UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

LULU operators on multidimensional arrays and applications

by
Inger Nicolette Fabris-Rotelli

Submitted in partial fulfilment of the requirements for the degree

Magister Scientiae

In the Department of Mathematics and Applied Mathematics
In the Faculty of Natural and Agricultural Sciences
University of Pretoria
Pretoria
November 2009

I, Inger Nicolette Fabris-Rotelli declare that the dissertation, which I hereby submit for the degree Magister Scientiae in Applied Mathematics at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

SIGNATURE:

DATE: 2 November 2009

Summary

The LULU operators, L_n and U_n, are smoothers, that is they smooth data received as a signal. They are nonlinear and this nonlinearity makes them more robust but also more complicated to study since the projection theorem does not hold. Their smoothing action is aimed at removing the impulsive noise present in any received signal. A signal can be of one or two dimensions, or of any higher dimension. In one dimension a signal is represented as a sequence and in two dimensions as an image. Higher dimensions include video feed and other more complex data streams. Carl Rohwer developed the LULU smoothers for sequences over the last three decades and the need for an extension to higher dimensions became more and more obvious as the applications of these smoothers were investigated. Perhaps the most important application is that of the Discrete Pulse Transform which is obtained via recursive application of the smoothers. In this dissertation the extension to dimensions higher than one is presented. All the essential properties developed for the one dimensional smoothers are replicated in this work. In addition, the Discrete Pulse Transform is used to illustrate some simple applications to image smoothing and feature detection.

Acknowledgements

Firstly, enormous thanks are due for my supervisor Prof. Roumen Anguelov. Without his unending patience and persistence to do better and to do more than the norm this work would never have developed to what it has. He has developed my research skills and thinking processes immensely since I started working with him. I thank him for this wonderful learning experience and look forward to future research and growth in myself with him.

Secondly, thanks are in order to the Department of Defence, Armscor and the CSIR for presenting their research opportunity to me. Thank you also for the biannual meetings with other researchers to share ideas and do some brainstorming. These meetings contributed greatly to my knowledge and ideas.

Lastly and most importantly, thank you to my husband Steven. His continuous support and urging on was much needed and I really appreciate it. Everybody needs someone to share your daily life with its challenges with you and Steven provides that constantly to me with undeniable love. Thank you my angel.

Contents

Chapter 1

Introduction

The LULU operators, L_n and U_n, are smoothers, that is they smooth data received as a signal. They are nonlinear and this nonlinearity makes them more robust but also more complicated to study since the projection theorem does not hold. Their smoothing action is aimed at removing the impulsive noise present in any received signal. A signal can be in one or two dimensions, or of any higher dimension. In one dimension a signal is represented as a sequence and in two dimensions as an image. Higher dimensions include video feed and other more complex data streams. Carl Rohwer developed the LULU smoothers for sequences over the last three decades. Their characterization as smoothers and separators, shape preservation properties and total variation preservation property make them powerful smoothers. The need for an extension to higher dimensions became more and more obvious as the applications of these smoothers were investigated, which included a possible use on images. Perhaps the most important application in that of the Discrete Pulse Transform which is obtained via recursive application of the smoothers. It results in a multiresolution decomposition of the sequence, that is into a number of resolution layers. At each layer, the respective resolution is represented, or in other words structures of the sequence at that resolution are represented. These powerful operators, their properties, as well as the Discrete Pulse Transform, now receive attention in higher dimensions, specifically for application in image processing. All the properties developed for the one dimensional smoothers are replicated making use of the morphological concept of a *connection*, [108]. In addition, the Discrete Pulse Transform is used to illustrate some simple applications to image smoothing and feature detection.

The nonlinearity of the LULU smoothers, as mentioned above, make theoretical development more complicated than for linear operators. However, taking on the additional complexity is justified since in two dimensions an image is basically the transformation of data by a human eye or measuring instrument. This transformation is significantly complicated to be considered nonlinear, [95]. Thus the analysis of images via nonlinear operators is more logical than that of linear.

We present first the background of Carl Rohwer and his collaborators' work on the LULU smoothers. This is presented in Chapter 2. Their connection with the field of Mathematical Morphology is also presented. In Chapter 3 we present the extension of the one dimensional operators on $\mathbb{Z}$ to higher dimensions, namely $\mathbb{Z}^d$, and in Chapter 4 the Discrete Pulse Transform in two dimensions, with applications to image processing, as well as the implementation and distributions of the operators in two dimensions.

Chapter 2

Carl Rohwer's LULU Theory

2.1 Introduction

Carl Rohwer's work on the so-called LULU theory began around 1983 while working on practical work at the Institute for Maritime Technology in Simonstown, South Africa. His work on these operators continued while he was at the BMI in Stellenbosch and has now been developed into an extensive theory, with collaborators such as Wild (theory) and Laurie (implementation), of which the two main resources are [93] and [95]. The first work on the LULU operators is presented in [86], the properties of these operators in [87], [88], [90], [91], [135], [92], [93] [29], [95], [28] and [26], a comparison with the median filters in [89] and [85], the resulting Discrete Pulse Transform in [38], [64], [66] and [63], the statistical distributions and properties in [56], [94] and [27], and the LULU operators on a continuous domain are dealt with in [73], [4] and [3]. An application of Rohwer's LULU operators for images in dealt with in [58].

LULU theory is based on two operators L_n and U_n which represent the words 'lower' and 'upper'. These operators are smoothers and are applied to a bi-infinite sequence of real numbers, $x = (x_n)_{n\in\mathbb{Z}}$ with the assumption that the ℓ_1 norm is finite, i.e. $\|x\|_1 = \sum_n |x_n| < \infty$. This sequence is more commonly referred to as a signal. It can also be viewed as a time series. Let the space of all such sequences be

$$X = \{x = (x_n)_{n\in\mathbb{Z}} : \|x\|_1 < \infty\}. \tag{2.1}$$

In applications x is not usually bi-infinite i.e. is finite, but zeros are appended to the left and right so that the subscript n goes from $-\infty$ to ∞. A complication arises when we have to smooth at the two ends of the finite sequence. There are a number of options, some of which are

(1) **Copy-on end-value rule, [115] [49]:** For those points which do not fit in a required window size, leave the values as they are.

(2) **Replicate end-value rule, [49]:** Replicate the end points with respect to the required window size so that all original sequence values are smoothed. This is the most common method.

(3) **Extrapolation end-value rule, [115]:** Linear extrapolation is used on the smoothed values to get the smoothed values outside of the sequence values.

(4) **Cyclic end-value rule, [49]:** For periodic data a cycle of values may be known in which case end points can be predicted using values at the some cycle point which have been smoothed.

(5) **Step-down end-value rule, [49]:** Decrease the window size near the end points so that only values in the sequence are used.

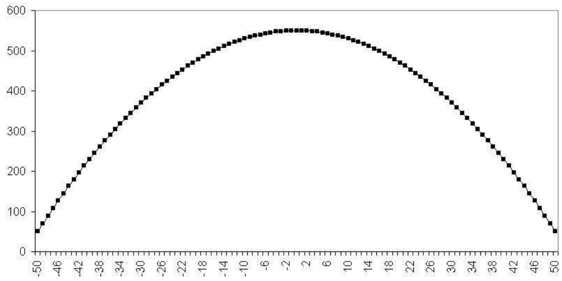
(6) **Omit end-point rule, [49]:** Simply remove the point which cannot be smoothed due to insufficient window size.

(7) **Velleman's end-value rule, [123]:**

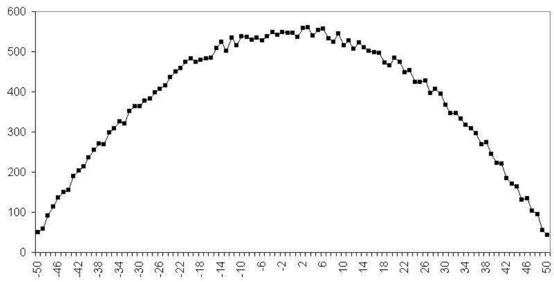
$$T(x)_i = median\{3Tx_{i+2} - 2Tx_{i+1}, x_i, Tx_{i+1}\}$$

As will be seen in Chapter 3, the definition for the multidimensional LULU operators allows for all possible windows (connected neighbourhoods) of size n of a specific point. In higher dimensions there are numerously more possible windows thus lessening the effect dealt with above at edge points. As a side note, the elements in the sequence need not be spaced evenly as nonlinear operators perform well in this situation, [123], that is, the underlying mesh need not be uniform.

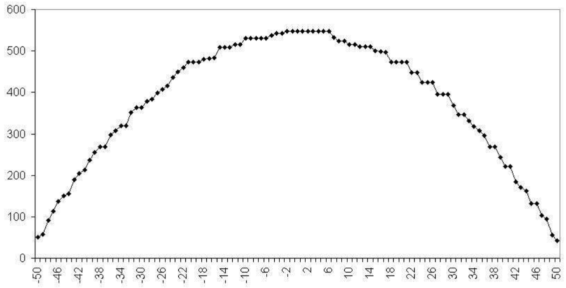
The data for a signal is obtained from a specific practical application and as with most real-world collected data sets there will be inherent noise intermingled with each true data value. Some examples, see [93], are the transmission of data from an instrument in outer space, the measurement of a specific pixel's intensity via a video camera, the measurement of a missiles trajectory via radar, the measurements of a speedometer of a small boat on rough seas (the waves will toss the boat into the air adding noise into the speed measurement), the measurement of speech, voice or sound, as well as numerous medical applications. As can be seen from these examples the inherent noise is usually added or introduced by the measuring instrument, whether it be a camera, speedometer, radar, sound recorder or simply human measurement. The following (artificial) example shows first, a pure signal uncontaminated by noise and then, the same signal with added random noise, to illustrate noise removal via the LULU operators. A noiseless parabola is shown in Figure 2.1(a). In Figure 2.1(b) we see the same parabola but with uniform noise added to it. The total variation (see Definition 40 later in this chapter) indicates how the noise adds to the data. The respective total variation in Figures 2.1(a) and (b) is 1000 and 1441. By applying the one-dimensional LULU operator $P_5 \circ P_4 \circ ... \circ P_1$ where $P_n = U_n \circ L_n$ we smooth the noisy data, see Figure 2.1(c). The total variation is also reduced close to the original total variation, namely 1001.



(a) Noiseless Parabola (TV = 1000)



(b) Noisy Parabola (TV = 1441)



(c) Smoothed Parabola i.e. $L_5 \circ U_5$ (TV = 1001)

Figure 2.1. Noise Removal via One-Dimensional LULU Operators

To get an indication of the pattern of the true data, smoothers are used to eliminate the noisy phenomena. The distribution of the noise can be assumed to be Gaussian or non-Gaussian. This process is known as linear smoothing or linear signal extraction when the smoothed values are a linear transformation of the observed values. Recall that a **linear transformation** [62] $T : \ell_1 \to \ell_1$ is one which, for all scalar sequences r and sequences $u, v \in \ell_1$, (1) preserves addition,

$$T(u + v) = T(u) + T(v), \tag{2.2}$$

and (2) preserves scalar multiplication,

$$T(ru) = rT(u). \tag{2.3}$$

A smoother (linear or nonlinear) does, however, not provide a perfect process for noise removal as the true nature of the noise will usually be unknown or only known to a certain degree. Thus the hope is that the noise is removed, whatever its form, to the best of our ability with the smoother concerned and this is therefore a mechanism in which different smoothers' capabilities can be compared. When comparing smoothers the following should be taken into consideration, [49],

1. The effectiveness of the smoothers ability to reduce the noise present while maintaining the true signal.
2. The smoother's response to outliers and large changes in the signal.

Let us look at a well known linear smoother, namely the Haar wavelets. The Haar wavelets were introduced in 1910 by the Hungarian mathematicain Alfred Haar, [51]. They form part of the larger theory of wavelet analysis which began in the 1980's as an improvement of Fourier analysis, [120]. The Fourier expansion involves an infinite series of sines and cosines, but provides only frequency resolution and not time resolution in the expansion. Wavelet analysis, however, provides both resolutions, thus belonging to the larger field of the aptly termed multiresolution analysis, [120].

A family of wavelets constructed from a function $\Psi(t)$ (called the 'mother wavelet'), is given by

$$\Psi_{a,b}(t) = |a|^{-1/2}\Psi\left(\frac{t-b}{a}\right) \tag{2.4}$$

where a is a the scaling parameter, b a translation parameter and the fraction $|a|^{-1/2}$ the normalization constant, [120] [134]. The wavelet for a function f is then given by, [134],

$$W_\Psi(f)(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t)\Psi\left(\frac{t-b}{a}\right)dt.$$

If $\Psi(t)$ satisfies the following admissibility criterion,

$$\int \frac{|\mathbf{\Psi}(\omega)|^2}{|\omega|} d\omega < \infty$$

where $\mathbf{\Psi}$ is the Fourier transform of Ψ, then Ψ can be used to analyze and reconstruct the signal without loss of information, [109]. This complete reconstruction is an important result which should be compared with the Discrete Pulse Transform's complete reconstruction, see Section 2.4.

The Haar wavelet is the simplest and oldest orthogonal wavelet and has applications in image coding [113] [110], edge extraction [113] [110], binary logic design [113] and multiplexing [110], amongst others. The Haar wavelets are obtained when Ψ is the following function,

$$\Psi(x) = \begin{cases} 1 & 0 \le x < 1/2 \\ -1 & 1/2 < x \le 1 \\ 0 & \text{otherwise,} \end{cases}$$

so the wavelet family is given by what follows below.

The Haar wavelet family on $[0,1]$ is given by the following,

$$\begin{aligned}
\Psi_0(t) &= 1 \text{ on } 0 \le t \le 1 \\
\Psi_1^1(t) &= \begin{cases} 1, & 0 \le t \le 1/2 \\ -1, & 1/2 < t \le 1 \end{cases} = \Psi(t) \\
\Psi_2^1(t) &= \begin{cases} \sqrt{2}, & 0 \le t \le 1/4 \\ -\sqrt{2}, & 1/4 < t \le 1/2 \\ 0, & 1/2 < t \le 1 \end{cases} \\
\Psi_2^2(t) &= \begin{cases} 0, & 0 \le t < 1/2 \\ \sqrt{2}, & 1/2 \le t \le 3/4 \\ -\sqrt{2}, & 3/4 < t \le 1 \end{cases} \\
&\vdots \\
\Psi_n^m(t) &= \begin{cases} 2^{(n-1)/2}, & \frac{m-1}{2^{n-1}} \le t < \frac{m-1/2}{2^{n-1}} \\ -2^{(n-1)/2}, & \frac{m-1/2}{2^{n-1}} \le t < \frac{m}{2^{n-1}} \\ 0, & \text{on the rest of the interval } [0,1] \end{cases}
\end{aligned}$$

Note that we use the notation $\Psi_n^m = \Psi_{a,b}$ with $a = 2^{n-1}$ and $b = m - 1$. Figure 2.2 illustrates the first few Haar wavelets.

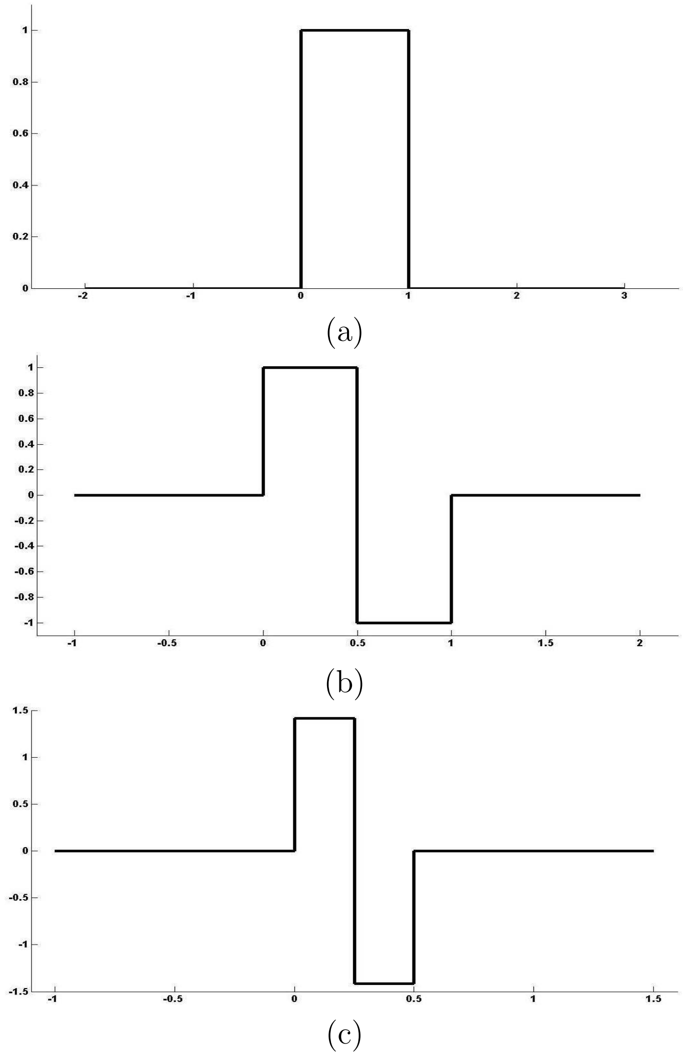


Figure 2.2. (a) Ψ_0 (b) Ψ_1^1 (c) Ψ_2^1

The Haar wavelets are orthogonal, [51], in the sense that

$$\int_0^1 \Psi_{n_1}^{m_1}(t)\Psi_{n_2}^{m_2}(t)dt = \delta_{(n_1,m_1),(n_2,m_2)},$$

where δ_{ij} is the Kronecker-Delta function. Any continuous function f can then be written as a series expansion, [110],

$$f(t) = c_0 + \sum_{n=0}^{\infty} \sum_{m=0}^{2^n-1} c_n^m \Psi_n^m(t),$$

where

$$c_n^m = \int_0^1 f(t)\Psi_n^m(t)dt.$$

The partial sums are then represented by S_N,

$$S_N(t) = c_0 + \sum_{n=0}^{N} \sum_{m=0}^{2^n-1} c_n^m \Psi_n^m(t).$$

If f is continuous then S_N converges to f uniformly and if f is discontinuous, but with all discontinuities as binary-rational points, the uniform convergence still holds, [110]. (A point x is binary-rational if there exist integers k and P such that $x = k/2^P$ where $k = 0, 1, ..., 2^P$.) Since the partial sum S_N is a step function with 2^N steps, the value of S_N at t is the mean value of f on the step interval containing t. The equation

$$\frac{d}{d\alpha} \int_{t_1}^{t_2} |f(t) - \alpha|^2 dt = 0$$

has the solution

$$\alpha = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} f(t)dt,$$

showing that S_N is the best approximation to f on the space spanned by $\{\Psi_n^m, m = 0, 1, ..., 2^n - 1, n = 1, 2, ..., N\}$ in the mean-square-error sense, [110].

A signal x is a sampling of some function f and is thus discrete. It also has a Haar wavelet expansion. We express this in terms of the Haar scaling function $\Psi_0(t)$, [93], [38]. The sequence $x = \{x_0, x_1, ..., x_{2^N}\} = \{f(0), f(1), ..., f(2^N)\}$ has the following representation

$$x = \sum_{i=0}^{N_0} \alpha_{0,i} \Psi_0(t - i),$$

where $N_0 = 2^N$.

We can see then that $\alpha_{0,i} = x_i$ for each i. Let $\mathcal{B}_0 = \text{span}\{\Psi_i : \Psi_i(t) = \Psi_0(t - i), i = 0, 1, ..., N_0\}$ and $\mathcal{B}_j = \text{span}\{\Psi_i : \Psi_i = \Psi_0(2^{-j}t - i), i =$

$0, 1, ..., N_j$ where $N_j = 2N_{j-1}\}$. Then $x \in \mathcal{B}_0$ and the Haar wavelets Ψ_k^m, $m = 1, ..., 2^{n-1}$, obtain the best least squares estimate for $x^{(k)} \in \mathcal{B}_k$ from $x^{(k-1)} \in \mathcal{B}_{k-1}$. Let P_k represent this operation into $\mathcal{B}_k$. Then we have the smoothed sequence,

$$x^{(k)} = P_k(x^{(k-1)}) = \sum_{i=0}^{N_k} \alpha_{k,i}\Psi_0(2^{-k}t - i),$$

with coefficients $\alpha_{k,i} = \frac{1}{2}(\alpha_{k-1,2i} + \alpha_{k-1,2i+1})$, and with $N_k = 2N_{k-1}$. The part which was removed is

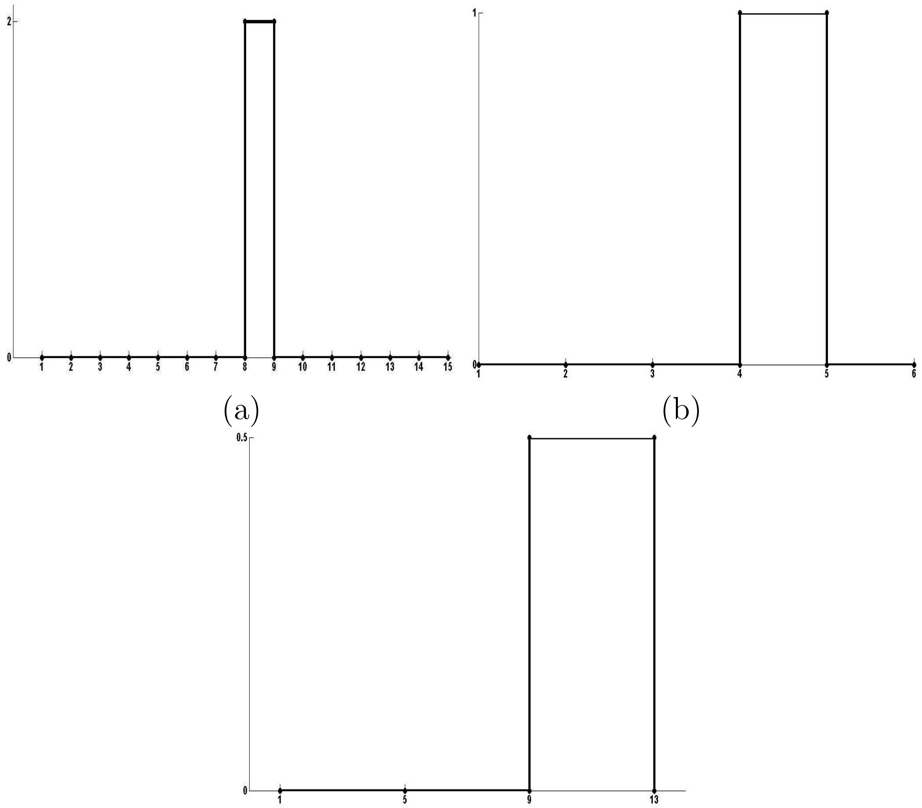
$$(I - P_k)(x^{(k-1)}) = \sum_{i=0}^{N_k} \beta_{k,i}\Psi_1^1(2^{-k}t - i)$$

where

$$\begin{aligned}\beta_{k,i} &= \alpha_{k-1,2i} - \alpha_{k,i}\\ &= \alpha_{k-1,2i} - \frac{1}{2}(\alpha_{k-1,2i} + \alpha_{k-1,2i+1})\\ &= \frac{1}{2}(\alpha_{k-1,2i} - \alpha_{k-1,2i+1}).\end{aligned}$$

Note that we are smoothing the sequence in the opposite direction away from f, from which the original sampling was taken. For example, if there were originally 8 elements in the sequence there will be 4 elements in $P_1(x)$, 2 elements in $P_2(P_1(x))$ etc. This allows a comparison with the LULU operators and their resulting Discrete Pulse transform.

This Haar smoother is linear and has a number of disadvantages. Firstly, the signal may be contaminated when large impulsive noise is involved. In Figure 2.3(a) an impulsive noise spike can be seen. By applying P_1 the spike is reduced but spread out, see Figure 2.3(b). In Figure 2.3(c) further reduction and spreading out can be seen. This indicates that the noise is actually not removed but rather spread out to contaminate the signal. The neighbour trend preservation property of the LULU operators prevent this disadvantage in their application, see Section 2.3. In addition the LULU operators allow for removal of such a pulse without contaminating the signal. Secondly, the total variation is not preserved. The total variation in the original sequence x is 2, however the total variation in $P_1(x)$ is 1 and in $(I - P_1)(x) = 2$. This gives us a total variation of 3 indicating an increase in variation and not variation preservation. Similarly the total variation in $P_2(P_1(x))$ is 0.5, and in $(I - P_2)(P_1(x))$ is 1, again greater than 1, the original total variation.



(a) (b) (c)

Figure 2.3. (a) Original Sequence x (b) $P_1(x)$ (c) $P_2(P_1(x))$

In Figure 2.4 we see another disadvantage of the Haar operators. In $(I-P)x$ we see that anomalies are removed at the two far edges of x which are not there. The LULU decomposition extracts pulses which represent information which *is* present in the signal.

A smoother can be classified as low-pass, high-pass and bandpass which respectively refer to those which remove high-frequency noise, low frequency noise and noise within two specified frequencies, [56]

A nonlinear smoother is one which is not required to satisfy equations 2.2 and 2.3 in general. Some examples of nonlinear smoothers are:

- Recursive Filters (specifically exponential smoothing), [75]:

$$\big(T(x)\big)_n = \alpha\big(T(x)\big)_{n-1} + (1-\alpha)x_n.$$

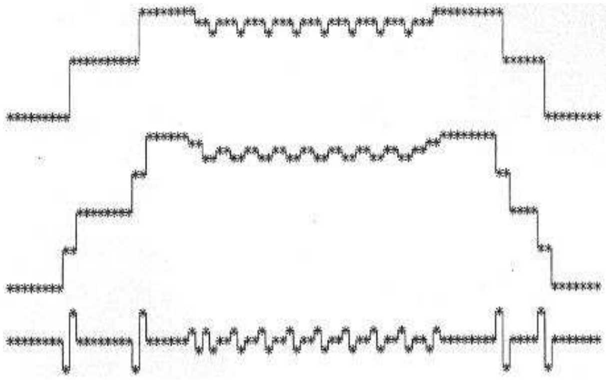


Figure 2.4. Figure 7.2 from [93]: Original sequence x (top) decomposed into Px (middle) and $(I-P)x$ (bottom).

- Reroughing, [122]: For a general nonlinear smoother, T we obtain the smoothed sequence $T(x)$ and the removed rougher part $(I-T)x$. In reroughing the rough part is smoothed i.e. $T(I-T)x$ and added to the smoothed sequence obtaining $T(x)+(T(I-T)x)$.

- Median smoother, [115]:

$$\big(M_n(x)\big)_i = median\{x_{i-n}, ..., x_i, ...x_{i+n}\},$$

as well as other order statistic smoothers. A number of modifications of the median smoothers have also been proposed, [115]: $M_2, M_1^2, 2M_n - M_n^2$, the recursive application of M_n: M_n^k , and the limiting median smoother

$$R(x) = \lim_{n\to\infty} M_n(x).$$

The weighted median filters have also been extensively investigated, [137]:

$$(M(x))_i = median\{w_i \diamond x_i, ..., w_n \diamond x_n\}$$

where $k \diamond x_i = \overbrace{x_i, ..., x_i}^{k\ times}$. In [137], another version of the recursive median is presented:

$$(R^*(x))_i = median\{(R^*(x))_{i-n}, ..., (R^*(x))_{i-1}, x_i, ..., x_{i+n}\}.$$

- '53H twice' smoother, [115]:

$$\begin{aligned}
\big(T_1(x)\big)_i &= median\{x_{i-2}, x_{i-1}, x_i, x_{i+1}, x_{i+2}\} \\
\big(T_2(x)\big)_i &= median\{\big(T_1(x)\big)_{i-1}, \big(T_1(x)\big)_i, \big(T_1(x)\big)_{i+1}\} \\
\big(T_3(x)\big)_i &= \frac{1}{4}\big(T_2(x)\big)_{i-1} + \frac{1}{2}\big(T_2(x)\big)_i + \frac{1}{4}\big(T_2(x)\big)_{i+1}
\end{aligned}$$

and then

$$\big(T(x)\big)_i = \big(T_3(x)\big)_i + \big(T_3(x - T_3(x))\big)_i.$$

- Cosine Bell of length v, [122]:

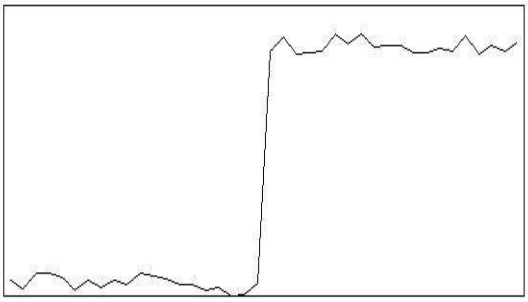
$$\big(T(x)\big)_{[(y+1)/2]+1-i} = 1 + cos\left(\frac{\pi_i}{v/2+1}\right) \text{ for } i = 0, 1, 2, ..., [v/2]$$

  where $[w]$ is the largest integer not exceeding w.

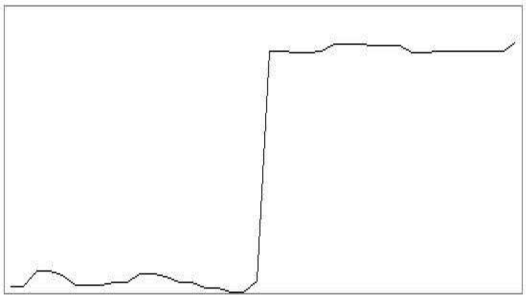
Specifically, our LULU operators L_n and U_n are nonlinear smoothers. They will be the focus of the rest of this chapter.

The strength of nonlinear smoothers over linear smoothers is that their application for signals and images is generally more powerful since most of the useful nonlinear smoothers are more robust (the output is not significantly affected by small changes in the input) and consistent (signal is mapped to signal and noise is mapped to noise) than linear smoothers. In an image or signal an edge is an impulse in the derivative i.e. a sharp change in the signal or image value in a certain direction. The edges and constant regions are important as they affect the image or signal quality and linear smoothers do not preserve these well, whereas nonlinear smoothers do, see Figure 2.5. Nonlinear smoothers also preserve image details and are insensitive to outliers, [75], which is termed **resistance**.

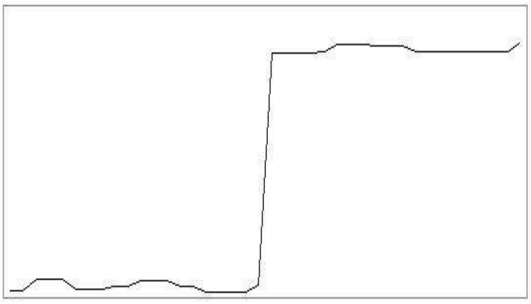
Linear smoothers aren't well suited to removing noise which arises from a long-tailed probability distribution, [122], which is characteristic when there are outliers present, (see Figure 2.6 for the effectiveness of the LULU smoothers), nor noise which is signal dependent, [26].



(a)



(b)

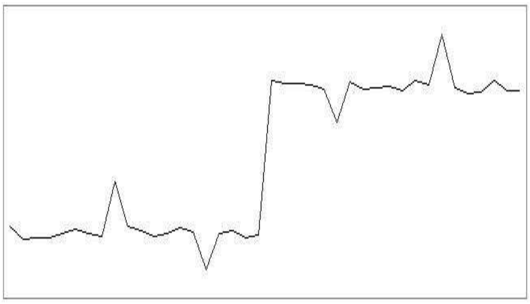


(c)

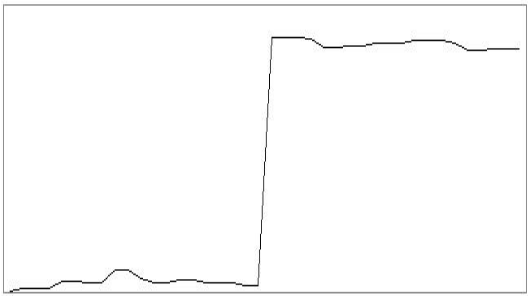
Figure 2.5. (a) Original Signal x with a Distinct Edge (b) $U_1 \circ L_1 x$ (c) $U_2 \circ L_2 \circ U_1 \circ L_1 x$

There are numerous applications in image processing to improve image quality by enhancing the edges and constant regions i.e. image enhancement, see [79], [117], [70], [50], [84], [103], [114] to name a few.

The theory and analysis of nonlinear smoothers has proven to be difficult up until LULU theory was developed, [122], [123]. In [75], a method is shown to develop a resistant nonlinear smoother which is almost linear. Here, Mallows shows that a nonlinear smoother is made up of a linear part and an orthogonal nonlinear residual part and measures its performance based on how linear it is i.e. how closely the operator resembles a linear operator,



(a)



(b)

Figure 2.6. (a) Original Signal x as in Figure 2.5 with isolated noise pulses (b) $U_1 \circ L_1 x$

with the mean squared error. His construction is quite strict on the allowed components though, bringing in statistical distributions for the various components. Mallows also proposes axioms for a smoother T. The operator E is called the **shift operator**. It is acts on sequence x as follows: $(Ex)_i = x_{i+1}$.

A1 T is stationary: $T(Ex) = E(Tx)$

A2 T is location invariant: $T(x + c) = Tx + c$

A3 T is centered: $T(0) = 0$ where 0 is the zero sequence.

A4 $\big(T(x)\big)_i$ depends only on finitely many components of x.

A5 $variance\big((T(x))_i\big)$ is finite

Axiom A1 relates to translation invariance in higher dimensions. In Axiom A4 he assumes $variance(x)$ is finite. These axioms are used as a starting point by Rohwer for developing the nonlinear theory of the LULU operators.

We will first provide some basic operator theory to set the background for the LULU operators, see [93]. Firstly we assume that a partial order exists for X.

Definition 1 *A relation T is a* **partial order** *on a set A if for $s, t \in A$,*

1. *T is reflexive: $(t, t) \in T$,*
2. *T is anti-symmetric: $(s, t) \in T$ and $(t, s) \in T$ implies $s = t$,*
3. *T is transitive: $(s, t) \in T$ and $(t, v) \in T$ implies $(s, v) \in T$.*

Definition 2 *The relation T is a* **total order** *if in addition to the properties in Definition 1, the following property also holds:*

$$\forall \; s, t \in A, \; \textit{either} \; (s, t) \in T \; \textit{or} \; (t, s) \in T.$$

This additional property for a total order is also known as **comparability** or the *Trichotomy Law*, [53].

Commonly, the partial order is denoted by '$\leq$'. Then $s \leq t$ refers to $(s, t) \in T$.

Definition 3 *A partially ordered set $\mathcal{L}$ is a* **lattice** *if any $\ell_1, \ell_2 \in \mathcal{L}$ admit a least upper bound $\ell_1 \vee \ell_2$ and a largest lower bound $\ell_1 \wedge \ell_2$. A lattice is* **complete** *if every subset of $\mathcal{L}$ has a least upper bound and a largest lower bound.*

It is well known that the set X which is under consideration, see (2.1), is a lattice with respect to the coordinate-wise defined partial order. More precisely, for $x = (x_n), y = (y_n) \in X$ we have $x \leq y \iff x_n \leq y_n \; \forall \; n \in \mathbb{Z}$.

Definition 4 *For two operators A and B on X, we have $A \leq B \iff Ax \leq Bx, \; \forall \; x \in X$.*

Definition 5 *An operator S is called* **syntone** *if $x \leq y$ implies $Sx \leq Sy$.*

This concept is also referred to as **increasing**, specifically in the field of mathematical morphology. We will expand more on this in Section 2.2.1. Below are some useful and easy to prove properties, from [93].

Theorem 6 *If A and B are two syntone operators then $A \circ B$ is also syntone.*

Theorem 7 *For operators A, B and C on X where A is syntone, $B \leq C$ implies $A \circ B \leq A \circ C$.*

We consider four criteria that are useful in the design and comparison of smoothers, [93], and in the sequel we denote the identity operator by I. The criteria are:

- EFFECTIVENESS: A smoother should separate a signal x into it's noise component $(I - P)x$ and its true signal Px.
- CONSISTENCY: Once the true signal and noise are extracted they should each be preserved i.e. $P(Px) = Px$ and $(I - P)((I - P)x)) = (I - P)x$. These two properties are referred to as *idempotent* and *co-idempotent* respectively. This means that the smoother separates x consistently.
- STABILITY: This criterion requires the smoother to be robust, that is, a small change in x should not result in a large change in the output, so that the signal is recovered well.
- EFFICIENCY: The computations should be economical as the desire is usually to have computations done in real time.

Based on these criteria and Mallows' axioms A1 - A5, now introduce the axioms for a smoother as well as a separator, developed by Rohwer.

Definition 8 Smoother Axioms *An operator P on X is a smoother if:*

1. $PE = EP$
2. $P(x + c) = P(x) + c$ *for each x and where $c \in X$ is a constant sequence.*
3. $P(\alpha x) = \alpha P(x)$ *for each x and scalar $\alpha \geq 0$.*

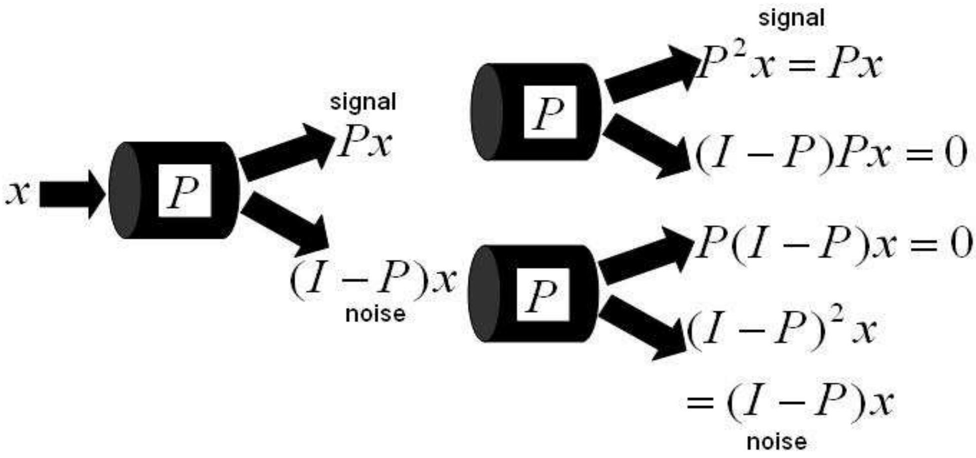


Figure 2.7. The Actions of a Separator

In [75] axiom 3 above is different in that $\alpha < 0$ is also allowed. The restriction introduced in Definition 8 is necessary as we have $L_n(-x) \neq -L_n(x)$ but rather that $L_n(-x) = -U_n(x)$ so that L_n and U_n are dual operators.

Definition 9 Separator Axioms *An operator P is a separator if it satisfies the smoother axioms as well as the following two axioms:*

4. $P^2 = P$ *(Idempotence)*
5. $(I - P)^2 = (I - P)$ *(Co-Idempotence)*

Figure 2.7 illustrates the effect of a separator. We have the following result, which can be seen from Figure 2.7.

Lemma 10 *An operator P is co-idempotent if and only if $P(I - P) = 0$.*

In Lemma 10 the operator 0 is the one which maps x onto the 0 sequence.

An important result is that any separator has only two eigenvalues, namely 0 and 1. Thus any sequence has the form $x = \alpha_0 e_0 + \alpha_1 e_1$ where e_0 and e_1 are the eigenvectors for the two eigenvalues. So we see that any sequence can be split into two eigensequences namely 'noise' $(\alpha_0 e_0)$ and 'signal' $(\alpha_1 e_1)$. This is called the Fundamental Separator Theorem, [95]. Note that an idempotent linear operator (a projection, [43]) is automatically co-idempotent, thus in linear signal processing there is no need for the concepts of a separator and co-idempotence. Since the operators L_n and U_n are nonlinear the opposite is true for them.

A simple class of examples of smoothers are those which select the output value via rank or order. They are applied to a n-window of subscripts from the sequence i.e. looking at the subscripts $\{i-n, i-n+1, ..., i, ..., i+n-1, i+n\}$, thus we look at the set of values $W_{i,n} = \{x_{i-n}, ..., x_i, ..., x_{i+n}\}$ from the sequence x. By using only this window of values the operator becomes a local operator because it only acts on values in the sequence close by.

Definition 11 *An operator S is a* **rank order selector** *if S acts on x such that $(Sx)_i = r(W_{i,n}) \in W_{i,n}$, where $r(W_{i,n})$ is a fixed pre-chosen rank position in $W_{i,n}$. Operator S is an* **rank based selector** *if it also makes use of the ranked values in $W_{i,n}$ and other information to select $r(W_{i,n})$.*

Theorem 12 *A rank order selector satisfies the smoother axioms.*

Proof
Axiom 1: $SE = ES$

$$(SEx)_i = (Sx)_{i+1} = r(W_{i+1,n}) \in W_{i+1,n}$$

and

$$(ESx)_i = E(r(W_{i,n})) = r(W_{i+1,n}) \in W_{i+1,n}.$$

Axiom 2: $S(x+c) = S(x) + c$

$$(S(x+c))_i = r(W_{i,n} + \{c_{i-n}, ..., c_i, ..., c_{i+n}\}) = r(W_{i,n}) + c_j$$

where $j \in \{i-n, ..., i, i+n\}$ and

$$(Sx)_i + c_i = r(W_{i,n}) + c_i = r(W_{i,n}) + c_j$$

since c is constant.
Axiom 3: $S(\alpha x) = \alpha S(x)$

$$(S(\alpha x))_i = r(\alpha W_{i,n}) = \alpha r(W_{i,n})$$

and

$$\alpha(Sx)_i = \alpha r(W_{i,n}).$$

■

An order selector is also syntone. For a proof see [93].

Due to the smoother axioms we have that the smoothers form a semi-group [95] with respect to composition. This will be dealt with in more detail later in this chapter. For now we will introduce some semi-group theory, [95].

Definition 13 *A* **semigroup** $(H, *)$ *is a set* H *endowed with an associative binary operation* $*$*, i.e.* $(a * b) * c = a * (b * c)$ *for all* $a, b, c \in H$*.*

We simplify the notation $a * b$ to ab. A semigroup H is a **band** if $Id(H) := \{a \in H : a^2 = a\} = H$. A **left zero semigroup** H is one for which $a * b = a$ for all $a, b \in H$ and a **right zero semigroup** is one for which $a * b = b$ for all $a, b \in H$. The LULU semigroup discussed later on is a band but is not a left or right zero semigroup.

Definition 14 *An* **ordered semigroup** *is a semigroup endowed with a partial order,* $\leq$*, which is compatible with the semigroup operation i.e. for all* $a, b, c \in H$*,* $a \leq b \Rightarrow (ca \leq cb$ *and* $ac \leq bc)$*.*

For an element $b \in H$, where H is an ordered semigroup, b is the **inverse** of $a \in H$ if $a = aba$ and $b = bab$. An **identity** element for a semigroup is an element, 1, such that $1a = a1 = a$ for all $a \in H$.

Theorem 15 *For an ordered semigroup* $(H, *, \leq)$*, any* $g \leq f$ *in* $Id(H)$ *generates an at most six-element semigroup.*

Proof
Since g and f belong to $Id(H)$ we have $g^2 = g$ and $f^2 = f$. Also, $fg \leq f^2 = f$ and $gf \geq g^2 = g$ since $f \leq g$. The idempotence of fg is proven as follows:

$$\begin{aligned} & (fg)(fg) \leq f(fg) = fg \\ \text{and} \quad & f(gf)g \geq fgg = fg \end{aligned}$$

so that $(fg)^2 = fg$. Similarly, $(gf)^2 = gf$. The idempotence of fgf is proven as follows:

$$\begin{aligned} & (fgf)(fgf) = (fg)(ffgf) \leq f(ffgf) = fgf \\ \text{and} \quad & (fgf)(fgf) = f(gf)(fg) \geq fg(gf) = fgf \end{aligned}$$

so that $(fgf)^2 = fgf$. Similarly, $(gfg)^2 = gfg$. Thus any composition of f, g, fg, gf, fgf, gfg will reduce to one of these original elements, producing an at most 6-element semigroup. ■

Criterion 6.6 in [78] provides the conditions for this 6-element semigroup to reduce to a 4-element semigroup. This will occur if we have $fgf = gf$ and $gfg = fg$, or $gf \geq fg$.

We will refer to this theorem later on again when we introduce the LULU semigroup.

Theorem 16 *The set of smoothers is a semi-group with respect to composition.*

Definition 17 *For two posets (partially ordered sets) P and Q, an increasing map $\delta : P \to Q$ is* **residuated** *if for each $y \in Q$ there exists a biggest element in P, say $\epsilon(y)$, that is mapped below y. Then ϵ is a map from Q to P, is also increasing and for each $x \in P$ there is a smallest element in Q, namely $\delta(x)$ which is mapped above x i.e.*

$$(\forall\ x \in P)(\forall\ y \in Q)\ \delta(y) \leq y \iff x \leq \epsilon(y).$$

We call ϵ the **residual** *of δ. The pair (δ, ϵ) is then a* **Galois connection** *between P and Q, [95].*

Clearly the roles of δ and ϵ is a Galois connection are interchangeable.

Definition 18 *An operator A is called* **extensive** *if $I \leq A$ and* **anti-extensive** *if $A \leq I$.*

Definition 19 *An operator (set of operators) is called a* **closing** *if it (they) is 1) increasing (syntone), 2) extensive, and 3) idempotent.*

Definition 20 *An operator (set of operators) is called an* **opening** *if it (they) is 1) increasing (syntone), 2) anti-extensive, and 3) idempotent.*

We have the following result regarding a Galois connection, [95].

Theorem 21 *If* (δ, ϵ) *is a Galois connection between* P *and* Q *then,*

1. $\delta\epsilon\delta = \delta$, $\epsilon\delta\epsilon = \epsilon$, $\delta\epsilon \leq I$ *is an opening on* P *and* $\epsilon\delta \geq I$ *is a closing on* Q.
2. *Any product of equally many* δ*'s and* ϵ*'s is idempotent.*

An example of a Galois connection is the pair erosion and dilation, see Section 2.2.1. We will define some specific operators before we continue to the LULU operators.

Definition 22 *Operators* A *and* B *are called* **dual** *operators if* $AN = NB$ *where* $Nx = -x$ *is the negation operator. An operator is* **self-dual** *if* $AN = NA$.

Definition 23 *The median operator is given by*

$$(M_n x)_i = median\{x_{i-n}, ..., x_i, ..., x_{i+n}\},$$

for any nonnegative integer n.

Theorem 24 *The* M_n *operator is a smoother.*
Proof
$M_n E = EM_n$**:**

$$\begin{aligned}(M_n Ex)_i &= M_n(Ex)_i \\ &= median\{x_{i-n+1}, ..., x_{i+1}, ..., x_{i+n+1}\} \\ &= (M_n x)_{i+1}\end{aligned}$$

and

$$\begin{aligned}(EM_n x)_i &= E(median\{x_{i-n}, ..., x_i, ..., x_{i+n}\}) \\ &= median\{x_{i-n+1}, ..., x_{i+1}, ..., x_{i+n+1}\} \\ &= (M_n x)_{i+1}.\end{aligned}$$

$M_n(x+c) = M_n(x) + c$**:**

$$\begin{aligned}(M_n(x+c))_i &= median\{x_{i-n} + c_{i-n}, ..., x_i + c_i, ..., x_{i+n} + c_{i+n}\} \\ &= median\{x_{i-n}, ..., x_i, ..., x_{i+n}\} + c_i \\ &= (M_n x)_i + c_i\end{aligned}$$

$M_n(\alpha x) = \alpha(M_n x)$**:**

$$\begin{aligned}(M_n(\alpha x))_i &= median\{\alpha x_{i-n}, ..., \alpha x_i, ..., \alpha x_{i+n}\} \\ &= \alpha median\{x_{i-n}, ..., x_i, ..., x_{i+n}\} \\ &= \alpha(M_n x)_i\end{aligned}$$

■

Definition 25 *A sequence x is n-**monotone** if $\{x_i, x_{i+1}, ..., x_{i+n+1}\}$ is monotone for each i.*

Definition 26 *$\mathcal{M}_n$ is the set of all sequences x which are n-monotone.*

This idea of local monotonicity has been a concept of smoothness in real analysis for a while, [98]. Obviously any sequence is 0-monotone i.e. belongs to $\mathcal{M}_0$, and then $\mathcal{M}_0 \supset \mathcal{M}_1 \supset \mathcal{M}_2 \supset ....$

The popularity of the median smoother is due to its good performance and logical simplicity, [47] [55] [7] [82] [60] [130] [139]. In addition a lot of theoretical background has been developed, [115] [7] [8] [46] [57] [116] [80] [11] [20] [81] [138] [42] [5] [68] [54] [137] [6]. It is also stable with respect to large impulse noise, since it does not distort the output excessively and has very good edge preservation, [56]. The output variance of the smoothed sequence is used to measure the capability of M_n under various white noise and non-white noise assumptions in [56] and the good edge preservation is shown using the mean squared error.

Definition 27 *The **root** of a smoother P is a sequence x which remains unchanged after smoothing, that is $Px = x$, [56].*

The roots of the median smoother M_n are those sequences x which are n-monotone i.e. $x \in \mathcal{M}_n$. Any non-root sequence of length L will become a root after a maximum of $\frac{1}{2}(L-2)$ passes of the median smoother with any window size, [45]. The window should not be too large though as important details may then be lost. This also means that the median smoother is not idempotent and is thus less computationally efficient than the LULU smoothers.

However, M_n is statistically unbiased [95] but it is not a separator. Thus M_n is iterated until convergence [115] to obtain M_n^∞, and in fact for $x \in \mathcal{M}_{n-1}$, $M_n x$ does not generally map into $\mathcal{M}_n$ although M_n^∞ does, [95]. It also has a lack of theory for the important aspects i.e. no Parseval identity or resolution level structure, and it is less economical computationally since to obtain the median of numbers is more complex than a max or min operation, [95].

We are now ready to introduce the LULU smoothers.

2.2 The LULU Operators in One Dimension

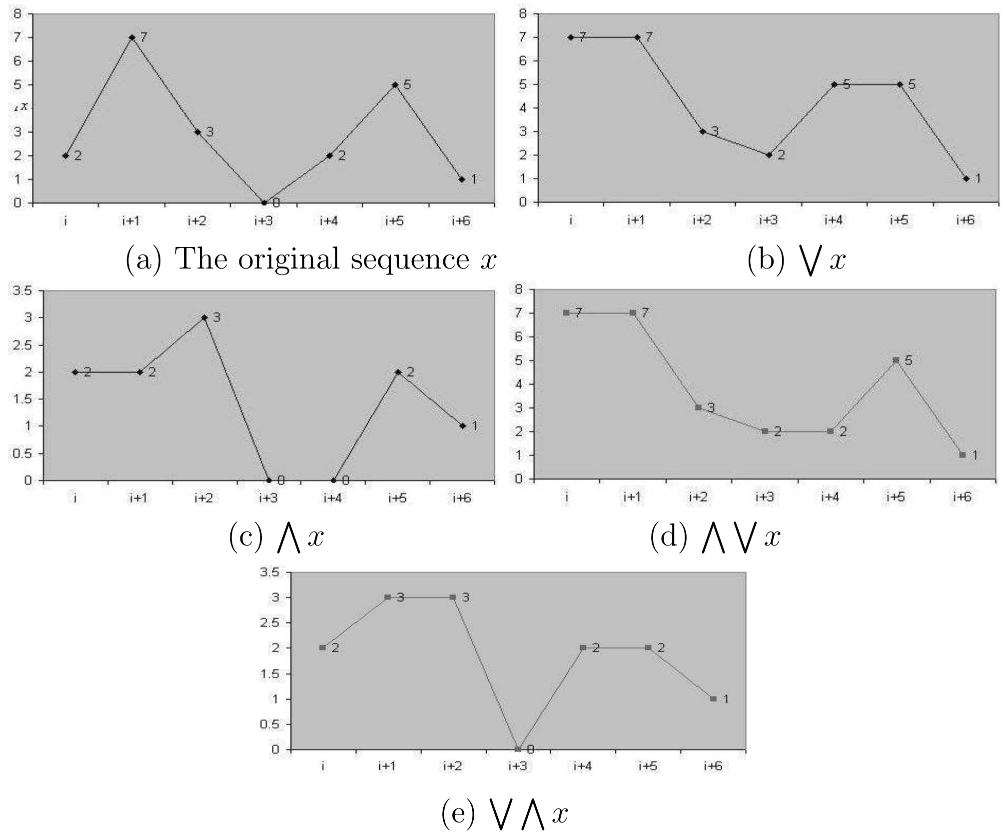
We first define the 'atoms' which make up the LULU smoothers.

Definition 28 *The operators* $\bigvee$ *and* $\bigwedge$ *are defined as*

$$\bigvee x = \max\{x_i, x_{i+1}\}$$

$$\bigwedge x = \min\{x_{i-1}, x_i\}.$$

We can relate $\bigvee$ to a union of sets because the union of a collection of sets represents the maximum set covered by the collection. Similarly, we can relate $\bigwedge$ to an intersection of sets because the intersection of sets results in the minimum set covered by the collection.



(a) The original sequence x

(b) $\bigvee x$

(c) $\bigwedge x$

(d) $\bigwedge \bigvee x$

(e) $\bigvee \bigwedge x$

Figure 2.8. Action of $\bigvee$ and $\bigwedge$

In fact the operators are more useful when applied together i.e. by applying $\bigvee\bigwedge$ or $\bigwedge\bigvee$, since $\bigvee$ will remove downward pulses and $\bigwedge$ will remove upward pulses. To illustrate this we use an example, see Figure 2.8. Consider the sequence $\{2, 7, 3, 0, 2, 5, 1\}$. (In reality we are considering the bi-infinite sequence $\{..., 0, 0, 0, 2, 7, 3, 0, 2, 5, 1, 0, 0, ...\} \in \ell_1$ by appending 0's to the left and right of the nonzero values.) The sequence is displayed in Figure 2.8(a). In Figure 2.8(b) we can see what happens when we apply $\bigvee$ to x. Notice how the sequence seems smoothed already. Compare Figure 2.8(b) to Figure 2.8(c). We now apply the opposite operator to the results in Figure 2.8(b) and Figure 2.8(c) and obtain $\bigwedge\bigvee x$ and $\bigvee\bigwedge x$ respectively, see Figures 2.8(d) and (e). Notice that $\bigwedge\bigvee x$ and $\bigvee\bigwedge x$ produce different results and that $\bigvee\bigwedge x \leq \bigwedge\bigvee x$.

We have the following results regarding these 'max' and 'min' operators (for the proofs see [93]). Note that $\bigvee^n = \max\{x_i, ..., x_{i+n}\}$ and similarly $\bigwedge^n = \min\{x_{i-n}, ..., x_i\}$.

RESULT 1 $\bigvee$ and $\bigwedge$ are syntone operators.

RESULT 2 $\bigwedge \leq I \leq \bigvee$ and $\bigvee\bigwedge \leq I \leq \bigwedge\bigvee$.

Result 2 can be seen in Figure 2.8.

RESULT 3 $\bigvee^{n+1}\bigwedge^{n+1} \leq \bigvee^n\bigwedge^n$ and $\bigwedge^{n+1}\bigvee^{n+1} \geq \bigwedge^n\bigvee^n$.

RESULT 4 For $n \leq m$, we have $\bigwedge^n\bigvee^n\bigwedge^m\bigvee^m = \bigwedge^m\bigvee^m$ and $\bigvee^n\bigwedge^n\bigvee^m\bigwedge^m = \bigvee^m\bigwedge^m$.

Result 4 is referred to in [93] as the **First Swallowing Theorem** as the repeated application of smaller number of operators will have no effect. If $n = m$ then it is called the **Idempotence Theorem** since then we have that $P^2 = P$ where $P = \bigwedge^m\bigvee^m$ or $\bigvee^m\bigwedge^m$.

RESULT 5 $\bigvee^n\bigwedge^n \leq \bigvee^n\bigwedge^n\left(\bigwedge^n\bigvee^n\right)\bigvee^n\bigwedge^n \leq \bigvee^n\bigwedge^n\left(\bigwedge^n\bigvee^n\right)$ and $\bigwedge^n\bigvee^n \leq \bigwedge^n\bigvee^n\left(\bigvee^n\bigwedge^n\right)\bigwedge^n\bigvee^n \leq \bigwedge^n\bigvee^n\left(\bigvee^n\bigwedge^n\right)$.

RESULT 6 $\left[\bigvee^n\bigwedge^n\left(\bigwedge^n\bigvee^n\right)\right]^2 = \bigvee^n\bigwedge^n\left(\bigwedge^n\bigvee^n\right)$ and $\left[\bigwedge^n\bigvee^n\left(\bigvee^n\bigwedge^n\right)\right]^2 = \bigwedge^n\bigvee^n\left(\bigvee^n\bigwedge^n\right)$.

Result 6 is called the **Second Idempotence Theorem**.

RESULT 7 $(\bigvee^n \bigwedge^n)(\bigwedge^n \bigvee^n)(\bigvee^n \bigwedge^n) \leq (\bigwedge^n \bigvee^n)(\bigvee^n \bigwedge^n)(\bigwedge^n \bigvee^n)$.

RESULT 8 $\left[(\bigvee^n \bigwedge^n)(\bigwedge^n \bigvee^n)(\bigvee^n \bigwedge^n)\right]^2 = (\bigvee^n \bigwedge^n)(\bigwedge^n \bigvee^n)(\bigvee^n \bigwedge^n)$
and
$\left[(\bigwedge^n \bigvee^n)(\bigvee^n \bigwedge)(\bigwedge^n \bigvee^n)\right]^2 = (\bigwedge^n \bigvee^n)(\bigvee^n \bigwedge^n)(\bigwedge^n \bigvee^n)$.

This result is called the **Third Idempotence Theorem**.

Definition 29 *We define the LULU operators as those finite compositions of the operators*

$$L_n = \bigvee^n \bigwedge^n \text{ and } U_n = \bigwedge^n \bigvee^n,$$

where $\bigvee^n = max\{x_i, x_{i+1}, ..., x_{i+n}\}$ *and* $\bigwedge^n = min\{x_{i-n}, x_{i-n+1}, ..., x_i\}$. *For example,* $L_2 \circ U_3 \circ L_1 \circ U_4$ *is a LULU operator.*

The operators L_n and U_n are applied as compositions because U_n smooths from below and L_n from above and thus are biased if not used together. The LULU operators are nonlinear operators. Consider the following counter-example to their linearity: For a sequence x such that $L_n x \neq U_n x$, let $y = -x$. Then $L_n(x+y) = L_n(0) = 0$ but $L_n x + L_n y = L_n x - U_n(-y) = L_n x - U_n x \neq 0$. This also implies that all compositions of the LULU operators will be nonlinear. The LULU operators are however smoothers since $U_n \circ L_n$ and $L_n \circ U_n$ are smoothers.

Theorem 30 $U_n \circ L_n$ *and* $L_n \circ U_n$ *are smoothers.*

Results 1 to 8 then imply the following results for L_n and U_n. Note that we use the notation $A \circ B = AB$ in what follows.

1. L_n and U_n are **syntone**: $L_{n+1} \leq L_n \leq L_0 = I = U_0 \leq U_n \leq U_{n+1}$ (by results 1,2 and 3).

2. $L_n L_m = L_m$ and $U_n U_m = U_m$ for $m \geq n$ (by result 4)

3. $(L_n)^2 = L_n$ and $(U_n)^2 = U_n$ (by result 4 with $m = n$). Thus L_n and U_n are idempotent.

4. $L_n \leq L_nU_nL_n \leq L_nU_n$ and $U_nL_n \leq U_nL_nU_n \leq U_n$ (by result 5).

5. $\left(L_nU_n\right)^2 = L_nU_n$ and $\left(U_nL_n\right)^2 = U_nL_n$ (by result 6). Thus the composition operators L_nU_n and U_nL_n are both idempotent.

6. $L_nU_nL_n \leq U_nL_nU_n$ (by result 7).

7. $\left(L_nU_nL_n\right)^2 = L_nU_nL_n$ and $\left(U_nL_nU_n\right)^2 = U_nL_nU_n$ (by result 8). Thus we also have that $L_nU_nL_n$ and $U_nL_nU_n$ are idempotent operator compositions.

Because the LULU operators are nonlinear we do not have the projection theorem available for a near-best approximation in the range in any norm via the Lebesgue inequality. However a projection has only two eigenvalues, 0 and 1, which simplistically correspond to the 'signal' and 'noise', and we have shown in the previous section this holds for a separator as well. Thus separators are nearly projections in this sense. The LULU operators can satisfy the same objectives as those when making use of a projection, [95]:

1. Our operators have useful ranges which help with the choice of operators. This range is that of $\mathcal{M}_n$.

2. The mappings onto this range are good approximations since a Lebesgue-type inequality exists. Recall that for a subspace S of a normed space X, the Lebesgue inequality for a linear mapping P into S such that $Ps = s \ \forall \ s \in S$, is

$$\|Px - x\| \leq \left(\|P\| + 1\right)\|x - s\| \ \forall x \in X, s \in S.$$

   In order to get a similar inequality for the LULU operators we associate $\mathcal{M}_n$ with S and use a Lipchitz constant get around the linearity. We can prove the following inequality, [93],

$$\|Px - x\|_p \leq \left(1 + (2n+1)^{1/p}\right)\|s - x\|_p,$$

   for $P = L_n \circ U_n$ or $U_n \circ L_n$ and $s \in \mathcal{M}_n$.

3. A preservation law exists, namely that of total variation: $T(x) = T(Px) + T(Px - x)$ where P is either $L_n \circ U_n$ or $U_n \circ L_n$, [93].

4. The operators act nearly linearly, [93, Theorem 6.20].

We also have the following further results for the LULU operators L_n and U_n.

RESULT 9 L_n and U_n are **dual** operators i.e. $L_n \circ N = N \circ U_n$. Note then that neither L_n nor U_n are self-dual. We also have that $L_n \circ U_n$ and $U_n \circ L_n$ are duals since

$$L_n \circ (U_n \circ N) = (L_n \circ N) \circ L_n = (N \circ U_n) \circ L_n,$$

and $L_n \circ U_n \circ L_n$ and $U_n \circ L_n \circ U_n$ are duals since

$$L_n \circ U_n \circ L_n \circ N = L_n \circ U_n \circ (N \circ U_n) = L_n \circ (N \circ L_n) \circ U_n = N \circ U_n \circ L_n \circ U_n.$$

RESULT 10 U_n is an **extensive** operator and L_n is an **anti-extensive** operator, see Definition 18.

RESULT 11 U_n is a **closing** and L_n is an **opening**, see Definitions 19 and 20.

RESULT 12 For each n, $L_n x = x = U_n x$ iff x is n-monotone and $(L_n \circ U_n)x = x = (U_n \circ L_n)x$ iff x is n-monotone.

RESULT 13 For each n, $L_n \circ U_n x$ and $U_n \circ L_n x$ are n-monotone.

RESULT 14 For each n, we have that $U_n \circ L_n \leq M_n \leq L_n \circ U_n$. This is an extremely important result as we then have that

$$L_n \leq U_n \circ L_n \leq L_n \circ U_n \leq U_n. \tag{2.5}$$

This provides the *fully* ordered (w.r.t. Definition 4) 4-element LULU semigroup, namely $\{L_n, U_n, L_n \circ U_n, U_n \circ L_n\}$. Note that although M_n belongs to the LULU-interval (see result 15 below) it does not replicate all of the nice properties of the LULU operators. The next result provides this reduction from the 6-element semigroup, see Lemma 15, to the 4-element semigroup displayed in Table 2.1.

RESULT 15 $U_n \circ L_n \circ U_n = L_n \circ U_n$ and $L_n \circ U_n \circ L_n = U_n \circ L_n$. This means that any composition of the operators L_n and U_n results in one of the following operators $\{L_n, U_n, L_n \circ U_n, U_n \circ L_n\}$ so that the set is closed with respect to composition with the ordering indicated with '$\leq$'. The 4-element LULU semigroup with respect to composition is displayed in Table 2.1. The interval $[U_n \circ L_n, L_n \circ U_n]$ is referred to as the LULU-interval and any smoothers that fall within this interval are called **n-LULU similar.**

	L_n	U_n	$U_n \circ L_n$	$L_n \circ U_n$
L_n	L_n	$L_n \circ U_n$	$U_n \circ L_n$	$L_n \circ U_n$
U_n	$U_n \circ L_n$	U_n	$U_n \circ L_n$	$L_n \circ U_n$
$U_n \circ L_n$	$U_n \circ L_n$	$L_n \circ U_n$	$U_n \circ L_n$	$L_n \circ U_n$
$L_n \circ U_n$	$U_n \circ L_n$	$L_n \circ U_n$	$U_n \circ L_n$	$L_n \circ U_n$

Table 2.1. The LULU semi-group

Thus we see also that there is a bias between the operators $U_n \circ L_n$ and $L_n \circ U_n$ which increases as n increases. This bias will be significant when a piece of Nyquist frequency is present, [95]. The Nyquist frequency is half the sampling frequency of the signal concerned. It should be greater than than the maximum frequency component in the signal in order to avoid aliasing. Aliasing refers to not being able to reconstruct the original signal due to the sampling, [52]. The semigroup $\{L_n, U_n, L_n \circ U_n, U_n \circ L_n\}$ is a band as its elements are all idempotent. See [73] for a full discussion on the reduction from a non-ordered 6 element to the fully ordered 4 element semigroup. We refer back to the inequality $U_n \circ L_n \leq M_n \leq L_n \circ U_n$ in Result 14. The median operator is thus LULU similar. Due to it's spurious roots however, it has associated odd behavior, [95].

<u>RESULT 16</u> $U_n, L_n, U_n \circ L_n$ and $L_n \circ U_n$ are co-idempotent.

Definition 31 *We define the operator C_n,*

$$C_1 = L_1 \circ U_1, \quad C_{n+1} = L_{n+1} \circ U_{n+1} \circ C_n$$

and the operator F_n as

$$F_1 = U_1 \circ L_1, \quad F_{n+1} = U_{n+1} \circ L_{n+1} \circ F_n.$$

The 'C' and 'F' refer to ***'ceiling'*** *and* ***'floor'*** *respectively.*

Definition 31 provides a better smoother than when using $L_n \circ U_n$ or $U_n \circ L_n$ for a fixed n, as we successively remove the pulses, see Result 17 which follows. Figure 2.9 shows this property. In Figure 2.9(c) we see the application of F_2. Compare this with the non-recursive application in Figure 2.9(d). It can clearly be seen that Figure 2.9(c) is the smoother result.

In [56], [38] and [93] we also find the following smoothers derived from L_n and U_n,

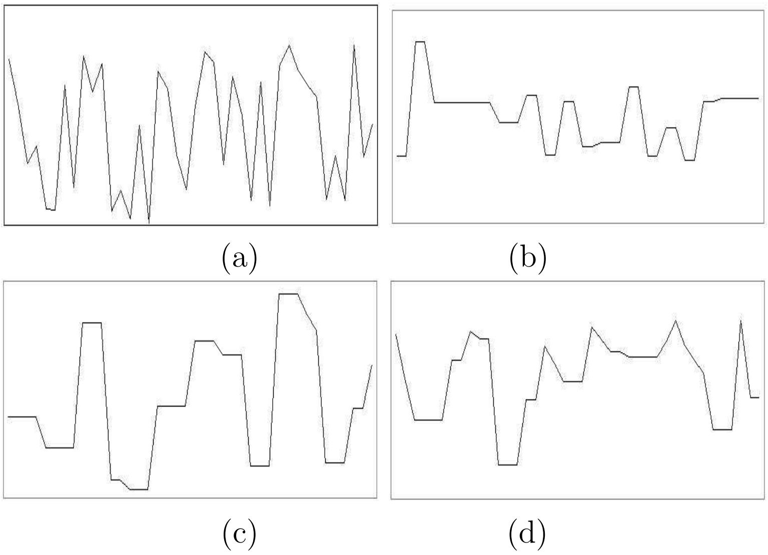


Figure 2.9. (a) Original signal x (b) $U_1 \circ L_1 x$ (c) $F_2 x = U_2 \circ L_2 \circ U_1 \circ L_1 x$ (d) $U_2 \circ L_2 x$

1. $G_n = \frac{1}{2}[L_n \circ U_n + U_n \circ L_n]$

2. G_n^∞ is obtained by repeated application of G_n until convergence.

3. $$(A_n x)_i = \begin{cases} x_i & if\ x_i \in [(U_n \circ L_n x)_i, (L_n \circ U_n x)_i] \\ (G_n x)_i & otherwise \end{cases}$$

4. A Winsorised smoother:

$$W_n^*(x)_i = \begin{cases} x_i & if\ x_i \in [(U_n \circ L_n x)_i, (L_n \circ U_n x)_i] \\ (U_n \circ L_n x)_i & if\ x_i < (U_n \circ L_n x)_i \\ (L_n \circ U_n x)_i & if\ x_i > (L_n \circ U_n x)_i \end{cases}$$

Then we have $W_n^* \in [U_n \circ L_n, L_n \circ U_n]$.

5. And it's compound smoother:

$$(W_n x)_i = \begin{cases} (W_{n-1} x)_i & if\ (W_{n-1} x)_i \in [(U_n \circ L_n x)_i, (L_n \circ U_n x)_i] \\ (U_n \circ L_n x)_i & if\ (W_{n-1} x)_i < (U_n \circ L_n x)_i \\ (L_n \circ U_n x)_i & if\ (W_{n-1} x)_i > (L_n \circ U_n)_i \end{cases}$$

with $W_0 = I$.

6. Another Winsorised smoother:

$$B_n^*(x)_i = \begin{cases} x_i & if\ x_i \in [(F_n x)_i, (C_n x)_i] \\ (F_n x)_i & if\ x_i < (F_n x)_i \\ (C_n x)_i & if\ x_i > (C_n x)_i \end{cases}$$

7. And it's related compound smoother:

$$B_n(x)_i = \begin{cases} (B_{n-1}x)_i & if\ (B_{n-1}x)_i \in [(F_n x)_i, (C_n x)_i] \\ (F_n x)_i & if\ (B_{n-1}x)_i < (F_n x)_i \\ (C_n x)_i & if\ (B_{n-1}x)_i > (C_n x)_i \end{cases}$$

   where $B_0 = I$.

8. The strange operator $Q_n = U_n + L_n - I \in [L_n, U_n]$.

9. The alternating bias smoothers:

$$Z_{n+1}^- = \begin{cases} L_{n+1} \circ U_{n+1} Z_n^- & if\ n\ is\ even \\ U_{n+1} \circ L_{n+1} Z_n^- & if\ n\ is\ odd \end{cases}$$

$$Z_{n+1}^+ = \begin{cases} L_{n+1} \circ U_{n+1} Z_n^+ & if\ n\ is\ odd \\ U_{n+1} \circ L_{n+1} Z_n^+ & if\ n\ is\ even \end{cases}$$

   These are used to balance the bias present between $L_n \circ U_n$ and $U_n \circ L_n$.

RESULT 17 For each n, $U_n \circ L_n \leq F_n \leq C_n \leq L_n \circ U_n$. This means that by using F_n or C_n we reduce the ambiguity present when using only L_n and U_n.

RESULT 18 $C_m \circ C_k = C_n$ and $F_m \circ F_k = F_n$ where $n = max\{m, k\}$.

RESULT 19 C_n and F_n are co-idempotent as well as idempotent.

RESULT 20 $L_n \leq G_n \leq U_n$, $G_n \circ L_n = U_n \circ L_n$, $G_n \circ U_n = L_n \circ U_n$, $(L_n \circ G_n)^2 = (L_n \circ G_n)^3$ and $(U_n \circ G_n)^2 = (U_n \circ G_n)^3$.

RESULT 21 $F_n \leq B_n \leq C_n$, $U_n \circ L_n \leq F_n \leq B_n \leq C_n \leq L_n \circ U_n$.

RESULT 22 W_n, W_n^*, B_n, B_n^* are all idempotent and co-idempotent.

RESULT 23 For each n,

- $L_n x = U_n x = x$
- $L_n \circ U_n x = U_n \circ L_n x = x$

- $W_n^* x = x$
- $W_n x = x$
- $C_n x = F_n x = x$
- $B_n^* x = x$
- $B_n x = x$

if and only if x is n-monotone.

Let us re-look at the LULU operators and the criteria mentioned earlier.

Let P represent a LULU operator.

- **EFFECTIVENESS** The LULU operators are required to separate x into the true signal, Px, and the noise $(I - P)x$, thus the output Px will contain less noise. From Result 13 we know that P maps into $\mathcal{M}_n$.
- **EFFICIENCY** The computations should be economical. We will introduce the Roadmaker's algorithm in Section 2.4 which greatly simplifies the computations.
- **CONSISTENCY** The LULU operators in the LULU semi-group are idempotent and co-idempotent, from results 4, 6 and 16. The co-idempotency of longer compositions than those in the semi-group remains an open problem, but the idempotent of all LULU operators holds, [95].
- **STABILITY** The stability of the LULU operators is shown in [93] since they have a Lipschitz constant.

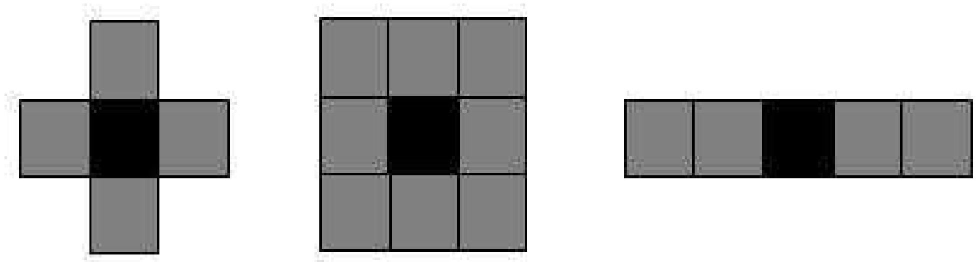
2.2.1 The LULU Operators as Morphological Filters

Morphological filters belong to the field of Mathematical Morphology. This field has its origins with Matheron while he was investigating porous media and Serra while looking into petrography of iron ore in 1964. It is a nonlinear branch of signal processing and is basically the application of set theory to image analysis, [31]. The basic morphological operators are the erosion and dilation, [104]. We first give the definitions based on binary sets. A **structuring element** is simply a set used to probe the set we wish to investigate. See Figure 2.10 for examples of some structuring elements.

Definition 32 *The* **erosion** *of a set X is the locus of centers x of structuring element B_x that are included in the set X. Here B_x is the translation of the structuring element B so that it is centered at x. So the erosion is given by*

$$\epsilon_B(X) = \bigcap_{b \in B} X_{-b} = \{x : B_x \subset X\}.$$

Here $x + b \in X \iff x \in X_{-b}$.



(a) (b) (c)

Figure 2.10. (a) A cross structuring element (4-connectivity) (b) A square structuring element (8-connectivity) (c) A line structuring element

Definition 33 *The* **dilation** *of a set X is the locus of centers x of structuring element B_x which hit the set X. So the dilation is given by*

$$\delta_B(X) = (\bigcap_{b \in B} (X_{-b})^c)^c = \{x : B_x \bigcap X \neq \emptyset\}$$

The morphological opening and closing are then the compositions of these two basic operators, [104].

Definition 34 *A* **morphological opening** *is given by*

$$\gamma_B(X) = \delta_{\check{B}} \circ \epsilon_B(X) = \bigcup_{B_y \subset X} B_y.$$

This is the domain swept out by all the translates of B which are included in the grains of X.

Definition 35 *A* **morphological closing** *is given by*

$$\rho_B(X) = \epsilon_{\check{B}} \circ \delta_B(X) = \Big(\bigcup_{B_y \subset X^c} B_y \Big)^c.$$

A point z will belong to the closing if all the translates of B, B_y, containing z hit X.

The morphological opening and closing are also algebraic openings and closings (see Definitions 19 and 20).

Definition 36 *An operator Φ is an (algebraic)* **opening** *if it is*

(i) Anti-extensive: $\Phi(X) \subset X \forall X$

(ii) Increasing: $X \subset Y \Rightarrow \Phi(X) \subset \Phi(Y)$

(iii) Idempotent: $\Phi(\Phi(X)) = \Phi(X)$

Definition 37 *An operator Φ is an (algebraic)* **closing** *if it is*

(i) Extensive: $X \subset \Phi(X) \forall X$

(ii) Increasing: $X \subset Y \Rightarrow \Phi(X) \subset \Phi(Y)$

(iii) Idempotent: $\Phi(\Phi(X)) = \Phi(X)$

Definition 38 *A* **morphological filter** *is an increasing, idempotent morphological operator.*

A morphological opening and closing are thus also morphological filters.

For grey-scale operations we have the following equivalent formulas for the morphological opening and closing, since $\epsilon_B(f)(x) = \min_{b \in B} f(x+b)$ and $\delta_B(f)(x) = \max_{b \in B} f(x+b)$.

$$\gamma_B(f)(x) = \max_{y \in \check{B}_x} \min_{z \in B_y} f(z). \tag{2.6}$$

$$\rho_B(f)(x) = \min_{y \in \check{B}_x} \max_{z \in B_y} f(z). \tag{2.7}$$

Consider a line structuring element of length $n+1$,

$$B = \boxed{0}\boxed{1}\boxed{\cdots}\boxed{n}$$

Then $\epsilon_B(f)(i) = \min\{f(i), f(i+1), ..., f(i+n)\}$ and $\delta_{\check{B}}(f)(i) = \max\{f(i-n), f(i-n+1), ..., f(i)\}$ so that,

$$\begin{aligned}
\gamma_B(f)(i) &= \delta_{\check{B}}\big(\epsilon_B(f)(i)\big) \\
&= \max\{\epsilon_B(f)(i-n), ..., \epsilon_B(f)(i)\} \\
&= \max\{\min\{f(i-n), f(i-n+1), ..., f(i)\}, ... \\
&\qquad ..., \min\{f(i), f(i+1), ..., f(i+n)\}\}.
\end{aligned}$$

So if we replace $f(i)$ with x_i, the i^{th} element in the sequence x, then we have $(\gamma_B(x))_i = \max\{\min\{x_{i-n}, ..., x_i\}, ..., \min\{x_i, ..., x_{i+n}\}\} = (L_n(x))_i$. Similarly, $(\rho_B(x))_i = \min\{\max\{x_{i-n}, ..., x_i\}, ..., \min\{x_i, ..., x_{i+n}\}\} = (U_n(x))_i$. Thus in one dimension L_n is an opening and U_n is a closing with the specific structuring element B, and the properties of anti-extensivity, extensivity, increasingness and idempotence follow automatically.

We have the following additional morphology concepts.

Definition 39 *A* **filter** *is an operator, F, which is syntone and idempotent. If $F^2 \leq F$ then the filter is called an* **underfilter** *or if $F^2 \geq F$ then the filter is called an* **overfilter**. *A* $\bigvee$**-filter** *is one for which $F(F \bigvee I) = F$ and a* $\bigwedge$**-filter** *if $F(F \bigwedge I) = F$. If F is a $\bigvee$-filter as well as a $\bigwedge$-filter, then it is called a* **strong filter**.

<u>RESULT 24</u> The operators $L_n, U_n, L_n \circ U_n$ and $U_n \circ L_n$ are strong filters.
Proof

$$\begin{aligned}
L_n(L_n \bigvee I) &= L_n(I) \text{ since } L_n \leq I \\
&= L_n \text{ and} \\
L_n(L_n \bigwedge I) &= L_n \circ L_n \\
&= L_n \text{ by the idempotence of } L_n.
\end{aligned}$$

The result holds similarly for U_n. Since U_n and L_n are attribute filters (see Chapter 3.3), the compositions $U_n \circ L_n$ and $L_n \circ U_n$ are also strong by Property 8 in [34]. ∎

2.3 Smoothing with LULU and Total Variation

By increasing the monotonicity of the sequence x, or equivalently by decreasing the variation within the sequence, are both methods which smooth the sequence x. This can also be viewed as improving the continuity (in the classical sense) of the sequence. Other measures of continuity are differentiability, integrability, measurability, bounded variation and the usual classical continuity. A function which is of bounded variation is differentiable almost everywhere, has only a countable number of discontinuities and is Riemann integrable, [98]. The LULU smoothers make use of monotonicity and total variation to measure the level of continuity of the original sequence as well as resulting sequences once the LULU operators are applied. We defined the monotonicity of a sequence in Definitions 25 and 26, and now define the concept of total variation for a sequence.

Definition 40 *The* **total variation** *for a sequence x is given by,*

$$T(x) = \|\Delta x\|_1 = \sum_{i=-N}^{N} |x_{i+1} - x_i|,$$

where the sum is from $-N$ to N since $x_i = 0$ for $i < -N$ and $i > N$. Note that since $x \in \ell_1$ we have $\Delta x \in \ell_1$ so $T(x)$ is well-defined.

Total variance is a semi-norm (1. $T(x) \geq 0$, 2. $T(\alpha x) = |\alpha| T(x)$, and 3. $T(x+y) \leq T(x) + T(y)$) but becomes a norm since $T(x) \leq 2\|x\|_1$. It is therefore a natural norm to use in our theory since $x \in \ell_1$, [95].

From [93] we have the following results:

<u>RESULT 25</u> $Tx \geq T(\bigvee x)$ (Theorem 6.2 in [93]) thus $\bigvee$ is a total variation diminishing operator. Similarly, $Tx \geq T(\bigwedge x)$ (Corollary on page 54 in [93]) thus $\bigwedge$ is also a total variation diminishing operator.

<u>RESULT 26</u> We also have that any composition of $\bigvee$ and $\bigwedge$ is also total variation diminishing, so that the LULU operators are total variation diminishing.

<u>RESULT 27</u> $T(\bigwedge \bigvee x) = T(\bigvee x)$ and $T(\bigvee \bigwedge x) = T(\bigwedge x)$. Thus we do not achieve any further reduction in total variation.

RESULT 28 $Tx = T(Ux)+T(x-Ux)$ and $Tx = T(Lx)+T(x-Lx)$. We also have $Tx = T(L \circ Ux) + T(x - L \circ Ux)$ and $Tx = T(U \circ Lx) + T(x - U \circ Lx)$. Thus the LULU smoothers preserve total variation.

The following definitions provide us with shape preservation properties.

Definition 41 *An operator P is* **neighbour trend preserving (ntp)** *if for each x,*

$$x_{i+1} \leq x_i \Rightarrow Px_{i+1} \leq Px_i$$

and

$$x_{j+1} \geq x_j \Rightarrow Px_{j+1} \geq Px_j$$

Definition 42 *An operator P is* **difference reducing** *if for each x and subscript i,*

$$|Px_{i+1} - Px_i| \leq |x_{i+1} - x_i|.$$

Definition 43 *An operator P is* **fully trend preserving (ftp)** *if it is neighbour trend preserving and difference reducing.*

Some important properties of ftp operators are:

1. For a ftp operator P we have that for each x,

   $$Tx = T(Px) + T(x - Px).$$

   The converse of this also holds, [95]. This means that a ftp operator is also total variation preserving, and vice versa.

2. For two ftp operators A and B,

   a) $A \circ B$ and $B \circ A$ are ftp (in fact any composition of ftp operators is ftp),

   b) $\alpha A + (1 - \alpha)B$ is ftp for $\alpha \in [0, 1]$, and

   c) $I - A$ is ftp.

   Statement (c) is in fact true if and only if.

3. A ftp operator also preserves n-monotone sequences.

RESULT 29 $\bigvee\bigwedge$, $\bigwedge\bigvee$ and all compositions of them are neighbour trend preserving. Note that $\bigvee$ and $\bigwedge$ are however not ntp.

RESULT 30 For each n, the LULU operators L_n and U_n are neighbour trend preserving and difference reducing. The LULU operators are therefore fully trend preserving. Consequently, the results in Result 28 are then proved in an alternative way. This means that our LULU smoothers never change the order of neighbours in the sequence. This is not a common result for linear operators, [95]. An alternative method for proving the neighbour trend preservation of L_n and U_n is presented in Theorem 12 of [136], since the LULU operators are also stack filters (although we will not present the proof of it here).

RESULT 31 L_n and U_n preserve n-monotone sequences.

The median smoother M_n is trend preserving in the sense that if any window $\{x_{i-n}, ..., x_i, ..., x_{i+n}\}$ is monotone then $(M_n(x))_i = x_i$. The local property in Definition 41 is much stronger than this property of the median operators enforcing the strength of the LULU operators.

If a measured signal is considered to be composed of the true signal s and the noise n, such that $x = s + n$, then in order to 'clean up' or smooth the signal we wish to remove the variation due to noise. We should be able to assume that $T(n) \ll T(x)$ i.e the signal-to-noise ratio (SNR), is not too high. If we use the LULU smoothers to remove the noise there should be an indication of when x is sufficiently smooth and an indication of this is when large reductions in variation stop, as this would mean we have removed the noise adequately. The variation spectrum will help in this regard.

Definition 44 *The* **variation spectrum** *of a sequence* $x \in \ell_1$ *is given by* $t(x) = \{t_i = T(r^i(x)), i \in \mathbb{Z}\}$.

In Figure 2.11 and Table 2.2 we illustrate this method. Notice the large proportion of total variation removed with F_1 (C_1 would give a similar result). The total variation subsequently removed reduces drastically at each step. We can see that as the total variation approaches the original total variation of the non-noisy original signal, the variation spectrum tapers off. A reduction in variation is indicative of a smoother signal.

A good indication of when sufficient noise has been removed is when the total variation reaches the expected total variation i.e. the variation we expect in the true signal. This true variation is obviously unknown but can perhaps be estimated? This has been looked into in [94]. Another approach is to fit a distribution to the removed pulses to acquire an indication of what has been removed.

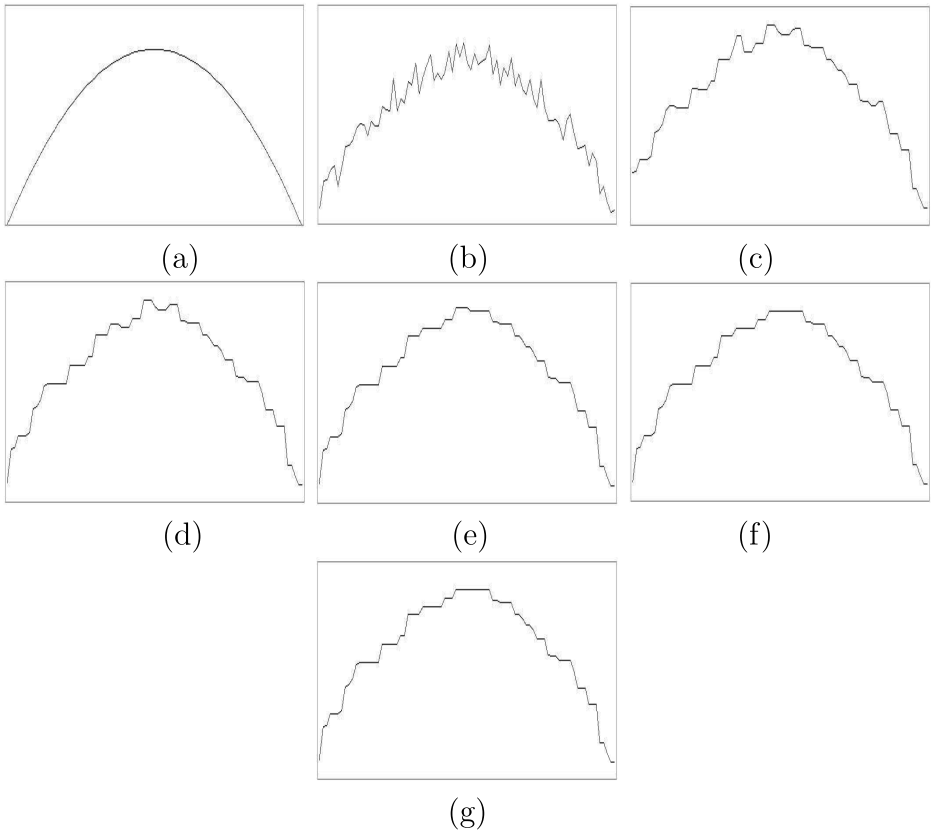


Figure 2.11. (a) Original Signal x (b) Signal with Added Noise $x + n$ (c) $U_1 \circ L_1(x + n)$ (d) $F_2(x + n)$ (e) $F_3(x + n)$ (f) $F_4(x + n)$ (g) $F_5(x + n)$

	Total Variation	Total Variation Still Present (%)	t_i
x	400	-	-
$x + n$	1486.82933	100	0
$F_1(x + n)$	505.2859	33.98	981.5434
$F_2(x + n)$	455.2960	30.62	49.9899
$F_3(x + n)$	422.1903	28.40	33.1057
$F_4(x + n)$	413.9632	27.84	8.2271
$F_5(x + n)$	413.9632	27.84	0

Table 2.2. Total Variation Reduction and the Variation Spectrum for Figure 2.11.

We end this section with two comments from [93]:

"Philosophically, we know that smoothing destroys information, unless that which is removed is stored separately."

Thus we will introduce the Discrete Pulse Transform in the next section. It involves the use of Px (the smoother part) and $(I - P)x$ (the noise component) when a LULU operator P is applied. The question to answer, once the transform has been applied, will be when to stop smoothing. A simple logical approach was shown in Figure 2.11, however,

"If some form of automated selection of degree of smoothing of data is required, it seems natural to smooth recursively, until come selected criterion is met."

The selection criterion chosen could be analogous to convergence, that is, when the variation reduction drops below a specified threshold the signal is considered sufficiently smooth.

2.4 The Discrete Pulse Transform in One Dimension

The Discrete Pulse Transform (DPT) is an application in Multiresolution analysis (MRA). It results in a decomposition of the sequence x, or when moving to two dimensions, an image f. The aim of this thesis is the investigation of reconstructing an image using the LULU operators and from the following quote in [93] we already feel on the right track.

"The claim is made that MRA is well suited when image reconstruction is done from a subset of the (additive-) decomposition for purposes of restoration, compression and partial reconstruction."

Definition 45 *The* **Discrete Pulse Transform** *is a mapping of a sequence x into a vector*

$$DPT(x) = [D_1(x), D_2(x), ..., D_N(x), D_0(x)].$$

$DPT(x)$ is a decomposition of x obtained using either $U_n \circ L_n$ or $L_n \circ U_n$. If using $U_n \circ L_n$, the decomposition is obtained as followed (and similarly for $L_n \circ U_n$):

1. **Apply** $U_1 \circ L_1$ **to** x. *Then*

   $$x = (U_1 \circ L_1)x + (I - U_1 \circ L_1)x = S_1(x) + D_1(x).$$

   So S_1 is the 'smoother' sequence and $D_1(x)$ is the noise removed by $U_1 \circ L_1$. The first component of the DPT is then D_1.

2. **Apply** $U_2 \circ L_2$ **to** $S_1(x)$. *Then*

   $$S_1(x) = (U_2 \circ L_2)S_1(x) + (I - U_2 \circ L_2)S_1(x) = S_2(x) + D_2(x).$$

   Again S_2 is the 'smoother' sequence (even smoother than $S_1(x)$) and $D_2(x)$ is the noise removed by $U_2 \circ L_2$. The second component of the DPT is then D_2.

3. *Continue this decomposition until $U_N \circ L_N$ is applied where N is the size of the signal x. This last application will result in*

   $$S_{N-1}(x) = (U_N \circ L_N)S_{N-1}(x) + (I - U_N \circ L_N)S_{N-1}(x) = S_N(x) + D_N(x).$$

> *D_N is then the second to last member of the DPT and S_N is a constant sequence so we denote it by $D_0(x)$. This is so that the decomposition contains all the information in the original sequence and the original sequence can be reconstructed using the DPT.*

The original sequence x can then be reconstructed using the DPT as follows:

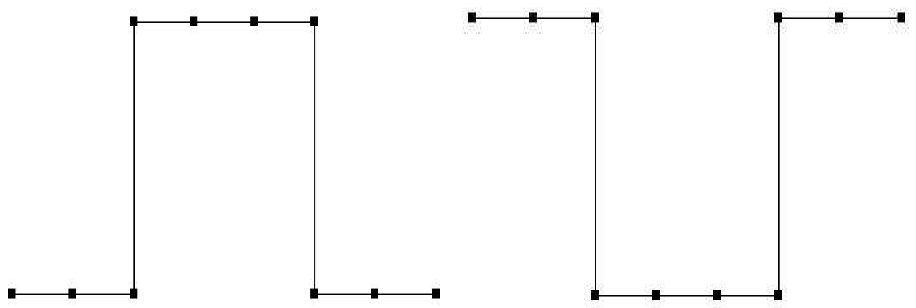
$$x = \sum_{n=1}^{N} D_n(x) + D_0(x).$$

The Discrete Pulse Transform for LULU operators is defined in [93] but its properties are particularly discussed in [63].

Each D_i is a discrete pulse which is made up of upward and downward pulses of size i which the LULU operator $U_i \circ L_i$ removed at the i^{th} iteration.

Definition 46 *A blockpulse of size n is such that $x_i = k$, a constant, for $i = j, j+1, ..., j+n-1$ but $x_{j-1} = k^* \neq k$ and $x_{j+n} = k^* \neq k$. The blockpulse is called* **upward** *if $k > k^*$ is positive or* **downward** *if $k < k^*$ is negative.*

The name blockpulse is perfectly descriptive as these are simply upward or downward blocks appearing in a sequence, see Figure 2.12.



(a) (b)

Figure 2.12. (a) An Upward Pulse (b) A Downward Pulse

Blockpulses are important as the LULU operators act directly on them. This has been termed the Roadmaker's Algorithm, [66]. The operator L_n removes all upward blockpulses of size n or less and U_n removes all downward blockpulses of size n or less. In addition, L_n does not remove downward pulses and U_n does not remove upward pulses. Thus when applying $L_n \circ U_n$ all blockpulses of size n or less are removed or sliced off. If we apply $L_n \circ U_n$ successively from $n = 1, 2, .., N$, as is explained in the DPT algorithm above, it means that $L_n \circ U_n$ will have removed a collection of upward and downward blockpulses of size n and none of size less than n. These would have already been removed by the previous applications $L_1 \circ U_1, L_2 \circ U_2, ...$ and $L_{n-1} \circ U_{n-1}$. We present an illustration of the DPT on a sequence x in Figure 2.13. Notice that when we sum $D_1(x), D_2(x)$ and $D_3(x)$ we obtain the original sequence x.

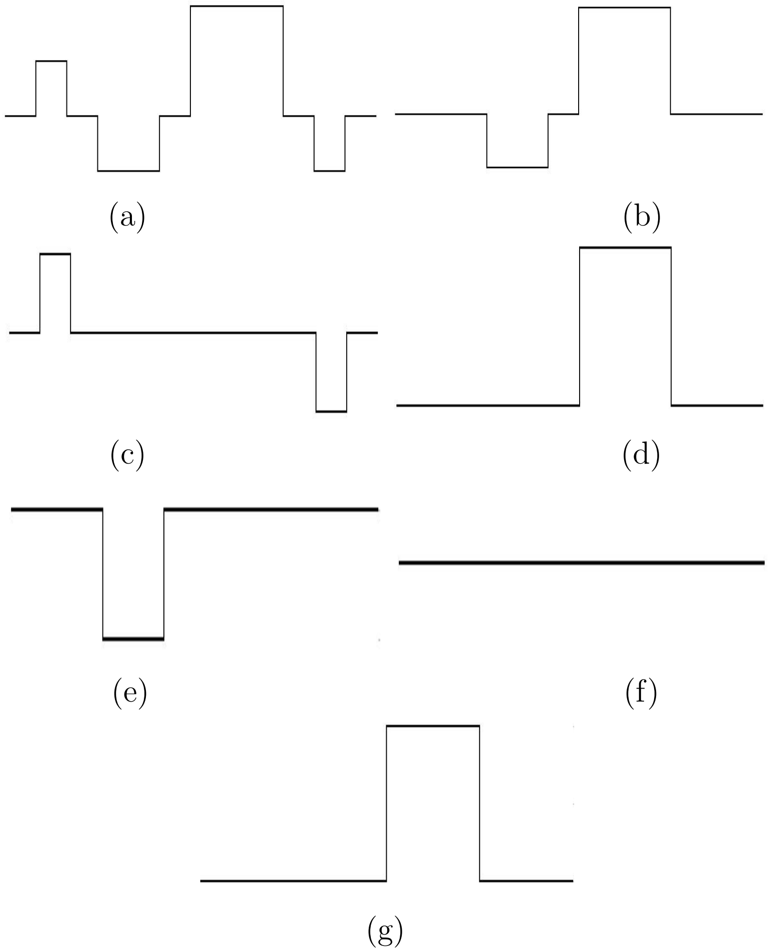
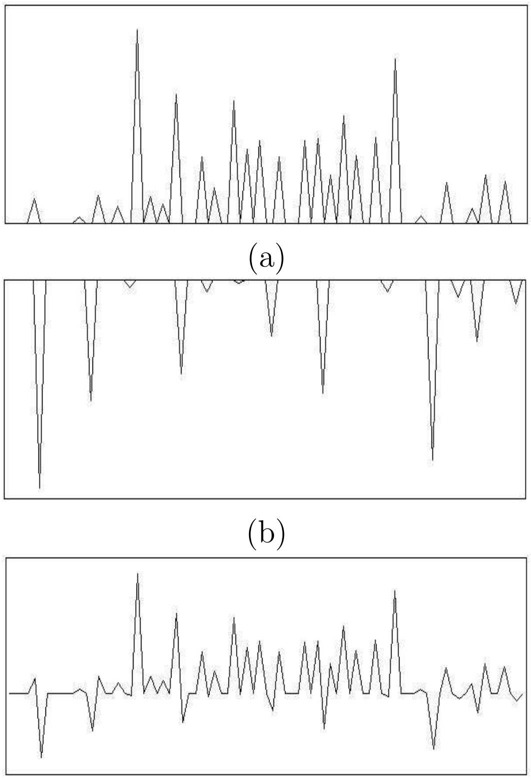


Figure 2.13. (a) Original sequence x (b) $U_1 \circ L_1 x$ (c) $D_1(x)$ (d) $U_2 \circ ... \circ L_1 x$ (e) $D_2(x)$ (f) $U_3 \circ ... \circ L_1 x$ (g) $D_3(x)$

Recall that the LULU operators are consistent i.e. they are idempotent and co-idempotent. We in fact have a more powerful result for the DPT, proven in [63], namely

The Highlight Result If $z = \sum_{n=1}^{N} \alpha_n^- D_n^-(x) + \alpha_n^+ D_n^+(x) + D_0(x)$, so that each resolution level is multiplied by it's own non-negative constant, and where $D_n(x) = D_n^-(x) + D_n^+(x)$ is separated into it's positive and negative pulses, then z is decomposed consistently i.e. $D_n^+(z) + D_n^-(z) = \alpha_n^+ D_n^+(x) + \alpha_n^- D_n^-(x)$.



(a)

(b)

(c)

Figure 2.14. For $x + n$ from Figure 2.11: (a) $D_1^+(x + n)$ (b) $D_1^-(x + n)$ (c) $D_1(x + n)$

In a simpler manner we can also write the above result as follows: If $z = \sum_{n=0}^{N} \alpha_n D_n(x)$ then $D_n(z) = \alpha_n D_n(x)$. This is a very important result because it means that multiples of pulses within a resolution level will appear in the same resolution level without distortion, enabling us to highlight certain pulses. Rohwer comments in [93] that this could have applications in image analysis, specifically feature detection.

Let us return to the smoothing of a sequence. Ultimately we wish to make use of the DPT to separate a sequence into its noise component and its true component. So how can we make use of the DPT? We need to decide which resolution levels contain the true information and then we are able reconstruct the signal with only those discrete pulses rendering a truer picture of the nature of the data. Noise usually occurs in lower resolution levels, thus the reason for applying $L_n \circ U_n$ from $n = 1, 2, ..., N$ rather than as $n = N, N-1, ..., 2, 1$. The noise in a signal can be removed when the lower resolution levels have been detected via the LULU operators for 'small' n. In [95] it is suggested that a distribution is fitted to the first resolution layer $r^{(1)}$ to determine the appropriate thresholding to remove the noise significantly. In Figure 2.14, this first resolution layer can be seen. We will look at this in more detail in the next section. We also know that $T(r^{(n)}) = \frac{2}{n}\|r^{(n)}\|$, [95], so we can use total variation to decide on the required resolution levels we keep.

An alternative method for using the LULU smoothers for removing noise is presented in [61]. There a LULU-filtration method is proposed by using the 3-point filtration operator

$$(S(x))_i = \frac{1}{2}\big((L_1 \circ U_1(x))_i + (U_1 \circ L_1(x))_i\big)$$

and on the edges (end points) of the signal the operations are $(L_1x)_0, (U_1x)_0$, $(L_1x)_N$ and $(U_1x)_N$. They also propose a 5-point filtration operator by using L_3 and U_3 similarly to above and by using windows of size 2 on the edges.

2.5 Distributions of the LULU Operators

In [27] the exact and asymptotic distributions (as n goes to infinity) of L_n, U_n, $L_n \circ U_n$ and $U_n \circ L_n$ are derived for one dimension. The first case considered is for a sequence of random variables $..., X_{-1}, X_0, X_1, ...$ which are independent and identically distributed with distribution function $F_X(x) = P[X \leq x]$. They are as follows for $n = 1, 2, 3, ...$,

$$\begin{aligned}
F_{L_n \circ U_n(X)}(x) &= F_X^{n+1}(x) + n(1 - F_X(x))F_X^{n+1}(x) + (1 - F_X(x))F_X^{2(n+1)}(x) \\
&\quad \frac{1}{2}(n-1)(n+2)(1 - F_X(x))^2 F_X^{2(n+1)}(x) \\
F_{U_n \circ L_n(X)}(x) &= 1 - \left[1 - F_X(x)\right]^{n+1} - nF_X(x)\left[1 - F_X(x)\right]^{n+1} \\
&\quad - F_X(x)\left[1 - F_X(x)\right]^{2(n+1)} \\
&\quad - \frac{1}{2}(n-1)(n+2)F_X^2(x)\left[1 - F_X(x)\right]^{2(n+1)} \\
F_{U_n(X)}(x) &= F_X^{n+1}(x) + n(1 - F_X(x))F_X^{n+1}(x) \\
F_{L_n(X)}(x) &= 1 - \left[1 - F_X(x)\right]^{n+1} - nF_X(x)\left[1 - F_X(x)\right]^{n+1}.
\end{aligned}$$

In [56] it is also stated that work is being done on determining the joint distribution of $L_n \circ U_n$ and $L_n \circ U_n$.

The second case considered was for $..., X_{-1}, X_0, X_1, ...$ independent but no longer identically distributed i.e. each X_j has distribution function $F_j(x)$ for $j = 0, \pm 1, \pm 2, ...$, [56]. Then we have

$$\begin{aligned}
F_{L_n \circ U_n(X)_i}(x) &= 1 - q_{i-n}(x) + \sum_{r=i-n}^{i} d_{r,n}(x) + \sum_{r=i+1}^{i+n-1} d_{r,n}(x)\left[q_{i-n}(x) - \sum_{s=i-n}^{r-n-2} d_{s,n}(x)\right] \\
&\quad - \sum_{k=i}^{i+n-1} d_{k,n}(x)\left[q_{k-2n}(x) - \sum_{r=k-2n}^{k-n-2} d_{r,n}(x)\right]
\end{aligned}$$

where

$$q_k(x) = 1 - \prod_{j=k}^{k+n} F_j(x)$$

and

$$d_{k,n}(x) = (1 - F_k(x)) \prod_{j=k+1}^{k+n+1} F_j(x),$$

and

$$F_{U_n \circ L_n(X)_i}(x) = q^*_{i-n}(x)) - \sum_{r=i-n}^{i} d^*_{r,n}(x) + \sum_{r=i+1}^{i+n-1} d^*_{r,n}(x)\left[q^*_{i-n}(x) - \sum_{s=i-n}^{r-n-2} d^*_{s,n}(x)\right] + \sum_{k=i}^{i+n-1} d^*_{k,n}(x)\left[q^*_{k-2n}(x) - \sum_{r=k-2n}^{k-n-2} d^*_{r,n}(x)\right]$$

where

$$q^*_k(x) = 1 - \prod_{j=k}^{k+n}(1 - F_j(x))$$

and

$$d^*_{k,n}(x) = F_k(x) \prod_{j=k+1}^{k+n+1}(1 - F_j(x)).$$

Lastly, they consider the asymptotic distributions i.e. as n tends to infinity. Since the LULU operators are based on the extreme values they make use of results from Extreme Value Theory. In this direction they make use of the Fisher-Tippett theorem, [133].

Theorem 47 Fisher-Tippett *For sequences of constants* $(a_n), a_n > 0$ *and* (b_n) *such that*

$$\frac{1}{a_n}(X_{(n)} - b_n) \underset{\rightarrow}{D} H$$

as $n \rightarrow \infty$ *where* $X_{(n)} = \max\{X_1, ...X_n\}$ *and* H *is a non-degenerate distribution function which can be either of the type Frechet, Gumbel or Weibull, [39]. If this holds* F_X *is said to belong to the maximum domain of attraction of* H *or* $F_X \in MDA(H)$.

The asymptotic distributions are then given by

$$F_{U_n}(a_n x + b_n) \underset{\rightarrow}{D} H(x) - H(x) log H(x),$$

$$F_{L_n \circ U_n}(a_n x + b_n) \underset{\rightarrow}{D} H(x) - H(x) log H(x) + \frac{1}{2}\left[H(x) log H(x)\right]^2,$$

$$F_{L_n}(a_n x + b_n) \quad \underset{\rightarrow}{D} \quad 1 - (1 - H(x))(1 - log(1 - H(x)))$$

and

$$F_{U_n \circ L_n}(a_n x + b_n) \quad \underset{\rightarrow}{D} \quad 1 - H(x) - (1 - H(x))(log(1 - H(x))) - \frac{1}{2}\left[(1 - H(x))log(1 - H(x))\right]^2$$

where a_n and b_n are normalizing constants determined by F_X.

They make use of the above distributions to detect an edge by looking at a jump as the result of two different underlying distribution functions. In other words, if the distributions suddenly change it means there is a difference in the characteristics of the data at that point perhaps indicating an edge of some object of interest. By extending this into two dimensions it will be useful to detect an object in an image by looking at the changes in distribution in the vicinity of an suspected object or target.

In [94] and [38] a procedure is investigated in which the variance of the noise is estimated. It is assumed that the noise forms part of the first resolution level and comes from a symmetric distribution. The average height of the negative pulses in the first resolution level using $L_1 \circ U_1$ is made use of. A third of the pulses are negative in this first level. The symmetry of the distribution may not always be the case in applications thus a more general method is required.

2.6 Conclusion

In this chapter we have presented a summary of the state of the art of the one dimensional LULU smoothers developed by Carl Rohwer and his collaborators. The properties of idempotence, total variation preservation, fully trend preserving, and the resulting Discrete Pulse Transform provide a framework for these nonlinear filters in which the benefits of linear filters are maintained and the disadvantages rectified. In one dimension these operators are applicable to sequences in ℓ_1. We quote from [93],

"The LULU-decompositions [in one dimension] are an alternative to the prevalent median transform of the same type. These seem generally to be good in the two-dimensional case of image processing, [9]. A disadvantage listed is the computational complexity. A more serious disadvantage seems to be the lack of theory."

The above mentioned lack of theory will be addressed in this thesis in the next two chapters. Digital images are discrete and are represented on a two dimensional integer grid. Can we extend the LULU operators to such a grid in a natural way, maintaining the properties developed by Rohwer as well as the all important Discrete Pulse Transform with its Highlighting Result? The implications for image processing will then provide numerous new insights into the multiresolution structure of images. We provide this extension, in a natural way, in the next chapter.

Chapter 3

LULU Theory on Multidimensional Arrays

3.1 Introduction

David Marr, [76], specifies the following as important for image analysis, [95]:

1. There exists a science of vision that must be developed for the understanding of human and robotic vision. The foundation must be found. Algorithms involved should be *physiologically realistic*.

2. The doubt in iterative loops, and the belief that the choice of *representation of an image* is crucial.

The extension of the LULU theory to multidimensional arrays has obvious implications in image processing, and in three dimensions, video processing. The algorithms and representation developed should thus receive attention when developing such a theory. The LULU theory, when extended precisely from one dimension, does provide a realistic algorithm, Section 4.3 and representation, Section 4.2.

In [95] the one-dimensional LULU operators are applied to an image by first decomposing each row separately, then each column separately and then finally taking averages to get a single resulting image. This implies a rotational symmetry of only π radians so we do not have a complete extension into two dimensions, [95]. Also see [87] for a proposed definition of L_1 and U_1 in two dimensions. Here their idempotence and co-idempotence are proven. However, for $n > 1$ there have previously been no results. In [58] an extension to two dimensions using 8-connectivity is proposed, but the construction does not satisfy the idempotence of the LULU operators. This chapter will provide the extension of the LULU operators to higher dimensions thus solving this problem of lack of theory. The need for this extension is also emphasized in the conclusion of [38].

The LULU operators are morphological filters, which are increasing, idempotent operators. However, unlike mainstream mathematical morphology, see Section 2.2.1, the emphasis in LULU theory is on what one may call structure preserving properties, like: consistent separation (separation of noise from signal) (Section 3.4), total variation (Section 3.6) and shape preservation (Section 3.5), and consistent hierarchical decomposition (Section 3.7). We discuss these properties briefly.

Consistent Separation

The issue of consistency of nonlinear filters is not easy to address in a straightforward manner. In fact, one may note that there is no established approach to this issue, with some authors only providing empirical evidence on the quality of their considered filters. Characterization of the quality of nonlinear filters is discussed at length in [75]. The concept of a smoother introduced there is based on preserving some linearity, namely, these are operators which are shift, location and scale invariant.

A common requirement for a filter P, linear or nonlinear, is its idempotence, i.e. $P \circ P = P$, for example, a morphological filter. As mentioned in Chapter 2, for linear operators the idempotence of P implies the idempotence of the complementary operator $I-P$. For nonlinear filters this implication generally does not hold so the idempotence of $I-P$, also called co-idempotence, [135], can be considered as an essential measure of consistency.

The above mentioned properties are all discussed in [93], [64], and in the previous chapter, where they are considered to collectively constitute what the we call a **consistent separation** and are absorbed into the concept of a **separator**. We will give the definition of a separator for operators on real functions defined on a domain with a group structure. Let Ω be an abelian group, so that commutativity always holds. Denote by $\mathcal{A}(\Omega)$ the vector lattice of all real functions defined on Ω with respect to the usual point-wise defined addition, scalar multiplication and partial order. For every $a \in \Omega$ the operator $E_a : \mathcal{A}(\Omega) \to \mathcal{A}(\Omega)$ given by $E_a(f)(x) = f(x-a),\ x \in \Omega$, is called a shift operator.

Definition 48 *An operator* $P : \mathcal{A}(\Omega) \to \mathcal{A}(\Omega)$ *is called a* **separator** *if*

(i)	$P \circ E_a = E_a \circ P,\ a \in \Omega$	*(Horizontal shift invariance)*
(ii)	$P(f+c) = P(f) + c,\ f, c \in \mathcal{A}(\Omega),$	c *-constant function* *(Vertical shift invariance)*
(iii)	$P(\alpha f) = \alpha P(f),\ \alpha \in \mathbb{R},\ \alpha \geq 0,$	$f \in \mathcal{A}(\Omega)$ *(Scale invariance)*
(iv)	$P \circ P = P$	*(Idempotence)*
(v)	$(I-P) \circ (I-P) = I - P.$	*(Co-idempotence)*

<u>Total Variation and Shape Preservation</u>

The total variation is a semi-norm on $\mathcal{A}(\Omega)$, as in Chapter 2. In the practically significant case of signals ($\Omega = \mathbb{Z}$), dealt with in Chapter 2, the total variation is a generally accepted measure for the amount of information present. Note that any separation of sequences may only increase the total variation. More precisely, for any operator $P : \mathcal{A}(\mathbb{Z}) \to \mathcal{A}(\mathbb{Z})$ we have

$$TV(f) \leq TV(P(f)) + TV((I-P)(f)). \tag{3.1}$$

Hence it is natural to expect that a good separator P should not create new variation, that is we have (in one dimension)

$$TV(f) = TV(P(f)) + TV((I-P)(f)). \tag{3.2}$$

An operator P satisfying property (3.2) is called **total variation preserving**, [88].

Shape preservation generally refers to the preservation of edges in the input, since the edges define shapes. Total variation preservation is closely linked to the shape preservation properties of the filter. In the case of signals the preservation of shape is actually preservation of trend. It is shown in [93] that a fully trend preserving operator on sequences is also total variation preserving. We will show that this is also true for images, that is, in two dimensions.

<u>Consistent Hierarchical Decomposition</u>

Similarly to the consistency of separation, characterizing the quality of hierarchical decompositions by nonlinear filters is also problematic. Indeed, as mentioned in [63], a decomposition by linear operators should typically recover the coordinates of any given linear combination of basis vectors, a property not at all applicable to the nonlinear case. A measure of the quality of any hierarchical decompositions is introduced in [63], namely:

A nonnegative linear combination of the output of the decomposition is decomposed into the same components. (3.3)

This was introduced in Chapter 2 as the Highlight Result. The aim of this chapter is to generalize the LULU operators to functions on $\Omega = \mathbb{Z}^d$ in such a way that their essential properties, mentioned above, are preserved.

3.2 Connectivity

The one-dimensional LULU operators act on sequences in the space ℓ_1. In this space one can see that we have an obvious ordering of the elements, namely x_{i+1} follows x_i and x_{i-1} precedes x_i. It is then natural to consider the elements x_{i+1} and x_{i-1} as the neighbours of x_i. The one-dimensional LULU operators act specifically on the n-neighbourhoods of each x_i, that is,

$$\{x_{i-n}, x_{i-n+1}, ..., x_i\}, ..., \{x_i, x_{i+1}, .., x_{i+n}\}. \tag{3.4}$$

So the order in a sequence in ℓ_1 is inherent in the definition of L_n and U_n in one dimension (Definition 29), since $\mathbb{Z}$ is totally ordered (see Definitions 1 and 2). Consider the case of images defined on a discrete grid in $\mathbb{Z}^2$.

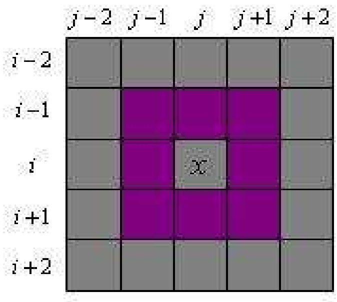


Figure 3.1. Discrete image grid in $\mathbb{Z}^2$ with pixel x at position (i, j).

Although it is natural to consider the 8 surrounding pixels for a pixel x as the neighbours, see Figure 3.1, there is no immediate ordering of the neighbours as is the case in one dimension. This is because $\mathbb{Z}^2$ is only partially ordered. In addition the n-neighbourhoods which contain the pixel x are numerous. In Figure 3.2 we see some possible 4-neighbourhoods, but there are obviously many more.

The ordering of these neighbourhoods cannot be done in a natural way, contrary to the natural ordering in (3.4). We could apply a raster scan to the grid, that is, starting with the first row move left to right from pixel to pixel and then repeat at next row and subsequent rows. This would however mean we have reduced the grid in $\mathbb{Z}^2$ to a sequence in ℓ_1 and we won't have achieved anything. Thus we see that the one dimensional ordering cannot be extended to two dimensions (or higher dimensions). We are aiming for a natural extension of the one-dimensional LULU operators into two dimensions, specifically images, and more generally onto an arbitrary dimensions $\mathbb{Z}^d$. When we take $d = 1$ the extended LULU operators should reduce to the one-dimensional LULU operators exactly. In addition, taking $d = 2$ we would like to have operators which naturally and logically extend from one dimension by preserving all the properties present in one dimension and having the equivalent higher dimensional concepts. We make use of the concept of connectivity to aid us in the extension.

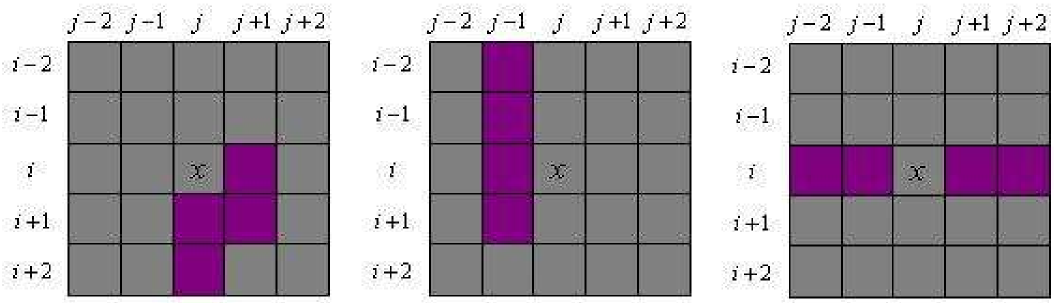


Figure 3.2. Possible 4-neighbourhoods of pixel x at position (i, j).

The concept of morphological connectivity was introduced by J. Serra and G. Matheron in the 1980's. They recognised the need for the concept of an axiomatic connectivity. There were two main connectivity concepts before they set up the axiomatic approach. These were arcwise connectivity and topological connectivity, both defined on a topological space E. Arcwise connectivity is based on paths, [129].

Definition 49 *A* **path** *in E joining two points $p, q \in E$ is a continuous mapping $f : [0, 1] \to E$ such that $f(0) = p$ and $f(1) = q$.*

Definition 50 *A subset D of E is* **arcwise connected** *if $\forall\, p, q \in D$ there exists a path in D joining p and q.*

Graph connectivity, [40], has a similar concepts to arcwise connectivity although it is defined on a graph and not a topological space. The two concepts cannot be shown to be equivalent.

Definition 51 *A graph $G = (V, E)$ is* **(graph-)connected** *if there exists a path $p = \{v_1, ..., v_n\}$ between every set of vertices. A path is a sequence of vertices in which each pair of consecutive vertices is joined by an edge.*

Topological connectivity, [129], is the weaker form of connectivity since arcwise connectivity implies topological connectivity, see Theorem 17 in [129, Chapter III, Section 4].

Definition 52 *Topological set E is* **(topologically) connected** *if there do not exist two open sets A and B such that $A \cap E$ and $B \cap E$ are disjoint, non-empty and have union E.*

Arcwise connectivity and topological connectivity are not in general equivalent. However, in $\mathbb{R}^d$ they are equivalent, [107]. Consider an arbitrary $V \in \mathbb{R}^d$ which is topologically connected, and let $p = (y_1, ..., y_d)$ and $q = (z_1, ..., z_d)$ be points in V. Define $f : [0, 1] \to V$ such that $f(t) = (x_1(t), ..., x_d(t))$ where $x_i(t) = (1 - t)y_i + tz_i$. Then since V is topologically connected, there do not exist open sets A and B such that $A \cup B = V$ but $A \cap V$ and $B \cap V$ are disjoint. Thus f is continuous, and $f(0) = p$ and $f(1) = q$, so that V is arcwise connected. By [129, Theorem 17, Chapter III] an arcwise connected space is topologically connected. However 8-connectivity in $\mathbb{Z}^2$, which is a kind of graph-connectivity, has no related topology for which its topology is equivalent to 8-connectivity, [96].

Thus the axiomatic approach to connectivity was introduced. In 1988 Serra, [105], and Matheron in 1985, [78], introduced the concept of a connectivity class, for use in Mathematical Morphology.

Definition 53 *$\mathcal{C}$ is a* **connectivity class** *or a* **connection** *on $\mathcal{P}(E)$ if the following axioms hold:*

(i) $\emptyset \in \mathcal{C}$

(ii) $\{x\} \in \mathcal{C}$ *for each* $x \in E$

(iii) For each family $\{C_i\}$ in $\mathcal{C}$ such that $\bigcap C_i \neq \emptyset$, we have $\bigcup C_i \in \mathcal{C}$.

A set $C \in \mathcal{C}$ is called ***connected****.*

Note that the intersection of connected sets is not necessarily connected. For example, consider Figure 3.3 which shows two connected sets (with respect to 4-connectivity). The intersection shown by the highlighted cells is clearly not 4-connected.

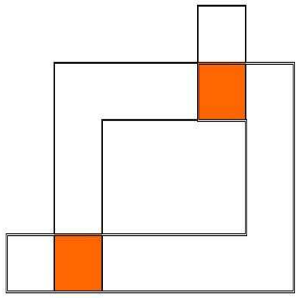


Figure 3.3. Intersection of connected sets

Axiom (iii) is required so that if we consider $\mathcal{C}_x = \{C : x \in C, C \in \mathcal{C}\}$, any union of a non-empty family in $\mathcal{C}_x$ is again in $\mathcal{C}_x$. Then the **point-connected opening**

$$\gamma_x(A) = \cup\{C : C \in \mathcal{C}, x \in C \subset A\} \tag{3.5}$$

has the invariant set $\beta_{\gamma_x} = \mathcal{C}_x \cup \{\emptyset\}$. The point-connected opening extracts the connected component of the set A which contains x. The point x is referred to as the point marker. Note that from this we can now extract a connected subset containing a specified point from a connected set.

The point-connected opening is an opening according to Definition 20 in Chapter 2. It is increasing since if $A \subseteq B$ then the connected component of B containing x will contain the connected component of A which contains x i.e. $\gamma_x(A) \subseteq \gamma_x(B)$. It is also anti-extensive, since the extracted connected component is contained in A, and idempotent, because the extracting the connected component a second time will not give a different connected component. In [59] an algorithm is presented, to determine $\gamma_x(A)$ in which x is dilated by a structuring element B but restricted to A, until convergence. The structuring element used is an elementary one, either the 6-pixel hexagon or 9-pixel square.

The following important result is proved in [105].

Theorem 54 *A connectivity class* $\mathcal{C}$ *on* $\mathcal{P}(E)$ *is equivalent to the family* $\{\gamma_x : x \in E\}$ *of openings such that*

(iv) $\gamma_x(\{x\}) = \{x\} \ \forall \ x \in E$

(v) $\forall \ A \subseteq E$, $\forall \ x, y \in E$, $\gamma_x(A)$ *and* $\gamma_y(A)$ *are either equal or disjoint.*

(vi) $\forall\ A \subseteq E, \forall\ x \in E,\ x \notin A \Rightarrow \gamma_x(A) = \emptyset.$

The dual operation of γ_x is the closing ρ_x, [31],

$$\rho_x(A) = E \backslash \gamma_x(A^c), A \in \mathcal{P}(E) \tag{3.6}$$

Indeed, $\rho_x(A^c) = E\backslash\gamma_x((A^c)^c) = E\backslash\gamma_x(A) = (\gamma_x(A))^c$. The point-connected closing ρ_x is a closing. It is increasing: If $A \subseteq B$, then

$$\begin{aligned} & \gamma_x(A) \subseteq \gamma_x(B) \\ \Rightarrow\ & (\gamma_x(A))^c \supseteq (\gamma_x(B))^c \\ \Rightarrow\ & \rho_x(A^c) \supseteq \rho_x(B^c) \\ & \text{where } B^c \subseteq A^c, \end{aligned}$$

it is extensive:

$$\begin{aligned} \gamma_x(A^c) &\subseteq A^c \\ (\rho_x(A))^c &\subseteq A^c \\ \text{so that } \rho_x(A) &\supseteq A, \end{aligned}$$

and ρ_x is idempotent:

$$\begin{aligned} \rho_x(\rho_x(A)) &= E\backslash\gamma_x((\rho_x(A))^c) \\ &= E\backslash\gamma_x(\gamma_x(A^c)) \\ &= E\backslash\gamma_x(A^c) \\ &= \rho_x(A). \end{aligned}$$

Figure 3.4 illustrates how the point-connected opening and closing work.

The concept of connectivity given in Definition 53 is applicable only for binary images, although it has been applied to grey scale and colour images, see [33] [100] [119]. Serra wanted to provide a unified concept which could be directly applied to whatever the application. He thus extended the connectivity class for lattices, specifically for lattices of functions. Recall from Chapter 2 that a complete lattice $\mathcal{L}$ is a set of ordered (partial or total) elements for which each family of elements possesses a supremum and an infimum. On binary images we consider the lattice of sets where the order is inclusion $\subset$, the supremum is the intersection $\cup$ and the infimum is the union $\cap$. On greyscale images we consider a lattice of functions with the order as the inequality $\leq$, the supremum as $\vee$ and the infimum as $\wedge$. We let the elements of $\mathcal{L}$ be denoted

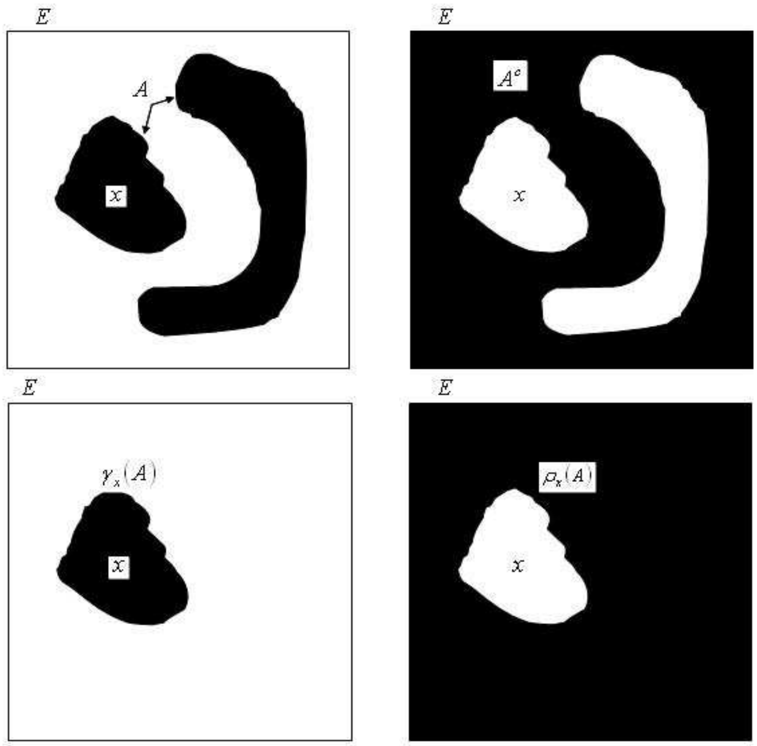


Figure 3.4. Operations of γ_x and ρ_x

by little letters $a, b, \ldots$ and families in $\mathcal{L}$ be denoted by capital letters $\mathcal{A}, \mathcal{B}, \ldots$. Let 0 be the smallest element in $\mathcal{L}$ and m be the largest element in $\mathcal{L}$. Then for $a \in \mathcal{L}$, let $M^a = \{x : x \in \mathcal{L}, x \leq a\}$ be the lower bound for a and $M_a = \{x : x \in \mathcal{L}, x \geq a\}$ be the upper bound for a.

Definition 55 *A family $\mathcal{X}$ in $\mathcal{L}$ is a* **sup-generator** *if for each $a \in \mathcal{L}$, a is the supremum of the elements of $\mathcal{X}$ that is majorates:*

$$a = \bigvee (\mathcal{X} \subset M^a) = \bigvee \{x \in \mathcal{X} : x \leq a.\}$$

We can now introduce the expanded definition for a connectivity class on lattices. The connectivity theory provided thus far is sufficient to allow for the extension in Section 3.3, so will only provide the expanded definition on lattices to indicate the full scope of the work that has been done by others.

Definition 56 *$\mathcal{C} \subset \mathcal{L}$ is a* **connectivity class** *or a* **connection** *on a lattice $\mathcal{L}$ if*

(I) $0 \in \mathcal{L}$

(II) $\mathcal{C}$ *is sup-generating for* $\mathcal{L}$*:* $\forall\ a \in \mathcal{L}, a = \bigvee \left(\mathcal{C} \bigcap M^a\right)$

(III) $\mathcal{C}$ *is conditionally closed under supremum:* $\mathcal{X} \subseteq \mathcal{C}, \bigwedge \mathcal{X} \neq 0 \Rightarrow \bigvee \mathcal{X} \in \mathcal{C}$*.*

Definitions 53 and 56 are equivalent. Indeed, axioms (1) and (3) in both cases are obvious, but axiom (2) needs some more thought. In Definition 56 axiom (II) implies that the connectivity class can represent any element of the whole lattice $\mathcal{L}$ as the supremum of elements belonging to the connectivity class and those which are lower bounds for the element. This brings to mind the concept of a dense subset. In Definition 53 axiom (ii) is more trivial, and simply implies that every individual element is in the connection, so that the connection makes up the whole space.

As before, we require axiom (III) so that for $C_x = \{c : x \leq c, c \in \mathcal{C}\} = \mathcal{C} \bigcap M_x$, the supremum of each nonempty family of elements in C_x is in $\mathcal{C}$. Thus the connected opening of origin x,

$$\gamma_x(a) = \bigvee \{c : c \in \mathcal{C}, x \leq c \leq a\}, a \in \mathcal{L}, \tag{3.7}$$

has invariant set $C_x \bigcup \{0\}$. We have a similar equivalence result as well.

Theorem 57 *If* $\mathcal{C}$ *is a sup-generator in* $\mathcal{L}$*, then* $\mathcal{C}$ *is a connectivity class* $\Longleftrightarrow$ $\mathcal{C}$ *coincides with the family* $\{\gamma_x : x \in \mathcal{C} \backslash \{0\}\}$ *of openings such that*

(IV) $\forall\ x \in \mathcal{C}\backslash\{0\},\ \gamma_x(x) = x$

(V) $\forall\ a \in \mathcal{L},\ \forall\ x, y \in \mathcal{C}\backslash\{0\},\ \gamma_x(a)$ *and* $\gamma_y(a)$ *are equal or disjoint*

(VI) $\forall\ a \in \mathcal{L},\ \forall\ x \in \mathcal{C}\backslash\{0\},\ x \not\leq a \Rightarrow \gamma_x(a) = 0.$

The connected sets that the LULU operators are applied to are the supports of certain sets so it is sufficient that we use the connectivity class presented in Definition 53. The extension of LULU operators for the connectivity in Definition 56 will be investigated in future work. For images we consider $E = \mathbb{Z}^2$ in Definition 53. However, a connectivity class may not contain sets of every size. For example, $\{\emptyset\} \cup \{\{x\} : x \in \mathbb{Z}^d\}$ and $\{\emptyset\} \cup \{\{x\} : x \in \mathbb{Z}^d\} \cup \{\mathbb{Z}^d\}$ are connections on $\mathbb{Z}^d$ but neither of them contain sets of finite size other than 0 and 1. In the definition of the operators L_n and U_n we need sets of every size. We therefore assume that $\mathbb{Z}^d$ is equipped with a connection $\mathcal{C}$ which satisfies the following conditions

- $\mathbb{Z}^d \in \mathcal{C}$ (3.8)
- For any $a \in \mathbb{Z}^d, E_a(C) \in \mathcal{C}$ whenever $C \in \mathcal{C}$ (3.9)
  so that $\mathcal{C}$ is translation invariant (3.10)
- If $V \subsetneq W,\ V, W \in \mathcal{C},$ then there exists $x \in W \setminus V$
  such that $V \cup \{x\} \in \mathcal{C}$ (3.11)

Condition 3.8 ensures that the whole space is connected, condition 3.9 ensures that the connected sets are translation invariant, and condition 3.11 ensures that the addition of a single point in the connection to a connected set preserves the connectivity. The aim of the conditions (3.8)–(3.11) is to define a connection which is sufficiently rich in connected sets. Suppose we have two connected sets V and W with $V \subsetneq W$. Then $k_1 = \text{card}(V) < k < \text{card}(W) = k_2$, for some k. By condition 3.11 there exists $x_1 \in W \backslash V$ such that $V \bigcup \{x_1\} \in \mathcal{C}$. Then $\text{card}(V \bigcup \{x_1\}) = k_1 + 1 \leq k$. Continue this application of condition 3.11 until you obtain $S = V \bigcup \{x_1\} \bigcup ... \bigcup \{x_m\}$ such that $\text{card}(S) = k_1 + 1 + ... + 1 = k < k_2$. Thus we have the following property:

$$\begin{array}{l} \text{Let } V \subsetneq W,\ V, W \in \mathcal{C}. \text{ For every } k \in \mathbb{N} \text{ such that} \\ \text{card}(V) < k < \text{card}(W) \text{ there exists } S \in \mathcal{C} \\ \text{such that } V \subseteq S \subseteq W \text{ and } \text{card}(S) = k. \end{array} \tag{3.12}$$

As usual, $\text{card}(V)$ is the number of the elements in the set V. Given a point $x \in \Omega$ and $n \in \mathbb{N}$ we denote by $\mathcal{N}_n(x)$ the set of all connected sets of size $n + 1$ that contain point x, that is,

$$\mathcal{N}_n(x) = \{V \in \mathcal{C} : x \in V,\ \text{card}(V) = n + 1\}. \tag{3.13}$$

In addition to (3.8) – (3.11) we assume that the connection $\mathcal{C}$ is such that

$$\text{card}(\mathcal{N}_n(x)) < \infty,\ \forall n \in \mathbb{N},\ \forall x \in \mathbb{Z}^d. \tag{3.14}$$

We should note that if a connection on $\mathbb{Z}^d$ is defined via graph connectivity, where all vertices are of finite degree, then it is trivial to see that properties (3.11), (3.13) and (3.14) hold automatically. In image analysis ($d = 2$) the connectivity is a graph connectivity defined via a neighbour relation, e.g. 4-connectivity, 8-connectivity, see Figure 3.5.

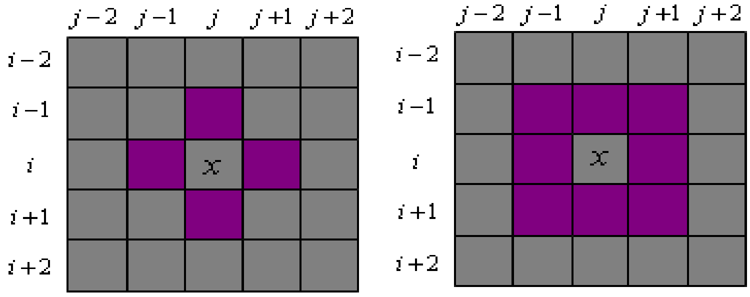


Figure 3.5. 4-Connectivity and 8-Connectivity.

In this case all properties (3.8) - (3.13) and (3.14) hold. However, in order to have maximum generality for the extension of the LULU operators to higher dimensions we adopt the present axiomatic approach with $\mathbb{Z}^d$ as the underlying space and with an arbitrary connection $\mathcal{C}$ satisfying the axioms in Definition 53 and the additional axioms (3.8)-(3.14).

3.2.1 Extended Results for Connectivity Classes

In addition to the Connectivity class introduced in Definitions 53 and 56, a number of additional concepts have been investigated in the literature. These are mostly for use in image processing. In [12], [13], [16], [17] and [18], σ-connectivity and multiscale connectivity are discussed. If the degree of connectivity with respect to a connectivity measure is at least σ then σ-connectivity holds.

In [112] and [30] constrained connectivity is investigated. Here, two pixels are considered connected if they satisfy a series of constraints. In [15] a type of greyscale connectivity is introduced. A greyscale image is considered connected if all the level sets below a predefined threshold are connected, termed level-k connectivity. In [14], fuzzy τ-connectivity is discussed, in [107] and [14] hyperconnectivity and hypoconnectivity, in [97] partial connectivity, and in [102] pseudo-connectivity.

3.3 The Multidimensional LULU Operators

We are now ready to define the operators L_n and U_n on $\mathcal{A}(\mathbb{Z}^d)$ as follows. Recall that $\mathcal{A}(\mathbb{Z}^d)$ is the vector lattice of all real functions defined on $\mathbb{Z}^d$ with respect to the usual point-wise defined addition, scalar multiplication and partial order (see Section 3.1).The extension to an arbitrary domain Ω should also be looked at in the future.

Definition 58 *Let $f \in \mathcal{A}(\mathbb{Z}^d)$ and $n \in \mathbb{N}$. Then*

$$L_n(f)(x) = \max_{V \in \mathcal{N}_n(x)} \min_{y \in V} f(y), \ \ x \in \mathbb{Z}^d, \tag{3.15}$$

$$U_n(f)(x) = \min_{V \in \mathcal{N}_n(x)} \max_{y \in V} f(y), \ \ x \in \mathbb{Z}^d. \tag{3.16}$$

Let us confirm that Definition 58 generalizes the definition of L_n and U_n for sequences. Suppose $d = 1$ and let $\mathcal{C}$ be the connection on $\mathbb{Z}$ generated by the pairs of consecutive numbers. Then all connected sets on $\mathbb{Z}$ are sequences of consecutive integers and for any $i \in \mathbb{Z}$ we have

$$\mathcal{N}_n(i) = \{\{i-n, i-n+1, ..., i\}, \{i-n+1, i-n+2, ..., i+1\}, ..., \{i, i+1, ..., i+n\}\}$$

Hence for an arbitrary sequence considered as a function on $\mathbb{Z}$ the formulas (3.15) and (3.16) are reduced to Definition 29.

The operators L_n and U_n in Definition 58, as well as their one-dimensional counterparts (see Definition 29), can also be presented in the general setting of Mathematical Morphology. Within this theory L_n is an area opening and U_n is an area closing, where the area of a set refers to the number of points in it. Let us recall that a morphological opening (closing) is a composition of an erosion and a dilation (dilation and erosion) with a specified structuring element, see Definitions 34 and 35. We introduce the area opening and closing, [111] to show the relationship between them and the LULU operators.

Definition 59 *An **area opening** is the union of all openings with connected structuring elements whose size (area) is λ pixels, namely*

$$\gamma_\lambda = \bigvee_i \{\gamma_{B_i} | B_i \textit{ is connected and } area(B_i) = \lambda\}.$$

An area opening removes all connected components whose area is smaller than λ. The area closing is the dual operation of the area opening, so that $\gamma_\lambda(f) = -\rho_\lambda(-f)$, thus it removes the same connected components in the background (that is those negative of the input).

Definition 60 *An* ***area closing*** *is given by*

$$\rho_\lambda = \bigwedge_i \{\rho_{B_i} | B_i \text{ is connected and } area(B_i) = \lambda\}.$$

It is important to notice that here the structuring elements can take on any shape as the only restriction is on their size. This is the important difference between the LULU operators, which are only concerned about size, and morphological filters, which operate in conjunction with a specified structuring element with a specified size. In the situation of an application which requires a specifically shaped structuring element, this can be taken into account in the Discrete Pulse Decomposition as all the connected components are obtained and their shapes or shape similarity can be determined.

Due to the simple structure of $\mathbb{Z}$ the one-dimensional L_n and U_n are morphological opening and closing, respectively, where the structuring element is a line segment of length $n+1$. This is trivial to see. However, the segments are the only connected sets under the considered connectivity and the considered dimension. In this sense, all connected sets have the same shape. One may consider morphological opening and closing as a generalization of the one-dimensional L_n and U_n to operators on $\mathcal{A}(\mathbb{Z}^d)$. However, an essential property of L_n and U_n for sequences is that they form a Matheron pair [95], that is we have

$$L_n \circ U_n \circ L_n = U_n \circ L_n \text{ and } U_n \circ L_n \circ U_n = L_n \circ U_n. \tag{3.17}$$

It is easy to see by examples that for a general structuring element on $\mathbb{Z}^d, d > 1$, one can find $f \in \mathcal{A}(\mathbb{Z}^d)$ such that (3.17) is violated. For example the morphological opening and closing by a structuring element B do not in general satisfy equation 3.17.

This motivates the proposed definition of L_n and U_n, when $d > 1$, which is independent of shape. We obtain the proof of (3.17) from known results on attribute filters of which L_n and U_n are particular cases, [32]. Firstly, an attribute opening is defined in terms of a trivial opening, [19]. We first provide the definitions for binary images and then show the extensions to grayscale images.

Definition 61 *For a connected set $C \subset \Omega$ and an increasing criterion T, a* **trivial opening** Γ_T *is given by*

$$\Gamma_T(C) = \begin{cases} C & \text{if } C \text{ satisfies criterion } T \\ \emptyset & \text{otherwise} \end{cases}$$

and $\Gamma_T(\emptyset) = \emptyset$.

Examples of an increasing criterion are (1) C must have an area (number of pixels) of at least λ, (2) the diagonal of the minimum enclosing rectangle in a given direction must be of length at least λ, and (3) the area of the largest circle that can fit inside the region must be at least λ. So these illustrate that a criterion T is increasing if every superset of C also satisfies T if C satisfies T. The maximum geodesic distance, [111], of the connected region is an example of a nonincreasing criterion.

Definition 62 *The* **geodesic distance** *between two pixels p and q in a connected set A is,*

$$d_A(p,q) = \min\{L(\mathcal{P}) : p_1 = p, p_\ell = q \text{ and } \mathcal{P} \subseteq A\},$$

the minimum of the length L of path(s) $\mathcal{P} = (p_1, p_2, ..., p_\ell)$ joining p and q in A.

Indeed, as discussed in [19], any criterion that is used to define shape is nonincreasing. Thus it is more adequate to use criterions which involve size. This is an important observation in relation to the LULU operators. It concretes the method of using only the size of the connected neighbourhoods and at a later stage (after the Discrete Pulse decomposition has been applied) shape attributes can be taken into account as well. For the increasing criterion given in (1) above, one obtains the area opening, see (59), [125] [126]. An attribute opening preserves only those connected regions that satisfy the criterion T.

Definition 63 *For any set $X \subset \Omega$ and an increasing criterion T, the* **attribute opening** *of X is given by,*

$$\Gamma^T(X) = \bigcup_{x \in X} \Gamma_T\big(\Gamma_x(X)\big)$$

where Γ_x is the connected opening from Equation 3.5.

It is proven in [19] that an attribute opening is indeed an algebraic opening, see Definition 20, so that it is increasing, idempotent and anti-extensive. Because attribute openings work wholely on the connected components they preserve the shape of regions. An attribute thinning is defined similarly in [19]. First an trivial thinning.

Definition 64 *For a connected set $C \subset \Omega$ and a criterion T (not necessarily increasing), a* **trivial thinning** Φ_T *is given by*

$$\Phi_T(C) = \begin{cases} C & \text{if } C \text{ satisfies criterion } T \\ \emptyset & \text{otherwise} \end{cases}$$

and $\Phi_T(\emptyset) = \emptyset$.

Note that the only difference when compared to the trivial opening is that the criterion is not required to be increasing for a trivial thinning. Then an attribute thinning is defined by the following.

Definition 65 *For any set $X \subset \Omega$ and any criterion T, the* **attribute thinning** *of X is given by,*

$$\Phi^T(X) = \bigcup_{x \in X} \{\Phi_T\big(\Gamma_x(X)\big)$$

where Γ_x is the connected opening from Equation 3.5.

The attribute thinning is not an algebraic closing but it does satisfy the two properties of idempotence and anti-extensivity. This is not ideal as we would like an algebraic closing to correlate with the attribute opening. This is simple to obtain by taking duals. Thus we define an attribute closing as follows.

Definition 66 *An* **attribute closing** Ψ^T *is defined as the dual of the attribute opening, that is*

$$\Psi^T(X) = \big(\Gamma_T(X^C)\big)^C.$$

The attribute closing is then an algebraic closing, see Definition 19.

For gray-scales images the corresponding definitions of the preceding concepts are as follows, [19].

Definition 67 *For a grayscale image f and an increasing criterion T, the* **grayscale attribute opening** *is given by*

$$\gamma^T(f)(x) = \max\{t : x \in \Gamma_T\big[\mathcal{H}_t(f)\big]\}$$

where $\mathcal{H}_t(f) = \{x \in \Omega : f(x) \geq t\}$ is the threshold set.

Definition 68 *For a grayscale image f and any criterion T, the* **grayscale attribute thinning** *is given by*

$$\phi^T(f)(x) = \max\{t : x \in \Phi_T\big[\mathcal{H}_t(f)\big]\}$$

where $\mathcal{H}_t(f) = \{x \in \Omega : f(x) \geq t\}$ is the threshold set.

Definition 69 *A* **grayscale attribute closing** *is defined as the dual of the grayscale attribute opening,*

$$\psi^T(f)(x) = -\gamma^T(-f)(x).$$

We can now present the very important result introduced in (3.17). We first prove the duality of the LULU operators and the ordering of the elements $\{L_n, U_n, L_n \circ U_n, U_n \circ L_n, U_n \circ L_n \circ U_n, L_n \circ U_n \circ L_n\}$.

Theorem 70 *The operators L_n and U_n are duals.*

Proof
For an arbitrary $f \in \mathcal{A}(\mathbb{Z}^d)$,

$$\begin{aligned} L_n(-f) &= \max_{V \in \mathcal{N}_n(x)} \min_{y \in V}(-f(y)) \\ &= \max_{V \in \mathcal{N}_n(x)} \big(-\max_{y \in V} f(y)\big) \\ &= -\min_{V \in \mathcal{N}_n(x)} \max_{y \in V} f(y) \\ &= -U_n(f). \end{aligned}$$

■

Corollary 71 *The operators $L_n \circ U_n$ and $U_n \circ L_n$ are duals.*

Proof
For an arbitrary $f \in \mathcal{A}(\mathbb{Z}^d)$,

$$\begin{aligned} U_n \circ L_n(-f) &= U_n \circ (-U_n(f)) \\ &= -L_n \circ U_n(f). \end{aligned}$$

■

The area opening and closing are then special cases of the attribute opening and closing and are therefore algebraic openings and closings respectively, [111]. Thus since the operators L_n and U_n are respectively an area opening and closing, the following holds,

$$\text{Increasingness: } f \leq g \Longrightarrow (\ L_n(f) \leq L_n(g),\ U_n(f) \leq U_n(g)\ ) \tag{3.18}$$

$$\text{Idempotence: } L_n \circ L_n = L_n, \qquad U_n \circ U_n = U_n \tag{3.19}$$

$$\text{Anti-Extensivity and Extensivity: } L_n(f) \leq f \leq U_n(f) \tag{3.20}$$

An alternative for the idempotence of L_n and U_n can be proven directly as well. The inequality

$$L_n \circ L_n \leq L_n$$

is an immediate consequence of (3.20). Then it is sufficient to prove the inverse inequality. Let $f \in \mathcal{A}(\mathbb{Z}^d)$ and $x \in \mathbb{Z}^d$. We have

$$L_n(L_n(f))(x) = \max_{W \in \mathcal{N}_n(x)} \min_{y \in W} \max_{V \in \mathcal{N}_n(y)} \min_{z \in V} f(z). \tag{3.21}$$

But $y \in W \in \mathcal{N}_n(x)$ implies $W \in \mathcal{N}_n(y)$. Therefore for every $W \in \mathcal{N}_n(x)$ and $y \in W$ we have

$$\max_{V \in \mathcal{N}_n(y)} \min_{z \in V} f(z) \geq \min_{z \in W} f(z).$$

Using that the right hand side is independent of y we further obtain

$$\min_{y \in W} \max_{V \in \mathcal{N}_n(y)} \min_{z \in V} \geq \min_{z \in W} f(z),\ \ W \in \mathcal{N}_n(x).$$

Then it follows from the representation (3.21) that

$$L_n(L_n(f))(x) \geq \max_{W \in \mathcal{N}_n(x)} \min_{z \in W} f(z) = L_n(f)(x).$$

This holds similarly for U_n. A similar method of proof is provided in [95, Corollary 28]. Here we see that L_n is idempotent if and only if

$$\forall\ x \in \mathbb{Z}^d, \forall\ W \in \mathcal{N}_n(x), \forall\ y \in W, \text{ then } W \in \mathcal{N}_n(y).$$

Furthermore, it is easy to see that these operators are monotone with respect to n.

Theorem 72 *For $n_1 < n_2$ we have that*

$$L_{n_1} \geq L_{n_2} \text{ and } U_{n_1} \leq U_{n_2}. \tag{3.22}$$

Proof
It follows from (3.13) that for every $x \in Z^d$ and $V \in \mathcal{N}_{n_2}(x)$ there exists a set $W \in \mathcal{N}_{n_1}(x)$ such that $W \subseteq V$. Therefore

$$\min_{y \in V} f(y) \leq \min_{y \in W} f(y) \leq \max_{S \in \mathcal{N}_{n_1}(x)} \min_{y \in S} f(y) = L_{n_1}(f)(x).$$

Hence

$$L_{n_2}(f)(x) = \max_{V \in \mathcal{N}_{n_2}(x)} \min_{y \in V} f(y) \leq L_{n_1}(f)(x), x \in \mathbb{Z}^d.$$

The inequality for U_n is proved in a similar way. ■

For next result we require the d-dimensional median operator. In one dimension the median smoother is given by

$$M_n(x)_i = median\{x_{i-n}, ..., x_i, ..., x_{i+n}\}. \tag{3.23}$$

Notice that the neighbourhood is $\{x_{i-n}, ..., x_i, ..., x_{i+n}\}$ which includes all the n-neighbourhoods of x_i, namely $\{x_{i-n}, ..., x_i\}$, $\{x_{i-n+1}, ..., x_{i+1}\}$, ..., $\{x_i, ..., x_{i+n}\}$, which are involved in the operators U_n and L_n in one dimension. The median smoother in two dimensions has been investigated in [36],

$$M(f)(x_{ij}) = median\{f(x_{i,j-1}), f(x_{i,j+1}), f(x_{ij}), f(x_{i-1,j}), f(x_{i+1,j})\}. \tag{3.24}$$

The formulation in (3.24) does not allow for variation in the neighbourhood size over $n \in \mathbb{N}$. Another alternative is presented in [83],

$$M(f)(x_{ij}) = median\{f(x_{i+r,j+s}) : (r, s) \in I\} \tag{3.25}$$

where $I = \{(r, s) \in \mathbb{Z}^2\} \subset \mathbb{Z}^2$. The formulation in (3.25) doesn't specify the neighbourhood over which the median is taken, that is the index set I, though. A possible neighbourhood is investigated in [55]. Here they use a neighbourhood of size $n \times m$ (where m, n are odd integers) and the median operation is the median of the gray levels of the picture elements lying in this $m \times n$ neighbourhood with the neighbourhood centered at the element x_{ij}. This allows for different neighbourhood sizes but is still restrictive since ideally the neighbourhood formulation should depend on the connectivity used. We thus propose the following formulation for the d-dimensional median operator, which incorporates the requirement that the neighbourhood should contain every possible connected set in $\mathcal{N}_n(x)$.

Definition 73 *For $f \in \mathcal{A}(\mathbb{Z}^d)$, the* **$d$-dimensional median smoother** *is given by,*

$$M_n(f)(x) = median\{f(y) : y \in B\}$$

where $B = \cup_V \{V \in \mathcal{N}_n(x)\}$.

Definition 73 is equivalent to (3.23) since then

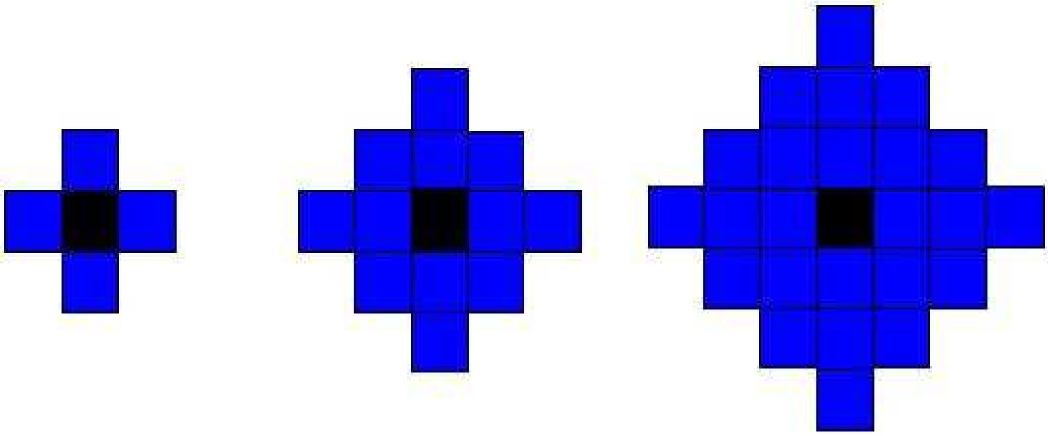
$$\begin{aligned} B &= \bigcup_V \{V \in \mathcal{N}_n(x_i)\} \\ &= \bigcup \{\{x_{i-n}, ..., x_i\}, \{x_{i-n+1}, ..., x_{i+1}\}, ..., \{x_i, ..., x_{i+n}\}\} \\ &= \{x_{i-n}, ..., x_i, ..., x_{i+n}\}. \end{aligned}$$

In two dimensions, using 4-connectivity, the neighbourhoods for $n = 1, 2, 3$ are given in Figure 3.6.

We can show that for each $n \in \mathbb{N}$ we have that $U_n \circ L_n \leq M_n \leq L_n \circ U_n$, similar to the one dimensional case presented in [93, Theorem 3.4]. This would prove the inequality

$$U_n \circ L_n \leq L_n \circ U_n. \tag{3.26}$$

We rather use the next result to prove (3.26).



(a) (b) (c)

Figure 3.6. Neighbourhoods for the d-dimensional median operator for (a) $n = 1$ (b) $n = 2$ (c) $n = 3$.

Theorem 74 *For $n \in \mathbb{N}$,*

$$L_n \circ U_n \circ L_n = U_n \circ L_n \text{ and } U_n \circ L_n \circ U_n = L_n \circ U_n. \tag{3.27}$$

Proof It follows from (3.20) that

$$L_n \circ U_n \circ L_n \le id \circ U_n \circ L_n = U_n \circ L_n. \tag{3.28}$$

Assume that $L_n \circ U_n \circ L_n = U_n \circ L_n$ is violated. In view of (3.28), this means that there exists $f \in \mathcal{A}(\mathbb{Z}^d)$ and $z \in \mathbb{Z}^d$ such that

$$L_n(U_n(L_n(f)))(z) < U_n(L_n(f))(z).$$

It follows from Theorem 78 (which follows in Section 3.4) that there exists $k \le n$ and $V \in \mathcal{N}_k(z)$ such that V is a local maximum set for $U_n(L_n(f))(z)$. Then, by Theorem 81, there exists $W \subseteq V$ such that W is a local maximum set of the function $L_n(f)$. We have card$(W) \le k \le n$. However, $L_n(f)$ does not have any local maximum sets of size less than or equal to n, see Theorem 79. This contradiction completes the proof. The second equality is proven in a similar manner. ■

The result

$$U_n \circ L_n \le L_n \circ U_n \iff \big(L_n \circ U_n \circ L_n = U_n \circ L_n \text{ and } U_n \circ L_n \circ U_n = L_n \circ U_n\big) \tag{3.29}$$

can be easily proven, see [93, Theorem 2.9]. We then also have that $U_n \circ L_n \le L_n \circ U_n$. The main consequence of L_n and U_n comprising a Matheron pair, as originally introduced by Matheron [78], is that L_n, U_n and all their compositions form a four element semi-group with respect to composition. This will be dealt with in the next section in more detail.

Next we relate the multidimensional U_n and L_n to the concept of a separator given in Definition 48. Indeed, conditions (i), (ii) and (iii) of Definition 48 hold for all openings and closings, [111, Chapter 2]. The idempotence was given in (3.19). Thus only the co-idempotence remains. We can remark that co-idempotence is seldom discussed in the standard literature on Mathematical Morphology. However, L_n and U_n are also min-max operators as defined by Wild, [135], since they are respectively a morphological opening and closing. As such, their co-idempotence follows from [135, Corollary 11]. Therefore,

$$L_n, U_n \text{ are separators for every } n \in \mathbb{N}. \tag{3.30}$$

3.4 The LULU Semigroup

Due to Theorem 74 a four-element semi-group, similar to the one shown in Chapter 2, is obtained

$$\{U_n, L_n, U_n \circ L_n, L_n \circ U_n\}.$$

The composition table is given in Table 2.1. Moreover, it follows from (3.18) - (3.20) that this semi-group is fully ordered as it is in the one-dimensional case i.e. we have

$$L_n \leq U_n \circ L_n \leq L_n \circ U_n \leq U_n. \quad (3.31)$$

Similar to their counterparts for sequences the operators, the multidimensional operators L_n and U_n smooth the input function by removing sharp peaks (the application of L_n) and deep pits (the application of U_n). The smoothing effect of these operators is made more precise by using the concepts of a local maximum set and a local minimum set given below.

Definition 75 *Let $V \in \mathcal{C}$. A point $x \notin V$ is called* **adjacent** *to V if $V \cup \{x\} \in \mathcal{C}$. The set of all points adjacent to V is denoted by $\mathrm{adj}(V)$, that is,*

$$\mathrm{adj}(V) = \{x \in \mathbb{Z}^d : x \notin V, V \cup \{x\} \in \mathcal{C}\}.$$

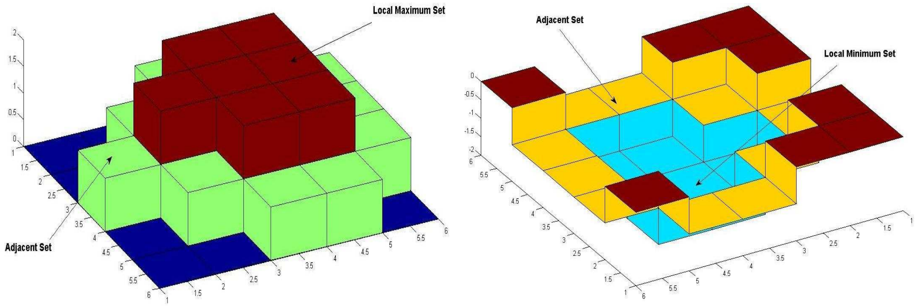
An equivalent formulation of the property (3.11) of the connection $\mathcal{C}$ is as follows:

$$V, W \in \mathcal{C},\ W \subsetneq V \implies \mathrm{adj}(W) \cap V \neq \emptyset. \quad (3.32)$$

We introduce the following simple Lemma.

Lemma 76 *a) Given $V, W \in \mathcal{C}$ with $W \subset V$. Then for $x \notin V$ but $x \in \mathrm{adj}(W)$, we have $x \in \mathrm{adj}(V)$.*
b) Given distinct $V, W \in \mathcal{C}$ with $V \cap W \neq \emptyset$, there exists an $x \in V \backslash W$ such that $x \in \mathrm{adj}(W)$.

Proof
a) V and $W + \{x\}$ are connected and have a nonempty intersection thus their union $V \cup \{x\}$ is also connected. Then by Definition 75, $x \in \mathrm{adj}(V)$.
b) Applying (3.32) to $V \cup W$, since $V \subsetneq V \cup W$, we get $\mathrm{adj}(W) \cap V =$



(a) (b)

Figure 3.7. (a) A Local Maximum Set (b) A Local Minimum Set

$\text{adj}(W) \cap (V \cup W) \neq \emptyset$. Thus there exists $x \in V \backslash W$ such that $x \in \text{adj}(W)$. ■

It follows further from condition (3.14) that

$$\text{card}(V) < \infty \implies \text{card}(\text{adj}(V)) < \infty. \tag{3.33}$$

Indeed if $\text{card}(V) = n$ then for an arbitrary $x \in V$ we have $\{\{a\} \cup V : a \in \text{adj}(V)\} \subset \mathcal{N}_{n+1}(x)$, so that $\text{card}(\text{adj}(V)) \leq \text{card}(\mathcal{N}_{n+1}(x)) < \infty$.

Definition 77 *A connected subset V of $\mathbb{Z}^d$ is called a* **local maximum set** *of $f \in \mathcal{A}(\mathbb{Z}^d)$ if*

$$\sup_{y \in \text{adj}(V)} f(y) < \inf_{x \in V} f(x).$$

Similarly V is a **local minimum set** *if*

$$\inf_{y \in \text{adj}(V)} f(y) > \sup_{x \in V} f(x).$$

The next four theorems deal with different aspects of the application of L_n and U_n to functions in $\mathcal{A}(\mathbb{Z}^d)$. They are followed by a discussion on their cumulative effect. All theorems contain statements a) and b). Due to the similarity we present only the proofs of a).

Theorem 78 *Let $f \in \mathcal{A}(\mathbb{Z}^d)$ and $x \in \mathbb{Z}^d$. Then we have*
a) $L_n(f)(x) < f(x)$ if and only if there exists a local maximum set V of f such that $x \in V$ and $\text{card}(V) \leq n$;
b) $U_n(f)(x) > f(x)$ if and only if there exists local minimum set V of f such that $x \in V$ and $\text{card}(V) \leq n$.

Proof

a) Implication to the left. Suppose that there exists a local maximum set $V \in \mathcal{N}_k(x)$, $k < n$. Consider an arbitrary $W \in \mathcal{N}_n(x)$ and let S be a connected component of $W \cap V$. Then $W \backslash V \neq \emptyset$ since $\mathrm{card}(S) < \mathrm{card}(W)$ by (3.13) and by (3.32) we have $\mathrm{adj}(S) \cap W \neq \emptyset$. Let $z \in \mathrm{adj}(S) \cap W$. If $z \in V$ then this means $z \in V \cup W$, and so $S \cup \{z\}$ is connected. This is a contradiction on S being a connected component of $V \cap W$. So $z \notin V$. Then using also that V is a local maximum set we obtain

$$\min_{y \in W} f(y) \leq f(z) < \min_{t \in V} f(t) \leq f(x).$$

Since the set $W \in \mathcal{N}_n(p)$ is arbitrary, this inequality implies that $L_n(f)(x) < f(x)$.

Implication to the right. Suppose $L_n(f)(x) < f(x)$. Let V be the greatest (in terms of $\subseteq$) connected set containing x such that

$$f(y) \geq f(x), \ \forall \ y \in V. \tag{3.34}$$

The set V is obviously unique and can be constructed as $V = \gamma_x(Y)$, where γ_x is the morphological point connected opening generated by x, see [106] or [108], and $Y = \{y \in \mathbb{Z}^d : f(y) \geq f(x)\}$.

Assume that $\mathrm{card}(V) > n$. It follows from (3.13) that there exists $W \in \mathcal{N}_n(x)$ such that $W \subset V$. Then

$$L_n(f)(x) = \max_{S \in \mathcal{N}_n(x)} \min_{y \in S} f(y) \geq \min_{y \in W} f(y) \geq \min_{y \in V} f(y) = f(x).$$

This contradicts the assumption $L_n(f)(x) < f(x)$. Therefore, $\mathrm{card}(V) \leq n$.

We have $f(z) < f(x)$, for all $z \in \mathrm{adj}(V)$, because otherwise (3.34) is satisfied on the larger connected set $\{z\} \cup V$. Then, also using (3.33), we obtain

$$\max_{z \in \mathrm{adj}(V)} f(z) < f(x) = \min_{y \in V} f(y).$$

Hence V is a local maximum set. ■

Theorem 79 *Let $f \in \mathcal{A}(\mathbb{Z}^d)$. Then*
a) the size of any local maximum set of the function $L_n(f)$ is larger than n;
b) the size of any local minimum set of the function $U_n(f)$ is larger than n.

Proof
a) Assume the opposite, that is, there exists a local maximum set V of $L_n(f)$ such that $\mathrm{card}(U) \leq n$. By Theorem 78 we have that

$$L_n(L_n(f))(x) < L_n(f)(x),\ x \in V.$$

Since L_n is idempotent, see (3.19), this implies the impossible inequality $L_n(f)(x) < L_n(f)(x)$, which completes the proof. ∎

Theorem 80 *Let $V \in \mathcal{C}$ and let $x \in \mathrm{adj}(V)$.*
a) If $f(x) \leq \inf_{y\in V} f(y)$ then $L_n(f)(x) \leq \inf_{y\in V} L_n(f)(y)$;
b) If $f(x) \geq \sup_{y\in V} f(y)$ then $U_n(f)(x) \geq \sup_{y\in V} U_n(f)(y)$.

Proof
a) For any $W \in \mathcal{N}_n(x)$ the set $W \cup V$ is connected and of size at least $n+1$. Therefore, by (3.13), for every $y \in V$ there exists $S_y \in \mathcal{N}_n(y)$ such that $S_y \subset W \cup V$. Then, using also the given inequality and since $\inf_{z\in V\cup\{x\}} f(z) = f(x)$, for every $y \in V$ and $W \in \mathcal{N}_n(x)$ we have

$$\min_{z\in W} f(z) = \inf_{z\in W\cup V} f(z) \leq \min_{z\in S_y} f(z) \leq L_n(f)(y).$$

Hence

$$L_n(f)(x) = \max_{W\in\mathcal{N}_n(x)} \min_{z\in W} f(z) \leq \inf_{y\in V} L_n(f)(y).$$

∎

Theorem 81 *Let $f \in \mathcal{A}(\mathbb{Z}^d)$ and let V be a finite connected set.*
a) If V is a local minimum set of $L_n(f)$ then there exists a local minimum set W of f such that $W \subseteq V$.
b) If V is a local maximum set of $U_n(f)$ then there exists a local maximum set W of f such that $W \subseteq V$.

Proof
a) Let V be a finite connected set which is a local minimum set of $L_n(f)$. Then by (3.33) the set $\mathrm{adj}(V)$ is finite and we have

$$\min_{y\in \mathrm{adj}(V)} f(y) \geq \min_{y\in \mathrm{adj}(V)} L_n(f)(y) > L_n(f)(x)\ \forall\ x \in V.$$

Let $q \in \text{adj}(V)$ be such that $f(q) = \min_{y \in \text{adj}(V)} f(y)$ and let

$$Y = \{y \in V : f(y) < f(q)\}.$$

Note that $Y \neq \emptyset$, since if $Y = \emptyset$, then $f(q) \leq \inf_{y \in V} f(y)$ and by Theorem 80 we have

$$L_n(f)(q) \leq \inf_{y \in V} L_n(f)(y).$$

This is a contradiction since V is a local minimum set of $L_n(f)$. Nevertheless, this result is an essential ingredient of the proof not least due to the fact that this is the only point where we use that V is a local minimum set of $L_n(f)$. Let $t \in Y$ and let W be the largest connected component of Y containing t so that $W = \gamma_t(Y)$ as in the proof of Theorem 78. For every $z \in \text{adj}(W)$, since $W \subseteq V$ then $z \in \text{adj}(V)$ as well, by Lemma 76. So we have $f(z) \geq f(q) > \max_{y \in W} f(y)$. Therefore W is a local minimum set of f. ■

Theorems 78–81 provide the following characterization of the effect of the operators L_n and U_n on a function $f \in \mathcal{A}(\mathbb{Z}^d)$:

1. The application of L_n (U_n) removes local maximum (minimum) sets of size smaller or equal to n.
2. The operator L_n (U_n) does not affect the local minimum (maximum) sets in the sense that such sets may be affected only as a result of the removal of local maximum (minimum) sets. However, no new local minimum (maximum) sets are created where there were none. This does not exclude the possibility that the action of L_n (U_n) may enlarge existing local minimum (maximum) sets or join two or more local minimum (maximum) sets of f into one local minimum (maximum) set of $L_n(f)$ ($U_n(f)$).
3. $L_n(f) = f$ ($U_n(f) = f$) if and only if f does not have local maximum (minimum) sets of size n or less.

Furthermore, as an immediate consequence of Theorem 79 and Theorem 81 we obtain the following corollary.

Corollary 82 *For every $f \in \mathcal{A}(\mathbb{Z}^d)$ the functions $(L_n \circ U_n)(f)$ and $(U_n \circ L_n)(f)$ have neither local maximum sets nor local minimum sets of size n or less. Furthermore,*

$$(L_n \circ U_n)(f) = (U_n \circ L_n)(f) = f$$

if and only if f does not have local maximum sets or local minimum sets of size less than or equal to n.

Theorem 83 *For $f \in \mathcal{A}(\mathbb{Z}^d)$,*
a) $L_n(f)$ is constant on any local maximum set W of f with card$(W) \leq n+1$
b) $U_n(f)$ is constant on any local minimum set W of f with card$(W) \leq n+1$

Proof
We only prove (a). Result (b) is proven by duality. Let Let W be a local maximum set of f with $\text{card}(W) \leq n+1$ and take arbitrary $p, q \in W$. Consider $V \in \mathcal{N}_n(p)$ such that $V \neq W$. Then $V \cup W$ is connected and so by Lemma 76(b) there exists $x \in V$ with $x \in \text{adj}(W)$. Since W is a local maximum set of f we have

$$f(x) \leq \inf_{z \in W} f(z),$$

and hence

$$\inf_{z \in V} f(z) = \inf_{z \in V \cup W} f(z).$$

By (3.13) there exists $U \in \mathcal{N}_n(q)$ such that $U \subset V \cup W$. Thus

$$\inf_{z \in V} f(z) = \inf_{z \in V \cup W} f(z) \leq \inf_{z \in U} f(z),$$

and since $U \in \mathcal{N}_n(q)$ we have

$$\inf_{z \in V} f(z) \leq \inf_{z \in U} f(z) \leq L_n(f)(q). \tag{3.35}$$

We we consider $V \in \mathcal{N}_n(p)$ such that $V = W$, we have

$$\inf_{z \in V} f(z) = \inf_{z \in W} f(z) \leq L_n(f)(q)$$

since $q \in W$ means that $W \in \mathcal{N}_n(q)$. Thus since (3.35) holds for all $V \in \mathcal{N}_n(p)$ we have $L_n(f)(p) \leq L_n(f)(q)$. By interchanging the role of p and q we obtain the other inequality and thus have equality. ■

We should remark that in the one dimensional setting, the sequences without local maximum sets or local minimum sets of size less than or equal to n are exactly the so-called n-monotone sequences. Hence Corollary 82 generalizes the respective results in the LULU theory of sequences, [93, Theorem 3.3].

3.5 Preservation Properties of the LULU Semi-group

The preservation of shape presented in Theorem 78 to Theorem 81 can be made more precise by generalising to $\mathcal{A}(\mathbb{Z}^d)$ the concepts of neighbour trend preserving and fully trend preserving introduced in [93, Chapter 6] for sequences.

Definition 84 *An operator P is* **neighbour trend preserving** *if for any points $p, q \in \Omega$, such that $\{p, q\} \in \mathcal{C}$, and for $f \in \mathcal{A}(\mathbb{Z}^d)$ we have*

$$f(p) \leq f(q) \implies P(f)(p) \leq P(f)(q).$$

The operator P is **fully trend preserving** *if both P and $I-P$ are neighbour trend preserving.*

In Definition 84, for P to be fully trend preserving the requirement on $I-P$, that is the neighbour trend preserving property, can be equivalently formulated as:

$$|P(f)(p) - P(f)(q)| \leq |f(p) - f(q)|. \tag{3.36}$$

In the context of sequences the property (3.36) is called **difference reducing**.

Theorem 85 *The operators L_n, U_n, $n = 1, 2, ...,$ and their compositions, are all fully trend preserving.*

Proof We prove the result for L_n. The case for U_n is dealt with similarly. Furthermore, it is easy to obtain that compositions of fully trend preserving operators are fully trend preserving, which proves the rest of the theorem, see [93, Theorem 6.10].

Since, the neighbour trend preserving property of L_n follows directly from Theorem 80, we only need to prove the neighbour trend preserving property of $I - L_n$ or equivalently, the inequality (3.36). Consider $p, q \in \mathbb{Z}^d$ such that $\{p, q\} \in \mathcal{C}$. We may assume without loss of generality that $f(p) \geq f(q)$. Then $L_n(f)(p) \geq L_n(f)(q)$ by the neighbour trend preservation. By (3.20) we have either (i) $L_n(f)(q) = f(q)$ or (ii) $L_n(f)(q) < f(q)$. If (i) holds then,

$$L_n(f)(p) - L_n(f)(q) = L_n(f)(p) - f(q) \leq f(p) - f(q),$$

again due to (3.20), so that $|L_n(f)(p) - L_n(f)(q)| \leq |f(p) - f(q)|$. If (ii) holds, by Theorem 78, q must belong to the support of a local maximum set, say W, of size at most n of f. Since $f(p) \geq f(q)$ and $\{p,q\} \in \mathcal{C}$, q must also belong to the support of W. This means that by Theorem 83 we have $L_n(f)(p) = L_n(f)(q)$ so that

$$|L_n(f)(p) - L_n(f)(q)| = 0 \leq |f(p) - f(q)|.$$

So L_n is difference reducing and thus fully trend preserving. ■

The next theorem generalizes the properties of L_n and U_n in Theorem 81 to arbitrary neighbor trend preserving operators.

Theorem 86 *Let $A : \mathcal{A}(\mathbb{Z}^d) \to \mathcal{A}(\mathbb{Z}^d)$ be a neighbor trend preserving operaror and let $f \in \mathcal{A}(\mathbb{Z}^d)$. For every finite local minimum (maximum) set V of $A(f)$ there exists a local minimum (maximum) set W of f such that $W \subseteq V$.*

Proof Let V be a finite local minimum set of $A(f)$. Let $q \in \mathrm{adj}(V)$ be such that $f(q) = \min_{y \in \mathrm{adj}(V)} f(y)$. It follows from (3.11) that there exists $p \in V$ such that $\{p,q\}$ is connected. By the local minimality of V we have

$$A(f)(p) < A(f)(q). \tag{3.37}$$

Due to the fact that A is neighbor trend preserving the inequality $f(p) \geq f(q)$ implies that $A(f)(p) \geq A(f)(q)$ which contradicts (3.37). Therefore, $f(p) < f(q)$. Let $Y = \{x \in V : f(x) < f(q)\}$. Clearly, $p \in Y$. Denote by W the largest connected component of Y which contains p. In terms of the notations used in the proof of Theorem 78, we have $W = \gamma_p(Y)$. We will show that W is a local minimum set of f. Let $z \in \mathrm{adj}(W)$. It follows from the construction of W that $f(q) > \max_{x \in W} f(x)$. Hence it is enough to show that $f(z) \geq f(q)$. By the Lemma 76a) we have two possibilities:

(i) $z \in V$. Then $z \notin Y$. Indeed, if $z \in Y$ then $W \cup \{z\}$ is a connected component of Y which contains W. This is impossible since W is the largest such component of Y. Since $z \notin Y$, then it violates the defining inequality of Y, that is we have $f(z) \geq f(q)$.

(ii) $z \in \mathrm{adj}(V)$. Then $f(z) \geq \min_{y \in \mathrm{adj}(V)} f(y) = f(q)$.

The proof for local maximum sets is carried out in a similar way. ■

3.6 Total Variation Preservation

Recall that we are assuming that the connection $\mathcal{C}$ on $\mathbb{Z}^d$ is defined via the so-called graph connectivity. More precisely, the points of $\mathbb{Z}^d$ are considered as vertices of a graph with edges connecting some of them. Equivalently, the connectivity of such a graph can be defined via a relation $r \subset \mathbb{Z}^d \times \mathbb{Z}^d$, where $p \in \mathbb{Z}^d$ is connected (by an edge) to $q \in \mathbb{Z}^d$ iff $(p,q) \in r$.

The relation r reflects what we consider neighbours of a point in the given context. For example, in image analysis ($d = 2$), it is common to use 4-connectivity (neighbours left, right, up and down) and 8-connectivity (in addition, the diagonal neighbours are considered). Let r be a relation on $\mathbb{Z}^d$. We call a set $C \subseteq \mathbb{Z}^d$ **connected**, with respect to the graph connectivity defined by r, if for any two pixels $p, q \in C$ there exists a set of pixels $\{p_1, p_2, ..., p_k\} \subseteq C$ such that each pixel is neighbour to the next one, p is neighbour to p_1 and p_k is neighbour to q. Here we assume that,

- r is reflexive, symmetric and shift invariant (3.38)
- $(p, p + e_k) \in r,$ for all $k = 1, 2, ..., d$ and $p \in \mathbb{Z}^d$ (3.39)

  where $e_k \in \mathbb{Z}^d$ is defined by $(e_k)_i = \begin{cases} 0 & \text{if } i \neq k \\ 1 & \text{if } i = k \end{cases}$

Conditions (3.38) and (3.39) ensure that the set of connected sets $\mathcal{C}$ defined through this relation is a connection in terms of Definition 53 and satisfies the conditions (3.8)–(3.11). Condition (3.39) is essential to the definition of total variation as will be seen in the sequel.

Since the information in an image is in the contrast, the total variation of the luminosity function is an important measure of the quantity of this information. Image recovery and noise removal via total variation minimization are discussed in [99], [121], [10], [67], [128], [35], [124], [23] [72], [22], [24] and [44]. It should be noted that there are several definitions of total variation for functions of multi-dimensional argument, namely Arzelá variation, Fréchet variation, Vitali variation, Pierpont variation, Hardy variation and Tonelli variation which are discussed in [1] and [25]. These definitions of total variation date back to the 1930's. However, in most of the applications cited above the total variation is the L^1 norm of a vector norm of the gradient of a function u defined on Ω, namely

$$TV(u) = \int_{\Omega} | \bigtriangledown u | d\Omega. \tag{3.40}$$

Note that defining total variation using the L^2 norm requires the function u to be smooth. This is unrealistic for our application in image processing as the image will always be in part noise and thus not smooth. In [22] the corresponding discretization of (3.40) in two dimensions is considered, leading us to the following definition.

Definition 87 *The* ***Total Variation*** *of $f \in \mathcal{A}(\mathbb{Z}^2)$ if given by*

$$TV(f) = \sum_{(i,j)\in\mathbb{Z}^2} \left(|f(x_{i,j+1}) - f(x_{ij})| + |f(x_{i+1,j}) - f(x_{ij})|\right)$$

and

Definition 88 *The* ***Total Variation*** *of $f \in \mathcal{A}(\mathbb{Z}^d)$ is given by*

$$TV(f) = \sum_{p\in\mathbb{Z}^d} \sum_{i=1}^{d} |f(p + (e_k)_i) - f(p)|.$$

If $TV(f) < \infty$, then $f \in \mathbb{Z}^d$ is said to be of **bounded variation**.

Table 3.1 gives the total variation of a few sample image seen in Figure 3.8. Notice that the pure noise image has the highest total variation and as the images become more homogenous their total variation reduces.

Image in Figure 3.8	Total Variation (standardized)
(a)	109173
(b)	132527
(c)	167011
(d)	193650
(e)	213530
(f)	235908
(g)	386408
(h)	703707

Table 3.1. Standardized Total Variation of Some Sample Images

As mentioned in Chapter 2, the LULU operators for sequences are total variation preserving. We show here that their d-dimensional counterparts

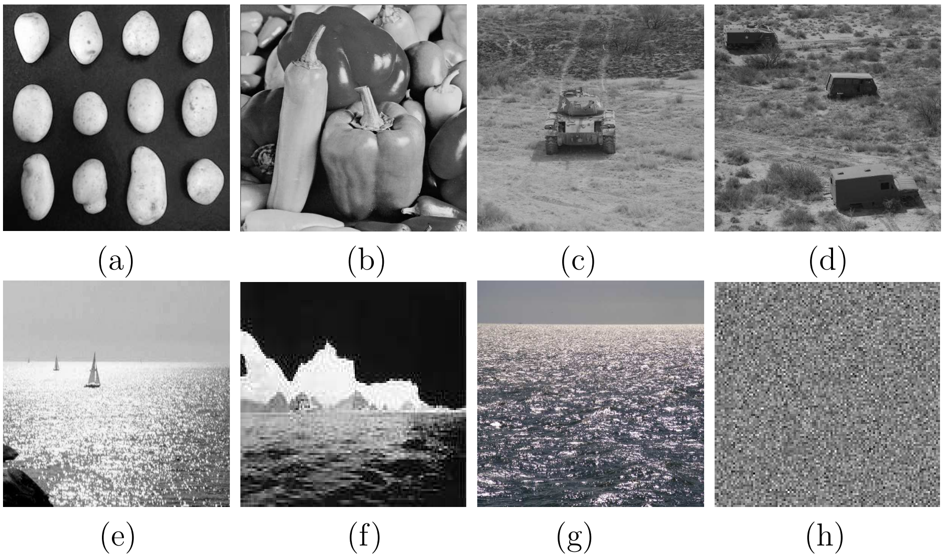


Figure 3.8. Sample Images

considered have the same property with respect to the total variation as given in Definition 88.

Let us denote by $BV(\mathbb{Z}^d)$ the set of all functions of bounded variation in $\mathcal{A}(\mathbb{Z}^d)$. Clearly, all functions of finite support are in $BV(\mathbb{Z}^d)$. For example, the luminosity functions of images are in $BV(\mathbb{Z}^2)$. Note that when $d = 1$ the Definition 88 gives that total variation of sequences as discussed in Chapter 2. Similar to sequences the total variation in Definition 88 is a semi-norm so we have that for a operator P

$$TV(f) \leq TV(P(f)) + TV((I - P)(f)).$$

Definition 89 *An operator P on $BV(\mathbb{Z}^d)$ is called* ***total variation preserving*** *if*

$$TV(f) = TV(P(f)) + TV((I - P)(f)).$$

There is a close connection between the property of total variation preservation and that of fully trend preserving.

Theorem 90 *If an operator $P : BV(\mathbb{Z}^d) \to BV(\mathbb{Z}^d)$ is fully trend preserving then it is also total variation preserving.*

Proof
For $p, q \in \Omega$ such that $\{p, q\} \in \mathcal{C}$, we have

$$|f(p) - f(q)| \leq \big|P(f)(p) - P(f)(q)\big| + \big|f(p) - P(f)(p) - \big(f(q) - P(f)(q)\big)\big|. \tag{3.41}$$

If $f(p) \geq f(q)$ then $P(f)(p) - P(f)(q) \geq 0$ since P is neighbour trend preserving. Then

$$\big|f(p) - P(f)(p) - \big(f(q) - P(f)(q)\big)\big| = \big(f(p) - f(q)\big) - \big(P(f)(p) - P(f)(q)\big) \geq 0$$

since P is difference reducing, see (3.36). Thus Equation (3.41) holds as an equality. Hence

$$\begin{aligned} TV(f) &= \sum_{p \in \mathbb{Z}^d} \sum_{i=1}^{d} \big|f\big(p + (e_k)_i\big) - f(p)\big| \\ &= \sum_{p \in \mathbb{Z}^d} \sum_{i=1}^{d} \Big(\big|P(f)\big(p + (e_k)_i\big) - P(f)(p)\big| + \big|f\big(p + (e_k)_i\big) \\ &\qquad - P(f)\big(p + (e_k)_i\big) - \big(f(p) - P(f)(p)\big)\big|\Big) \\ &= TV\big(P(f)\big) + TV\big((I - P)(f)\big). \end{aligned}$$

■

Thus as an easy consequence of Theorem 85 and Theorem 90 we have

Theorem 91 *The operators L_n, U_n, $n = 1, 2, ...$, and all their compositions, are total variation preserving.*

Note that if an operator P is total variation preserving then the complementary operator $I - P$ is also total variation preserving by (3.2). Similarly to the case for compositions of fully trend preserving operators, it is easy to show that compositions of operators which individually preserve the total variation maintain the preservation as well.

3.7 Conclusion

In this chapter we have extended the LULU operators L_n and U_n from one dimension, where they act on sequences, to higher dimensions, specifically for multidimensional arrays. The extension is done via the morphological concept of a connection due to the loss of order from one dimension, $\mathbb{Z}$, to higher dimensions, $\mathbb{Z}^d$. The extended operators are smoothers and separators, and are fully trend preserving and total variation preserving. In the next chapter we will present the application of this theoretical work to two dimensions for use in image processing.

Chapter 4

Applications to Image Processing

4.1 Introduction

In Chapter 3 we presented the extension of Carl Rohwer's LULU operators for sequences onto multidimensional arrays, namely $\mathbb{Z}^d$. The first and most obvious application after sequences is image processing thus the applications of this work are focused on images, that is for $d = 2$. Note that in the vast majority of fields in mathematics the nontrivial results and properties appear in two dimensions as this is generally considered the most important application domain. It thus seems logical to start the investigation of applications of our extension in two dimensions. We thus now present the extension for the Discrete Pulse Transform, its implementation and the distributional properties of the LULU operators in two dimensions. We also provide some illustrations of the use of the DPT in image processing.

4.2 The DPT

The Discrete Pulse Transform based on the LULU operators for sequences was derived in [93], [63], [95] and discussed in detail in Section 2.4. Multiresolution Analysis is effective for analyzing the information content in images, [74]. This is because varying structure sizes within an image make it difficult to analyze the content from only the grey-level pixel intensities. Using the extension of the LULU operators to functions on $\mathbb{Z}^d$ in the preceding sections we derive the DPT for functions in $\mathcal{A}(\mathbb{Z}^d)$ now. Following the success of the DPT for sequences in signal processing one may expect that the DPT on $\mathcal{A}(\mathbb{Z}^d)$ can play an important role in the analysis of these functions, due to the ability to represent the image (when $d = 2$) at all the resolution levels.

Similar to the case of sequences we obtain a decomposition of a function $f \in \mathcal{A}(\mathbb{Z}^d)$, with finite support. As usual $\text{supp}(f) = \{p \in \mathbb{Z}^d : f(p) \neq 0\}$. Let $N = \text{card}(\text{supp}(f))$. We derive the DPT of $f \in \mathcal{A}(\mathbb{Z}^d)$ by applying iteratively the operators L_n, U_n with n increasing from 1 to N as follows

$$DPT(f) = (D_1(f), D_2(f), ..., D_N(f)), \tag{4.1}$$

where the components of (4.1) are obtained through

$$D_1(f) = (I - P_1)(f) \tag{4.2}$$
$$D_n(f) = (I - P_n) \circ Q_{n-1}(f), \ n = 2, ..., N, \tag{4.3}$$

and $P_n = L_n \circ U_n$ or $P_n = U_n \circ L_n$ and $Q_n = P_n \circ ... \circ P_1$, $n \in \mathbb{N}$. We will show that this decomposition retains the properties of the decomposition (45) in the sense that each component D_n in (4.1) is a sum of discrete pulses with disjoint supports of size n, where in this setting a discrete pulse is defined as follows.

Definition 92 *A function $\phi \in \mathcal{A}(\mathbb{Z}^d)$ is called a* ***pulse*** *if there exists a connected set V and a real number α such that*

$$\phi(x) = \begin{cases} \alpha & \text{if} \quad x \in V \\ 0 & \text{if} \quad x \in \mathbb{Z}^d \setminus V. \end{cases}$$

The set V is the support of the pulse ϕ, that is $\text{supp}(\phi) = V$.

Note that a pulse as defined in Definition 92 is similar to the idea of a **flat zone** from mathematical morphology, [101]. It should be remarked that the

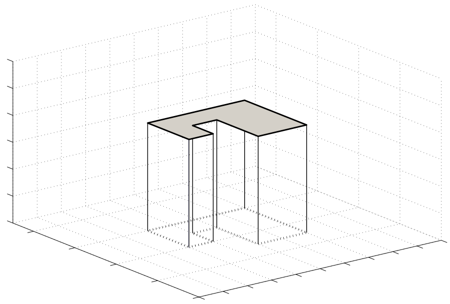


Figure 4.1. A Pulse

support of a pulse may generally have any shape, the only restriction being that it is connected.

It follows from (4.2)–(4.3) that

$$f = \sum_{n=1}^{N} D_n(f). \tag{4.4}$$

The usefulness of the representation (4.4) of a function $f \in \mathcal{A}(\mathbb{Z}^d)$ is in the fact that all terms are sums of pulses as stated in the theorem following the next lemma.

In the results that follow we will use the next lemma.

Lemma 93 *Let $f \in \mathcal{A}(\mathbb{Z}^d)$, $\text{supp}(f) < \infty$, be such that f does not have local minimum sets or local maximum sets of size smaller than n, for some $n \in \mathbb{N}$. Then we have the following two results.*

a)

$$(I - P_n)f = \sum_{i=1}^{\gamma^-(n)} \phi_{ni} + \sum_{j=1}^{\gamma^+(n)} \varphi_{nj}, \tag{4.5}$$

where $V_{ni} = \text{supp}(\phi_{ni}), i = 1, 2, ..., \gamma^-(n)$, are local minimum sets of f of size n, $W_{nj} = \text{supp}(\varphi_{nj}), j = 1, 2, ..., \gamma^+(n)$, are local maximum sets of f of size n, ϕ_{ni} and φ_{nj} are negative and positive discrete pulses *respectively, and we also have that*

- $V_{ni} \cap V_{nj} = \emptyset$ *and* $\text{adj}(V_{ni}) \cap V_{nj} = \emptyset, \ i, j = 1, ..., \gamma^-(n), \ i \neq j,$ (4.6)
- $W_{ni} \cap W_{nj} = \emptyset$ *and* $\text{adj}(W_{ni}) \cap W_{nj} = \emptyset, \ i, j = 1, ..., \gamma^+(n), i \neq j,$ (4.7)
- $V_{ni} \cap W_{nj} = \emptyset \ i = 1, ..., \gamma^-(n) \ , j = 1, ..., \gamma^+(n).$ (4.8)

b) For every fully trend preserving operator A

$$U_n(I - AU_n) = U_n - AU_n,$$
$$L_n(I - AL_n) = L_n - AL_n.$$

Proof

a) Let $V_{n1}, V_{n2}, ..., V_{n\gamma^-(n)}$ be all local minimum sets of size n of the function f. Since f does not have local minimum sets of size smaller than n, then f is a constant on each of these sets, by Theorem 83. Hence, the sets are disjoint, that is $V_{ni} \cap V_{nj} = \emptyset$, $i \neq j$. Moreover, we also have

$$\text{adj}(V_{ni}) \cap V_{nj} = \emptyset, \ i, j = 1, ..., \gamma^-(n). \tag{4.9}$$

Indeed, let $x \in \text{adj}(V_{ni}) \cap V_{nj}$. Then there exists $y \in V_{ni}$ such that $(x, y) \in r$. Hence $y \in V_{ni} \cap \text{adj}(V_{nj})$. From the local minimality of the sets V_{ni} and V_{nj} we obtain respectively $f(y) < f(x)$ and $f(x) < f(y)$, which is clearly a contradiction. For every $i = 1, ..., \gamma^-(n)$ denote by y_{ni} the point in $\text{adj}(V_{ni})$ such that

$$f(y_{ni}) = \min_{y \in \text{adj}(V_{ni})} f(y). \tag{4.10}$$

Then we have

$$U_n f(x) = \begin{cases} f(y_{ni}) & \text{if } x \in V_{ni}, \ i = 1, ..., \gamma^-(n) \\ f(x) & \text{otherwise (by Theorem 78)} \end{cases}$$

Therefore

$$(I - U_n)f = \sum_{i=1}^{\gamma^-(n)} \phi_{ni} \tag{4.11}$$

where ϕ_{ni} is a discrete pulse with support V_{ni} and negative value (down pulse).

Let $W_{n1}, W_{n2}, ..., W_{n\gamma^+(n)}$ be all local maximum sets of size n of the function $U_n f$. By Theorem 81(b) every local maximum set of $U_n f$ contains a local maximum set of f. Since f does not have local maximum sets of size smaller than n, this means that the sets W_{nj}, $j = 1, ..., \gamma^+(n)$, are all local maximum sets of f and f is constant on each of them. Similarly to the local minimum sets of f considered above we have $W_{ni} \cap W_{nj} = \emptyset$, $i \neq j$, and $\text{adj}(W_{ni}) \cap W_{nj} = \emptyset$, $i, j = 1, ..., \gamma^+(n)$. Moreover, since $U_n(f)$ is constant on any of the sets $V_{ni} \cup \{y_{ni}\}$, $i = 1, ..., \gamma^-(n)$, see Theorem 83, we also have

$$(V_{ni} \cup \{y_{ni}\}) \cap W_{nj} = \emptyset, \ \ i = 1, ..., \gamma^-(n), \ \ j = 1, ..., \gamma^+(n), \tag{4.12}$$

which implies (4.8).

Further we have

$$L_n U_n f(x) = \begin{cases} U_n f(z_{nj}) & \text{if } x \in W_{nj},\, j = 1, ..., \gamma^+(n) \\ \\ U_n f(x) & \text{otherwise} \end{cases}$$

where $z_{nj} \in \text{adj}(W_{nj})$, $j = 1, ..., \gamma^+(n)$, are such that $U_n f(z_{nj}) = \max\limits_{z \in \text{adj}(W_{nj})} U_n f(z)$. Hence

$$(I - L_n) U_n f = \sum_{j=1}^{\gamma^+(n)} \varphi_{nj} \tag{4.13}$$

where φ_{nj} is a discrete pulse with support W_{nj} and positive value (up pulse). Thus we have shown that

$$(I - P_n) f = (I - U_n) f + (I - L_n) U_n f = \sum_{i=1}^{\gamma^-(n)} \phi_{ni} + \sum_{j=1}^{\gamma^+(n)} \varphi_{nj}.$$

b) Let the function $f \in \mathcal{A}(\mathbb{Z}^d)$ be such that it does not have any local minimum or local maximum sets of size less than n. Denote $g = (I - AU_n)(f)$. We have

$$g = (I - AU_n)(f) = (I - U_n)(f) + ((I - A)U_n)(f). \tag{4.14}$$

As in a) we have that (4.11) holds, that is we have

$$(I - U_n)(f) = \sum_{i=1}^{\gamma^-(n)} \phi_{ni}, \tag{4.15}$$

where the sets $V_{ni} = \text{supp}(\phi_{ni})$, $i = 1, ..., \gamma^-(n)$, are all the local minimum sets of f of size n and satisfy (4.6). Therefore

$$g = \sum_{i=1}^{\gamma^-(n)} \phi_{ni} + ((I - A)U_n)(f). \tag{4.16}$$

Furthermore,

$$U_n(f)(x) = \begin{cases} f(x) & \text{if} \quad x \in \mathbb{Z}^d \setminus \bigcup_{i=1}^{\gamma^-(n)} V_{ni} \\ \\ v_i & \text{if} \quad x \in V_{ni} \cup \{y_{ni}\},\ i = 1, ..., \gamma^-(n), \end{cases}$$

where $v_i = f(y_{ni}) = \min_{y \in \mathrm{adj}(V_{ni})} f(y)$. Using that A is fully trend preserving, for every $i = 1, ..., \gamma^-(n)$ there exists w_i such that $((I - A)U_n)(f)(x) = w_i$, $x \in V_{ni} \cup \{y_{ni}\}$. Moreover, using that every adjacent point has a neighbour in V_{ni} we have that $\min_{y \in \mathrm{adj}(V_{ni})} ((I - A)U_n)(f)(y) = w_i$. Considering that the value of the pulse ϕ_{ni} is negative, we obtain through the representation (4.16) that V_{ni}, $i = 1, ..., \gamma^-(n)$, are local minimum sets of g.

Next we show that g does not have any other local minimum sets of size n or less. Indeed, assume that V_0 is a local minimum set of g such that $\mathrm{card}(V_0) \le n$. Since $V_0 \cup \mathrm{adj}(V_0) \subset \mathbb{Z}^d \setminus \bigcup_{i=1}^{\gamma^-(n)} V_{ni}$ it follows from (4.16) that V_0 is a local minimum set of $((I - A)U_n)(f)$. Then using that $(I - A)$ is neighbour trend preserving and using Theorem 86 we obtain that there exists a local minimum set W_0 of $U_n(f)$ such that $W_0 \subseteq V_0$. Then applying again Theorem 86 or Theorem 81 we obtain that there exists a local minimum set $\tilde{W}_0$ of f such that $\tilde{W}_0 \subseteq W_0 \subseteq V_0$. This inclusion implies that $\mathrm{card}(\tilde{W}_0) \le n$. Given that f does not have local minimum sets of size less than n we have $\mathrm{card}(\tilde{W}_0) = n$, that is $\tilde{W}_0$ is one of the sets V_{ni} - a contradiction. Therefore, V_{ni}, $i = 1, ..., \gamma^-(n)$, are all the local minimum sets of g of size n or less. Then using again (4.11) we have

$$(I - U_n)(g) = \sum_{i=1}^{\gamma^-(n)} \phi_{ni} \tag{4.17}$$

Using (4.15) and (4.17) we obtain

$$(I - U_n)(g) = (I - U_n)(f)$$

Therefore

$$\begin{aligned} (U_n(I - AU_n))(f) &= U_n(g) = g - (I - U_n)(f) \\ &= (I - AU_n)(f) - (I - U_n)(f) \\ &= (U_n - AU_n)(f). \end{aligned}$$

This proves the first identity. The second one is proved in a similar manner.
■

Theorem 94 *Let* $f \in \mathcal{A}(\mathbb{Z}^d)$.

a) For every $n \in \mathbb{N}$ *the function* $D_n(f)$ *is a sum of discrete pulses with disjoint support, that is, there exist* $\gamma(n) \in \mathbb{N}$ *and discrete pulses* ϕ_{ns}*,* $s = 1, ..., \gamma(n)$*, such that*

$$D_n(f) = \sum_{s=1}^{\gamma(n)} \phi^*_{ns} = \sum_{i=1}^{\gamma^-(n)} \phi_{ni} + \sum_{j=1}^{\gamma^+(n)} \varphi_{nj}, \gamma(n) = \gamma^-(n) + \gamma^+(n), \quad (4.18)$$

and

$$\operatorname{supp}(\phi^*_{ns_1}) \cap \operatorname{supp}(\phi^*_{ns_2}) = \emptyset \text{ for } s_1 \neq s_2. \quad (4.19)$$

b) Let $n_1, n_2, s_1, s_2 \in \mathbb{N}$ *be such that* $n_1 < n_2$*,* $1 \leq s_1 \leq \gamma(n_1)$ *and* $1 \leq s_2 \leq \gamma(n_2)$*. Then*

$$\operatorname{supp}(\phi^*_{n_1 s_1}) \cap \operatorname{supp}(\phi^*_{n_2 s_2}) \neq \emptyset \Longrightarrow \operatorname{supp}(\phi^*_{n_1 s_1}) \subset \operatorname{supp}(\phi^*_{n_2 s_2}) \quad (4.20)$$

Proof
a) $D_n(f)$ is applied to the function $Q_{n-1}(f)$ which, by Corollary 82, does not have local maximum or minimum sets of size less than n. Thus by Lemma 93(a) we have that $D_n(f) = (I - P_n)Q_{n-1}(f)$ is a sum of disjoint discrete pulses as given in (4.6) and (4.7).

b) Let $\operatorname{supp}(\phi^*_{n_1 s_1}) \cap \operatorname{supp}(\phi^*_{n_2 s_2}) \neq \emptyset$. It follows from the construction of (4.18) derived in (a) that the functions $Q_n(f)$ and $L_{n+1}(Q_n(f))$, $n \geq n_1$, are constants on the set $\operatorname{supp}(\phi^*_{n_1 s_1})$. Furthermore, the set $\operatorname{supp}(\phi^*_{n_2 s_2})$ is a local maximum set of $Q_{n_2-1}(f)$ or a local minimum set of $L_{n_2}(Q_{n_2-1}(f))$. From the definition of local maximum set and local minimum set it follows that $\operatorname{supp}(\phi^*_{n_1 s_1}) \subset \operatorname{supp}(\phi^*_{n_2 s_2})$. ■

Using Theorem 94, the identity (4.4) can be written in the form

$$f = \sum_{n=1}^{N} \sum_{s=1}^{\gamma(n)} \phi_{ns}. \quad (4.21)$$

The equality (4.21) is a decomposition of the image which we call the **Discrete Pulse Transform** of f, where the pulses have the properties (4.19) – (4.20).

Although the importance of total variation preservation for separators cannot be doubted, it is even more so for hierarchical decompositions like the

Discrete Pulse Transform, due to the fact that they involve iterative applications of separators. Using Theorem 91 it is easy to obtain the statement of the following theorem, which shows that, irrespective of the length of the vector in (4.1) or the number of terms in the sum (4.21), no additional total variation, or noise, is created via the decomposition.

Theorem 95 *The discrete pulse decomposition (4.1) is total variation preserving, that is*

$$TV(f) = \sum_{n=1}^{N} \sum_{s=1}^{\gamma(n)} TV(\phi_{ns}). \tag{4.22}$$

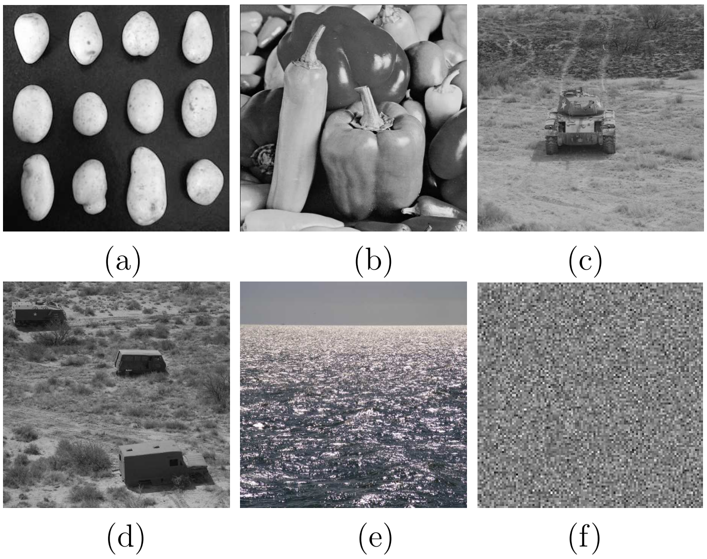


Figure 4.2. Sample Images

We should remark that representing a function as a sum of pulses can be done in many different ways. However, in general, such decompositions increase the total variation, that is, we might have strict inequality in (4.22) instead of equality. The variation spectrum of the decomposition also provides useful information in analyzing the image content. In Figure 4.3 the variation spectrum of the images in Figure 4.2 are shown. The total variation at each resolution level is standardized to a 100 × 100 image and a log scale is used to make the variation spectrum more obvious. Notice that the noise image in Figure 4.2(f) has a spectrum where all the variation is in the lower half of the resolution levels. The images in Figure 4.2(a)-(d) all can be seen to have variation in the lower half of the spectrum as well as ‘scattered’ variations amongst the rest of the spectrum as well. This is an indication of underlying noise in the images as well as features of interest. In the image in Figure 4.2(e) we see a large amount of variation in the lower and upper halves of the spectrum. This is indicative of random noise (lower resolution levels) as well as larger noise features or the presence of many features in the upper half of the spectrum.

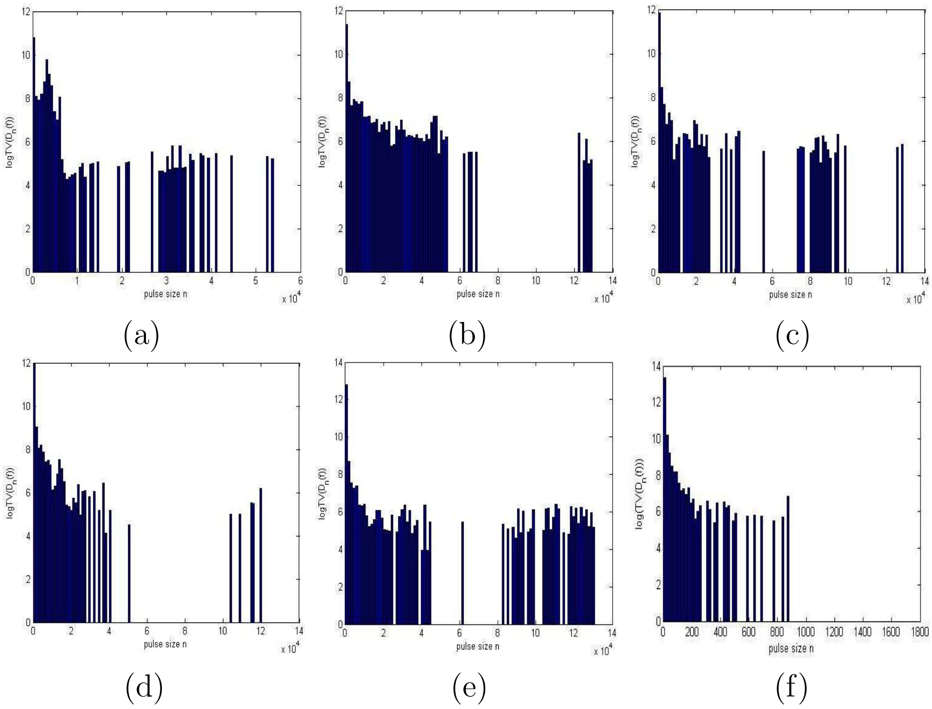


Figure 4.3. Variation Spectrums of Images displayed in Figure 3.9.

As discussed in the introduction of this chapter, the quality of a nonlinear hierarchical decomposition, such as the Discrete Pulse Transform given in (4.1), can be characterized through the concept of **consistent decomposition** (also called **strong consistency** [63]) given in (3.3). Whether or not the multidimensional Discrete Pulse Transform in (4.1) is strongly consistent is still an open problem. However, the next theorem shows that the Discrete Pulse Transform in (4.1) satisfies a slightly weaker form of strong consistency involving only the sums of the input layers, that is, a linear combination.

Theorem 96 *Let $f \in \mathcal{A}(\mathbb{Z}^d)$. For any two integers m and n such that $m < n$ the function $g = D_m(f) + D_{m+1}(f) + ... + D_n(f)$ decomposes consistently, that is*

$$D_j(g) = \begin{cases} D_j(f) & \text{for } m \le j \le n \\ 0 & \text{otherwise} \end{cases}$$

The proof uses the following Lemma.

Lemma 97 *Let* $Q_n = P_n P_{n-1} ... P_1$ *where* $P_k = L_k U_k$ *or* $P_k = U_k L_k$. *We have*

a) $Q_n Q_m = Q_{\max\{n,m\}}$

b) $Q_m(I - Q_n) = Q_m - Q_n = (I - Q_n)Q_m$ *for all integer* m, n *such that* $m \leq n$.

Proof
We consider only $P_k = L_k U_k$ as the other case is dealt with by symmetry. Let $f \in \mathcal{A}(\mathbb{Z}^d)$.

a) It follows from Corollary 82 that $Q_n(f)$ does not have any local minimum or local maximum sets of size n or less. Hence $P_k(Q_n(f)) = Q_n(f)$ for $k = 1, ..., n$. For $m \leq n$ this implies that $Q_m(Q_n(f)) = Q_n(f)$. If $m > n$ then we have $(Q_m Q_n)(f) = (P_m ... P_{n+1} P_n ... P_1)(Q_n(f)) = (P_m ... P_{n+1})(Q_n(f)) = C_m(f)$.

b) We use induction on j as in the proof of this property in the one dimensional case, see [92]. Let $j = 1$. Using the result in Lemma 93b), the full trend preservation property of the LULU operators established in Theorem 85 and the absorbtion property in a) we have

$$Q_1(I - Q_n) = L_1(U_1(I - Q_n L_1 U_1)) = L_1(U_1 - Q_n L_1 U_1)$$
$$= L_1(I - Q_n L_1)U_1 = (L_1 - Q_n L_1)U_1 = Q_1 - Q_n = (I - Q_n)Q_1.$$

Assume now that the statement is true for some $m = j < n$. From the inductive assumption we have

$$\begin{aligned} Q_{j+1}(I - Q_n) &= P_{j+1} Q_j (I - Q_n) = P_{j+1}(Q_j - Q_n) \\ &= P_{j+1}(Q_j - Q_n Q_j) = P_{j+1}(I - Q_n)Q_j. \end{aligned}$$

Using Lemma 93b), Theorem 85 and a) as for $j = 1$ we obtain further

$$P_{j+1}(I - Q_n)Q_j = L_{j+1}(U_{j+1}(I - Q_n L_{j+1} U_{j+1})Q_j$$
$$= (L_{j+1} U_{j+1} - Q_n L_{j+1} U_{j+1})Q_j = Q_{j+1} - Q_n = (I - Q_n)Q_{j+1}.$$

■

Proof of Theorem 96. Using Lemma 97, function g can be written in the following equivalent forms

$$\begin{aligned} g &= ((I-P_m)Q_{m-1}+(I-P_{m+1})Q_m+...+(I-P_n)Q_{n-1})(f) \\ &= (Q_{m-1}-Q_n)(f)=(I-Q_n)Q_{m-1}=Q_{m-1}(I-Q_n) \end{aligned} \tag{4.23}$$

It follows from Corollary 82 and Theorem 86 that g does not have any local maximum or local minimum sets of size less than m. Hence $P_k(g)=g$ for $k=1,...,m-1$ and therefore $Q_k(g)=g$ for $k=1,...,m-1$. Then it follows from (4.23) that $D_j(g)=(I-P_j)(g)=0$ for $j<m$. Let $m\le j\le n$. Then using again Lemma 97 we obtain

$$\begin{aligned} D_j(g) &= (Q_{j-1}-Q_j)(g)=(Q_{j-1}(I-Q_n)Q_{m-1}-Q_j(I-Q_n)Q_{m-1})(f) \\ &= ((I-Q_n)Q_{j-1}Q_{m-1}-(I-Q_n)Q_jQ_{m-1})(f) \\ &= ((I-Q_n)Q_{j-1}-(I-Q_n)Q_j)(f)=(Q_{j-1}-Q_n-Q_j+Q_n)(f) \\ &= (Q_{j-1}-Q_j)(f)=D_j(f). \end{aligned}$$

Finally, for $k\ge n$ we have

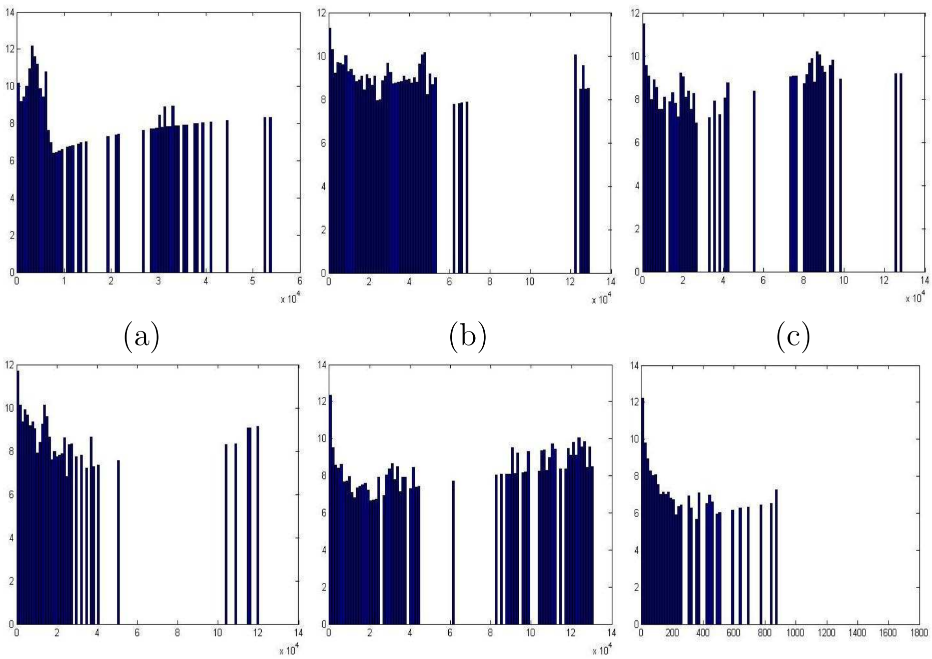
$$Q_k(g)=(Q_k(I-Q_n)Q_{m-1})(f)=(Q_kQ_n(I-Q_n)Q_{m-1})(f)=0,$$

which implies that $D_j(g)=0$ for $j>n$. ■

It is instructive to look at the connection between the DPT and mathematical morphology. We know that L_n and U_n are attribute filters, specifically area openings and closings. In [127] granulometries, anti-granulometries and granulometric curves are discussed. Granulometries also incorporate the scale inherent in an image in their mechanism. First introduced by Matheron in 1967, [77], and investigated further in [111, Chapter 4.6], [36], [37], [127], [119] using morphological reconstruction and area openings and closings, they provide a pattern spectrum with which texture analysis, feature extraction, object recognition and shape analysis, to mention a few, can be investigated. A granulometry is defined as follows.

Definition 98 *A* ***granulometry*** *(of openings)* $\{\gamma_\lambda\}$ *with size parameter* λ*, is an antiextensive, increasing and absorbing (*$\gamma_{\lambda_1}\circ\gamma_{\lambda_2}=\gamma_{\lambda_2}\circ\gamma_{\lambda_1}=\gamma_{\max\{\lambda_1,\lambda_2\}}$*) set of operators, [111, Chapter 4.6]. The first two properties are also those of an opening. An* ***anti-granulometry*** *(of closings) differs only in the first property in that it is instead extensive, thus specifying a closing with the first two properties.*

The L_n and U_n operators are respectively the area opening and closing. They therefore form a granulometry and anti-granulometry with scale parameter n respectively. In order to preserve contours, edges and thus shapes, within the image, the operator used as the granulometry should be connected, [118]. This means that the operator acts on flat zones (connected components with constant value) and not the individual pixels, [100]. The area granulometries provided by L_n and U_n provide this property, since they work directly on the connected local maximum and local minimum sets. In [118] a more effective method is introduced by defining the degree of connectedness, which will also be looked into in the future. The area granulometry is however more efficient to apply computationally.



(a) (b) (c)

(d) (e) (f)

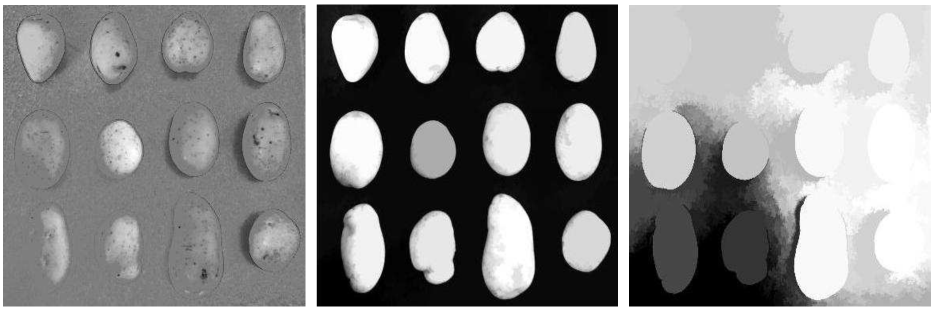
Figure 4.4. Granulometries (Pattern Spectrum) of the Images Displayed in Figure 4.2.

As mentioned before, using either L_n or U_n individually introduces a bias as L_n removes connected components of area n with positive height and U_n removes those with negative height, namely peaks and valleys respectively. We thus use either $P_n = L_n \circ U_n$ or $P_n = U_n \circ L_n$. We then construct the pattern spectrum as the loss between P_n and P_{n-1} vs. n. The volume is the sum of the volumes of all the pulses (negative and positive) extracted by P_n from $Q_{n-1}(f)$. The individual pulse volumes are simply the area

(n) multiplied by the height (relative luminosity). A large impulse in the resulting pattern spectrum indicates the presence of many structures at that scale. Figure 4.4 give the pattern spectrum for the images in Figure 4.2. Notice that the granulometries in Figure 4.4 are very similar to the variation spectrums in Figure 4.3.

Another option to extract information in a similar manner to the pattern spectrum is to plot the total variation $TV(P_n(f) - P_{n-1}(f))$ vs. n. This is exactly the variation spectrum dealt with before.

In Figures 4.5-4.10 we have reconstructed the images in Figure 4.2 using only certain pulses obtained via the Discrete Pulse Transform of the respective image. We note the connection between the partially reconstructed images and the corresponding pattern spectrums in Figure 4.4. Figures 4.11-4.16 provide the number of pulses of each size in the decomposition. It can clearly be seen that the numbers taper off as n increases. At certain values of n there may be an increase in the numbers of pulses of that size. This indicative of many (important) features at that scale.



(a) (b) (c)

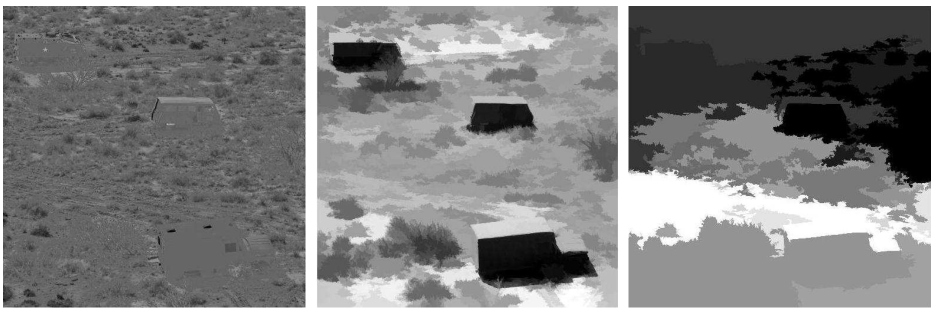
Figure 4.5. Reconstruction via DPT of Figure 4.2(a): Only pulses of size (a) 1 to 2500 (b) 2501 to 7000 (c) 7001 to 57478

(a) (b) (c)

Figure 4.6. Reconstruction via DPT of Figure 4.2(b): Only pulses of size (a) 1 to 100 (b) 101 to 53000 (c) 53001 to 130678

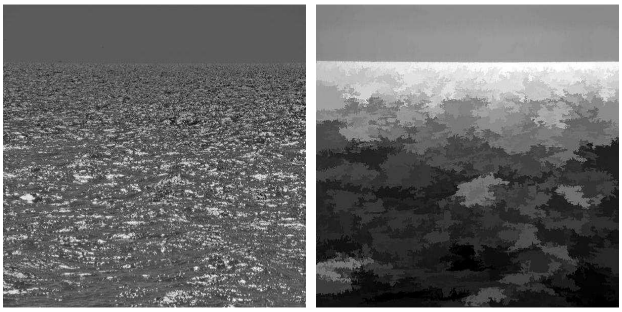
(a) (b) (c)

Figure 4.7. Reconstruction via DPT of Figure 4.2(c): Only pulses of size (a) 1 to 50 (b) 51 to 8000 (c) 8001 to 130139



(a) (b) (c)

Figure 4.8. Reconstruction via DPT of Figure 4.2(d): Only pulses of size (a) 1 to 1000 (b) 1001 to 28000 (c) 28001 to 121447



(a) (b)

Figure 4.9. Reconstruction via DPT of Figure 4.2(e): Only pulses of size (a) 1 to 1000 (b) 1001 to 132370

(a) (b) (c)

Figure 4.10. Reconstruction via DPT of Figure 4.2(f): Only pulses of size (a) 1 to 100 (b) 101 to 1000 (c) 1001 to 1660

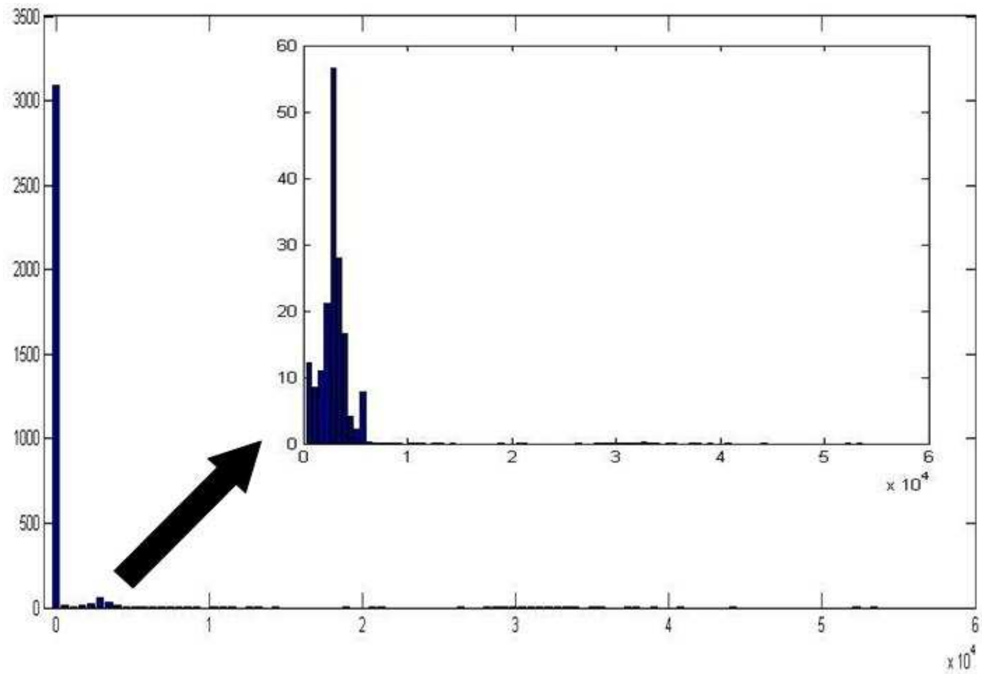


Figure 4.11. Number of Pulses of each Size for Figure 4.2(a).

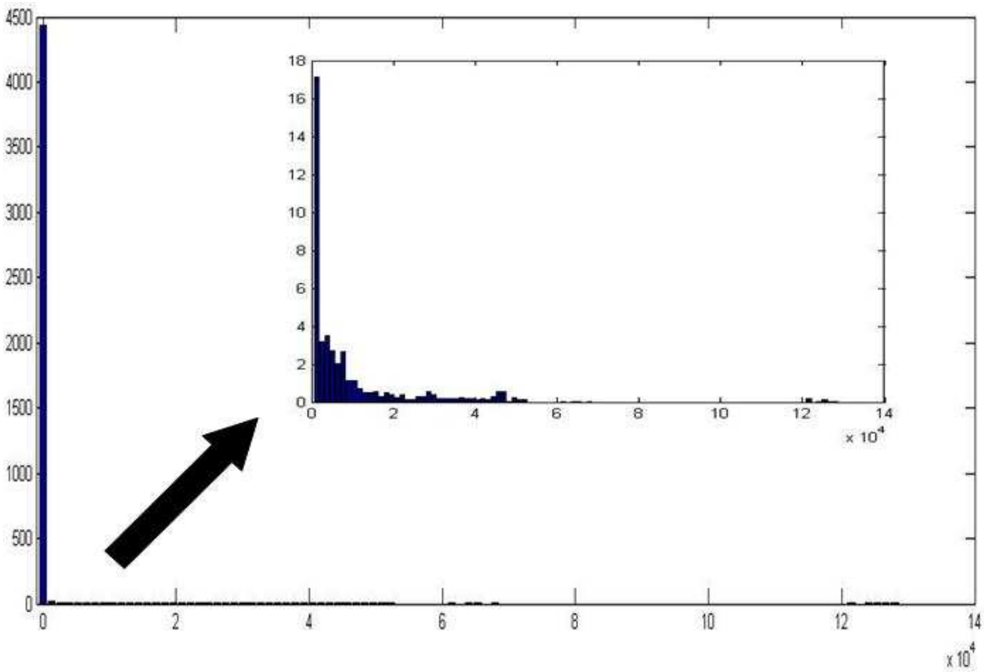


Figure 4.12. Number of Pulses of each Size for Figure 4.2(b).

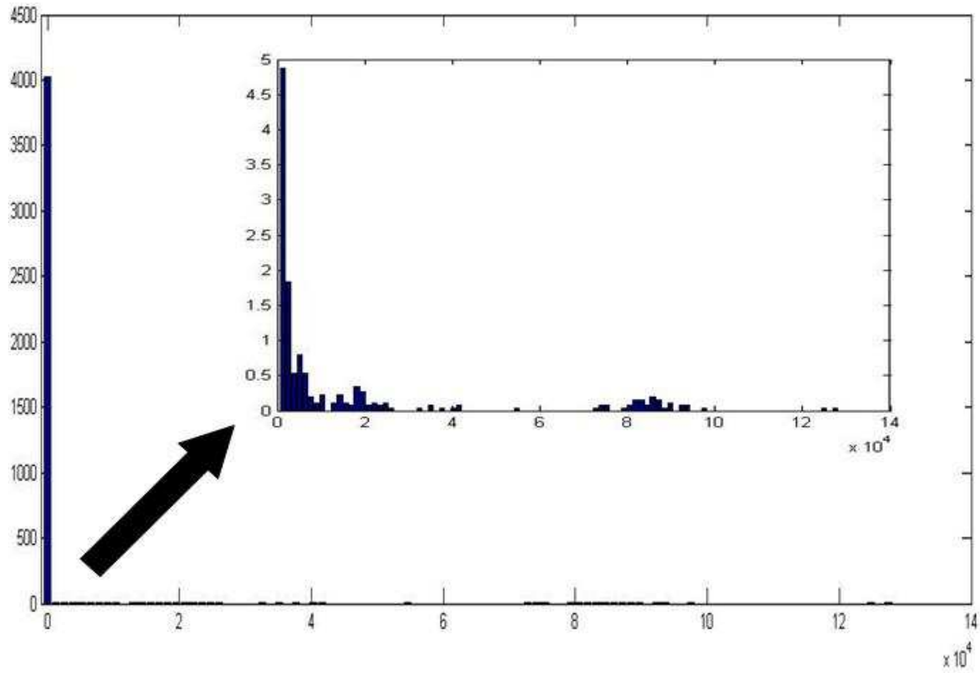


Figure 4.13. Number of Pulses of each Size for Figure 4.2(c).

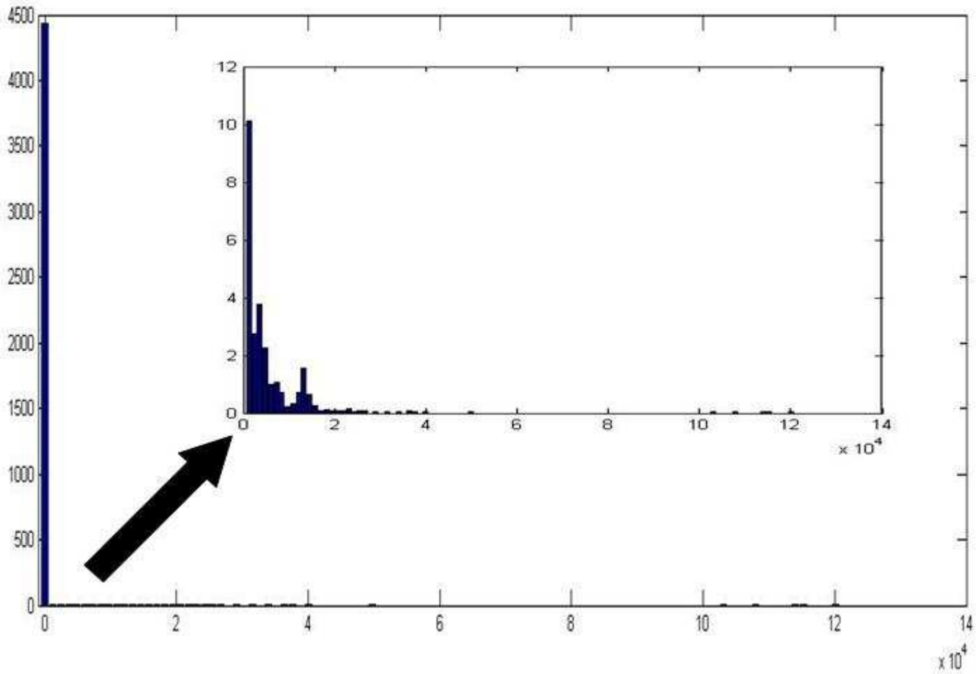


Figure 4.14. Number of Pulses of each Size for Figure 4.2(d).

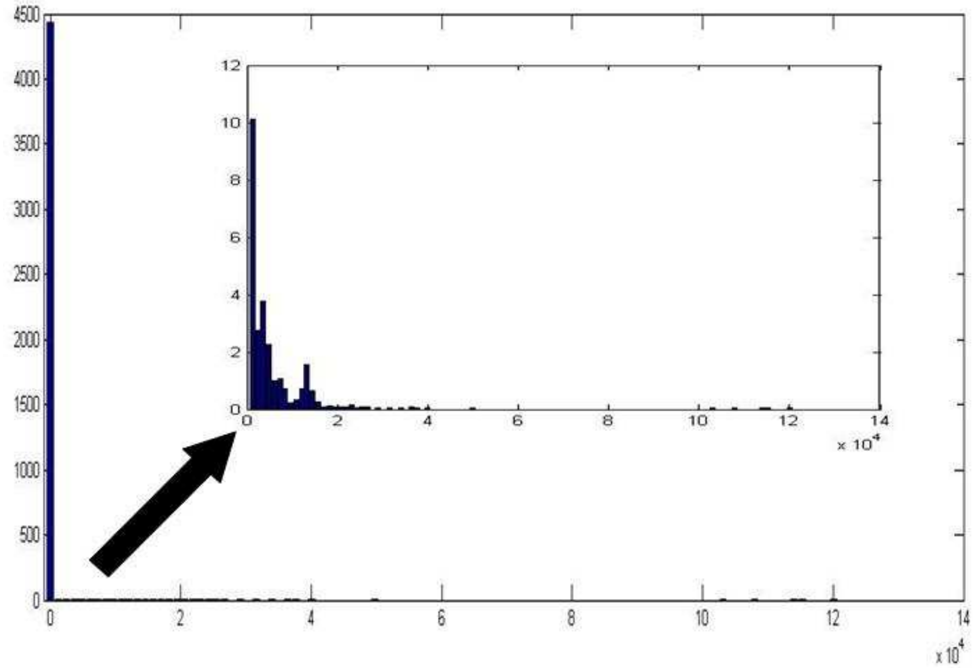


Figure 4.15. Number of Pulses of each Size for Figure 4.2(e).

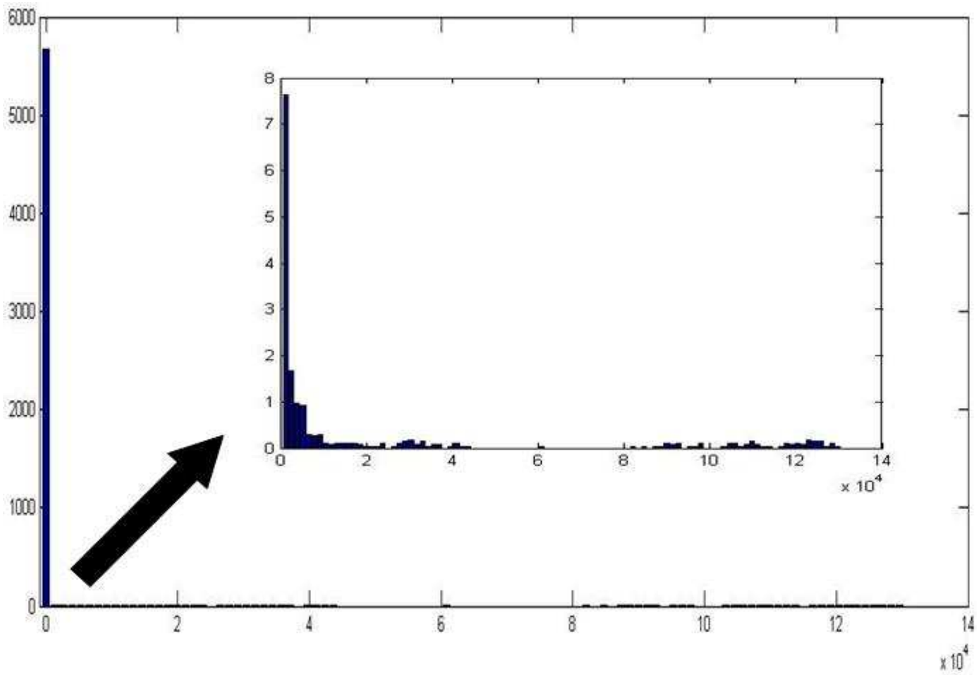


Figure 4.16. Number of Pulses of each Size for Figure 4.2(f).

4.3 The Roadmaker's Algorithm and its Implementation

The implementation of the LULU smoothers for sequences was developed in [63], [64] and [66] and is termed the **Roadmaker's algorithm** due to the removal of peaks and filling of valleys. The application of $L_n \circ U_n$ or $U_n \circ L_n$ to obtain the DPT from first principles, requires $O(n^3)$ time where n is the length of the sequence. The Roadmaker's algorithm however requires only $O(n)$ time.

The hope is thus to obtain an algorithm to extract the DPT of images in $O(nm)$ time where the size of the image is $n \times m$. This algorithm is under development with Stéfan van der Walt from Stellenbosch University. We include our current working code (using MATLAB) for completeness, which has also been converted to C++ code by B Anguelov in [2].

Function *dpt* obtains the DPT of image *ff* (of type double)

```
function [dp,background]=dpt; global ff N M global set csind setlum
undo adj cset ancestors

%detecting the prime critical sets:
tic k=0; cset=[]; [N,M]=size(ff); csind=zeros(N,M); for i=1:N
    for j=1:M
        if csind(i,j)==0
            pnbr=nbr([i j]);
            setlum=ff(i,j);
            v=[];
            for p=1:size(pnbr,1)
                v(p)=ff(pnbr(p,1),pnbr(p,2));
            end
            minv=min(v);
            maxv=max(v);
            if setlum<=minv|setlum>=maxv
                k=k+1;
                set=[i j];
                csind(i,j)=k;
                adj=[];
                undo=[];
                rgrow(pnbr);
```

```
                for s=1:size(undo,1)
                    csind(adj(s,1),adj(s,2))=undo(s,1);
                end
                cset(k).set=set;
                cset(k).adj=adj;
                cset(k).alife=1;
                cset(k).ancestors=[];
            end
        end
    end
end active=([1:k])';

%pulse extraction by size

size(cset,2) dp=[]; dpcounter=0; n=0; while size(active,1)>1
    n=n+1;
    newactive=[];
    for u=1:size(active,1)
        k=active(u);
        if cset(k).alife>0;
            if size(cset(k).set,1)<=n
                set=cset(k).set;
                setlum=ff(set(1,1),set(1,2));
                v=[];
                for i=1:size(cset(k).adj,1)
                    v(i)=ff(cset(k).adj(i,1),cset(k).adj(i,2));
                end
                minv=min(v);
                if minv>setlum
                    dpcounter=dpcounter+1;
                    dp(dpcounter).set=set;
                    dp(dpcounter).adj=cset(k).adj;
                    dp(dpcounter).rellum=setlum-minv;
                    dp(dpcounter).cumlum=setlum;
                    ancestors=cset(k).ancestors;
                    for s=1:size(ancestors,1)
                        dp(ancestors(s)).heir=dpcounter;
                    end
                    dp(dpcounter).ancestors=ancestors;
                    cset(k).ancestors=dpcounter;
                    for s=1:size(set,1)
```

```
            ff(set(s,1),set(s,2))=minv;
        end
        newactive=[newactive;k];
        setlum=minv; adj=[]; undo=[];
        ancestors=cset(k).ancestors;
        rgrow(cset(k).adj);
        for s=1:size(undo,1)
            csind(adj(s,1),adj(s,2))=undo(s,1);
        end
        cset(k).adj=adj;
        cset(k).set=set;
        cset(k).ancestors=ancestors;
    else
        maxv=max(v);
        if maxv<setlum
            dpcounter=dpcounter+1;
            dp(dpcounter).set=set;
            dp(dpcounter).adj=cset(k).adj;
            dp(dpcounter).rellum=setlum-maxv;
            dp(dpcounter).cumlum=setlum;
            ancestors=cset(k).ancestors;
            for s=1:size(ancestors,2)
                dp(ancestors(s)).heir=dpcounter;
            end
            dp(dpcounter).ancestors=ancestors;
            cset(k).ancestors=dpcounter;
            for s=1:size(set,1)
                ff(set(s,1),set(s,2))=maxv;
            end
            newactive=[newactive;k];
            setlum=maxv; adj=[]; undo=[];
            ancestors=cset(k).ancestors;
            rgrow(cset(k).adj);
            for s=1:size(undo,1)
                csind(adj(s,1),adj(s,2))=undo(s,1);
            end
            cset(k).adj=adj;
            cset(k).set=set;cset(k).ancestors=ancestors;
        else
            cset(k).alife=0;
        end
```

```
                end
            else
                newactive=[newactive;k];
            end
        end
    end
    active=newactive;
end background=ff(1,1); toc
```

Function *nbr* obtains the neighbours for pixel p with respect to 4-connectivity (called by *dpt*).

```
function y=nbr(p) global N global M k=0; i=p(1,1); j=p(1,2); if i>1
    k=k+1;
    y(k,1)=i-1;
    y(k,2)=j;
end if j>1
    k=k+1;
    y(k,1)=i;
    y(k,2)=j-1;
end if i<N
    k=k+1;
    y(k,1)=i+1;
    y(k,2)=j;
end if j<M
    k=k+1;
    y(k,1)=i;
    y(k,2)=j+1;
end
```

Function *rgrow* is used to obtain the n-neighbourhoods (is called by *dp*).

```
function rgrow(new) global ff setlum k
%input
global set undo adj global ancestors
%cumulative output
global cset csind
%input and update
for s=1:size(new,1)
```

```
    switch csind(new(s,1),new(s,2))
        case 0
            if ff(new(s,1),new(s,2))==setlum;
                set=[set;new(s,:)];
                csind(new(s,1),new(s,2))=k;
                rgrow(nbr(new(s,:)));
            else
                adj=[adj;new(s,:)];
                csind(new(s,1),new(s,2))=k;
                undo=[undo;0];
            end
        case k
        otherwise
            kk=csind(new(s,1),new(s,2));
            if ff(new(s,1),new(s,2))==setlum;
                set=[set;cset(kk).set];
                ancestors=[ancestors;cset(kk).ancestors];
                for q=1:size(cset(kk).set,1)
                    csind(cset(kk).set(q,1),cset(kk).set(q,2))=k;
                end
                cset(kk).alife=0;
                rgrow(cset(kk).adj);
            else
                adj=[adj;new(s,:)];
                csind(new(s,1),new(s,2))=k;
                undo=[undo;kk];
            end
        end
    end
end
```

Function *reconstructandscale* reconstructs the image with only pulses of size *minsize* to *maxsize*.

```
function rec=reconstructandscale(dp,minsize,maxsize,N,M)
rec=zeros(N,M); k=1; while (size(dp(k).set,1)<minsize)
    k=k+1;
end while (k<=size(dp,2))&(size(dp(k).set,1)<=maxsize)
    for i=1:size(dp(k).set,1)
        rec(dp(k).set(i,1),dp(k).set(i,2))=
```

```
            rec(dp(k).set(i,1),dp(k).set(i,2))+dp(k).rellum;
    end
    k=k+1;
end l=min(min(rec)); u=max(max(rec));
rec=20+round((rec-l)*200/(u-l));
```

4.4 Image Processing

In this section we will expand on the application of the DPT to feature detection in images. We will investigate the images in Figure 4.17 in the order they appear.

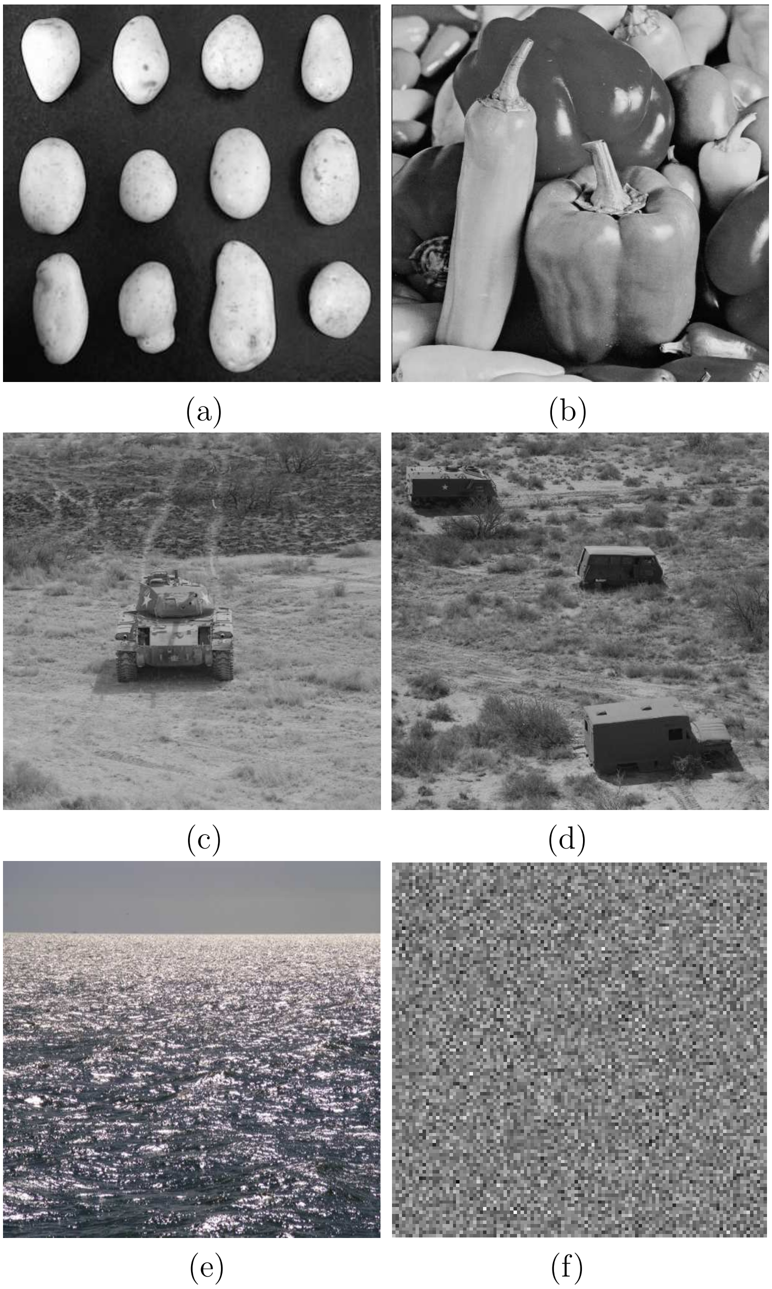


Figure 4.17. Sample Images

The variation spectrum for Figure 4.17(a), the 'Potatoes' image, is given in Figure 4.18.

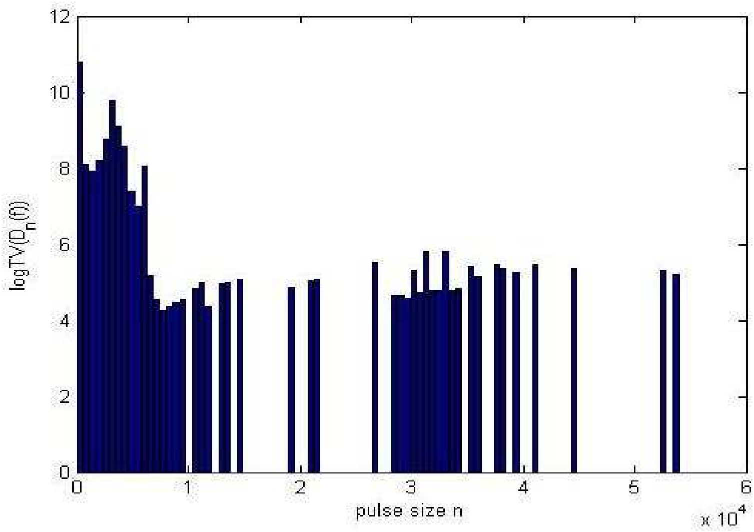


Figure 4.18. The Variation Spectrum of Figure 4.17(a).

The first thing to note in Figure 4.18 is that there is a decrease in the variation as n increases. We thus naturally assume noise in the image appears in the lower resolution layers (n small). This phenomenon can be seen in all the variation spectrums of the images in Figure 4.17. Recall also that the graphs of the variation spectrums are on a log scale so the variation decrease with n is less pronounced than it is in reality. The reduction is so drastic on a non-log scale that the higher resolution layers (n large) can barely be seen.

There seems to be three clusters of variation in Figure 4.18: (1) 1 to 22000, (2) 25000 to 45000 and (3) 50000 to 57478. If we reconstruct the image into only these three groups pulse sizes we obtain the results in Figure 4.19. We see that the potatoes feature very clearly in Figure 4.19(a) and without the background illumination seen in Figure 4.17(a). Figures 4.19(b) represents the illumination pulses and Figure 4.19(c) represents larger noise pulses. Figure 4.20 shows the further separation of Figure 4.19(a) into the noise (Figure 4.20(a)) and the features, namely the potatoes (Figure 4.20(b)).

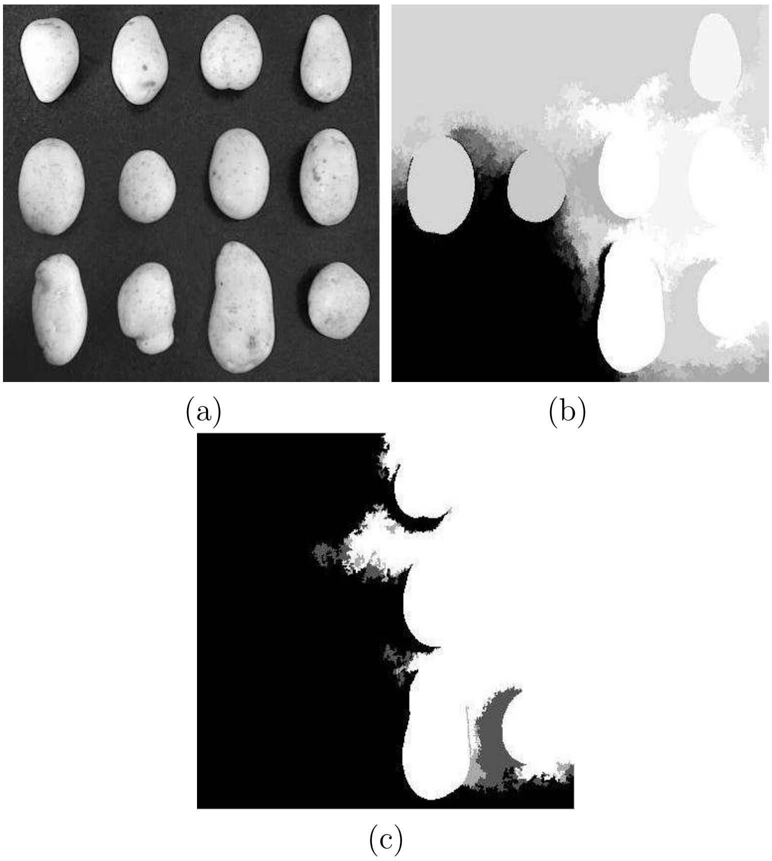


Figure 4.19. Reconstruction via DPT of Figure 4.17(a) with only pulses of size (a) 1 to 22000 (b) 22001 to 45000 (c) 45001 to 57478

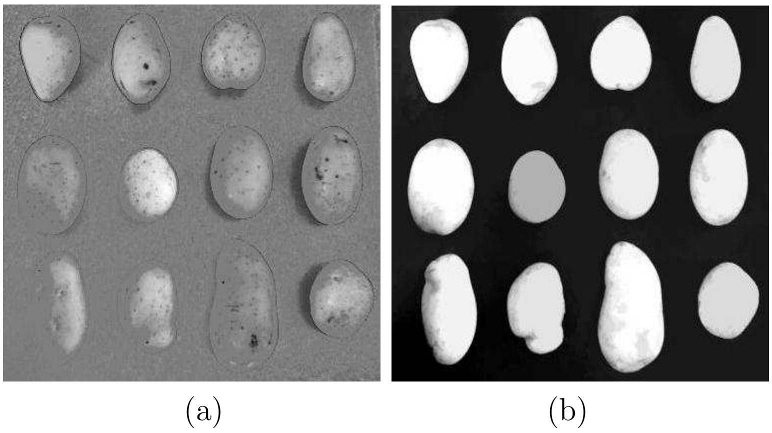


Figure 4.20. Reconstruction via DPT of Figure 4.17(a) with only pulses of size (a) 1 to 2500 (b) 2501 to 22000

The variation spectrum for Figure 4.17(b), the 'Veggies' image, is given in Figure 4.21.

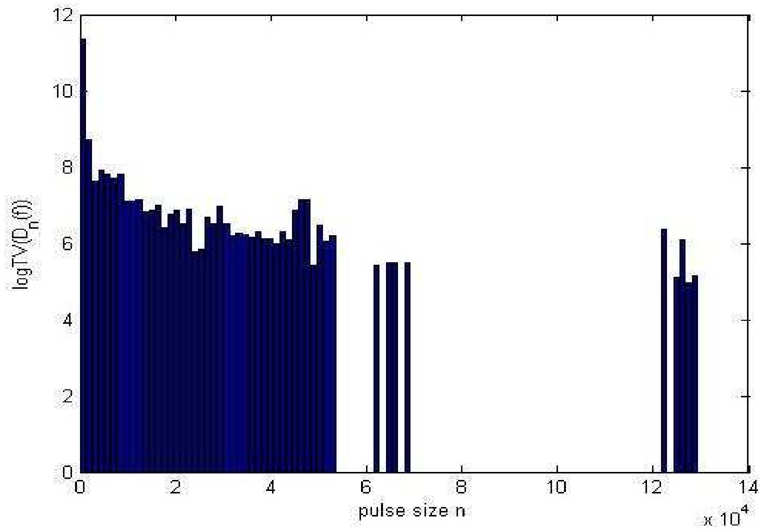
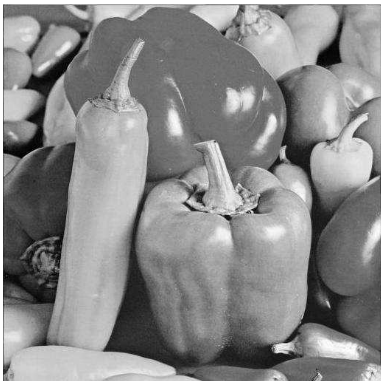


Figure 4.21. The Variation Spectrum of Figure 4.17(b).

There seems to be three clusters of variation in Figure 4.21: (1) 1 to 53000, (2) 60000 to 70000 and (3) 120000 to 130678. If we reconstruct the image into only these three groups pulse sizes we obtain the results in Figure 4.22.

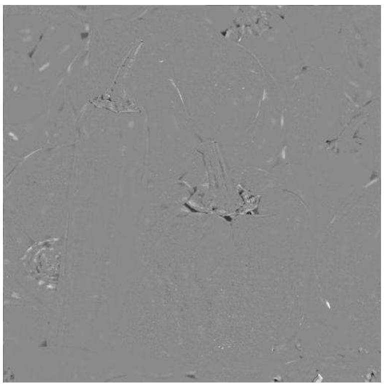
We see that the full 'Veggies' image in Figure 4.22(a). The respective variation clusters for Figures 4.22(b) and (c) are relatively small indicating there isn't a lot of information in those resolution layers. They only represent the larger noise pulses. This veggies image do not have a lot of noise in it. See Figure 4.23 for the further separation of Figure 4.22(a) into the noise (Figure 4.23(a)) and the rest of the image (Figure 4.23(b)).


(a)


(b)


(c)

Figure 4.22. Reconstruction via DPT of Figure 4.17(b) with only pulses of size (a) 1 to 53000 (b) 60000 to 70000 (c) 120000 to 130678


(a)


(b)

Figure 4.23. Reconstruction via DPT of Figure 4.17(b) with only pulses of size (a) 1 to 100 (b) 101 to 53000

The variation spectrum for Figure 4.17(c), the 'Tank' image, is given in Figure 4.24.

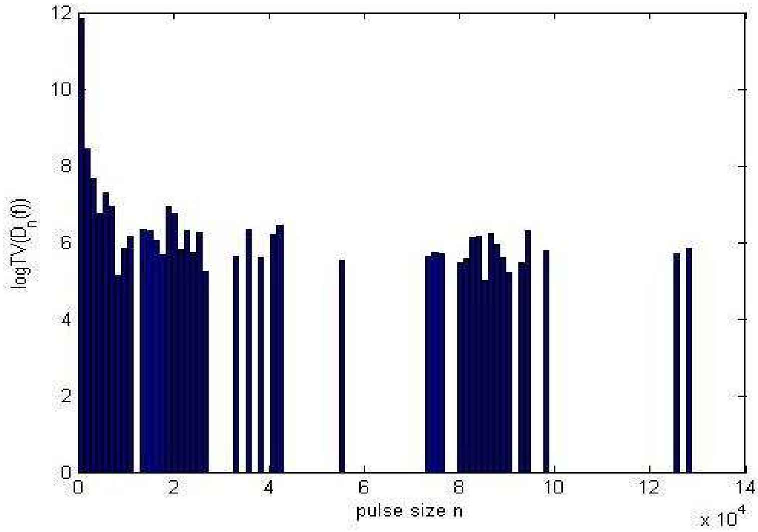
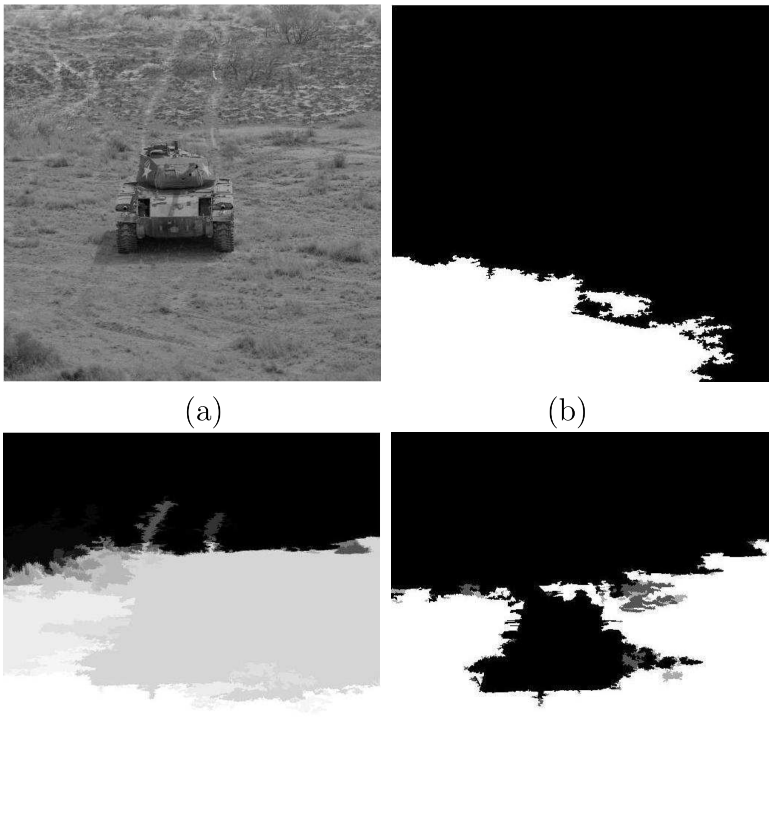


Figure 4.24. The Variation Spectrum of Figure 4.17(c).

There seems to be four clusters of variation in Figure 4.24: (1) 1 to 43000, (2) 50000 to 60000, (3) 75000 to 100000 and (4) 120000 to 130139. If we reconstruct the image into only these three groups pulse sizes we obtain the results in Figure 4.25.

We see that the full 'Tank' image in Figure 4.25(a). The respective variation clusters for Figures 4.25(b) and (c) represent the illumination in the image, and Figure 2.25(d) represents the larger noise pulses. We can separate the reconstruction in Figure 4.25(a) - see Figure 4.26. Figure 4.26(a) shows the underlying noise, Figure 4.26(b) shows the tank, but not in its entirety, and Figure 4.26(c) the rest of the tank from Figure 4.25(a). This image is very interesting because we are dealing with a camouflaged target. What is camouflage? It's simply obtained by adding the background pattern onto the feature. The image analysis technique thus finds it difficult to distinguish between background and the feature. From Figures 4.25 and 4.26 it is clear that we can't pick out the tank separately from other resolution layers. Further analysis for a camouflaged image is thus required.



(a) (b) (c) (d)

Figure 4.25. Reconstruction via DPT of Figure 4.17(c) with only pulses of size (a) 1 to 43000 (b) 50000 to 60000 (c) 75000 to 100000 (c) 120000 to 130139

(a) (b) (c)

Figure 4.26. Reconstruction via DPT of Figure 4.17(b) with only pulses of size (a) 1 to 50 (b) 51 to 8000 (c) 8001 to 43000

We turn to using shape analysis to detect the tank as a feature due to the camouflage in the image. The most logical choice for a shape descriptor is rectangularity as this is the distinguishing characteristic between the camouflaged tank and the background. Rectangularity is measured as the ratio between the pulse and its minimum bounding rectangle i.e. the number of pixels in the pulse divided by the number of pixels in the pulse's minimum bounding rectangle. This then gives a measure of how rectangular the pulse is. In Figure 4.27 we show how the rectangularity measure, applied to pulses of size 51 to 8000, picks out only the tank (although not in it's entirety) as the measure increases towards one, but removes the background.

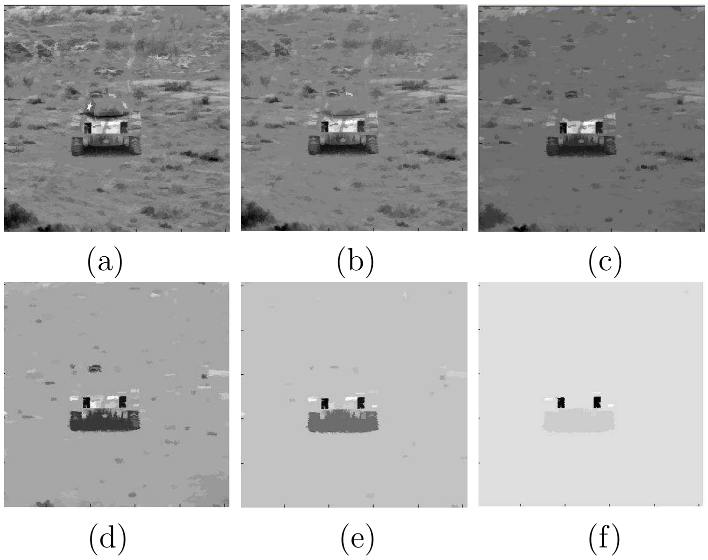
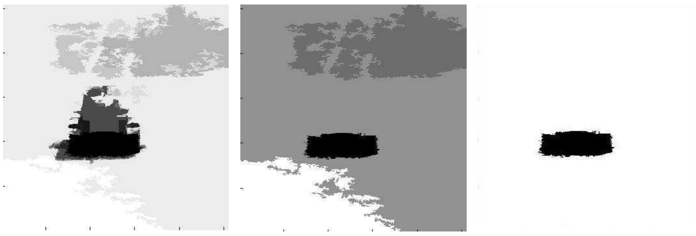


Figure 4.27. Restricting the Rectangularity of the Pulses represented in Figure 4.26(b) to (a) ≥ 0.3 (b) ≥ 0.4 (c) ≥ 0.5 (d) ≥ 0.6 (e) ≥ 0.7 (f) ≥ 0.8

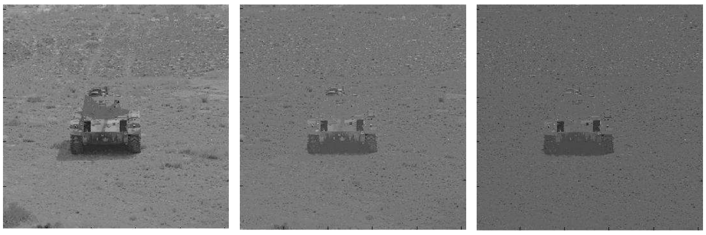
In Figure 4.28 we apply the same approach to pulses of size 8001 to 43000. Notice that again the tank is singled out and the large illumination pulses are removed.

To illustrate the importance of first removing the 'noise' and singling out the resolution layers which contain the feature of interest, we investigated the effect of the rectangularity measure on pulses 1 to 43000 (Figure 4.29) and on the original image (Figure 4.30). In both these figures we see that the tank is 'lost', thus illustrating the strength of the Discrete Pulse Transform in separating the features of interest from the rest of the image.



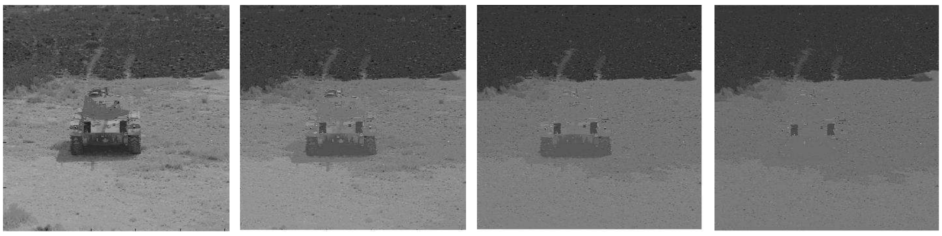
(a) (b) (c)

Figure 4.28. Restricting the Rectangularity of the Pulses represented in Figure 4.26(c) to (a) ≥ 0.5 (b) ≥ 0.6 (c) ≥ 0.7



(a) (b) (c)

Figure 4.29. Restricting the Rectangularity of the Pulses in Figure 4.25(a) to (a) ≥ 0.5 (b) ≥ 0.6 (c) ≥ 0.7



(a) (b) (c) (d)

Figure 4.30. Restricting the Rectangularity of all the Pulses to (a) ≥ 0.5 (b) ≥ 0.6 (c) ≥ 0.7 (d) ≥ 0.8

The variation spectrum for Figure 4.17(d), the 'Three Vehicles' image, is given in Figure 4.31. There seems to be three clusters of variation in Figure 4.31: (1) 1 to 41000, (2) 50000 to 60000 and (3) 100000 to 121447. If we reconstruct the image into only these three groups pulse sizes we obtain the results in Figure 4.32.

We can further remove the noise in Figure 4.32(a), see Figures 4.33 and 4.34. Notice that removing the smaller pulses removes the detail in the image but keeps the features of interest in tact.

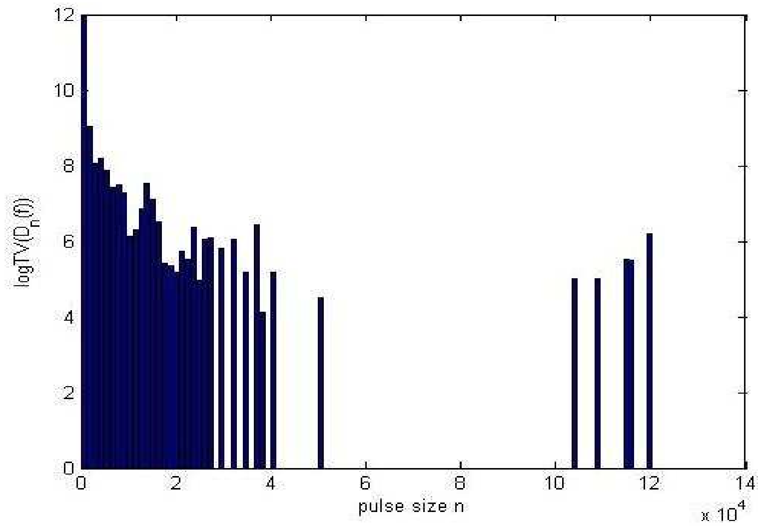
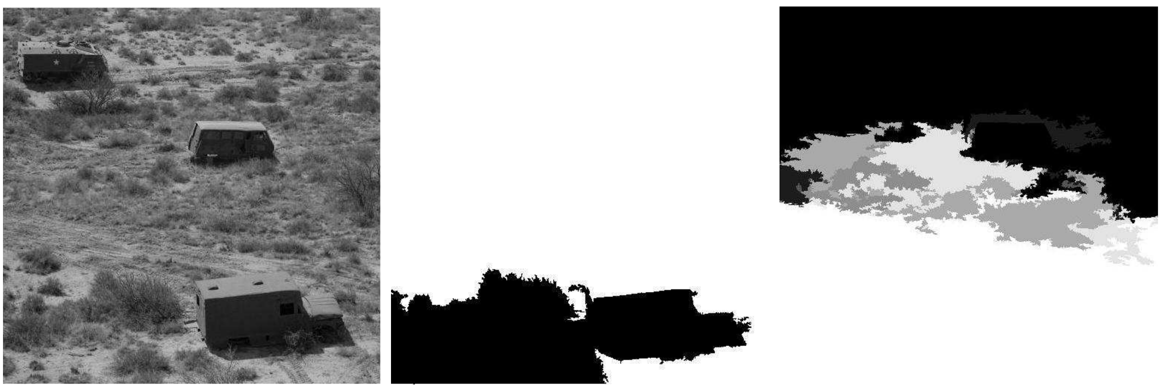
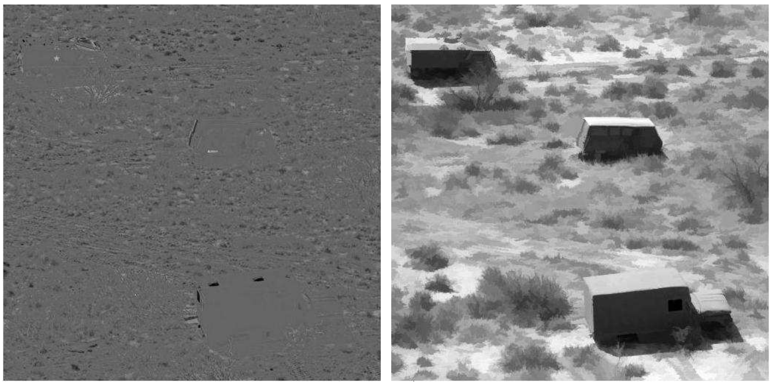


Figure 4.31. The Variation Spectrum of Figure 4.17(d).



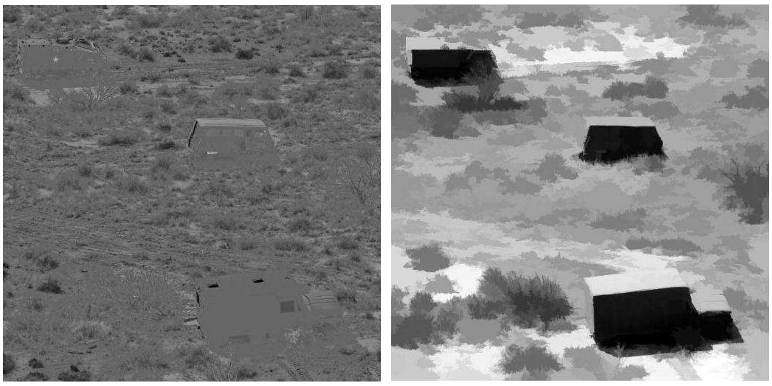
(a) (b) (c)

Figure 4.32. Reconstruction via DPT of Figure 4.17(d) with only pulses of size (a) 1 to 41000 (b) 50000 to 60000 (c) 100000 to 121477



(a) (b)

Figure 4.33. Reconstruction via DPT of Figure 4.17(d) with only pulses of size (a) 1 to 100 (b) 101 to 41000



(a) (b)

Figure 4.34. Reconstruction via DPT of Figure 4.17(d) with only pulses of size (a) 1 to 1000 (b) 1001 to 41000

The variation spectrum for Figure 4.17(e), the 'Ocean' image, is given in Figure 4.35. There seems to be three clusters of variation in Figure 4.31: (1) 1 to 45000, (2) 60000 to 70000 and (3) 80000 to 132370. If we reconstruct the image into only these three groups pulse sizes we obtain the results in Figure 4.36. An ocean image such as this one produces interesting complications. The glint on this water is considered noise when working at detecting a feature on the surface of the ocean. We thus would like to remove it prior to feature detection. We are able to separate the glint from the ocean, see Figure 4.37, by detecting the pulses which are the correct size as well luminosity.

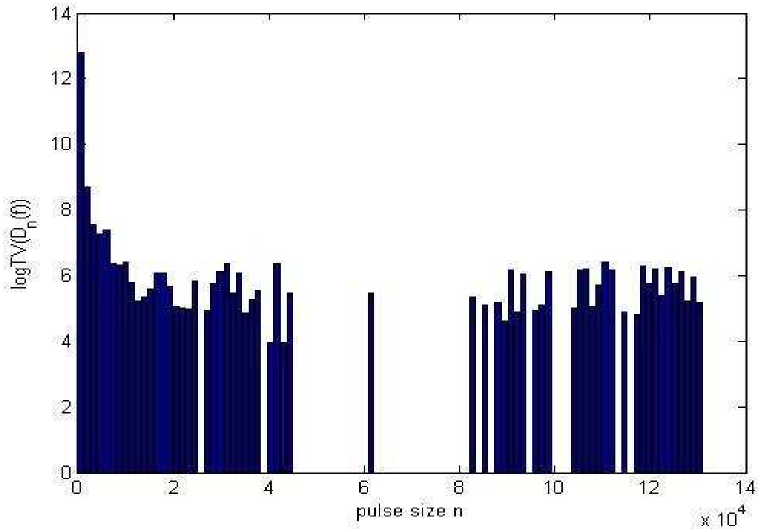
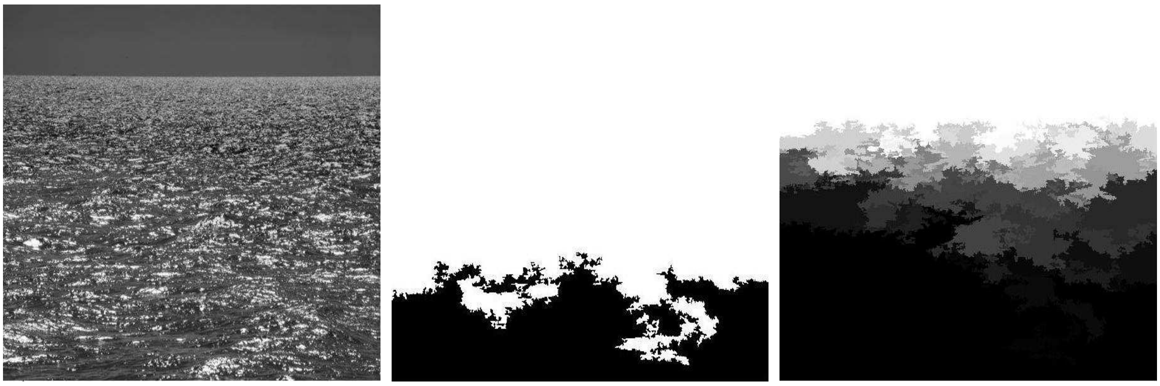
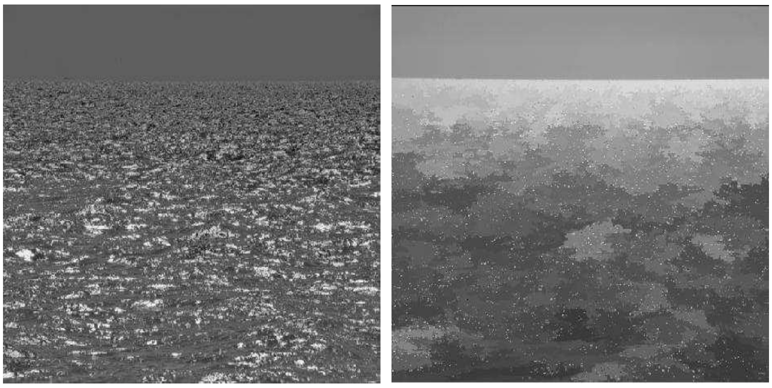


Figure 4.35. The Variation Spectrum of Figure 4.17(e).



(a) (b) (c)

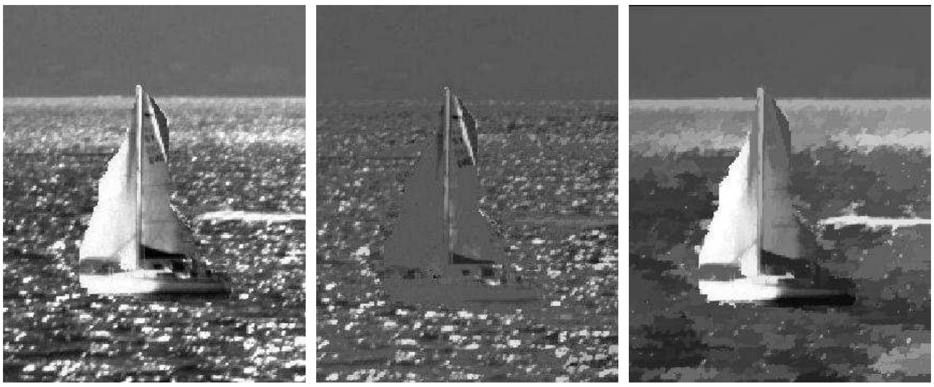
Figure 4.36. Reconstruction via DPT of Figure 4.17(e) with only pulses of size (a) 1 to 45000 (b) 60000 to 70000 (c) 80000 to 132379



(a) (b)

Figure 4.37. (a) Reconstruction via DPT of Figure 4.17(e) with only pulses of size 1 to 1000 and luminosity less than 25 (b) Removal of (a) from the original image.

We now apply this same method to an ocean image with glint and a target of interest. See Figure 4.38(a). We first detect the pulses small enough (in this case 1 to 500 as the image is a different size) and then restrict the luminosity to 25, see Figure 4.38(b). In Figure 4.38(c) we subtract (b) from the original to obtain the boat.



(a) (b) (c)

Figure 4.38. (a) Original Image (b) Reconstruction via DPT of Figure 4.38(a) with only pulses of size 1 to 500 and luminosity less than 25 (c) Removal of (b) from the original image.

The variation spectrum for Figure 4.17(f), the 'Noise' image, is given in Figure 4.39.

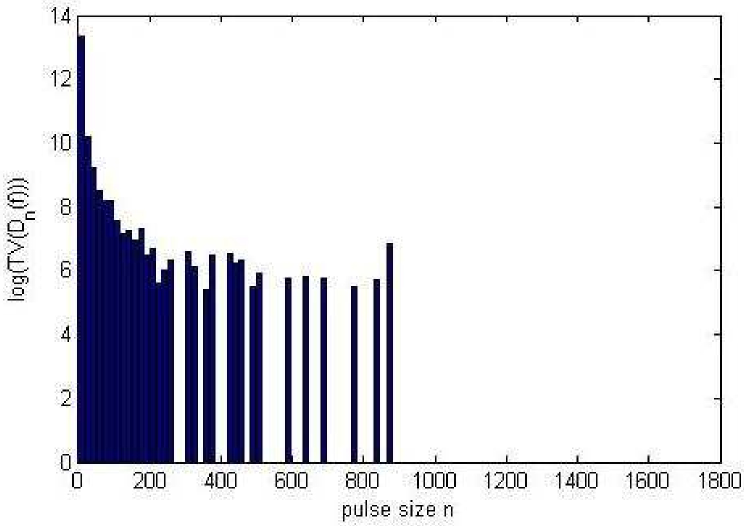


Figure 4.39. The Variation Spectrum of Figure 4.17(f).

There seems to be only one cluster of variation which gives a good indication of the structure of noise in general. We can however separate the noise into small noise, medium-sized noise and larger noise pulses, see Figure 4.40.

(a) (b) (c)

Figure 4.40. Reconstruction via DPT of Figure 4.17(f): Only pulses of size (a) 1 to 100 (b) 101 to 1000 (c) 1001 to 1660

4.5 Distributions of the LULU Operators

All the results of Section 2.5 (from [27]) can be extended into 2 dimensions (or more). We consider the cases

(i) $\{X_{ij}\}_{i,j=-\infty}^{\infty}$ independent and identically distributed with distribution function F_X.

(ii) $\{X_{ij}\}_{i,j=-\infty}^{\infty}$ independent and each with distribution function $F_{X_{ij}}$.

(iii) The asymptotic results are applied to the notation in case (i).

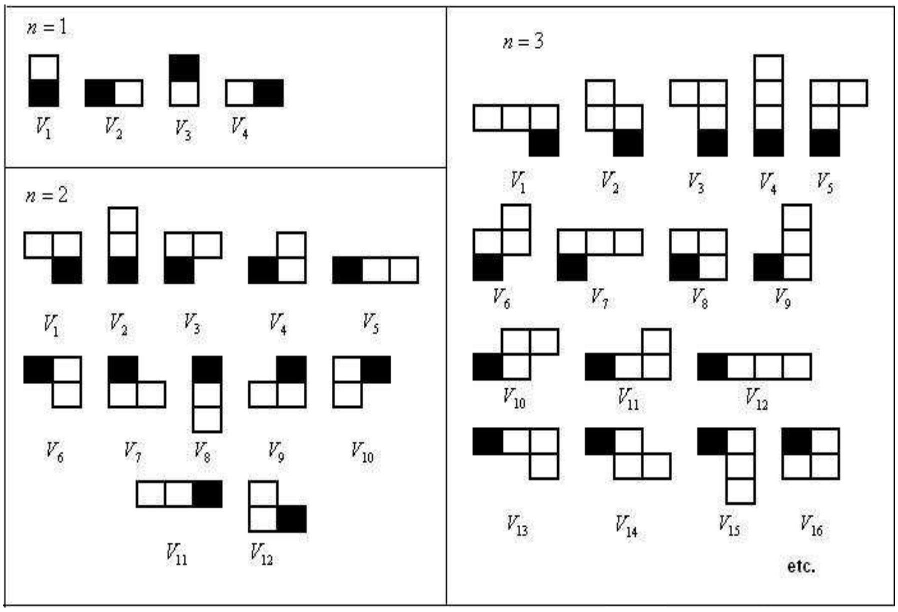


Figure 4.41. The ordering construction of the n-neighbourhoods

The order in the sequence ..., X_{-1}, X_0, X_1, X_2, ... is often made use of in the proofs for the one dimensional distributions thus we need to incorporate an 'order' to the two dimensional theory. This could easily done by ordering the neighbours in an n-neighbourhood in order of increasing distance away from the principal point, where distance is measured by the number of pixels between the two points along a connected set of neighbours. This method will introduce ambiguities though, as some points may lie at the same distance in a different direction. Thus, we can also order the N n-neighbourhoods in

$\mathcal{N}_n(i,j)$ by listing them in an order so that any two consecutive neighbourhoods in the order differ only by 1 element i.e. the 'movement' of one pixel. Note that for a fixed n, N will remain constant over each pixel (i,j). In Figure 4.41 we see that a single pixel in each neighbourhood is changed from one set to the next.

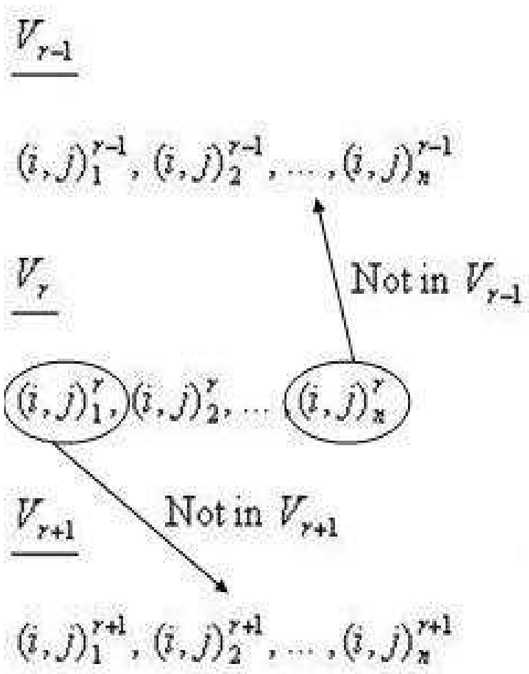


Figure 4.42. Structure of the n-neighbourhoods

Let $\{(i,j)^r_1, (i,j)^r_2, ..., (i,j)^r_n\}$ be the n points in the connected neighbourhood V_r with principal point (i,j) (with $X^{(1)}_{ij_r}, X^{(2)}_{ij_r}, ..., X^{(n)}_{ij_r}$ as the function values at these points), and $V_1, V_2, ..., V_N$ be the N neighbourhoods of $\mathcal{N}_n(i,j)$. Also, let $(i,j)_{(V_r^+)}$ be the element in V_r that is not present in the next set V_{r+1} and $(i,j)_{(V_r^-)}$ be the element in V_r that is not present in the previous set V_{r-1}. For convenience we let $(i,j)^r_1 = (i,j)_{(V_r^+)}$ and $(i,j)^r_n = (i,j)_{(V_r^-)}$, see Figure 4.42.

Now, let the observed points in the image be denoted by x_{ij} where (i,j) is the row and column position. Then the LULU operators are given by

$$\begin{aligned}(L_n x)_{ij} &= \max_{V\in\mathcal{N}_n(i,j)} \min_{(k,l)\in V} x_{kl},\\ (U_n x)_{ij} &= \min_{V\in\mathcal{N}_n(i,j)} \max_{(k,l)\in V} x_{kl}.\end{aligned}$$

The proofs for the distributions of $L_n(X)$, $U_n(X)$, $L_n \circ U_n(X)$ and $U_n \circ L_n(X)$, then follow similarly as in [27], but with some slight changes in the results and some changes in the notation to indicate the two-dimensionality. Note that we are not incorporating the concept of connectivity in the development of this theory. As a first step we work with the ordering developed above.

4.5.1 Exact Distributions of the LULU Operators

In this section we derive the distributions for $L_n(X)$, $U_n(X)$ and their compositions for an image of random variables X_{ij} where the X_{ij}'s are independent and identically distributed with distribution given by $F_X(x)$ i.e. where $X = X_{ij}\ \forall\ (i,j)$. In [21] the exact distributions for C_n and F_n where $n \leq 6$ are derived. We first define the following random variables. For an arbitrary pixel (i,j), consider $V_s \in \mathcal{N}_n(i,j)$ and let,

$$Y_{ij}^{(V_s)} = \max\{X_{ij_s}^{(1)}, X_{ij_s}^{(2)}, ..., X_{ij_s}^{(n)} : (i,j)^{(m)} \in V_s, m = 1,2,...,n\},$$

and then let

$$Z_{ij} = \min\{Y_{ij}^{(V_1)}, Y_{ij}^{(V_2)}, ..., Y_{ij}^{(V_N)} : V_s \in \mathcal{N}_n(i,j), s = 1,2,...,N\}$$

For each $V_s \in \mathcal{N}_n(i,j)$ let

$$A_{ij}^{(V_s)} = \min\{Z_{ij_s}^{(1)}, Z_{ij_s}^{(2)}, ..., Z_{ij_s}^{(n)} : (i,j)^{(m)} \in V_s, m = 1,2,...,n\}$$

(here Z represents the new two dimensional data after the application of max and min) and then let

$$B_{ij} = \max\{A_{ij}^{(V_1)}, A_{ij}^{(V_2)}, ..., A_{ij}^{(V_N)} : V_s \in \mathcal{N}_n(i,j), s = 1,2,...,N)\}.$$

These definitions mean then that $(L_n \circ U_n(X))_{ij} = B_{ij}$. This is the first difference for the procedure in [27] as there is only one possible forward window of size n in 1D for element X_i. The result for $U_n \circ L_n(X)$ can be derived from the result obtained for $L_n \circ U_n(X)$ since $L_n \circ U_n(-X) = -U_n \circ L_n(X)$ because L_n and U_n are dual operators, Theorem 70,

$$\begin{aligned} F_{U_n \circ L_n(X)}(x_0) &= P[U_n \circ L_n(X) \leq x_0] \\ &= P[-L_n \circ U_n(-X) \leq x_0] \\ &= P[L_n \circ U_n(-X) \geq -x_0] \\ &= 1 - P[L_n \circ U_n(-X) \leq -x_0] \\ &= 1 - F_{L_n \circ U_n(-X)}(-x_0). \end{aligned} \tag{4.24}$$

We will make use of the following definitions as well.

$$\begin{aligned} m_r^Y(x) &= P[Y_{ij}^{(V_1)} > x, Y_{ij}^{(V_2)} > x, ..., Y_{ij}^{(V_{r-1})} > x, Y_{ij}^{(V_r)} \leq x] \\ d_n(x) &= (1 - F_X(x))[F_X(x)]^n \\ q_r^Y(x) &= P[Y_{ij}^{(V_1)} > x, Y_{ij}^{(V_2)} > x, ..., Y_{ij}^{(V_r)} > x], \end{aligned}$$

These are thus for r of the N n-neighbourhoods, $V_1, V_2, ..., V_r$, in their order as mentioned before. (Note that the order is periodic since the neighbourhood relation is symmetric.) The Y in the definitions above can just as easily be replaced by Z, A or B. To get the distribution of $L_n \circ U_n(X)$ we first prove two lemmas.

Lemma 99

$$m_r^Y(x) = \begin{cases} d_n(x) & \textit{if } r = 2 \\ q_{r-2}^Y(x) d_n(x) & \textit{if } r \geq 3. \end{cases}$$

Proof
We have that

$$\begin{aligned} & \{Y_{ij}^{(V_r)} \leq x\} \cap \{Y_{ij}^{(V_{r-1})} > x\} = \{Y_{ij}^{(V_r)} \leq x\} \cap \{X_{ij_{r-1}}^{(1)} > x\} \\ = \; & \{X_{ij_{r-1}}^{(2)} \leq x, X_{ij_{r-1}}^{(3)} \leq x, ..., X_{ij_{r-1}}^{(n)} \leq x, X_{ij_{r-1}}^{(1)} > x\} \end{aligned}$$

Since the element $(i,j)_{r-1}^{(1)}$ is present in V_k for $k = r - n - 1, ..., r - 1$,

$$\{X_{ij_{r-1}}^{(1)} > x\} \subset \{Y_{ij}^{(V_k)} > x\} \text{ for } k = r - n - 1, ..., r - 1.$$

Thus we have

$$\{Y_{ij}^{(V_{r-n-1})} > x, ..., Y_{ij}^{(V_{r-1})} > x\} \cap \{X_{ij_{r-1}}^{(1)} > x\} = \{X_{ij_{r-1}}^{(1)} > x\}.$$

So

$$\begin{aligned} & \{Y_{ij}^{(V_1)} > x, ..., Y_{ij}^{(V_{r-1})} > x, Y_{ij}^{(V_r)} \leq x\} \\ = \; & \{Y_{ij}^{(V_1)} > x, ..., Y_{ij}^{(V_{r-n-2})} > x\} \cap \{Y_{ij}^{(V_{r-n-1})} > x, ..., Y_{ij}^{(V_{r-2})} > x\} \\ & \cap \{X_{ij_{r-1}}^{(1)} > x\} \cap \{Y_{ij}^{(V_r)} \leq x\} \\ = \; & \{Y_{ij}^{(V_1)} > x, ..., Y_{ij}^{(V_{r-n-2})} > x\} \\ & \cap \{X_{ij_{r-1}}^{(2)} \leq x, ..., X_{ij_{r-1}}^{(n)} \leq x, X_{ij_{r-1}}^{(1)} > x\}. \end{aligned} \tag{4.25}$$

So we have by (4.25) that for $r \geq 3$, $m_r^Y(x) = q_{r-n-2}^Y(x)d_n(x)$. For $r = 2$,

$$\{Y_{ij}^{(V_1)} > x, Y_{ij}^{(V_2)} \leq x\} = \{X_{ij_{r-1}}^{(2)} \leq x, ..., X_{ij_{r-1}}^{(n)} \leq x, X_{ij_{r-1}}^{(1)} > x\},$$

so that $m_r^Y(x) = d_n(x)$. ■

NOTE: For any events $A_1, A_2, ..., A_r$

$$\begin{aligned} & \{A_1 \cap ... \cap A_r\} \backslash \{A_1 \cap ... \cap A_{r-1} \cap A_r^C\} \\ = & \{A_1 \cap ... \cap A_r\} \cap \{A_1 \cap ... \cap A_{r-1} \cap A_r^C\}^C \\ = & \{A_1 \cap ... \cap A_r\} \cap \{A_1^C \cup ... \cup A_{r-1}^C \cup A_r\} \\ = & \{A_2 \cap ... \cap A_r\} \cup \{A_1 \cap A_3 \cap ... \cap A_{r-1}\} \\ & \cup ... \cup \{A_1 \cap ... \cap A_r\} \\ = & \{A_1 \cap ... \cap A_r\}. \end{aligned} \tag{4.26}$$

This result is in fact trivial because $A_1 \cap ... \cap A_{r-1} \cap A_r^C$ is disjoint from $A_1 \cap ... \cap A_r$.

Lemma 100 *For* $n = 1, 2, ...$

$$\begin{aligned} q_r^Y(x) &= P[Y_{ij}^{(V_1)} > x, ..., Y_{ij}^{(V_r)} > x] \\ &= \begin{cases} 1 - [F_X(x)]^n & \textit{for } r = 1 \\ q_1^Y(x) - d_n(x) & \textit{for } r = 2 \\ q_2^Y(x) - d_n(x) \sum_{k=1}^{r-2} q_k^Y(x) & \textit{for } r \geq 3. \end{cases} \end{aligned}$$

Proof
For $r = 1$,

$$q_r^Y(x) = P[Y_{ij}^{(V_1)} > x] = 1 - P[Y_{ij}^{(V_1)} \leq x] = 1 - P[X_{ij_1}^{(1)} \leq x, ..., X_{ij_1}^{(n)} \leq x] = 1 - [F_X(x)]^n.$$

For $r \geq 2$

$$\begin{aligned} & \{Y_{ij}^{(V_1)} > x, ..., Y_{ij}^{(V_r)} > x\} \backslash \{Y_{ij}^{(V_1)} > x, ..., Y_{ij}^{(V_{r-1})} > x, Y_{ij}^{(V_r)} \leq x\} \\ = & \{Y_{ij}^{(V_1)} > x, ..., Y_{ij}^{(V_r)} > x\} \text{ by (4.26)}, \end{aligned}$$

thus

$$q_r^Y(x) = q_{r-1}^Y(x) - m_r^Y(x)$$

and we recursively obtain

$$q_r^Y(x) = (q_{r-2}^Y(x) - m_{r-1}^Y(x)) - m_r^Y(x) = \ldots = q_s^Y(x) - \sum_{k=s+1}^{r} m_k^Y(x)$$

where $s \leq r-1$. For $r=2$ we take $s=1$:

$$q_r^Y(x) = q_1^Y(x) - m_2^Y(x) = q_1^Y(x) - d_n(x) \text{ by } 4.25$$

and for $r \geq 3$ take $s=2$, and again by 4.25,

$$q_r^Y(x) = q_2^Y(x) - \sum_{k=3}^{r} q_{k-2}^Y(x) d_n(x) = q_2^Y(x) - d_n(x) \sum_{k=1}^{r-2} q_k^Y(x)$$

■

And now for the main result. The result looks different to that obtained by [27] as Lemma's 99 and 100 are slightly different.

Theorem 101 *For $n = 1, 2, ...$*

$$F_{L_n \circ U_N(X)}(x) = 1 - q_2^Y(x) - N d_n(x) \sum_{k=nN-1}^{nN+N-2} q_k^Y(x) + d_n(x) \sum_{k=1}^{nN+N-2} q_k^Y(x)$$

where $N = 4(3^{n-2})$.

Proof

$$\begin{aligned} F_{L_n \circ U_n(X)}(x) &= P[B_{ij} \leq x] \\ &= P[A_{ij}^{(V_1)} \leq x, A_{ij}^{(V_2)} \leq x, ..., A_{ij}^{(V_N)} \leq x, \text{ for } V_s \in \mathcal{N}_n(i,j), s = 1, 2, ..., N] \\ &= p_N^A(x), \text{ say.} \end{aligned}$$

Now

$$\begin{aligned} p_N^A(x) &= P[A_{ij}^{(V_1)} \leq x, A_{ij}^{(V_2)} \leq x, ..., A_{ij}^{(V_N)} \leq x] \\ &= P[A^{(V_1)} \leq x, A^{(V_2)} \leq x, ..., A^{(V_N)} \leq x] \\ &\quad \text{for each } (i,j) \text{ since they are identically distributed} \\ &= P[A^{(V_1)} \leq x, A^{(V_2)} \leq x, ..., A^{(V_{N-1})} \leq x] \\ &\qquad - P[A^{(V_1)} \leq x, A^{(V_2)} \leq x, ..., A^{(V_{N-1})} \leq x, A^{(V_N)} > x] \text{ by } (4.26) \\ &= p_{N-1}^A(x) - P[A^{(V_2)} \leq x, A^{(V_3)} \leq x, ..., A^{(V_{N-1})} \leq x, A^{(V_N)} > x] \\ &\qquad + P[A^{(V_1)} > x, A^{(V_2)} \leq x, ..., A^{(V_{N-1})} \leq x, A^{(V_N)} > x] \text{ by } (4.26). \end{aligned}$$

The last term is equal to 0 by looking at the definition of A. Therefore, applying this recursively,

$$p_N^A(x) = p_{N-1}^A(x) - P[A^{(V_{N-1})} \leq x, A^{(V_N)} > x].$$

Now applying this recursively,

$$
\begin{aligned}
& p_N^A(x) \\
= \ & p_{N-2}^A(x) - P[A^{(V_{N-2})} \le x, A^{V_{N-1}} > x] - P[A^{(V_{N-1})} \le x, A^{(V_N)} > x] \\
= \ & p_1^A(x) - \sum_{k=1}^{N-1} P[A^{(V_k)} \le x, A^{(V_{k+1})} > x] \\
= \ & 1 - P[A^{(V_1)} > x] - \sum_{k=1}^{N-1} P[Z_{ij_k}^{(1)} \le x, Z_{ij_{k+1}}^{(1)} > x, ..., Z_{ij_{k+1}}^{(n)} > x] \\
= \ & 1 - P[A^{(V_1)} > x] - \sum_{k=1}^{N-1} P[Z_{ij_1}^{(1)} \le x, Z_{ij_2}^{(1)} > x, ..., Z_{ij_2}^{(n)} > x] \\
& \quad \text{since they are identically distributed} \\
= \ & 1 - q_n^Z(x) - (N-1)P[Z_{ij_1}^{(1)} \le x, Z_{ij_2}^{(1)} > x, ..., Z_{ij_2}^{(n)} > x] \\
= \ & 1 - q_n^Z(x) - (N-1)\big(P[Z_{ij_2}^{(1)} > x, ..., Z_{ij_2}^{(n)} > x] \\
& \quad - P[Z_{ij_1}^{(1)} > x, Z_{ij_2}^{(1)} > x, ..., Z_{ij_2}^{(n)} > x]\big) \\
= \ & 1 - q_n^Z(x) - (N-1)q_n^Z(x) + (N-1)q_{n+1}^Z(x) \\
= \ & 1 - Nq_n^Z(x) + (N-1)q_{n+1}^Z(x).
\end{aligned}
$$

$$
\begin{aligned}
\text{Now } q_n^Z(x) \ &= \ P[Z_{ij_s}^{(1)} > x, ..., Z_{ij_s}^{(n)} > x] \\
& \quad \text{i.e. for } n \text{ data points in the } s^{th} \text{ neighbourhood of } (i,j) \\
&= \ P[Y_{ij}^{(1,V_1)} > x, ..., Y_{ij}^{(1,V_N)} > x, ..., Y_{ij}^{(n,V_1)} > x, ..., Y_{ij}^{(n,V_N)} > x] \\
&= \ q_{nN}^Y(x)
\end{aligned}
$$

and similarly $q_{n+1}^Z(x) = q_{(n+1)N}^Y(x)$. So by Lemma 100

$$
\begin{aligned}
p_N^A(x) \ &= \ 1 - Nq_{nN}^Y(x) + (N-1)q_{(n+1)N}^Y(x) \\
&= \ 1 - N[q_2^Y(x) - d_n(x) \sum_{k=1}^{nN-2} q_k^Y(x)] \\
& \quad + (N-1)[q_2^Y(x) - d_n(x) \sum_{k=1}^{nN+N-2} q_k^Y(x)] \\
&= \ 1 - q_2^Y(x) - Nd_n(x) \sum_{k=nN-1}^{nN+N-2} q_k^Y(x) + d_n(x) \sum_{k=1}^{nN+N-2} q_k^Y(x).
\end{aligned}
$$

Since $d_1^Y(x) = 1 - [F_X(x)]^n$ and $q_2^Y(x) = q_1^Y(x) - d_n(x)$ these terms are known. The summations are functions of $q_r^Y(x)$ for $r \ge 3$ and can be found recursively using Lemma 4.25 since $nN + N - 2$ is finite. ■

Theorem 102 *For* $n = 1, 2, ...$

$$F_{U_n \circ L_n(X)}(x) = 1 - g(1 - F_X(x))$$

where $g(F_X(x)) = F_{L_n \circ U_n(X)}(x)$.

Proof

$$F_{U_n \circ L_n(X)}(x) = 1 - F_{L_n \circ U_n(-X)}(-x) = 1 - g(F_{-X}(-x))$$

and

$$F_{-X}(-x) = P[-X \leq -x] = P[X \geq x] = 1 - F_X(x).$$

■

Corollary 103 *For* $n = 1, 2, ...$

$$F_{U_n(X)}(x) = 1 - q_N^Y(x) \text{ and } F_{L_n(X)}(x) = 1 - F_{U_n(-X)}(-x).$$

Proof

$$F_{U_n(X)}(x) = P[Z_{ij} \leq x] = 1 - P[Z_{ij} > x] = 1 - P[Y_{ij}^{(V_1)} > x, ..., Y_{ij}^{(V_N)} > x] = 1 - q_N^Y(x)$$

where $q_N^Y(x)$ can be obtained recursively by Lemma 100. Then, we also have

$$F_{L_n(X)}(x) = P[-U_n(-X) \leq x] = P[U_n(-X) \geq -x] = 1 - F_{U_n(-X)}(-x).$$

■

Thus we have the distributions of $L_n \circ U_n(X)$ and $U_n \circ L_n(X)$ and the marginal distributions for $U_n(X)$ and $L_n(X)$.

As an illustration we show the distribution for $U_n(X), n = 2, 3, ..., 6$.

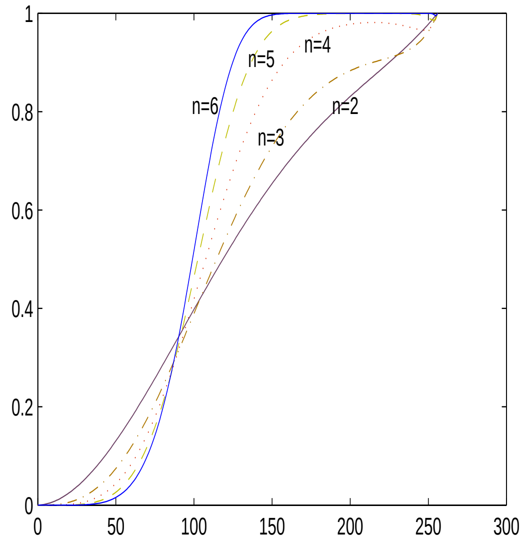


Figure 4.43. Distribution of U_n for $n = 2, 3, ..., 6$.

From Figure 4.43 we can see that the distribution becomes more vertical as n increase, indicating the smoothing nature of the nonlinear operator U_n.

4.5.2 General Distributions of the LULU Operators

For the case where $\{X_{ij}\}_{i,j=-\infty}^{\infty}$ are independent but non-identically distributed so that $P[X_{ij} \leq x] = F_{X_{ij}}(x) = F_{ij}(x)$, say, for $i, j = 0, \pm 1, ...$, a more sophisticated method than that used in the previous section is required to extend the results from [56] into two dimensions. This will be investigated in future research work.

4.5.3 Asymptotic Distributions of the LULU Operators

We now look at the asymptotic distributions. They are asymptotic in that $n \to \infty$ i.e. in the sense that we apply the LULU operators infinitely many times. The results from the one dimensional case are applied to the extreme order statistic $X_{(n)} = \max\{X_{ij}^{(1)}, ..., X_{ij}^{(n)}\}$ where $X_{ij}^{(1)}, ..., X_{ij}^{(n)}$ are the n neighbours of X_{ij} some neighbourhood of X_{ij} and the X_{ij}'s are i.i.d. The

Fisher-Tippett theorem can then be applied to $X_{(n)}$ in the same manner as in the one dimensional case so that if there exists sequences $(a_n), a_n > 0$ and (b_n) such that

$$\frac{1}{a_n}(X_{(n)} - b_n) \underset{\to}{D} H$$

(where H is a non-degenerate distribution) then H is one of three types: Fréchet, Gumbel or Weibull. The necessary and sufficient conditions for these three types are given in [71] and [48]. The Gumbel types have distribution

$$H_G(y) = e^{-e^{-y}},$$

the Fréchet types have distribution

$$H_F(y) = e^{-y^{-\alpha}} \text{ for } y > 0 \text{ and } \alpha > 0$$

and the Weibull types have distribution

$$H_W(y) = \begin{cases} e^{-(-y)^{\alpha}} & \text{for } y \leq 0 \text{ and } \alpha < 0 \\ 1 & \text{for } y > 0 \end{cases}$$

for some constant α which is the shape parameter and reflects the weight of the tail for F_X. The Fisher-Tippett theorem is to maxima what the Central limit theorem is to averages. The most common type are the Gumbel types which have (normalized) density function, [133],

$$h_G(x) = \frac{1}{b} e^{\frac{a-x}{b} - e^{\frac{a-x}{b}}}$$

and distribution function

$$H_G(x) = e^{-e^{\frac{a-x}{b}}},$$

where a is the location parameter and b is the scale parameter, see Figure 4.44, [133].

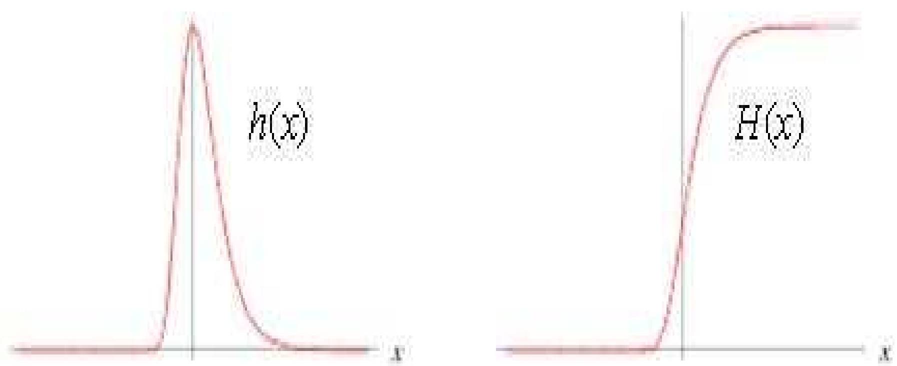


Figure 4.44. Density Function $h_G(x)$ and Distribution Function $H_G(x)$.

The distribution function of X_{ij}, namely F_X, is said to then belong to the maximum domain of attraction of H, i.e. $F_X \in MDA(H)$. The asymptotic distribution H has mean $a + b\gamma$, where γ is the Euler-Mascheroni constant, variance $\frac{1}{6}b^2\pi^2$, skewness (degree of symmetry) $\frac{12\sqrt{6}\xi(3)}{\pi^3}$, where $\xi(3) = 1.2020569032$ is Apéry's constant correct to 10 decimal places, [131], and kurtosis (degree of peakedness) $\frac{12}{5}$. The Euler-Mascheroni constant is the limit of

$$\sum_{k=1}^{n} \frac{1}{k} - \ln n$$

as $n \to \infty$ and correct to 10 decimal places is given by 0.5772156649, [132]. So the result in [27] holds in two dimensions as well.

Theorem 104 *If $F_X \in MDA(H)$, and sequences $(a_n), a_n > 0$ and (b_n) exist, then as $n \to \infty$ we have*

$$J_n(x) = F_{U_n(X)}(a_n x + b_n) \underrightarrow{D} J(x) = H(x) - H(x)logH(x)$$

and

$$G_n(x) = F_{L_n \circ U_n(X)}(a_n x + b_n) \underrightarrow{D} H(x) - H(x)logH(x) + \frac{1}{2}[H(x)logH(x)]^2 = G(x)$$

4.6 Conclusion

The Discrete Pulse Transform was derived for two dimensions and applications of it on images presented. The separation of an image into noise and signal is clearly presented in Section 4.4. In fact, throughout Section 4.4 we emphasized target or feature detection. This was achieved by two means, namely extracting pulses of the desired size and using shape descriptors. In the first method, the feature size must then be known to some degree, and in the second, similarly, the shape of what we are looking for needs to be known. Automatic target detection will be looked at in the future, as the obvious extension of these initial findings. In each of the examples presented we could additionally highlight the desired pulses (or whole resolution layers) to emphasize the detected features. When the Highlight result from Section 2.4 is extended to two dimensions (this will, to our knowledge, be published soon) it will provide further advantages for image processing, although the applications already available, presented in this work, provide powerful results.

Chapter 5

Conclusion

We have presented the extension of Carl Rohwer's LULU theory in higher dimensions than that of sequences, specifically for multidimensional arrays. Prior to this, we did a review of Rohwer's one dimensional theory from the point of view of our extension, that is, the theory developed in one dimension which we extended to higher dimensions. This included the following: the basic definitions and their relationship with Mathematical Morphology, the smoother and separator properties and the resulting action on signal due to these properties, the LULU semigroup, the preservation properties of fully trend preserving and total variation preserving, the distributions of the operators, and finally the Discrete Pulse Transform. All these concepts were extended in Chapters 3 and 4 and are new results and original work. Specifically, the new results are

- The Multidimensional LULU Operators (Section 3.3)
- The LULU Semigroup (Section 3.4)
- Preservation Properties of the LULU Semigroup (Section 3.5)
- Total Variation Preservation (Section 3.6)
- The DPT (Section 4.2)
- The Roadmaker's Algorithm and it's Implementation (Section 4.3)
- Image Processing (Section 4.4)
- Distributions of the LULU Operators (Section 4.5)

The importance and need of the extension has been mentioned extensively in the literature, as mentioned at many places in the previous chapters.

The applicability of the extension for images, and more complex domains e.g. video, is now open for investigation. We provided simple applications of the Discrete Pulse Transform to images in Chapter 4, but there are many image processing ideas that have sprung up during this research which will be further investigated in the near future.

We have one publication thus far from this work, namely [41] and in addition the following future research will be investigated:

1. The extension of the DPT onto $\mathbb{Z}^d$ will be investigated. This will most likely be achieved with the definition of total variation on $\mathbb{Z}^d$ as defined in [65] on a graph. The extension of the distributions on $\mathbb{Z}^d$ could also be investigated in the setting of a graph, possibly removing the requirement of independence between pixels. In addition, the new method introduced in [21] for deriving the distributions of the output of the LULU operators on sequences will be investigated for an extension to $\mathbb{Z}^d$.

2. The distributions of the LULU operators for the non-identically distributed case will be investigated. The distribution theory and results will also be applied to image processing applications such as edge detection.

3. The extension of the LULU theory to an arbitrary domain Ω, which could be discrete or continuous, will provide an even wider field of applications.

4. Various image processing applications, making use of the DPT, can be looked into. The most important relationship with existing image techniques may be that of scale spaces. We will investigate in what manner we can cast the DPT in the setting of scale-space theory, [69]. This setting is suitable since there exists a natural scale space structure that can be associated with the extensive current theory on scale spaces. Making use of scale spaces or multi-resolution methods to analyze an image for feature detection allows the use of more information than the pixel luminosity only. The most commonly used scale space is the Gaussian scale space. One drawback of the Gaussian scale space is that it removes small scale features (noise) very well but results in spatial distortions as scale increases, i.e. reduced sharpness of edges

and shapes. The DPT is expected to avoid this disadvantage due to the shape preservation properties of the LULU operators. Other image processing applications that should be looked into are

- Further applications of the DPT to important problems in image segmentation: Will segmentation making use of the DPT provide better results than ordinary segmentation?
- Granulometries: This was discussed in this work but more detailed work will be looked into.
- Image Compression: Can the DPT reduce the storage space required for an image?
- Pattern Recognition and template matching: The DPT contains more information from the image than the original image. Can this be exploited? Pattern recognition should also be looked into by investigating shape descriptors and their relationship with the pulses. A pattern should be able to be described with a specific shape, which in turn should differentiate it from another pattern.

There are numerous image processing techniques which investigate the above applications. The aim would thus be to compare the DPT's effectiveness with known techniques and to indicate where it is superior.

Bibliography

[1] C R Adams and J A Clarkson, Properties of Functions $f(x, y)$ of Bounded Variation, Transactions of American Mathematical Society, **36**(4), 1934.

[2] B Anguelov, Discrete Pulse Transform of Images: Algorithm and Applications, Proceedings of ICPR 2008: 19th International Conference on Pattern Recognition, Tampa, Florida, USA, 8-11 December 2008.

[3] R Anguelov, LULU Operators and Locally δ-Monotone Approximations, *Constructive Theory of Functions*, Marin Drinov Academic Publishing House, Sofia, 2006, 22-34.

[4] R Anguelov and C H Rohwer, LULU Operators for Functions of Continuous Argument, Quaestiones Mathematicae, **32**(2) (2009) 187–202.

[5] G R Arce, N C Gallager and T A Nodes, Median Filters: Theory for One- and Two-Dimensional Filters, *Advances in Computer Vision and Image Processing*, T S Huang (editor), Springer Verlag, JAI Press, **Vol. 2**, 1986.

[6] G R Arce, A General Weighted median Filter Structure Admitting Negative Weights, IEEE Transactions on Signal Processing, **46**(12) (1998) 3195–3205.

[7] E Ataman, V K Aatre and K M Wong, A Fast Method for Real-Time Median Filtering, IEEE Transactions on Acoustics, Speech and Signal Processing, **28**(4) (1980) 415–421.

[8] E Ataman, V K Aatre and K M Wong, Some Statistical Properties of Median Filters, IEEE Transactions on Acoustics, Speech and Signal Processing, **29**(5) (1981) 1073–1075.

[9] A Bijaoui, A Murtagh and J L Starck, *Image Processing and Data Analysis - The Multiscale Approach*, Cambridge Univeristy Press, 1998.

[10] P Blomgren, T F Chan, P Mulet and C K Wong, Total Variation Image Restoration: Numerical Methods and Extensions, Proceedings of ICIP 1997: International Conference on Image Processing, Washington DC, USA, 26-29 October 1997, 384–387.

[11] A Bovik, T Huang and D Munson, A Generalization of Median Filtering using Linear Combinations of Order Statistics, IEEE Transactions on Acoustics, Speech and Signal Processing, **31**(6) (1983) 1342–1350.

[12] U M Braga-Neto and J Goutsias, An Axiomatic Approach to Multiscale Connectivity in Image Analysis, Technical Report, (2001).

[13] U M Braga-Neta and J Goutsias, A Multiscale Approach to Connectivity, Computer Vision and Image Understanding, **89** (2003) 70–107.

[14] U M Braga-Neto and J Goutsias, A Theoretical Tour of Connectivity in Image Processing and Analysis, Journal of Mathematical Imaging and Vision, **19** (2003) 5–31.

[15] U M Braga-Neto and J Goutsias, Grayscale Level Connectivity: Theory and Applications, IEEE Transactions on Image Processing, **13**(12) (2004) 1567–1580.

[16] U M Braga-Neta, Multiscale Connected Operators, Journal of Mathmatical Imaging and Vision, **22** (2005) 199–216.

[17] U M Braga-Neto, Object-Based Image Anlaysis Using Multiscale Connectivity, IEEE Transactions on Pattern Analysis and Machine Intelligence, **27**(6) (2005).

[18] U M Braga-Neto and J Goutsias, Constructing Multiscale Connectivities, Computer Vision and Image Understanding, **99** (2005) 126–150.

[19] E J Breen and R Jones, Attribute Openings, Thinnings, and Granulometries, Computer Vision and Image Understanding, **64**(3) (1996) 377–389.

[20] D R K Brownrigg, The Weighted Median Filter, Communications of the ACM, **27**(8) (1984) 807–818.

[21] P W Butler, The Transfer of Distributions of LULU Smoothers, Masters thesis, University of Stellenbosch, 2008.

[22] A Chambolle and P-L Lions, Image Recovery via Total Variation Minimization and Related Problems, Numerische Mathematik, **76** (1997) 167–188.

[23] T F Chan, G H Golub and P Mulet, A Nonlinear Primal-Dual Method for Total Variation-Based Image Restoration, Lecture Notes in Control and Information Systems, **219** (1996) 241–252.

[24] T Chan, S Esegoglu, F Park and A Yip, Recent Developments in Total Variation Image Restoration, *Mathematical Models of Computer Vision*, N Paragios, Y Chena and O Faugeras (editors), Springer Verlag, 2005.

[25] J A Clarkson and C R Adams, On Definitions of Bounded Variation for Functions of Two Variables, Transactions of American Mathematical Society, **35**(4) (1933).

[26] W J Conradie, T de Wet and M D Jankowitz, An Overview of LULU Smoothers with Application to Financial Data, Journal for Studies in Economics and Econometrics, **29**(1) (2005) 97–121.

[27] W J Conradie, T de Wet, M Jankowitz, Exact and Asymptotic Distributions of LULU Smoothers, Journal of Computational and Applied Mathematics, **186** (2006) 253–267.

[28] W J Conradie, T de Wet and M D Jankowitz, Preformance of Nonlinear Smoothers in Signal Recovery, Applied Stochastic Models in Business and Industry, **25**(4) (2009) 425–444.

[29] T de Wet and W Conradie, Smoothing Sequences of Data by Extreme Selectors, Proceedings of ICOTS7 2006: The 7th International Conference on Teaching Statistics, Salvador, Bahia, Brazil, 2-7 July 2006.

[30] J Crespo and R W Schafer, Locality and Adjacency Stability Constraints for Morphological Connected Operators, Journal of Mathematical Imaging and Vision **7** (1997) 85-102.

[31] J Crespo and V Maojo, New Results on the Theory of Morphological Filters by Reconstruction, Pattern Recognition, **31**(4) (1998) 419–429.

[32] J Crespo, Adjacency Stable Connected Operators and Set Levelings, Proceedings of ISMM'07: The 8th International Symposium on Mathematical Morphology, Rio de Janeiro, Brazil, 10-13 October 2007, 215-226.

[33] J Crespo, J Serra and R W Schafer, Theoretical Aspects of Morphological Filters by Reconstruction, Signal Processing, **47** (1995) 201–225.

[34] J Crespo and V Maojo, The Strong Property of Morphological Connected Alternated Filters, Journal of Mathematical Imaging and Vision, **32**(3) (2008) 251-263.

[35] D C Dobson and C R Vogel, Convergence of an Iterative Method fo Image Denoising, SIAM Journal of Numerical Analysis, **34**(5) (1997) 1779–1791.

[36] E R Dougherty and J Astola, *An Introduction to Nonlinear Image Processing*, SPIE Optical Engineering Press, Washington, 1994.

[37] A Doulamis, N Doulamis and P Maragos, Generalized Multiscale Connected Operators with Applications to Granulometric Image Analysis, Proceedings of ICIP: The International Conference on Image Processing, Thessaloniki, Greece, 7-10 October 2001, **3** 684–687.

[38] J P du Toit, The Discrete Pulse Transform and Applications, Msc. Dissertation at the University of Stellenbosch, South Africa, 2007.

[39] P Embrechts, C Klüppelberg and T Mikosch, *Modelling Extremal Events for Insurance and Finance*, Springer, New York, 1997.

[40] S S Epp, *Discrete Mathematics with Applications*, International Thomson Publishing, 1995.

[41] R Anguelov and I N Fabris-Rotelli, Discrete Pulse Transform of Images, Proceedings of ICISP 2008: International Conference on Image and Signal Processing, Cherbourg-Octeville, Normandy, France, 1-3 July 2008, Lecture Notes in Computer Science, **5099** (2008) 1-9.

[42] J Fitch, E Coyle and N Gallagher, Root Properties and Convergence Rates of median Filters, IEEE Transactions on Acoustics, Speech and Signal Processing, **33**(1) (1985) 230–240 .

[43] J B Fraleigh and R A Beauregard, *Linear Algebra*, Third Edition, Addison-Wesley Publishing Company, 1995.

[44] C Frohn-Schauf, S Henn and K Witsch, Nonlinear Multigrid Methods for Total Variation Image Denoising, Computing and Visualization in Science, **7** (2004) 199-206.

[45] N C Gallagher and G L Wise, A Theoretical Analysis of the Properties of Median Filters, IEEE TRansactions on Acoustics, Speech and Signal Processing, **29**(6) (1981) 1136–1141.

[46] N C Gallagher and G Wise, A Theoretical Analysis of the Properties of Median Filters, IEEE Transactions on Acoustics, Speech and Signal Processing, , **29**(6) (1981) 1136–1141.

[47] G Garibotto and L Lambarelli, Fast On-Line Implementation of Two-Dimensional Median Filtering, Electronic Letters, **15**(1) (1979) 24–25.

[48] B V Gnedenka, Sur la Distribution Limite du Terme Maximum d'une Série Aléatoire, Annals of Mathematics, **44** (1943) 423–453.

[49] C Goodall, A Survey of Smoothing Techniques, *Modern Methods of Data Analysis*, J Fox and J S Long (editors), Sage Publications, 1990.

[50] H Greenspan, C H Anderson and S Akber, Image Enhancement by Nonlinear Extrapolation in Frequency Space, IEEE Transactions on Image Processing, **9**(6) (2000) 1035–1048.

[51] A Haar, Zur Theorie der Orthogonalen Funktionsysteme, Mathematische Annalen, **69** (1910) 331–371.

[52] R W Hamming, *Digital Filters*, Prentice Hall, N.J., 1956.

[53] G Hellman, Mathematical Constuctivism in Spacetime, The British Journal for the Philosophy of Science, **49**(3) (1998) 425–450.

[54] H Hwang and R A Haddad, Adaptive Median Filters: New Algorithms and Results, IEEE Transactions on Image Processing, **4**(4) (1995) 499–502.

[55] T S Huang, G J Yang and G Y Tang, A Fast Two-Dimensional Median Filtering Algorithm, IEEE Transactions on Acoustics, Speech, and Signal Processing, **27**(1) (1979) 13–18.

[56] M D Jankowitz, Some Statistical Aspects of LULU Smoothers, PhD Dissertation for the Department of Statistics and Actuarial Science at Stellenbosch University, South Africa, 2007.

[57] B I Justusson, Median Filtering: Statistical Properties, *Topics in Applied Physics: Two-Dimensional Digital Signal Processing*, T S Huang (editor), Springer Verlag, 1981.

[58] O Kao, Modification of the LULU Operators for Preservation of Critical Image Details, Proceedings of CISST'02: International Conference on Imaging Science, Systems and Technology, Las Vegas, Nevada, USA, 24-27 June 2002.

[59] J-C Klein, Conception et rëalisation d'une unitë logique pour l'anakyse quantitative d'images, PhD thesis, University of Nancy, 1976.

[60] S-J Ko, Y H Lee and A T Fam, Efficient Implementation of One-Dimensional Recursive Median Filters, IEEE Transactions on Circuits and Systems, **37**(11) (1990) 1447–1450.

[61] I V Kolpakov, Application of Atomic Functions in Problems of Signal's Preprocessing, Proceedings of MSMW'04: The Fifth International Khardov Symposium on Physics and Engineering of Microwaves, Millimeter, and Submillimeter Waves, Kharkov, Ukraine, 21-26 June 2004.

[62] E Kreyzig, *Introductory Functional Analysis with Applications*, John Wiley & Sons, 1978.

[63] D P Laurie and C H Rohwer, The Discrete Pulse Transform, SIAM Journal on Mathematical Analysis, **38**(3) 2007.

[64] D P Laurie and C H Rohwer, Fast Implementation of the Discrete Pulse Transform, Extended Abstracts of ICAAM 2006: International Conference of Numerical Analysis and Applied Mathematics, Crete, Greece, 15-19 September 2006, 484–487.

[65] D P Laurie, The Roadmaker's Algorithm and the Discrete Pulse Transform, private communication with the author on a paper submitted for publication, 2009.

[66] D P Laurie, The Roadmaker's Algorithm, Technical Report, (2005).

[67] Y Li and F Santosa, A Computational Algorithm for Minimizing Total Variation in Image Restoration, IEEE Transactions on Image Processing, **5**(6) (1996) 987–995.

[68] H-M Lin and A N Willson, Median Filters with Adaptive Length, IEEE Transactions on Circuits and Systems, **35**(6) (1988) 675–690.

[69] T Lindeberg, Scale-Space: A Framework for Handling Image Structures at Multiple Scales, Proceedings of CERN School of Computing, Egmond aan Zee, The Netherlands, 821 September, 1996.

[70] M Lindenbaum, M Fischer and A Bruckstein, On Gabor's Contribution to Image Enhancement, Pattern Recognition, **27**(1) (1994) 1–8.

[71] F M Longin, Asymptotic Distribution of Extreme Stock Market Returns, The Journal of Business, **69**(3) (1996) 383–408.

[72] F Malgouyres, Mathematical Analysis of a Model which Combines Total Variation and Wavelet for Image Restoration, Workshop on Image Processing and Related Mathematical Topics, **2**(1) (2002) 1–10.

[73] E Malkowsky and C H Rohwer, The LULU-Semigroup for Envelopes of Functions, Quaestiones Mathematicae **27** (2004) 89–97.

[74] S G Mallet, A Theory for Multiresolution Signal Decomposition: The Wavelet Representation, IEEE Transactions on Pattern Analysis and Machine analysis, **11**(7) (1989) 674–693.

[75] C L Mallows, Some Theory of Nonlinar Smoothers, Annals of Statistics **8** (1980) 695–715.

[76] D Marr, *Vision*, W H Freeman and Company, 1982.

[77] G Matheron, *Eléments pour une Théorie des Milieux Poreux*, Masson, Paris, 1967.

[78] G Matheron, Filters and lattices, *Image Analysis and Mathematical Morphology, Volume II: Theoretical Advances*, J Serra (editor), Academic Press: London, Chap. 6, 115–140, 1988.

[79] M Nitzberg and T Shiota, Nonlinear Image Filtering with Edge and Corner Enhancement, IEEE TRansactions on Pattern Analysis and Machine Intelligence, **14**(8) (1992) 826–833.

[80] T Nodes and N Gallagher, Median filters: Some modifications and Their Properties, IEEE Transactions on Acoustics, Speech and Signal Processing, **30**(5) (1982) 739–746.

[81] T Nodes and N Gallagher, The Output Distribution of Median Type Filters, IEEE Transactions on Communications, **32**(5) (1984) 532–541.

[82] M Omair Ahmed, D Sundararajan, A Fast Algorithm for Two-Dimensional Median Filtering, IEEE Transactions on Circuits and Systems, **34**(1) (1987) 1364–1374.

[83] I Pitas and A N Venetsanopoulos, Order Statistics in Digital Image Processing, Proceedings of the IEEE, **80**(12) (1992) 1893–1921.

[84] I Pollak, A S Willsky and H Krin, Image Segmentation and Edge Enhancement with Stabalized Inverse Diffusion Equations, IEEE Transactions on Image Processing, **9**(2) (2000) 256–266.

[85] C H Rohwer, Idempotent One-Sided Approximation of Median Smoothers, Journal of Appoximation Theory, **58**(2) (1989) 151–163.

[86] C H Rohwer and L M Toerien, Locally Monotone Robust Approximation of Sequences, Journal of Computational and Applied Mathematics, **36** (1989) 399–408.

[87] C H Rohwer, Projections and Separators, Quaestiones Mathematicae, **22** (1999) 219–230.

[88] C H Rohwer, Variation reduction and $LULU$-smoothing, Quaestiones Mathematicae **25** (2002) 163–176.

[89] C H Rohwer and M Wild, Natural Alternatives for One Dimensional Median Filtering, Quaestiones Mathematicae, **25** (2002) 135–162.

[90] C H Rohwer, Multiresolution Analysis with Pulses, International Series of Numerical Mathematics, **142** (2002) 165–186.

[91] C H Rohwer, Fast Approximation with Locally Monotone Sequences, Proceedings of the Fourth International Conference on Functional Analysis and Approximation Theory, Potenza, Italy, 22-28 September 2000, **68**(2) 777–790.

[92] C H Rohwer, Fully trend preserving operators, Quaestiones Mathematicae **27** (2004) 217–230.

[93] C H Rohwer, *Nonlinear Smoothers and Multiresolution Analysis*, Birkhäuser, 2005.

[94] C H Rohwer, The Estimation of Moments of an Unknown Error Distribution in the Discrete Pulse Transform, Numerical Algorithms, **45** (2007) 239–251.

[95] C H Rohwer and M Wild, *LULU* Theory, Idempotent Stack Filters, and the Mathematics of Vision of Marr, Advances in Imaging and Electron physics **146** (2007) 57–162.

[96] C Ronse, Set-Theoretic Algebraic Approaches to Connectivity in Continuous or Digital Spaces, Journal of Mathematical Imaging and Vision, **8** (1998) 41–58.

[97] C Ronse, Partial Partitions, Partial Connections and Connective Segmentation, Journal of athematical Imaging and Vision, **32** (2008) 97–125.

[98] H L Royden, *Real Analysis*, Macmillan, 1969.

[99] L I Rudin, S Osher and E Fatemi, Nonlinear Total Variation Based Noise Removal Algorithms, Physica D, **60** (1992) 259–268.

[100] P Salembier and J Serra, Multiscale Image Segmentation, Proceedings of SPIE: Visual Communications and Image Processing, Boston, MA, 18-20 November 1992, **1818** 620–631.

[101] P Salembier and J Serra, Flat Zones Filtering, Connected Operators and Filters by Reconstruction, IEEE Transactions on Image Processing, **4**(8) (1995).

[102] P Salembier and A Oliveras, Practical Extensions of Connected Operators, *Mathematical Morphology and its Applications to Image and Signal Processing*, P A Maragos, R W Schafer and M A Butt (editors), 1996.

[103] J G M Schavemaker, M J T Reinders, J J Gerbrands and E Backer, Image Sharpening by Morphological Filters, Pattern Recognition, **33** (2000) 997–1012.

[104] J Serra, *Image Analysis and Mathematical Morphology*, Academic Press, London, 1982.

[105] J Serra, Mathematical Morphology for Boolean Lattices, *Image Analysis and Mathematical Morphology, Volume II: Theoretical Advances*, Academic Press, 1988.

[106] J Serra, *Image Analysis and Mathematical Morphology, Volume II: Theoretical Advances*, J. Serra (editor), Academic Press, London, 1988.

[107] J Serra, Connectivity on Complete Lattices, Journal of Mathematical Imaging and Vision, **9** (1998) 231–251.

[108] J Serra, A Lattice Approach to Image Segmentation, Journal of Mathematical Imaging and Vision, **24** (2006) 83–130.

[109] Y Sheng, Wavelet Transform, *The Electrical Engineering Handbook Series: The Transforms and Applications Handbook*, A D Poularikas (editor), Boca Raton, Florida, USA, CRC Press, 1996, 747–827.

[110] J E Shore, On the Application of Haar Functions, NTIS, US Department of Commerce (NRL Report 7467), 1973.

[111] P Soille, *Morphological Image Analysis*, Springer, 1999.

[112] P Soille, Constrained Connectivity for Hierarchical Image Partitioning and Simplification, IEEE Transactions on Pattern Analysis and Machine Intelligence, **36**(7) (2008).

[113] R S Stanković and B J Falkowski, The Haar Wavlet Transform: It's Status and Achievments, Computers and Electrical Engineering, **29** (2003) 25–44.

[114] J-L Starck, F Murtagh, E J Candes and D L Donoho, Gray and Color Image Contrast Enhancemnet by Curvelet Transform, IEEE Transactions on Image Processing, **12**(6) (2003) 706–717.

[115] J W Tukey, *Exploratory Data Analysis*, Addison Wesley, Reading, Mass, 1977.

[116] S G Tyan, Median Filtering: Deterministic Properties, *Topics in Applied Physics: Two-Dimensional Digital Signal Processing*, T S Huang (edior), Springer Verlag, 1981.

[117] C Y Tyan and P P Wang, Image Processing - Enhancement, Filtering and Edge Detection Using Fuzzy Logic Approach, Proceedings of Second IEEE International Conference on Fuzzy Systems, San Francisco, California, 28 March - 1 April 1993, **1** 600–605.

[118] C S Tzafestas and P Maragos, Shape Connectivity: Multiscale Analysis and Application to Generalized Granulometries, Journal of Mathematical Imaging and Vision, **17** (2002) 109–129.

[119] C Vachier and F Meyer, Extinction Value, a New Measurement of Persistence, IEEE Workshop on Nonlinear Signal and Image Processing, Neos Marmaras, Greece, 20-22 June 1995.

[120] C Valens, A Really Friendly Guide to Wavelets, Technical Report, (1999).

[121] P M van den Berg and R E Kleinman, A Total Variation Enhanced Modified Gradient Algorithm for Profile Reconstruction, Inverse Problems, **11**(L5-L10) (1995).

[122] P F Velleman, Robust Nonlinear Data Smoothers: Definitions and Recommendations, Proceedings of the National Academy of Sciences, **74**(2) (1977) 434–436.

[123] P F Velleman, Definition and Comparison of Robust Nonlinear Data Smoothing Algorithms, Journal of the American Statistical Association, **75**(371) (1908) 609–615.

[124] L A Vese and S J Osher, Modeling Textures with Total Variation Minimization and Oscillating Patterns, Journal of Scientific Computing, **19**(13) (2003).

[125] L Vincent, Morphological Area Openings and Closings for Greyscale Images, Proceedings of NATO Shape in Picture Workshop, Driebergen, The Netherlands, 1992, Springer-Verlag, 197–208.

[126] L Vincent, Grayscale Area Openings and Closings, Thier Efficient Implementation and Applications, *Mathematical Morphology and Its Applications to Signal Processing*, J Serra and P Salembier (editors), UPC Publications, 1993, 22–27.

[127] L Vincent, Fast Greyscale Granulometry Algorithms, Proceedings of ISMM'94: International Symposium on Mathematical Morphology, Fontainebleau, France, 5-9 September 1994, Kluwer Academic Publishers, 265–272.

[128] C R Vogel and M E Oman, Fast, Robust Total Variation-Based Reconstruction of Noisy, Blurred Images, IEEE Transactions on Image Processing, **7**(6) (1998) 813–824.

[129] A H Wallace, *An Introduction to Algebraic Topology*, Pergamon Press, 1963.

[130] B Weiss, Fast Median and Bilateral Filtering, Association for Computing Machinery, Inc., (2006).

[131] E W Weisstein, "Apry's Constant." From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/AperysConstant.html

[132] E W Weisstein, "Euler-Mascheroni Constant." From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/Euler-MascheroniConstant.html

[133] E W Weisstein, "Extreme Value Distribution." From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/ExtremeValueDistribution.html.

[134] E W Weisstein, "Wavelet." From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/Wavelet.html.

[135] M Wild, Idempotent and Co-idempotent Stack Filters and Min-Max Operators, Theoretical Computer Science, **299** (2003) 603–631.

[136] M Wild, The many benefits of putting stack filters into disjunctive or conjunctive normal form, TDiscrete Applied Mathematics, **149** (2005) 174-191.

[137] L Yin, R Yang, M Gabbouj and Y Neuvo, Weighted Median Filters: A Tutorial, IEEE TRansactions on Circuits and Systems - II: Analog and Digital Signal Processing, **43**(3) (1996).

[138] L Yong and S Kassam, Generalized Median Filtering and Related Nonlinear Filtering Techniques, IEEE Transactions on Acoustics, Speech and Signal Processing, **33**(3) (1985) 672–683.

[139] Z Yong and G Taubin, Real-Time Median Filtering for Embedded Smart Cameras, Proceedings of ICVS 2006: IEEE International Conference on Computer Vision Systems, Manhattan, New York City, New York, USA, 5-7 January 2006.