Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

Quarterly, Issue no. 4
September 1984

CWI NEWSLETTER

Number 4, September 1984

Editors

Arjeh M. Cohen   Richard D. Gill   Jan Heering

The CWI Newsletter is published quarterly by the Centre for Mathematics and Computer Science (Centrum voor Wiskunde en Informatica), Kruislaan 413, 1098 SJ Amsterdam, The Netherlands. The Newsletter will report on activities being conducted at the Centre and will also contain articles of general interest in the fields of Mathematics and Computer Science, including book reviews and mathematical entertainment. The editors encourage persons outside and in the Centre to contribute to the Newsletter. Normal referee procedures will apply to all articles submitted.

The Newsletter is available free of charge to all interested persons. The Newsletter is available to libraries on an exchange basis.

All correspondence should be addressed to: *The CWI Newsletter, Centre for Mathematics and Computer Science, Kruislaan 413, 1098 SJ Amsterdam, The Netherlands.*

ISSN 0168-826x

Contents

CWI

Centre for Mathematics and Computer Science

Centrum voor Wiskunde en Informatica

STOCHASTIC GEOMETRY AND IMAGE ANALYSIS*

by

A.J. Baddeley

School of Mathematics, University of Bath,
Claverton Down, Bath, BA2 7AY,
U.K.

Summary

We list recent ideas in stochastic geometry which are closely related to image analysis. These include the synthesis of stochastic models of images, techniques for evaluating models and algorithms, general concepts of 'geometrical information' and the theory of random sets, problems of image irregularity and errors in observation, techniques of geometric integration theory, and fractional dimensional irregularity.

1. Introduction

The development of computerized image processing and image analysis already seems to have prompted considerable study of the relations between geometry, probability theory and computer science. Rosenfeld [29, preface] observes that all image processing algorithms must be based explicitly or implicitly on mathematical models of the images to be processed. Some of the newer stochastic image models presented in [29] are based on Markov processes, random fields, random mosaics (tessellations) and stochastic grammars. Apart from image modeling, we imagine other mathematical contributions should include a theoretical background for the comparison of algorithms, and mathematical techniques for the treatment of image models.

Independently of such requirements, many concepts related to image analysis have evolved in other areas, notably in stochastic geometry, stereology and geometric integration theory. *Stochastic geometry* is that part of probability theory dealing with random subsets of a geometrical space, and interactions between probability and geometry. This includes all stochastic image models, at least in principle, but some frequently studied models are: elementary constructions of random lines, circles or triangles; spatial schemes such as random mosaics and random coverings of the plane; and general random processes and random sets. The main body of theory concentrates on *uniformly random* models, for which there are simple explicit solutions. However, the last decade

* This paper is a contribution to the 'Proceedings of the CWI symposium on mathematics and computer science', CWI monograph 1.

has seen the introduction of more flexible techniques and a completely general theoretical foundation for random sets.

This paper summarizes some recent work in stochastic geometry (drawing also on stereology and geometric integration theory) which could be connected with image analysis. Section 2 introduces the range of random image models in stochastic geometry, and outlines the classical theory of uniformly random models. The more recent combinatorial theory (section 3) has an application to problems of image complexity. Section 4 discusses the Kendall-Matheron abstract theory of random sets, which has many similarities to tenets of image analysis. J. Serra's mathematical morphology and image analysis theory is touched upon in Section 5. Recent thoughts about image irregularity and observation errors (Section 6) are developed using geometric integration theory. Finally Section 7 speculates on the usefulness of fractal (fractional dimensional) models of image irregularity.

2. Classical stochastic geometry

Detailed surveys of stochastic geometry can be consulted in the literature [24, 3, 7, 32, 35] and we shall give here a very brief sketch. Probability models available for generating random geometrical objects (hence random image models) can be classified as

(a) elementary constructions;
(b) stochastic processes;
(c) theory of random sets.

(a) Elementary constructions are the simple geometrical figures of Euclid with an added component of randomness, as for example the output of a computer graphics program when the input is a random number generator. Points, lines, triangles, circles and other figures are determined by $n<\infty$ real parameters so that a random figure can be defined as a probability distribution on the n-dimensional parameter space. Of course we may also construct the random line joining two random points, and so on. Using parametrisations of the rotation and translation groups we may generate random positions of an arbitrary object. Typical problems include finding the probability that two random figures (or a random figure and a fixed figure) will intersect; the mean area of length of overlap between figures; and the probability that N random figures will completely cover a specified region.

Even the simplest problems for random figures lead to difficult multiple integrals. An exception to this rule is that *uniformly distributed* random figures often lead to simple explicit solutions. For example, a random two-dimensional point $X = (x_1, x_2)$ is a *uniformly random* (UR) point in the region $A \subset \mathbb{R}^2$ if it has constant probability density $f(x_1, x_2) = K$. The constant must be $K = 1/\text{area}(A)$ since probability integrates to 1. For any measurable subset $B \subset A$ we find the probability

$$P(X \text{ falls in } B) = \frac{\text{area}(B)}{\text{area}(A)}, \tag{1}$$

which is what we understand by a 'simple explicit solution'. Now consider a random circle $C(X,r)$ of fixed radius r obtained by randomizing the centre point X. Let X be a uniformly random point in the disc D_{R+r} of radius $R+r$ and centre 0. Then the circle $C(X,r)$ always intersects D_R, the disc of radius R about 0. We say $C(X,r)$ is a uniformly random circle hitting D_R. For any (fixed) point $x \in D_R$,

$$P(C(X,r) \text{ contains } x) = P(X \text{ falls in } C(x,r)) = \frac{\pi r^2}{\pi(R+r)^2}$$

by (1), which does not depend on x. Furthermore, the mean or expected area of overlap between $C(X,r)$ and D_R is by Fubini's theorem

$$\mathbb{E}(\text{area } C(X,r) \cap D_R) = \int_{D_R} P(x \text{ lies in } C(X,r))dx$$

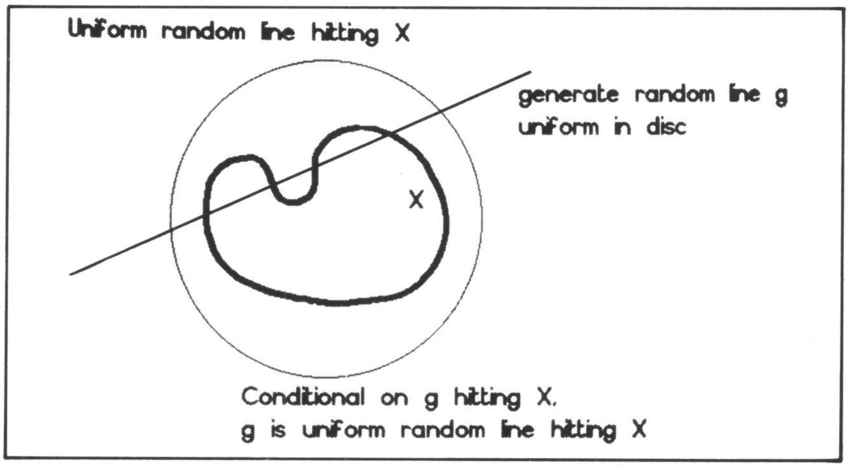
$$= \pi R^2 \frac{r^2}{(R+r)^2},$$

i.e. proportional to the product of areas of $C(X,r)$ and D_R.

Definition of a uniformly random line is less intuitive. Let parameters (p,θ) specify the line

$$\{(x,y): x\cos\theta + y\sin\theta = p\},$$

i.e. $|p|$ is the distance of the line from the origin, and θ determines its direction. A *uniformly random* (UR) line is such that (p,θ) is a uniformly distributed point in some bounded region of $\mathbb{R} \times [0,\pi)$. For example a UR line hitting the disc D_R is obtained when p and θ are independent random variables uniformly distributed over $[-r,+r]$ and $[0,\pi)$ respectively. In general for $X \subset \mathbb{R}^2$ the set of lines intersecting X is some irregular set of (p,θ) points in the allowable region. To generate a UR line hitting X, in practice, find a disc D_R circumscribing X. Generate a UR line L hitting D_R; if $L \cap X = \varnothing$, reject this attempt and generate another line L; until L hits X. Then L is UR hitting X.



Uniform random lines have the invariance property that if L is a UR line hitting X, and if $Y \subset X$, then the probability $P(L \text{ hits } Y)$ does not depend on the position or orientation of Y within X. All parts of X are equally likely to be 'sampled' by L. This fair sampling property, which characterizes the uniform distribution, can be recognised as invariance under the euclidean group of rigid motions. Another nice characterization of UR lines is based on the two-person game where A 'hides' a set Y inside X and player B draws a line L through X to find Y. Optimal strategy for B is to generate a uniform random line.

We state two fundamental results concerning UR lines. Let L be a UR line through X, a bounded measurable plane set. If $A \subset X$ is measurable then

$$\mathbb{E}\ \text{length}\ (L \cap A) = \frac{\pi.\text{area}(A)}{K} \tag{2}$$

where $\mathbb{E}$ again denotes expected (mean) value, and K is a constant depending on X. If $C \subset X$ is a plane curve then

$$\mathbb{E} n(L \cap C) = \frac{2.\text{length}(C)}{K} \tag{3}$$

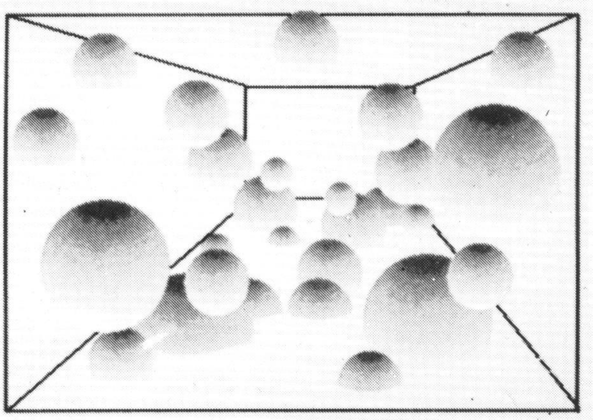
where $n(L \cap C)$ is the number of intersection points between L and C. Thus, the mean amount of overlap between a UR line and a fixed figure is given by (2), (3) *regardless* of the geometrical configuration of the figure. This generality is the basis of the classical theory. Corresponding formulae hold in higher dimensions and noneuclidean spaces [30].

Apart from the obvious application of (2)-(3) to stochastic image models, we can interpret them to give methods for measurement of length and area. If an image consists of several curves, their total length can be statistically estimated by randomly rotating the image, superimposing a grid of parallel lines and counting the number of crossing points.

Statements about image complexity also follow from (2)-(3). Suppose the image consists of curves of total length l, the screen is divided into an $n \times n$ square grid, and we wish to estimate the number of grid squares which contain part of the image. Assuming the image and grid are randomly superimposed, the mean number of grid-image intersections is $4(n-1)l$. For large n this approximates the mean number of *squares* crossed, i.e. the mean complexity.

Stochastic image models may also be based on (b) *stochastic processes.* To generate a random pattern extending over the entire plane, divide $\mathbb{R}^2$ into squares, and place a random number of random points in each square. A random pattern of lines is a random pattern of (p,θ) points in $\mathbb{R} \times [0,\pi)$, and so on. Thus we define a *random point process* in space S as a random locally finite set of points in S, where 'locally finite' means each bounded region of S only contains a finite (random) number of (random) points. *A random line process* 'is' a random point process in $\mathbb{R} \times [0,\pi)$, or more intrinsically, is a random locally finite set of lines in $\mathbb{R}^2$. In calculations one uses the correspondence between a random point process and the system of random variables

$N(A)$ = (number of points in A), $A \subset S$, which constitute a *random measure* $N(\cdot)$ on S. A random line process is a random measure on $\mathbb{R} \times [0,\pi)$, or intrinsically, corresponds to a random capacity function $H(A)$ = (number of lines *intersecting* A), $A \subset \mathbb{R}^2$. See [18,12].



A finite population of spheres in $\mathbb{R}^3$
or
a realisation of a random sphere process

Explicit calculations are usually unsuccessful except for *uniform Poisson processes,* in which each bounded part of the process consist of independent uniformly random points/lines, and $N(A)$, $N(B)$ are independent when $A \cap B = \emptyset$. Equations (1)-(3) yield the expected values of $N(A)$, $H(A)$, the number of crossings of a fixed curve, the total length of lines overlapping A, and the number of line-line crossings inside A.

General random point processes and line processes have been studied using moments [12,19,32] and Palm probabilities [26]. For a point process the first two moment measures are the intensity measure $\mu(A) = \mathbb{E}[N(A)]$ on $\mathbb{R}^2$, and the second moment measure $\mu^{(2)}$ on $\mathbb{R}^2 \times \mathbb{R}^2$ defined by $\mu^{(2)}(A \times B) = \mathbb{E}[N(A)N(B)]$, which together contain all variance-covariance information. If the process is statistically stationary, then $\mu(A) = \lambda.\text{area}(A)$ where $\lambda > 0$ is the intensity, while $\mu^{(2)}$ 'disintegrates',

$$d\mu^{(2)}(x,y) = d\gamma(y-x)d\mu(x) \quad x,y \in \mathbb{R}^2$$

and the measure γ on $\mathbb{R}^2$ describes correlations between points in the process. The correlation characteristics can be estimated from observations of the process, furnishing a general empirical approach to point- and line- processes [33]. Second-order statistics characterize many of the visible characteristics of an image or pattern [11], but are not infallible [28,5]. A direct analysis of dependence between points or lines in a process is obtained using the Palm probabilities P^x, essentially the conditional probability distribution of the random process *given* that there is a random point at x.

A random line process or circle process subdivides the plane into a random tessellation. This is a potentially important model of random images [14, 24, 31]. Characteristics of the polygons formed by a *Poisson* line process have been determined by Miles [23], in particular the means and variances of polygon area, perimeter length and number of sides. Another important random tessellation is the Dirichlet or Voronoi tessellation: if $\{x_i, i \in Z\}$ are the points in a point process, let the tile corresponding to x_i be

$$T_i = \{y \in \mathbb{R}^2 : |y - x_i| \leq |y - x_j|, j \neq i\}.$$

The T_i are polygons tessellating $\mathbb{R}^2$. Characteristics of the Voronoi tessellation induced by a Poisson point process are given by Miles [21].

Finally, random image models can be based on (c) *the theory of random sets.* This is discussed in Section 4.

3. Combinatorial theory

More results have recently been obtained for classical problems, by simplifying geometry and applying combinatorial probability methods [1]. We will first prove the curve length formula (3),

$$\mathbb{E}n(L \cap C) = \frac{2 \text{ length } (C)}{K}$$

where C is a plane curve, L is a UR line hitting $X \supset C$, and $n(L \cap C)$ = number of intersection points in $L \cap C$. Suppose C is a *polygonal* curve consisting of line segments $S_1, S_2, \ldots, S_n$. Let $[S_i]$ denote the event $L \cap S_i \neq \varnothing$, that is L hits S_i. Put

$$1_{[S_i]} = \begin{cases} 1 & \text{if} \quad L \cap S_i \neq \varnothing \\ 0 & \text{if} \quad L \cap S_i = \varnothing . \end{cases}$$

Clearly we have

$$n(L \cap C) = \sum_{i=1}^{n} 1_{[S_i]}$$

with probability 1, since $P(L$ contains $S_i) = 0$. But immediately

$$\mathbb{E}n(L \cap C) = \sum_{i=1}^{n} \mathbb{E} 1_{[S_i]} = \sum_{i=1}^{n} P([S_i]).$$

It can easily be argued that uniform random lines have $P([S_i])$ proportional to length (S_i);

$$\mathbb{E}n(L \cap C) = \alpha \sum_{i=1}^{n} \text{length } (S_i) = \alpha. \text{ length } (C)$$

which proves (3) up to the constant factor.

The proof reveals importance of *additivity,* meaning both the linearity of the integral $\mathbb{E}$ and the additivity of the counting function $n(L \cap C)$. Together with the natural properties of uniform distributions, this property forms the basis of stochastic geometry.

Suppose now we want the *distribution* of the variable $n(L \cap C)$: computation of $P\{n(L \cap C) = k\}$ is not obvious. Consider two segments S_1, S_2 and evaluate $P([S_1] \cap [S_2])$, the probability that L intersects *both* S_1, S_2. Case 1: if S_1, S_2 have a common point, let T be the third side of the triangle. Then

$$1_{[S_1] \cap [S_2]} = \tfrac{1}{2}(1_{[S_1]} + 1_{[S_2]} - 1_{[T]}) \qquad \text{a.s.}$$

since if L intersects both S_1, S_2 the sum in brackets equals 2, and otherwise is zero. Case 2: if S_1, S_2 have no common point we can derive a similar expression

$$1_{[S_1] \cap [S_2]} = \tfrac{1}{2}(1_{[A_1]} + 1_{[A_2]} - 1_{[B_1]} - 1_{[B_2]}) \qquad \text{a.s.}$$

where A_1, A_2, B_1, B_2 are segments joining the four endpoints of S_1, S_2. But this implies that *every expression* $1_{[S_i] \cap [S_j]} = 1_{[S_i]} \cdot 1_{[S_j]}$ *can be written as a linear combination* of variables $1_{[T_k]}$, where T_k are line segments joining vertices of C.

Theorem. *Let* $x_1, \ldots, x_n$ *be points in* $\mathbb{R}^2$, *and* s_{ij} *the line segment joining* x_i *with* x_j. *For a random line* L, *let* $[s_{ij}]$ *be the event* $L \cap s_{ij} \neq \varnothing$. *Let* $\mathcal{Q}$ *be the ring of events generated (through unions, intersections, set differences) by* $[s_{ij}]$, $1 \leq i < j \leq n$. *Then for any* $A \in \mathcal{Q}$ *there exist constants* $c_{ij}(A)$ *such that*

$$1_A = \sum_{i<j} c_{ij}(A) 1_{[s_{ij}]} \tag{4}$$

holds except when L *contains a vertex* x_i.

If L is uniformly distributed we take mean values in (4) to get

$$P(A) = 2/K \sum_{i<j} c_{ij}(A) \|x_i - x_j\| \tag{5}$$

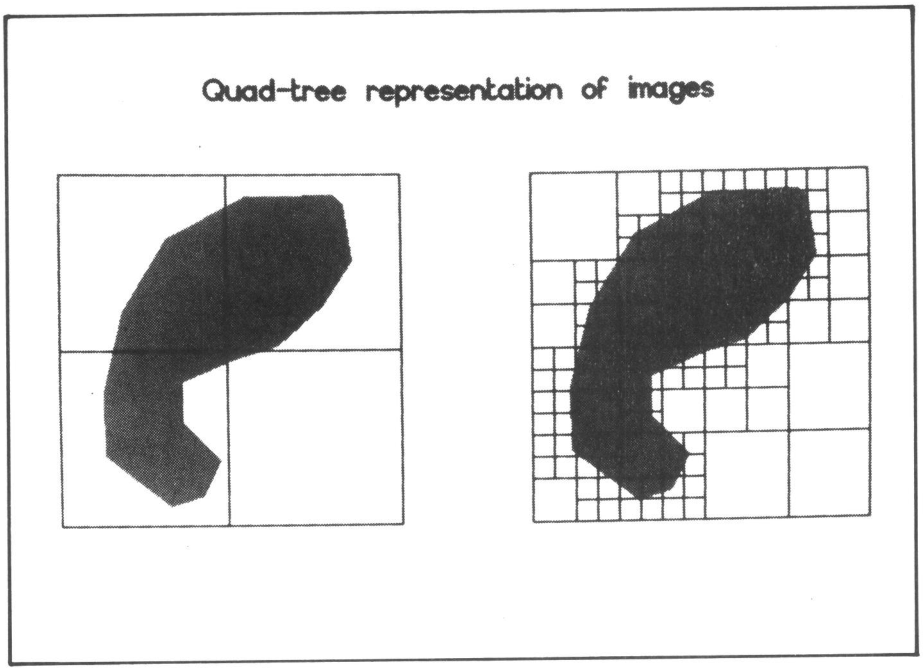
i.e. all combinatorial probabilities for UR lines are expressible as sums of lengths of segments s_{ij}. For example, the distribution of $n(L \cap C)$ is expressible in terms of the distances between each pair of vertices of C. This is a great advance, in principle, on the classical theory which was restricted to mean values. An algorithm for the $c_{ij}(A)$ is known, and practicable for small n.

One can also take non-uniform random lines in (4), say with probability distribution Q, to obtain

$$Q(A) = \sum_{i<j} c_{ij}(A) Q[s_{ij}] \tag{6}$$

and note the coefficients $c_{ij}(A)$ are the same as above. The quantity $Q[s_{ij}]$ serves as a generalized length of s_{ij}. Thus, again in principle, nonuniform random lines are no more computationally difficult than UR lines.

Finally we present another application to image complexity, concerning the quad-tree representation of images. An image can be recorded or transmitted as tree structure, as follows. Divide the image field into four equal squares and note which squares, if any, consist of a single colour. The remaining, multicoloured squares are subdivided again into four, and the process repeats until a predetermined level of subdivision is reached. The record of subdivisions and



colours forms the *quad tree.* Important questions include the average complexity (number of nodes) of the quad tree, and estimating the increase in complexity if a deeper level (finer subdivision) is added. Both problems depend on the image, but it is reasonable to suppose that in a sufficiently small square, the image boundary can be regarded as a uniformly random line. Consider a UR line hitting a square subdivided into $k \times k$ equal squares. According to (3) the mean number of subsquares crossed equals k. Furthermore using (5) we can compute the distribution of the number N of subsquares crossed. In the interesting case $k = 2$, we have $P(N = 1) = \frac{1}{2}(\sqrt{2}-1)$, $P(N = 2) = 2-\sqrt{2}$, $P(N = 3) = \frac{1}{2}(\sqrt{2}-1)$. Thus the cost of adding one extra level of subdivision is to double the number of terminal nodes, on average. One fifth of the new branches will be triple.

4. Random set theory

In addition to the constructive examples of random geometry in Section 2, one can propose others such as the zero-set (or contours) of a random function. Foundations of a general theory of random sets were laid by G. Matheron [18] and D.G. Kendall [13]. Matheron's theory of random closed sets was expressly developed as a mathematical background to image analysis as well as stochastic geometry. Kendall's theory takes an abstract view of the construction of probability spaces of random sets, emphasising the variety of structures which can be chosen. The two approaches are complementary [27] and both make use of Choquet's capacity theorem.

To introduce the theory we generalize the random events $[S_i]$ which played a formative role in section 3. For the Matheron approach, let $\mathfrak{F}$ be the class of all closed sets in $\mathbb{R}^n$. If $T \subset \mathbb{R}^n$ define the *hitting set*

$$[T] = \{F \in \mathfrak{F} : F \cap T \neq \varnothing\}.$$

Endow $\mathfrak{F}$ with the (weakest) topology such that $[U]$ is an open subset of $\mathfrak{F}$ for all open sets $U \subset \mathbb{R}^n$, and $[K]$ is closed for all compact $K \cap \mathbb{R}^n$ (see [20]). Then $\mathfrak{F}$ becomes a Polish space. Define a *random closed set* as a random element of $\mathfrak{F}$ with the Borel σ-algebra. Under this structure the *events* $[T]$, $T \subset \mathbb{R}^n$ are measurable when T is open, closed or indeed Borel. Intersections and unions of random closed sets are random closed sets. Area, length (where defined) and number of points (where finite) are random variables.

Kendall's approach emphasises that the definition of a random set depends on the geometrical information which is assumed to be observable. Its basic constituents are the random events $[T] = \{X \cap T \neq \varnothing\}$ where X is the random set and T is a fixed set called a 'trap'. The associated random variable

$$h(T) = \begin{cases} 1 & \text{if} \quad X \cap T \neq \varnothing \\ 0 & \text{if} \quad \text{not} \end{cases} \tag{7}$$

corresponds to a 'bit' or 'flag' indicating whether X was detected by the trap T. From the observer's point of view, the random set X is characterized by the information $\{h(T), T \in \mathfrak{T}\}$ where $\mathfrak{T}$ is the class of all traps available to the observer. Define a *trapping system* $\mathfrak{T}$ on a space S to be a class of nonempty subsets of S, which cover S, satisfying certain properties analogous to separability and local compactness. A *random* $\mathfrak{T}$*-set* in S is a random function

$$h : \mathfrak{T} \rightarrow \{0,1\}$$

i.e. a stochastic process of 0-1 variables $h(T)$, $T \in \mathfrak{T}$, subject to a consistency condition which enables h to be interpreted in the form (7). Note the probability structure depends completely on the choice of trapping-system. If $S = \mathbb{R}^n$ and $\mathfrak{T} =$ open sets, a random $\mathfrak{T}$-set is a random closed set in Matheron's sense. Smaller trapping-systems may be inadequate to distinguish all closed sets. A set X is indistinguishable (to the observer) from its $\mathfrak{T}$*-closure,*

$$\text{clos}\,(X, \mathfrak{T}) = \left[\bigcup_{X \cap T = \varnothing} T\right]^c = \bigcap_{X \cap T = \varnothing} T^c$$

(c denotes complement) and we need only consider $\mathfrak{T}$-*closed* sets $X = \text{clos}(X,\mathfrak{T})$. For example if $\mathfrak{T} = \{\text{open halfplanes of } \mathbb{R}^2\}$ the $\mathfrak{T}$-closed sets are the convex sets of $\mathbb{R}^2$. Thus random $\mathfrak{T}$-sets in this case 'are' random convex sets; and the customary representation of convex sets by support functions can be derived from $h(T)$.

Random set theory provides solid foundations for investigating both stochastic geometry and the observation and processing of images. For example, convergence of random sets is a natural concept in the general theory which has been applied to assess errors committed in digitizing an image, approximation of one image by another, and the stability of image processing transformations [31, Chapter VII] and to derive the statistically important laws of large numbers and a central limit theorem for repeated observations of images [2,36]. The general setting also permits more involved discussion of the probabilistic properties of image models, such as infinite divisibility and the semi-Markov property [18,19]. It is a basic result that the probability distribution of a random set X is determined by its *avoidance function*

$$Q(A) = \text{Prob}\,(X \cap A = \varnothing), \quad A = \bigcup_{i=1}^{n} T_i, \quad T_i \in \mathfrak{T}$$

and the introduction of Q makes for a coherent approach to image models [31,18].

The strongest link between image analysis and random set theory is surely the trapping system. Any image is given to us through an array of detectors (and perhaps subjected to edge detection processing, etc.) which can be formalised as a trapping system. Further, the relationships between various forms of image information (e.g. digitized versions on different lattices; grey tones) can be studied by varying $\mathfrak{T}$ in the stochastic model. The author feels that the great potential of this method is yet unexplored.

5. Mathematical morphology

The work of J. Serra [31] establishes a coherent methodology for image analysis which avoids the fragmentary character of most other approaches. Mathematical morphology developed in parallel with random set theory, beginning with Matheron's [17] geostatistical work and Serra's invention of the 'texture analyzer' image processing devices. The result is a combination of sound theoretical criteria with practical experience. We can only convey the flavour of the subject here.

Transformations of sets arise in many stochastic geometry problems. Consider the probability distribution of the random distance $d(x, A)$ from a fixed set $A \subset \mathbb{R}^2$ to a random point $x \notin A$. Clearly $P\{d(x, A) \leq r\}$ equals the probability that X falls in the region $A_{(r)} = \{x \in \mathbb{R}^2 : d(x, A) \leq r\}$ which we dub the *r-envelope* of A. Equivalently $A_{(r)}$ is the set formed by placing a disc of radius r around every point $a \in A$. The envelope transformation $A \to A_{(r)}$ is the simplest example of a set transformation. If $A = D_R$ is a disc then $A_{(r)} = D_{R+r}$, while in general the shape of $A_{(r)}$ is more rounded (with smaller holes) than that of A. It is argued that the function $f_A(r) = \text{area}\,(A_{(r)})$ reflects essential characteristics of the geometry of A. If A is convex then $f_A(r) = \pi r^2 + r.\text{length}\,(\partial A) + \text{area}\,(A)$, while if A is a finite set of points then f_A is piecewise quadratic with a behaviour reflecting the sizes of gaps between the points. A series of images $A_1, \ldots, A_n$ could be differentiated or discriminated using the derived functions $f_{A_1}(r), \ldots, f_{A_n}(r)$.

The envelope operation can be performed on a discrete grid of points. A simple algorithm is to scan the entire grid and, for each point x whose digital neighbourhood includes a point of the current image A, we mark x for inclusion in the new image $A_{(r)}$. Furthermore we can watch this process of expansion for increasing r by repeating the algorithm, since $(A_{(r)})_{(s)} = A_{(r+s)}$. This is done by texture analyzers.

The *Minkowski sum* of two sets $A, B \subset \mathbb{R}^2$ is defined as

$$A \oplus B = \{a + b : a \in A,\ b \in B\}$$

in the sense of vector addition. If B is the disc D_r then $A \oplus D_r = A_{(r)}$, the r-envelope. More generally $A \oplus B$ is the superposition of translated copies of B centred on each of the points of A, if we take the origin 0 as the 'centre' of B. Shifted copies of A are obtained when B is a single point, $A \oplus \{b\} = \{a + b : a \in A\}$. Defining $\check{B} = \{-b : b \in B\}$ one can interpret $A \oplus B = \{x \in \mathbb{R}^2 : (x \oplus \check{B}) \cap A = \varnothing\}$, the set of all 'centres' of shifted copies of $\check{B}$which intersect A. Hence the transformation $A \rightarrow A \oplus B$ also has a clear interpretation in stochastic geometry, and can be claimed to reflect important characteristics of the geometry of A. This and other set transformations can be implemented on a discrete grid by including or removing points x according to the state of the entire digital neighbourhood of x.

Minkowski subtraction of $A, B \subset \mathbb{R}^2$ is defined by

$$A \ominus B = (A^c \oplus B)^c$$

i.e. the complement A^c is enlarged by B. For example, if $B = D_r$ is a disc, $A \ominus D_r = \{x \in A : d(x, A^c) \geqslant r\}$ is the *inner parallel set.* In general $A \ominus B = \{x \in A : x \oplus \check{B} \subset A\}$ is the set of all centres of copies of $\check{B}$ *contained* in A. This has a natural interpretation and the function $g(r) = \text{area}\,(A \ominus D_r)$ is claimed to contain essential information about the geometry of A. Define two further set transformations, the closure

$$A^B = (A \oplus \check{B}) \ominus B$$

and opening

$$A_B = (A \ominus \check{B}) \oplus B.$$

Thus A_B is the union of all copies of B contained in A; and A^B is the result of a similar operation on A^c. A set is B-closed, $A^B = A$, iff it is $\mathfrak{T}$-closed in the sense of Section 4 where $\mathfrak{T}$ is the class of all translated copies of B. Apart from their natural interpretation in stochastic geometry, A^B and A_B can be used to develop a rigorous definition of size and size distribution for images [18,31].

The mathematical morphology approach to an image processing problem is to select an image transformation (built from $\oplus, \ominus$, A^B, A_B etc.) suitable to the application, and make numerical analyses of the transformed images. One chooses transformations either by experience, intuition about the scientific problem, or by setting down criteria which the transformation must satisfy.

Some limitations of mathematical morphology as it currently stands call for brief comments. The texture analyser is designed on a hexagonal point lattice for the digitized image. Naturally the theory is strongly dependent on this choice of instrumentation, and probably does not answer all questions about random image models that are required in different applications. Associated with the choice of instrumentation is the adoption [31, pp 8-15] of a list of theoretical principles which notably excludes *rotational* stability. A hexagonal grid has only three basic directions and there have been difficulties with the analysis of image orientation or directionality. There may also be practical reasons for employing a rectangular grid or other system of image detection - for example, satellite data may already be in this form. Another problem with all image analysis based on stochastic geometry is that images are not sharply divided black and white sets, but grey tone functions. This is a drawback to the widespread use of texture analyzers. Mathematical morphology for grey-tone functions is under development [31, Chapter XII].

The author suspects one can be led astray by excessive analysis of a single image, when this image is to be representative of a larger population. This applies particularly in stereology, where the planar image is a random plane section $X \cap E$ of a three-dimensional body X which is the real object of interest. It is then important that the sampling procedure used to generate $X \cap E$ should be known, and appropriate. Statistical inferences depend on the sampling method used. It is not quite sufficient to base image analysis on considerations of the trapping-system and other geometrical structures, without incorporating statistical models for the origins of data.

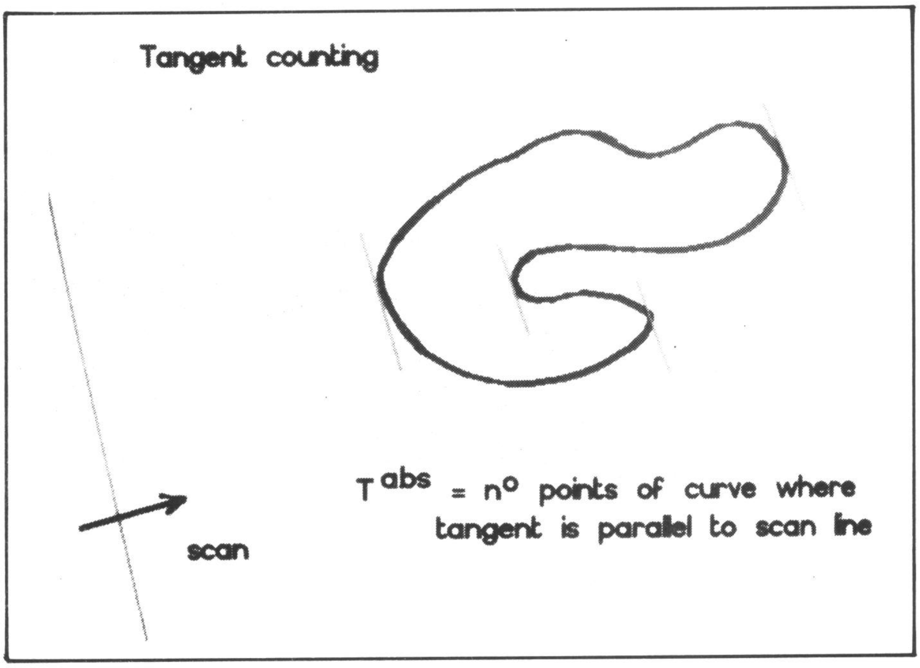
6. Image irregularity, observation errors and geometric measure theory

Elementary formulae from stochastic geometry (see (1)-(3) in Section 2) are widely used in stereology for measuring curve lengths, estimating surface areas and so on. Yet these results were derived for ideal smooth curves and it is a priori doubtful whether they apply to irregular images or images observed under error.

An extreme example is *tangent counting*. Let C be a twice differentaible plane curve, $\theta \in [0,\pi)$ and $T^{abs}(\theta)$ = number of tangents to C parallel to direction θ. This would be found by scanning a straight line across the image (parallel to θ) and counting the positions where the image is tangent to C. We have

$$\int_0^{\pi} T^{abs}(\theta) d\theta = \int_C |\kappa(s)| ds \tag{8}$$

where $\kappa(s)$ is the curvature of C at point s. If the scan direction θ is generated at random (uniformly), $\pi T^{abs}(\theta)$ is a statistically unbiased estimator of the total absolute curvature of C. Additionally if C is itself a random plane section of a curved surface, then T^{abs} yields an estimate of the total 'absolute' surface curvature.



Even assuming that real images are differentiable, the tangent count is unstable in the sense that small perturbations (kinks, ripples) in C may cause large changes in T^{abs} and κ. More realistically if C is the boundary of a finite union of convex compact sets (hence, almost everywhere differentiable) T^{abs} does not share the properties usually required of a good statistic. Serra [31, p.141 ff] nevertheless shows that a precise and useful interpretation can be given to the tangent count or 'convexity number' of such curves, and that this can be approximately determined from a digitized image.

Practical stereologists and image analysts follow procedures for counting 'tangents' to image curves, even when these are irregular, thick or fuzzy, broken or digitized. A tangent counting algorithm may be built into the image analyzing device. Mathematicians should be discussing the performance of such algorithms, their relation to real geometry, and the effects of observation errors.

Standard proofs of (1)-(3) and (8) do not accommodate a discussion of perturbations or errors, being applications of Fubini's theorem to simple geometrical models. We need the more powerful methods of geometric measure theory [6], principally the *coarea formula.* Briefly, let M, N be m- and n-dimensional domains (rectifiable surfaces), $m \geqslant n$, and let $p: M \to N$ be a Lipschitz-continuous map. For almost every $x \in N$, $p^{-1}\{x\} = \{z \in M : p(z) = x\}$ is an $m-n$ dimensional rectifiable set. If $m = n$, then $p^{-1}\{x\}$ is a finite set. There is a function $J^n p$ defined on M called the approximate Jacobian of p, such that the *coarea formula*

$$\int_M f(z)(J^n p)(z) d\mathcal{H}^m z = \int_N \int_{p^{-1}\{x\}} f(z) d\mathcal{H}^{m-n} z \, d\mathcal{H}^n x \tag{9}$$

holds for any $\mathcal{H}^m$-integrable function $f: M \to \mathbb{R}$, where $\mathcal{H}^k$ is the k-dimensional Hausdorff measure ('k-dimensional volume integration', see Section 7).

Thus (9) is a kind of generalization of Fubini's theorem which incorporates the Jacobian for a change of variables.

To prove (8), for example, let C be a twice differentiable curve, and introduce

$$C^* = \{(s,l) : s \in C, \, l \text{ is the tangent to } C \text{ at } s\}.$$

This is a one-parameter set of points in $\mathbb{R}^2 \times \mathbb{R} \times [0,\pi)$. Apply the coarea formula (9) to the map

$$p: C^* \to C, \quad p(s,l) = s.$$

This has $(J^1 p)(s,l) = (1+\kappa^2)^{-\frac{1}{2}}$ where $\kappa = \kappa(s)$ is the curvature of C, and since $p^{-1}\{s\}$ is a single point (s,l) we get

$$\int_{C^*} f(s,l)(1+\kappa^2)^{-\frac{1}{2}} d\mathcal{H}^1(s,l) = \int_C f(s,l) ds$$

for any function f. Similarly, for the map

$$q: C^* \to [0,\pi), \quad q(s,l) = \text{direction of line } l,$$

we have $(J^1 q)(s,l) = (\kappa^2/1+\kappa^2)^{\frac{1}{2}}$. Since $q^{-1}\{\theta\}$ consists of all pairs (s,l) where l is parallel to θ, we get

$$\int_{C^*} \tilde{f}(s,l)(\kappa^2/1+\kappa^2)^{\frac{1}{2}} d\mathcal{H}^1(s,l) = \int_0^\pi \sum_{q^{-1}\{\theta\}} \tilde{f}(s_i,l_i) d\theta.$$

If $\tilde{f} = 1$, the sum on the right hand side above is $T^{abs}(\theta)$. Choosing $f(s,l) = |\kappa|$ so that the two left hand sides agree, we get equation (8).

Now suppose that C is nondifferentiable, and that the experimenter has some algorithm for counting or detecting apparent tangents to C. Let

$$\tilde{C} = \{(s,l) : s \in \mathbb{R}^2, l \text{ is a line; the algorithm counts } l \text{ as a tangent to at } s\}.$$

Then under suitable conditions we may replace C^* above by $\tilde{C}$ and perform the same calculations to get

$$\int_0^{\pi} \tilde{T}^{abs}(\theta)d\theta = \int_{\Gamma} \tilde{\kappa}(s)ds,$$

where $\tilde{T}^{abs}$ is the experimentally observed tangent count, $\Gamma = p(\tilde{C})$ is the set of points at which tangents are detected, and $\tilde{\kappa} = J^1q/J^1p$ is a kind of generalized curvature. For example, let C be an irregular curve $C(t) = A(t)+\epsilon(t)$, $0 \leqslant t \leqslant 1$, where curve A is smooth and $\|\epsilon(t)\| < r$. If the tangent algorithm is such that $s \in A$ and l is tangent to $A \oplus D_r$, then $\Gamma = A$, and $\tilde{\kappa}$ is a function of r and the curvature of A. Thus Γ is a rectified version of C. Secondly, if C is smooth, but a tangent where $\kappa(s)$ is small may not be observed, we get

$$\mathbb{E}(\pi \tilde{T}^{abs}) = \int_C |\kappa(s)| u(\kappa(s))ds$$

where $u(\kappa)$ = probability of detecting a given tangent at curvature κ. Further examples are explored in [4].

Thus we still have a geometrical interpretation of the image analysis algorithm when it is applied to non-ideal images. This is achieved by concentrating on intrinsic behaviour of the algorithm or observation method, encapsulated in the projection maps p, q. More generally we can regard an image analysis algorithm as an *operator* on images in the sense of generalized functions, and the mathematical prerequisites for such an approach already exist [6].

7. Fractals

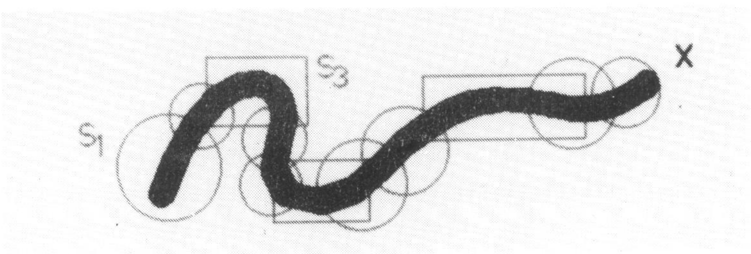
Mandelbrot [15,16] explored the concept of fractal (fractional dimensional) sets initiated by Besicovitch, which have wide mathematical associations and seem to be useful models for real images. The simplest kind of fractal set is *self-similar:* if X can be divided into k disjoint sets each of which is congruent to X after magnification by a factor α, then $\Delta = \log\alpha / \log k$ is the similarity dimension of X. For curves $\Delta = 1$; for a disc $\Delta = 2$; but for the Cantor set, $k = 2$, $\alpha = 3$, $\Delta = \log 3/\log 2$ is fractional. When X is magnified, its content increases by a fractional power of the magnification. This extreme form of fractal behaviour is not generally required (except in the limit of small scale). Define for each real $t \geqslant 0$ the t-dimensional Hausdorff measure $\mathcal{H}^t$,

$$\mathcal{H}^t(X) = \lim_{\epsilon \downarrow 0} c_t . \inf\{\sum_{i=1}^{N} (\text{diam } S_i)^t : S_1, \ldots, S_N \text{ cover } X, \text{diam } S_i < \epsilon\}$$

where the infimum ranges over (say) all families of compact sets S_i with diameters less than ϵ. The limit may be infinite. Define the Hausdorff-Besicovitch dimension of X as

$$D(X) = \sup\{t \geqslant 0 : \mathcal{H}^t(X) < \infty\} = \inf\{t \geqslant 0 : \mathcal{H}^t(X) = \infty\}.$$

Hausdorff measure H^t



$$H^t(X) = c_t \lim_{\delta \downarrow 0} \inf\Big\{\sum (\text{diam } S_i)^t : S_1, \ldots, S_n \text{ cover } X, \ \text{diam } S_i < \delta\Big\}$$

Then X is fractal if $D(X)$ is not an integer. If X is a random closed set (Section 4), the topology of $\mathcal{F}$ is such that $D(X)$ is a random variable.

Other examples of fractals include the graphs and zero sets of random continuous functions (the graph of Brownian motion is *statistically* self similar) and limit sets of iterations of quadratic maps in the complex plane. The viewer's impression of a fractal curve is one of sharp irregularity and unbounded oscillation.

Real objects and images do often behave non-linearly with magnification. Coastlines are the best-known example. Given a picture of a fractal curve ($1<D<2$) we could estimate D as the slope of the regression line relating $\log L(\alpha)$ to $\log\alpha$, where $L(\alpha)$ is an estimate of length obtained at magnification α. Applied to coastlines this has produced a range of fractional dimensions, which seem to reflect degrees of irregularity. A more serious application concerns the measurement of lung membrane surface area [34, p. 156] from plane section curves. Conflicting estimates based on different magnifications have been reconciled and a consistent estimate of D obtained.

Many real phenomena and images have been described as 'fractal' and their empirical values of D determined. The theory of ideal fractals has not kept pace with this development of approximate fractals. Any empirical value of D is a partial description of the image, at certain scales only, and over different scales the 'dimension' may vary. This should not be an objection to the use of fractals as a geometrical model (naturally any model is confined to a chosen scale), but the meaning of a fractal approximation needs to be clarified [10]. We have already observed that Hausdorff dimension fits into the general theory of random closed sets, and indeed $D(X)$ represents an asymptotic index of the frequency of intersections between X and small *traps* $D_r, r \to 0$. It seems to the author that fractional dimensional irregularity could be better understood from the empirical and statistical viewpoint of stochastic geometry.

References

[1] AMBARTZUMIAN, R.V. (1982). *Combinatorial integral geometry,* J. Wiley and Sons, Chichester.

[2] ARTSTEIN, Z. and VITALE, R.A. (1975). A strong law of large numbers for random compact sets. *Ann. Probability 5,* 879-882.

[3] BADDELEY, A. (1982). Stochastic geometry: an introduction and reading list. *Internat. Statist. Review 50,* 179-193.

[4] BADDELEY, A. (1983). Applications of the coarea formula to stereology. *In [8],* 1-17.

[5] BADDELEY, A. and SILVERMAN, B.W. A cautionary example on the use of second order methods for analyzing point patterns. *To appear in Biometrics.*

[6] FEDERER, H. (1969). Geometric measure theory. *Springer,* Heidelberg.

[7] FORTET, R. and KAMBOUZIA, M. (1975). Ensembles aléatoires, répartitions ponctuelles aléatoires, problèmes de recouvrement. *Ann. Inst. Henri Poincaré 11(B),* 299.

[8] GUNDERSEN, H-J.G. AND JENSEN, E.B. (1983). (Eds.) Proceedings of the Second International Workshop on Stochastic Geometry and Stereology, *Memoir No. 6, Department of Theoretical Statistics,* University of Aarhus.

[9] HARDING, E.F. AND KENDALL, D.G. (1973). (Eds.) *Stochastic Geometry: a tribute to the memory of Rollo Davidson,* J. Wiley and Sons, Chichester.

[10] HEYDE, C.C. On some new probabilistic developments of significance to statistics: martingales, long range dependence, fractals and random fields. *To appear.*

[11] JULESZ, B. (1975). Experiments in the visual perception of texture. *Scientific American 232, 4,* 34-43.

[12] KALLENBERG, O. (1983). The invariance problem for stationary line and flat processes. *IN [8],* 105-114.

[13] KENDALL, D.G. (1974). Foundations of a theory of random sets. *In [9]* 322-376.

[14] KENDALL, M.G. and MORAN, P.A.P. (1963). Geometrical probability. *Statist. Monographs and Courses No. 10, Griffin,* London.

[15] MANDELBROT, B.B. (1976). *Fractals: form, chance and dimension,* W.H. Freeman, San Francisco.

[16] MANDELBROT, B.B. (1982). *The fractal geometry of nature,* W.H. Freeman, San Francisco.

[17] MATHERON, G. (1967). *Eléments pour une Théorie des Milieux Poreux,* Masson, Paris.

[18] MATHERON, G. (1974). *Random sets and integral geometry,* J. Wiley and Sons, New York.

[19] MATTHES, K., KERSTAN, J., MECKE, J. (1978). *Infinitely divisible point processes,* J. Wiley and Sons, New York.

[20] MICHAEL, E. (1951). Topologies on spaces of subsets. *Trans. Amer. Math. Soc. 71,* 152-182.

[21] MILES, R.E. (1970). On the homogeneous planar Poisson process. *Math. Biosci., 6,* 85-127.

[22] MILES, R.E. (1972). The random division of space, *Suppl. Advan. Appl. Prob. 4,* 243-266.

[23] MILES, R.E. (1973). The various aggregates of random polygons determined by random lines in a plane. *Advan. Math. 10,* 256-290.

[24] MILES, R.E. (1981). A survey of geometric probability in the plane, with emphasis on stochastic image modeling. *In [23],* 277-300.

[25] MILES, R.E. and SERRA, J. (Eds.) (1978). Geometrical probability and biological structures: Buffon's 200th Anniversary. *Lecture Notes in Biomathematics, No. 23. Springer,* Heidelberg.

[26] PAPANGELOU, F. (1974). The conditional intensity of general point processes and an application to line processes. *Z. Wahrsch'theorie 28,* 207-226.

[27] RIPLEY, B.D. (1976). Locally finite random sets: foundations for point process theory. *Annals of Probability 45,* 983-994.

[28] RIPLEY B.D. (1981). *Spatial Statistics,* J. Wiley and Sons, New York.

[29] ROSENFELD, A. (Ed.) (1981). Image Modeling. Academic Press, New York. *Note:* largely reprinted in *Computer Graphics and Image Processing,* Vol. 12.

[30] SANTALÓ, L.A. (1976). *Integral geometry and geometric probability,* Addison-Wesley, New York.

[31] SERRA, J. (1982). *Image analysis and mathematical morphology,* Academic Press, New York.

[32] STOYAN, D. (1979). Applied stochastic geometry: a survey. *Biom. Journal 21,* 693-715.

[33] STOYAN, D. and MECKE, J. (1983). *Stochastische Geometrie,* Wissenschaftlicher Taschenbücher Bd. 275, Akademie-Verlag, Berlin (DDR).

[34] WEIBEL, E.R. (1979). *Stereological methods, Vol. 1.,* Academic Press, New York.

[35] WEIL, W. (1983). Stereology: a survey for geometers. *In: Convexity and Its Applications,* Birkhäuser, Basel, edited by P. Gruber and J.M. Wills.

[36] WEIL, W. (1982). An application of the central limit theorem for Banach space valued random variables to the theory of random sets. *Z. Wahrsch'theorie 60,* 203-208.

Typesetting at the CWI — Part 2

by Jaap Akkerhuis

Handling of tabular material

Although it is possible to do this "by hand", the easiest way of handling tabular material is using the TBL [1] preprocessor. This preprocessor turns a simple description of a table into TROFF commands.

I will illustrate TBL with some examples. The TBL preprocessor will only process lines between the delimiters, `.TS` and `.TE`, and so, in general, a table will look like

```
.TS
options ;
format .
data
.TE
```

The symbol Ⓣ in the input represents a tab character.

The first example is a three part table, with the first items centered, and the rest of the items left adjusted.

Input:

```
.TS
box;
c c c
l l l.
Language Ⓣ Authors Ⓣ Runs on

Fortran Ⓣ Many Ⓣ Almost anything
PL/I Ⓣ IBM Ⓣ 360/370
C Ⓣ BTL Ⓣ 11/45,H6000,370
BLISS Ⓣ Carnegie-Mellon Ⓣ PDP-10,11
IDS Ⓣ Honeywell Ⓣ H6000
Pascal Ⓣ Stanford Ⓣ 370
.TE
```

The first part of this article appeared in CWI Newsletter no. 3 (June 1984).

Output:

Language	Authors	Runs on
Fortran	Many	Almost anything
PL/I	IBM	360/370
C	BTL	11/45,H6000,370
BLISS	Carnegie-Mellon	PDP-10,11
IDS	Honeywell	H6000
Pascal	Stanford	370

The following table has all entries boxed, and the first entry centered and spanned. The complete table is expanded over the available line length as well.

Input:

```
.TS
expand allbox;
c c c c c c c c c
n n n n n n n n n.
Type ⓣ Vf ⓣ If ⓣ Wa ⓣ Wg2 ⓣ Ea ⓣ Eg2 ⓣ Wo ⓣ Mhz
807 ⓣ 6.3 ⓣ .9 ⓣ 30 ⓣ 3.5 ⓣ 750 ⓣ 300 ⓣ 50 ⓣ 60
813 ⓣ 10 ⓣ 5 ⓣ 100 ⓣ 22 ⓣ 2000 ⓣ 400 ⓣ 260 ⓣ 30
815 ⓣ 6.3 ⓣ 1.6 ⓣ 25 ⓣ 4 ⓣ 500 ⓣ 200 ⓣ 56 ⓣ 125
829A ⓣ 6.3 ⓣ 2.5 ⓣ 40 ⓣ 7 ⓣ 750 ⓣ 240 ⓣ 87 ⓣ 200
832A ⓣ 6.3 ⓣ 1.6 ⓣ 15 ⓣ 5 ⓣ 750 ⓣ 250 ⓣ 26 ⓣ 200
```

Output:

Type	Vf	If	Wa	Wg2	Ea	Eg2	Wo	Mhz
807	6.3	.9	30	3.5	750	300	50	60
813	10	5	100	22	2000	400	260	30
815	6.3	1.6	25	4	500	200	56	125
829A	6.3	2.5	40	7	750	240	87	200
832A	6.3	1.6	15	5	750	250	26	200

The next table is more complicated. To get the complicated header right, it uses the **.T&** (table continue) command. This table is boxed again, and has some entries boxed, by using the | specifier. The = in the data part of the table denotes a double line over the full width of the column. A _ specifies a single line. The \^ specifies a vertical aligned column. Note the difference between this table and the "all boxed" one in the previous example.

Input:

```
.TS
box;
cfB s s s s s.
Ranges of Typical Commercial Spark Stations
_
.T&
c | c s s s | c
c | c s s s | c
c | c | c s s | c
c | c | c s s | c
c | c | c | c | c | c.
 Ⓣ Range in nautical miles Ⓣ Wave-
Power Ⓣ _ Ⓣ \^
\^Ⓣ Over sea Ⓣ Over Land Ⓣ length
required Ⓣ \^Ⓣ _ Ⓣ \^
\^Ⓣ \^Ⓣ Flat Ⓣ Hilly Ⓣ Mountainous Ⓣ (metres)
=
.sp 0.5
.T&
n | n | n | n | n | n.
300 watts Ⓣ 100 Ⓣ 77 Ⓣ 30 Ⓣ 13 Ⓣ 300
\^Ⓣ 100 Ⓣ 95 Ⓣ 73 Ⓣ 52 Ⓣ 1200
1\(12 kw Ⓣ 220 Ⓣ 170 Ⓣ 67 Ⓣ 28 Ⓣ 300
\^Ⓣ 220 Ⓣ 210 Ⓣ 160 Ⓣ 115 Ⓣ 1200
3 kw Ⓣ 280 Ⓣ 220 Ⓣ 84 Ⓣ 36 Ⓣ 300
\^Ⓣ 280 Ⓣ 270 Ⓣ 200 Ⓣ 145 Ⓣ 1200
5 kw Ⓣ 340 Ⓣ 260 Ⓣ 100 Ⓣ 43 Ⓣ 300
\^Ⓣ 340 Ⓣ 325 Ⓣ 240 Ⓣ 175 Ⓣ 1200
10 kw Ⓣ 470 Ⓣ 360 Ⓣ 138 Ⓣ 59 Ⓣ 300
\^Ⓣ 470 Ⓣ 450 Ⓣ 330 Ⓣ 240 Ⓣ 1200
.nr x \n(.n-7n \" Remember position
.TE
.ta \nxuL              \" set a tab
 Ⓣ \fI(Sankey)\fP      \" goto tab and print
```

Output:

Ranges of Typical Commercial Spark Stations					
Power required	Range in nautical miles				Wave-length (metres)
	Over sea	Over Land			
		Flat	Hilly	Mountainous	
300 watts	100	77	30	13	300
	100	95	73	52	1200
1½ kw	220	170	67	28	300
	220	210	160	115	1200
3 kw	280	220	84	36	300
	280	270	200	145	1200
5 kw	340	260	100	43	300
	340	325	240	175	1200
10 kw	470	360	138	59	300
	470	450	330	240	1200

(Sankey)

A more exotic use of TBL is to generate a picture with it, as in the next example. The `ce0` stands for `c`entered `e`qualwidth column with space of `0` ens* between the columns. Note that the complete table is centered on the page as well.

Input:

```
.TS
center;
ce0 ce0 ce0 ce0 ce0 ce0 ce0 ce0 ce0 ce0 ce0 ce.
- Ⓣ - Ⓣ - Ⓣ - Ⓣ - Ⓣ - Ⓣ - Ⓣ - Ⓣ - Ⓣ - Ⓣ - Ⓣ -
| Ⓣ . Ⓣ . Ⓣ . Ⓣ . Ⓣ . Ⓣ . Ⓣ . Ⓣ . Ⓣ . Ⓣ . Ⓣ +
| Ⓣ . Ⓣ . Ⓣ @ Ⓣ . Ⓣ . Ⓣ . Ⓣ . Ⓣ ] Ⓣ . Ⓣ . Ⓣ |
| Ⓣ . Ⓣ . Ⓣ . Ⓣ . Ⓣ B Ⓣ . Ⓣ . Ⓣ . Ⓣ . Ⓣ . Ⓣ |
| Ⓣ . Ⓣ . Ⓣ . Ⓣ . Ⓣ . Ⓣ . Ⓣ . Ⓣ . Ⓣ . Ⓣ . Ⓣ |
- Ⓣ - Ⓣ - Ⓣ - Ⓣ - Ⓣ + Ⓣ - Ⓣ - Ⓣ - Ⓣ - Ⓣ - Ⓣ -
.TE
```

* An en is the width on the character n in the current font.

Output:

```
 - - - - - - - - - - - -
| . . . . . . . . . . . +
| . . @ . . . . ] . . |
| . . . . . B . . . . . |
| . . . . . . . . . . . |
 - - - - - + - - - - - -
```

Handling of mathematics material

Mathematics material is handled by the EQN [2] preprocessor. Just as the TBL preprocessor it arranges low level TROFF commands.
The input resembles a programming language which allows specification of the equation to be typeset by describing the desired output in words. It is claimed that it is easy to use and one can learn to typeset an in-line formula like $\int_0^1 \sin\pi x \, dx$ or displayed equations like

$$D_n = \frac{1}{n} \sum_{k=1}^{n} \left| \exp\left(\frac{k}{n} 2\pi i\right) - 1 \right| =$$

$$= \frac{1}{n} \sum_{k=1}^{n} \left| \left(\cos\frac{2\pi k}{n} + i \sin\frac{2\pi k}{n}\right) - 1 \right| =$$

$$= \frac{1}{n} \sum_{k=1}^{n} \left(2 - 2\cos\frac{2\pi k}{n}\right)^{\frac{1}{2}} = \frac{2}{n} \sum_{k=1}^{n} \sin\frac{\pi k}{n}$$

in an hour or so.

EQN will process everything in the input between `.EQ`, `.EN` and pairs of dollar signs (`$`), the most used *delimiters*. The `.EQ` and `.EN` with their optional arguments are copied through untouched.

The —*ms*-package will center the given equation. A `.EQ L` will make the displayed equation a left justified block, and `.EN I` somewhat indented. An optional third argument gives the equation number placed towards the right margin.

```
.EQ I (3a).
a = b + c
.EN
```

$$a = b + c \qquad \text{(3a).}$$

Note that this is not a feature of EQN, but of the macro package actually used. So if you want to have the equation number in front, you have to rewrite the `.EQ` macro.

One of the problems of the beginning user is that input spaces are nearly always ignored, so the following different input lines

```
x=y+z

x = y + z

x    =    y
   + z
```

will all result in the same output:

$x = y + z$

The (input) spaces are only used as obligatory keyword delimiters, so `$x sub2$` will give you $xsub2$ instead of x_2.

However, if you want to have spaces in the output, the ~ character will give you a space and the ^ half a space. So

```
x~=~y~+~z
```

will look like

$x \ = \ y \ + \ z.$

Subscripts and superscripts, special names and brackets
As you will have noticed before, subscripts are easily generated by the `sub` command. Superscripts are made in the same way, so

```
x sup 2 + y sub k
```

gives

$x^2 + y_k$

Of course, more complicated matter can be treated in the same way:

```
e sup {i pi sup {rho +1}}
```

$e^{i\pi^{\rho+1}}$

The braces `{}` are used here to group things together. Everything between them will be treated by EQN as a block. Also `pi` and `rho` are translated into the corresponding Greek character. To get the braces itself one uses `{` and to get large braces in the output, the `left` and `right` constructs are used. See the examples in the following paragraph.

Fractions and roots

In the following example, you see the use of big brackets and the way fractions are made with the `over` construct.

```
left { a over b + 1 right }
 ~=~ left ( c over d right )
 + left [ e right ]
```

$$\left\{ \frac{a}{b} + 1 \right\} = \left(\frac{c}{d} \right) + \left[e \right]$$

Square roots are made just as easily

```
sqrt a+b
```

$$\sqrt{a+b}$$

One should avoid big square roots, they don't look nice.

```
sqrt {a sup 2 over b sub 2}
```

$$\sqrt{\frac{a^2}{b_2}}$$

One can better rearrange the formula so as to get

$$(a^2 / b_2)^{\frac{1}{2}}$$

This is done by

```
(a sup 2 /b sub 2 ) sup half
```

Large operators and matrices

Large operators like integrals and summations are made straightforwardly as in:

```
prod from { i != j } to n
     { ( 1-x sub i x sub j sup -1 ) } sup k
```

$$\prod_{i \neq j}^{n} (1 - x_i x_j^{-1})^k$$

A matrix is made by

```
matrix {
  ccol { x sub i above y sub i }
  ccol { x sup 2 above y sup 2 }
  }
```

which results in

x_i x^2

y_i y^2

There are more ways of placing things on top of each other, and also for letting parts of an equation line up with each other. Also, font changes can take place on demand.

As a last example, here follows the input of the big display from the beginning. Note the use of defining macros with `define name @ macro body @`.

```
.EQ I
define T1 @ 1 over n sum from { k = 1 } to n @
D sub n ~ mark = ~  T1 left | ~ exp
     left ( k over n ~ 2 pi i  right ) ~-1 right | ~=
.EN
.EQ I
lineup = ~ T1 left | ~ left ( cos { 2 pi k } over n
     ~ + ~ i ~ sin { 2 pi k } over n right )
     ~ -1 right | ~=
.EN
.EQ I
lineup = ~ T1 left ( 2 ~ - ~ 2 cos  {2 pi k } over n
     right ) sup { 1 over 2 } ~=~ 2 over n
     sum from { k = 1 } to n sin { pi k } over n
.EN
```

Bibliographic references

An application of inverted indices on the UNIX system is the refer [3] preprocessor. This program makes it possible to give an imprecise citation in the text, which is translated into a proper reference. The database of citations searched may be tailored to each system, and individual users may specify their own citation files.

For example, the reference for TEXTSCHAAF was specified by

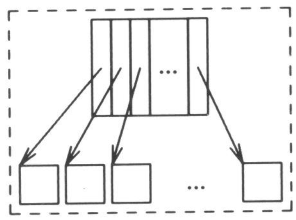
```
. . .
handled by the EQN
.[
tekstschaaf Grune
.]
preprocessor
. . .
```

There are various styles in which the references can appear, for instance, in a footnote, or in a list which can be sorted on various items at the end of the paper (as done here). Also the style of the labels used can be varied. Although refer provides the necessary information to do this, the actual implementation of how the output looks, as usual, is dependent on the macros package actually used.

Handling of graphics material

Typesetters are not really made for this, but it is sometimes possible to turn them into drawing machines. For drawing pictures there exists the PIC [4] preprocessor. It implements a language of the same name for specifying simple figures. The basic objects in PIC are boxes, circles, ellipses, lines, arrows, arcs, spline curves, and text. These may be placed anywhere, at positions specified absolutely or in terms of previous objects. Just like the other preprocessors, PIC only looks at the part between certain macros, `.PS` and `.PE`.

I will not explain all the details of the language in this article but I hope the next example, which has been taken straight out of the user manual, will give a general idea of how pictures are made.



This picture was generated with the following input:

```
   .PS
1  h = .5i
2  dh = .02i
3  dw = .1i
4  [
5      Ptr: [
6          boxht = h; boxwid = dw
7          A: box
8          B: box
9          C: box
10         box wid 2*boxwid "..."
11         D: box
12     ]
13     Block: [
14         boxht = 2*dw; boxwid = 2*dw
15         movewid = 2*dh
16         A: box; move
17         B: box; move
18         C: box; move
19         box invis "..." wid 2*boxwid; move
20         D: box
21     ] with .t at Ptr.s - (0,h/2)
22     arrow from Ptr.A to Block.A.nw
23     arrow from Ptr.B to Block.B.nw
24     arrow from Ptr.C to Block.C.nw
25     arrow from Ptr.D to Block.D.nw
26 ]
27 box dashed ht last [].ht+dw wid last [].wid+dw at last []
   .PE
```

The line numbers don't belong to the input, they have been inserted for the sake of the following explanation only.

The first three lines give values to later used constants. These values are in inches, but could have been scaled to centimetres if the `scale=2.54` had been issued first.

Any sequence of PIC statements may be enclosed in brackets `[...]` to form a block, which can than be treated as a single object, just as the braces are used for grouping with EQN. At line 4 a big block starts, consisting of two other blocks made again with the square brackets and some arrows.

These two blocks are labeled with `Ptr` and `Block`, which allows them to be referenced later by name. The labels refer to the center of the blocks.

The block `Ptr` is made out of several boxes which will get their sizes specified in line 6. At line 10 a box is specified twice the size of the default box (in the current block) with a text consisting of periods in the middle of it.

The second block `Block` is specified in a similar fashion. Note that on line 19 an invisible box is specified, which contains `...` as text.

Line 21 will place the block `Block` with its top (`.t`), at the South of block Ptr (`Ptr.s`), moved over the vector `(0,h/2)` downwards.

Line 22 specifies that an arrow be drawn from the center of box `A` in block `Ptr` to the North-West corner of the box `A` in block `Block`.

At the last line (27), a box is specified with a dashed line style and a height (`ht`) of the height of the last block plus something extra, (`[].ht+dw`). Also, the width is specified and finally this box is placed on top of the previous (conceptual) block, with the centers aligned (`at last []`).

More pictures

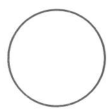
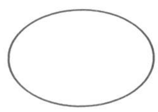
Yet another preprocessor is IDEAL [5]. It implements a language for describing pictures with the same name. It has similar capabilities as PIC.

Some conclusions

The UNIX typesettings tools produce a fine quality of printed material with a relatively small claim on the financial and manpower budget of a medium scale research institute like the CWI.

What really made the system a success is its modularity and flexibility. Separate modules are used for separate functions which implies that if you are not using a certain function, you leave out the module, and so you won't get a surprising effect when you accidentally trigger a function in the system you are not aware of.

An advantage of this separated functionality is that the next complicated looking output can easily be made. Mathematics as well as graphics are used inside the following table.

Text	Equation	Graphics
circle	$x^2 + y^2 = r^2$	
ellipse	$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$	

To have a text processing system as described here around, has its influence on the people dealing with it. Although the quality of the printed material is superior to what was produced before, there tends to be more criticism of it than before. Partly this criticism is justified. There are some defects in the system that need to be changed. The programs currently in use have actually never officially been released for use in a production environment. There is still work going on to make them more stable and to increase the quality of the output.

Also people tend to have difficulties proofreading material to be typeset. A lot of the time there are complaints about the appearance of the text, while the contents should be checked. Some people have strong objections to checking the contents of a text when the output is produced by simulating the typesetter on a cheaper raster style device. I guess that this is the price to pay for "in house" text processing.

Acknowledgements
I would like to thank the editors for commenting on this paper, their suggestions for improving my English and some of the used material.

References

[1] M. E. Lesk, "TBL—A program to format tables", *UNIX Programmer's manual* **2**, section 10, 1979.

[2] B. W. Kernighan, L. L. Cherry, "A System for Typesetting Mathematics", *CACM*, **51**(1975), 151-156.

[3] M. E. Lesk, "Some applications of inverted indexes on the UNIX system", *UNIX Programmer's manual* **2**, section 11, 1979.

[4] Brian W. Kernighan, *PIC — A Crude Graphics Language for Typesetting*, Computing Science Technical Report 85, Murray Hill, NJ, (revised edition), 1982. Also in: *SIGPLAN Symposium on Text Manipulation,* Portland, Oregon, 1981.

[5] Christopher J. van Wyk, "A Graphics Typesetting Language", *SIGPLAN Symposium on Text Manipulation,* Portland, Oregon, 1981.

Order Statistics, Quantile Processes and Extreme Value Theory

Oberwolfach Meeting from 25th to 31st March 1984

by R. Helmers

This meeting, organized by R.D. Reiss (Siegen) and W.R. van Zwet (Leiden), brought together 43 researchers working in three different, though closely related, fields in probability and statistics: order statistics, quantile processes and extreme value theory. In this report I shall survey a number of recent developments that were discussed at the conference. Specifically I shall deal with the following subjects: spacings theory, empirical and quantile processes, estimation of the tail of a distribution, extreme value theory, and a miscellaneous category.

Spacings theory

The talks in this section dealt with functions of uniform k-spacings. Statistics of this type play an important role in a number of different contexts, such as tests for uniformity, Poisson processes and non parametric density estimation.

Consider a sequence $U_1, U_2, \ldots$ of independent random variables (observations), each of them distributed according to the uniform distribution on [0,1], and let, for each $n \geq 1$, $U_{1:n} \leq \cdots \leq U_{n:n}$ denote the first n U_i's ordered in ascending order of magnitude. Set $U_{0:n} = 0$ and $U_{n+1:n} = 1$. *Uniform spacings* are defined by $D_{in} = U_{i+1:n} - U_{i:n}$ $(0 \leq i \leq n)$, the gaps induced by the 'random points' $U_1, \ldots, U_n$ in the interval [0,1], and, more generally, *uniform k-spacings* by $D_{ink} = U_{i+k:n} - U_{i:n}$ $(0 \leq i \leq n+1-k)$. Let $M_n = \max_{0 \leq i \leq n} D_{in}$, the maximal spacing or maximal gap, and $M_{nk} = \max_{0 \leq i \leq n+1-k} D_{ink}$ the maximal k-spacing. Of course $M_{n1} = M_n$.

As early as in 1939 P. Levy found the limit distribution of M_n: $\lim_{n \to \infty} P(nM_n \leq \log n + x) = \exp(-e^{-x})$ for all $x \in \mathbb{R}$. It is well-known that the distribution function $\exp(-e^{-x})$ has mean $\gamma = 0.5772 \cdots$, Euler's constant, and variance $\frac{\pi^2}{6}$. In Devroye (1981) it was noted that indeed the expected value $E(nM_n - \log n) \to \gamma$ and the variance $\sigma^2(nM_n) \to \frac{\pi^2}{6}$, as $n \to \infty$, as one may expect from Levy's result. However, the question remains: how small or large can the maximal gap M_n be as n gets large? In Devroye (1982) it was proved that, with probability 1,

$$\liminf_{n\to\infty} (nM_n - \log n + \log\log\log n) = -\log 2$$

and (1)

$$\limsup_{n\to\infty} \frac{nM_n - log\ n}{2\log\log n} = 1$$

At the conference Deheuvels (joint work with L. Devroye) proved similar strong limit laws ($n\to\infty$) for M_{nk} and related statistics, both when k is fixed and when it is allowed to increase with n at a rate not exceeding $\log n$. If k is *fixed*, then, with probability 1,

$$\liminf_{n\to\infty} \frac{nM_{nk} - \log n - (k-1)\log\log n}{\log\log\log n} = -1$$

and (2)

$$\limsup_{n\to\infty} \frac{nM_{nk} - \log n - (k-1)\log\log n}{2\log\log n} = 1$$

Note that (2) includes the case $k=1$, for which (2) is implied by the more refined result (1). Clearly k affects only the second order terms in (2). This means that if the maximal spacing M_n is either 'small' or 'large', the maximal k-spacing M_{nk} is likely to be of the same order of magnitude. If, on other hand, $k = k(n) \to \infty$, *but* $k = o(\log n)$, then, with probability 1,

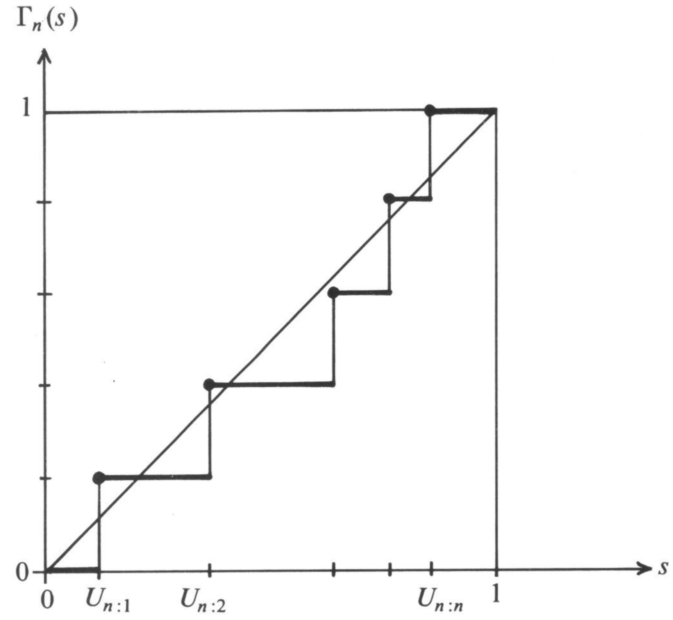
$$\frac{nM_{nk} - \log n}{(k-1)\log(\frac{e\log n}{k})} \to 1 \qquad (3)$$

Again, as in (2), nM_{nk} is of the order of $\log n$. A correction term $(k-1)\log(\frac{e\log n}{k})$ is also established. In reference [3] the deviation of $nM_{nk} - \log n$ from this correction term is studied. Also the case $k = c\log n$, for some constant $c > 0$, was considered in the talk as well as results for the non-uniform case.

Berry-Esseen bounds and Edgeworth expansions for statistics of the form $\Sigma_{i=0}^{n} g((n+1)D_{in})$ for a fixed function g, were derived by Does (joint work with R. Helmers and C.A.J. Klaassen). Klaassen proved a general limit theorem for conditional statistics, with an application to uniform k-spacings.

Empirical and quantile processes

In a three part talk M. Csörgö, S. Csörgö and D.M. Mason (joint work with L. Horvath), introduce a new Brownian Bridge approximation to the uniform empirical and quantile processes and discuss applications in probability and statistics. To explain their result we need a bit of notation. Let Γ_n denote



the empirical distribution function based on $U_1,...,U_n$; i.e. $\Gamma_n(s)$ is the proportion of the U_i's $(1 \leq i \leq n)$ which are $\leq s$ $(0 \leq s \leq 1)$. The *uniform empirical process* α_n is given by $\alpha_n(s) = n^{1/2}(\Gamma_n(s) - s)$, $0 \leq s \leq 1$. Also let Q_n be the uniform empirical quantile function; $Q_n(s) = U_{i:n}$ for $\frac{i-1}{n} < s \leq \frac{i}{n}$ $(1 \leq i \leq n)$ and $Q_n(0) = 0$. So Q_n is the left-continuous inverse of Γ_n. The *uniform quantile process* β_n is given by $\beta_n(s) = n^{1/2}(s - Q_n(s))$, $0 \leq s \leq 1$. Finally let $B(s)$, $0 \leq s \leq 1$, denote a *Brownian Bridge,* i.e. a real-valued, zero-mean Gaussian process with continuous sample paths and covariance function $EB(s)B(t) = \min(s,t) - st$, $0 \leq s,t \leq 1$.

A probability space is constructed with a sequence of independent uniform (0,1) random variables $U_1, U_2,...$ and a sequence of Brownian Bridges $B_1, B_2,...$ defined on it, such that for all $0 \leq \nu < 1/4$

$$\sup_{0 \leq s \leq 1} |\alpha_n(s) - \bar{B}_n(s)| / (s(1-s))^{1/2-\nu} = O_p(n^{-\nu}) \tag{4}$$

where $\bar{B}_n(s) = B_n(s)$ for $n^{-1} \leq s \leq 1 - n^{-1}$ and zero elsewhere; in addition one also has for all $0 \leq \nu < 1/2$

$$\sup_{1/n+1 \leq s \leq n/n+1} |\beta_n(s) - B_n(s)| / (s(1-s))^{1/2-\nu} = O_p(n^{-\nu}) \tag{5}$$

Here $O_p(n^{-\nu})$ has the standard meaning that n^{ν} times the l.h.s. of (4) and (5) remain bounded in probability as n gets large.

In a way these Brownian Bridge approximations improve upon the well-known 'KMT-embedding' for α_n, due to Komlos, Major and Tusnady and also known as the Hungarian embedding, and the parallel result for β_n; the improvement is in the 'tails', i.e. in neighbourhoods of zero and one. The constructions (4) and (5) were successfully applied to certain asymptotic problems involving the tails of α_n and β_n. Examples of such problems, which were mentioned in the talks, include: a refined Chibisov-O-Reilly theorem for the weak convergence of *weighted* empirical and quantile processes, a new proof of the Jaeschke-Eicker limit theorems, and central limit theorems for sums of extreme order statistics.

In the talk of Révész a very delicate strong invariance principle was presented for the *local time* γ_n of α_n; here γ_n is the number of zero crossings of α_n. Révész proved that one has with probability 1,

$$|n^{-\frac{1}{2}}\gamma_n - \eta(n)| = 0(n^{-\frac{1}{4}+\epsilon}) \tag{6}$$

for any $\epsilon > 0$, where the process $\eta(t)$, $t \geqslant 0$ denotes the (carefully defined) local time of a Kiefer process. As an application of (6) we have, with probability 1,

$$\limsup_{n\to\infty} \frac{\gamma_n}{\sqrt{n \log\log n}} = \frac{1}{\sqrt{2}} \tag{7}$$

Ruymgaart (joint work with J. Einmahl and J.A. Wellner) established a Chibisov O-Reilly result for the weak convergence of weighted empirical processes for the multidimensional case, both when the process is indexed by points (the classical case) and when it is indexed by rectangles. Steinebach proved an improved Erdös - Rényi strong law for moving quantiles.

The talks in this category discussed so far emphasise a probabilistic point of view. In contrast, the two part talk of Basset and Koenker contributes significantly to problems of statistical applications, e.g. in econometrics. Their idea is to estimate the error structure in linear models with the aid of an empirical regression quantile function. The behaviour of associated quantile processes is studied and applications discussed. Other talks in this category were by Boos (estimation of large quantiles) and by Falk (kernel type estimators of a population quantile).

Estimation of the tail of a distribution

Consider a distribution function F with regularly varying upper tail, i.e.

$$1 - F(x) = x^{-\lambda}L(x), \quad x > 0 \tag{8}$$

where $\lambda > 0$ denotes the tail index of F and L is slowly varying at infinity. The problem is to estimate λ on the basis of n independent observations $X_1,...,X_n$ from F. Hill [4] proposed the estimator

$$\lambda_n = (k^{-1}\,\Sigma_{i=1}^{k} \log X_{n-i+1:n} - \log X_{n-k:n})^{-1} \tag{9}$$

where $X_{1:n} \leqslant \cdots \leqslant X_{n:n}$ are the ordered X_i's. For simplicity we suppose $F(0) = 0$. The integer k $(1 \leqslant k \leqslant n)$ depends on n in such a way that

$$k = k(n) \to \infty, \quad k(n) = o(n) \tag{10}$$

Three talks were devoted to the problem of the asymptotic normality of λ_n. Haüsler showed that $\sqrt{(n)}\,(\lambda_n - \lambda)$ is asymptotically normally distributed, provided $k(n) \to \infty$ sufficiently slow. If some knowledge about L is available then the sequence $k = k(n)$ can be determined explicitly. Smith dealt with the same problem from a different point of view: $k = k(n)$ is now the random number of exceedances of a given threshold level x_n. Again to determine x_n information about L is required. S. Csörgö (joint work with D.M. Mason) showed that the Brownian Bridge approximation (5) can be applied to establish the asymptotic normality of λ_n.

While attending the meeting S. Csörgö, P. Deheuvels and D.M. Mason wrote paper [5]. In this paper these authors introduced a new class of estimators of λ, which can be viewed as the inverse of the convolution of a kernel function with the logarithm of the empirical quantile function of the X_i's. Asymptotic normality is established with the aid of (5) and optimal kernels are selected that give more weights to the extreme observations then Hill's λ_n.

Extreme value theory

Extreme value theory in concerned with the asymptotic behaviour of sample extremes. The central result in this area is of course Gnedenko's theorem stating that the limit distribution of the normalized maximum $X_{n:n}$ of independent and identically distributed random variables $X_1,...,X_n$, if it exists, must be one of three types: of the well-known extreme value distributions. Balkema (joint work with L. de Haan and S. Resnick) gave rate of convergence results for the case that the X_i's are in the domain of attraction of the extreme value distribution $\exp(-e^{-x})$.

It is known that under appropriate 'long range' and 'local' dependence conditions, classical extreme value theory remains true for dependent (stationary) sequences. Leadbetter discusses what happens if the 'local' dependence condition is relaxed. The answer turns out to be that the asymptotic distribution of the maximum is essentially unchanged, whereas the distributions of the other extreme order statistics are altered in a specific way. This phenomenon is caused by the fact that the point processes of high exceedances, which are classically Poisson, now involve clustering.

Other talks in this category were by de Haan (records from an improving population), Hüsler (extreme value theory for non-stationary sequences), Meizler (extreme value theory for independent but not identically distributed random variables), Lindgren (optimal prediction of upcrossing of a critical level by a stationary Gaussian process), Rootzén (behaviour of extremes of moving averages) and Tiago de Oliveira (extreme value theory for bivariate random variables).

Miscellaneous

Daniels discussed the joint distribution of the maximum of a random walk and the time at which it is attained. Asymptotically the problem is related to one involving Brownian motion in presence of a parabolic boundary. This bears some resemblance with recent work at CWI by P. Groeneboom and N.M. Temme on the statistical problem of estimating a monotone density.

The other talks dealt with the saddle point method and M-estimators (Dinges), the asymptotic behaviour of the L_1-error of density estimators (Györfi), generalized L-statistics (Helmers; joint work with P. Janssen and R.J. Serfling), order statistics in the non i.i.d. case and the influence of outliers (David, Gather, Mathar), order statistics in insurance mathematics (Teugels), order statistics and symmetric functions (Rüschendorf), test for a change point model (Mittal), and a multivariate two sample test based on nearest neighbours (Henze).

References

1 DEVROYE, L. (1981). Laws of the iterated logarithm for order statistics of uniform spacings. *Ann. of Probability 9,* 860-876.

2 DEVROYE, L. (1982). A log log law for maximal uniform spacings. *Ann. of Probability 10,* 863-868.

3 DEHEUVELS, P. & L. DEVROYE (1984). Strong laws for the maximal k-spacing when $k \leqslant c \log n$. *Z. Wahrscheinlichheitstheorie verw. Gebiete 66,* 315-334.

4 HILL, B.M. (1975). A simple approach to inference about the tail of a distribution. *Ann. of Statistics 3,* 1163-1174.

5 CSÖRGÖ, S., P. DEHEUVELS, D.M. MASON (1984). *Kernel estimates of the tail index of a distribution,* Techn. report Laboratoire de Statistique Theorique et Applique, Université Pierre et Marie Curie, Paris.

6 REISS, R.D. (1984). *Tagungsbericht 14,* Mathematisches Forschungs institut, Oberwolfach.

Abstracts

of Recent CWI Publications

When ordering any of the publications listed below please use the order form at the back of this issue.

MC Tract 164. F.A. van der Duyn Schouten. *Markov Decision Processes with Continuous Time Parameter.*
AMS 60B10, 60G17, 60J25, 90B05, 90B25, 90C47, 93E20; 195pp.
Abstract: We deal with continuous time Markov decision drift processes (CTMDP), which permit both controls affecting jump rates of the process and impulsive controls causing immediate transitions. Between two successive jump epochs the state of the process evolves according to a deterministic drift function. Given a CTMDP we construct a sequence of discrete time Markov decision drift processes (DTMDP) with decreasing distance between the successive decision epochs. Sufficient conditions are given under which the law of the CTMDP controlled by a fixed policy is the limit (in the sense of weak convergence of probability measures) of the laws of the approximating DTMDPs controlled by fixed discrete time policies. These conditions concern both the parameters of the CTMDP and the relation between the discrete time and continuous time policies. Based on these conditions for weak convergence we derive relations between the discrete time and the continuous time optimal policies for the discounted cost criterion. In this way discounted optimality results for a CTMDP are obtainable from the corresponding results of a DTMDP. These discounted optimality results for a CTMDP are in turn the base for the analysis of the average cost criterion. Applications to an *M/M/1* queueing model, a maintainance-replacement model and an inventory model are given.

CWI Tract 1. D. Epema. *Surfaces with Canonical Hyperplane Sections.*
AMS 14C30, 14J10, 14J15, 14J17, 14J25; 106pp.
Abstract: We study a special case of surfaces, characterized by the property that they can be embedded in some projective space in such a way, that a general hyperplane section is a canonical curve. The main examples are quartics in $\mathbb{P}^3$ with isolated singularities and double covers of $\mathbb{P}^2$ branched along a sextic. First, we derive general properties of these surfaces, but for the greatest part we focus on those surfaces with this property, which are birational to a ruled surface over an elliptic curve. This leads to a detailed description and classification of all quartics in $\mathbb{P}^3$ with either a singularity of genus 2, or with two simple elliptic singularities, and so the construction of a moduli variety for the double covers of $\mathbb{P}^2$. We conclude with a discussion of the mixed Hodge structure on the cohomology groups of a certain open part of the minimal resolution of these surfaces, and the description of a period map for the double planes.

CWI Tract 3. A.J. van der Schaft. *System Theoretic Descriptions of Physical Systems.*
AMS 93-02, 93B05, 93B10, 93C05, 93C10, 58-02, 58F05, 70D05, 70H05; 256pp.
Abstract: In this monograph the dynamical behaviour of physical systems is described *including* their interaction with the environment. This interaction or external behaviour is given as the time-evolution of the external variables, just as the internal behaviour is given by the state variables. In order to do so the usual framework of mathematical system theory is enlarged by not requiring a priori that the external variables are split into inputs and outputs. This framework is explored for

set-theoretic, linear and smooth nonlinear systems, emphasizing a "geometric" approach. The central part of the monograph is concerned with the treatment of Hamiltonian systems from this point of view. Roughly speaking the symplectic formulation of the Hamilton and Euler-Lagrange equations without external forces is extended to the original equations *with* external forces. Gradient systems are similarly treated. Furthermore, symmetries, conservation laws and time-reversibility are dealt with in this setting, and the developed theory of Hamiltonian systems is applied to optimal control.

CWI Tract 4. J. Koene. *Minimal Cost Flow In Processing Networks, A Primal Approach.*

AMS 90B10, 90C35; 157pp.

Abstract: In the past decades the so-called pure and generalized network flow problems have been investigated thoroughly. This tract is concerned with a more general type of network flow problems, namely processing network problems. In a processing network it is allowed that a given flow splits up proportionally in a number of components (a refining process), or in reverse, that a number of components is blended in given proportions (a blending process). Moreover, the same requirements hold as in pure or generalized networks, giving rise to two types of processing networks: pure and generalized networks. The processing network structure appears among others in production planning environments in the process industry and in energy, assembly and economic models. Since processing network problems can be formulated as LP-problems, they can be solved by standard LP-programs. However, in large size applications, much computer time and storage requirements might be saved if solution techniques were available in which the (processing) network structure is exploited. Primal simplex algorithms for both pure and generalized processing network problems are developed. We present a characterization of a basis in terms of the network and yield a natural way to partition a basis. The steps of the simplex algorithm can partially be done by using techniques developed for pure and generalized networks. On the other hand a so-called working basis is required. This working basis has a size equal to the number of basis refining and blending processes and is maintained in block triangular form. Using these results we provide an algorithm for solving processing network problems with additional linear constraints. The relation between processing networks and general linear programming is discussed. It is shown that any linear programming problem can be transformed to a pure processing network problem.

CWI Tract 5. B. Hoogenboom. *Intertwining Functions on Compact Lie Groups.*

AMS 17B05, 22F30, 22F46, 33A65, 33A75, 43A90; 86pp.

Abstract: We consider a generalization, called intertwining functions, of spherical functions on a compact semisimple Lie group. These are matrix coefficients of some irreducible representation of U, which are left - K -, and right - H - invariant; K and H being the fixed point groups of two commuting involutions on U. It is shown that the intertwining functions on U can be considered as orthogonal polynomials on some region in $\mathbb{R}^l$ with respect to a certain positive weight function. This generalizes results of Vretare in the spherical case. In the course of this, a structure theory for real semisimple Lie algebras with two commuting involutions is developed. We also prove a generalized Cartan decomposition for U, and finally the orthogonal polynomials constructed before are characterized as eigenfunctions of the K,H-radial part of the Laplace-Beltrami operator on U.

CWI Tract 6. A.P.W. Böhm. *Dataflow Computation.*

AMS 68A05, 68F10, 68F20; CR C.1.3, D.3.2, F.1.1, F.2.1, F.3.2; 208 pp.

Abstract: This monograph is devoted to dataflow computation, a particular kind of parallel computation. Chapter One gives a short overview of parallel computers and their underlying model of computation. The dataflow model of computation is discussed in some detail. A dataflow net is a directed graph in which the nodes represent processing elements and the edges represent data paths. An existing dataflow machine, the Manchester Dataflow Machine, is discussed.
Chapter Two introduces an elementary dataflow model, which differs from the widely accepted model of Rodriquez and Adams in that it mirrors the time-dependent, non-functional behaviour of

dataflow machines. It is shown that for so called well-formed dataflow nets the asynchronous, parallel execution mode does not lead to non-functional behaviour. The model is shown to have universal computing power. Other models of computation are simulated.
Chapter Three introduces a programming language with explicit parallelism at the procedure level. A program in execution is a dynamically changing network of processes which communicate with each other via channels (hence the name of the language: DNP for Dynamic Networks of Processes).
Chapter Four presents a number of algorithms in DNP, which are believed to be prototypical for dataflow computing. The complexity of the algorithms is analysed. It is shown that not all classes of computation graphs can be generated by DNP. Ways to overcome this are indicated. A comparison is made with the standard complexity classes.
In Chapter Five DNP programs from Chapter Four are proved correct. The proofs are based upon a semantics of DNP according to Kahn's ideas.

CWI Tract 7. A. Blokhuis. *Few-Distance Sets.*
AMS 05-02, 05B25, 05B35, 06C10, 51K99, 51M10, 52A40; 70pp.
Abstract: In this tract bounds are derived for the cardinality of point sets with few distances in Euclidean space and on the "unit-sphere" of inner product spaces with arbitrary signature. Especially the case of hyperbolic space, R^d, yields sharp bounds, also for the related problem of equiangular lines. A theorem of Frankl and Wilson concerning few distances modulo a prime is generalized to Delsarte spaces. A problem of Erdös on few-distance point sets is shown to be related to 2-distance sets and solved. It is shown that graphs satisfying: 1) $\exists k$ such that each maximal clique has k points; 2) $\exists e$ such that each point p has e neighbours in any maximal clique not containing p, have a highly geometric structure, resembling that of polar spaces.

CWI Syllabus 1. *Vacation-course 1984: Hewet-plus mathematics.*
233pp. (in Dutch).
Abstract: This syllabus, for high-school teachers of mathematics, contains fourteen articles on operations-research, matrix algebra, statistics and probability theory, geometry in three-dimensional space, mathematical modeling, and computer science.

CS-R8408. J.A.M. van de Graaf. *Towards a specification of the B programming environment.*
AMS 68B20; 23 pp.; **key words:** programming environments, personal computing, *B*.
Abstract: A primary aim in the design of a dedicated environment for the *B* programming language has been to make the environment easy to use and to learn. We tried to achieve this by tuning the features to the expected use, by integrating the language and the environment and by allowing only a limited number of powerful, coherent concepts which have in varying context a similar meaning. This report contains an informal description and a tentative specification of the environment.

CS-R8409. J.A. Bergstra & J.W. Klop. *Process algebra for communication and mutual exclusion. Revised version.*
AMS 68B10; 33 pp.; **key words:** concurrency, communicating processes, process algebra, merge, critical regions, asynchronous cooperation, synchronous cooperation.
Note: This report is a revised version of IW 218/83.
Abstract: Within the context of an algebraic theory of processes we provide an equational specification of process cooperation. We consider four cases: free merge or interleaving, merging with communication, merging with mutual exclusion of critical sections and synchronous process cooperation. The rewrite system behind the communication algebra is shown to be confluent and

terminating (modulo its permutative reductions). Furthermore, some relationships are shown to hold between the four concepts of merging.

CS-R8410. J.A. Bergstra, J.W. Klop & J.V. Tucker. *Process algebra with asynchronous communication mechanisms*
AMS 68B10; 23 pp.; **key words:** concurrency, communicating processes, asynchronous communication, stimulus-response mechanism.
Abstract: Algebraic specifications are given for a stimulus-response mechanism, for asynchronous non-order-preserving send and receive actions and for asynchronous order-preserving send and receive statements.

CS-R8411. J.A. Bergstra, J. Heering & J.W. Klop. *Object-oriented algebraic specification: proposal for a notation and 12 examples.*
AMS 68C01; 25 pp.; **key words:** object-oriented specification, algebraic specification, configuration transition system, transformation rule.
Abstract: A notation is introduced for expressing the dynamic behaviour of configurations of objects. At each instant of time a configuration is just a multi-set of objects which themselves are points (values) from some algebraically specified abstract data type. Several examples should convince the reader of the attractive expressive power of our notation.

CS-N8402. L.J.M. Geurts. *Computer programming for beginners. Introducing the B language, Part I.*
CR D.3.3; 85 pp.; **key words:** programming, programming languages, *B*.
Abstract: This report contains part 1 of a first course in programming based on the new programming language *B*. *B* is a programming language designed specifically with the beginner, and all other 'non-professional' computer users, in mind. Most elementary programming techniques and most of the features of *B* are presented. The focus is on designing and writing programs, not on actually entering programs in the computer, changing those programs, etc.. Many short programs are shown, or asked to be written by the reader as exercises. The text is self-contained, and may be used in courses or for self-study.

CS-N8403. C.L. Blom, A. Chaudry, P.J.W. ten Hagen, L.O. Hertzberger, A. Janssen, A.A.M. Kuyk, F. Tuynman, & W.E. van Waning. *A strategy for computer integrated manufacturing systems: Processing and communication.*
AMS 69L60; 132 pp.; **key words:** Computer Integrated Manufacturing (CIM), CAD/CAM, Flexible Manufacturing System (FMS), sensor systems, robot vision, solid modeling, design, computer networks, processing technology, communication technology.
Abstract: In 1983 the European Economic Commission launched a research program called 'European Strategic Program for Research and Development in Information Technology' (ESPRIT). Since manufacturing is an important application area of Information Technology, attention has to be paid to the evolving integration of Computer Aided Design (CAD), and Computer Aided Manufacturing (CAM). In order to understand and optimize this process, a research project was initiated to establish 'Design Rules for Computer Integrated Manufacturing (CIM) Systems'. As a first step, the state-of-the-art in two main areas of Information Technology (Processing and Communication Technology) was investigated with respect to key modules of both CAD and CAM, in order to identify well-founded strategies aimed at optimal support for the various Computer Integrated Manufacturing activities, especially from the point of view of Processing and Communication Technology.

CS-N8404. S. Pemberton. *A user's guide to the B system.*

AMS 69D49; 10 pp.; **key words:** programming languages, *B*, programming environments.
Abstract: *B* is a new interactive programming language being developed at the CWI. This report gives a brief introduction to the use of the current *B* implementation. It does not teach about the language, for which you should refer to CS-N8402.

CS-N8406. L.G.L.T. Meertens & S. Pemberton. *An implementation of the B programming language*
AMS 69D44; 8 pp.; **key words:** programming language implementation, programming environments, *B*.
Abstract: *B* is a new programming language designed for personal computing. We describe some of the decisions taken in implementing the language, and the problems involved.

OS-R8406. P.J.C. Spreij. *An on-line parameter estimation algorithm for counting process observations.*
AMS 60G55; **key words:** counting process, martingale, estimation, stochastic Lyapunov function.
Abstract: The parameter estimation problem for counting process observation is considered. It is assumed that the intensity of the counting process is adapted to the family of σ-algebras generated by the counting process itself. Furthermore we assume that the intensity depends linearly on some deterministic constant parameters. An on-line parameter estimation algorithm is then presented for which convergence is proved by using a stochastic approximation type lemma.

OS-R8407. J.P.C. Blanc, E.A. van Doorn. *Relaxation times for queueing systems.*
AMS 60K25; 20 pp.; **key words:** queueing system, relaxation time, time-dependent behaviour.
Abstract: When a stochastic queueing model is used for performance analysis of, e.g., a computer or communication system, the steady-state situation is usually assumed to prevail. Since many systems exist where the validity of this assumption is questionable, while determination of the time-dependent behaviour of the system is difficult or even impossible, some simple means of characterizing the speed with which system performance measures tend to their steady-state values are called for. In this paper the concept of relaxation time is put forward to provide such a characterization. We give a survey of results pertaining to relaxation times for a variety of queueing models. Also, some conjectures and open problems are mentioned.

OS-R8408. J.P.C. Blanc. *The transient behaviour of networks with infinite server nodes.*
AMS 90B22; 18 pp.; **key words:** Jackson network, transient behaviour, relaxation time.
Abstract: How long does it take before the number of jobs in a Jackson network is within, say, 1% of its steady-state value, when the system starts working at time 0 with given numbers of jobs at the various nodes? How is this amount of time related to the relaxation time of the mean number of jobs in the network? These questions are discussed in detail for networks with infinitely many servers at each node. Conjectures are formulated for networks with finitely many servers at the nodes on the basis of similarities with the infinite case for networks with one or two nodes.

OS-N8401. J.M. Anthonisse. *Proportional representation in a regional council.*
AMS 90B10; 16 pp. (in Dutch); **key words:** Proportional representation.
Abstract: The members of a regional council are appointed by and from the local councils which

participate in the regional cooperation. The regional council should constitute a fair representation of the local interests and also a fair representation of the political views. A method is presented for determining the number of members to be appointed by each local council from each political party. The method consists of solving two (maxflow-mincost) network flow problems.

NM-R8404. J.G. Verwer & J.M. Sanz-Serna. *Convergence of method of lines approximations to partial differential equations.*
AMS 65X02; 14 pp.; **key words:** initial value problems, partial differential equations, method of lines, stiff differential equations, nonlinear problems, convergence analysis.
Abstract: Many existing numerical schemes for evolutionary problems in partial differential equations (PDEs) can be viewed as method of lines (MOL) schemes. This paper treats the convergence of one-step MOL schemes. Our main purpose is to set up a general framework for a convergence analysis applicable to nonlinear problems. The stability materials for this framework are taken from the field of nonlinear stiff ordinary differential equations. In this connection, important concepts are the logarithmic matrix norm and C-stability. A nonlinear parabolic equation and the cubic Schrödinger equation are used for illustrating the ideas.

NM-R8405. J.M. Sanz-Serna & J.G. Verwer. *Conservative and nonconservative schemes for the solution of the nonlinear Schrödinger equation.*
AMS 65M10; 24 pp.; **key words:** numerical analysis, nonlinear Schrödinger equation, time integration, conservation of energy.
Abstract: Five methods for the integration in time of a semi-discretization of the nonlinear Schrödinger equation are extensively tested. Three of them (a partly explicit scheme and two splitting procedures) are found to perform poorly. The reasons for their failure, including the so-called nonlinear blow-up, are analyzed. We draw general conclusions on the advantages and drawbacks associated with the use of time-integrators which exactly conserve energy.

NM-R8406. F.W. Wubs. *Evaluation of time integrators for shallow-water equations.*
AMS 65M20; 32 pp.; **key words:** method of lines, linear stability, shallow-water equations, efficiency of time integrators, storage reduction.
Abstract: In this report some well-known time integrators for shallow-water equations are described and compared with each other with respect to efficiency. In particular, the efficiencies of low-order versus high-order methods and of implicit methods versus explicit methods are considered. Numerical experiments for a problem posed by Grammeltvedt showed a good performance of the classical fourth-order Runge-Kutta method. However, conclusions should be drawn with care because the influence of the error due to the space discretization is not taken into account.

NM-R8407. J. Kok & G.T. Symm. *A proposal for standard basic functions in Ada.*
AMS 69D49; 9 pp.; **key words:** Ada, high level language, basic mathematical functions, scientific libraries, portability.
Abstract: This paper contains a proposal for a standard basic mathematical functions package for scientific computation in Ada. The package is transportable to machines with different floating-point types and its availability will enhance the portability of numerical software.

MS-R8409. A.W. Ambergen. *Asymptotic distributions of estimators for posterior probabilities in a classification model with both continuous and discrete variables.*
AMS 62H30; 26 pp.; **key words:** estimating posterior probabilities, discri-

minant analysis.

Abstract: This paper is devoted to the asymptotic distribution of estimators for the posterior probability that an observation vector originates from one of k populations. The estimators are based on training samples. The random vectors contain both continuous and discrete variables. Observation vector and prior probabilities are regarded as given constants. The continuous part of the random vector has conditional on the discrete part a multivariate normal distribution. Several assumptions about homogeneity of the dispersion matrices are considered.

MS-R8410. R.D. Gill *A note on two papers in dependent central limit theory.*
AMS 60F17; 9 pp.; **key words:** martingale central limit theorems, Poisson random measures, approximation of population processes, Markov processes, semi-Markov processes.

Abstract: In this note a discussion is given of two papers on dependent central limit theory which were presented at the 44th meeting of the International Statistical Institute in Madrid, 1983. The papers are by I.S. Helland, "Applications of Central Limit Theorems for Martingales with Continuous Time", and by T.G. Kurtz, "Gaussian Approximations for Markov Chains and Counting Processes". The note addresses itself to the main differences between the approaches described by Helland and Kurtz, called the martingale approach and the Poisson process approach respectively, and discusses their application to statistical problems in survival, life-testing, demography, epidemiology, etc.

AM-R8409. N.M. Temme. *A convolution integral equation solved by Laplace transformation.*
AMS 45A05; 5 pp.; **key words:** integral equations of convolution type, Laplace transformations, analytic continuation, Airy functions.

Abstract: We consider the integral equation

$$p(t) = \int_0^t K(t-\tau)p(\tau)d\tau + r(t)$$

where both $K(t)$ and $r(t)$ behave as $exp(\alpha t^3)$ as $t \to \infty$ ($\alpha > 0$). So straightforward application of the Laplace transform technique is not possible. By introducing a complex parameter the equation is solved in the complex domain. Analytic continuation with respect to this parameter yields the desired solution. For a particular example (which arose in a statistical problem on estimating monotone densities) we describe the construction of the explicit solution of the equation.

AM-N8401. J. Grasman & R. van der Horst. *Automatic data-handling in studies of cell-proliferation using time lapse cinematography.*
AMS 92-04; 4 pp. (in Dutch); **key words:** cell-proliferation, branching process.

Abstract: Data of cell proliferation is registered in the form of a family tree. For the mathematical modeling of the cell cycle it is necessary to have this data available for automatic handling. This report gives a method for constructing such a data file and for tracing possible errors. Moreover, a graphical reconstruction of the family tree from the data is made.

AM-N8402. L.L.M. van der Wegen. *Asymptotic expansions for modified Bessel functions of large order.*
AMS 33A40; 39 pp.; **key words:** asymptotic expansion, modified Bessel function, error bound for the remainder.

Abstract: In this report asymptotic expansions for modified Bessel functions of large order are derived by applying the saddle-point method to well-known integral representations for $K_\nu(x)$ and $I_\nu(x)$, in the case $\nu > 0$ and $x \geqslant 0$. Moreover, error bounds for the remainders are given. Finally, a comparison is made between these error bounds and the real errors that are made by

approximating the Bessel functions by parts of their asymptotic series.

PM-R8406. C.F. Dunkl. *Orthogonal polynomials and a Dirichlet problem related to the Hilbert transform.*
AMS 33A65; 31 pp.; **key words:** Pollaczek polynomials, Hilbert transform, shift operator, Fourier series, harmonic functions, Dirichlet problem, Poisson kernel, groups of Möbius transformations, Heisenberg group.
Abstract: An operator closely related to the Hilbert transform on the circle is shown to be unitarily equivalent to a shift realized on a basis of Pollaczek polynomials, a family of orthogonal polynomials with weight supported by all of the real line. There is an associated Dirichlet problem for the disk, where one wants to find harmonic functions with specified boundary values on the upper half circle and with specified constant (possibly complex) directional derivative on the real diameter. The Poisson kernel is found and is used to obtain continuous and L^p function existence and boundedness results. The Dirichlet problem is a limiting case of boundary value problems for certain special functions coming from the Heisenberg groups.

PM-N8401. M. Hazewinkel. *Notes on (the philosophy of) linearization.*
AMS 93C10; 14 pp.; **key words:** linearization of differential equations, linearization of control systems, embedding control systems.
Abstract This note, originally a handout for the special session on linearization at the 1983 Conference on Decision and Control in San Antonio, consists of a discussion and initial guide to the literature on linearization in various parts of mathematics. Special attention is paid to linearization of control systems.

CWI Activities

Autumn 1984

With each activity we mention its frequency and (between parentheses) a contact person at CWI. Sometimes some additional information is supplied, such as the location if the activity will not take place at CWI.

Topology Day CWI.
25 years ago Prof.dr. P.C. Baayen, now director of CWI, joined the Mathematical Centre. To celebrate this anniversary we are holding a 'Topology Day' on 28 September.
Topology is a relatively young science. This makes it all the more remarkable that its basic concepts, results and vocabulary play a major role in most of the major branches of mathematics. Indeed, it is nowadays very hard to conceive of doing analysis, geometry, algebra or even logic without a substantial involvement of topology. This makes it a prime example for illustrating what is probably the most important present trend in the mathematical sciences: the increasing interweaving of formerly separated disciplines. The day aims at illustrating this multi-usefulness of topology.
Invited speakers:
R.D. Anderson: Infinite dimensional topology and analysis.
J. van Mill: Set theory and topology.
W.T. van Est: Algebraic topology or combinatorial topology?
M. Hazewinkel: Topology in algebra.
N.H. Kuiper: How convex can knots and surfaces be?

Study group on Analysis on Lie groups. Semisimple pseudo-Riemannian symmetric spaces. Joint with University of Leiden. Biweekly. (T.H. Koornwinder)

Seminar on Algebra and Geometry. Distance regular graphs. Biweekly. (A.E. Brouwer)

Study group on Cryptography. Biweekly. (A.E. Brouwer)

Colloquium 'STZ' on System Theory, Applied and Pure Mathematics. Twice a month. (J. de Vries)

Study group 'Biomathematics'. Lectures by visitors or members of the group. Joint with University of Leiden. (J. Grasman)

Study group 'Nonlinear Analysis'. Lectures by visitors or members of the group. Joint with University of Leiden. (O. Diekmann)

Progress Meetings of the Applied Mathematics Department. New results and open problems in biomathematics, mathematical physics and analysis. Weekly. (N.M. Temme)

Study group 'Semiparametric Estimation Theory'. Lectures by members of the group on non-parametric maximum likelihood estimators, density estimation, etc. Biweekly. (R.D. Gill)

National Study Group on Statistical Mechanics. Joint with Technological University of Delft, Universities of Leiden and Groningen. Monthly. University of Amsterdam. (H. Berbee)

Progress meetings of the Mathematical Statistics Department. New results in research and consultation projects. Monthly. (R.D. Gill)

Lunteren meeting on Stochastics. 12,13,14 November 1984 at 'De Blije Werelt', Lunteren. (R. Helmers)

Progress meetings on Combinatorial Optimization. Biweekly. (J.K. Lenstra)

System Theory Days. Irregular. (J.H. van Schuppen)

Study group on System Theory. Biweekly. (J.H. van Schuppen)

National colloquium on Optimization. Irregular. (J.K. Lenstra)

Study group on Differential and Integral Equations. Lectures by visitors or group members. Biweekly. (H.J.J. te Riele)

Study group Numerical Flow Dynamics. Lectures by group members. Every Wednesday. (J.G. Verwer)

Study group Hyperbolic systems. Every Wednesday. (P.W. Hemker)

Conference on Numerical Mathematics. 15,16,17 October 1984 at Zeist.
Invited speakers:
B. van Leer (Technological University Delft),
J.J. Chattot (MATRA, Velizy, France),
H. Viviand (ONERA, Paris),
R. Beauwens (Free University, Brussel),
J. Periaux (Avions Marcel Dassault, Paris),
D. Young (University of Texas, Austin),
A.O.H. Axelsson (University of Nijmegen),
H. Schippers (NLR, Emmeloord). (J.G. Verwer)

Seminar National Concurrency Project. Joint with Universities of Leiden, Utrecht, Nijmegen and Amsterdam. 5 October, 2 November and 7 December. University of Leiden. (J.W. de Bakker)

National Study Group 'Concurrency'. Joint with Universities of Leiden, Utrecht, Nijmegen and Amsterdam. 28 September, 19 October, 16 November and 14 December. University of Utrecht. (J.W. de Bakker)

Post-academic course on Modern Techniques in Software Engineering. 1,2 and 15,16 November. (J.C. van Vliet)

Post-academic course on Design of Interactive Graphical Systems. (P.J.W. ten Hagen)

Post-academic course on *B*. 3 September - 13 December. Twice a week. (L. Geurts)

Study group on Graphics Standards. Monthly. (M. Bakker)

Study group 'Dialogue programming'. (P.J.W. ten Hagen)

Study group 'Data Flow Club'. Irregular. (J. Heering)

Visitors to CWI from abroad

R.C. Bradley (Indiana University, USA) 1-5 July. **M.M. Chawla** (University of New Dehli, India) 18-19 July. **M.H.A. Davis** (Imperial College, London, UK) 1 August. **A. Feldstein** (Arizona State University, USA) 27 July. **N. Frances** (Technion, Haifa, Israel) 1 August - 7 September. **P.R. Freeman** (University of Leicester, UK) 9 July. **R.L. Griess Jr.** (University of Michigan, Ann Arbor, USA) August. **D. Griffiths** (University of Dundee, Scotland) 26 September. **M. Gyllenberg** (Helsinki University of Technology, Finland) August 1984 - July 1985. **Jiang Furu** (Fudan University, Shanghai, China) 25-26 June. **P.R. Kumar** (University of Maryland, Baltimore, USA) 13 July. **C. Meaney** (University of Adelaide, Australia) 19-21 June. **N. Megiddo** (Stanford University, USA) 30 August. **B. Neta** (Texas Techn. University, Lubbock, USA) 16-20 July. **A. Neumaier** (Albert-Ludwig University, Freiburg, West Germany) 15 August - 12 October. **C.F. Nourani** (G.T.E. Labs., Waltham, USA) 14 September. **J.B. Orlin** (MIT, Cambridge, USA) 1 September 1984 - 1 May 1985. **L. Petzold** (Sandia Laboratories, Albuquerque, USA) 7 September. **I.H. Sloan** (University of Canberra, Australia) 23 July. **K. and R.P. Soni** (University of Tennessee, Knoxville, USA) 5-6 July. **A.M. Spaccamela** (University of Rome, Italy) 31 July. **H.R. Thieme** (University of Heidelberg, West Germany) August 1984 - July 1985. **K. Tomoeda** (Osaka Institute of Techn., Japan) 5-6 July. **C. Vercellis** (University of Milan, Italy) 22 June. **J. Walrand** (University of California, Berkeley, USA) 27 July. **H.O. Walther** (Ludwig-Maximillians University, Munich, West Germany) 19-20 July. **J.A. Wellner** (University of Washington, Seattle, USA) 19 July - 20 August. **Zhu Hong** (Fudan University, Shanghai, China) 11-13 September.

Order form for CWI publications

Centre for Mathematics and Computer Science
Kruislaan 413
1098 SJ AMSTERDAM
The Netherlands

☐ Please send the reports marked below on an exchange basis
☐ Please send the reports marked below with an invoice

	Publication code	Price per copy	Number of copies wanted
☐	MC Tract 164	Dfl. 28.60	....
☐	CWI Tract 1	15.50	....
☐	CWI Tract 3	36.90	....
☐	CWI Tract 4	22.70	....
☐	CWI Tract 5	13.20	....
☐	CWI Tract 6	29.80	....
☐	CWI Tract 7	10.80	....
☐	CWI Syllabus 1	33.50	....
☐	CS-R8408	3.70	....
☐	CS-R8409	4.80	....
☐	CS-R8410	3.70	....
☐	CS-R8411	3.70	....
☐	CS-N8402	11.90	....
☐	CS-N8403	19.10	....
☐	CS-N8404	3.70	....
☐	CS-N8406	3.70	....
☐	OS-R8406	3.70	....
☐	OS-R8407	3.70	....
☐	OS-R8408	3.70	....
☐	OS-N8401	3.70	....
☐	NM-R8404	3.70	....
☐	NM-R8405	3.70	....
☐	NM-R8406	4.80	....
☐	NM-R8407	3.70	....
☐	MS-R8409	3.70	....
☐	MS-R8410	3.70	....
☐	AM-R8409	3.70	....
☐	AM-N8401	3.70	....
☐	AM-N8402	4.80	....

	Publication code	Price per copy	Number of copies wanted
☐	PM-R8406	4.80	....
☐	PM-N8401	3.70	....

If you wish to order any of the above publications please tick the appropriate boxes and return the completed form to our Sales Department.
Don't forget to add your name and address !
Prices are given in Dutch guilders and are subject to change without notice. Foreign payments are subject to a surcharge per remittance to cover bank, postal and handling charges.

Name ....

Street ....

City ....

Country ....

CWI NEWSLETTER CWI NEWSLETTER CWI NEWSLETTER

Contents