SRC TR 87-133

Morphological Filters - Part 1: Their Set-Theoretic Analysis and Relations to Linear Shift-Invariant Filters

by

P. Maragos and R.W. Schafer

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **20 FEB 1987**	2. REPORT TYPE	3. DATES COVERED **00-00-1987 to 00-00-1987**

4. TITLE AND SUBTITLE **Morphological Filters - Part 1: Their Set-Theoretic Analysis and Relations to Linear Shift-Invariant Filters**	5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Harvard University,Division of Applied Sciences,Cambridge,MA,02138**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
see report

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**49**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

Morphological Filters - Part 1: Their Set-Theoretic Analysis and Relations to Linear Shift-Invariant Filters

Petros Maragos
Division of Applied Sciences
Harvard University
Cambridge, MA 02138, U.S.A.
(617) 495-4390

Ronald W. Schafer
School of Electrical Engineering
Georgia Institute of Technology
Atlanta, GA 30332, U.S.A.
(404) 894-2917

March 24, 1987

Accepted for publication in the IEEE Transactions on ASSP, March 1987.

First submission: May 6, 1986

Revised and resubmitted: February 20, 1987.

EDICS Categories: 3.2.1, 3.2.3

---

*This work was supported by the Joint Services Electronics Program under Contract DAAG-84-K-0024 at the Georgia Institute of Technology, and in part by the National Science Foundation under Grant CDR-85-00108 at Harvard University.

Abstract

This paper examines the set-theoretic interpretation of morphological filters in the framework of mathematical morphology and introduces the representation of classical linear filters in terms of morphological correlations, which involve supremum/infimum operations and additions. Binary signals are classified as sets and multilevel signals as functions. Two set-theoretic representations of signals are reviewed. Filters are classified as set-processing (SP) or function-processing (FP). Conditions are provided for certain FP filters that pass binary signals as binary to commute with signal thresholding, because then they can be analyzed and implemented as SP filters.

The basic morphological operations of set erosion, dilation, opening, and closing are related to Minkowski set operations and are used to construct FP morphological filters. Emphasis is then given to analytically and geometrically quantifying the similarities and differences between morphological filtering of signals by sets and functions; the latter case allows the definition of morphological convolutions and correlations. Toward this goal various properties of FP morphological filters are also examined.

Linear shift-invariant filters (due to their translation-invariance) are uniquely characterized by their kernel, which is a special collection of input signals. Increasing linear filters are represented as the supremum of erosions by their kernel functions. If the filters are also discrete and have a finite-extent impulse response, they can be represented as the supremum of erosions only by their minimal (with respect to a signal ordering) kernel functions. Stable linear filters can be represented as the sum of (at most) two weighted suprema of erosions. These results demonstrate the power of mathematical morphology as a unifying approach to both linear and nonlinear signal-shaping strategies.

1 INTRODUCTION

Morphological filters are nonlinear signal transformations that locally modify *geometric* features of signals. They stem from the basic operations of a set-theoretical method for image analysis, called *mathematical morphology*, which was introduced by Matheron [1] and Serra [2]. In this method each signal is viewed as a *set* in a Euclidean space, and the morphological filters are set operations that transform the graph of the signal and can provide a quantitative description of its geometrical structure. For binary signals (viewed as sets), the erosion (contraction), dilation (expansion), opening, and closing are the simplest morphological operations. Set erosion and dilation are actually Minkowski set subtraction and addition [3,4], respectively. These filters were extended to multilevel signals in [2,5,6,7,8] by using mainly the correspondence between the shrinking/expanding of binary signals and the local min/max of multilevel signals [9]. Serra [2] used the *cross-sections* (thresholded versions) of the signal to generalize the morphological filtering of multilevel signals. Sternberg [6] further generalized morphological filters for multilevel signals by considering graytone images as surfaces of 3-D volumes (the *umbrae*). Ξ-filters [10] are related to morphological filters via Sternberg's approach. Lantuejoul and Serra [11] studied properties of generalized (algebraic) openings and closings, which they called *M-filters*.

The applications of morphological filters in image processing and analysis are numerous [2,12,13]. Areas of applications include biomedical image processing [2,6,8,14,15]; automated industrial inspection [16,17]; shape recognition [18]; nonlinear filtering [19,20]; edge detection [5,20,13]; noise suppression [5,21,20,13]; thinning [2,5,21,20,22]; enhancement [2,21]; representation and coding [20,22]; texture analysis [23]; and shape smoothing [2,20,22,24]. Currently, there are several commercialized image analyzers or other pipelined or parallel computer architectures [25]-[28] that use morphological filters (mainly for binary signals) among their main operations to extract pictorial information.

In this paper, first (in Section 2) we introduce a classification of signals and filters suitable for morphological filtering. Then we discuss the basic morphological concepts for representing signals by sets. In Section 3 we give the basic definitions and properties of the four simplest morphological filters. Although [2] is an excellent treatment of mathematical morphology, we feel that it is worthwhile to present some review material in Sections 2 and 3 for completeness and clarity, as well as for familiarizing the signal processing society with morphological signal analysis. Throughout Section 3 we attempt to clarify the transition and oscillation between morphological filtering of binary and multilevel signals. Some analysis and examples are provided to quantify the difference between morphologically convolving a signal with another binary or multilevel signal. In addition, some deterministic properties of morphological filters for multilevel signals are investigated concerning their commutability with thresholding, fixed points, and invertibility. Finally, in Section 4 we introduce the representation of linear shift-invariant filters in terms of morphological filters. That is, we define the *kernel* of a linear translation-invariant filter as a special collection of input signals and construct a *basis* of this kernel based on a signal ordering. The kernel or basis functions are then used to express a linear convolution as a supremum of subtractive morphological correlations (erosions).

This paper is the first in a sequence of two papers (Part 1 and 2). The results in Section 4 of this paper (Part 1) and the results in Part 2 [32] are all special cases of our earlier work in [19,20], which showed that a large class of nonlinear and linear translation-invariant filters can be represented exactly in terms of a minimal combination of morphological erosions or dilations. Both Parts 1 and 2 demonstrate the power of mathematical morphology as a unifying approach to both linear and nonlinear signal-shaping strategies.

2 REPRESENTATION OF SIGNALS BY SETS

An m-dimensional (m-D) signal can be represented mathematically by a function of m independent variables (an m-D function), where m is any positive integer. This function may assume only two distinct values, in which case we can represent the signal as a set in an m-D Euclidean space. For example, a binary image can be represented by a function $f(x)$ that assumes only two values, i.e., zero and one. Alternatively, the image *foreground* can be represented by the set $S = \{x : f(x) = 1\}$ and the image *background* by the set complement $S^c = \{x : f(x) = 0\}$. This function f is called the *characteristic function* of S.

Henceforth, both functions and sets will be used as mathematical representations of signals, with the distinction that an m-D **function** implies a multilevel m-D signal, whereas an m-D **set** refers to a binary m-D signal. Thus, a binary image will be represented by a 2-D set, whereas a graytone image by a 2-D function. This classification of signals induces a similar classification for filters into set-processing and function-processing filters. An m-D **set-processing (SP)** filter is a filter capable of accepting m-D binary signals as inputs and producing m-D binary signals as outputs. An m-D **function-processing (FP)** filter is any filter capable of accepting m-D functions as inputs and producing m-D functions as outputs. A subclass of m-D FP filters can produce an m-D binary signal whenever the input is also an m-D binary signal; these are called **function-and-set-processing (FSP)** filters. For example, on sampled graytone images a moving local minimum operation is a 2-D FSP filter, whereas a moving local averager is a 2-D FP (but not a FSP) filter.

In the above classification *set is the primary notion and function is just a particular case*, because the prototype morphological filters are defined through set operations. In addition, some of the FSP filters that this paper examines commute with thresholding of functions (explained later) and, hence, can be reduced to SP filters. In this light then, any FP or SP filter is viewed as a *set transformation* from one class of sets into another class of sets. The concept of a set, however, is more general than needed to represent signals. Therefore, we restrict ourselves to a class of sets that is just sufficiently general. Assuming [1] that every set representing a signal (viewed as an image object) contains its boundary, results in selecting the class $\mathcal{C}(\mathbf{E})$ of all *closed* subsets of a Euclidean space $\mathbf{E}$ as our general signal space. However, before any detailed discussion, it is necessary to introduce some notation.

Notation: The set of real numbers is denoted by $\mathbf{R}$, and the set of integers by $\mathbf{Z}$. Capital letters "$A, B, C, \ldots X, Y$" mainly denote sets; points of sets are denoted by lower case letters "$a, b, c, \ldots, x, y, z$". X^c denotes the set complement of X. The set of points x satisfying a property "P" is denoted by $\{x : \mathrm{P}\}$. If $X \subseteq \mathbf{R}$, then $\sup(X)$, $\inf(X)$, $\max(X)$, and $\min(X)$ denote, respectively, the supremum, infimum, maximum, and minimum of X. (See [29] for the differences between sup/inf and max/min, respectively.) Functions are denoted by "f, g, h". Capital Greek letters, e.g., "Ψ, Φ" denote SP filters, whereas lowercase Greek letters, e.g., "ψ, ϕ" denote FP filters. If X or f are input signals to filters Ψ or ψ, then $\Psi(X)$ and $\psi(f)$ denote output signals.

2.1 Cross-sections and Umbra of a Function

We assume that the domain of an m-D function $f(x)$ is a subset of the *domain space* $\mathbf{D} = \mathbf{Z}^m$ or $\mathbf{R}^m$, depending on whether the function is sampled or not, respectively. We also assume that the range of $f(x)$ is a subset of the *range space* $\mathbf{V}$=$\mathbf{R}$ or $\mathbf{Z}$, depending on whether the amplitude of $f(x)$ varies continuously or discretely. Our general Euclidean space $\mathbf{E}$ will be equal to the cartesian product $\mathbf{D} \times \mathbf{V}$. Thus, all binary m-D signals will be subsets of $\mathbf{D}$, whereas all multilevel m-D signals will be subsets of $\mathbf{E}$.

Since signals can be represented either by functions or by sets, and set is the primary notion, the main issue is to *represent functions by sets.* This is done by following two different but equivalent approaches. That is, an m-D function can be represented either by an ensemble of m-D sets called its *cross-sections* or by a single $(m+1)$-D set called its *umbra.*

Figure 1 shows a 1-D function f, one of its cross-sections, and its umbra. The set

$$X_t(f) = \{x \in \mathbf{D} : f(x) \geq t\} \quad , \quad t \in \mathbf{V}, \tag{1}$$

is called the **cross-section** of f at level t and is obtained by *thresholding* f at level t. By considering all different levels t we can associate f with a family of sets, which decrease monotonically as t increases. Since we work in the class of closed sets, all the cross-sections of f must be closed. The corresponding class of functions with which we will always deal is the class of *upper semicontinuous*[1] *(u.s.c.)* functions on $\mathbf{D}$, denoted as $USC(\mathbf{D})$. This correspondence is established because a real-valued function f defined on $\mathbf{R}^m$ is u.s.c. if and only if (iff) its cross-sections $X_t(f)$ are closed sets in $\mathbf{R}^m$ for all $t \in \mathbf{R}$ [31]. Qualitatively, we can think of u.s.c. (resp. l.s.c.) functions as resulting from continuous functions after the addition of some positive (resp. negative) jumps. A sampled function is trivially both u.s.c. and l.s.c., because all its cross-sections are subsets of $\mathbf{Z}^m$ and, hence, they are both closed and open.

[1]These are defined in [29,30] together with the *lower semincontinuous* functions *(l.s.c.)*

If we know all the cross-sections of an u.s.c. function, then we can uniquely reconstruct it through a supremum operation. This is illustrated in Fig. 2 which shows a 1-D function $f(x)$ and two of its cross-sections at levels t_1 and t_2. In this figure we observe that, for a given point $x \in \mathbf{R}$, $x \in X_{t_1}(f)$ iff $f(x) \geq t_1$. By contrast, we observe that $x \notin X_{t_2}(f)$ iff $f(x) < t_2$. Thus, the value of f at x is equal to the "largest" (supremum) of t's such that $f(x) \geq t$, or, equivalently, $x \in X_t(f)$. The above discussion is formalized by

THEOREM 1 (Serra [2]). *(a) Let $f(x)$ be an u.s.c. real-valued function on $\mathbf{R}^m$ and let $X_t(f)$, $t \in \mathbf{R}$, be its cross-sections. Then, the X_t's are closed sets in $\mathbf{R}^m$ that are decreasing, i.e.,*

$$t_1 < t_2 \Longrightarrow X_{t_1} \supseteq X_{t_2}, \tag{2}$$

obey a monotonic continuity

$$X_\tau = \bigcap_{t<\tau} X_t \ , \tag{3}$$

and for each $x \in \mathbf{R}^m$

$$f(x) = \sup\{t \in \mathbf{R} \ : \ x \in X_t\} \ . \tag{4}$$

(b) Conversely, a collection $\{X_t : t \in \mathbf{R}\}$, of closed sets satisfying conditions (2) and (3) generates an u.s.c. function $f(x)$ through (4). Then, the cross-sections of the resulting function $f(x)$ are identical with the initial sets X_t's for all t.

Note that if $\mathbf{V} = \mathbf{Z}$, then (3) becomes $X_\tau = \bigcap_{n \leq \tau} X_n$ with $\tau, n \in \mathbf{Z}$, and hence (3) is trivially satisfied; (4) becomes $f(x) = \sup\{t \in \mathbf{Z} : x \in X_t\}$.

Another way of establishing a link between sets and functions involves the concept of the umbra due to Sternberg [6,7]. As shown in Fig. 1, the **umbra** $U(f)$ of f is a subset of $\mathbf{E}$ and consists of all those points that occupy the space below the graph of f down to $-\infty$. We can also define the umbra of a set. For instance, Fig. 3a shows a closed set B in $\mathbf{R}^2$. Its umbra $U(B)$ is the closed set formed as follows: Let the points in $\mathbf{E}$ be parametrized by their projection x on $\mathbf{D}$ and their altitude t perpendicular to $\mathbf{D}$ (in Fig. 3 $\mathbf{D} = \mathbf{R}$). Suppose that the opaque set B is "illuminated" from above by a point source located at $t = +\infty$; then the shadow of B is its umbra $U(B)$ shown in Fig. 3a. Analytically, the umbra of the closed set B is equal to the Minkowski sum (defined later in (12)) of B and the negative axis $(-\infty, 0]$ of the amplitudes t's [2]. Similarly, the umbra of a function f is the Minkowski sum of the graph of the function and $(-\infty, 0]$, where the *graph* of f is the set $G(f) = \{(x,t) \in \mathbf{E} : f(x) = t\}$ shown in Fig. 3b. Analytically, the umbra of f is the subset

$$U(f) = \{(x,t) \in \mathbf{D} \times \mathbf{V} : f(x) \geq t\} \tag{5}$$

of $\mathbf{E}$. Obviously, the umbra is a set of higher dimensionality than the function. Figure 3b shows that a point (x,t) belongs to $U(f)$ iff $f(x) \geq t$. The second property of $U(f)$ is that if $(x,t) \in U(f)$, then $(x,a) \in U(f)$ for all $a < t$ (see also Fig. 3b), since the set $\{x\} \times (-\infty, f(x)]$ is a subset of $U(f)$. From $U(f)$ we can uniquely reconstruct f, because $f(x)$ is the "largest" altitude, i.e., the supremum of all t's such that $(x,t) \in U(f)$. The third property of $U(f)$ is that it is a closed set. The class of functions whose umbra is a closed set in $\mathbf{E}$ is the class of u.s.c. functions on $\mathbf{D}$; further, for each u.s.c. function there corresponds a unique umbra, and vice-versa. Next we formalize the above discussion with

THEOREM 2 . *(a) To any real-valued u.s.c. function $f(x)$, $x \in \mathbf{R}^m$, there corresponds a unique umbra $U(f)$. This umbra U is a closed set in $\mathbf{R}^{m+1}$ such that:*

$$(x,t) \in U \Longleftrightarrow t \leq f(x) \Longleftrightarrow x \in X_t(f) \, . \tag{6}$$

$$(x,t) \in U \Longrightarrow (x,a) \in U \quad , \quad \forall a < t \, . \tag{7}$$

Also for each $x \in \mathbf{R}^m$

$$f(x) = \sup\{t \in \mathbf{R} \; : \; (x,t) \in U\} \, . \tag{8}$$

(b) Conversely, to any closed subset U of $\mathbf{R}^{m+1}$ satisfying (7) there corresponds a unique u.s.c. function $f(x)$, which can be constructed from (8). The umbra then of $f(x)$ is equal to U.

A proof of Theorem 2 can be found in [20, p.54]. Theorem 2 also holds for functions with discrete argument or amplitude. By comparing (1) and (5) with (6), we see that $U(f)$ is the union of the sets $[X_t(f)] \times \{t\}$ for all $t \in \mathbf{V}$.

2.2 Isomorphic Operations between Sets and Functions

Since functions are exactly represented by their umbrae, union and intersection of umbrae must isomorphically induce two equivalent operations between the corresponding functions. Figure 4 shows two functions f and g and their respective umbrae $U(f)$ and $U(g)$. The *intersection* of the two umbrae is an umbra corresponding to a new u.s.c. function. This function, shown by the dashed curve in Fig. 4, is equal to the *pointwise minimum* of f and g. Likewise, the *union* of the two umbrae is the umbra of a new u.s.c. function, which is shown by the solid curve in the right part of Fig. 4, equal to the *pointwise maximum* of f and g. We denote these new functions as:

$$\begin{aligned}(f \wedge g)(x) &= \min\{f(x), g(x)\} \quad , \quad x \in \mathbf{D}. \\ (f \vee g)(x) &= \max\{f(x), g(x)\} \quad , \quad x \in \mathbf{D}.\end{aligned}$$

Finally, the set inclusion of umbrae, which is an ordering relation, induces an ordering relation between functions too. *Set inclusion* between the umbrae of two functions f and g corresponds to an **ordering of functions**. That is, we say that "f is less than g", denoted as "$f \leq g$", iff $f(x) \leq g(x)$ $\forall x \in \mathbf{D}$. Moreover,

$$f \leq g \Longleftrightarrow X_t(f) \subseteq X_t(g) \;\; \forall t \in \mathbf{V} \Longleftrightarrow U(f) \subseteq U(g) . \tag{9}$$

Table I shows the types of function operations induced by set operations on umbrae or cross-sections; e.g., $\wedge(\vee)$ between f and g is equivalent to $\cap(\cup)$ between $X_t(f)$ and $X_t(g)$ for all $t \in \mathbf{V}$.

2.3 Upper Semicontinuous (u.s.c.) Filters

A detailed formal definition of a general u.s.c. filter (viewed as set mapping) lies beyond the scope of this paper. Intuitively speaking, a filter is u.s.c. if it is continuous "from above". However, if the SP or FP filter is *increasing* (see Table II for definitions), then we can easily verify whether it is u.s.c. as follows. If (X_n) is a decreasing sequence of closed sets in $\mathbf{E}$, let $X_n \downarrow X$ denote the monotonic set convergence where $X_{n+1} \subseteq X_n$ for all n and $X = \bigcap_n X_n$. Then, if Ψ is an increasing SP filter in $\mathcal{C}(\mathbf{E})$, Ψ is u.s.c. iff $X_n \downarrow X \Longrightarrow \Psi(X_n) \downarrow \Psi(X)$

[1]. An obvious way to extend these concepts to FP filters is to consider the umbrae or cross-sections of functions. Leaving details aside, if (f_n) is a decreasing sequence of u.s.c. functions, we denote by $f_n \downarrow f$ the monotonic convergence toward the u.s.c. function f, where $f_{n+1} \leq f_n$ for all n and $f(x) = \inf_n\{f_n(x)\}$ for all $x \in \mathbf{D}$. Then, if ψ is an increasing FP filter in $USC(\mathbf{D})$, ψ is u.s.c. iff $f_n \downarrow f \Longrightarrow \psi(f_n) \downarrow \psi(f)$.

The u.s.c. condition is necessary whenever we require filters to be insensitive to the fine variations in both the amplitude and region of support of signals.

2.4 FSP Filters that Commute with Thresholding

Let ϕ be an m-D FSP filter and let Φ be the respective m-D SP filter of ϕ. Then ϕ is said to **commute with thresholding** iff, for any u.s.c. function f,

$$X_t[\phi(f)] = \Phi[X_t(f)] \quad , \quad \forall t \in \mathbf{V}. \tag{10}$$

Thus, for a general FSP filter ϕ satisfying (10), transforming the input *function* f and then thresholding $\phi(f)$ at any level t is equivalent to thresholding f at level t and then transforming by Φ the *set* $X_t(f)$. This allows us to interpret such a FP filter as a SP filter, which, especially for sampled signals, is simpler to analyze and implement. That is, from Theorem 1 and (10) we can synthesize the output function $\phi(f)$ from its SP-filtered cross-sections:

$$[\phi(f)](x) = \sup\{t \in \mathbf{V} : t \in \Phi[X_t(f)]\} \quad , \quad \forall x \in \mathbf{D}. \tag{11}$$

A necessary condition for a FSP filter to commute with thresholding is given by

THEOREM 3 . *Let $\phi : \mathcal{F} \longrightarrow USC(\mathbf{D})$ be a FSP filter defined on a subclass $\mathcal{F} \subseteq USC(\mathbf{D})$ of u.s.c. functions closed under pointwise infimum. If ϕ commutes with thresholding and Φ is its respective SP filter, then both ϕ and Φ are increasing and u.s.c.*

Proof. (a)-(Increasing): Let A, B be two closed sets in $\mathbf{D}$ such that $A \subseteq B$. Let h be an u.s.c. function defined as $h(x) = 2$ for $x \in A$, $h(x) = 1$ for $x \in B \cap A^c$ and $h(x) = 0$ for $x \in B^c$. Then $\Phi(A) = \Phi[X_2(h)] = X_2[\phi(h)] \subseteq X_1[\phi(h)] = \Phi[X_1(h)] = \Phi(B)$. Thus, Φ is increasing for sets. Now let $f \leq g \iff X_t(f) \subseteq X_t(g)\ \forall t \in \mathbf{V}$. Then

$X_t[\phi(f)] = \Phi[X_t(f)] \subseteq \Phi[X_t(g)] = X_t[\phi(g)]$ $\forall t \in \mathbf{V}$, which implies that $\phi(f) \leq \phi(g)$. Hence, ϕ is increasing for functions too.

(b)-(u.s.c): Let $\mathcal{S}$ be the set class of all cross-sections of functions in $\mathcal{F}$; then $\mathcal{S}$ is closed under intersection, because $\mathcal{F}$ is closed under infimum. Let (X_n) be any decreasing sequence of closed sets in $\mathcal{S}$; then $X_n \downarrow X \in \mathcal{S}$. By generalizing the method of proof in (a), we can set $X_n = X_{t_n}(f)$ for some u.s.c. function f where $t_{n+1} \geq t_n$ for all n, $t = \sup_n\{t_n\}$, and $X = X_t(f)$. From (2) and (3), since $\phi(f)$ is an u.s.c. function, we have $\Phi(X_n) = X_{t_n}[\phi(f)] \downarrow X_t[\phi(f)] = \Phi(X)$. Thus Φ is u.s.c. Likewise, let (f_n) be a decreasing sequence of u.s.c. functions such that $f_n \downarrow f \in \mathcal{F}$. Then $f_n \downarrow f \Longrightarrow X_t(f_n) \downarrow X_t(f) \Longrightarrow X_t[\phi(f_n)] = \Phi[X_t(f_n)] \downarrow \Phi[X_t(f)] = X_t[\phi(f)]$ $\forall t \in \mathbf{V} \Longrightarrow \phi(f_n) \geq \phi(f_{n+1})$ and $[\phi(f)](x) = \sup\{t : x \in X_t[\phi(f)] \Leftrightarrow [\phi(f_n)](x) \geq t \;\forall n\} = \inf_n\{[\phi(f_n)](x)\}$. Thus $\phi(f_n) \downarrow \phi(f)$ and hence ϕ is u.s.c. *Q.E.D.*

Theorem 3 suggests a straightforward way to construct a FSP filter that commutes with thresholding from an *increasing and u.s.c. SP filter* Φ. That is, if Φ operates individually on all the cross-sections of an u.s.c. function f, the family of filtered cross-sections $\{\Phi[X_t(f)] : t \in \mathbf{V}\}$ satisfies both conditions (2) and (3). Hence Theorem 1 guarantees that the function $g(x) = \sup\{t \in \mathbf{V} : x \in \Phi[X_t(f)]\}$ is an u.s.c. function and that $X_t(g) = \Phi[X_t(f)]$ for all $t \in \mathbf{V}$. Thus, by setting $g = \phi(f)$, we construct a FSP filter ϕ that commutes with thresholding and whose respective SP filter is Φ.

Theorem 3 becomes a necessary and sufficient condition by requiring ϕ to be *translation-invariant.* Let $A_z = \{a + z : a \in A\}$ denote the *translate* of a set A by the vector z. A SP filter Ψ is translation-invariant iff $\Psi(A_z) = [\Psi(A)]_z$, for all $z \in \mathbf{E}$ and all sets A in a set collection $\mathcal{S}$ closed under vector translation. If $A = U(f)$, translation of A by $z = (y, c)$ corresponds to a shift of the argument of $f(x)$ by $y \in \mathbf{D}$ and to a shift of the amplitude of f by $c \in \mathbf{V}$. Thus, we define herein a FP filter ψ as translation-invariant iff $\psi[f(x-y)+c] = [\psi(f)](x-y)+c$, for all $(y,c) \in \mathbf{D} \times \mathbf{V}$ and all functions $f(x)$ in a function class $\mathcal{F}$ closed under such translation. Then we have

THEOREM 4 . *A translation-invariant FSP filter $\phi : \mathcal{F} \longrightarrow USC(\mathbf{D})$, where $\mathcal{F}$ is a class of u.s.c. functions closed under translation and pointwise infimum, commutes with thresholding iff it is increasing and u.s.c.*

Proof. The necessity was proven in Theorem 3. *Sufficiency:* Throughout this proof we will make use of the SP, FSP, and FP erosions filters which are defined in Section 3. Let Φ be the respective m-D SP filter of the m-D filter ϕ. Φ is defined on the set class $\mathcal{S}$ of the cross-sections of the functions in $\mathcal{F}$. Since ϕ is a translation-invariant and increasing FP filter, it is equal to the supremum of FP erosions by a class $\mathcal{K}$ of functions [20]. These functions must be binary, i.e., m-D sets, because otherwise ϕ would not pass an m-D binary signal as binary. Thus, $\phi(f)(x) = \sup_{B\in\mathcal{K}} \{f \ominus B^s(x)\}$, where $f \ominus B^s(x) = \inf\{f(y) : y \in B_x\}$ is the FSP erosion of f by B and commutes with thresholding. Hence, if f is binary and thus equal to the characteristic function of a set Y, then the above expression for $\phi(f)$ becomes $\Phi(Y) = \bigcup_{B\in\mathcal{K}} Y \ominus B^s$, where $Y \ominus B^s = \{z : B_z \subseteq Y\}$ is the SP erosion of Y by B. However, Φ is a translation-invariant and increasing SP filter and, hence, equal to the union of SP erosions by all sets B in its *kernel* $\mathcal{K}(\Phi) = \{X \in \mathcal{S} : \mathbf{0} \in \Phi(X)\}$ [1]. Thus the class $\mathcal{K}$ in the above representation of ϕ is actually the kernel of its SP filter Φ, and hence

$$[\phi(f)](x) = \sup_{B\in\mathcal{K}(\Phi)} \{f \ominus B^s(x)\}, \quad \forall x \in \mathbf{D},\ f \in \mathcal{F}.$$

In addition, $\forall t \in \mathbf{V}$,

$$\Phi[X_t(f)] = \bigcup_{B\in\mathcal{K}(\Phi)} [X_t(f)] \ominus B^s = \bigcup_{B\in\mathcal{K}(\Phi)} X_t(f \ominus B^s).$$

Since Φ is increasing and u.s.c., the set class $\{\Phi[X_t(f)] \ : \ t \in \mathbf{V}\}$ creates an u.s.c. function $g(x) = \sup\{t \in \mathbf{V} : x \in \Phi[X_t(f)]\} = \sup\{t : x \in X_t(f \ominus B^s) \iff f \ominus B^s(x) \geq t$, for some $B \in \mathcal{K}(\Phi)\} = \sup_{B\in\mathcal{K}(\Phi)}\{f \ominus B^s(x)\}$. Hence $g = \phi(f)$. The cross-sections of g are (Theorem 1) $X_t(g) = \Phi[X_t(f)]$, and thus ϕ commutes with thresholding. *Q.E.D.*

Not all FSP filters commute with thresholding. Examples of FSP filters that commute with thresholding include median and order-statistic filters [32] as well as erosion, dilation, opening, and closing of functions by sets, which are defined in Section 3.3.

3 MORPHOLOGICAL FILTERS

In morphological filtering each signal is viewed as a set, and its geometrical features are modified by morphologically convolving the signal with a **structuring element**, which is another set of simple shape and size [2]. By varying the structuring element we can extract different types of information from the signal. According to the four quantification principles of mathematical morphology [2], each morphological filter (viewed as a set mapping) must be: 1) translation-invariant, 2) scale-invariant (in $\mathbf{R}^m$), 3) dependent only on local knowledge of the signal, and 4) u.s.c. A sufficient condition for a morphological convolution to be u.s.c. is to use *compact* sets or functions with a compact region of support as structuring elements.

In a morphological convolution the signal and the structuring element could be either sets or functions. Thus, in this section we present the similarities and differences between the basic SP and FP morphological filters and examine some of their properties.

3.1 Morphological Filters of Sets by Sets

Morphological filters of sets by sets are SP filters processing input m-D sets by interacting them via Minkowski set addition or subtraction with structuring elements that are compact n-D sets ($n \leq m$). The **Minkowski set addition** [3,4] of the sets $A, B \subseteq \mathbf{R}^m$ is the set

$$A \oplus B = \{a + b \ : \ a \in A, b \in B\} = \bigcup_{b \in B} A_b \ , \tag{12}$$

where $A_b = \{a + b \ : \ a \in A\}$. The **Minskowski set subtraction** [4] of B from A is the set

$$A \ominus B = (A^c \oplus B)^c = \bigcap_{b \in B} A_b \ . \tag{13}$$

Let $B^s = \{-b : b \in B\}$ denote the *symmetric* set of B with respect to the origin, and $\emptyset$ denotes the empty set. The basic SP morphological filters are the **erosion** $X \ominus B^s$, **dilation** $X \oplus B^s$, **opening** X_B, and **closing** X^B of X by B, defined in [1] as:

$$X \ominus B^s = \{z : B_z \subseteq X\} = \bigcap_{b \in B} X_{-b} \tag{14}$$

$$X \oplus B^s = \{z : B_z \cap X \neq \emptyset\} = \bigcup_{b \in B} X_{-b} \tag{15}$$

$$X_B = (X \ominus B^s) \oplus B \tag{16}$$

$$X^B = (X \oplus B^s) \ominus B \ . \tag{17}$$

From (14)-(17) and Fig. 5 we observe the following: Geometrically, the erosion of X and B is defined as the set of all points z such that the translate B_z is contained in the original set X; the dilation of X by B is defined as the set of all points z such that B_z intersects X. Algebraically, the erosion of X by B is equal to the Minkowski set subtraction of B^s from X; the dilation of X by B is the Minkowski sum of X and B. Dilating X is equivalent to eroding X^c and complementing the result as implied by (13). The opening of X by B is the set resulting from erosion of X by B followed by Minkowski sum with B; this cascade does not generally recover X, but rather a subset of X which is the morphologically most essential part with respect to B. From (12), (14), and (16) it follows that

$$X_B = \bigcup_{B_z \subseteq X} B_z \,. \tag{18}$$

Similarly, the closing of X by B results from dilating X by B and then Minkowski subtracting B from the result; in general, the closing of X is a set containing X. Closing X is equivalent to opening X^c and complementing the result.

To visualize geometrically these morphological filters we assume that we deal with 2-D sets, which may represent binary images. Thus, let the closed set X represent a binary image and the compact set B a structuring element, such as the island and the disk, respectively, of Fig. 5. Then, Fig. 5 shows that erosion shrinks the set X, whereas dilation expands X. The opening suppresses the sharp capes and cuts the narrow isthmuses of X, whereas the closing fills in the thin gulfs and small holes. Thus, if the structuring element B has a regular shape, both opening and closing can be thought of as nonlinear filters which smooth the contours of the input signal.

Parallel to the evolution of all these morphological operations in [1,2], since the 1960's there have been many other researchers who have been using similar operations of the shrink/expand type (or cascades of shrink/expand) for digital (binary) image processing and with cellular array computers designed for image analysis. Surveys of these approaches can be found in [21,33]. In addition to its original contributions, mathematical morphology formalized all the above collections of image operations.

Figure 6 shows examples of dilations and erosions of discrete sets. It also illustrates the difference between Minkowski addition/subtraction and dilation/erosion, respectively. Of course, there is no difference between these two pairs of operations if B is symmetric. Discrete erosions and dilations accept simple parallel implementations. From (14) and (15), the erosion or dilation of X by B is equal to the set *intersection* or *union*, respectively, of all *translates* X_b of X, where the vector b sweeps B^s. Using this property, Sternberg [34] suggested parallel computer architectures to implement the SP erosion and dilation filters, which actually involve parallel logical AND's and OR's between binary image planes. We discussed the computational complexity of these implementations in [20,22]. In [32] we also give an alternative non-parallel implementation for the SP erosion and dilation using linear convolution concepts.

3.2 Umbra Interpretation of FP Morphological Filters

The four morphological SP filters can also be used for functions by viewing the morphological FP filters as special cases of SP filters that process the umbrae or cross-sections of the input functions. The most general case of a morphological transformation of a function f is the transformation of $U(f)$ by a structuring element B that is a compact subset of $\mathbf{E}$. For example, if f is a 1-D function, B could be a 2-D disk, such as the one shown in Fig. 7a. In general, B is a set of the same or lower dimensionality than that of $U(f)$, but not necessarily a function. Thus, *Minkowski addition* $f \oplus B$ and *subtraction* $f \ominus B$ of the u.s.c. m-D function f with a compact $(m+1)$-D set B can be defined [2] by finding their respective umbrae:

$$U(f \oplus B) = U(f) \oplus B = U(f) \oplus U(B) \tag{19}$$

$$U(f \ominus B) = U(f) \ominus B^r = U(f) \ominus [U(B)]^r \tag{20}$$

where $B^r = \{(x, -t) \in \mathbf{D} \times \mathbf{V} \; : \; (x,t) \in B\}$ is the *reflected* set of B with respect to $\mathbf{D}$. Thus transforming $U(f)$ by B is equivalent to transforming $U(f)$ by $U(B)$.

Sternberg [6,7] investigated the morphological filtering of 2-D graytone images by isotropic 3-D compact sets such as spheres, cones, paraboloids, and cylinders. In this paper, however, we focus on only two special cases for B. First, B becomes the graph of an m-D u.s.c. function g with compact region of support, as shown for example in Fig. 7b, and, hence, we transform $U(f)$ by the umbra of g. We shall call this case a *morphological filter of a function f by a (structuring) function g*. Second, B becomes an m-D compact set S lying at $t = 0$, as Fig. 7c shows. In this case, transforming $U(f)$ by S is equivalent to transforming $U(f)$ by the umbra of S, which is a half-infinite solid of top S formed by the Minkowski sum of S and the negative t-axis. We henceforth call this case a *morphological filter of a function f by a (structuring) set S.* Transforming f by a set $S \subseteq \mathbf{D}$ is a special case of transforming f by a function g. That is, if g is flat (binary), then it can be represented by a set S.

Concluding, (19) and (20) *geometrically* interpret the erosion/dilation of a function f by a function g or by a set S as the erosion/dilation of the umbra of f by the umbrae of g or S, respectively. In what follows, we will *analytically* define the morphological filters of a function by a structuring function or set.

3.3 Morphological Filters of Functions by Sets

These are FSP filters where the input signals are m-D u.s.c. functions and the structuring elements are *compact* n-D sets with $n \leq m$. Their analysis is easy, since they commute with thresholding because of their construction. For example, if Φ is the SP erosion filter by B, then [2] defines a FSP erosion by B as follows. Since Φ is increasing and u.s.c., for any input function f, the set class $\{\Phi[X_t(f)] = X_t(f) \ominus B^s : t \in \mathbf{V}\}$ satisfies (2) and (3) and, hence, through (4) it creates an output function h by setting $X_t(h) = \Phi[X_t(f)]$. This output function is the *erosion of f by B*, denoted by $f \ominus B^s$. Likewise, we can define the *dilation* $f \oplus B^s$, the *opening* f_B, and the *closing* f^B of an u.s.c. function f by a compact set B. Below we give their analytical definitions, which show how these filters operate both on the function f as a whole and on each one of its cross-sections:

CROSS-SECTIONS (all t) FUNCTIONS

$$X_t(f \ominus B^s) = X_t(f) \ominus B^s \iff (f \ominus B^s)(x) = \inf_{y \in B_x} \{f(y)\} \tag{21}$$

$$X_t(f \oplus B^s) = X_t(f) \oplus B^s \iff (f \oplus B^s)(x) = \sup_{y \in B_x} \{f(y)\} \tag{22}$$

$$X_t(f_B) = [X_t(f)]_B \iff (f_B)(x) = [(f \ominus B^s) \oplus B](x) \tag{23}$$

$$X_t(f^B) = [X_t(f)]^B \iff (f^B)(x) = [(f \oplus B^s) \ominus B](x) \tag{24}$$

Thus, the erosion (resp. dilation) of f by B at any point x is obtained by shifting the set B to location x and taking the infimum (resp. supremum) of f inside this shifted set. This structuring *set* B plays the same role as a moving window.

For *sampled* signals, the set B is discrete; B must also be compact. Hence, in the discrete case, B is a *finite* set (window). Therefore, *erosion* (resp. *dilation*) of a sampled m-D function by a finite set $B \subseteq \mathbf{Z}^m$ is equal to the *moving local minimum* (resp. *maximum*) of the function inside the window B. The erosion or dilation of the characteristic function of a set shrinks or expands respectively the set. The correspondence between shrink/expand of binary images and local min/max of graytone images, as well as the commuting of min/max with thresholding, was proven in [9] and later used in [5].

Figure 8 shows some examples of morphologically filtering a 1-D sampled function f of 80 samples by a small convex set $B = \{-2, -1, 0, 1, 2\}$. We see that erosion of a function by such a set reduces the peaks and enlarges the minima of the function. The dilation of f by B increases the valleys and enlarges the maxima of the function. Figure 8 shows that the opening by B smooths the graph of f from below by cutting down its peaks. The closing smooths the graph of f from above by filling up its valleys. Subtracting from f its opening or closing by B gives the peaks and valleys, respectively, of f. The width of these peaks and valleys depends on the size of B. Thus, opening and closing of functions by sets can be used for detection of peaks and valleys in signals. In addition, opening and closing by a convex set B can be used effectively to suppress *impulse noise* in signals [2,5,9,14,21,20]. By impulse noise is meant that a signal is corrupted by impulses (spikes), i.e., very large positive or negative values of short duration. Opening or closing by B can eliminate, respectively, such positive or negative noise impulses, if the impulse width does not exceed the size of B.

3.4 Morphological Filters of Functions by Functions

These are FP filters where the input signal is an m-D u.s.c. function f and the structuring element is a n-D u.s.c. function g ($n \leq m$) with a compact region of support. A cross-section interpretation of such filters can be found in [2]. Here, we limit our discussion only to an umbra interpretation [2]. That is, if we replace B with $U(g)$ in (19), the Minkowski sum of $U(f)$ and $U(g)$ gives the umbra of the Minkowski function addition $f \oplus g$ of f and g. (See [35] for a proof.) The Minkowski function subtraction $f \ominus g$ of g from f is defined similarly. Thus from (8), (19), and (20) we have, for each $x \in \mathbf{D}$,

$$(f \oplus g)(x) = \sup_{y \in \mathbf{D}} \{f(y) + g(x-y)\} \tag{25}$$

$$(f \ominus g)(x) = \inf_{y \in \mathbf{D}} \{f(y) - g(x-y)\} \tag{26}$$

The u.s.c. functions may be real-valued only on a subset of $\mathbf{D}$. However, for each $x \in \mathbf{D}$ $f(x) = \sup[A(x)]$, where $A(x) = \{t \in \mathbf{V} \ : \ x \in X_t(f)\}$. If $f(x_0) \notin \mathbf{V}$ for some $x_0 \in \mathbf{D}$, then we can still define $f(x_0) = \sup[A(x_0) = \emptyset] = -\infty$. If $A(X_0) = \mathbf{V}$, we define $f(x_0) = \sup(\mathbf{V}) = +\infty$. Thus we can assume that f and g are defined over all $\mathbf{D}$. We call the **region of support** of such a function f, denoted by $Ros(f)$, the subset of $\mathbf{D}$ on which $f(x) \neq -\infty$. For the morphological filtering of f by g to be an u.s.c. filter, it is sufficient that $Ros(g)$ be a *compact* (*finite*, for sampled signals) subset of $\mathbf{D}$.

The relation between Minkowski subtraction/addition and erosion/dilation, respectively, of functions is the same as for sets. That is, if $g^s(x) = g(-x)$ denotes the *symmetric* function of g with respect to the origin, Serra [2] defines:

$$\text{Erosion of } f \text{ by } g \ : \ (f \ominus g^s)(x) = f(x) \ominus g(-x) \tag{27}$$

$$\text{Dilation of } f \text{ by } g \ : \ (f \oplus g^s)(x) = f(x) \oplus g(-x) \tag{28}$$

$$\text{Opening of } f \text{ by } g \ : \ f_g(x) = [(f \ominus g^s) \oplus g](x) \tag{29}$$

$$\text{Closing of } f \text{ by } g \ : \ f^g(x) = [(f \oplus g^s) \ominus g](x) \ . \tag{30}$$

Thus the Minkowski sum of f and g is defined through an additive convolution between f and g and then taking a supremum, as opposed to summation or integration for linear convolutions. Likewise, the Minkowski subtraction of g from f is defined through a subtractive convolution between f and g and then taking an infimum. We shall call the Minkowski addition and subtraction of two functions **morphological convolutions**. From (25) and (28) we see that the relation between Minkowski function addition and dilation is the same as between *convolution* and *correlation* of two functions, respectively; likewise for Minkowski function subtraction and erosion. Thus, we shall call the dilation and erosion of two functions **morphological correlations**. Of course, if $g(x) = g(-x)$ is even, the above difference disappears, because then $f \oplus g = f \oplus g^s$ and $f \ominus g = f \ominus g^s$.

The morphological filtering of a function f by a set B is a simple case of filtering f by a function g. That is, if $g(x) = 0$ for all $x \in Ros(g)$, then (27)-(30) reduce respectively to (21)-(24) by setting $B = Ros(g)$. The difference between a binary and a multilevel structuring function g is also depicted in Fig. 9. Consider a cosine function $f(x) = \cos \omega x$, $x \in \mathbf{R}$, and a structuring function

$$g(x) = \begin{cases} A[\cos(\omega x) - \cos(\omega L)] & , \ \| x \| \leq L \\ -\infty & , \ \| x \| > L \end{cases}$$

where $\| \cdot \|$ denotes absolute value, $0 \leq A \leq 1$, and $0 \leq L \leq \pi/(2\omega)$. Thus $g(x)$ is a scaled top of the cosine. If $A = 0$, g is binary; if $A > 0$, g is multilevel. Figure 9a shows the opening f_g for three different structuring functions corresponding to three different values of $A = 0, 0.5, 1$. If $A = 0$, the FP opening f_g reduces to the FSP opening f_B, where $B = Ros(g) = [-L, L]$. This opening cuts down the peaks of the cosine. We can symmetrize this by considering the **open-closing** (opening followed by closing by the same structuring function) $(f_g)^g$ shown in Fig. 9b. Similar results can be observed if we consider the **clos-opening** (closing followed by opening by the same structuring function) $(f^g)_g$. (The open-closing and clos-opening are new morphological filters introduced in [20] and used for image noise suppression and for providing fixed points of median filters; see also [32].) For $A = 0$ the FSP open-closing cuts down the peaks and fills up the valleys of f. This clipping effect of the open-closing is very similar to the behavior of the analog median filter [36]. Figures 9a,b ($A \neq 1$) suggest that opening and open-closing have both "low-pass" and "high-pass" filtering characteristics. Indeed both attenuate the basic frequency ω of f and introduce higher harmonics; the opening also introduces a dc-offset [20]. However, the frequency analysis of the effects of opening or closing is of very limited importance, because these effects are dependent on the input signal, and they cannot be generalized since linear superposition does not apply. Of general importance, though, is the following geometric interpretation that we give for the opening.

For any functions f and g, the umbra of the opening f_g is equal to the union of all the translates of $U(g)$ that can "fit" inside $U(f)$; i.e., for $z = (y, c) \in \mathbf{D} \times \mathbf{V}$,

$$U(f_g) = \bigcup_{[U(g)]_z \subseteq U(f)} [U(g)]_z \Longleftrightarrow f_g(x) = \sup_{[g(x-y)+c] \leq f(x)} \{g(x-y) + c\} \tag{31}$$

A similar geometrical interpretation can be given for the closing by g, since $(f^g) = -(-f_g)$. For example, in Fig. 9 the closer A is to one, the more g resembles the cosine peak, the closer the fitting of g under the peaks or above the valleys of f, and hence the closer f_g or $(f_g)^g$ is toward f. In the limit when $A = 1$, g becomes equal to the cut-off cosine peak, and both f_g and $(f_g)^g$ are equal to f for all $L \leq \pi/(2\omega)$.

Next we provide some examples of *discrete* FP morphological filtering. Figure 10 shows a 1-D original sampled function representing 250 samples of a graytone image intensity profile (after the addition of a dc-offset); its median filtered version by an 11-point window; and its filtered versions by six morphological filters (erosion, dilation, opening, closing, open-closing, and clos-opening), each operated by three differently shaped discrete structuring elements with a convex 11-point region of support. The first structuring element is a set B, and the other two are functions g, h:

$$\begin{aligned} B &= \{n \in \mathbf{Z} : -5 \leq n \leq 5\} \\ g(n) &= G \cdot \sqrt{25 - n^2} \ , \ \ -5 \leq n \leq 5 \\ h(n) &= H \cdot (25 - n^2) \ , \ \ -5 \leq n \leq 5 \ , \end{aligned}$$

and $g(n) = h(n) = -\infty$ if $\| n \| > 5$. In Fig. 10 the scaling factors of g and h were $G = 5$ and $H = 2$. The umbra $U(B)$ is a semi-infinite rectangle of top B, and (if $G = H = 1$) $U(g)$ and $U(h)$ are the umbrae of a disk and a parabola, respectively. As explained in Section 3.2, morphological filtering of f by B, g, h results in a function whose umbra is identical to the umbra resulting from morphological filtering of $U(f)$ by $U(B), U(g), U(h)$. Therefore, in Fig. 10 we called these three cases filtering by "rectangle, disk, parabola", respectively. The following qualitative observations are evident from Fig. 10: 1) The erosion and dilation produce, respectively, a smaller (with respect to function $\leq$) and larger function. 2) The opening or closing by a rectangle produce signals whose peaks or valleys, respectively, consist of flat plateaus not smaller than the size of the rectangle top; this flatness increases with the size of the rectangle top. They also preserve the vertical boundaries of the signal they transform and commute with thresholding. 3) Morphological filtering by g or h does not commute with thresholding and tends to shape the original signal f similarly to the shape of g or h. Specifically, opening or closing by g and h tends to penetrate inside the peaks or valleys of f. The amount and shape of this penetration depends on the amplitude range, support width, and shape of g and h. Thus the parabola h penetrates more than the disk g since it has a narrower and sharper shape. 4) Both the open-closing and clos-opening offer a balanced mixing of the properties of opening and closing, and they transform f very similarly to the median (see also [20,32]). Finally, the original function in Fig. 10 was selected 1-D for an easier visualization of the geometrical effects of morphological filtering, but the above observations are general and apply to signals of higher dimensionality too.

The principal point of the above discussion and examples is that, for an intuitive understanding of morphological filtering, *any structuring function g should be seen first as a geometrical pattern.* For instance, (31) implies that the opening of a signal by g is the geometrical (inclusive) content of the signal in this pattern g. Thus, if the signal is not smooth and g is smooth, then the opening by g will be smoother than the original signal. However, if the signal is smooth and g is less smooth, then the opening by g may be less smooth than the original signal.

3.5 Properties of Morphological Filters

Some properties of arbitrary SP and FP filters are defined in Table II. Referring to these definitions, both SP and FP erosions, dilations, and all their cascades (e.g., opening, closing, open-closing, and clos-opening) or parallel combinations (using $\cap(\cup)$ for sets or pointwise $\wedge$ $(\vee)$ for functions) are *translation-invariant* and *increasing* filters. Further, if $\mathbf{0}$ denotes the origin of $\mathbf{D}$, we have:

Property 1. *Ordering.* The opening is always *anti-extensive* filter $(f_g \leq f)$, whereas the closing is always *extensive* $(f^g \leq f)$. If $g(\mathbf{0}) \geq 0$, then the erosion by g is anti-extensive, whereas the dilation by g is an extensive filter. Thus, if $g(\mathbf{0}) \geq 0$,

$$f \ominus g \leq f_g \leq f \leq f^g \leq f \oplus g \,. \tag{32}$$

Proof. If $g(\mathbf{0}) \geq 0$, $\forall x \in \mathbf{D}$, $f \ominus g(x) = \inf\{f(z) - g(x-z) \ : \ z \in \mathbf{D}\} \leq f(x) - g(\mathbf{0}) \leq f(x) \Longrightarrow f \ominus g \leq f$. Likewise, $f \oplus g \geq f$.

From (27),(28),(29), and $\forall g$, $f_g(x) = \sup_z\{\inf_y\{f(y) - g(y-z)\} + g(x-z)\} \Longrightarrow$

$$f_g(x) \leq \inf_y\{\sup_z\{f(y) - g(y-z) + g(x-z)\}\}.$$

(See also [29, p.43].) Selecting $y = x$ in the above inequality yields: $f_g(x) \leq \sup_z\{f(x)\} = f(x) \Longrightarrow f_g = f$. Likewise, $f^g \geq f$ can be proved. *Q.E.D.*

Property 1 also holds for SP filters. That is, $X_B \subseteq X \subseteq X^B$ for any sets X, B; moreover, if $\mathbf{0} \in B$, $X \ominus B \subseteq X_B \subseteq X \subseteq X^B \subseteq X \oplus B$. The validity of Property 1 can be observed in Fig. 5 for SP filters and in Fig. 10 for FP filters.

The SP and FP opening and closing are *idempotent*; i.e., $(f_g)_g = f_g$ and $(f^g)^g = f^g$. Similarly we have:

Property 2. The SP and FP open-closing and clos-opening filters are *idempotent*.

Proof. SP open-closing: Let $Y = (X_B)^B$. Then $Y = Y^B \supseteq (Y_B)^B = [((X_B)^B)_B]^B \supseteq [(X_B)_B]^B = Y \Longrightarrow Y = (Y_B)^B$. Similarly we prove that the SP clos-opening $(X^B)_B$ is idempotent. By replacing X with a function f, B with a structuring set or function, and set $\subseteq$ with function $\leq$, the property is proved for functions too. *Q.E.D.*

Equations (14),(15) and (16),(17) imply that there is a *duality* with respect to set complementation between set erosion and dilation as well as between set opening and closing. Likewise, there is a duality with respect to function negation between the corresponding FP filters. That is, $(-f) \ominus g = -(f \oplus g)$ and $(-f)_g = -(f^g)$, where $(-f)(x) = -f(x)$ for all x. More details about properties of SP filters can be found in [2,22]. Henceforth, we will focus only on some properties of FP morphological filters.

Dilation of f by g is both commutative and associative. Erosion is neither commutative nor associative. Moreover, we have:

Property 3. *Distributivity.* $(f \vee g) \oplus h = (f \oplus h) \vee (g \oplus h)$ and $(f \wedge g) \ominus h = (f \ominus h) \wedge (g \ominus h)$.

Property 4. *Parallel Composition.* $f \ominus (g \vee h) = (f \ominus g) \wedge (f \ominus h)$.

Property 5. *Serial Composition.* $(f \oplus g) \oplus h = f \oplus (g \oplus h)$ and $(f \ominus g) \ominus h = f \ominus (g \oplus h)$.

The proofs of Properties 3,4,5 result from interchanging sup/inf with max/min, respectively, and other similar properties of sup/inf. Properties 3,4,5 also hold for FSP erosions and dilations of functions by sets if we replace the $g \vee h$ or $g \wedge h$ of two functions by the union $A \cup B$ or intersection $A \cap B$, respectively, of two sets.

Property 5 makes erosion and dilation inherently *separable* filters. For example, the erosion of a 2-D set or function by the discrete 3×3-pixel set B of Fig. 6e, can also be obtained by eroding first by B_1 (the 3-pixel horizontal segment) and then eroding the result by B_2 (the 3-pixel vertical segment), because $B = B_1 \oplus B_2$. Thus, this 2-D erosion becomes a cascade of 1-D erosions.

Difference between structuring function and set. The difference in geometrical effects between opening (or closing) of a function by a binary and multilevel structuring function can be seen in Figs. 9,10. Next we quantify analytically this difference.

Property 6. If g is a bounded real-valued structuring function with a compact region of support $B = \mathrm{Ros}(g)$, then, for any function f, $\forall x \in \mathbf{D}$,

$$\| f_g(x) - f_B(x) \| \leq \sup_{z \in B}\{g(z)\} - \inf_{z \in B}\{g(z)\} \,. \tag{33}$$

Proof. Both $s = \sup\{g(z) : z \in B\}$ and $r = \inf\{g(z) : z \in B\}$ are are real numbers because g is bounded. Then, $f_g(x) = \sup\{f \ominus g^s(z) + g(x-z) : z \in \mathbf{D}\} \Longrightarrow f_g(x) \leq s + \sup\{f \ominus g^s(z) : z \in (B^s)_x\}$. Also, $f \ominus g^s(z) = \inf\{f(y) - g(y-z) : y \in B_z\} \leq f \ominus B^s(z) - r$. Thus, $f_g(x) \leq \sup\{f \ominus B^s(z) : z \in (B^s)_x\} + s - r = f_B(x) + s - r$. Likewise, we can prove that $f_g(x) \geq f_B(x) - s + r$, and the proof is complete. *Q.E.D.*

The result (33) applies also to the closing of f by g since $f^g = -(-f)_g$. For *sampled* signals, the sup/inf of g over the compact $Ros(g) = B$ become the max/min of g over the finite $Ros(g)$. For example, referring to the original function f and the structuring elements B, g, h of Fig. 10, $\| f_g(x) - f_B(x) \| \leq 25$ and $\| f_h(x) - f_B(x) \| \leq 50$ for each x.

Fixed points of opening and closing by a structuring function. In [1,11] it was shown that the opening and closing filters can be completely specified from their fixed points, i.e., signals invariant to these filters. The interest in the fixed points of the openings and closings also arises from their direct relations to the fixed points of median filters [20,32]. The following theorem classifies the fixed points of the FP opening and closing filters. (A proof of a similar theorem for SP openings/closings can be found in [2].)

THEOREM 5 . *A function f is a fixed point of the opening (resp. closing) filter by a function g iff $f = h \oplus g$ (resp. $f = h \ominus g$), where h is an arbitrary function. Likewise, for any set B, $f = f_B \Longleftrightarrow f = h \oplus B$ and $f = f^B \Longleftrightarrow f = h \ominus B$, for some function h.*

Proof. If $f = f_g$, then we simply set $h = f \ominus g^s$. Conversely, if $f = h \oplus g$ for some function h, from Property 1 and since dilation is increasing filter,

$$h \oplus g \geq (h \oplus g)_g = [(h \oplus g) \ominus g^s] \oplus g = h^{g^s} \oplus g \geq h \oplus g \, .$$

Thus, $f_g = (h \oplus g)_g = h \oplus g = f$. Likewise for the closing, if $f = f^g$, we set $h = f \oplus g^s$. Conversely, if $f = h \ominus g$,

$$h \ominus g \leq (h \ominus g)^g = [(h \ominus g) \oplus g^s] \ominus g = h_{g^s} \ominus g \leq h \ominus g \, .$$

Thus $f^g = (h \ominus g)^g = h \ominus g = f$. Nothing also changes in the proof if we replace g with B (and hence g^s with B^s). *Q.E.D.*

Invertibility of erosion and dilation. In general, erosion and dilation are noninvertible operations. However, from the erosion of f by g we can recover the original f iff $f = f_g$. Specifically, since $(f \ominus g^s) \oplus g = f_g$, the Minkowski sum of the erosion $f \ominus g^s$ and g recovers f iff $f = f_g$. Likewise, if we have the dilation $f \oplus g^s$ and we Minkowski subtract g from it, we obtain $(f \oplus g^s) \ominus g = f^g$. Thus, if $f = f^g$, Minkowski subtraction of g from the dilation $f \oplus g^s$ recovers the original f. However, in the general case, where $f_g \neq f \neq f^g$, we have:

Property 7. If $g(0) \geq 0$, then

$$f_g \leq h \leq f \implies h \ominus g^s = f \ominus g^s \implies h_g = f_g \, , \tag{34}$$

$$f \leq h \leq f^g \implies h \oplus g^s = f \oplus g^s \implies h^g = f^g \, . \tag{35}$$

Proof. For any f and g, because of Theorem 5,

$$f_g \ominus g^s = [(f \ominus g^s) \oplus g] \ominus g^s = (f \ominus g^s)^{g^s} = f \ominus g^s \, .$$

Since $g(0) \geq 0$, the erosion by g is increasing filter. Hence, $f_g \leq h \leq f \implies f_g \ominus g^s \leq h \ominus g^s \leq f \ominus g^s \implies h \ominus g^s = f \ominus g^s \implies h_g = f_g$. Similarly,

$$f^g \oplus g^s = [(f \oplus g^s) \ominus g] \oplus g^s = (f \oplus g^s)_{g^s} = f \oplus g^s \, .$$

Since $g(0) \geq 0$, the dilation by g is increasing filter. Hence, $f \leq h \leq f^g \implies f \oplus g^s \leq h \oplus g^s \leq f^g \oplus g^s \implies h \oplus g^s = f \oplus g^s \implies h^g = f^g$. *Q.E.D.*

Thus, in the general case when $f \neq f_g$, from the erosion $f \ominus g^s$ we cannot recover f, because there is not a unique function whose erosion by g is equal to $f \ominus g^s$. Furthermore, the Minkowski sum of $f \ominus g^s$ and g will recover only the function f_g, but none of the functions h between f and f_g. Likewise, if $f \neq f^g$, the dilation of f by g cannot be inverted.

4 RELATIONS BETWEEN LINEAR AND MORPHOLOGICAL FILTERS

A *linear shift-invariant (LSI)* filter is viewed in our analysis as a FP filter that commutes only with a shift with respect to the ***argument*** of its input functions. A *linear translation-invariant (LTI)* filter (see Table II) is a LSI filter that passes constant signals unchanged, viz., whose dc-gain is equal to one. Now suppose that a LTI filter ψ is defined on a class $\mathcal{F}$ of real-valued u.s.c. functions that is closed under translation. This class $\mathcal{F}$ could be, for instance, the class of continuous functions, or the class of bounded functions, or the class of functions with a compact region of support. Let $h(x)$, $x \in \mathbf{D}$, denote the *impulse response* of ψ. Then, $\psi(f) = h * f$ for any $f \in \mathcal{F}$, where "$*$" denotes linear convolution, discrete or continuous. The **kernel** of ψ is defined as the following set of input functions:

$$\mathcal{K}(\psi) = \{g \in \mathcal{F} \ : \ h * g(\mathbf{0}) \geq 0\} . \tag{36}$$

The above kernel uniquely characterizes the LTI filter ψ and can reconstruct it, as explained in Maragos [20, chap.5], because $\forall z \in \mathbf{D}$

$$[\psi(f)](z) = \sup\{t \in \mathbf{R} : [f(x+z) - t] \in \mathcal{K}(\psi)\} . \tag{37}$$

In addition, if ψ is also *increasing*, then it can be represented exactly as the supremum of erosions by all its kernel functions. The following theorem provides a necessary and sufficient condition for a LSI filter to be increasing.

THEOREM 6 . *A linear shift-invariant filter is increasing iff its impulse response is nonnegative everywhere.*

Proof. Let $h(x)$, $x \in \mathbf{D}$, be the impulse response of the LSI filter. *Sufficiency*: Let $h(x) \geq 0\ \forall x$. If $f \leq g$, then $p(x) = g(x) - f(x) \geq 0\ \forall x \Longrightarrow h * p(x) \geq 0\ \forall x \Longrightarrow h * g \geq h * f$.

Necessity: Let $h * f \leq h * g$ for any $f \leq g$. Then, $h * p(x) \geq 0\ \forall x$, if $p(x)$ is any function with $p(x) \geq 0\ \forall x$. 1) *Discrete Filters.* Let the nonnegative function $p(n)$, $n \in \mathbf{Z}^m$, of the previous discussion be equal to the discrete unit impulse $\delta(n)$. Then $h(n) = h(n) * \delta(n) \geq 0$ $\forall n$. 2) *Analog Filters.* Express the Dirac unit impulse function as the limit $\delta(x) = \lim[p_k(x)]$ for $k \to \infty$, where $p_k(x)$, $x \in \mathbf{D} = \mathbf{R}^m$, is a nonnegative triangular pulse whose width goes to zero and height goes to $+\infty$ as $k \to \infty$ in a way such that $\int_{\mathbf{D}} p_k(x)dx = 1$ for all k. Then, if $k \to \infty$, $\forall x$, $h(x) = \int_{\mathbf{D}} h(z)\delta(x-z)dz = \int_{\mathbf{D}} h(z)[\lim p_k(x-z)]dz = \lim[\int_{\mathbf{D}} h(z)p_k(x-z)dz] \geq 0$, because $h * p_k(x) \geq 0$ since $p_k(x) \geq 0\ \forall k, \forall x$. *Q.E.D.*

For the analysis in this section we also need the following

THEOREM 7 (Maragos [20, p.126]). *Any translation-invariant and increasing FP filter defined on a class of u.s.c. functions closed under translation can be represented exactly as the supremum of erosions by all its kernel functions.*

Theorems 6 and 7 prepared the groundwork for the first representation of linear convolutions using morphological correlations. Formally, we have

THEOREM 8 . *Let ψ be a linear shift-invariant filter defined on a class $\mathcal{F}$ of u.s.c. functions closed under translation. Let also its impulse $h(x)$ satisfy the following two conditions:*

(a) $h(x) \geq 0$ *for all* $x \in \mathbf{D} = \mathbf{R}^m$ *(or* $\mathbf{Z}^m$*),*

(b) $\int_{\mathbf{R}^m} h(x)dx = 1$ *(or* $\sum_{n \in \mathbf{Z}^m} h(n) = 1$*).*

Then ψ is exactly represented as the supremum of erosions by all its kernel functions $g \in K(\psi)$; thus, for any $f \in \mathcal{F}$ and $x \in \mathbf{D}$,

$$(h * f)(x) = \sup_{g \in K(\psi)} \{ \inf_{z \in \mathbf{D}} \{f(z) - g(z-x)\}\} \tag{38}$$

Proof. ψ is increasing because of Theorem 6 and condition (a). Condition (b) makes ψ translation-invariant. Thus Theorem 7 completes the proof of (38). *Q.E.D.*

For *linear convolutions* with $h(x)$ we assume that $h(x) = 0$ iff $x \notin Ros(h)$, whereas for *morphological convolutions* with the structuring function $g(x)$ we assume that $g(x) = -\infty$ iff $x \notin Ros(g)$. Thus, when the input function f is linearly convolved with h in (38), we assume that $f(x) = 0$ outside $Ros(f)$. However, when f is morphologically convolved with g, we assume that $f(x) = -\infty$ outside $Ros(f)$.

Theorem 8 may have some theoretical interest, but its practicality is rather small because, in general, it is not a simple task to analytically find and describe all the (infinite in number) kernel functions that (38) requires. Our goal then is to find a subset of the kernel which is mathematically more tractable and can still represent the filter. We summarize below our approach. (The complete theoretical analysis is contained in [20].) The kernel of the LTI filter ψ, equipped with the function ordering $\leq$, becomes a *partially ordered set* (a *poset*). A **minimal element** of the poset $(\mathcal{K}(\psi), \leq)$ is a function $g \in \mathcal{K}(\psi)$ that is not preceded (with respect to $\leq$) by any other function of $\mathcal{K}(\psi)$. We define the set $\mathcal{B}(\psi)$ of all the minimal elements of $\mathcal{K}(\psi)$ as the **basis** of ψ. We have shown that the basis of any translation-invariant, increasing, and u.s.c. filter is nonempty and can exactly represent it. The basis of such a filter may be *finite* (as is the case for morphological, median, and order-statistic filters [20,13,32]), in which case the filter is realized exactly as the maximum of a finite number of erosions. For *discrete* LTI filters a *sufficient* condition to find a nonempty basis in their kernel is to have an impulse response of *finite extent*:

THEOREM 9 . *Let $h(n)$, $n \in \mathbf{Z}^m$, be the finite-extent impulse response of an increasing LTI discrete filter ψ which is defined on a class $\mathcal{F}$ of real-valued sampled functions closed under translation. Then the basis of ψ is equal to*

$$\mathcal{B}(\psi) = \{g \in \mathcal{F} : \sum_{k \in \mathrm{Ros}(h)} h(k) g(-k) = 0 \ \text{ and } \ g(n) = -\infty \Longleftrightarrow h(-n) = 0\} \tag{39}$$

Further, ψ can be represented exactly as the supremum of erosions by all its basis functions $g \in \mathcal{B}(\psi)$. That is, for any $f \in \mathcal{F}$ and $n \in \mathbf{Z}^m$,

$$(h * f)(n) = \sum_{k \in \mathrm{Ros}(h)} h(k) f(n-k) = \sup_{g \in \mathcal{B}(\psi)} \{ \min_{k \in [\mathrm{Ros}(g)]_n} \{ f(k) - g(k-n) \} \} \tag{40}$$

Proof. (a) *Basis*: Call $\mathcal{B}$ the class of functions given by (39), and let $\mathcal{M}$ be the true basis of ψ. We must show that $\mathcal{M} = \mathcal{B}$. $\mathcal{B}$ is nonempty, because $g^* \in \mathcal{B}$, where $g^*(n) = 0$

iff $h(-n) \neq 0$ and $g^*(n) = -\infty$ otherwise. Let now $g \in \mathcal{B}$. Then, $g \in \mathcal{K}(\psi)$ because $h * g(0) = 0$. Is g minimal? Suppose it is not. Then there is $f \in \mathcal{K}(\psi)$ such that $f \leq g$ and $f \neq g$. Since $h(n) \geq 0\ \forall n$, $0 \leq h * f(0) \leq h * g(0) = 0 \Longrightarrow h * f(0) = 0$. Since $f \leq g$ and $g(n) = -\infty\ \forall n \notin Ros(g)$, there exists $k \in Ros(g)$ such that $f(k) < g(k)$; this implies $h(-k)f(k) < h(-k)g(k)$ and thus $h * f(0) < h * g(0) = 0$: Contradiction! Hence, $g \in \mathcal{M}$, and thus $\mathcal{B} \subseteq \mathcal{M}$.

Let now $g \in \mathcal{M}$. All the basis functions g must have a minimal region of support $G = [Ros(h)]^s$, because only the indices $n \in G$ are required for g(n) in computing $h * g(0)$. Thus g satisfies one of the two requirements of (39). Suppose that $g \notin \mathcal{B}$. Let $h * g(0) = p > 0$ and consider the function $f(n) = g(n) - p$, $n \in G$. Then, $f \leq g$ and $f \neq g$. However, $h * f(0) = h * g(0) - p = 0$, and hence f is a kernel function of ψ that precedes g. Hence, g is not a minimal element: Contradiction! Therefore, $g \in \mathcal{B}$, and thus $\mathcal{M} \subseteq \mathcal{B} \subseteq \mathcal{M} \Longrightarrow \mathcal{M} = \mathcal{B}$.

(b) *Representation*: Since ψ is translation-invariant, increasing, and u.s.c. (due to the finite extent of h), it can be represented exactly as the supremum of erosions by its basis functions [20, p.135]; this proves (40). An alternative proof proceeds as follows: Let $Ros(h) = \{k_1, k_2, \ldots, k_N\}$ be the N-point finite region of support of $h(n)$. Let also $h_i = h(k_i)$, $f(n - k_i) = f_i$, and $g(-k_i) = g_i$, with $i = 1, 2, \ldots, N$ and $k_i \in Ros(h)$. Then we must prove that, $\forall n$,

$$\sum_i h_i f_i = \sup_{g \in \mathcal{B}(\psi)} \{\min_i \{f_i - g_i\}\} \tag{41}$$

subject to

$$\sum_i h_i g_i = 0 \ , \ \sum_i h_i = 1 \ , \ h_i \geq 0 \ .$$

For any n, f_i, $i = 1, \ldots, N$, and $\sum_i h_i f_i$ are arbitrary but fixed real numbers. Hence, among all functions in $\mathcal{B}(\psi)$ we can always find a basis function $g^\star$ defined by $g_i^\star = g^\star(-k_i) = f_i - \sum_i h_i f_i$, $i = 1, \ldots, N$. For each $g \in \mathcal{B}(\psi)$ there is a $j \leq N$ such that $f_j - g_j \leq f_i - g_i$ $\forall i \Longrightarrow h_i(f_j - g_j) \leq h_i(f_i - g_i)\ \forall i$. Summing the latter inequality for all i, we obtain $f_j - g_j = \min_i\{f_i - g_i\} \leq \sum_i h_i f_i$, $\forall g \in \mathcal{B}(\psi)$. However, $g^\star$ achieves this upper bound since $\min_i\{f_i - g_i^\star\} = \sum_i h_i f_i$, and thus the proof of (41) is complete. *Q.E.D.*

Thus, if an increasing LTI discrete filter ψ has a finite impulse response h, then its basis exists, but it contains an infinite number of functions g. These basis functions can be found by solving the linear equation $h * g(0) = 0$ subject to three constraints: (1) $h(n) \geq 0$, all n, (2) $\sum_n h(n) = 1$, and (3) $Ros(g) = [Ros(h)]^{\bullet}$. If $Ros(h)$ has exactly N points, then both h and all the g's are vectors inside the N-D Euclidean space $\mathbf{R}^N$. Since, in solving $h * g(0) = 0$, we can choose freely the $N-1$ values of g, the basis $\mathcal{B}$ is isomorphic to the $(N-1)$-D vector space $\mathbf{R}^{N-1}$ [20]. Moreover, $\mathcal{B}$ is the hyperplane of $\mathbf{R}^N$ that is perpendicular to the vector h, because $h * g(0) = 0$ corresponds to a zero inner-product of the vectors h and g. This is depicted in Fig. 11 for the spaces $\mathbf{R}^N$ and $\mathbf{R}^2$. In $\mathbf{R}^2$, the 2-point impulse response vector has only two components. Because the filter is LTI and increasing, all components of h are nonnegative and sum up to one. Hence, h is confined to move only along the line connecting the points (0,1) and (1,0) in $\mathbf{R}^2$. The basis of these filters is the 1-D space (line) of vectors g that is perpendicular to h and passes through the origin, as shown in Fig. 11. The next two simple examples clarify these concepts.

Examples. Consider the increasing LTI filter ψ_1 (a moving average) whose impulse response is $h(n) = 0.5[\delta(n) + \delta(n-1)]$, $n \in \mathbf{Z}$. Its kernel is equal to $\mathcal{K}(\psi_1) = \{f : f(n) + f(n-1) \geq 0\}$. Its basis functions g can be found from $h(0)g(0) + h(1)g(-1) = 0$; hence, $g(n) = r[\delta(n) - \delta(n+1)]$, $r \in \mathbf{R}$, and $g(n) = -\infty$ if $n \notin \{-1, 0\}$. Thus, from (40), we can exactly express ψ_1 by a supremum of minima:

$$0.5[f(n) + f(n-1)] = \sup_{r \in \mathbf{R}} \{\min\{f(n) - r, f(n-1) + r\}\} \tag{42}$$

for any input signal $f(n)$. Another way (independent of Theorem 9) to prove (42) can be found in [20, p.155].

Consider now a 3-point increasing LTI filter ψ_2 with impulse response $h(n) = a\delta(n-1) + b\delta(n) + c\delta(n+1)$, $n \in \mathbf{Z}$, where $a + b + c = 1$ and $a, b, c > 0$. Its basis functions g satisfy $ag(-1) + bg(0) + cg(1) = 0$; hence, $g(-1) = r \in \mathbf{R}$, $g(0) = s \in \mathbf{R}$, $g(1) = \frac{-ar-bs}{1-a-b}$, and $g(n) = -\infty$ if $n \notin \{-1, 0, 1\}$. Thus ψ_2 can be realized by both linear or morphological convolutions:

$$af(n-1)+bf(n)+cf(n+1) = \sup_{(r,s) \in \mathbf{R}^2} \{\min\{f(n-1) - r,\ f(n) - s,\ f(n+1) + \frac{ar+bs}{1-a-b}\}\} \tag{43}$$

for any input $f(n)$.

Theorems 8 and 9 require some constraints on the impulse response of the LSI filter, i.e., nonnegativity and area equal to one. These constraints are relaxed by the following

THEOREM 10 . *Any LSI discrete (resp. continuous) filter whose impulse response is absolutely summable (resp. integrable) can be represented exactly as the sum of two suprema of erosions, each followed by a gain factor.*

Proof. Assume that the filter ψ is discrete, because for continuous filters we need only to replace summation with integration in this proof. If $h(n)$, $n \in \mathbf{Z}^m$, is the impulse response of ψ, then we can represent ψ as the sum of two other filters ψ_P and ψ_N with impulse responses h_P and h_N, respectively, such that $h(n) = h_P(n) + h_N(n)$, $h_P(n) \geq 0$, and $h_N(n) \leq 0$ for all n. Since $h(n)$ is absolutely summable, the two sums $G_P = \sum_n h_P(n) \geq 0$ and $G_N = \sum_n h_N(n) \leq 0$ are finite real numbers. If $G_P = 0$, $h_P(n) = 0$ for all n, and $\psi = \psi_N$. Likewise, if $G_N = 0$, then $\psi = \psi_P$. Avoiding the trivial case where $h(n) = 0$ for all n, at least one of G_P and G_N must be nonzero. If $G_P > 0$, then we can realize ψ_P as a filter ψ_P^* with impulse response $h_P^*(n) = h_P(n)/G_P \geq 0 \; \forall n$, followed by the gain factor G_P. Likewise, if $G_N < 0$, we can realize ψ_N as a filter ψ_N^* with impulse response $h_N^*(n)/G_N \geq 0$ $\forall n$, followed by the gain factor G_N. The filters ψ_P^* and ψ_N^* satisfy the requirements of Theorem 8 or 9 and, hence, can be represented exactly as suprema of erosions. Thus the proof is complete. *Q.E.D.*

The requirement for absolute summability of the impulse response in Theorem 10 is equivalent to requiring that the LSI filter be *stable* in the bounded-input bounded-output sense. In addition, the gains G_P and G_N mentioned in the above proof may be equal to one, in which case we can rephrase Theorem 10 as follows. *Any stable LSI filter can be represented exactly by the sum of (at most) two suprema of erosions.*

The representations in (38) and (40) might be useful in *analysis of LSI* filters. They are not useful, however, for exact practical realization because the LSI filters have an infinite number of kernel or basis functions. In the case where a LSI filter admits a representation upon basis functions, if we quantize their amplitude and bound their range between certain limits, the supremum operation in (40) will be replaced by the maximum over a finite ensemble of basis functions. Of course, this realization will only *approximate* the true response of the LSI filter. The advantages of such a realization of the linear filter would be to realize it only by using max-min and additions and, thus, avoiding multiplications. Naturally, there are many questions that arise: How to quantize and/or bound the amplitude range of the basis functions, and how to quantify the approximation error? The answers to these questions are still the object of continuing research.

5 CONCLUSIONS

Signals can be classified into *sets* (binary signals) and *functions* (multilevel signals). Filters can be classified into *set-processing* (SP) and *function-processing* (FP). Certain FP filters are *function- and set-processing* (FSP). We provided necessary and sufficient conditions for FSP filters to commute with thresholding, because then they can be interpreted as SP filters. In this analysis we considered the set as the primary concept. Thus, a function can be represented by an ensemble of sets (its *cross-sections*) or by a single set (its *umbra*). Set operations on umbrae or cross-sections induce some isomorphic operations on the respective functions.

The four basic morphological filters are the *erosion, dilation, opening,* and *closing*, and they all stem from Minkowski set addition. We introduced two new morphological filters, the *open-closing* and *clos-opening*, which behave similarly to median filters (see also Part2). These six filters are translation-invariant (with respect to shifts of both the argument and the amplitude of signals), increasing, nonlinear, and generally noninvertible. Several properties of FP morphological filters have been studied concerning the conditions for commuting with thresholding, the fixed points of opening and closing, the invertibility of erosions and dilations, and the difference between binary and multilevel structuring functions. Some attractive features of morphological filters are: 1) *Simplicity* and *parallel* implementation of FSP filters (Pointwise min/max or Boolean AND/OR of shifted versions of the signal); 2) *Separability*; i.e., if a 2-D function g is the Minkowski sum of two 1-D functions, then the 2-D morphological filtering by g reduces to a cascade of two 1-D morphological filters; 3) Numerous *applications* in image processing and analysis; 4) Systematic detection and quantification of the shape and size of geometrical features in signals; 5) Operate on signals of both discrete and continuous argument; 6) *Duality*; for each morphological filter operating on a signal there is a dual filter operating on the background of the signal; 7) Increasing morphological FSP filters *commute with thresholding*; this reduces a multilevel to a binary signal filtering, which is easier to analyze and implement; 8) Erosions or dilations are the prototypes of a large class of linear and nonlinear filters; and 9) Morphological filters look at signals under study as sets and are defined through logical relations rather than arithmetic ones; consequently, if a signal processing problem is stated in terms of logical relations, then it can be expressed directly in terms of mathematical equations containing morphological operations.

The main difficulties in their analysis or design arise from their *nonlinearity* and the lack of *analytic* criteria to choose a structuring set or function.

We have related morphological to linear filters as follows. Linear translation-invariant (LTI) filters are linear shift-invariant (LSI) filters with dc-gain equal to one. LSI filters are increasing iff they have nonnegative impulse response. A LTI filter can be uniquely characterized by the set of its *kernel functions*. An increasing LTI filter can be exactly represented as a supremum of erosion by all its kernel functions; if, in addition, the filter is *discrete* and its impulse response is of *finite* extent, then it can be represented as the suprema of erosions only by its *minimal* kernel functions, which can be found by solving a linear equation. Finally, any *stable* LSI filer is exactly represented as the sum of two weighted suprema of erosions. The practical disadvantage of the above representation is that it uses an *infinite* number of kernel functions. The advantages are that linear convolutions can be realized by using only min/max and additions (without any multiplications), and that linear filters have been related to a large class of nonlinear filters.

ACKNOWLEDGEMENTS

We wish to thank the anonymous reviewer A whose detailed comments helped us to improve the readability and contents of this paper.

References

[1] G. Matheron, *Random Sets and Integral Geometry*, J. Wiley, N.Y., 1975.

[2] J. Serra, *Image Analysis and Mathematical Morphology*, Acad. Press, N.Y., 1982.

[3] H. Minkowski, "Volumen und Oberflache," *Math. Annalen*, Vol. 57, 1903, pp. 447-495.

[4] H. Hadwiger, *Vorlesungen uber Inhalt, Oberflache und Isoperimetrie*, Springer Verlag, Berlin, 1957.

[5] V. Goetcherian, "From Binary to Grey Tone Image Processing Using Fuzzy Logic Concepts," *Pattern Recog.*, Vol. 12, 1980, pp. 7-15.

[6] S. R. Sternberg, "Biomedical Image Processing" *IEEE Comput. Magaz.*, Jan. 1983, pp. 22-34.

[7] S. R. Sternberg, "Grayscale Morphology," *Comput. Vision, Graph., Image Processing*, **35**, pp.333-355, 1986.

[8] K. Rodenacker, P. Gais, U. Jutting, G. Burger, "Mathematical Morphology in Grey Images," in *Proc. 1983 European Signal Processing Conference.*

[9] Y. Nakagawa and A. Rosenfeld, "A Note on the Use of Local Min and Max Operations in Digital Picture Processing," *IEEE Trans. Syst., Man, Cybern.*, Vol. SMC-8, Aug. 1978, pp. 632-635.

[10] K. Preston, Jr., "Ξ-Filters," *IEEE Trans. Acoust., Speech, and Signal Processing*, Vol. ASSP-31, Aug. 1983, pp. 861-876.

[11] C. Lantuejoul and J. Serra, "M-Filters," in *Proc. 1982 IEEE Int. Conf. Acoust., Speech, Signal Processing*, Paris, France, May 1982, pp. 2063-2066.

[12] P. Maragos and R.W. Schafer, "Applications of Morphological Filtering to Image Processing and Analysis," in *Proc. 1986 IEEE Int. Conf. Acoust., Speech, Signal" Processing,* Tokyo, Japan, April 1986, pp. 2067-2070.

[13] P. Maragos, "Tutorial on Advances in Morphological Image Processing and Analysis," *Proc. SPIE* Vol. 707 Visual Commun. and Image Processing, pp.64-74, 1986.

[14] F. Meyer, "Iterative Image Transformations for an Automatic Screening of Cervical Smears," *J. Histochem. and Cytochem.*, Vol. 27, 1979, pp. 128-135.

[15] M. M. Skolnick, "Application of Morphological Transformations to the Analysis of Two-Dimensional Electrophoretic Gels of Biological Materials," *Comput. Vision, Graph., Image Processing*, **35**, pp.306-332, 1986.

[16] S. R. Sternberg and E. S. Sternberg, "Industrial Inspection by Morphological Virtual Gauging," in *Proc. IEEE Workshop Comput. Archit. Pattern Anal. Image Datab. Manag.*, Pasadena, CA, Oct. 1983.

[17] J. R. Mandeville, "Novel Method for Analysis of Printed Circuit Images," *IBM J. Res. Develop.*, Vol. 29, Jan. 1985, pp. 73-86.

[18] T. R. Crimmins and W. M. Brown, "Image Algebra and Automatic Shape Recognition," *IEEE Trans. Aerosp. Electron. Syst.*, Vol. AES-21, Jan. 1985, pp. 60-69.

[19] P. Maragos and R.W. Schafer, "A Unification of Linear, Median, Order-Statistics, and Morphological Filters under Mathematical Morphology," in *Proc. 1985 IEEE Int. Conf. Acoust., Speech, Signal Processing*, Tampa, FL, March 1985, pp. 1329-1332.

[20] P. Maragos, *A Unified Theory of Translation-Invariant Systems with Applications to Morphological Analysis and Coding of Images*, Ph.D. Thesis, School Electr. Enginr., Georgia Inst. Technology, Atlanta, GA, July 1985. Also *Tech. Rep. DSPL-85-1.*

[21] A. Rosenfeld and A.C. Kak, *Digital Picture Processing,* Vols. 1&2, Acad. Press, N.Y., 1982.

[22] P. Maragos and R.W. Schafer, "Morphological Skeleton Representation and Coding of Binary Images," *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. ASSP-34, Oct. 1986, pp. 1228-1244.

[23] M. Werman and S. Peleg, "Min-Max Operators in Texture Analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. PAMI-7, Nov. 1985, pp. 730-733.

[24] S. Suzuki and K. Abe, "New Fusion Operations for Digitized Binary Images and Their Applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. PAMI-7, Nov. 1985, pp. 638-651.

[25] J. C. Klein and J. Serra, "The Texture Analyzer," *J. Microscopy*, Vol. 95, Pt. 2, April 1972, pp. 349-356.

[26] R. M. Lougheed, D.L. McCubbrey, and S.R. Sternberg, "Cytocomputers: Architectures for Parallel Image Processing," *Proc. Workshop Picture Data Descr. Manag.*, Pacific Grove, CA, Aug. 1980.

[27] P.E. Danielsson and S. Levialdi, "Computer Architectures for Pictorial Information Systems," *IEEE Comput. Magaz.*, Nov. 1981, pp. 53-67.

[28] M.J. Kimmel, R.S. Jaffe, J.R. Mandeville, M.A. Lavin, "MITE: Morphic Image Transform Engine," *IBM Res. Rep. RC 11438*, Oct. 1985.

[29] R.G. Bartle, *The Elements of Real Analysis*, J. Wiley, N.Y. 1968.

[30] H.L. Royden, *Real Analysis*, MacMillan Publ. Co., N.Y., 1968.

[31] G. Choquet, *Topology*, Acad. Press, N.Y., 1966.

[32] P. Maragos and R. W. Schafer, "Morphological Filters-Part 2: Their Relations to Median, Order-Statistic, and Stack Filters," in this issue.

[33] K. Preston, Jr., M. J. B. Duff, S. Levialdi, P. E. Norgren, and J-I. Toriwaki, "Basics of Cellular Logic with Some Applications to Medical Image Processing," *Proc. IEEE*, Vol. 67, May 1979, pp. 826-856.

[34] S.R. Sternberg, "Parallel Architectures for Image Processing," in *Proc. IEEE Conf. Comput. Softw. Applic.*, Chicago, 1979, pp. 712-717.

[35] R. M. Haralick, S. R. Sternberg, X. Zhuang, "Grayscale Morphology," in *Proc. IEEE Conf. Comput. Vision Pattern Recogn.*, Miami, FL, June 1986, pp.543-550.

[36] J. P. Fitch, E. J. Coyle, and N. C. Gallagher, "The Analog Median Filter," *IEEE Trans. Circ. and Syst.*, Vol. CAS-33, pp.94-102, Jan. 1986.

TABLE I

Isomorphic operations between functions, their cross-sections, and their umbrae

FUNCTIONS	CROSS-SECTIONS (all $t \in \mathbf{V}$)	UMBRAE
$f \wedge g$	$X_t(f) \cap X_t(g)$	$U(f) \cap U(g)$
$f \vee g$	$X_t(f) \cup X_t(g)$	$U(f) \cup U(g)$
$minf(x) = -\infty^*$	$\emptyset$	$\emptyset$
$pinf(x) = +\infty^*$	$\mathbf{D}$	$\mathbf{E} = \mathbf{D} \times \mathbf{V}$
$f \leq g$	$X_t(f) \subseteq X_t(g)$	$U(f) \subseteq U(g)$

*for all $x \in \mathbf{D}$.

TABLE II

Properties of Set- and Function-Processing Filters

Properties	SP filter Ψ	FP filter ψ
translation-invar.	$\Psi(A_z) = [\Psi(A)]_z$	$\psi[f(x-y)+c] = [\psi(f)](x-y)+c^*$
increasing	$A \subseteq B \Longrightarrow \Psi(A) \subseteq \Psi(B)$	$f \leq g \Longrightarrow \psi(f) \leq \psi(g)$
extensive	$\Psi(A) \supseteq A$	$\psi(f) \geq f$
anti-extensive	$\Psi(A) \subseteq A$	$\psi(f) \leq f$
idempotent	$\Psi[\Psi(A)] = \Psi(A)$	$\psi[\psi(f)] = \psi(f)$

*$x, y \in \mathbf{D}$ and $c \in \mathbf{V}$.

CAPTIONS OF FIGURES

Figure 1. A function f, its cross-section $X_t(f)$ at level t, and its umbra $U(f)$.

Figure 2. Reconstruction of a function from its cross-sections.

Figure 3. (a) Umbra of a set B; (b) Umbra and graph of a function f.

Figure 4. Isomorphic operations between sets and functions.

Figure 5. Erosion, dilation, opening, and closing of X by B (the shaded areas correspond to the interior of the sets, the dark solid curve to the boundary of the transformed sets, and the dashed curve to the boundary of the original set).

Figure 6. Dilations and erosions of discrete sets: (a) Minkowski subtraction, (b) Erosion, (c) Minkowski addition, (d) Dilation, (e) Forming larger sets as the Minkowski sum of simpler sets. ($\bullet$ = set points; + marks origin (0,0) of $\mathbf{Z}^2$.)

Figure 7. A 2-D structuring element B, a 1-D structuring function g, a 1-D structuring set S, and their umbrae.

Figure 8. Erosion, dilation, opening, and closing of a function by a set $B = \{-2, -1, 0, 1, 2\}$. (The dashed curve refers to the original function.)

Figure 9. Morphological filtering of a cosine $\cos \omega x$ by the structuring function $g(x) = A(\cos \omega x - \cos \omega L)$, $-L \leq x \leq L$. ($L = 0.39\pi/\omega$.)

Figure 10. A 1-D original function f and its morphological filtering by a rectangle ($B = \{n \in \mathbf{Z} : -5 \leq n \leq 5\}$), a disk ($g(n) = 5\sqrt{25 - n^2}$, $n \in B$), and a parabola ($h(n) = 2(25 - n^2)$, $n \in B$).

Figure 11. Geometric interpretation of the relation between the finite impulse response h and the basis functions g of an increasing LTI discrete filter.

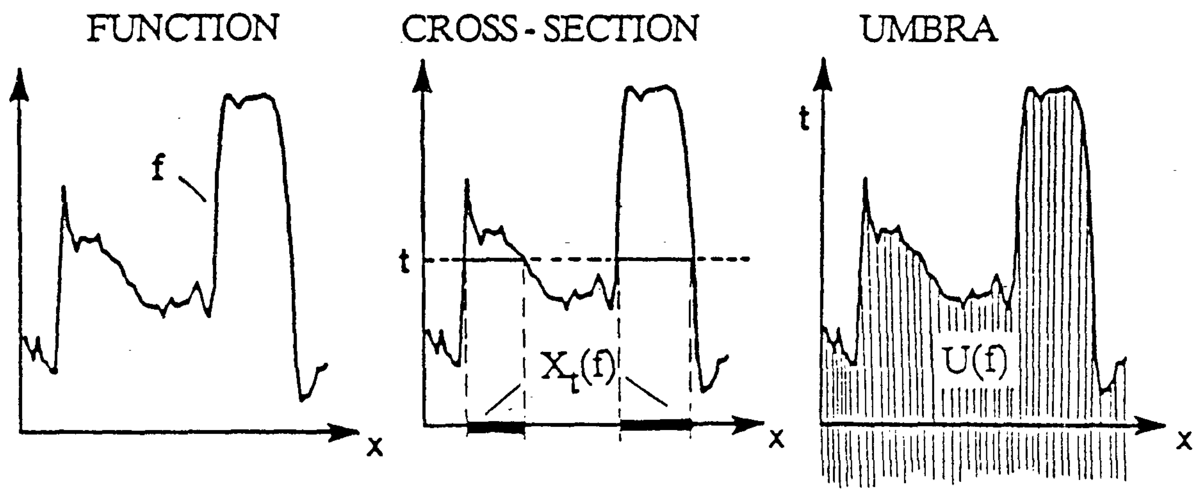


FIGURE 1

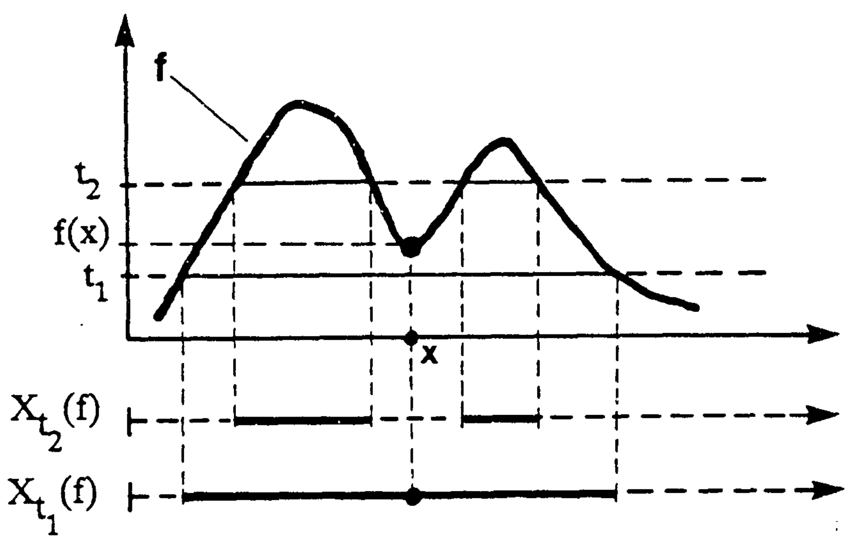
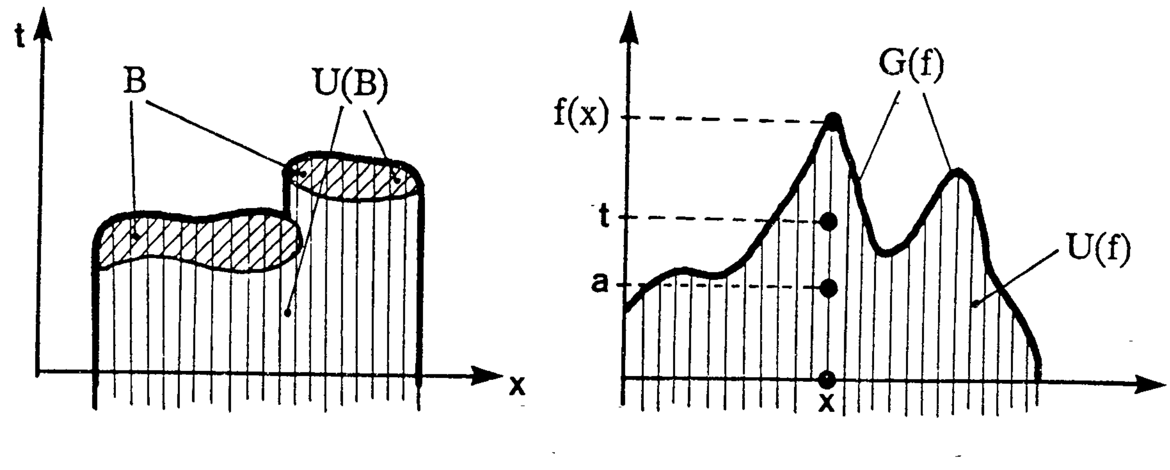


FIGURE 2



(a) (b)

FIGURE 3

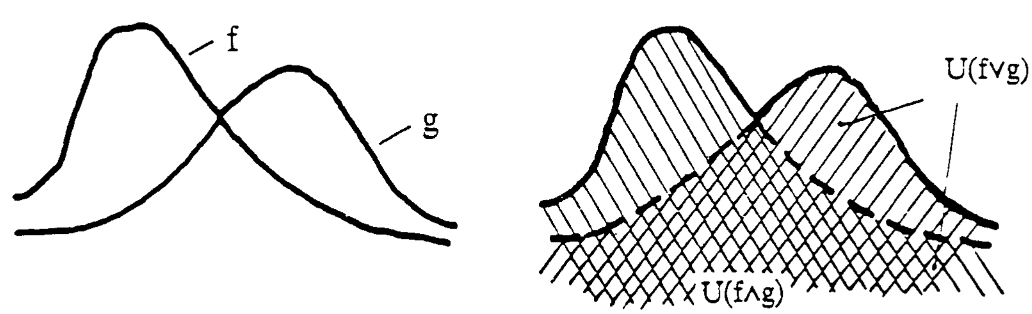


FIGURE 4

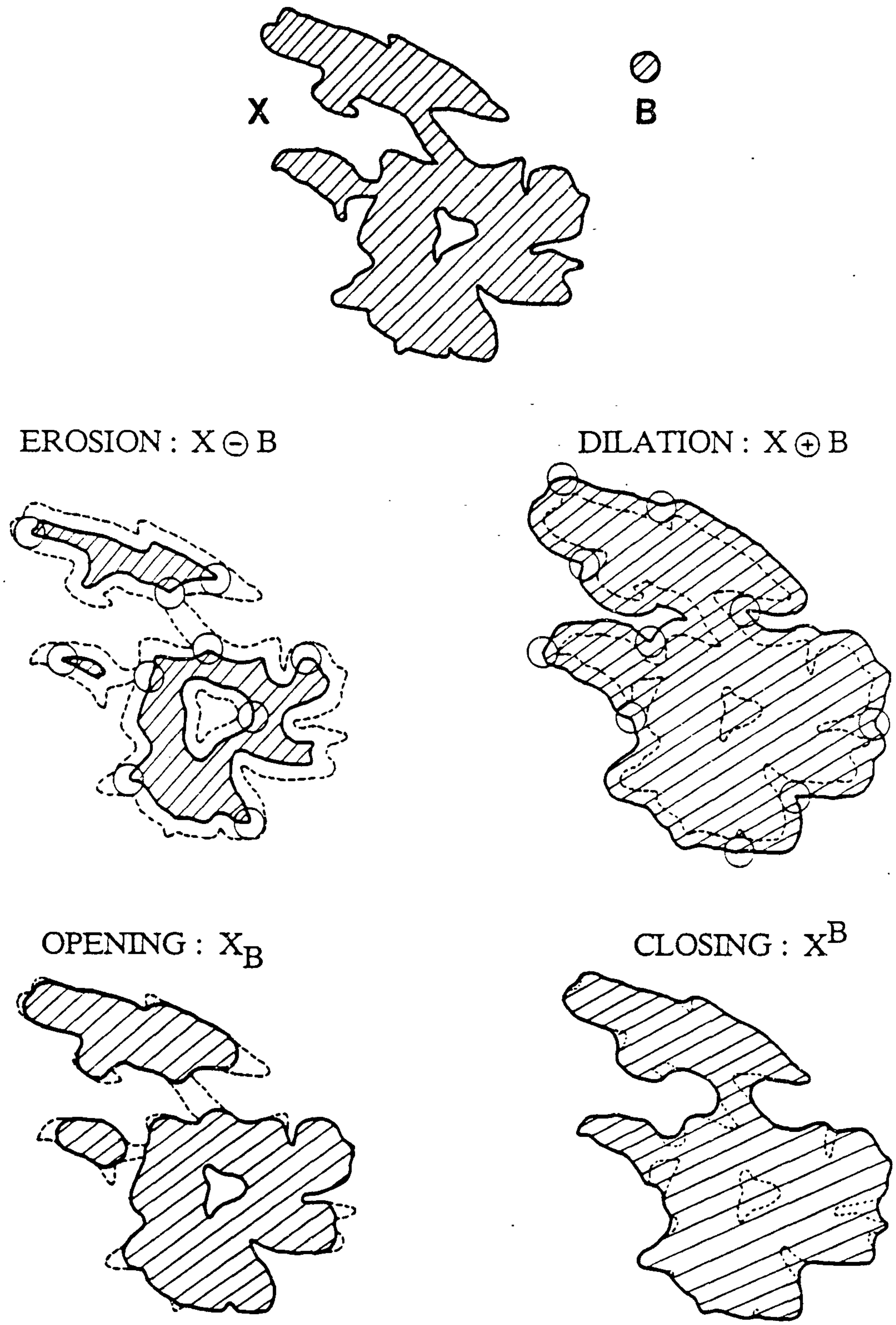


FIGURE 5

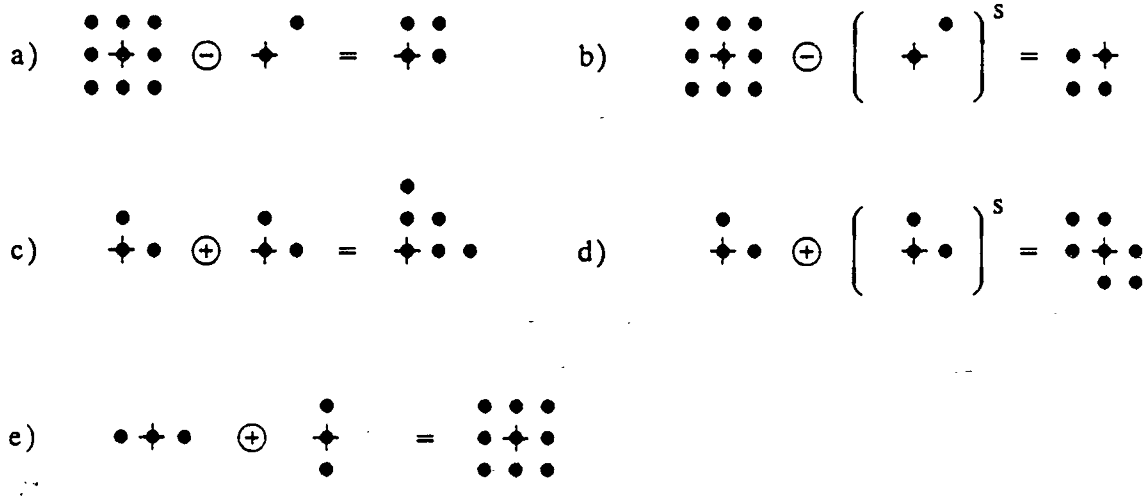


FIGURE 6

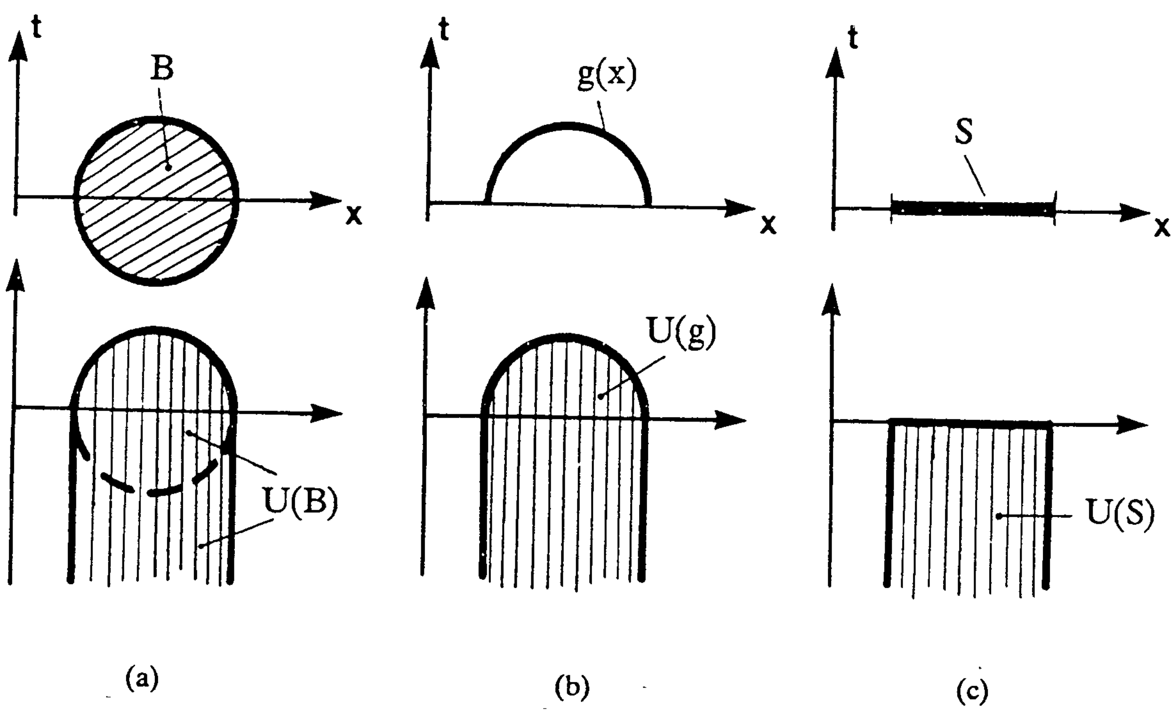


FIGURE 7

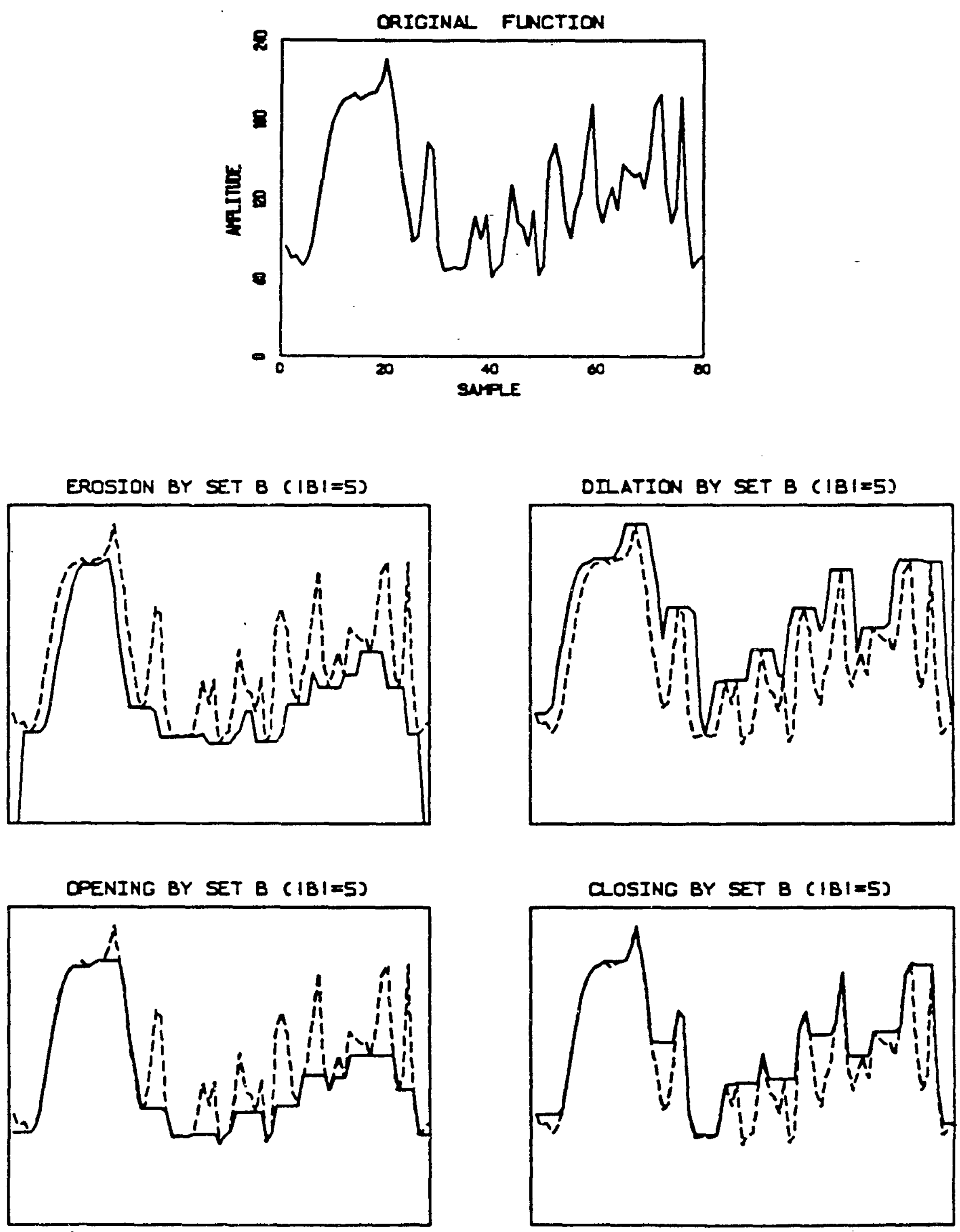

ORIGINAL FUNCTION
AMPLITUDE
0
20
40
60
80
SAMPLE
EROSION BY SET B (|B|=5)
DILATION BY SET B (|B|=5)
OPENING BY SET B (|B|=5)
CLOSING BY SET B (|B|=5)

FIGURE 8

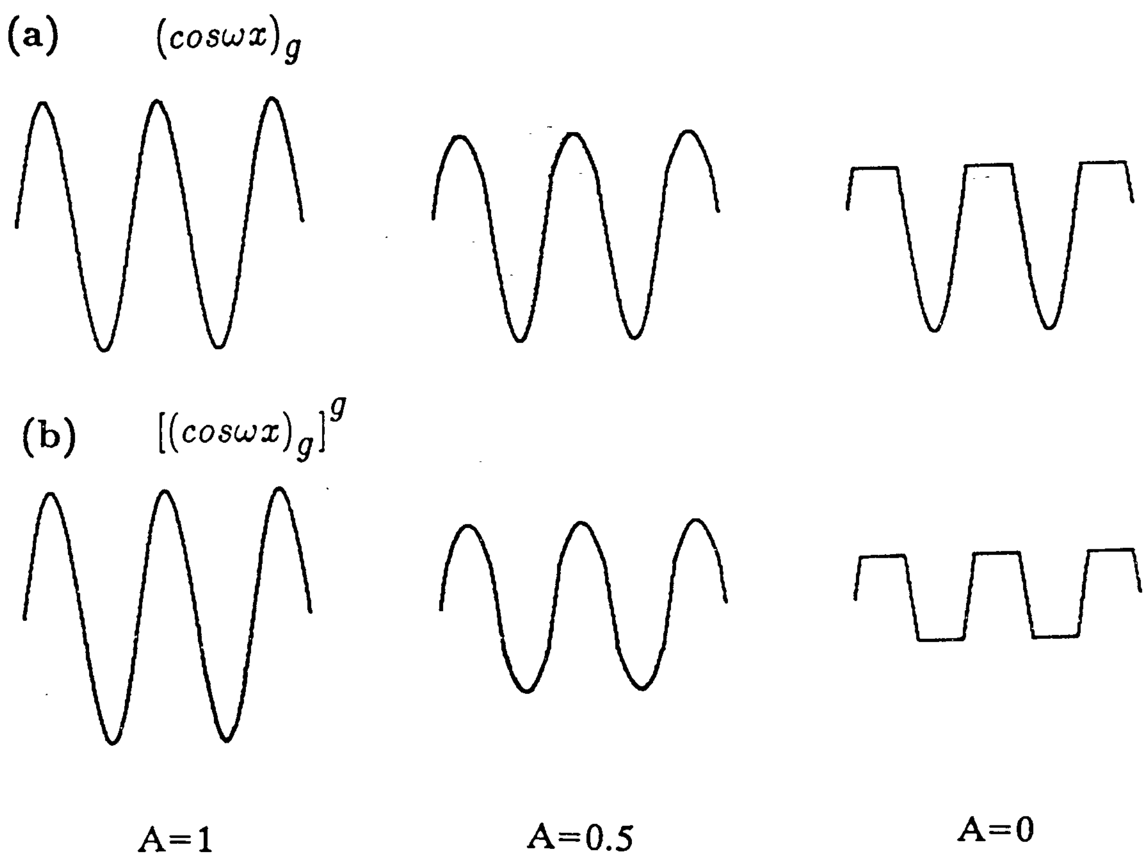
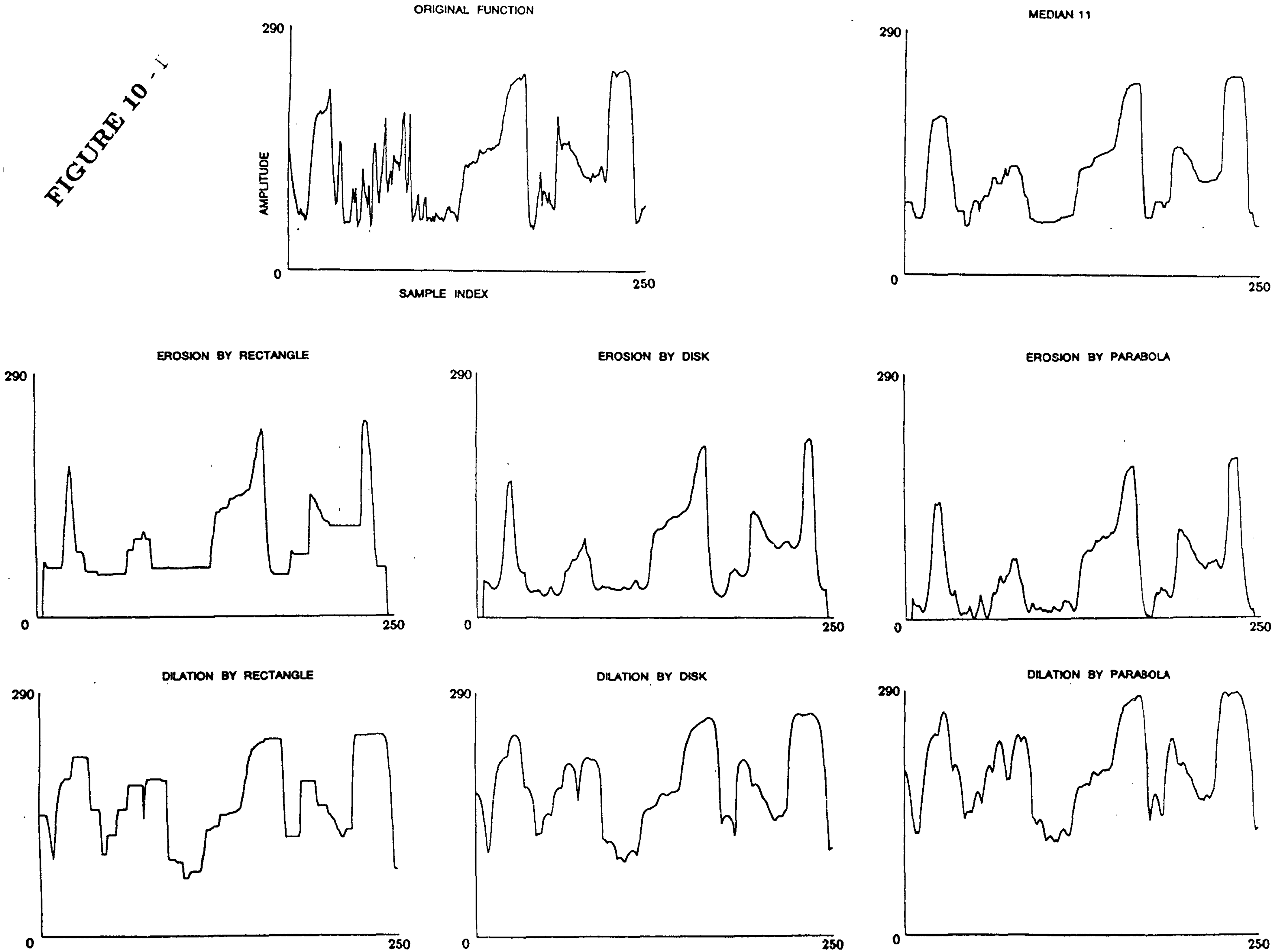


FIGURE 9

FIGURE 10 - 1


ORIGINAL FUNCTION
290
AMPLITUDE
0
SAMPLE INDEX
250
MEDIAN 11
290
0
250
EROSION BY RECTANGLE
290
0
250
EROSION BY DISK
290
0
250
EROSION BY PARABOLA
290
0
250
DILATION BY RECTANGLE
290
0
250
DILATION BY DISK
290
0
250
DILATION BY PARABOLA
290
0
250

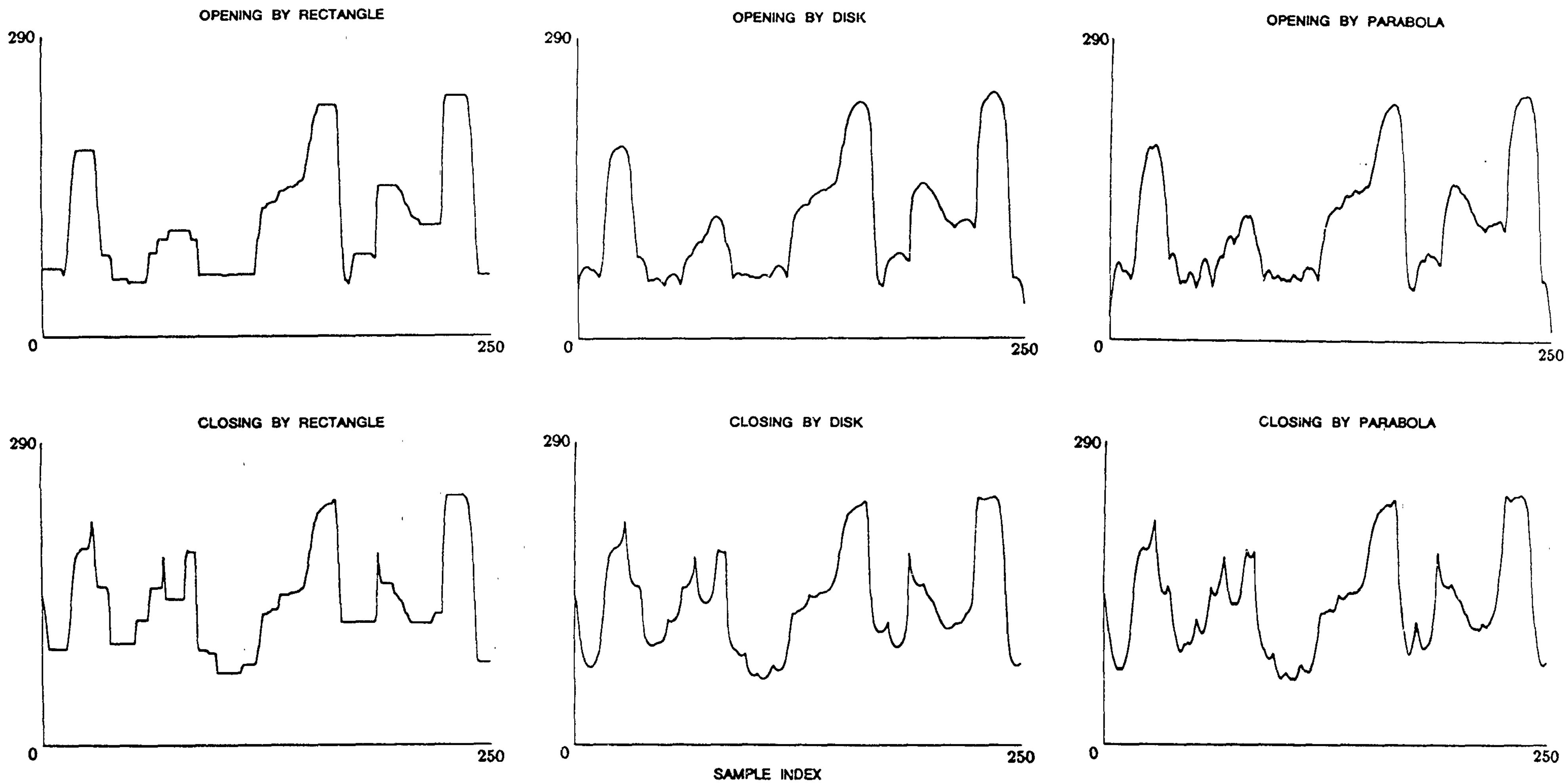


FIGURE 10

(CONTINUED)

II

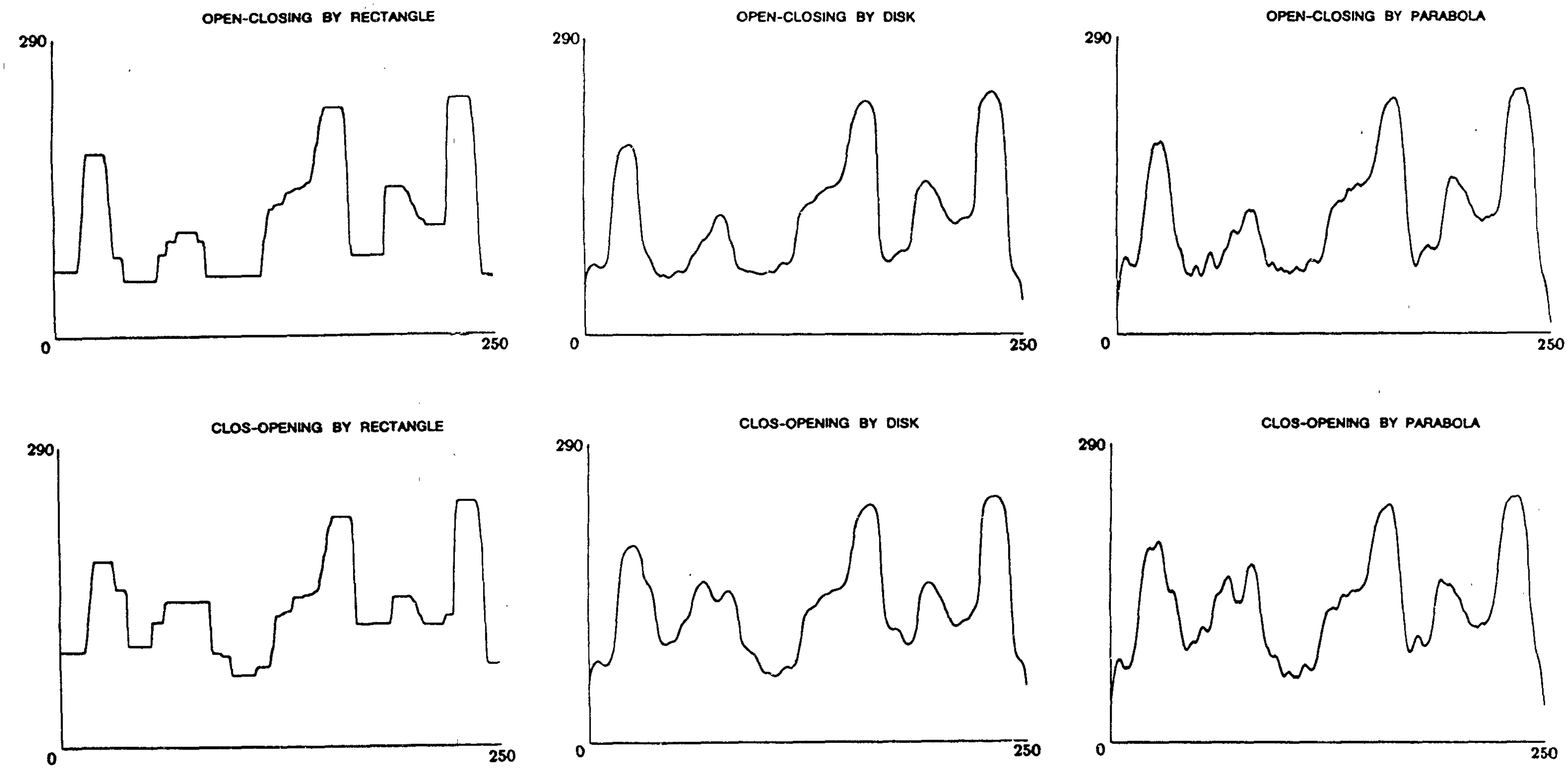
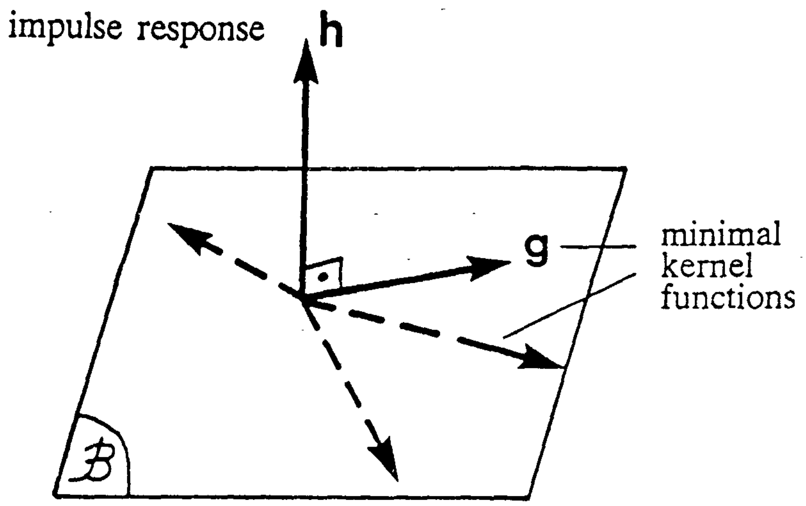


FIGURE 10

(CONTINUED)

III

(a) space R^N



(b) space R^2

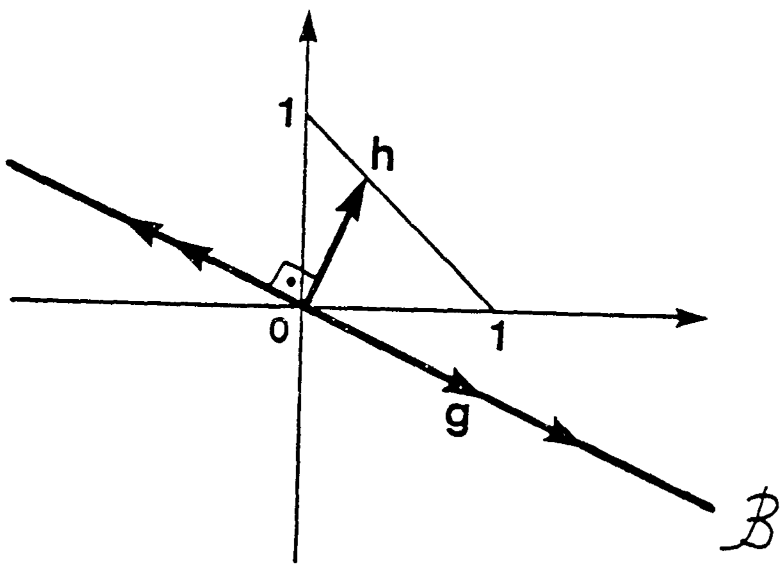


FIGURE 11