(E)^*n$, where $C(E)^*$ maps $M(\Omega)$ into $\mathbb{R}^m$ for some m (one may call $C(E)^*$ the output mapping). If a family $\Gamma(E)$ of $k \times m$ matrices exists such that

$$C(E)^* = \Gamma(E)P \tag{4.7}$$

then $C(E)^*n = \Gamma(E)Pn = \Gamma(E)N$ and consequently a coupled *finite dimensional* system for N and E describes the interaction completely. Once E is determined by solving this reduced system we can consider

$$\frac{dn}{dt} = A(E)^*n \tag{4.8}$$

as a non-autonomous (i.e. time-dependent) but *linear* equation. A special case of particular importance arises if one can conclude from the (N,E)-system that E approaches a limit (or a periodic solution) as $t \to \infty$. In that case the linear equation for n is asymptotically autonomous (periodic) and one can base further conclusions on the known asymptotic behaviour for these special situations.

Example 1. Consider a cell population with size structure and assume that a mother cell divides into two exactly equal daughter cells. Then the forward equation for densities reads

$$\frac{\partial n}{\partial t}(t,x) = -\frac{\partial}{\partial x}(g(x,E)n(t,x)) - \beta(x,E)n(t,x) + 4\beta(2x,E)n(t,2x) \tag{4.9}$$

(see HEIJMANS, 1984, METZ & DIEKMANN, 1986, sections I.4 and VI.5 and the references given there). Introducing the total biomass operator

$$P\psi = \int_{x_{\min/2}}^{x_{\max}} x\psi(x)dx \tag{4.10}$$

we note that

$$P(-\beta(\cdot,E)\psi(\cdot) + 4\beta(2\cdot,E)\psi(2\cdot)) = 0$$

since mass is conserved in the division process. In other words, $PB(E)^* = 0$. Assuming

$$g(x,E) = h(E)x$$

we get

$$PA_0(E)^*\psi = -\int_{x_{\min/2}}^{x_{\max}} x\frac{\partial}{\partial x}(h(E)x\psi(x))dx = h(E)\int_{x_{\min/2}}^{x_{\max}} x\psi(x)dx = h(E)P\psi$$

(provided $\psi(\frac{1}{2}x_{\min}) = \psi(x_{\max}) = 0$). Combining these two observations we have

$$PA(E)^* = h(E)P.$$

Next interpret E as the concentration of the limiting substrate (so E is one-

dimensional) and assume that the individual growth rate is proportional to the individual nutrient uptake rate. Then

$$C(E)^* n = kh(E)Pn$$

with k a conversion factor. Finally, assume chemostat dynamics then

$$\frac{dN}{dt} = h(E)N - DN$$

$$\frac{dE}{dt} = D(E^i - E) - kh(E)N$$

where D is the dilution rate and E^i is the concentration of the limiting substrate in the inflowing fluid (note that we have now added a term $-Dn(t,x)$ to the right hand side of (4.9)). Under appropriate assumptions on h the ode system has a globally stable steady state. □

Under further conditions one can obtain stronger results. The idea is that one can write down the zero'th generation, i.e. the solution of $\frac{dn_0}{dt} = A_0(E)^* n_0$, explicitly and that n_0 tends to zero as $t \to \infty$ (recall that A_0 incorporates only i-state change and death, no birth; in fact the weaker condition that $B(E)^* n_0(t) \to 0$ as $t \to \infty$ is sufficient for our purposes). So whenever $B(E)^* = \Lambda(E)P$ for some family of mappings $\Lambda(E): \mathbb{R}^k \to M(\Omega)$ one can write down the solution $n(t)$ explicitly in the form of a variation - of - constants formula involving $n_0(t)$ and $N(t)$. Motivated by this observation we first look for $P: M(\Omega) \to \mathbb{R}^k$ and $k \times k$ matrices $H(E)$ such that

$$PA_0(E)^* = H(E)P. \tag{4.11}$$

If then, moreover, $B(E)^* = \Lambda(E)P$ and $C(E)^* = \Gamma(E)P$ the infinite dimensional dynamical system is reduced to a finite dimensional one and, conversely, one can recover the solution of the original system completely from the solution of the ode system by using the (explicit) variation - of - constants formula for the growth-death part $A_0(E)^*$.

5. Sufficient conditions for linear chain trickery

Assume that the i-state space is one dimensional. Then

$$(A_0(E)\phi)(x) = v(x,E)\phi'(x) - \mu(x,E)\phi(x) \tag{5.1}$$

where v is the individual growth rate and μ the per capita death rate.

We first restrict ourselves to the case where P has one-dimensional range. The question then is: under which conditions on v and μ can we find a (continuous) function $\phi(x)$ and a function $\lambda(E)$ such that

$$v(x,E)\phi'(x) - \mu(x,E)\phi(x) = \lambda(E)\phi(x) \quad ? \tag{5.2}$$

If we rewrite (5.2) in the form

$$\frac{\mu(x,E) + \lambda(E)}{v(x,E)} = \frac{\phi'(x)}{\phi(x)} \tag{5.3}$$

we see that a necessary as well as sufficient condition is that for some function $\lambda(E)$ the function $\frac{\mu(x,E)+\lambda(E)}{\nu(x,E)}$ is independent of E (Indeed, if we baptize this function $f(x)$ we can choose $\phi(x) = \exp\int^x f(\xi)d\xi$).

Example 2. Age-dependent population dynamics.
If $\nu(x,E) = 1$ the condition is simply that μ is the sum of a function of x (in this context we prefer to call x a) and a function of E. Note that λ is determined modulo additive constants only and that, as a consequence, we have a one-parameter family of ϕ's. This gives greater flexibility for satisfying the remaining conditions for $B(E)^*$ and $C(E)^*$. □

Example 3. Size-dependent population dynamics.
If, for instance, μ is independent of x we can choose $\lambda(E) = -\mu(E)$ and $f(x) = 0$, hence $\phi(x) = 1$, which simply means that N is total population. The conditions on B and C then require that the per capita contribution to the birth rate as well as the per capita consumption rate are independent of x. □

As a second special case we don't restrict the dimension of the range of P but concentrate on age a as the i-state. Then

$$\Phi'(a) - \mu(a,E)\Phi(a) = H(E)\Phi(a) \tag{5.4}$$

Let us first take some arbitrary but fixed E_0 then

$$\Phi(a) = e^{\int_0^a \mu(\alpha,E_0)d\alpha} e^{H(E_0)a}\Phi(0). \tag{5.5}$$

Substituting this expression into (5.4) we find

$$(\mu(a,E_0) - \mu(a,E))\Phi(a) = (H(E) - H(E_0))\Phi(a) \tag{5.6}$$

or, in other words, $\Phi(a)$ is an eigenvector of $H(E) - H(E_0)$ with corresponding eigenvalue $\mu(a,E_0) - \mu(a,E)$. For fixed E_0 and E the eigenvalues form a discrete set. On the other hand it is reasonable to assume that $a \mapsto \mu(a,E_0) - \mu(a,E)$ is continuous. But a continuous function taking values in a discrete set is constant. So necessarily

$$\mu(a,E) = \mu(a,E_0) + \lambda(E), \quad \text{with } \lambda(E_0) = 0. \tag{5.7}$$

Then we can simply take

$$H(E) = H(E_0) - \lambda(E)I \tag{5.8}$$

and verify that (5.5) implies (5.4) when (5.7) and (5.8) hold. We conclude that in the age-dependent case (5.7) is a necessary and sufficient condition for this part of the linear chain trickery and that we still have a vector $\Phi(0)$ and a matrix $H(E_0)$ at our disposal to satisfy the further conditions imposed by the birth and output operators.

If we try to do the same thing for general one-dimensional i-state x, the

analogue of (5.6) is

$$\left[\frac{v(x,E)}{v(x,E_0)}\mu(x,E_0)-\mu(x,E)\right]\Phi(x) = (H(E)-\frac{v(x,E)}{v(x,E_0)}H(E_0))\Phi(x) \tag{5.9}$$

and our argument breaks down since the matrix at the right hand side is, in general, not independent of x (it is only in the case of physiological age, i.e. $v(x,E) = a(E)f(x)$). The following example shows that this indeed makes an essential difference.

Example 4. Once more size dependent population dynamics.
We still assume that μ is independent of x. We choose $k = 2$ and $\phi_1(x) = 1$, $\phi_2(x) = x$. We want that, for some 2×2 matrix $H(E)$

$$v(x,E)\begin{bmatrix}0\\1\end{bmatrix}-\mu(E)\begin{bmatrix}1\\x\end{bmatrix} = H(E)\begin{bmatrix}1\\x\end{bmatrix} = \begin{bmatrix}h_{11}(E)+h_{12}(E)x\\h_{21}(E)+h_{22}(E)x\end{bmatrix} \tag{5.10}$$

Choosing $h_{12}(E) = 0$ and $h_{11}(E) = \mu(E)$ we have an identity for the first component without any requirement for v. However, the second component requires that v is of the form

$$v(x,E) = a(E)+b(E)x. \tag{5.11}$$

We then choose $h_{21}(E) = a(E)$ and $h_{22}(E) = b(E)-\mu(E)$. So

$$H(E) = \begin{bmatrix}0 & 0\\ a(E) & b(E)\end{bmatrix}-\mu(E)I$$

which is not of the form (5.8). □

Note that we can immediately extend the example to any number of dimensions by choosing

$$\Phi(x) = \begin{bmatrix}1\\ x\\ \frac{1}{2}x^2\\ .\\ .\\ .\end{bmatrix}$$

Moreover, one can employ an E independent change of i-state variable to bring a growth rate v in the form (5.11). For example, the growth laws most commonly encountered in the literature

(i) von Bertalanffy: $\frac{dy}{dt} = \alpha y^{2/3}-\beta y$

(ii) logistic: $\frac{dy}{dt} = \alpha y-\beta y^2$

(iii) Gompertz: $\frac{dy}{dt} = \alpha y-\beta y\log y$

can all be linearized:

$$\text{(i)} \quad x = y^{1/3} \Rightarrow \quad \frac{dx}{dt} = \frac{1}{3}(\alpha - \beta x)$$

$$\text{(ii)} \quad x = \frac{1}{y} \Rightarrow \quad \frac{dx}{dt} = \beta - \alpha x$$

$$\text{(iii)} \quad x = \log y \Rightarrow \quad \frac{dx}{dt} = \alpha - \beta x$$

(we thank Y. Iwasa for bringing (ii) and (iii) to our attention).

Modulo such changes of i-state variable Example 4 is the only example we know in which a nontrivial higher dimensional Φ occurs. So far our attempts to formulate and demonstrate *necessary* conditions (perhaps under some a priori restrictions) which would imply that there are no other examples have not been successful.

In conclusion of this section we present one example with a higher dimensional i-state space. The analogue of (5.1) is

$$(A_0(E)\phi)(x) = v(x,E).\nabla\phi(x) - \mu(x,E)\phi(x) \tag{5.12}$$

or, in words: $v\phi'$ is replaced by the directional derivative in the direction of the vector field v.

Example 5. Consider two-dimensional x and let v be given by

$$v(x,E) = \begin{bmatrix} a(E) + b(E)x_1 \\ c(E) \end{bmatrix}$$

Define

$$\Phi(x) = \begin{bmatrix} 1 \\ x_1 \\ x_1^2 \\ e^{-kx_2} \\ x_1 e^{-kx_2} \\ x_1^2 e^{-kx_2} \end{bmatrix}$$

and

$$H(E) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ a & b & 0 & 0 & 0 & 0 \\ 0 & 2a & 2b & 0 & 0 & 0 \\ 0 & 0 & 0 & -kc & 0 & 0 \\ 0 & 0 & 0 & a & (b-kc) & 0 \\ 0 & 0 & 0 & 0 & 2a & 2b-kc \end{bmatrix}$$

A straightforward calculation then shows that

$$v(x,E)\cdot\nabla\Phi(x) = H(E)\Phi(x)$$

which is the required relation for $\mu = 0$. When μ is non-zero but still independent of x we can redefine $H(E)$ by subtracting $-\mu(E)I$.

The biological interest of this example is that we may interpret x_1 as size and x_2 as physiological age. Moreover Φ is chosen in such a way that we can choose

$$\beta(x,E) = f(E)(1-e^{-kx_2})x_1^2$$

as an age and size dependent birth rate of individuals.

6. Discussion

Understanding the precise nature of necessary and sufficient conditions for linear chain trickery may be important for two reasons. First of all it should yield a *complete* catalogue of cases in which a reduction to finite dimensions is possible. No doubt this catalogue will contain useful cases which thus far escaped our attention (like the case from Example 5). Secondly it will show which (classical) ode models can be reinterpreted as reduced structured models (In our, admittedly somewhat biased, opinion the justification of any ode model should derive from the fact that such an interpretation is possible).

In this paper we have made a first step towards the derivation of necessary and sufficient conditions but we are still far from reaching our final goal. However, we plan to keep on working at it.

References

F. van den Bosch, A.M. de Roos & W. Gabriel, *Cannibalism as a life boat mechanism*, to appear in J. Math. Biol.

Ph. Clement, O. Diekmann, M. Gyllenberg, H.J.A.M. Heijmans & H.R. Thieme, *Perturbation theory for dual semigroups.*
I. *The sun-reflexive case*, Math. Ann. **277** (1987) 709-725
II. *Time-dependent perturbations in the sun-reflexive case*, Proc. Roy. Soc. Edinburgh **109 A** (1988) 145-172.
III. *Nonlinear Lipschitz continuous perturbations in the sun-reflexive case.* Proceedings of the meeting 'Volterra Integro-Differential Equations in Banach Spaces and Applications' Trento, to appear.
IV. *The intertwining formula and the canonical pairing*, Proceedings of the meeting 'Trends in Semigroup Theory and Applications' Trieste, to appear.

O. Diekmann, *On semigroups and populations.* Proceedings of the meeting 'Advanced Topics in the Theory of Dynamical Systems' Trento, to appear.

H.J.A.M. Heijmans (1984), *On the stable size distribution of populations reproducing by fission into two unequal parts*, Math. Biosc. **72:** 19-50.

S.A.L.M. Kooijman & J.A.J. Metz (1984), *On the dynamics of chemically stressed populations: the deduction of population consequences from effects on individuals*, Ecotox. Env. Saf. **8:** 254-274.

J.A.J. Metz & O. Diekmann (eds) (1986), *The Dynamics of Physiologically Structured Populations*, Springer Lect. Notes in Biomath. **68.**

J.A.J. Metz, A.M. de Roos & F. van den Bosch (1988), *Population models*

incorporating physiological structure: a quick survey of the basic concepts and an application to size-structured population dynamics in waterfleas. In: B. Ebenman & L. Persson, eds., Size-structured Populations: Ecology and Evolution, Springer, 106-124.

A.M. DE ROOS (1988), *Numerical methods for structured population models: the escalator boxcar train*, Num. Meth PDE **4:** 173-195.

A.M. DE ROOS, O. DIEKMANN & J.A.J. METZ (preprint), *The escalator boxcar train: basic theory and an application to Daphnia population dynamics.*

Publikaties H.A. Lauwerier

Axiomatische onderzoekingen over de vlakke meetkunde, (1948) Dissertatie TH Delft.

Colloquium asymptotische ontwikkelingen 1947-1950, 3f; Confluent hypergeometric functions, (1949) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TC 12f/49.

Mathematisch problemen uit de warmtegeleidingstheorie, Cursus 1954/55 (1955) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TC 24/55. Syllabus cursus 1954.

De invloed van wind op zeestroming, (1954) Voorlopig rapport, 1, Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TD 1/54.

Enige randproblemen met een scheve afgeleide, (1954) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 29/54 Mathematisch Centrum / Actualiteiten * Voordracht 27.11.1954 ; ZW 1954-016.

Note on a problem of S. Goldstein, (1955) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 30/55.

The motion of a shallow sea under influence of a non-stationary wind-field, (1955) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 31/55

The motion of a half-plane sea under influence of a non-stationary wind, (1955) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 32/55.

The expansion of a function into a Fourier series with prescribed phases, valid in the half-period interval, (1955) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 33/55.

The North Sea problem, 6; non stationary wind-effects in a rectangular bay; theoretical part, (1955) (Verschenen in: Proceedings KNAW Series A, 64(1961), nr 1, Indagationes Mathematicae, 23 (1961), p 104-122), Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 55/55/R.

Integraalvergelijkingen, Cursus voorjaar 1956, (1956) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TC 27/56. Syllabus cursus voorjaar 1956.

met D. J. Hofsommer & B.R. Damsté, *The influence of a disturbance upon an infinitely large shallow sea of constant depth,* (1956) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 35/56.

The wind-effect in the southern parts of the North Sea due to a single storm and the influence of the Channel, (1956) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 36/56.

Diffusion from a point source into a space bounded by an impenetrable plane, (1956) (Verschenen in: Applied scientific research a, 6 (1956), p 197-204), Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 38/56/R.

Alcuni aspetti della matematica del mare del nord, (1956) (Verschenen in: Rendiconti dell'academia nazionale dei Lincei, classe di scienze fisiche, matematiche e naturali 8, 19 (1955), nr 3-4, p 116-124), Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TN 2/56/R.

On the non-stationary behaviour of a rotating shallow sea under influence of a windfield, (1956) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TN 5/56.

met C. G. Lekkerkerker, *Functietheorie; speciale onderwerpen, Theorie en toepassingen der Laplacetransformatie,* (1956) Mathematisch Centrum, Amsterdam, Zuivere Wiskunde, ZC 37/56. Syllabus cursus Sittard 1956; zie ZC 36.

A note on the formula of Plemelj, (1957) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 39/57.

The influence of a homogeneous wind upon an infinitely wide North Sea, (1957) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 41/57.

Exponential wind-fields, (1957) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 42/57.

On certain trigonometrical expansions, (1957) (Verschenen in: Journal of mathematics and mechanics, 8 (1959), p 419-432), Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 43/57/R.

The surface charge of a semi-infinite cylinder due to an axial point charge, (1958) (Verschenen in: Applied scientific research b, 8 (1960), p 277-289), Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 44/58/R.

A note on flexible hexagons, (1958) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 45/58.

Free motions in a rotating sea which has the form of a semi-infinite strip, (1958) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 46/58.

A uniform wind-field on a rotating sea in presence of a semi-infinite barrier, (1958) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 50/58.

The influence of some uniform wind-fields upon a sea having the form of a semi-infinite strip, (1958) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 52/58.

A diffusion problem with chemical reactions, (1958) (Verschenen in: Applied scientific research a, 8 (1959), p 366-376), Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 54/58/R.

Der Einfluss eines homogenen zeit-abhängigen Windfeldes auf ein rotierendes Meer mit einem halb-unendlichen Hindernis, (1958) (Verschenen in: Zeitschrift für angewandte Mathematik und Mechanik, 38 (1958), nr 7-8, p 1-2), Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TN 9/58/R.

Solutions of the equation of Helmholtz in an angle, 1; introduction, the F-problem, (1959) (Verschenen in: Proceedings KNAW Series A, 62(1959), nr 5, Indagationes Mathematicae, 21 (1959), p 475-488), Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 61/59/R.

met D. van Dantzig, *The North Sea problem, 1; general considerations concerning the hydrodynamical problem of the motion of the North Sea,* (1959) (Verschenen in: Proceedings KNAW Series A, 63(1960), nr 2, Indagationes Mathematicae, 22 (1960), p 170-180), Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 62/59/R.

Asymptotic expansion of Stirling numbers, (1959) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TN 13/59.

met D.J. Hofsommer & B.R. Damsté, *The influence of an exponential wind-field upon a semi-circular bay,* (1959) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 56/59.

Warmte-overdracht in een gietblok, (1959) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 63/59.

Het Noordzeeprobleem, (1960) Rede uitgesproken bij de aanvaarding van het ambt van bijzonder hoogleraar in de toegepaste wiskunde aan de Universiteit van Amsterdam.

met D. van Dantzig, *The North Sea problem, 4; free oscillations of a rotating rectangular sea,* (1960) (Verschenen in: Proceedings KNAW Series A, 63 (1960), nr 4, Indagationes Mathematicae, 22 (1960), p 339-354), Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 64/60/R.

The North Sea problem, 5; free motions of a rotating rectangular bay, (1960) (Verschenen in: Proceedings KNAW Series A, 63 (1960), nr 5, Indagationes Mathematicae, 22 (1960), p 423-438), Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 65/60/R.

Solutions of the equation of Helmholtz in an angle, 2; the G-problem, (1960) (Verschenen in: Proceedings KNAW Series A, 63 (1960), nr 4, Indagationes Mathematicae, 22 (1960), p 355-372), Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 66/60/R.

Solutions of the equation of Helmholtz in an angle, 3; the case of a half-plane, (1960) (Verschenen in: Proceedings KNAW Series A, 64 (1961), nr 1, Indagationes Mathematicae, 23 (1961), p 123-140), Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 67/60/R.

Randwaardeproblemen uit de mathematische fysica, (1960) College toegepaste wiskunde 1959/60 Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TC 33/60, College-diktaat 1959.

The generalized potential of an ellipsoid, (1960) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TN 14/60.

A note on the zeta-function of Riemann-Hurwitz, (1960) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TN 15/60.

A note on a paper by H. Levine, (1960) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TN 16/60.

The North Sea problem, 2; influence of a stationary wind-field upon a bay with a uniform depth, (1960) (Verschenen in: Proceedings KNAW Series A, 63 (1960), nr 3, Indagationes Mathematicae, 22 (1960), p 266-278), Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 57/60/R.

The North Sea problem, 3; influence of a stationary wind-field upon a bay with an exponentially increasing depth, (1960) (Verschenen in: Proceedings KNAW Series A, 63(1960), nr 3, Indagationes Mathematicae, 22 (1960), p 279-290), Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 58/60/R.

Perturbation of orbits of earth-satellites by atmospheric contact, (1960) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TN 17/60, ZW 1960-005.

Niet-lineaire trillingen, College 1960/61 (1961) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TC 34/61 College-diktaat 1960.

Laplacetransformaties, College toegepaste wiskunde 1960/61 (1961) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TC 35/61, College-diktaat 1960.

Distributies, College toegepaste wiskunde 1960/61 speciale onderwerpen (1961) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TC 36/61, College-diktaat 1960.

Colloquium gegeneraliseerde functies, Distributies (1961) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TC 37/61, Syllabus colloquium 1961.

Solutions of the equation of Helmholtz in an angle, 4; the F- and G-problem in boundary conditions, (1961) (Verschenen in: Proceedings KNAW Series A, 64 (1961), nr 3, Indagationes Mathematicae, 22 (1961), p 348-359), Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 70/61/R.

The North Sea problem, 7; non-stationary wind-effects in a rectangular bay, numerical part, (1961) (Verschenen in: Proceedings KNAW Series A, 64 (1961), nr 4, Indagationes Mathematicae, 23 (1961), p 418-431), Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 73/61/R.

Some recent work of the Amsterdam Mathematical Centre on the hydrodynamics of the North Sea, (1961) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 74/61.

Solutions of the equation of Helmholtz in an angle, 6; the case of a semi-infinite barrier, (1961) (Verschenen in: Proceedings KNAW Series A, 65 (1962), nr 4, Indagationes Mathematicae, 24 (1962), p 473-483), Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 75/61/R.

Topological foundations of the theory of distributions, 1, (1961) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 80/61.

Topological foundations of the theory of distributions, 2, (1961) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 81/61.

Inleiding tot de methoden van de mathematische fysica, (1962) College toegepaste wiskunde 1961/62, Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TC 38/62, College-diktaat 1961.

Besselfuncties, (1962) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TN 21/62.

A note on the summation of some series of Bessel functions, (1962) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TN 22/62.

A note on the effect of a return surge, (1962) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TN 23/62.

Solutions of the equation of Helmholtz in an angle, 5; the case of a half-plane in boundary conditions, (1962) (Verschenen in: Proceedings KNAW Series A, 65 (1962), nr 1, Indagationes Mathematicae, 24 (1962), p 93-99), Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 71/62/R.

The analytical treatment of storm surges, (1962) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 79/62.

The theory of distributions (1962) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 82/62.

The Hilbert problem for h generalized functions, (1962) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 83/62.

met B. R. Damsté, *The North Sea problem, 8; a numerical treatment,* (1962) (Verschenen in: Proceedings KNAW Series A, 66 (1963), nr 2, Indagationes Mathematicae, 25 (1963), p 167-184), Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 84/62/R.

The diffraction of a cylindrical pulse by a half-plane, (1962) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 86/62.

The asymptotic expansion of the statistical distribution of N.V. Smirnov, (1962) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 90/62.

Partiële differentiaalvergelijkingen van het hyperbolische type, Karakteristiekenmethodes. College speciale onderwerpen 1962/63 (1963) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TC 40/63, College-diktaat 1962.

Inleiding tot de vector- en tensorrekening, College toegepaste wiskunde 1962/63 (1963) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TC 41/63. College-diktaat 1962.

met J.F. Frankena, *Capita selecta uit de theorie der lineaire operatoren,* (1964) College toegepaste wiskunde 1963/64, Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TC 43/64, College-diktaat 1963.

met S.A. Wouthuysen, *Colloquium representatie van de draaiingsgroep en de Lorentzgroep,* 1963/64 (1964) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TC 44/6, Colloquium 1963.

samen met W. Eckhaus, *Partiële differentiaalvergelijkingen met kleine parameter, 1,* Colloquium 1964/65 (1965) W. Eckhaus, H. A. Lauwerier, Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TC 45.1/65, Colloquium 1964.

samen met W. Eckhaus, *Partiële differentiaalvergelijkingen met kleine parameter, 2,* Colloquium 1964/65 (1965) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TC 45.2/65, Colloquium 1964.

The deformation of a square wave by a linear filter, (1965) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TN 39/65.

A note on Landen's transformation (1965) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TN 44/65.

met M. P. van Ouwerkerk-Dijkers & J. Nuis, *Discussion of some mathematical aspects of an anodic stripping voltammetry problem,* (1966) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 103/66.

Asymptotic analysis, MC Tract 13, Mathematisch Centrum, Amsterdam, 1966.

Randwaarde problemen, 1,2,3, MC Syllabus 2,3,4, Mathematisch Centrum, Amsterdam, 1967-1968.

Syllabus naar het college speciale onderwerpen Brown'se beweging en gemengde randwaardeproblemen, Univ. van Amsterdam, Amsterdam, 1968.

Representaties van groepen, MC Syllabus 4, Mathematisch Centrum, Amsterdam, 1968.

Niet-lineaire differentiaalvergelijkingen, Colloquium 1968 (1968) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TC 47/68, Colloquium 1968.

met H. Bavinck & J. Grasman, *Relaxatietrillingen,* (1968) Voordracht colloquium 1968: niet-lineaire differentiaalvergelijkingen H. A. Lauwerier, Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TC 47a/68, Supplement TC 47.

Syllabus vector- en tensorrekening; appendix 2, (1969) Samenvatting algemene tensorrekening, Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TC 49/69, Colleges 1968; App. TC 41, colleges 1962.

Toepassingen van de functietheorie, Univ. van Amsterdam, Amsterdam, 1970.

Elasticiteitstheorie, Univ. van Amsterdam, Amsterdam, 1971.

Gewone differentiaalvergelijkingen, Univ. van Amsterdam, Amsterdam, 1971.

Vector- en tensorrekening, Univ. van Amsterdam, Amsterdam, 1971.

A note on the van Wijngaarden transformation, (1972). (Verschenen in SIAM J. Math. Anal., 6 (1975) p 96-104), Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 136/72.

A limit case of a Volterra-Lotka system, (1975) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TN 79/75.

Polymerisatie en gelvorming, (1976) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TN 83/76.

Polycondensation and gelation of trifunctional monomers, (1976) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 157/76.

Polycondensation and gelation: the general case, (1976) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 160/76.

Technieken van de mathematische fysica, Univ. van Amsterdam, Amsterdam, 1977.

Asymptotic Analysis, MC Tract 54, Mathemstisch Centrum, Amsterdam, 1977.

A class of problems in hydrodynamics, (1978) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 175/78.

Vraagstukken toegepaste analyse, Univ. van Amsterdam, Amsterdam, 1980.

Inleiding partiële differentiaalvergelijkingen, Univ. van Amsterdam, Amsterdam, 1980.

Calculus of variations in mathematical physics, MC Tract 14, Mathematisch Centrum, Amsterdam, 1980.

College niet-lineaire differentiaalvergelijkingen Univ. van Amsterdam, Amsterdam, 1981.

Mathematical models of epidemics, MC Tract 138, Mathematisch Centrum, Amsterdam, 1981.

Entire functions for the logistic map, deel I, (1982) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 228/82, CWI. Department of Applied Mathematics.

Growth, fission and the stable size distribution, (1983) (O. Diekmann, H. A. Lauwerier, T. Aldenberg, J.A.J. Metz, (Verschenen in J. Math. Biol. 18 (1983) p 135-148), Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 235/83.

The parametrisation of the unstable invariant manifold for a class of horseshoe maps, (1983) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 237/83.

Chaos and order, (1983) (Verschenen in Nieuw Archief voor Wiskunde IV, 2 (1984) p 373-401), Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 243/83.

Bifurcation of a map at resonance 1:4, (1983) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 245/83.

Local bifurcation of a logistic delay map, (1983) Mathematisch Centrum, Amsterdam, Toegepaste Wiskunde, TW 246/83.

Entire functions for the logistic map, deel II, (1984) CWI, Department of Applied Mathematics; R 8404.

Global bifurcation of a logistic delay map, (1984) CWI. Department of Applied Mathematics, R 8402.

A case of a not so strange strange attractor, (1984) CWI. Department of Applied Mathematics, R 8406.

Dynamical systems and numerical integration, (1984) CWI. Department of Applied Mathematics, R 8413.

De bloeiende boom van Pythagoras, (1984) CWI. Department of Applied Mathematics, N 8403.

met M. B. van der Mark, *Chaos and order in an optical ring cavity,* (1984) CWI. Department of Applied Mathematics, R 8416.

The Pythagoras tree as a Julia set, (1985) (Verschenen in CWI Newsletter, 6 (1985) p 2-18), CWI. Department of Applied Mathematics, R 8503.

Hopf bifurcation in host-parasitoid models, (1985) (Verschenen (met J.A.J. Metz) in IMA J. Math. Appl. Med. Biol. 3 (1986) p. 191-210), CWI. Department of

Applied Mathematics, R 8516.

The structure of a strange attractor, Physica 21D (1986) p. 146-154.

met J. A. Kaandorp, *Fractals (mathematics, programming and applications,* (1987) CWI. Department of Computer Science, R 8762. 1968.

Fractals, Aramith, Amsterdam, 1987.

Analyse met de microcomputer, Epsilon, Utrecht, 1987.

Meetkunde met de microcomputer, Epsilon, Utrecht, 1987.

Symmetrie, Aramith, Amsterdam, 1988.