IPAM GBM Tutorials
March 27 – April 6, 2001

Image Analysis and P.D.E.'s

Frédéric Guichard and Jean-Michel Morel

Image Analysis and P.D.E.'s

Frédéric Guichard and Jean-Michel Morel

Abstract.

It is well known that a conveniently rescaled iterated convolution of a linear positive kernel converges to a gaussian. As a consequence, all iterative linear smoothing methods of a signal or an image boil down to the application to the signal of the heat equation. This book explains how a similar analysis can be performed for image iterative smoothing by a wide class of nonlinear operators, the contrast invariant operators. These monotone operators have a property which makes them suitable for image analysis: they commute with contrast changes of the images. Among them, the median operator which computes a local "mean value" independent of constrast changes. We prove that all monotone and contrast invariant operators, are asymptotically equivalent (when they become more and more local) to a motion of the image by its curvature. The iteration of these filters is equivalent to the application to the image of a nonlinear heat equation. We give in parallel a classification of all image multiscale smoothing methods (the so called "scale space" methods in Computer Vision). It is shown that both approaches (classification of iterative filtering methods or of "scale spaces") yield the same, recently discovered, partial differential equations. Experiments are presented with both classical and new, contrast invariant and monotone numerical schemes.

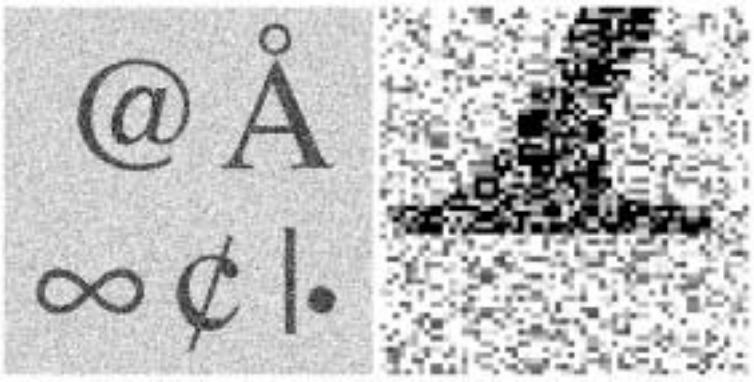

Figure 1: Zoom on Noise.

Caution:

This is a working and uncomplete version, subject to errors, of "Image Analysis and PDE's" - Frédéric Guichard and Jean-Michel Morel. No diffusion is allowed and all rights reserved.

Since the manuscript is subject to changes, the authors encourage to send an email to
`fguichard@poseidon.fr`
Emails will be sent back concerning progress of the manuscript when subtantial changes (resp. errors) will be made (resp. fixed).

Date : 22/02/2001

Contents

Chapter 1

Introduction

This book adresses a possible theory of image *low level* analysis. Image "low level" analysis aims at extracting reliable, local geometric informations from a digital image. Such informations are often called "features" and they are used in order to compare an image to other images. For instance, these features can be used for motion estimation, or to retrieve shapes, or to build the still hypothetic "high level" vision. The observed image is the result of a smoothing of the original photon flux and is therefore continuous. It is nontheless well admitted that the subjacent "real image", namely the focused photon flux, is either a measure or, for more optimistic authors, a function which presents strong discontinuities. Rudin and de Giorgi proposed independently in 1984 the space BV of functions with bounded variation as the right function space for "real" images. More recently (1999), however, Gousseau and Alvarez used a statistical device on digital images to estimate how their real subjacent images oscillate. They deduced, by geometric measure arguments, that the "real" physical images have in fact unbounded variation. We may therefore accept the idea that the subjacent high resolution image behaves in a strongly oscillatory way. Although the digital images present a averaging of this oscillatory phenomenon, common sense tells us that they must have anyway strong discontinuities at transitions between different observed objects, i.e. on the apparent contours of physical objects. The BV space looked at first well adapted to that aim because it contains functions having step discontinuities.
One of the goals of image analysis has ever been to find such discontinuities in an image. This search is called "edge detection" because early vision research played with images of cubes. Along the edges of the cubes, the light intensity behaved, in a first approximation, as a step function. Unfortunately, the early research in vision led to the sad discovery that one could find edges "everywhere" in a digital image (Marr, Vision), due to the oscillations remaining in the digital image after the digitization problem. As a consequence, the image analysis process was conceived as a smoothing process, permitting to decluster the true "edges" from the inherent noise. As in Distribution theory, a smoothing was necessary before computing any derivative. This is why the heat equation was proposed and a new doctrine proposed : the "scale space". Scale space means that, instead of talking of features of an image at a given location, we talk of them at a given location and at a given scale. The scale quantifies the amount of smoothing performed on the image before computing the feature. We shall therefore see in experiments "edges at scale 4" and "edges at scale 7" as different outcomes of an edge detector.

Which kind of smoothing should be performed ? Three terms associated with image analysis opera-

tors arise here, to which we will give a more and more precise meaning.
The first one, "locality", is related to the occlusion problem : Most optical images are made of a superposition of different objects partly hiding each other. It is plain that we must avoid mixing them in the analysis, as would do e.g. a wide convolution. Thus, the analysis must be made as local as possible. As we shall see, the heat equation is the worst candidate to the task, since it makes a wide-range mélange of grey levels.
The second key word is "iteration". Indeed, we shall see that it is generally better, from the locality viewpoint, to iterate a very local smoothing operator than to apply it directly at a large scale. This is precisely not true for the heat equation ! Iterating the convolution of small gaussian kernels is stricly the same as convolving directly the image with a big gaussian. Now, iteration of very local filters will bring a significant improvement for some of the most relevant nonlinear filters which we shall consider, namely the median filter and the affine erosion-dilations. At this point, it must be immediately announced that the combination of smoothing, locality and iteration implies that we are talking about parabolic partial differential equations. This announcement is heuristic and the object of the book is precisely to formalize the necessity and the role of several P.D.E.'s in image analysis.
Our last key word is "invariance". The invariance requirements play a central role in image analysis because the objects to be recognized have to be recognized under varying conditions of illumination (contrast invariance) and from different points of view (projective invariance). Contrast invariance is one of the key requirements of a famous image analysis theory, the Mathematical Morphology (Matheron, Serra). This theory proposed a list of contrast invariant image analysis operators (dilations, erosions, median filters, openings, closings,...) We shall involve this theory, as we shall attempt to localize as much as possible the "morphomath" operators to extract their behaviour at small scale, and then iterate them. As an outcome, we shall prove that several geometric partial differential equations, namely the curvature motions, can be considered as the common asymptotic denominators to many "morphomath" operators. These P.D.E.'s permit therefore to fuse the Scale Space doctrine and Mathematical Morphology. In particular, affine invariant morphomath operators, which looked unpractical, turn out to yield in their local iterated version a very affordable P.D.E., the so called "affine morphological scale space" (A.M.S.S.).
In the next section, we shall make a survey of most P.D.E.'s which have been proposed for image analysis (Section 1) and thereafter give a detailed mathematical account of how this book is organized.

We would like to end this short foreword with a light warning to the reader. We do not claim that what will be developped here is a necessary future for image analysis. What the mathematical analysis can provide is more hypothetical, as noticed Von Neumann and Morgenstern in their book on game theory. We would say, by imitation : "If Image Analysis requires a smoothing theory, then here is how it should be done, and here is the proof that there is no other way to do it." This statement does not exclude the possibility of other theories, based on different principles, or even the impossibility of making any theory. We have tried to prove every single mathematical statements, assuming only a two or three years mathematical training. Thus, most of the P.D.E's adressed, and all of the relevant ones, will be proved existence, uniqueness and given invariant monotone approximation schemes. This has been technically possible by combining tools from the recent and remarkbly simple theory of viscosity solutions on the one side, and of the Matheron formalism for monotone set and function operators. Thus, the really necessary mathematical knowledge amounts to elementary differential calculus, linear algebra, and some notions of integration in

the chapter on the heat equation. We nether put mathematical statements *pour l'amour de l'art* ; all of them are directed at proving the rightfulness of a model, of its properties and of the associated numerical schemes. Many experiments are spread out in the text, with detailed comments. They can provide an independent, parallel, reading to the mainstream of the text. There are no direct industrial applications treated. Now, many of the image representation tools commented in the text are being used in the public software MegaWave (http://www.cmla.ens-cachan.fr) developped jointly by several university research groups in France and Spain. They are also extensively used in the industrial video and image processing software Investigator, sold by Cognitech, Inc., Pasadena, California.

1.1 The when, why and how of P.D.E. models in image processing

Image processing is divided in three parts, corresponding to as many different goals. The first one derives from the discrete nature of images and the search of their minimal representation in terms of digital memory. This discipline is called *image compression* (see Figure 1.1). The second goal is the *restoration* of

Figure 1.1: Compression. From left to right : An original image and its more and more compressed versions : compression factor 7, 10 and 25 respectively. One of the first goals of image processing is the definition of algorithms permitting high compression factors without visible alteration. Compression may, however alter the image.

a better version of an image, given a generation model with noise and blur, or other perturbations. This is illustrated in Figure 1.2. The image on the right is apparently destroyed: more than 75% of its pixels

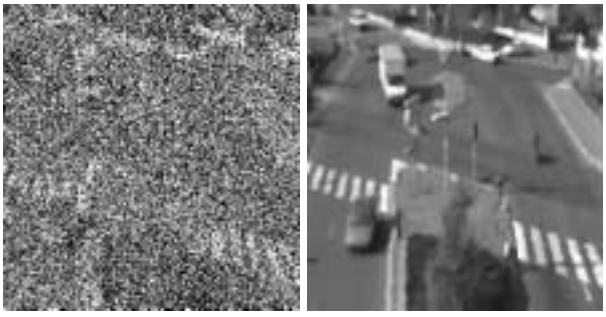

Figure 1.2: Denoising. A second goal of image processing : the restoration. Left: original noisy image (simulated salt and pepper noise up to 75%), right: denoised version

have been put at a random value. We can nontheless restore it significantly: here is, on the right, such a restored version. The third goal is analysis , which means in Greek "breaking into parts". Look at the

level curve of Figure 1.3, extracted from a hand image: it is full with a mix of details and noise. What if we ask for a sketchy version, where, however, all essential details are kept ? The curve on the right is such a sketchy version, where most of the spurious details have disappeared, but the main structures are maintained. This is what we shall call *image analysis* . The aim is not denoising or compression : it is to construct an invariant code putting in evidence the main parts (here, for instance, the fingers) and permitting a fast recognition in a large database of shapes.

Figure 1.3: Analysis of a shape. Left : Original scanned shape, then some simplified versions : the aim here is not restoration, but analysis, that is, to define more and more sketchy versions of the shape. Those sketchy versions may permit a very short and invariant encoding of the shape. Notice how the number of inflexion points of the shape has decreased in the simplification process (Chapter 22).

The heat equation arises naturally in the image generation process. Indeed, according to Shannon's theory, an image can be correctly represented by a discrete set of values, the "samples", only if it has been previously smoothed. This is illustrated in Figure 1.4 : Let us call the original baby image "Victor". If we attempt to reduce the size of Victor by a mere subsampling, that is by taking a point of each sixteen, we obtain a new and smaller image, in which the subsampling has created new and unstable patterns : see how new stripes have been created, with a frequency an direction which has nothing to do with the original ! If, instead of being steady, the camera moved, those newly created patterns would move and flicker in a totally uncontrolled way. This kind of moving pattern appears often in recent commercial DVD's. They have simply been subssampled against the Shannon rule. Let us now comment briefly how the subsampling should be done. According to Shannon's theory, a previous smoothing must to be done before the subsampling. We start with u_0, the original image. Then a blur kernel k is applied, i.e. we convolve u_0 with k to obtain a new image $k * u_0$. A subsequent subsampling is thereafter possible, where the distance between samples is related to the band-width of the blur kernel by the Nyquist rule. Stability of the image representation is maintained.

This simple remark, that smoothing is a necessary part of image formation, leads us to our first PDE's. Gabor remarked in 1960 that the difference between the original and the blurred image is roughly proportional to its laplacian. In order to formalize this remark, we have to notice that k is spatially concentrated, and that we may introduce a scale parameter for k, namely $k_h(\mathbf{x}) = \frac{1}{h} k(\frac{\mathbf{x}}{h^{\frac{1}{2}}})$. Then

$$\frac{u_0 * k_h(\mathbf{x}) - u_0(\mathbf{x})}{h} \to \Delta u_0(\mathbf{x}),$$

so that when h gets smaller, the blur process looks more and more like the heat equation

$$\frac{\partial u}{\partial t} = \Delta u, \quad u(0) = u_0.$$

Conversely, Gabor deduced that we can, in some extent, deblur an image by reversing time in the heat

Figure 1.4: Shannon theory and subsampling. From left to right: original image, smoothed image, subsampling of the original image and subsampling of the smoothed image. In the subsampling, one point of each 4 is taken in the horizontal and vertical directions. In order to make the reduced image still visible, we have zoomed back the subsampled versions by a zoom factor 4. We clearly see that subsampling an image without previous smoothing creates aliasing : high frequencies are projected onto lower frequencies and therefore generate new patterns. Shannon theory tells us how to remove those potentially parasite high frequencies before subsampling. This results in the necessity of smoothing the image before subsampling.

Figure 1.5: Heat equation and blur. Left : original image, right : the heat equation has been applied to some scale and the resulting image is blurred (Chapter 2).

equation :

$$\frac{\partial u}{\partial t} = -\Delta u, u(0) = u_{observed}.$$

Numerically, this amounts to substracting its laplacian from the observed image :

$$u_{restored} = u_{observed} - h\Delta u_{observed}.$$

This operation can be repeated several times with some small values of h, until it... blows up. Indeed, the reverse heat equation is extremely ill-posed. All the same, this Gabor method is efficient and can be applied with some success to most digital images obtained from an optical device. Let us examine what happens with Victor (Figure 1.6). We see that the method yields some improvement at the beginning and then blows up. We can also simulate a blur on Victor and try to go back : again, the process blows up but yields a significant improvement at some scale.

We therefore see two directions. One is to improve, to stabilize, the reverse heat equation. We shall see that this is doable by nonlinear models. The second direction is to go on with the heat equation : we can numerically simulate a further blurring of the image. Why should we do so ? Because, first, this leads to the wavelet theory and its applications to optimal multiscale sampling and compression. Second,

Figure 1.6: Gabor's deblurring. Gabor proposed in 1960 to deblur an image by substracting its laplacian : this means inverting the heat equation ! Left : original image, middle : three iterations of Gabor's algorithm, right : ten iterations. As is well known, and can be observed in the right image, the inverse heat equation blows up. A few iterations can, as we see in the middle, nicely enhance the image.

Figure 1.7: Gabor's deblurring again. Same deblurring experiment as in Figure 1.6, but applied on a much more blurred image

iterated linear and nonlinear smoothing (that is, nonlinear PDE's) will be relevant to our main goal : image analysis.

We can indeed improve the time-reverse heat equation. The first example, due to Rudin and Osher in 87 and 92 proposes an pseudoinverse, where the propagation term —Du— is tuned by the sign of the laplacian.

$$\frac{\partial u}{\partial t} = -sign\Delta u|Du|.$$

The equation is called "shock filter". As we shall see, this equation propagates, with a constant speed, the level lines of the image in the same direction as the reverse heat equation would do. It therefore enhances the image. The equation is more or less equivalent to a good old nonlinear filter due to Kramer in the seventies. Kramer's filter can be interpreted as a partial differential equation, by the same kind of heuristic arguments which Gabor developed to derive the heat equation. This equation is

$$\frac{\partial u}{\partial t} = -signD^2u(Du, Du)|Du|.$$

Thus, the laplacian is replaced by a directional second derivative of the image, $D^2u(Du, Du)$. We shall later on interpret this differential operator as an "edge detector". Kramer's equation yields a slightly better version of shock filter as is illustrated in Figure 1.8. Both deblurring equations work... to some extent. They experimentally do not blow up and attain steady states ! The third deblurring method we

Figure 1.8: Image deblurring by shock filters and by a variational method. From left to right : blurred image, Rudin-Osher shock filter which is a pseudoinverse of the heat equation attaining a steady state, Kramer's improved shock filter, also attaining a steady state and the Rudin, Osher, Fatemi restoration method, obtained by deblurring with a controlled image total variation. This last method is very efficient when the noise and blur models are known. It is currently being used by the French Space Agency (CNES) for satellite image restoration (Chapter 16).

can mention here is, to our knowledge, the best version. It poses the deblurring problem as an inverse problem. Given the observed image u_0, we try to find a restored version u such that $k * u$ is as close as possible to u_0 and the oscillation of u is nontheless bounded :

$$u_{restored} = Argmin(\int |Du| + \lambda(k * u - u_0)^2.$$

The parameter λ tunes the oscillation we allow for the restored version. If λ is large, the restored version will satisfy accurately the equation $k * u = u_0$, but may be very oscillatory. If instead λ is small, we get a smooth but unaccurate solution. This parameter can be computed in principle as a Lagrange multiplier. The obtained restoration can be remarkable. We display the best result we can obtain with the blurred Victor in Figure 1.8-right. This total variation restoration method also has fast wavelet packets versions. It recently won a benchmark in satellite image deblurring organised by the French Space Agency (CNES).

The original remark of Gabor, about image generation being related to the Laplacian of the image, leads to the wavelet theory as well. Here is how it works : if we convolve the image with some smoothing kernel and thereafter make the difference, we obtain a new image, actually a laplacian, which turns out to be faded with respect to the original. In Figure 1.9, the last image on the right shows in black the values of this laplacian image of Victor which differ significantly from zero : In most natural images, as here, this representation is sparse and adapted to compression. This is why one of the first wavelet representations, due to Burt and Adelson in 83, was called "Laplacian pyramid". It boils down to the iteration of a convolution followed by subsampling. We only keep the differences between images smoothed at different scales, i.e. their laplacians. The objective is a compressed representation, but to the price of a loss of invariance due to the multiscale subsampling.

In image analysis, the heat equation has had a very different use: Marr, Hildreth, Canny, Witkin, Koenderink proposed in the eighties to analyse an image by applying the heat equation. As Rudin and Florack noticed, this is related to distribution theory. Indeed, details of the images, like boundaries, corners and other singularities cannot be computed without some previous smoothing because they are derivatives of a nonsmooth function. And this smoothing has to be multiscale because the image is multiscale ! The heat equation is easily proved to be the only good candidate to the task if image analysis has to be linear. What derivatives should be computed in an image ? The early research in computer vision

Figure 1.9: The "laplacian pyramid" of Burt and Adelson. From left to right : original image, image blurred by gaussian convolution, then difference between the original image and the blurred version, which simulates the laplacian of the original image. In black in the last image, points where this laplacian image is large. This experiment simulates the first step of the laplacian pyramid. The laplacian image is, for most digital images, a sparse representation, therefore well adapted to compression.

proposed "edge detection" as a main tool : it is assumed that the apparent contours of the objects and also the boundaries of the facets of objects, result in step discontinuities in the image, while, inside those boundaries, the image oscillates mildly. The apparent contour points, or "edges points" will be computed as points where the gradient is is some sense largest. Two ways to do so : Hildreth and Marr proposed the points where Δu crosses zero. A significant improvement was done by Canny, who proposed to compute the points where Du is maximal on the gradient lines. Such points satisfy $D^2u(Du, Du) = 0$. Figure 1.10 displays what happens when we smooth the image with the heat equation and compute the points where $D^2u(Du, Du) = 0$ and $|Du|$ is large enough. At first, everything in the image is boundary : the image, being a very oscillatory function, has inflexion points everywhere ! After some evolution of the heat equation, we can see what happens : we are able to extract some structure.

Figure 1.10: Heat equation and Canny's edge detector. Boundaries, or "edges" of the image can be defined as points where the gradient attains a maximal and large value along the gradient lines. This amounts to say that edge points are points where $D^2u(Du, Du)$ crosses zero and $|Du|$ is large. Canny's edge detector computes those points. On the left, the original image, followed by the edge points found. They make a very dense set, because of the oscillatory character of the image. Next, the image blurred by the Gauss kernel (heat equation) and the Canny edges found. The heat equation has removed the "irrelevant" edges (Chapter 3)

If the heat equation is, under sound invariance requirements, the only good linear smoother, there are instead many nonlinear ways to smooth an image. The first one was proposed by Perona and Malik in 87. The idea is roughly to smooth out what has to be smoothed, the irrelevant, homogeneous, regions and to enhance instead the boundaries. Thus, the diffusion should look like the heat equation when $|Du|$ is small and an inverse heat equation should instead be applied when $|Du|$ is large. Here is the equation in

divergence form.

$$\frac{\partial u}{\partial t} = div(g(|Du|^2)),$$

where $g(s) = \frac{1}{1+\lambda s^2}$ decreases when s increases. It is easily checked that we have a diffusion equation when $|Du| \leq \lambda$ and an inverse diffusion equation when $|Du| > \lambda$. In order to do so, we rewrite the equation in the following way. We consider the second derivative of u in the direction of Du,

$$u_{\eta\eta} = D^2u(\frac{Du}{|Du|}, \frac{Du}{|Du|})$$

and the second derivative in the orthogonal direction,

$$u_{\xi\xi} = D^2u(\frac{Du\perp}{|Du|}, \frac{Du^\perp}{|Du|}),$$

where $Du = (u_x, u_y)$ and $Du^\perp = (-u_y, u_x)$. The laplacian can be rewritten in the intrinsic coordinates (ξ, η) as $\Delta u = u_{\xi\xi} + u_{\eta\eta}$. The Perona-Malik equation rewrites

$$\frac{\partial u}{\partial t} = \frac{u_{\xi\xi}}{1+\lambda^2|Du|^2} + \frac{(1-\lambda^2|Du|^2)u_{\eta\eta}}{(1+\lambda^2|Du|^2)^2}.$$

So the first term always appears as a one-dimensional heat equation in the direction orthogonal to the

Figure 1.11: Perona-Malik equation and edge detection : same experiment as in Figure **??**, but the heat equation has been replaced by the Perona-Malik equation. Notice that the edge map looks slightly better localized as with the heat equation.

gradient, tuned by the size of the gradient though. The second term can be a directional heat equation, or reverse heat equation in the direction of the gradient. So we indeed mix in this model the heat equation and the reverse heat equation ! We compare in Figure 1.11 the Perona-Malik with the classical heat equation in terms of accuracy on the boundaries obtained by Canny's edge detector : at a comparable scale of smoothing, we clearly gain some accuracy in the boundaries and get rid of more "spurious" boundaries. The representation is both more sparse and more accurate.

Now, this ambitious model attempts to put in a single operator two very different goals which we already mentionned, namely restoration and analysis. This has a cost : the model contains a "constrast threshold" which can only be fixed manually. Mathematical existence and uniqueness are not guaranteed, despite some attempts by Kichenassamy and Weickert. Let us summarize the involved parameters : we need to fix both λ and the smoothing scale(s) t and the threshold on the gradient in Canny's detector as

well. We obviously must take the same gradient threshold λ in Canny's detector and in the Perona-Malik equation. All the same, we have a two parameters game : how will this be dealt with in automatic image analysis ? This question seems to have no general answer for the time being. An interesting attempt based on statistical arguments is made, however, in Black, Sapiro.

If any nonlinear diffusion can be an image analysis model, why not trying them all ? This is exactly what has happened in the past ten years. We can claim that almost all possible nonlinear parabolic equations have been proposed. The logic in this proliferation of models is this : each attempt fixes one intrinsic diffusion direction and tunes the diffusion by the size of the gradient or by the size of a nonlinear estimate of the gradient. Sometimes, the proposed models are even systems of PDE's, but in order not to blow up this preface, we shall focus on the simplest proposed examples. We can start with Rudin-Osher-Fatemi's model, which consists, for the smoothing term, of minimizing the total variation of u. The gradient descent for $\int |Du|$ writes

$$\frac{\partial u}{\partial t} = div(\frac{Du}{|Du|}) = \frac{1}{|Du|} u_{\xi\xi}.$$

Written in that way, the method appears as diffusion in the direction orthogonal to the gradient, tuned by the magnitude of the gradient. Caselles, ? ? and ? ? proved that this equation is indeed well posed in the space of bounded variation. A variant was proposed independently by Alvarez and al.,

$$\frac{\partial u}{\partial t} = \frac{|Du|}{|k * Du|} div(\frac{Du}{|Du|}) = \frac{1}{|k * Du|} u_{\xi\xi},$$

where the tuning of the gradient is nonlocal. Kimia, Tannenbaum and Zucker proposed, endowed in a more general shape analysis framework, the simplest equation of the list,

$$\frac{\partial u}{\partial t} = |Du| div(\frac{Du}{|Du|}) = D^2u(\frac{Du}{|Du|}^{\perp}, \frac{Du^{\perp}}{|Du|}) = u_{\xi\xi}.$$

This equation had been proposed some time before in another context by Sethian as a tool for front propagation algorithms. This equation, which we call in continuation "curvature equation", is a "pure" diffusion in the direction orthogonal to the gradient. The Weickert equation is a variant of the curvature equation, with nonlocal estimate of the direction orthogonal to the gradient : the diffusion direction $d = SEigen(k * (Du \otimes Du))$ is computed as the eigenvector of the least eigenvalue of $k * (Du \otimes Du)$: if the convolution kernel is removed, this eigenvector simply is $Du^{\perp}$. The three mentionned models can be interpreted as diffusions in a direction orthogonal to (an estimate of) the gradient, tuned by the magnitude of the gradient (Figure 1.14). Other diffusions have been considered as well : For interpolation goals, Caselles et al. proposed a diffusion which may be interpreted as the strongest possible image smoothing,

$$\frac{\partial u}{\partial t} = D^2u(Du, Du).$$

This equation is not used as the other ones as a preprocessing of the image, but a way to interpolate between the level lines an image with sparse level lines (Figure 1.13). Zhong and Carmona proposed a diffusion in the direction $d = SEigen(D^2u)$ of the eigenvector with least eigenvalue of D^2u (Figure 1.12). Sochen, Kimmel and Malladi propose instead a nondegenerate diffusion, associated with a minimal

Figure 1.12: A proliferation of diffusion models. From left to right: Original image, Perona and Malik equation 1987, Zhong, Carmona 1998 (diffusion along the least eigenvector of D^2u) and Sochen, Kimmel and Malladi, 1998 (minimization of the image graph area.

Figure 1.13: A proliferation of diffusion models (III) Here, the diffusion is made in the direction of the gradient and the model is applied for image interpolation when level lines are sparse. From left to right : original image, quantized image (only 10 levels are kept - 3.32 bits/pixel) and reinterpolated image by the Caselles and Sbert (1998)algorithm. They apply a diffusion on the quantized image, with values on the remaining level lines as boundary conditions.

Figure 1.14: A proliferation of diffusion models (II). From left to right: Osher, Sethian 1988: curvature equation, Rudin, Osher and Fatemi 1992: minimization of the image total variation, Alvarez, Lions et al. : nonlocal variant of the preceding, 1992, Weickert 1994 : nonlocal variant of the curvature equation. All of these models only diffuse in the direction orthogonal to the gradient, with a more or less local estimate of this direction.

surface variational formulation : their idea was to make a gradient descent for the area of the graph of u, $\int \sqrt{1+|Du|^2}$, which leads to the diffusion equation (Figure 1.12).

$$\frac{\partial u}{\partial t} = div(\frac{Du}{\sqrt{1+|Du|^2}}).$$

Among the mentionned models, only the curvature motion was explicitly proposed by Kimia, Tannenbaum and Zucker as shape analysis tool. We shall now explain why.
In order to do so, we have to give a definition of image analysis. There might be as many ways to define

this discipline as they are applicational goals involving digital images. Now, the range of applications is as wide as the human activity, since most of the scientific and technical human activity, including even sound analysis (visual sonagrams), involves the perceptual analysis of images. Fortunately, we have at hand a mathematical shortcut to avoid an endless list of partial and specific requirements : This shortcut, well known in Mechanics, consists of stating invariance requirements. Invariance requirements will be a short list and they will, as we shall see, give a possible classification of models and point out the ones which are the most adequate for all purposes image analysis tools. The first invariance requirement is the Wertheimer principle according to which visual perception (and therefore, may we add, image analysis) should be independent of the image contrast. We state this in the following way :

Contrast invariant classes
u and v are said to be (perceptually) equivalent if there is a continuous increasing function g such that $v = g(u)$.

Contrast invariance requirement : An image analysis operator T must act directly on the equivalence class. As a consequence, we may ask that $T(g(u)) = g(Tu)$, i.e. a commutation of the image analysis operator with contrast changes.
newline

The contrast invariance requirement rules out the heat equation and all models stated before, except the curvature motion. Contrast invariance led Matheron in seventy-five to reduce image analysis to a set analysis, namely the analysis of level sets. We call upper level set with level λ of an image u the set

$$\mathcal{X}_\lambda u = \{\mathbf{x}, u(\mathbf{x}) \geq \lambda\}.$$

We can define in exactly the same way the lower level sets, by changing "$\geq$" into "$\leq$". The main point to retain here is the global invariance of level sets under contrast changes, namely, if g is a continuous increasing contrast change, then,

$$\mathcal{X}_{g(\lambda)} g(u) = \mathcal{X}_\lambda u.$$

According to Mathematical Morphology, this image analysis doctrine founded by Matheron, all of the image shape information is therefore contained in the level sets : it can be proved that an image can be reconstructed, up to a contrast change, from its set of level sets (Figure 1.15 : an image and some of its level sets).

The contrast invariance requirement yields powerful and simple denoising operators as the so called "Extrema killer" defined by Vincent and Serra in 1993. This image operator simply removes all connected components of upper and lower level sets with area smaller than some fixed scale. This is not a PDE, actually it's much simpler ! Now, its effect is amazingly good for impulse noise i.e. local destructions of the image, spots. In Figure 1.16, we see a image degraded up to 75%. Below, its restoration by the extrema killer. Left, result of the same operator applied to the original.

Caselles and Coll localized farther in 1996 the contrast invariance requirement in image analysis. They proposed as the main object of analysis the level lines of the images, that is, the boundaries of level sets.

Figure 1.15: An image and one of its level sets. Right : level set 140 of the left image. This experiment illustrates Matheron's thesis that the main shape information is contained in the level sets of the image. Level sets are contrast invariant (Chapter 4).

Figure 1.16: The "extrema killer" filter : all connected components of the upper or lower level sets with small area are removed from the image. From left to right : original image, extrema killer applied with area 80 pixels, then 75% salt and pepper noise added to the original image and the same filter applied (Chapter 7).

This proposition makes sense for a digital image, which is assumed to be a sampling of a very smooth function as the result of the optical smoothing. We can therefore define the level lines if, e.g., the interpolated image is C^1 as is guaranteed by the canonical Shannon interpolation. There may be other interpolation methods, and even interpolations into a discontinuous functions : this is the case if, e.g., we consider the digital image as constant on each pixel. We must then for each interpolation method make clear how the level lines are computed and what their structure is. Two properties are desirable : that the level curves indeed are curves in some affordable sense (Jordan rectifiable curves) and that they are nested, i.e. never cross, so that they make an inclusion tree. A study of Kronrod (1950) shows that if the function u is continuous, then the isolevels sets $\{\mathbf{x}, u(\mathbf{x}) = \lambda\}$ are nested : they build a tree ordered by inclusion. Now, these isolevel sets need not be really curves. Monasse (2000) generalized recently the preceding result to lower semicontinuous or upper semicontinous functions. His result implies that the simplest, piecewise constant, interpolation of an image yields a nested set of Jordan curves bounding the pixels. Thus, we have two good ways to associate with the digital image a set of nested Jordan curves. We call this set "topographic map". [1] We display in Figure 1.17 the level lines of a digital image at some fixed level. By the introduction of the topographic map, the search for image smoothing, which we had already reduced to set smoothing, is further reduced to curve smoothing, provided of course this smoothing preserves curve

[1] This point of view also is coherent with the "BV assumption" which we mentionned at the beginning of the introduction, according to which the right function space for images should be the space BV of functions with bounded variation. By coarea formula, we can then describe the image by a bunch of Jordan level curves (see Ambrosio et al.) Now, it is in general false for BV functions that boundaries of lower and upper level sets make a nested set of curves : these curves may cross.

inclusion.

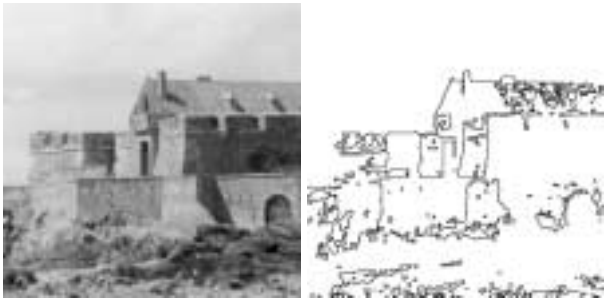

Figure 1.17: Level lines of an image. Level lines, defined as the boundaries of level sets, can be defined to be a nested set of Jordan curves. They give a contrast invariant representation of the image. Right : level lines with level 183 of the left image (Chapter 5).

Chen, Giga and Goto and Alvarez, et al. proved that, under the usual invariance requirements for image processing, including the contrast invariance, all image multiscale analyses should have the form of a curvature motion, namely

$$\frac{\partial u}{\partial t} = F(curv(u), t)|Du|,$$

where F is increasing with respect to its first argument. This equation can be interpreted as this : we consider a point $\mathbf{x}$ on a given level curve of $u(t)$, at time t. We call $n(\mathbf{x})$ the normal vector to the level curve and $curv(\mathbf{x})$ its curvature. Then the preceding equation turns out to be associated with the curve motion equation

$$\frac{\partial \mathbf{x}}{\partial t} = F(curv(\mathbf{x}))n(\mathbf{x}),$$

which describes how the point $\mathbf{x}$ moves in the direction of the normal. Not much more can be said at this level of generality on F. Now, two particular cases happen to play a prominent role. First, the case $F(curv(u), t) = curv(u)$, the so called curvature equation which we already met, and second the case $F(curv(u), t) = curv(u)^{\frac{1}{3}}$.

This particular form for the curvature dependence, the power one third, permits to get a very relevant additional invariance, the affine invariance. We would like to have a full projective invariance, but a theorem proved by Alvarez et al. shows that this is impossible. The best we can have is invariance with respect to the so called chinese perspective, which preserves parallelism. Most of the mentionned equations, particularly when F is a power of curvature, have a viscosity solution in the sense of Crandall and Lions, as shown by recent works of Ishii and Souganidis.

As we already mentionned, contrast invariant processing boils down to level set, and finally level curve processing. The above mentionned equations indeed are equivalent to curve evolution models, provided strong existence results are at hand. This is the case for the most important cases, namely the power 1, the so called "curve shortening" and the power 1/3, known as "affine shortening". Grayson proved existence, uniqueness and analycity for the first equation,

$$\frac{\partial \mathbf{x}}{\partial t} = curv(\mathbf{x})n(\mathbf{x})$$

and Angenent, Sapiro and Tannenbaum for the affine shortening

$$\frac{\partial \mathbf{x}}{\partial t} = curv(\mathbf{x})^{\frac{1}{3}} n(\mathbf{x}).$$

Those results are very relevant to image analysis as they ensure that the diffusion process indeed reduces

Figure 1.18: The Affine and Morphological Scale Space (AMSS model). From left to right : original image, level lines of the images (16 levels only), smoothed image by using the affine and morphological scale space, and its level lines (Chapter 6).

the curve to a more and more sketchy version. We check the affine invariance in Figure 1.19 . The numerical test we shall make here is as follows : we apply a very fast and fully affine invariant numerical scheme designed by Lionel Moisan. In the middle, the initial shape is an affine transform of the first one ; the shape on the right will be an inverse affine transform of the middle shape. If everything is correct, we can expect that, after processing, the shape on the right will be identical to the shape on the left. We make the experiment with both the the curve shortening and the affine shortening. So, it works !

Evans-Spruck and Chen-Giga-Goto proved in 1991 that a continuous function moves by curvature mo-

Figure 1.19: Experimental check of the affine invariance of the affine shortening (AMSS). We display on the left image three shapes. The second one is an affine transform AS of the first shape S. The third one is obtained from the second by the inverse affine transform. It therefore initially is $A^{-1}AS = S$. On the right image : result after application of AMSS to the two first shapes : are viewed $S(t)$,, $(AS)(t)$ and $A^{-1}((AS)(t))$. If the numerical scheme is affine invariant, this third shape must coincide with $S(t)$, which is indeed the case. Middle : the same procedure applied with the curvature equation, which proves not to be affine invariant, as expected (Chapter 24).

tion if and only almost all of its level curves move by curve shortening. This yields, in that case, a mathematical justification of the now classical Osher-Sethian numerical method for moving fronts by moving a distance function to the front. The same result is true for the affine invariant curve evolution. The Osher-Sethian point of view is just converse to the point of view adopted here : they associate with some

curve C or surface its signed distance function u , so that the curve or surface is handled indirectly as the zero isolevel set of u. Then u is evolved by, say, the curvature motion with a classical numerical difference scheme. In that way, the curve evolution is dealt with efficiently and accurately. From our point view, the image can be viewed as a distance function to all and each of its level sets, since we are interested in all of them.

We show in Figure 1.18 an application of this numerical method, with both curvature and affine invariant

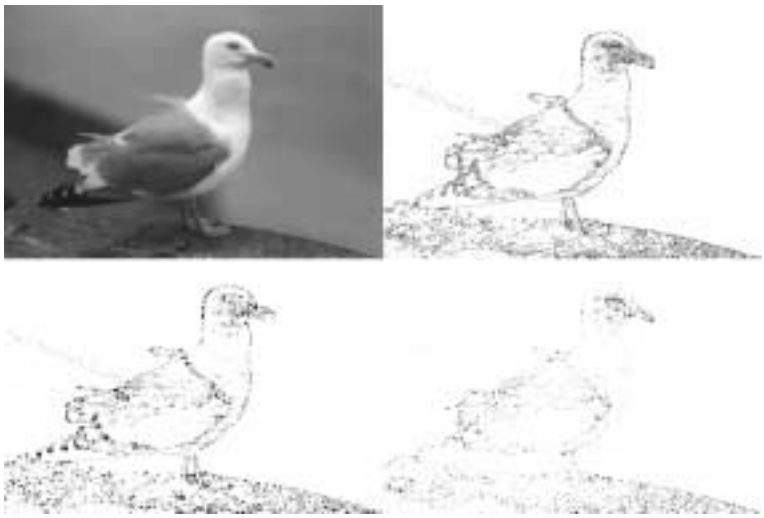

Figure 1.20: Computation of the curvature of the original sea bird image after it has been smoothed by curvature motion at calibrated scale 1. The first image displays the smoothed version of the sea bird at a small scale. The second image displays the absolute value of the curvature, with the convention that the darkest points have the largest curvature. We have displayed the curvature only at points where the gradient of the image was larger than 6. (The image grey levels range from 0 to 255). In continuation, we display on separate images the positive part of the curvature and the negative part. The curvature motion can be used as a nonlinear means to compute a "multiscale" curvature of the original image. Compare with Figure 1.21, where the calibrated scale of smoothing is 4. (A calibrated scale t means that at this scale a disk with radius t disappears)

curvature motions. In order to gain visibility, we do not display all level curves, but only for about eighteen levels. Notice that the aim is not here subsampling ; we keep the same resolution. It is not either restoration : the processed image is clearly worse than the original. The aim is invariant simplification leading to shape recognition.

Before proceeding to shape recognition, let us mention that a well adapted variant of curvature equation can be used for shape detection. It's a by now famous method of contour detection in an image, initially proposed by Kass, Witkin and Terzopoulos. This method was very unstable and the winning method turns out to be a variant of curvature motion proposed by Caselles, Catt, Coll, Dibos and improved simultaneously by Caselles, Kimmel, Sapiro, and Malladi, Sethian. Here is how it works. The user draws roughly what are the contours he wants in the image and the algorithm then finds the best possible contour in terms of some variational criterion. This turns out to be very useful in medical imaging. The motion of the contour is a tuned curvature motion which tends to minimize the energy E which we will now explain. Given an original image u_0 containing some circular contours which we wish to approximate, we start with an "edge map"

$$g(\mathbf{x}) = \frac{1}{1 + |Du_0(\mathbf{x})|^2},$$

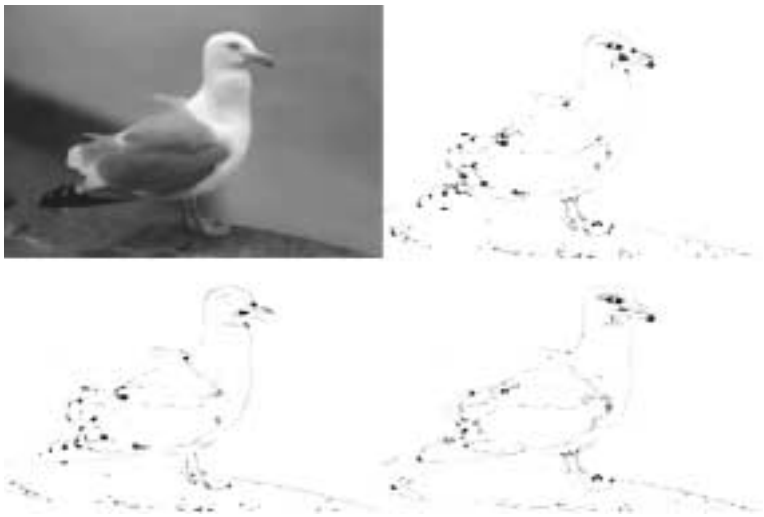

Figure 1.21: Computation of the curvature of the original sea bird image after it has been smoothed by curvature motion. The first image displays the smoothed version of the sea bird at calibrated scale 4. The second image displays the absolute value of the curvature, with the convention that the darkest points have the largest curvature. We have displayed the curvature only at points where the gradient of the image was larger than 6. (The image grey levels range from 0 to 255). In continuation, we display on separate images the positive part of the curvature and the negative part. The curvature motion can be used as a nonlinear means to compute a "multiscale" curvature of the original image. Compare with Figure 1.20, where the calibrated scale of smoothing is 1.

that is, a function which vanished on the edges of the image. The user then points out the contour he is interested in, by drawing a polygon γ_0 surrounding roughly the desired contour. The "geodesic snake" algorithm then builds a distance function v_0 to this initial contour, so that γ_0 is the zero level set of v_0. The energy to be minimized is

$$E(\gamma) = \int_\gamma g(\mathbf{x}(s))ds,$$

where g is the edge map associated with the original image u_0 and s denotes the length parameter on γ. The motion of the "analysing image" v is governed by

$$\frac{\partial v}{\partial t} = g|Dv|curv(v) - Dv.Dg.$$

We display an example in Figure 1.22.

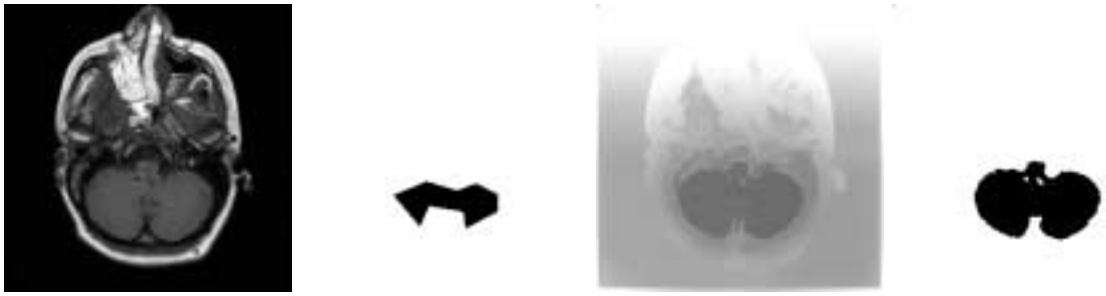

Figure 1.22: Active contour, or "snake" From left to right : original image, initial contour, evolved distance function, final contour (Chapter 19).

The main obvious application of invariant PDE's seems to be shape retrieval in large databases. There are thousands of different definitions of shapes, and of shape recognition algorithms. Now, the real bottleneck has ever been extraction of the relevant shapes. The discussion above points to a brute force strategy : all contrast invariant local elements, are are the level lines of the image, are candidates to be "shape elements". Of course, this name of shape element suggests the contours of some object, but there is no way to give a simple geometric definition of objects. We must give up the hope of jumping from the geometry to the common sense world. We may instead simply ask the question : given two images, can we retrieve all level lines similar in both ? This would give a factual, a posteriori definition of shapes : they would be defined as pieces of level lines common to two different images, no matter what their relationships to real physical objects are. Of course, this brute force strategy would be impossible without the previous

Figure 1.23: Level lines based shape parser. Shape extraction has ever been the bottleneck of shape recognition algorithms. With the presented algorithm, this problem is solved by a brute force method : it compares all level lines of the images to be compared. Left pair of images : two images of a desk taken from different angles. In the left hand desk image, one level line has been put in white. We display, also in white, in the right image of the pair, the matching level lines. The match is ambiguous, as must be expected when the same object is repeated twice in the scene ! In the right pair of images, we display in white all matching pairs of level lines. (Experiment : J.-L. Lisani).

invariant filtering (AMSS). It is instead doable if the level lines have been significantly simplified. This simplification entails the possibility of compressed invariant encoding. In Figure 1.23, we present an experiment due to Lisani et al.. Two images of a desk are taken from different angles, and then, in white, all of the pieces of level lines in Image 1 and in Image 2 which found a match in the other image. In continuation, we present some of the matches. We notice that several of these matches are doubled : indeed, there are two similar chairs in each images ! A Gestalt law comes immediately to mind. This law states that human perception tends to group similar shapes. We now see the numerical necessity of this perceptual grouping : a previous self-matching of each image, with grouping of similar shapes, must be performed before we can compare it to other images !

1.2 The mathematical organisation

The mathematical organisation of this book derives from the above discussion of PDE models. We chose to focus on image analysis models, so that invariance requirements will lead the mathematical analysis. Most chapters include numerical experiments and indications on the right way to make them.

Chapter 2 We extensively analyse the heat equation, because it is a useful mathematical and algorithmic tool throughout the book. We first prove existence and uniqueness results for the heat equation and then prove that iterated linear smoothing filters are asymptotically equivalent to the heat equation. This is a first way to show its uniqueness as a linear smoothing filter.

Chapter 3 We then use the heat equation as a tool to :

- edge detection (Hildreth-Marr, Witkin)
- set smoothing (Koenderink-Van Dorn, Bence-Merriman-Osher)
- curve smoothing (Mackworth-Mokhtarian)

Chapter 4 We prove one of the main principles of mathematical morphology : a contrast invariant class of images is completely described by its set of level sets. Thus, contrast invariance reduces image analysis to set analysis.

Chapter 5 Level lines and level surfaces are introduced and the definitions and main formulae of curvature of level lines and, in higher dimension, of the principal curvatures of an isolevel surface given and proved. We also give a definition of level lines for discontinuous digital images and adress their visualization.

Chapter 7 We prove that monotone image operators that are contrast invariant boil down to set monotone operators. We explain how to define the set operator from the image operator and, conversely, how to construct a contrast invariant image operator from a set operator. Commutation of the monotone contrast invariant operators with thresholds is proved. As a first application, the Vincent-Serra "extrema killer" is formally defined.

Chapter 8 An analytic form is given to monotone contrast invariant and translation invariant image operators : If T is such an operator, it has a canonical "sup-inf" form,

$$Tu(\mathbf{x}) = \sup_{B \in I\!B} \inf_{\mathbf{y} \in B} u(\mathbf{x} + \mathbf{y}),$$

where $I\!B$ is a set of subsets of $I\!R^2$ called "set of structuring elements". Conversely, every image operator in the sup-inf form is monotone and contrast invariant. This theorem is a nonlinear version of the Riesz theorem, according to which a linear, continuous, translation invariant operator can be associated a convolution kernel k, so that

$$Tu(\mathbf{x}) = \int k(\mathbf{y}) u(\mathbf{x} - \mathbf{y}) d\mathbf{y}.$$

The kernel k is called "impulse response". In the same way, $I\!B$ is an impulse response for the nonlinear operator.

Chapter 9 The simplest sup-inf operators of mathematical morphology, the so called "dilations (only sup)" and "erosions (only inf)" are presented and analysed. Their subjacent PDE are identified as propagation equations

$$\frac{\partial u}{\partial t} = c|Du|,$$

where $c = 1$ for dilations and $c = -1$ for erosions.

Chapter 10 The median filter is one of the most emblematic and efficient contrast invariant monotone operator. It is defined and its numerical implementation discussed.

Chapter 11 This chapter focuses on the asymptotic expansion of contrast invariant operators when their scale tends to zero. We associate with each operator T its scaled version T_h associated with the shrinked set of structuring elements $\mathbb{B}_h = \{hB, B \in \mathbb{B}\}$. It is then proved roughly that

$$T_h u(\mathbf{x}) - u(\mathbf{x}) = ch|Du| + hH(hcurv(u))|Du|,$$

where c is a constant and H a nondecreasing function, both associated with T by simple formulas. In the case of the median filter it is proved that $c = 0$ and $H(s) = s$. We therefore have two main classes of contrast invariant operators :

- $c \neq 0$:the operator has the same asymptotic behaviour as a dilation if $c > 0$ and an erosion if $c < 0$.
- $c = 0$: the operator corresponds to a curvature dependent motion. The overall conclusion of this chapter is that all contrast invariant monotone operators boil down to erosions, dilations, and curvature motions.

Chapter 12 : performs the same analysis in arbitrary dimension.

Chapter 13 Because of the affine invariance requirement, a special attention is paid to affine invariant operators : an affine invariant distance of a point to a set is defined. In continuation, affine dilation and erosion operators are deduced and their set of structuring elements identified.

Chapter 14 Affine invariant families of structuring elements are not bounded, since by an affine map with determinant 1 we can arbitrarily stretch any set of the family. It is shown that the asymptotic behaviour of the affine invariant operator is not altered if we impose adequate bounds on the affine structuring elements.

Chapter 15 It is shown that if T is an affine invariant operator such that $T(-u) = -T(u)$, then its asymptotic behaviour is

$$T_h u(\mathbf{x}) - u(\mathbf{x}) = ch^{\frac{4}{3}}|Du|curv(u)^{\frac{1}{3}}.$$

In summary, all affine invariant contrast invariant operators are asymptotically equivalent to the affine PDE,

$$\frac{\partial u}{\partial t} = |Du|curv(u)^{\frac{1}{3}}.$$

Chapter 16 : adresses the extension of the preceding theory to the so called "nonflat" mathematical morphology, i.e. when the operators are no more contrast invariant, but only monotone. It is shown that an asymptotic expansion of a monotone operator can lead either to any of the Hamilton-Jacobi equations

$$\frac{\partial u}{\partial t} = H(Du)$$

or to parabolic equations (e.g. the heat equation). Actually, all of the intrinsic parabolic equations can derived from the Mathematical Morphology framework ! As applications, we analyse Rudin-Osher and Kramer's shock filters.

Chapter 17 The curvature equations are not smooth enough to be given classical solutions : we have a need for a concept of solutions compatible with contrast changes. Now, if u is a solution of the curvature equation and g a continuous increasing nondifferentiable contrast change, $g(u)$ should be another solution and is clearly not more than continuous ! The concept of viscosity solutions in the framework of Crandall-Lions, permits to solve this difficulty. We define this concept of solution and prove the main needed properties. In particular, we check that the erosions are dilations are viscosity solutions of the propagation equations

$$\frac{\partial u}{\partial t} = c|Du|$$

. We also prove a theorem, due to Barles and Souganidis, which permits to prove that the iteration of a contrast invariant operators T_h yields a discrete approximation of the subjacent curvature equation. We also give a brief account of the uniqueness theory for viscosity solutions.

Chapter 18 is devoted to the application of the mathematical techniques developped in the former chapter. We consider the main relevant contrast invariant operators, namely the median filter and the affine invariant erosion-dilation. We consider adequate rescaled versions of both, which we denote generically by T_h. Then we prove that $(T_h)^n u_0 \to u(t)$ when $nh \to t$ and $u(t)$ is a viscosity solution of the associated PDE, $\frac{\partial u}{\partial t} = |Du| curv(u)^{\frac{1}{3}}$ with $u(0) = u_0$, when T is an affine contrast invariant operator and $\frac{\partial u}{\partial t} = |Du| curv(u)$ when T is a median filter. Since the viscosity solution of these equations has been proved to be unique, we have both constructed this unique solution and proved that the considered iterated filter converges towards this unique solution !

Chapter 19 is an application of the results and techniques of the former chapters to the classical "active contour" or "geodesic snakes" method. We explain this method, show its structural properties and prove existence, uniqueness and contrast invariance of the motion of the analysing function v.

Chapters 20 and 21 : image scale space theory introduces a different point of view on image analysis, namely the Scale Space theory. A scale space is family of smoothing operators T_t, depending upon a scale parameter t, which associate with the original image u_0 more and more sketchy images $u(t)$. The original scale space theory proposes to compute $u(t)$ as the solution of the heat equation with u_0 as initial condition. Clearly, the viscosity solutions $u(t)$ of curvature equations give other scale space theories, more invariant. Now, aren't we missing other possibilities ? The axiomatic analysis performed in Chapters 18 and 19 permits first to explain which requirements (locality, invariance, comparison principle...) lead to a parabolic P.D.E. and then to classify all possible P.D.E.'s according to their invariance properties. The affine invariant equations are given in any dimension ; in dimension 2, the AMSS model (power 1/3 is proved to be the only possibility.

Chapter 22 A similar axiomatics analysis is performed on shape Scale Spaces. Here are also stated (without proof for once) the existence, uniqueness and regularity results for the curve evolutions to the powers 1 and 1/3. This shape analysis axiomatics turns out to be simpler than the general image scale space axiomatics, because it is based on the shape inclusion principle. It gives also a formal justification of the algorithm which, instead of moving the image by a curvature P.D.E., moves all level lines by the corresponding curve evolution equation. It is shown that the viscosity solution for the first equation indeed is obtained by moving all level lines with the second

Chapter 23 : Scale Space of movies. Here, a new axiom is introduced, the galilean invariance, which permits to single out a single most invariant P.D.E. for movies.

Chapter 24 is devoted to several different numerical strategies with finite difference schemes to implement image evolution by a curvature motion.

Part I

Linear Image analysis

Chapter 2

The Heat Equation

2.1 Image linear smoothing and the Laplacian

We note $\mathbf{x} = (x_1, ..., x_N)$ a point of $I\!R^N$, $\mathbf{x} = (x, y)$ if $N = 2$ and $|\mathbf{x}| = (x_1^2 + ... + x_N^2)^{\frac{1}{2}}$. If $\mathbf{x}, \mathbf{y} \in I\!R^N$, we denote by $\mathbf{x}.\mathbf{y} = x_1 y_1 + ... + x_N y_N$ their scalar product. Consider a bounded function $u_0(\mathbf{x})$, which we interpret as an image : $u_0(\mathbf{x})$ is the observed "grey level" at $\mathbf{x}$. We write, when $u(x, y)$ is a smooth enough function,

$$u_x = \frac{\partial u}{\partial x}, \quad u_y = \frac{\partial u}{\partial y}, \quad u_{xy} = \frac{\partial^2 u}{\partial x \partial y}$$

and, in the same way, if $u_0(\mathbf{x}) = u_0(x_1, ..., x_N)$,

$$u_i = \frac{\partial u}{\partial x_i}, \quad u_{ij} = \frac{\partial^2 u}{\partial x_i \partial x_j}, \quad \text{etc.}$$

We denote the gradient of u by

$$Du = (u_x, u_y)$$

when $u(\mathbf{x}) = u(x, y)$ and

$$Du = (u_1, ..., u_N)$$

when $u(\mathbf{x}) = u(x_1, ..., x_N)$. The Laplacian of u is denoted by

$$\Delta u = u_{xx} + u_{yy}$$

if $N = 2$ and

$$\Delta u = u_{11} + ... + u_{NN}$$

in general.

Assume we wish to establish a more reliable value $M_h u_0(\mathbf{x})$ of $u_0(\mathbf{x})$ as a mean value of u_0 over a neighborhood of $\mathbf{x}$ with size h. As for an obvious (but useful) example, let us mention the case where $Tu_0(\mathbf{x}) = (M_h u_0)(\mathbf{x})$ is obtained as the mean value of u_0 in a neighborhood of $\mathbf{x}$ and let us take $N = 2$. We set

$$M_h u_0(\mathbf{x}) = \frac{1}{\pi h^2} \int_{D(\mathbf{x},h)} u_0(\mathbf{y}) d\mathbf{y}, \tag{2.1}$$

In this case, the locality means that only values of u_0 inside a disk $D(\mathbf{x}, h)$ around $\mathbf{x}$ matter. When h tends to zero, we get a more and more local estimate of the value of u_0 at $\mathbf{x}$. The parameter h characterizes the "locality" of the considered operator, that is, the size of the neighborhood involved in the filtering operation. The question arises of whether such an operation can be made independent of h. A first answer is to analyze the asymptotic expansion of $M_h u_0(\mathbf{x})$ as h tends to zero. Clearly, if u_0 is continuous at $\mathbf{x}$, we have $M_h u_0(\mathbf{x}) \to u_0(\mathbf{x})$. If we assume that u_0 is twice continuously differentiable (C^2) at $\mathbf{x}$, then it is easily seen that in fact [1]

Figure 2.1: Local averaging algorithm. Left : original image, right : result of replacing the grey level at each pixel by the average grey level over the neighboring pixels. The shape of the neighborhood is defined by the black spot displayed up-right.

$$\frac{M_h u_0(\mathbf{x}) - u_0(\mathbf{x})}{h^2} = \frac{1}{8}\Delta u_0(\mathbf{x}) + \varepsilon(h) \tag{2.2}$$

In order to prove this formula, let us set, without loss of generality, $\mathbf{x} = 0$. Then for $\mathbf{y} = (x, y)$ in $D(0, h)$, we have by Taylor formula

$$u_0(\mathbf{y}) = u_0(0) + Du_0(0).\mathbf{y} + \frac{1}{2}((u_0)_{xx}x^2 + (u_0)_{yy}y^2 + 2(u_0)_{xy}xy) + o(h^2).$$

Taking the mean value over $D(0, h)$, that is, applying M_h, we obtain

$$(M_h u_0)(0) = u_0(0) + \frac{1}{2\pi h^2}((u_0)_{xx}(0)\int_{D(0,h)} x^2 dxdy + (u_0)_{yy}(0)\int_{D(0,h)} y^2 dxdy) + o(h^2).$$

Since

$$\frac{1}{2\pi h^2}\int_{D(0,h)} x^2 dxdy = \frac{1}{4\pi h^2}\int_{D(0,h)} (x^2 + y^2)dxdy$$

$$= \frac{1}{4\pi h^2}\int_0^h 2\pi r^3 dr = \frac{h^2}{8},$$

we obtain the announced formula (2.2). This formula suggests the following result : consider for any t a sequence of real numbers h and integers n such that $nh^2 \to t$. Set $u^n = M_h^n u_0$. Then it is to be expected that $u^n(\mathbf{x}) \to u(t, \mathbf{x})$ where $u(t, \mathbf{x})$ satisfies the heat equation associated with (2.2),

[1] We always denote in this text by $o(h)$ a function of h which tends to zero faster than h, by $O(h)$ a function such that, for some constant C, $|O(h)| \leq C|h|$, and by $\varepsilon(h)$ a function tending to zero as h tends to zero. Thus an $o(h)$ function can also be written : $h\varepsilon(h)$.

$$\frac{\partial u}{\partial t}(t,\mathbf{x}) = \frac{1}{8}\Delta u(t,\mathbf{x}), \quad u(0) = u_0. \tag{2.3}$$

This convergence result will be obtained in Section 3.1. We need some preliminaries on the heat equation. In Section 2.2 below, we shall prove that the heat equation has a unique solution for a given initial datum u_0 and is equivalent to the convolution of u_0 with Gaussian kernels of increasing width.

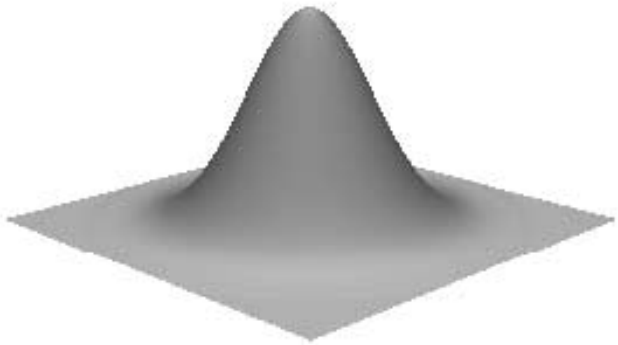

Figure 2.2: The Gauss kernel in two dimensions.

2.2 Existence and uniqueness of solutions of the heat equation

Define an image $u_0(\mathbf{x})$, $\mathbf{x} \in I\!R^N$, as a real function which is primarily defined on the hypercube $[0,1]^N$, and subsequently extended to $C = [-1,1]^N$ by symmetry across the coordinate hyper-planes. This extension satisfies for every $\mathbf{x} = (x_1, ..., x_N)$ in C,

$$u(\varepsilon_1 x_1, ..., \varepsilon_N x_N) = u(x_1, ..., x_N) \tag{2.4}$$

where $(\varepsilon_1, ..., \varepsilon_N)$ takes all possible values in $\{-1,1\}^N$. We then extend $u(\mathbf{x})$ by periodization into a function on all of $I\!R^N$ which is 2-periodic with respect to all variables, so that

$$u(x_1 + 2n_1, ..., x_N + 2n_N) = u(x_1, ..., x_N) \tag{2.5}$$

for all $(n_1, ..., n_N) \in Z\!\!Z^N$.

The aim of these successive extensions is first to preserve the continuity properties of u across the boundary of $[0,1]^N$ and second to have u defined on all of $I\!R^N$, so that (e.g.) convolutions of u with another function can be defined easily. This way of extending u is classical in image processing and used in most compression and transmission standards. It is easily checked that if u is continuous on $[0,1]$, its extension to $I\!R^N$ defined in the preceding way also is continuous.

Definition 2.1 *Whenever a function u_0 defined on $I\!R^N$ satisfies the periodicity condition (2.5), we shall say that u_0 belongs to L_C. If it is bounded, we say that it belongs to L_C^∞. If it is integrable, that is,*

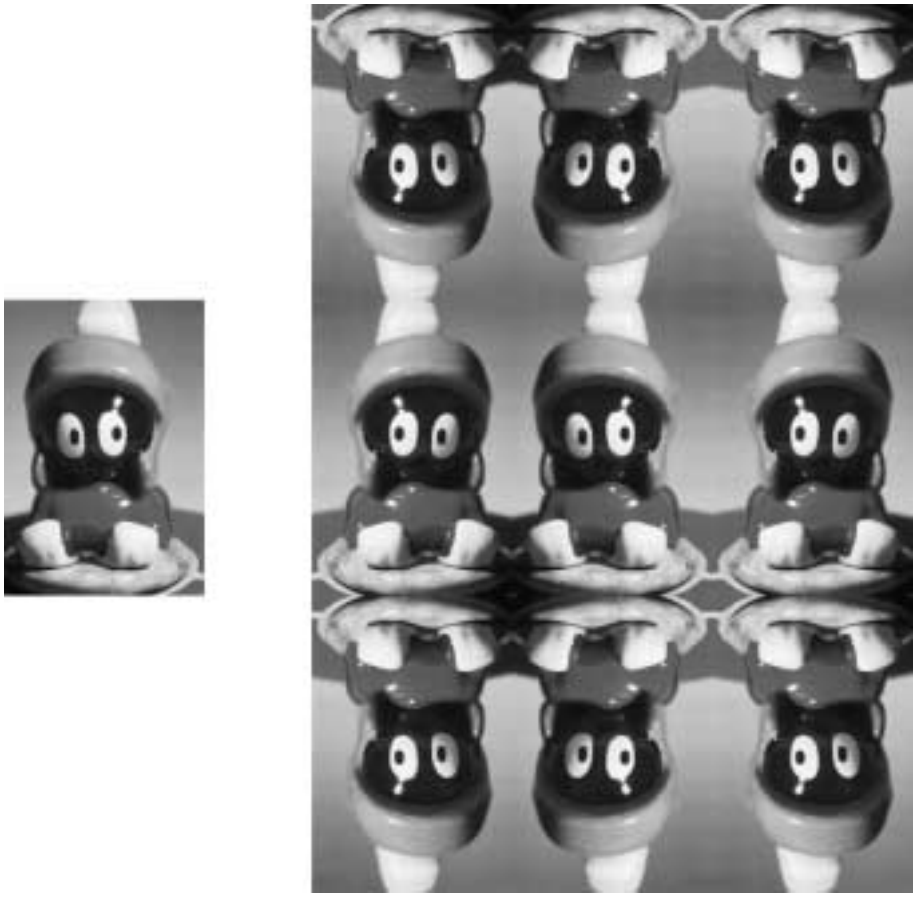

Figure 2.3: Image extension by symmetry and then periodization.

$\int_C |u_0(\mathbf{x})| d\mathbf{x} < \infty$, *we say that* u_0 *belongs to* L^1_C . *We endow theses spaces with norms, which we shall use alternatively in the statements to come.*

$$\textit{The } L^\infty_C \textit{ norm } : ||u||_\infty = \sup_{\mathbf{x}\in C} |u(\mathbf{x})|$$

$$\textit{and the } L^1_C \textit{ norm } : ||u||_{L^1(C)} = \int_C |u(\mathbf{x})| d\mathbf{x}.$$

If in addition the function satisfy the symmetries (2.4), we say that it is symmetric.
When $||u-v||_\infty \to 0$, we say that u converges to v uniformly. When $||u-v||_{L^1(C)} \to 0$, we say that u converges to v in L^1_C. When we assume that a function $u \in L^\infty_C$ is continuous, or C^m, or C^∞, this means that u has such properties on all of $\mathbb{R}^N$ and not only on C. If u is continuous, then it is uniformly continuous. Indeed, by periodicity, $\sup_{\mathbf{X}\in\mathbb{R}^N} |u(\mathbf{x}+h) - u(\mathbf{x})| = \sup_{\mathbf{X}\in C} |u(\mathbf{x}+h) - u(\mathbf{x})|$ and by the compactness of C, $\sup_{\mathbf{X}\in C} |u(\mathbf{x}+h) - u(\mathbf{x})| = o(1)$. In the same way, if u is C^1, notice that $\sup_{\mathbf{X}\in\mathbb{R}^N} |Du|(\mathbf{x}) = \sup_{\mathbf{X}\in C} |Du|(\mathbf{x})$, etc.
We shall also consider functions g integrable on all of $\mathbb{R}^N$; in such a case, we write

$$||g||_{L^1} = ||g||_{L^1(\mathbb{R}^N)} = \int_{\mathbb{R}^N} |g(\mathbf{x})| d\mathbf{x}.$$

We recall a classical density result in L^1 : (See e.g. [360])

Proposition 2.2 *If $g \in L^1(\mathbb{R}^N)$, then there exists a sequence g_n of continuous functions which are zero outside a compact set such that $g_n \to g$ in $L^1(\mathbb{R}^N)$, that is, $\int_{\mathbb{R}^N} |g_n(\mathbf{x}) - g(\mathbf{x})| d\mathbf{x} \to 0$ as $n \to +\infty$.*
If $u \in L^1_C$, there exists a sequence of continuous functions $u_n \in L^1_C$ such that $||u_n - u||_{L^1(C)} \to 0$.

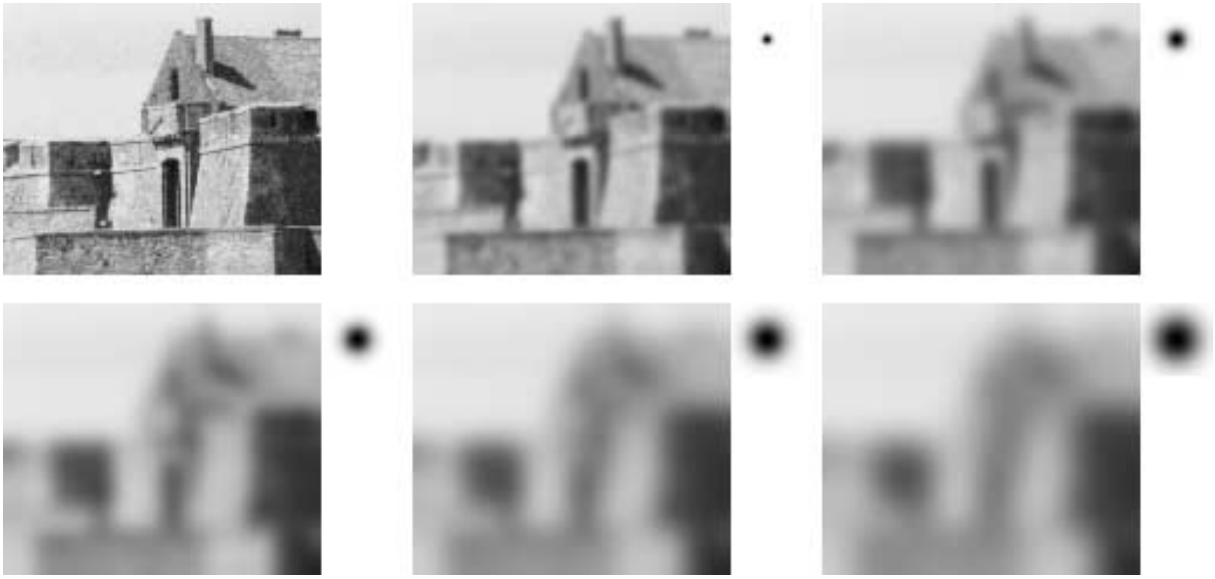

Figure 2.4: Convolution by gaussian kernels (heat equation). From top-left to bottom-right, we display the original image, and the results of the convolutions with gaussians of increasing variance. A grey level representation of the convolution kernel is put on the right of each convolved image ; it gives an idea of the size of the involved neighborhood. The resulting image is more and more blurry.

A consequence of the density result, also classical is :

Proposition 2.3 *If $u \in L^1_C$, then*

$$\int_C |u(\mathbf{x} - \mathbf{y}) - u(\mathbf{x})| d\mathbf{x} = \varepsilon(|\mathbf{y}|).$$

If $u \in L(\mathbb{R}^N)$, then $u \in L^1_C$, then

$$\int_{\mathbb{R}^N} |u(\mathbf{x} - \mathbf{y}) - u(\mathbf{x})| d\mathbf{x} = \varepsilon(|\mathbf{y}|).$$

Proof Fix $\varepsilon > 0$. By Proposition 2.2, we can find $v \in L^\infty_C$, continuous, such that $||u - v||_{L^1(C)} \leq \varepsilon$. Since v is uniformly continuous, we have $\sup_{\mathbf{x} \in \mathbb{R}^N} |v(\mathbf{x} - \mathbf{y}) - v(\mathbf{x})| \leq \varepsilon$ for $|\mathbf{y}| < \eta$ small enough. Thus

$$\int_C |u(\mathbf{x} - \mathbf{y}) - u(\mathbf{x})| d\mathbf{x} \leq$$

$$\int_C |u(\mathbf{x} - \mathbf{y}) - v(\mathbf{x} - \mathbf{y})| d\mathbf{x} + \int_C |v(\mathbf{x} - \mathbf{y}) - v(\mathbf{x})| d\mathbf{x} + \int_C |v(\mathbf{x}) - u(\mathbf{x})| d\mathbf{x} \leq 3\varepsilon.$$

The proof of the second statement is an obvious adaptation of the first. □

We shall use multi-indices for derivation, which we denote by $\alpha = (\alpha_1, ..., \alpha_N) \in I\!N^N$, or $\beta = (\beta_1, ..., \beta_N) \in I\!N^N$. We write $\mathbf{x}^\alpha$ and $|\mathbf{x}|^\alpha$ for

$$x_1^{\alpha_1} x_2^{\alpha_2} ... x_N^{\alpha_N} \text{ and } |x_1|^{\alpha_1} |x_2|^{\alpha_2} ... |x_N|^{\alpha_N}$$

respectively. We shall abbreviate the partial derivatives of a function g by setting

$$\partial^\alpha g = \frac{\partial^{\alpha_1 + ... + \alpha_N}}{\partial x_1^{\alpha_1} ... \partial x_N^{\alpha_N}} g.$$

We write $g \in C^\infty$ if g is differentiable of any order.

Definition 2.4 *We say that a function g defined on $I\!R^N$ belongs to the Schwartz class $\mathcal{S}$ if $g \in C^\infty$ and for each pair of multi-indices α, β there is a constant C such that*

$$|\mathbf{x}|^\beta |\partial^\alpha g(\mathbf{x})| \leq C.$$

Proposition 2.5 *If $g \in \mathcal{S}$, then g is integrable on $I\!R^N$: $\int_{I\!R^N} |g(\mathbf{x})| d\mathbf{x} < \infty$. For all multi-indices α, β the function $\mathbf{x}^\beta \partial^\alpha g$ also belongs to $\mathcal{S}$. In addition, for all α, $\partial^\alpha g$ is uniformly continuous on $I\!R^N$.*

Proof The second statement immediately follows from the Leibnitz rule for differentiating a product. The first one follows from the fact that we can write $|g(x)| \leq \frac{C}{1+|\mathbf{X}|^{N+2}}$, which is an integrable function on $I\!R^N$. Finally, every continuous function on $I\!R^N$ tending to zero at infinity also is uniformly continuous, so that the last statement is valid. □

Proposition 2.6 (The Gaussian and the heat equation)
For all $t > 0$, the function $\mathbf{x} \to G_t(\mathbf{x}) = \frac{1}{(4\pi t)^{\frac{N}{2}}} e^{-\frac{|\mathbf{X}|^2}{4t}}$ belongs to $\mathcal{S}$ and satisfies the heat equation

$$\frac{\partial G_t}{\partial t} - \Delta G_t = 0.$$

Proof It is enough to prove the first statement for the function $g(\mathbf{x}) = e^{-|\mathbf{x}|^2}$. An easy induction argument shows that $\partial^\alpha g(\mathbf{x}) = P_\alpha(\mathbf{x}) e^{-|\mathbf{x}|^2}$, where P_α is a n-variable polynomial. Using the fact that for every k, $x^k e^{-x^2} \to 0$ as $x \to \infty$ concludes the argument. An easy calculation shows that G_t satisfies the heat equation. □

The main operation in linear image filtering is the convolution of the image with positive integrable functions. (The prototype of such a "convolution kernel", is the Gauss function).

Proposition 2.7 (and definition of the convolution) *Let u be a function in L^1_C and $g \in L^1(I\!R^N)$. Set $(g * u)(\mathbf{x}) = \int_{I\!R^N} u(\mathbf{x} - \mathbf{y}) g(\mathbf{y}) d\mathbf{y}$. Then $(g * u)(\mathbf{x})$ is defined for almost every $\mathbf{x} \in C$. The convolution function $g * u$ belongs to L^1_C and we have*

$$||g * u||_{L^1(C)} \leq ||g||_{L^1(I\!R^N)} ||u||_{L^1(C)}. \tag{2.6}$$

If in addition u is bounded, then $g * u$ is also bounded and

$$||g * u||_{L^\infty(C)} \leq ||g||_{L^1(I\!R^N)} ||u||_{L^\infty(I\!R^N)}. \tag{2.7}$$

Proof Let us start by proving that $g * u$ is well defined and belongs to L^∞_C and that (2.6) holds. This is an direct consequence of Fubini Theorem ([?]), which yields

$$||g * u||_{L^1(C)} \leq \int_C (\int_{I\!R^N} |u(\mathbf{x}-\mathbf{y})g(\mathbf{y})|d\mathbf{y})d\mathbf{x} = \int_{I\!R^N} (\int_C |u(\mathbf{x}-\mathbf{y})g(\mathbf{y})|d\mathbf{x})d\mathbf{y}.$$

Using the periodicity of u, this last integral is equal to

$$\int_{I\!R^N} \int_C |u(\mathbf{x})|d\mathbf{x})|g(\mathbf{y})|d\mathbf{y} = ||u||_{L^1(C)}||g||_{L^1(I\!R^N)}.$$

Relation (2.7) is an obvious consequence of the definition of $g * u$ and the 2-periodicity is easily checked. □

Exercise 2.1 *Check that $g * u$ is 2-periodic. Show that if, in addition to the assumptions of Proposition 2.7, g is even and u is symmetric, i.e. satisfies the symmetry relations (2.4), then $g * u$ also is symmetric.*

We shall also be led to consider the convolution of a function in L^1_C with an integrable function $g \in L^1(I\!R^N)$. The next lemma gives a useful general condition on g in order that the convolution of a function u in L^1_C with $g \in L^1(I\!R^N)$ be bounded.

Lemma 2.8 *Let $g \in L^1(I\!R^N)$ be an integrable and locally bounded function such that for any $R > 0$, $g^R(\mathbf{x}) = \sup_{\mathbf{y}\in B(\mathbf{x},R)} g(\mathbf{y})$ also belongs to $L^1(I\!R^N)$. Then there exists a constant $C(g)$ such that for any $u \in L^1_C$, $g * u$ is in L^∞_C and*

$$||g * u||_{L^\infty(C)} \leq C(g)||u||_{L^1(C)}. \tag{2.8}$$

Proof We set $\mathbf{z} = (z_1, ..., z_N) \in (2Z\!\!\!Z)^N$. The hypercubes $\mathbf{z} + C$ cover $I\!R^N$ and for every $\mathbf{y} \in \mathbf{z} + C$, we have $|\mathbf{y} - \mathbf{z}| \leq \text{diam}(C)$, where

$$\text{diam}(C) = \sup_{\mathbf{y},\mathbf{z}\in C} |\mathbf{y} - \mathbf{z}| = 2\sqrt{N}.$$

Thus, using the 2-periodicity of u,

$$|(g * u)(\mathbf{x})| \leq \int_{I\!R^N} |u(\mathbf{x}-\mathbf{y})g(\mathbf{y})|d\mathbf{y} = \sum_{\mathbf{z}\in(2Z\!\!\!Z)^N} \int_{\mathbf{z}+C} |u(\mathbf{x}-\mathbf{y})g(\mathbf{y})|d\mathbf{y} \leq$$

$$\leq (\int_C |u(\mathbf{y})|d\mathbf{y}) \sum_{\mathbf{z}\in(2Z\!\!\!Z)^N} \sup_{\mathbf{y}\in\mathbf{z}+C} |g(\mathbf{y})|,$$

which implies

$$|(g * u)(\mathbf{x})| \leq (\int_C |u(\mathbf{y})|d\mathbf{y}) \int_{I\!R^N} |g^{\text{diam}(C)}(\mathbf{y})|d\mathbf{y} = C(g)||u||_{L^1(C)}.$$

□

Let us now focus on the case where the convolution kernel, like the Gauss function, belongs to $\mathcal{S}$.

Proposition 2.9 *If $u \in L^1_C$ and $g \in \mathcal{S}$, then $g * u \in C^\infty(I\!R^N) \cap L^\infty_C$ and*

$$\partial^\alpha(g * u) = (\partial^\alpha g) * u \tag{2.9}$$

for every multi-index α.

Proof By Proposition 2.5, $||g||_{L^1} = \int_{I\!R^N} |g(\mathbf{x})|d\mathbf{x} < \infty$. Thus by Proposition 2.7, $g * u$ belongs to L^1_C. In order to prove (2.9), we notice that it is enough to prove it for $\alpha = (1, 0, ...0)$. Indeed, we know by Proposition 2.5 that $\partial^\alpha g$ is in $\mathcal{S}$ if g is, so that the general case for (2.9) follows from the case $\alpha = (1, 0, ...0)$ by an obvious induction. Using the Taylor expansion formula and setting $e_1 = (1, 0, ..., 0)$,

$$(g * u)(\mathbf{x} + he_1) - (g * u)(\mathbf{x}) = \int_{I\!R^N} (g(\mathbf{x} + he_1 - \mathbf{y}) - g(\mathbf{x} - \mathbf{y}))u(\mathbf{y})d\mathbf{y} =$$

$$= h \int_{I\!R^N} \frac{\partial g}{\partial x_1}(\mathbf{x} - \mathbf{y})u(\mathbf{y})d\mathbf{y} + \frac{h^2}{2} \int_{I\!R^N} \frac{\partial^2 g}{\partial x_1^2}(\mathbf{x} + \theta he_1 - \mathbf{y})u(\mathbf{y})d\mathbf{y},$$

where $0 \leq \theta = \theta(\mathbf{y}) \leq 1$. Since $g \in \mathcal{S}$, $|\frac{\partial^2 g}{\partial x_1^2}(\mathbf{x})| \leq \frac{C}{1+|\mathbf{x}|^{N+2}}$ and therefore, for $|h| \leq 1$:

$$\frac{h^2}{2} | \int_{I\!R^N} \frac{\partial^2 g}{\partial x_1^2}(\mathbf{x} + \theta he_1 - \mathbf{y})u(\mathbf{y})d\mathbf{y}| \leq \frac{h^2}{2} \int_{I\!R^N} \frac{C|u(\mathbf{y})|d\mathbf{y}}{1 + ((|\mathbf{x} - \mathbf{y}| - 1)^+)^{N+2}} = Ch^2.$$

(We note $r^+ = \sup(r, 0)$.) Indeed, the last integral is finite, being the convolution of $u \in L^1_C$ with an integrable function satisfying the assumptions of Lemma 2.8. So we deduce that $g * u$ is differentiable in x_1 and $\frac{\partial(g*u)}{\partial x_1} = (\frac{\partial g}{\partial x_1}) * u$. □

Proposition 2.10 *Let $g \in \mathcal{S}$, $g \geq 0$, $\int_{I\!R^N} g(\mathbf{x})d\mathbf{x} = 1$ and set $g_t = \frac{1}{t^N} g(\frac{\mathbf{x}}{t})$ for $t > 0$.*
i) If $u_0 \in L^\infty_C$ is continuous, $g_t * u_0$ converges uniformly to u_0, when $t \to 0$. In addition, we have a maximum principle :

$$\inf_{\mathbf{x} \in C} u_0(\mathbf{x}) \leq g_t * u_0 \leq \sup_{\mathbf{x} \in C} u_0(\mathbf{x}) \tag{2.10}$$

ii) If we only know $u_0 \in L^1_C$, then

$$\int_C |(g_t * u_0)(\mathbf{x}) - u_0(\mathbf{x})|d\mathbf{x} = ||g_t * u_0 - u_0||_{L^1(C)} \to 0 \quad as \quad t \to 0.$$

Proof of i) We remark that

$$\int_{I\!R^N} g_t(\mathbf{y})d\mathbf{y} = 1 \tag{2.11}$$

and

$$\forall \eta > 0, \int_{\mathbf{y} \in I\!R^N,\ |\mathbf{y}| \geq \eta} g_t(\mathbf{y})d\mathbf{y} \to 0 \ \text{ as } t \to 0. \tag{2.12}$$

Using (2.11), we have

$$g_t * u_0(\mathbf{x}) - u_0(\mathbf{x}) = \int g_t(\mathbf{y})(u_0(\mathbf{x} - \mathbf{y}) - u_0(\mathbf{x}))d\mathbf{y}.$$

Now, as already mentionned, since u_0 is continuous on the compact C and periodic, it is in fact uniformly continuous, so that for $|\mathbf{y}| \leq \eta(\varepsilon)$ we have $|u_0(\mathbf{x} - \mathbf{y}) - u_0(\mathbf{x}))| \leq \varepsilon$. Using this inequality and (2.12) we obtain

$$|g_t * u_0(\mathbf{x}) - u_0(\mathbf{x})| \leq \int_{|\mathbf{y}| \leq \eta} |g_t(\mathbf{y})(u_0(\mathbf{x} - \mathbf{y}) - u_0(\mathbf{x}))|d\mathbf{y} + \int_{|\mathbf{y}| \geq \eta} |g_t(\mathbf{y})(u_0(\mathbf{x} - \mathbf{y}) - u_0(\mathbf{x}))|d\mathbf{y}$$

$$\leq \varepsilon + 2||u_0||_{L^\infty(C)} \int_{|\mathbf{y}| \geq \eta} g_t(\mathbf{y})d\mathbf{y} \leq 2\varepsilon \text{ for } t \text{ small enough.}$$

Relation (2.10) immediately follows from the assumptions $g_t \geq 0$, $\int g_t = 1$.

Proof of ii) By Proposition 2.3, if $u \in L^1(C)$, then we can find for every ε a real number $\eta(\varepsilon)$ such that

$$\int_C |u(\mathbf{x}-\mathbf{y}) - u(\mathbf{x})| d\mathbf{x} \leq \varepsilon \text{ for } |\mathbf{y}| \leq \eta(\varepsilon).$$

Thus, by Fubini Theorem again, using $\int_{I\!R^N} g_t(\mathbf{y}) d\mathbf{y} = 1$ and (2.12):

$$\int_C (\int_{I\!R^N} |g_t(\mathbf{y})(u(\mathbf{x}-\mathbf{y}) - u(\mathbf{x}))| d\mathbf{y}) d\mathbf{x} \leq$$

$$\int_{|\mathbf{y}|\leq\eta(\varepsilon)} g_t(\mathbf{y})(\int_C |u(\mathbf{x}-\mathbf{y}) - u(\mathbf{x})| d\mathbf{x}) d\mathbf{y} + \int_{|\mathbf{y}|>\eta(\varepsilon)} g_t(\mathbf{y})(2\int_C |u(\mathbf{x})| d\mathbf{x}) d\mathbf{y} \leq 2\varepsilon$$

for t small enough. □

We now have all tools at hand to prove the main theorem of this section.

Theorem 2.11 (Existence and Uniqueness of solutions for the heat equation)
Let $u_0 \in L^1_C$. Then
i) $u(t) = G_t * u_0$ satisfies, for all $t > 0$ and $\mathbf{x} \in I\!R^N$, the heat equation with initial value u_0, that is

$$\frac{\partial u}{\partial t} = \Delta u \text{ and } \int_C |u(t,\mathbf{x}) - u_0(\mathbf{x})| d\mathbf{x} \to 0 \text{ as } t \to 0. \tag{2.13}$$

In addition, $u(t,\mathbf{x})$ is C^∞ for $t > 0$ and $\mathbf{x} \in I\!R^N$, it belongs for every $t > 0$ to L^1_C and is uniformly bounded for $t \in [t_1, +\infty[$ where t_1 is any positive real number:

$$\sup_{\mathbf{x}\in I\!R^N, t\geq t_1} |u(t,\mathbf{x})| \leq C(t_1)||u_0||_{L^1(C)}. \tag{2.14}$$

ii) Conversely, given u_0 as above, there is a unique solution $u(t,\mathbf{x})$ of (2.13) belonging to L^∞_C for each $t > 0$, bounded on $[t_1, +\infty[$ for each $t_1 > 0$ and C^3 on $]0, +\infty[\times I\!R^N$.

Proof i) By Propositions 2.5 and 2.6, G_t and its derivatives belong to $\mathcal{S}$. By Proposition 2.9, we have

$$\frac{\partial u}{\partial t} - \Delta u = u * (\frac{\partial G_t}{\partial t} - \Delta G_t),$$

which is zero by Proposition 2.6. Relation (2.14) follows from Lemma 2.8 applied to u and $g = G_t$. It is easily seen that $C(G_t) \leq C(t_1)$ is a bounded function for $t \geq t_1$.
ii) Let us now show the uniqueness. Let v and w be two solutions of the heat equation (2.13) with the same initial datum $u_0 \in L^1_C$ and bounded for $t \geq t_1 > 0$. Thus $u = v - w$ satisfies the heat equation with initial datum $u_0 = 0$ and is bounded again on each interval $[t_1, +\infty[$, $t_1 > 0$. Let us assume by contradiction that $u(t,\mathbf{x}) \neq 0$, denote by $u(t)$ the partial function $\mathbf{x} \to u(t,\mathbf{x})$ and set $u_h(t,\mathbf{x}) = (G_h * u(t))(\mathbf{x})$. Then, by Propositions 2.7 and 2.9, u_h satisfies the same properties as u, is again solution of the heat equation and we have by Lemma 2.8 the additional property that $u_h(t,\mathbf{x}) \to 0$ uniformly as $t \to 0$. Indeed, using the initial value condition in (2.13),

$$||u_h(t)||_{L^\infty(C)} = ||G_h * u(t)||_{L^\infty(C)} \leq C(G_h)||u(t)||_{L^1(C)} \to 0 \text{ as } t \to 0.$$

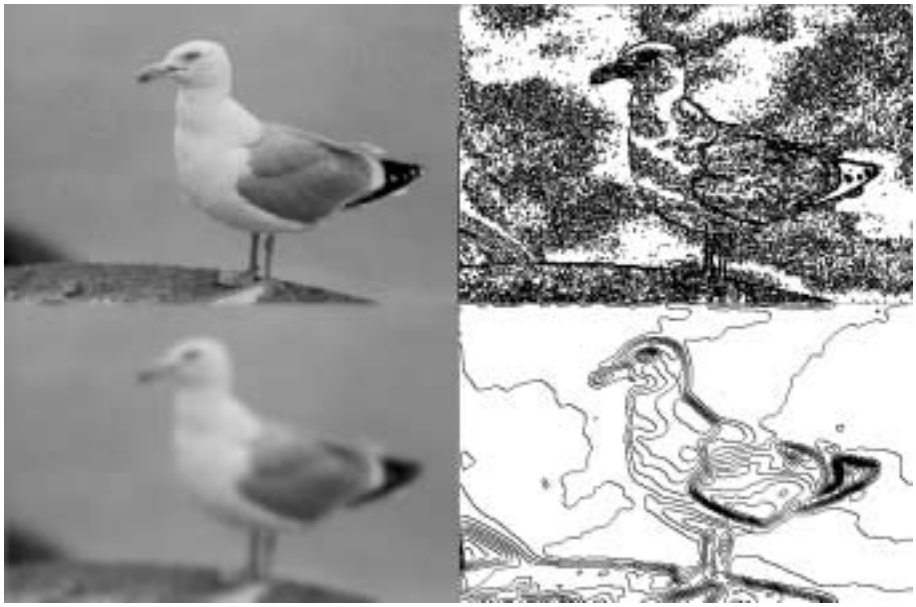

Figure 2.5: Level lines and the heat equation. On the first row, we display on the left a 410×270 grey level image, photograph of a sea bird and on the right : its level lines for levels multiple of 12. On the second row, the heat equation, i.e. a convolution with a gaussian, has been applied to the original image. The standard deviation of the gaussian is 4, which means that its spatial range is comparable to a disk with radius 4. The image gets blurred by the convolution, which mixes grey level values and removes all sharp edges. This can be appreciated on the right, where we have displayed all level lines for levels multiple of 12. We can see how the level lines on the boundaries of the image have split into parallel level lines which have gone away from each other. The image has become smooth, but is losing its structure.

In addition, $u_h(t, \mathbf{x})$ is bounded. Choosing h small enough and changing u into $-u$ if necessary, we can assume by Proposition 2.10 that $u_h(t, \mathbf{x}) > 0$ at some $(t, \mathbf{x})$. We now consider the new function $u^\varepsilon(t, \mathbf{x}) = e^{-\varepsilon t} u_h(t, \mathbf{x})$. This continuous and periodic (in $\mathbf{x}$) function belongs for every $t \geq 0$ to L^∞_C and tends uniformly to zero as $t \to \infty$ or 0. Thus, its supremum is attained at a point $(t_0, \mathbf{x}_0)$ such that $t_0 > 0$ and $\mathbf{x}_0 \in C$. At $(t_0, \mathbf{x}_0)$, we must have $\frac{\partial u^\varepsilon}{\partial t}(t_0, \mathbf{x}_0) = 0$ and $\Delta u^\varepsilon(t_0, \mathbf{x}_0) = e^{-\varepsilon t} \Delta u_h \leq 0$. Since u_h is a solution of the heat equation, we have

$$0 = \frac{\partial u^\varepsilon}{\partial t}(t_0, \mathbf{x}_0) = (-\varepsilon u^\varepsilon + e^{-\varepsilon t} \frac{\partial u_h}{\partial t})(t_0, \mathbf{x}_0)$$

$$= -\varepsilon u^\varepsilon(t_0, \mathbf{x}_0) + e^{-\varepsilon t} \Delta u_h(t_0, \mathbf{x}_0) \leq -\varepsilon u^\varepsilon(t_0, \mathbf{x}_0) < 0.$$

This yields a contradiction and we conclude that the solution of the heat equation is unique. □

References.

The Riesz theorem [360] states that every continuous translation invariant function $u \to Tu$ operator is a convolution $u \to g * u$ where g is a smoothing kernel. This is one of the fundamental theorems of both mechanics and signal processing [277]. If, in addition, $u \geq 0 \Rightarrow Tu \geq 0$, then $g \geq 0$ is a nonnegative kernel. The Gauss kernel is then paragdimatic, since, as we shall show in the next chapter, it is the only one to be stable by iterated convolutions. There are several types of continuity for T, leading to several

types of properties for g. For instance, if T is continuous from $L^2(\mathbb{R}^N)$ into $C^0(\mathbb{R}^N)$ then $g \in L^2(\mathbb{R}^N)$. Thus, all linear stable image filtering operators are described by their convolution kernel g. Hence the analysis of convolutions performed in Section 2.2 as a necessary tool for image processing. The existence and uniqueness proof of Theorem 2.11 is classical and to be found in all classical treatises on Partial Differential Equations like Evans [141], or Brezis [64]. In this chapter, we have used the classical Lebesgue integral. A short and comprehensive presentation can be found in Rudin [360].

Chapter 3

Applications of the heat equation to image analysis.

3.1 Convergence of iterated smoothing filters to the heat equation

In this section, we prove roughly that the heat equation (or the convolution by a Gauss kernel at different scales) is the asymptotic state of any iterative linear isotropic smoothing.

Definition 3.1 *Let $g(\mathbf{x}) \in L^1(I\!R^N)$ be a real function, which we understand as a smoothing kernel with which we intend to convolve images. We say that g is radial if $g(\mathbf{x}) = g(|\mathbf{x}|)$ only depends on the norm of $\mathbf{x}$. We say that g is pseudo-radial if it satisfies the following relations :*

$$\int_{I\!R^N} g(\mathbf{x}) d\mathbf{x} = 1 \tag{3.1}$$

and, for every $i, j = 1, .., N$, such that $i \neq j$,

$$\int_{I\!R^N} \mathbf{x}_i g(\mathbf{x}) = \int_{I\!R^N} \mathbf{x}_i \mathbf{x}_j g(\mathbf{x}) d\mathbf{x} = 0, \tag{3.2}$$

$$\int_{I\!R^N} \mathbf{x}_i^2 g(\mathbf{x}) d\mathbf{x} = \sigma. \tag{3.3}$$

Exercise 3.1 *Show that if g is a radial function, then an adequate rescaling $ag(\frac{\mathbf{x}}{b})$ of g satisfies the relations (3.1) and (3.3) with $\sigma = 2$ and preserves the zero moment relations (3.2). Thus, there is no loss of generality in assuming that these relations are true for g.*

We consider rescalings of g,

$$g_h(\mathbf{x}) = \frac{1}{h^{\frac{N}{2}}} g(\frac{\mathbf{x}}{h^{\frac{1}{2}}}), \tag{3.4}$$

which concentrate g and maintain (3.1) and (3.2). In the following, we note $g^{n*} = g * g * ... * g$, the n-times convolution of a function g. Our main concern is the behavior of g_h^{n*} as $n \to \infty$ and $h \to 0$. Let us first see what happens with the convolution by g_h as $h \to 0$.

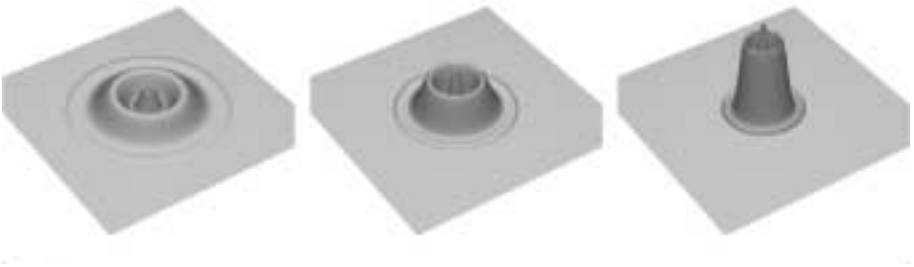

Figure 3.1:
A kernel g and its rescalings $g_t = \frac{1}{t^2} g(x/t)$, for t=2, 3, 4.

Theorem 3.2 *Let $g(\mathbf{x}) \in L^1(\mathbb{R}^N)$ be a radial or pseudo-radial function satisfying the conditions (3.1) to (3.3). Assume further that*

$$\int_{\mathbb{R}^N} |g(\mathbf{z})||\mathbf{z}|^3 d\mathbf{z} = C < +\infty. \tag{3.5}$$

Then for every C^3 function $u \in L^\infty_C$,

$$(g_h * u)(\mathbf{x}) - u(\mathbf{x}) = h\frac{\sigma}{2}\Delta u(\mathbf{x}) + O(h^{\frac{3}{2}}), \tag{3.6}$$

where $|O(h^{\frac{3}{2}})(\mathbf{x})| \leq C h^{\frac{3}{2}} \max_{\mathbf{x}\in C} ||D^3 u(\mathbf{x})||$.

Proof Using (3.1), a rescaling inside the integrals and a Taylor expansion of u,

$$(g_h * u)(\mathbf{x}) - u(\mathbf{x}) = \int_{\mathbb{R}^N} h^{-\frac{N}{2}} g(\frac{\mathbf{y}}{h^{\frac{1}{2}}})(u(\mathbf{x}-\mathbf{y}) - u(\mathbf{x}))d\mathbf{y} = \int_{\mathbb{R}^N} g(\mathbf{z})(u(\mathbf{x} - h^{\frac{1}{2}}\mathbf{z}) - u(\mathbf{x}))d\mathbf{z}$$

$$= \int_{\mathbb{R}^N} g(\mathbf{z})(-h^{\frac{1}{2}} Du(\mathbf{x}).\mathbf{z} + \frac{h}{2} D^2 u(\mathbf{x})(\mathbf{z},\mathbf{z}))d\mathbf{z} - \frac{1}{6} h^{\frac{3}{2}} \int_{\mathbb{R}^N} g(\mathbf{z}) D^3 u(\mathbf{x} - h^{\frac{1}{2}}\theta\mathbf{z})(\mathbf{z},\mathbf{z},\mathbf{z})d\mathbf{z},$$

where $\theta = \theta(\mathbf{x},\mathbf{z},h)$ belongs to $[0,1]$. Using the moment information (3.2-3.3) and the bound (3.5), we obtain

$$|(g_h * u)(\mathbf{x}) - u(\mathbf{x}) - h\frac{\sigma}{2}\Delta u(\mathbf{x})| \leq C h^{\frac{3}{2}} \max_{\mathbf{x}\in C} ||D^3 u(\mathbf{x})||,$$

where C is defined in (3.5). □

The former theorem shows a direct relation between the convolution with a smoothing kernel and the heat equation. This link will be completed by the next theorem : It essentially states that if we set $(T_h u_0) = g_h * u_0$, then iterates of T_h, $((T_h)^n u_0)(\mathbf{x})$, tend to $u(t,\mathbf{x})$, where $u(t,\mathbf{x})$ is solution of the heat equation.

Theorem 3.3 *Let $g(x) \in L^1(\mathbb{R}^N)$ be a radial or pseudo-radial nonnegative function satisfying the moment conditions (3.1 to 3.3), (3.5) and set $g_h(\mathbf{x}) = \frac{1}{h^{\frac{N}{2}}} g(\frac{\mathbf{x}}{h^{\frac{1}{2}}})$ and $T_h u = g_h * u$. Then*

$$((T_h)^n u_0)(\mathbf{x}) \to u(t,\mathbf{x}) \text{ in } L^1(C) \text{ as } n \to +\infty,\ nh \to t, \tag{3.7}$$

where $u(t,\mathbf{x}) = G_t * u_0 \in L^\infty_C$ is the solution of the heat equation (2.13) with initial value u_0,

$$\frac{\partial u}{\partial t} = \frac{\sigma}{2}\Delta u \text{ and } \int_C |u(t,\mathbf{x}) - u_0(\mathbf{x})|d\mathbf{x} \to 0 \text{ as } t \to 0.$$

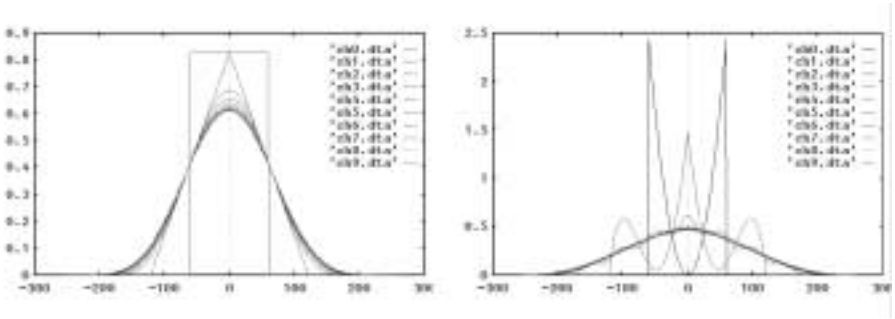

Figure 3.2: Iterated linear smoothing converges towards the heat equation. In this experiment on one-dimensional functions, it can be appreciated how fast an iterated convolution of a positive kernel converges to a gaussian. On the left, we display nine iterations of the convolution of the characteristic function of an interval with itself, with the adequate rescaling. On the right, the same experiment is repeated with a by far more irregular kernel. Convergence to the gaussian is almost equally fast. (See Theorem 3.3).

The preceding theorem means that from the asymptotic, scale independent viewpoint, all local linear isotropic "low pass" iterated filters are equivalent to the heat equation (and to the convolution with Gaussian kernels with increasing width).

Proof Without loss of generality (see Exercise 3.1), we take $\sigma = 2$. In this proof, $O(h^\alpha)$ denotes any function of t and $\mathbf{x}$ such that $|O(h^\alpha)| \leq Ch^\alpha$ where C does not depend upon $\mathbf{x}$ and $t \in [t_1, t_2]$, a fixed interval of $]0, +\infty[$. We shall write $T_h u(t, \mathbf{x})$ for what would be more properly written $(T_h(u(t)))(\mathbf{x})$, where $u(t)$ denotes the function $\mathbf{x} \to u(t, \mathbf{x})$. In the same way, we shall write $T_h^2 u(t, \mathbf{x})$ for $(T_h^2(u(t)))(\mathbf{x})$.

Using Formula (3.6) in Theorem 3.2, we have

$$T_h u(t, \mathbf{x}) - u(t, \mathbf{x}) = g_h * u(t, \mathbf{x}) - u(t, \mathbf{x}) = h\Delta u(t, \mathbf{x}) + O(h^{\frac{3}{2}}), \tag{3.8}$$

because $||D^3 u(t, \mathbf{x})||$ can be bounded independently of $(t, \mathbf{x})$ on the compact set $[t_1, t_2] \times C$. Since $u(t, \mathbf{x})$ is a solution of the heat equation, we also have

$$u(t + h, \mathbf{x}) - u(t, \mathbf{x}) = h\Delta u(t, \mathbf{x}) + O(h^2), \tag{3.9}$$

and, since u is C^∞, the behavior of $O(h^2)$ is again uniform on the compact set $[t_1, t_2] \times C$. By substracting (3.9) from (3.8), we obtain

$$T_h u(t, \mathbf{x}) - u(t + h, \mathbf{x}) = O(h^{\frac{3}{2}}). \tag{3.10}$$

We notice that if $u \leq C$, then $T_h u \leq C$. As a consequence, $T_h(O(h^\alpha))$ also is an $O(h^\alpha)$, uniformly in $t \in [t_1, t_2[, \mathbf{x} \in I\!R^N$. Thus, applying T_h to both sides of (3.10), we obtain

$$(T_h)^2 u(t, \mathbf{x}) - T_h u(t + h, \mathbf{x}) = O(h^{\frac{3}{2}}). \tag{3.11}$$

We can use Relation (3.10) with $t + h$ instead of t and we then get

$$T_h u(t + h, \mathbf{x}) - u(t + 2h, \mathbf{x}) = O(h^{\frac{3}{2}}). \tag{3.12}$$

By adding (3.11) to (3.12) we obtain

$$T_h^2 u(t, \mathbf{x}) - u(t + 2h, \mathbf{x}) = 2O(h^{\frac{3}{2}}). \tag{3.13}$$

We can iterate this process and get

$$T_h^n u(t,\mathbf{x}) - u(t+nh,\mathbf{x}) = nO(h^{\frac{3}{2}}), \tag{3.14}$$

provided $t_1 \leq t+nh \leq t_2$. Letting $n \to \infty$ and setting $h = \frac{\tau}{n}$ yields

$$T_h^n u(t,\mathbf{x}) - u(t+\tau,\mathbf{x}) = O((\frac{\tau}{n})^{\frac{1}{2}}), \tag{3.15}$$

provided $t_1 \leq t+\tau \leq t_2$. This would end the proof, could we take $t = t_1 = 0$. Now, we can't, but we can have it very small. Fix $\varepsilon > 0$ and t_1 small enough to have

$$||u(t_1) - u_0||_{L^1(C)} = \int_C |u(t_1,\mathbf{x}) - u_0(\mathbf{x})| d\mathbf{x} < \varepsilon. \tag{3.16}$$

By Proposition 2.7, we have for every $u \in L^1_C$

$$||g_h * u||_{L^1(C)} \leq ||g_h||_{L^1(I\!R^N)} ||u||_{L^1(C)}.$$

Since $\int g_h = 1$, we deduce from (3.16) and this relation applied to $u = u(t_1) - u_0$ that

$$||T_h^n u(t_1,.) - T_h^n u_0||_{L^1(C)} \leq \varepsilon. \tag{3.17}$$

By integrating Relation (3.15) on C with $t = t_1$, we also have

$$||T_h^n u(t_1,.) - u(t_1+\tau,.)||_{L^1(C)} = O((\frac{\tau}{n})^{\frac{1}{2}}) < \varepsilon \tag{3.18}$$

for n large enough. Combining (3.17) and (3.18) yields

$$||T_h^n u_0 - u(t_1+\tau)||_{L^1(C)} < \varepsilon$$

for t_1 small enough and $hn = \tau$. By (??), we conclude that

$$||T_h^n u_0 - u(\tau)||_{L^1(C)} \leq 2\varepsilon,$$

for $nh = \tau$ and n large enough, which concludes the proof. □

3.2 Directional averages and directional heat equations

The aim of this section is to show how to approximate directional second derivatives by using local averages. We consider a C^2 function u from $I\!R^N$ into $I\!R$, and a vector $\mathbf{z}$ of $I\!R^N$ with norm 1. We can apply the Taylor formula around a point $\mathbf{x}$ by considering a perturbation of $\mathbf{x}$, $\mathbf{x}+h\mathbf{z}$:

$$u(\mathbf{x}+h\mathbf{z}) = u(\mathbf{x}) + hDu(\mathbf{x}).\mathbf{z} + \frac{h^2}{2} D^2u(\mathbf{x})(\mathbf{z},h\mathbf{z}) + o(h). \tag{3.19}$$

We define $T_h^{\mathbf{z}}$, the operator which computes the mean value of u on a straight segment oriented by $\mathbf{z}$,

$$T_h^{\mathbf{z}} u(\mathbf{x}) = \frac{1}{2h} \int_{h\in[-1,1]} u(\mathbf{x}+h\mathbf{z}) dh$$

Integrating (3.19) for $h \in [-1,1]$, we obtain.

Proposition 3.4

$$T_h^{\mathbf{z}} u(\mathbf{x}) - u(\mathbf{x}) = \frac{1}{3} h^2 D^2 u(\mathbf{x})(\mathbf{z}, \mathbf{z}) + o(h^2)$$

As a consequence, the iteration of the operator $T_h^{\mathbf{z}}$ intuitively corresponds to a directional heat equation, in the direction of $\mathbf{z}$,

$$\frac{\partial u}{\partial t}(t, \mathbf{x}) = \frac{\sigma}{2} D^2 u(\mathbf{x})(\mathbf{z}, \mathbf{z}). \tag{3.20}$$

Notice that the directional averages, and the preceding equation, treat u independently on each line paralel to $\mathbf{z}$. Exercise 3.2 formalizes this.

Exercise 3.2 *Let, in arbitrary dimension N, $u_0 \in L^1(C)$ be an initial datum. Consider the function $g(s) = \frac{1}{2} \mathbb{1}_{[-1,1]}(s)$.*

(i) Show that it satisfies the the moment conditions (3.1 to 3.3), (3.5). Compute the variance σ of g.

(ii) We set $g_h(s) = \frac{1}{h^{\frac{1}{2}}} g(\frac{s}{h^{\frac{1}{2}}})$ and $T_h u(\mathbf{x}) = \int_I Ru(\mathbf{x} + s\mathbf{z}) g_h(s) ds$. By applying, in dimension $N = 1$, Theorem 3.3, show that for any function u_0 in L^1_C,

$$((T_h)^n u_0)(\mathbf{x}) \to u(t, \mathbf{x}) \text{ in } L^1(C) \text{ as } n \to +\infty,\ nh \to t, \tag{3.21}$$

where $u(t, \mathbf{x})$ is the solution of the directional heat equation (3.20) with initial value u_0,

If we tune the vector $\mathbf{z}$ adequately at each point $\mathbf{z}$, the direction diffusion depending upon u and $\mathbf{x}$, the operator is then no more linear but the same asymptotic analysis applies at each point. The main choices we shall consider in this book are $\mathbf{z} = \frac{Du}{|Du|}$ and $\mathbf{z} = \frac{Du^\perp}{|Du|}$, where $Du = (u_x, u_y)$ and $Du^\perp = (-u_y, u_x)$. We then get by applying Proposition 3.4 the following asymptotic behaviours :

- Average in the direction of the gradient We choose $\mathbf{z}(\mathbf{x}) = \frac{Du(\mathbf{x})}{|Du(\mathbf{x})|}$, so that

$$\frac{1}{|Du|^2} Du(Du, Du) = 3 \lim_{h \to 0} T_h^{Du/|Du|} u.$$

  We shall interpret in the next section this differential operator as an "edge detector": the Canny operator.

- Average in a direction orthogonal to the gradient We choose $\mathbf{z}(\mathbf{x}) = \frac{Du(\mathbf{x})^\perp}{|Du(\mathbf{x})|}$, so that

$$\frac{1}{|Du|^2} Du(Du^\perp, Du^\perp) = 3 \lim_{h \to 0} T_h^{Du/|Du|} u.$$

  This last operator will arise as the second term of the curvature motion.

Exercise 3.3 *We consider a C^2 function u, from $\mathbb{R}^2$ into $\mathbb{R}$. Prove that there $\exists C > 0$, such that*

$$D^2 u(Du, Du)(\mathbf{x}) = C \lim_{h \to 0} \frac{1}{2h^2} (\min_{\mathbf{y} \in B(\mathbf{x}, h)} u(\mathbf{y}) + \max_{\mathbf{y} \in B(\mathbf{x}, h)} u(\mathbf{y}) - 2u(\mathbf{x})).$$

3.3 Edge detection and linear scale space

3.3.1 The edge detection doctrine

One of the uses of the linear theory is, in dimension 2, "edge detection". The assumption of the edge detection doctrine is that relevant information in an image is contained in the trace left in the image by the apparent contours of physical objects which have been photographed. If an object with some constant color, say, black, is photographed on a bright background, then it is expected that its silhouette in the image is a closed curve across which the light intensity $u_0(\mathbf{x})$ varies strongly. Let us call this curve an "edge". The local detection of an edge can *a priori* be done by computing the gradient $Du_0(\mathbf{x})$. This gradient should have a large intensity $|Du_0(\mathbf{x})|$ and a direction $\frac{Du_0(\mathbf{x})}{|Du_0(\mathbf{x})|}$ which indicates the direction normal to the silhouette curve. It therefore looks sound to simply compute the gradient of u_0 and choose the points where this gradient is large. This conclusion is a bit irrealistic for two reasons.
a) The points where the gradient is larger than a given threshold are likely to form regions, and not curves.
b) Many points may have a large gradient because of tiny oscillations of the image, not related to the real objects. In fact, the digital images being always noisy, there is no reason to assume the existence or computability of any gradient at all.
Objection b) is solved by defining a smoothing process : we associate with the image a smoothed version $u(t)$, depending of course upon a scale parameter t measuring the amount of smoothing. In the classical linear doctrine, this smoothing is made by convolving the image with gaussians of increasing widths.
Objection a) is solved by defining edge points not as points where the gradient is large only, but as points where some maximality property of the gradient is observed. Let us take an example in dimension one. Let $u(x)$ be a C^2 real function on $\mathbb{R}$ and consider points where $|u'(x)|$ is maximal. At these points, the second derivative $u''(x)$ changes sign because extrema of the gradient correspond to a change from concave to convex or conversely. Thus we can look for the "edge points" of the smooth signal among the points where $u''(x)$ crosses zero. Generalizing this in dimension 2 leads to the Hildreth-Marr edge detection theory, or alternatively to the Canny edge detection theory. The only difference is that Hildreth and Marr replace, in dimension 2, $u''(x)$ by $\Delta u(\mathbf{x}) = \frac{\partial^2 u}{\partial x^2}(\mathbf{x}) + \frac{\partial^2 u}{\partial y^2}(\mathbf{x})$, which is the only isotropic linear differential operator of order 2 generalizing u''.

Canny ([71]) instead, gives up the linearity and defines edge points as edges where the gradient is maximal in the direction of gradient. In other terms, an edge point satisfies $g'(0) = 0$, where $g(t) = |Du|(\mathbf{x} + t\frac{Du}{|Du|})$. This implies $D^2u(\frac{Du}{|Du|}, \frac{Du}{|Du|}) = 0$.

Exercise 3.4 *Compute $g'(0)$ and check the mentionned implication.*

So the algorithms of Hildreth-Marr and Canny are as follows.

Edge detection algorithm 1 (Hildreth-Marr)
• Convolve u_0 with gaussians G_t of increasing widths. We obtain a multiscale image $u(t, \mathbf{x})$.
• At each scale t compute all points where $Du \neq 0$ and Δu changes sign. Such points are called "zero-crossings of the Laplacian", or shortly "zero-crossings".
• (Optional) Eliminate the zero-crossings at which the gradient is small.

In practice, the found "edges" are displayed for a series of dyadic scales, $t = 2$, 4, 8, 16, etc.

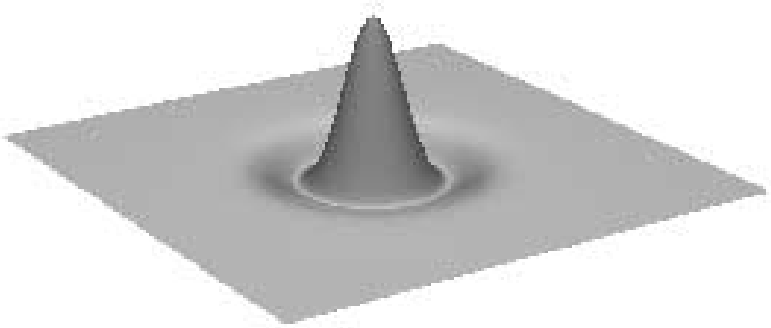

Figure 3.3: Three-dimensional representation of the laplacian of the Gauss kernel. This convolution kernel, actually a wavelet, is used to estimate the laplacian of the image at different scales of linear smoothing. (See Proposition 2.9, and Section 3.3).

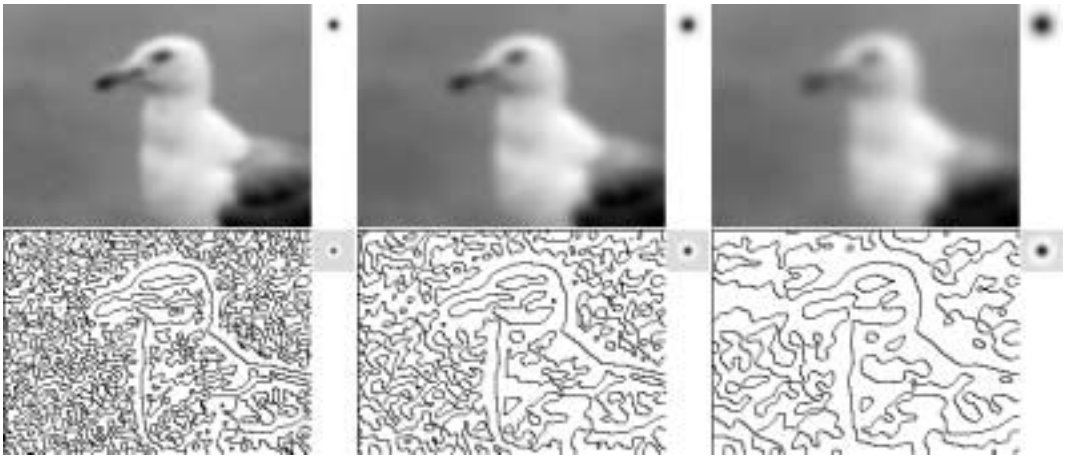

Figure 3.4: "Zero-crossings of the laplacian" at different scales. In this display, we illustrate the original Scale Space theory, as it is for instance developed in the founding book by David Marr, Vision. In order to extract more global structure, the image is convolved with gaussians whose variances are powers of 2. One computes the laplacian of the resulting smooth images and displays the lines along which this laplacian changes sign : the so called "zero-crossings of the laplacian". According to David Marr, those zero-crossings represent the "raw primal sketch" of the image, that is, the information on which further vision algorithms should be based.
Above : from left to right, we display the results of the smoothing, and the associated gaussian kernels, of scales 1,2 and 4 respectively. Below : we display the zero-crossings of the laplacian, and the corresponding kernels, i.e. the laplacians of the gaussians used above.

Edge detection algorithm 2 (Canny's edge detector) • Here again, convolve u_0 with gaussians G_t of increasing widths. We obtain a multiscale image $u(t, \mathbf{x})$.
• At each scale t, find all points $\mathbf{x}$ where $Du(\mathbf{x}) \neq 0$ and $D^2u(\frac{Du}{|Du|}, \frac{Du}{|Du|})(\mathbf{x})$ crosses zero. At such points, the function $t \to u(\mathbf{x} + t\frac{Du}{|Du|})$ changes, when t crosses 0, from concave to convex or conversely.
• At each scale t : fix a threshold $\theta(t)$ and retain as "edge points at scale t" only the points which satisfy the preceding conditions and in addition $|Du(\mathbf{x})| > \theta(t)$.

Figure 3.5: Zero crossings of the laplacian of a synthetic image. From left to right : the original image, the image linearly smoothed by a convolution with a Gauss function, the sign of the laplacian of the filtered image (the gray color corresponds to values close to 0, black to clearcut negative values, white to clearcut positive values) and the zero-crossings of the Laplacian. This experiment clearly shows the drawbacks of the Laplacian as edge detector.

3.3.2 Discussion and objections

The Canny edge detector is generally preferred for its accuracy to the Marr-Hildreth theory. Their use and characteristics are, however, essentially the same. There are many variants and tentative improvements to the edge detection theory. Now, the discussion which follows adapts easily to the variants. The first thing to be noticed is that, thanks to Theorem 2.11 in the former section, we know that $u(t, \mathbf{x}) = G_t * u_0$ is a C^∞ function provided u_0 is bounded. Thus we can indeed compute second order differential operators of $u(t, \mathbf{x})$ for $t > 0$. In the case of linear operators like the Laplacian or the gradient, the task is facilitated by the formula proved in Proposition 2.9. We have $\Delta u(t, \mathbf{x}) = \Delta(G_t * u_0) = (\Delta G_t) * u_0$, where (in dimension 2)

$$\Delta G_t(\mathbf{x}) = \frac{|\mathbf{x}|^2 - 4t}{16\pi t^3} e^{-\frac{|\mathbf{x}|^2}{4t}}.$$

In the same way, the Canny edge detector makes sense because u is C^∞ at all points where $Du(\mathbf{x}) \neq 0$. Such points cannot be edge points.

Thus, we can doubtless compute edge points thanks to those filters. Let us now list the drawbacks of such methods. They are in fact well explained in the *Scale Space* theory developed by Witkin and Koenderink.

• **The Scale Space theory.** A first severe problem is the addition of an extra dimension : the scale t in computations and image understanding. We get no absolute definition of edges. We only can talk about "edges at a certain scale". An answer to this problem would be to try to track the edges across scales. Indeed, as is noticeable in experiments, the "main edges" resist a convolution with a wide filter, but loose much of their spatial accuracy. On the opposite side, if one makes a sharp low filtering, with a gaussian with small variance, these edges keep their correct location. Now, the "main" edges are then lost in a crowd of "spurious" edges due to noise, texture, etc... The "scale space" theory of Witkin [440] proposes therefore to identify the main edges at a low scale, and then to "follow them backward" by making the scale decrease again. This method could in theory give the exact location of all "main edges". However, its implementation is rather heavy from the computational viewpoint and unstable, because of the following-up of edges across scales and the multiple thresholdings involved in the edge detection at

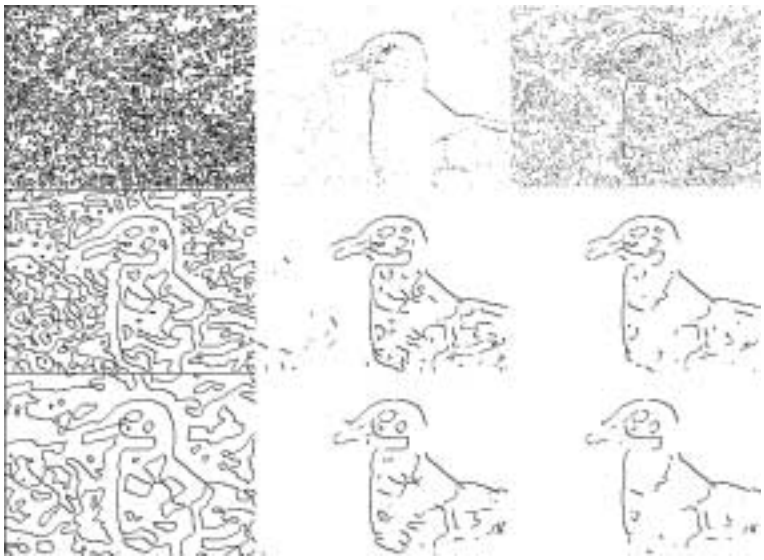

Figure 3.6: Canny's edge detector. These images illustrate the Canny edge detector, which attempts to find boundaries in an image. The Canny filter involves the following operations :

1. Convolution of the image by a gaussian kernel (from top to bottom, scale = 0.1, 0.5, 1.0).
2. Estimation of $D^2u(Du, Du)$ (we use here a finite difference scheme).
3. Convolution of $D^2u(Du, Du)$ with a small Gauss kernel (sigma = 0.001).
4. Thresholding of the gradient of the result of Step 1.
5. Zero-crossings of the result of Step 3, only displayed when the threshold of Step 4 is achieved.

Left Column : result of the Canny filter without the threshold on the gradient, (Step 4 removed)
Middle column : result with a visually " optimal" scale and an image depending threshold. (from top to bottom : 15, 0.5, 0.6)
Right column : result with a fixed gradient threshold equal to 0.5.
Notice that this edge detection "theory" depends upon not less than three parameters which have to be fixed by the user.

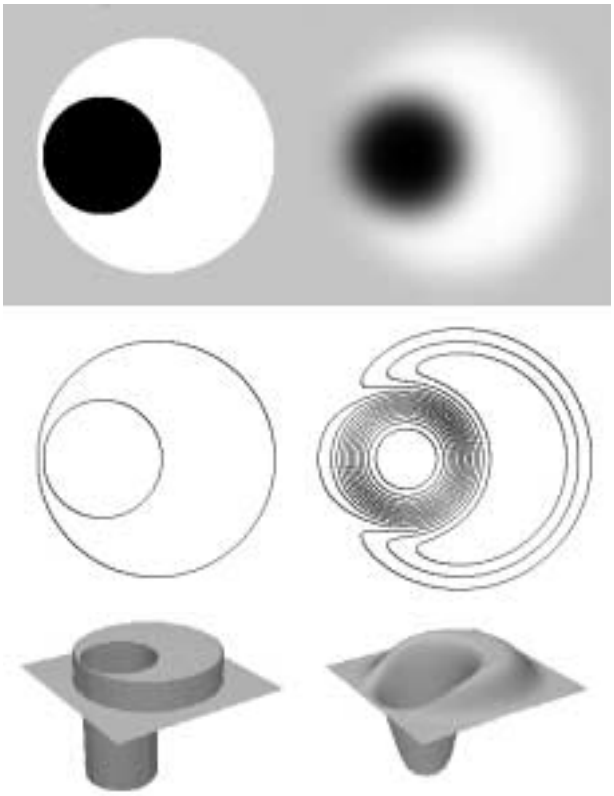

Figure 3.7: Violation of the inclusion by the linear scale-space. The linear scale-space does not maintain the inclusion between objects. Top-left: an image that contains a black disk enclosed by a white disk. At a certain scale, the black and white circles mix together (top-right). The level lines (middle) show that the inclusion is no longer preserved. Bottom : three-dimensional representation of both images, where the vertical coordinate corresponds to the gray-level. For the shape inclusion principle, see Chapter 22.

each scale.

The tracking of edges across scales excludes any thresholding of the gradient. Indeed, such a thresholding may remove edges at certain scales and not at other ones. So one should trace across scales all "zero-crossings", without consideration to their likeliness to be edges or not. This makes the edge matching across scales very difficult : indeed, the experimental sparseness of zero-crossings associated with sharp edges is no more true for zero-crossings at large. Small fluctuations of the image may generate "spurious" zero-crossing edges. The same is equally true for Canny's detector, as is again quite obvious in experiments.

In conclusion, the "edge detection theory" is more an attempt than an established theory. After more than 30 years of existence, it has also become clear that no robust technology can be based on it. Since the edge detection devices depend upon multiple thresholdings on the gradient, followed by "filling the holes" algorithms, there can be no scientific agreement on the identity of "edge points" in a given image.

Contrast invariance

As we shall see in the next chapters, the assumption of contrast invariance for image operators will solve most of the technical problems associated with linear operators. We say that an image operator $u \to Tu$ is contrast invariant if it commutes with all increasing functions g, i.e.,

$$g(Tu) = T(g(u)) \tag{3.22}$$

In image processing, most images are known up to a contrast change because we ignore both the lightning conditions and the nonlinear response of the sensors. The commutation relation (3.22) ensures that the filtered image $Tu = g^{-1}(T(g(u)))$ does not depend upon g. Convolutions *are not* contrast invariant, i.e. we have in general

$$g(k * u) \neq k * g(u).$$

Exercise 3.5 *Construct two simple functions u and g such that the above commutation indeed fails.*

In the same way, the operator $T_t : u_0 \to u(t)$ associated with the heat equation *is not* contrast invariant : one has

$$\frac{\partial(g(u))}{\partial t} = g'(u)\frac{\partial u}{\partial t}$$

and

$$\Delta(g(u)) = g'(u)\Delta u + g''(u)|Du|^2$$

if g is C^2. Thus $g(u)$ is not necessarily a solution of the heat equation if u is. (See Figure 20.5).

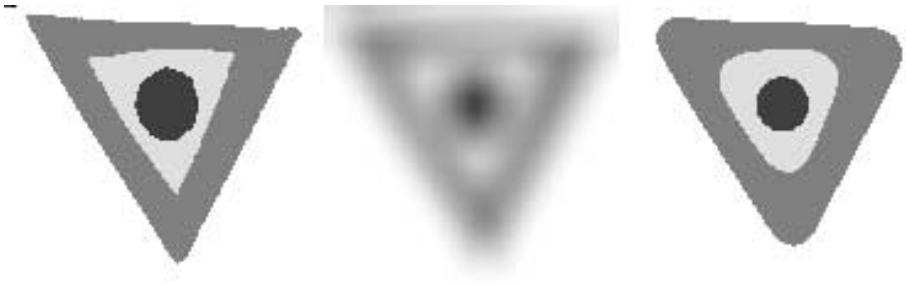

Figure 3.8: The heat equation creates structure. This experiment shows that the linear scale-space can create structure, i.e. increase the complexity of the image. On the left : original synthetic image, three grey levels, two black regional minima, one grey regional maximum, one white regional maximum. Middle : result when applying the heat equation : the grey regional minimum is split into three regional minima. Compare with the application of a contrast invariant local filter, the iterated median filter introduced in Chapter 10.

3.4 Dynamic Shape

In [252], Koenderink and Van Dorn define a "shape" in $\mathbb{R}^N$, a closed subset X of $\mathbb{R}^N$. They propose to smooth the shape by applying directly the heat equation $\frac{\partial u}{\partial t} - \Delta u = 0$ to the characteristic function $\mathbb{1}_X$ of X. (We set $\mathbb{1}_X(\mathbf{x}) = 1$ if $\mathbf{x} \in X$, $= 0$ otherwise). Of course, the solution $G_t * \mathbb{1}_X$ is no more a characteristic function. The authors define the evolved shape at scale t by

$$X_t = \{\mathbf{x}, u(t, \mathbf{x}) \geq \frac{1}{2}\}$$

The value $\frac{1}{2}$ is chosen by an obvious requirement. Let $X = \{(x, y) \in I\!R^2, x \geq 0\}$, then we ask that

$$X_t = \{G_t * X \geq \lambda\} = X$$

which is only true if $\lambda = \frac{1}{2}$.

Figure 3.9: Nonlocal behaviour of shapes with the Dynamic Shape method. This image displays the smoothing of two irregular shapes by the Dynamic Shape method (Koenderink and Van Dorn). Top left : initial image, made of two irregular shapes. Top to bottom and left to right: smoothing with increasing scales. Notice how, the convolution being made with gaussians of increasing variance, the shapes merge more and more. We do not have a separate analyis of both shapes but a "joint analysis" of both. This joint analysis depends of course a lot upon the initial distance between both shapes.

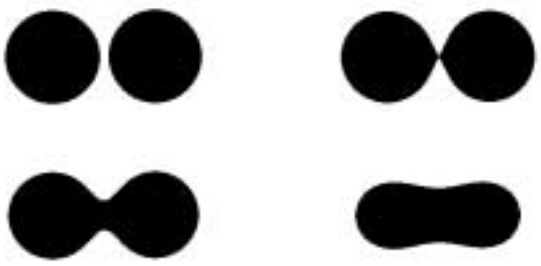

Figure 3.10: Non local interactions in the Dynamic Shape method. Two close disks (top-left) interact as scale grows and create a qualitatively different and new shape. The change of topology, at the scale where both shapes merge into one, entails the appearance of a singularity (a cusp) on the shape(s) boundaries. Top right : Dynamic Shape evolution at the small critical scale where the cusp appears. Bottom : further evolution when scale grows : creation of a new global shape.

Let us mention some drawbacks of this shape evolution.

The non-local interactions: Take two close disks $D(\mathbf{x}_0, 1)$ and $D(\mathbf{x}_1, 1)$, with $|\mathbf{x}_0 - \mathbf{x}_1| = 1 + \epsilon$. Then the evolution of the union of both disks, considered as a single shape, is quite different from the evolution of each disk separately, (see figures 3.9 and 3.10).

Creation of singularities. As another consequence, singularities of the orientation and curvature of the boundary of the shape may appear with the evolution. Thus, the smoothing creates new salient features!

3.5 Curve evolution by the heat equation.

Let us go back to shapes whose boundary can be described by a finite set of rectifiable Jordan curves. We call rectifiable Jordan curve a simple closed curve of the plane, i.e. a curve without self-crossings and with finite length. We also assume that these Jordan curves do not meet, so that the Jordan curve structure is uniquely defined. Let us then focus on the smoothing of a single Jordan curve, with finite length.

We parameterize the initial curve by length

$$s \in [0, L] \to \mathbf{x}_0(s),$$

where s is the curve length between $\mathbf{x}(0)$ and $\mathbf{x}(s)$. We have $\mathbf{x}_0(s) = (x(s), y(s)) \in I\!R^2$. A first obvious idea in order to smooth the curve is to convolve $x(s)$ and $y(s)$ by a smoothing kernel. The kernel must be as local as possible in order to ensure the analysis of local curve features. Thus, we are naturally led to iterate a smoothing kernel M_h, with $h \to 0$. By Theorem 3.3, we know that all such processes boil down to the application of the heat equation.

$$\frac{\partial \mathbf{x}}{\partial t}(t,s) = \frac{\partial^2 \mathbf{x}}{\partial s^2}(t,s), \qquad s \in [0, L], \quad t \in [0, +\infty[. \tag{3.23}$$
$$\mathbf{x}(0,s) = \mathbf{x}_0(s)$$

Here, we must notice two very important facts which advocate against the method.

1. When $t > 0$, s is no more a length parameter of the evolved curve $\mathbf{x}(t)$.
2. Maximum principle holds for $x(t,s)$ and $y(t,s)$ as scalar solutions of the heat equation. In addition, $x(t,s)$ and $y(t,s)$ are, as proved in chapter 2, C^∞ functions of (t,s), for $t > 0$. **Now, this does not imply that the curve $\mathbf{x}(t,s)$ behaves smoothly!** In fact, it can easily be seen (see Figure 3.5) that a smooth curve may generate by this evolution self-crossings which by further smoothing entail the appearance of singularities.

Exercise 3.6 *Construct a C^∞ map : $\sigma \in [0,1] \to \mathbf{x}(\sigma) \in I\!R^2$ such that the curve $\{\mathbf{x}(\sigma),\ \sigma \in [0,1]\}$ is a square. Deduce that a curve can have a C^∞ parameterization without being smooth.*

3.6 How to restore locality and causality ?

3.6.1 Localization of the "Dynamic Shape" method.

The main objective of this treatise is to redefine the preceding smoothing processes in such a way that they are local and do not create new singularities. This can be done, as we shall prove further on, by *alternating* a small scale linear convolution with a natural normalization process.

In the case of the Dynamic Shape analysis, we define an alternate Dynamic Shape algorithm in the following way.

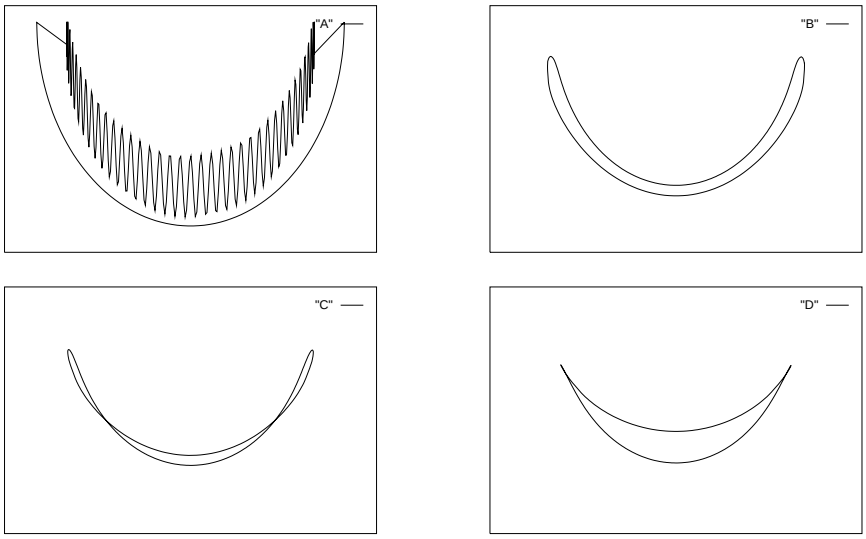

Figure 3.11: Curve evolution by the heat equation. The coordinates of the curves are parameterized by the arc length, and then smoothed as real functions of the length, by applying to them the heat equation. From A to D, the coordinates are smoothed with an increasing scale. Each coordinate function therefore is C^∞ ; the evolving curve can, all the same, generate self-crossings (C) or singularities (D).

Algorithm. Iterated local Dynamic Shape algorithm, or Merriman-Bence-Osher algorithm.

1. Convolve the initial shape $\mathbb{1}_{X_0}$ with G_h, h small.
2. Define $X_1 = \{\mathbf{x}, G_h * \mathbb{1}_{X_0}(\mathbf{x}) \geq \frac{1}{2}\}$.
3. Set $X_0 = X_1$ and go back to 1.

In fact the Dynamic Shape method is a *median filter* (see Chapter 10) and the preceding Merriman-Bence-Osher algorithm an *iterated median filter*. We can anticipate that the application of an alternate median filter yields, as $h \to 0$ and the number of iterations tends to infinity, a "motion by mean curvature" which will be defined in Chapter 6 as one of the main objects of Chapters 11 and following.

3.6.2 Renormalized heat equation for curves.

In [272] Mackworth-Mokhtarian noticed the loss of causality of the heat equations applied to curves. Their solution, at least formally, looks like the solution given to the nonlocality of the Dynamic Shape method. Instead of applying the heat equation for long times (or, equivalently, to convolve the curve $\mathbf{x}(s)$ with Gaussians $G_t(s)$ of arbitrary width, they do the following.

Algorithm : Renormalized heat equation for curves.

1. Convolve the initial curve $\mathbf{x}_0(s_0)$, parameterized by its length parameter $s_0 \in [0, L_0]$, with a Gaussian G_h ; h is small.

Figure 3.12: The Bence-Merriman-Osher shape smoothing method is a localized and iterated version of the Dynamic Shape method. A convolution of the binary image with small-sized gaussians is alternated with mid-level thresholding. It uses (top, left) as initial data the same shapes as in Figure 3.4. From top to bottom and left to right: smoothing with increasing scales. Notice that the shapes keep separate. In fact, their is no interaction between their evolutions. Each one evolves as it would do alone.

2. Let L_n be the length of the curve $\mathbf{x}_n$ obtained after n iterations and $s_n \in [0, L_n]$ its arc length parameter. Set, for $n \geq 1$, $\tilde{x}_{n+1}(s_n) = (G_h * \mathbf{x}_n)(s_n)$. Then reparameterize $\tilde{x}_{n+1}(s_n)$ by its arc length parameter $s_{n+1} \in [0, L_{n+1}]$, which yields a curve $\mathbf{x}_{n+1}(s_{n+1})$.

3. Iterate.

Theorem 3.5 *Let $\mathbf{x}_0(s_0)$ be a C^2 curve parameterized by length. When h tends to zero, there is a constant c such that*

$$(G_h * \mathbf{x}_0)(s_0) - \mathbf{x}_0(s_0) = ch^2 \frac{\partial^2 \mathbf{x}_0}{\partial s_0^2} + o(h^2) \tag{3.24}$$

Proof Direct application of Theorem 3.2. □

Exercise 3.7 *Compute the constant c of Theorem 3.5.*

We notice that (3.24) is consistent with the following evolution equation(as $h \to zero$):

$$\frac{\partial \mathbf{x}}{\partial t} = \frac{\partial^2 \mathbf{x}}{\partial s^2} (= Curv(\mathbf{x})(s)) \tag{3.25}$$

This equation is not the heat equation (3.23). Indeed, in (3.25), we assume that s denotes the length parameter of evolved curve $\mathbf{x}(t)$ at time t (and not the initial parameterization). We shall consider this last, nonlinear, curve evolution in Chapter 6.

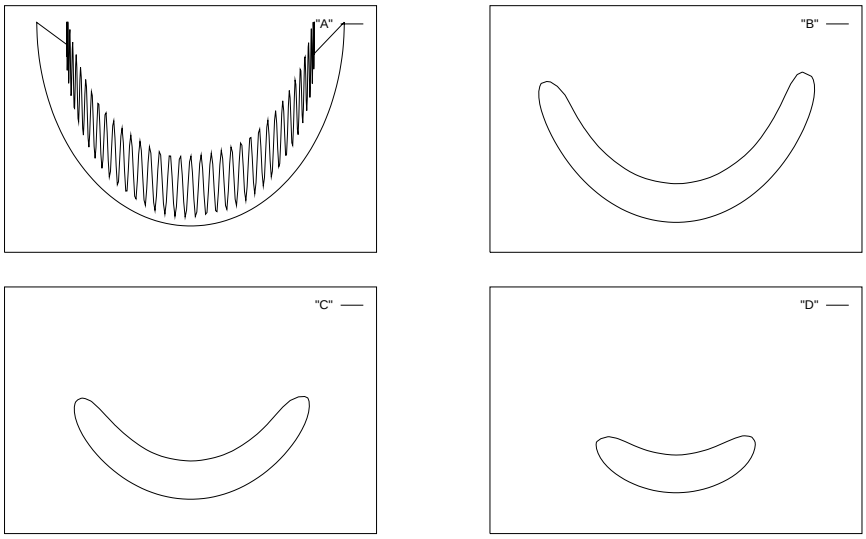

Figure 3.13: Curve evolution by the " renormalized heat equation" (Mackworth-Mokhtarian). At each smoothing step, the coordinates of the curves are reparameterized by the arc length of the smoothed curve. From A to D, the curve is smoothed with an increasing scale. Note that, in contrast with the linear heat equation (Figure 3.5), the evolving curve shows no singularities and does not cross itself.

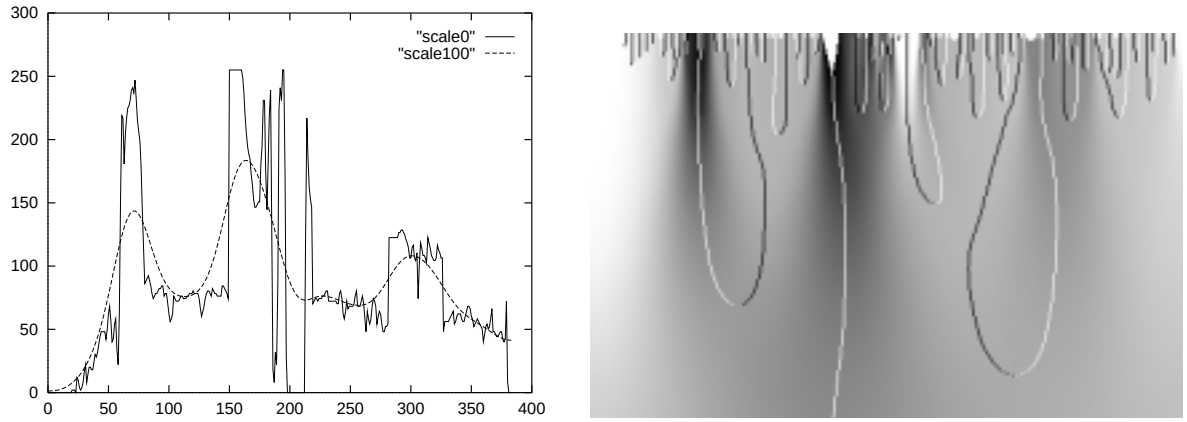

Figure 3.14: Multiscale histogram modes by using the linear scale space (heat equation). Histograms are strongly oscillatory functions, presenting many peaks due to flat image regions and quantization effects. We call *modes*of the histogram all the intervals whose endpoints are two successive local minima. We call central point of the mode the unique maximal point between them. In order to get a more and more global analysis of the histogram, the heat equation may be applied. We then talk about "modes at scale t". On the left, we display the original histogram (scale 0) and the histogram at scale 100, corresponding to a convolution with the gaussian of standard deviation 50. Only three modes are left. On the right, we display the so-called "fingerprints" of the histogram, a scale-space representation where scale increases downwards. The white curves display the minima of the smoothed histogram and the black curves the maxima of the histogram. By a classical property of the heat equation, maxima and minima collapse by pairs, thus yielding the observed "fingerprint" organization of multiscale modes : each mode at scale t can be followed upwards to the fine scales and contains more and more submodes. Modes are therefore organized in a tree. No new mode is ever created when the scale increases. This is an instance where the Witkin linear scale space works nicely !

References.

Section 3.1 : To make local averages of the grey level in order to restore an image is one of the first tools proposed in early image processing in the sixties [176], [203]. The classical remark that the difference between an image and its local average is proportional to its laplacian is one of the most fruitful formulae in image processing. Gabor (quoted in [266]) used it as an image restoration tool, as we pointed out in the introduction. See also Hummel [218], [219]. As we commented in the introduction, Burt and Adelson [69] based on this simple remark their Laplacian pyramid representation, one of the ancestors to the theory of wavelets. Our proof that iterated convolutions of radial kernels converge to a gaussian is a version of the De Moivre-Laplace formula (Central Limit Theorem [63]) adapted to Image Processing. Its relevance for image processing is high, since it leads to recommend gaussian filtering as the unavoidable linear smoother. We shall prove this fact from the "scale space" point of view in Section 20.5.

Section 3.2 : Directional diffusions have a long story since Hubel and Wiesel showed the existence of direction sensitive cells in the visual areas of the neocortex. There has been a blow-up in publications on directional linear filters since (e.g.) the seminal paper of Daugman [**?**]

Section 3.3 : The use of gaussian filtering is so widespread in image analysis that it is hard to find a "first paper". The Hildreth-Marr [286] paper on edge detection is seminal as well as the famous David Marr book, Vision [287]. Now, the term "edge detection" in TV image transmission is present as early as 1959 [231]. The idea that the computation of derivatives in an image necessitates a previous smoothing has been strongly developped in the Dutch Image Analysis school ([158], [59]. See also the books of Florack [153], Romeny [403] and Lindeberg [265]. See also [143]. Now, the Canny edge detector [71] is probably the most famous image analysis operator, and is based on a second derivative. Deriche [119] derived a recursive implementation. The general image analysis framework by which an image is associated its smoothed version at several scales is called "scale space". The term originates in Witkin [440], 1983, who proposed to track the zero-crossings of the Laplacian of the smoothed image accross scales. Yuille and Poggio [447], proved that this tracking works for 1-D signals. Hummel [215], [221] and Yuille and Poggio [448] analysed the conjectures of Marr and Witkin according to which the image is completely recoverable from its edges at different scales. A wavelet-based attempt to the same aim is proposed in Mallat [277]. A general and insightful theory of image scale space is presented in Koenderink [249].

Section 3.4 : This section is based on the famous paper of Koenderink and Van Doorn [252], proposing to analyse a shape by convolving its characteristic function with a gaussian.

Section 3.5 : We use the first version of curve analysis proposed by Mackworth and Mokhtarian [271]. See also Horn and Weldon [208].

Section 3.6 : The solution to a restoration of causality and locality for the Dynamic Shape was found by Merriman, Bence and Osher [291] who devised this algorithm for totally independent reasons : they looked for a clever numerical implementation of the mean curvature motion equation. In [272], the model errors of [271] were corrected and the right intrinsic equation proposed. This 1992 paper contains, however, several inexact statements about the properties of the intrinsic equation. The right theorems and proofs can be found in Grayson [180], 1987.

Part II

Contrast Invariant Image Analysis

Chapter 4

Contrast invariant classes of functions and their level sets

Let $u(\mathbf{x})$ a real function, which we interpret again as an image. Define the level set of u with level λ by $\mathcal{X}_\lambda u = \{\mathbf{x}, u(\mathbf{x}) \geq \lambda\}$, for $\lambda \in \overline{I\!R}$. (We denote by $\overline{I\!R}$ the set $I\!R \cup \{-\infty, +\infty\}$). Obviously,

$$\mathcal{X}_{-\infty} u = I\!R^N, \;\; \mathcal{X}_{+\infty} u = \emptyset.$$

The level sets of a function have two striking properties. The first one is that they give a complete account of the function. Indeed, we can reconstruct u from its level sets $\mathcal{X}_\lambda u$ by the formula

$$u(\mathbf{x}) = \sup\{\lambda, \mathbf{x} \in \mathcal{X}_\lambda u\}.$$

Exercise 4.1 *Show this last formula, which is true for any real function u.*

The second one is their global invariance by contrast change. We say that two functions u and v have globally the same level sets if for every λ there is μ such that $\mathcal{X}_\mu v = \mathcal{X}_\lambda u$, and conversely. If we apply to u a contrast change understood as a continuous increasing function g, then it is easily checked that $v = g(u)$ and u have globally the same level sets. (Every level set of u is a level set of v and conversely).

Exercise 4.2 *Check this fact, for any real valued fonction u and any increasing continuous function g.*

We shall investigate more general contrast changes, which are nondecreasing, but neither continuous nor increasing. This is justified by the technology where such contrast changes are used systematically for image digitization. In that case, we shall see that the level sets of $g(u)$ are either level sets of u or strict level sets of u, that is, sets of the form $\{\mathbf{x}, u(\mathbf{x}) > \lambda\}$. We shall prove a converse statement : assume that a function v has the same level sets as another function u in the preceding sense. Then v is obtained from u by a contrast change $v = g(u)$.

4.1 From an image to its level sets, and conversely.

In the following proposition, we give structure properties of the family of level sets $\mathcal{X}_\lambda u$ of a function u. Conversely, given a family of sets $(X_\lambda)_{\lambda \in I\!R}$ satisfying the structure properties, we construct a function u

such that $\mathcal{X}_\lambda u = X_\lambda$ for every λ. This explains a main technique of mathematical morphology, consisting in handling directly the level sets of a function and ensuring reconstruction ([290]).

In the whole book, we take the conventions on infima and suprema of subsets of $\overline{I\!R}$ that

$$\inf(\emptyset) = +\infty, \quad \sup(\emptyset) = -\infty.$$

Proposition 4.1 *If $u : I\!R^N \to \overline{I\!R}$ is a function and $X_\lambda = \mathcal{X}_\lambda u$ denote its level sets, then*

(4.1.i) $X_\lambda \subset X_\mu$ if $\lambda > \mu$, $X_{-\infty} = I\!R^N$.

(4.1.ii) $X_\lambda = \cap_{\mu<\lambda} X_\mu$ for every $\lambda > -\infty$.

Conversely, if $(X_\lambda)_{\lambda \in I\!R}$ is a family of subsets of $I\!R^N$ satisfying (4.1i) and (4.1.ii), then the function with values in $\overline{I\!R}$,

(4.1.iii) $u(\mathbf{x}) = \sup\{\mu, \mathbf{x} \in X_\mu\}$ satisfies $\mathcal{X}_\lambda u = X_\lambda$ for every $\lambda \in \overline{I\!R}$.

Figure 4.1: Level sets of a digital image. In this figure, we first show a grey level image, which has a range of grey levels from 0 to 255, and then eight level sets, in decreasing order from 225 to 50, the grey scale step being 25. Notice how essential features of the shapes are contained in the boundaries of level sets, the level lines. Each level set (in gray) is contained in the next, (Formula 4.1.i).

Proof of Proposition 4.1 Relation (4.1.i) is obvious and so is Relation (4.1.ii), since $u(\mathbf{x}) \geq \lambda$ if and only if $u(\mathbf{x}) \geq \mu$ for every $\mu < \lambda$. Conversely, we define a function u from the set of sets $(X_\lambda)_{\lambda \in I\!R}$ by

$$u(\mathbf{x}) = \sup\{\lambda \in \overline{I\!R}, \mathbf{x} \in X_\lambda\} \tag{4.1}$$

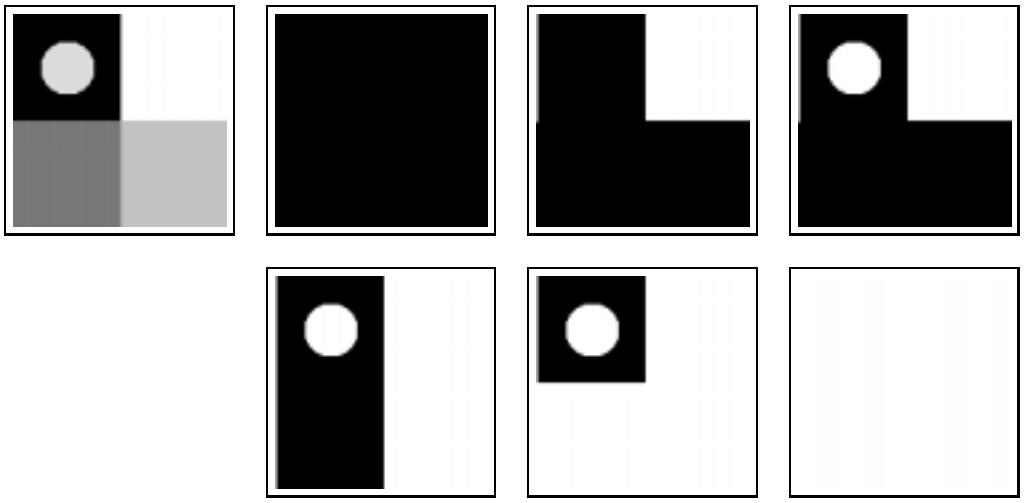

Figure 4.2: A simple synthetic image and all of its level-sets (in gray) with decreasing levels from top to bottom and from left to right.

Let us prove that for every $\lambda \in \overline{I\!R}$,

$$\mathcal{X}_\lambda u = X_\lambda \tag{4.2}$$

Indeed, if $\mathbf{x}$ belongs to X_λ, then from (4.1) we deduce that $u(\mathbf{x}) \geq \lambda$, so that $\mathbf{x}$ belongs to $\mathcal{X}_\lambda u$. Thus, $X_\lambda \subset \mathcal{X}_\lambda u$. Conversely, if $\mathbf{x}$ belongs to $\mathcal{X}_\lambda u$ with $\lambda > -\infty$, then $u(\mathbf{x}) \geq \lambda$ and therefore $\sup\{\mu, \mathbf{x} \in X_\mu\} \geq \lambda$. Thus for any $\mu < \lambda$, there exists μ' such that $\lambda \geq \mu' \geq \mu$ and $\mathbf{x} \in X_{\mu'}$. By (4.1.i) we then have $\mathbf{x} \in X_\mu$. Thus, $\mathbf{x}$ belongs to X_μ for every $\mu < \lambda$. From (4.1.ii), we conclude that $\mathbf{x}$ belongs to X_λ. The case $\lambda = -\infty$ is easily checked : Since $X_{-\infty} = I\!R^N$, we obviously have $\mathcal{X}_{-\infty} u \subset X_{-\infty}$. □

Remark 4.2 *Notice that if $u(\mathbf{x}) < \infty$ for every $\mathbf{x}$, then $\mathcal{X}_\infty u = \emptyset$. Conversely, if $X_\infty = \emptyset$, the reconstructed function u from the X_λ satisfies $u(\mathbf{x}) < +\infty$ for every $\mathbf{x}$. In the same way, $u(\mathbf{x}) > -\infty$ for every x if and only if*

$$\bigcup_{\lambda > -\infty} X_\lambda = I\!R^N.$$

Exercise 4.3 *Show that if $X_\lambda = I\!R^N$ for $\lambda < \lambda_0$, then $u \geq \lambda_0$ and if $X_\lambda = \emptyset$ for $\lambda > \lambda_0$, then $u \leq \lambda_0$.*

Proposition 4.3 *Let $u(\mathbf{x})$ be a real function and $\mathcal{X}_\lambda u = \{\mathbf{x},\ u(\mathbf{x}) \geq \lambda\}$ its level sets for $\lambda \in I\!R \cup \{+\infty\}$.*

$$\text{If } \mu_n \nearrow \lambda, \text{ then } \mathcal{X}_\lambda u = \bigcap_n \mathcal{X}_{\mu_n} u, \tag{4.3}$$

$$\{\mathbf{x},\ u(\mathbf{x}) > \lambda\} = \bigcup_{\mu > \lambda} \mathcal{X}_\mu u. \tag{4.4}$$

Proof Relation (4.4) follows from the obvious equivalence $u(\mathbf{x}) > \lambda \Leftrightarrow (\exists \mu > \lambda,\ u(\mathbf{x}) \geq \mu)$. □

Figure 4.3: Image reconstruction from its level sets. Illustration of Proposition 4.1. We reconstruct an image from its level sets. We used top-left : all level sets, top-right : all level sets whose grey level is a multiple of 8, bottom-left : multiple of 16 and bottom-right : multiple of 32. Notice the relative stability of the image shape content under these drastic quantizations of the grey levels.

4.1.1 Functions and level sets defined almost everywhere

Let us give a statement simpler than Proposition 4.1, provided we consider functions and sets defined almost everywhere. We say that a set X is contained in a set Y almost everywhere if $meas(X \setminus Y) = 0$, where $meas$ denotes the usual Lebesgue measure in $\mathbb{R}$ (length in dimension 1, area in dimension 2, volume in dimension 3...). We say that $X = Y$ almost everywhere if $X \subset Y$ and $Y \subset X$ almost everywhere. We say that two functions u and v are almost everywhere equal if $meas(\{\mathbf{x}, u(\mathbf{x}) \neq v(\mathbf{x})\}) = 0$. More generally, we say that a property $P(\lambda)$, $\lambda \in \mathbb{R}^N$, is true " almost everywhere" or "for almost every λ" if it is true for every λ, with the exception of a set with zero N-dimensional Lebesgue measure.

Lemma 4.4 *Let $(X_\lambda)_{\lambda \in \mathbb{R}}$ be a nonincreasing family of sets, i.e. $X_\lambda \subset X_\mu$ if $\lambda \geq \mu$. Then, for almost every λ in $\mathbb{R}$,*

$$X_\lambda = \bigcap_{\mu < \lambda} X_\mu, \qquad \textit{almost everywhere} \tag{4.5}$$

Proof Let us consider an integrable strictly positive continuous function $h \in L^1(\mathbb{R}^N)$. Set $m(X) = \int_X h(\mathbf{x})d\mathbf{x}$. We notice that $m(X) = 0$ if and only if $meas(X) = 0$. The function $\lambda \to m(X_\lambda)$ is nonincreasing. Thus, it has a countable set of jumps. Since every countable set has zero Lebesgue measure, we deduce that for almost every λ,

$$\lim_{\mu \to \lambda} m(X_\mu) = m(X_\lambda).$$

As a consequence, for those λ's, $m(\bigcap_{\mu<\lambda} X_\mu \setminus X_\lambda) = 0$, which implies (4.5).

□

Exercise 4.4 *Show the property used in the former proof : if h is a positive continuous integrable function on $\mathbb{R}^N$ and if we set $m(X) = \int_{\mathbb{R}^N} h(\mathbf{x})d\mathbf{x}$, then for every measurable set X, $m(X) = 0$ if and only if*

$meas(X) = 0$.

Corollary 4.5 *Let $(X_\lambda)_{\lambda \in \overline{\mathbb{R}}}$ a family of subsets of $\mathbb{R}^N$ such that $X_{-\infty} = \mathbb{R}^N$, $X_\lambda \subset X_\mu$ for $\lambda \geq \mu$. Then the function u defined by*

$$u(\mathbf{x}) = \sup\{\lambda, \mathbf{x} \in X_\lambda\}$$

satisfies for almost every λ, $X_\lambda = \mathcal{X}_\lambda u$ almost everywhere. This function has values in $\overline{\mathbb{R}}$.

Proof We proceed as in the proof of Proposition 4.1. We have

$$\mathcal{X}_\lambda u = \{\mathbf{x}, \sup\{\mu, \mathbf{x} \in X_\mu\} \geq \lambda\}$$

Now, if $\mathbf{x} \in X_\lambda$, we have $\sup\{\mu, \mathbf{x} \in X_\mu\} \geq \lambda$ which implies $\mathbf{x} \in \mathcal{X}_\lambda u$. Thus, $X_\lambda \subset \mathcal{X}_\lambda u$ everywhere. Conversely, let λ be chosen such that $X_\lambda = \cap_{\mu<\lambda} X_\mu$ almost everywhere, which by Lemma 4.4 is true for almost every $\lambda \in \mathbb{R}$. Then if $\mathbf{x} \in \mathcal{X}_\lambda u$, we have by definition of u, $\mathbf{x} \in X_\mu$ for every $\mu < \lambda$. Thus $\mathbf{x} \in \bigcap_{\mu<\lambda} X_\mu$. We conclude that $X_\lambda u \subset \bigcap_{\mu<\lambda} X_\mu$ and therefore $\mathcal{X}_\lambda u \subset X_\lambda$ almost everywhere. □

We end this subsection with a last useful lemma ensuring that if we know almost everywhere the level sets of a function for almost all levels, then the function itself can be retrieved, up to a set with measure zero.

Lemma 4.6 *Let v be a function and $(Y_\lambda)_{\lambda \in \mathbb{R}}$ be a family of sets such that*

$$\mathcal{X}_\lambda v = Y_\lambda, \quad a.e.\ in\ \lambda,\ a.e.\ in\ \mathbf{x}.$$

Then $v(x) = \sup\{\lambda,\ \mathbf{x} \in Y_\lambda,\}$ a.e. in $\mathbf{x}$.

Proof Let N be the negligible subset of $\mathbb{R}$ such that $\mathcal{X}_\lambda u = Y_\lambda$ almost everywhere and for all $\lambda \in \mathbb{R} \setminus N$. We choose $\Lambda \subset \mathbb{R} \setminus N$, a countable, dense subset of $\mathbb{R}$. We then still have

$$v(\mathbf{x}) = \sup\{\lambda \in \Lambda,\ u(\mathbf{x}) \in \mathcal{X}_\lambda v\}.$$

Let now $N_\lambda = (\mathcal{X}_\lambda u \setminus Y_\lambda) \cup (Y_\lambda \setminus \mathcal{X}_\lambda u)$ for $\lambda \in \Lambda$ and $M = \bigcup_{\lambda \in \Lambda} N_\lambda$. We have $meas(M) = 0$ and, for $\mathbf{x} \in \mathbb{R}^N \setminus M$, $v(\mathbf{x}) = \sup\{\lambda \in \Lambda,\ \mathbf{x} \in \mathcal{X}_\lambda v,\ \} = \sup\{\lambda \in \Lambda,\ \mathbf{x} \in Y_\lambda.\}$.

□

Exercise 4.5 *Construct a simple example of family $X_\lambda \subset \mathbb{R}^2$, $\lambda \in \overline{\mathbb{R}}$ for which u attains the values $+\infty$ and $-\infty$ on subsets of $\mathbb{R}^2$ with positive measure.*

4.2 Contrast changes and level sets.

Definition 4.7 *We call contrast change any nondecreasing function $g : \mathbb{R} \to \mathbb{R}$. We also consider its natural extension from $\overline{\mathbb{R}}$ into $\overline{\mathbb{R}}$ obtained by setting*

$$g(+\infty) = \sup_{\mathbb{R}} g, \quad g(-\infty) = \inf_{\mathbb{R}} g.$$

Without risk of confusion, we systematically adopt this extension.

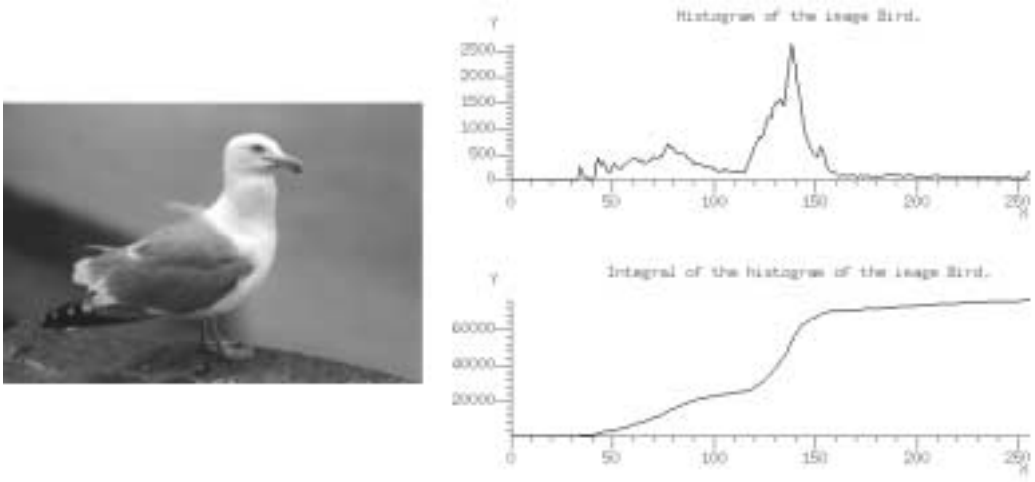

Figure 4.4: Histogram of an image : for each $i \in \{0, 1, ..., 255\}$, we display (above, right) the function $h(i) = Card\{\mathbf{x}, u(\mathbf{x}) = i\}$. The function below is the repartition function of u, $g(i) = Card\{\mathbf{x}, u(\mathbf{x}) \leq i\}$, that is, a primitive of h. It gives an indication on the overall contrast of the image and on the contrast change imposed by the sensor. The inverse function g^{-1} can be used as inverse contrast change, in order to restore an image $g^{-1}(u)$ with flat histogram. On the left column, original image and, below, the result of this "histogram equalization".

The next lemma defines a pseudoinverse $g^{(-1)}$ to any nondecreasing function g.

Lemma 4.8 *Let $g : \overline{I\!R} \to \overline{I\!R}$ be a real nondecreasing function and set for every $\lambda \in \overline{I\!R}$,*

$$g^{(-1)}(\lambda) = \inf\{r,\ g(r) \geq \lambda\}.$$

Then

$$\text{if } \lambda \leq g(g^{(-1)}(\lambda)),\ \text{then } g(s) \geq \lambda \Leftrightarrow s \geq g^{(-1)}(\lambda). \tag{4.6}$$

$$\text{if } \lambda > g(g^{(-1)}(\lambda)),\ \text{then } g(s) \geq \lambda \Leftrightarrow s > g^{(-1)}(\lambda). \tag{4.7}$$

Proof Assume first that $\lambda \leq g(g^{(-1)}(\lambda))$. If $g(s) \geq \lambda$, then $s \geq g^{(-1)}(\lambda)$ by the definition of $g^{(-1)}(\lambda)$. Conversely, if $s \geq g^{(-1)}(\lambda)$, we obtain

$$g(s) \geq g(g^{(-1)}(\lambda)) \geq \lambda \text{ and therefore } \ g(s) \geq \lambda.$$

This yields (4.6).

Assume now that $\lambda > g(g^{(-1)}(\lambda))$. Then $g(s) \geq \lambda$ implies $g(s) > g(g^{(-1)}(\lambda))$. Thus, g being nondecreasing, $s > g^{(-1)}(\lambda)$. Conversely, assume that $s > g^{(-1)}(\lambda)$. Then there exists by definition of $g^{(-1)}$ some $r < s$ such that $g(r) \geq \lambda$ and therefore $g(s) \geq g(r) \geq \lambda$. □

Exercise 4.6 *Compute $g^{(-1)}$ for the following functions :*

- $g(s) = \max(0, s)$
- $g(s) = 1$ *if* $s \geq 0$, s *otherwise.*

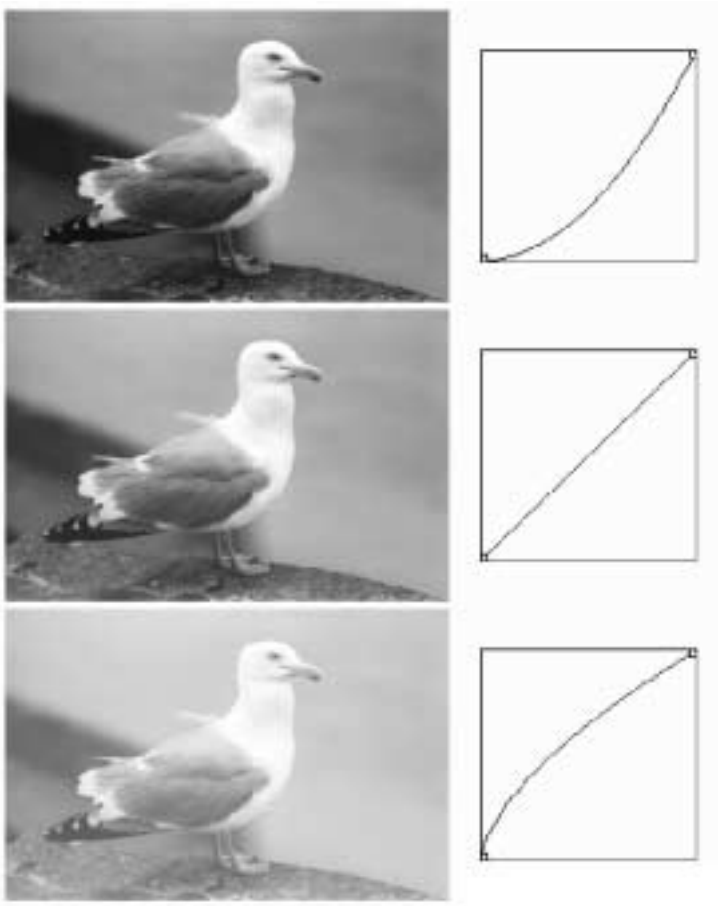

Figure 4.5: Contrast changes, and an equivalence class of images. A grey level image with different contrasts. The three displayed images have exactly the same level sets and level lines, but are associated with three different scales of grey levels. The graphs on the right are the graphs of the functions $u \to g(u)$ which have been applied to the initial grey levels. The first one is convex and enhances the brighter parts of the image. The second one is the identity, leaving thus the image unaltered. The third one is concave and enhances the darker parts of the image. When we dispose of a digital image, we spontaneously arrange the constrast in order to see better such or such shape information. Thus, from the image analysis viewpoint, the image data should be considered as an equivalence class under all possible contrast changes.

The next theorem explains in which sense level sets are conserved by a nondecreasing contrast change g.

Theorem 4.9 *Let g be a real nondecreasing function and $u(\mathbf{x})$ a real function defined on $\mathbb{R}^N$. Then every level set of $g(u)$, $\mathcal{X}_\lambda(g(u))$, satisfies one of the following properties*

$$\exists \mu,\ \mathcal{X}_\lambda(g(u)) = \{\mathbf{x}, u(\mathbf{x}) \geq g^{-1}(\lambda)\} = \mathcal{X}_{g^{(-1)}(\lambda)} u, \tag{4.8}$$

$$\exists \mu,\ \mathcal{X}_\lambda(g(u)) = \{\mathbf{x}, u(\mathbf{x}) > g^{-1}(\lambda)\}. \tag{4.9}$$

Proof Following Lemma 4.8, we have for each λ the alternative (4.6-4.7), that is

$$(g(s) \geq \lambda \Leftrightarrow s \geq g^{(-1)}(\lambda)) \quad \text{or} \quad (g(s) \geq \lambda \Leftrightarrow (s > g^{(-1)}(\lambda)).$$

This yields the alternative

$$\mathcal{X}_\lambda(g(u)) = \{\mathbf{x}, g(u(\mathbf{x})) \geq \lambda\} = \{\mathbf{x}, u(\mathbf{x}) \geq g^{(-1)}(\lambda)\} = \mathcal{X}_{g^{(-1)}(\lambda)} u,$$

or

$$\mathcal{X}_\lambda(g(u)) = \{\mathbf{x}, g(u(\mathbf{x})) \geq \lambda\} = \{\mathbf{x}, u(\mathbf{x}) > g^{(-1)}(\lambda)\}$$

□

Exercise 4.7 *Let $g : \mathbb{R} \to \mathbb{R}$ be a nondecreasing function and $u : \mathbb{R}^N \to \mathbb{R}$ a real function. Show that*

$$\forall \lambda \in \mathbb{R},\ \forall \varepsilon > 0,\ \exists \mu,\ \ \textit{such that } \mathcal{X}_\lambda(g(u)) \supset \mathcal{X}_\mu u \textit{ and } meas(\mathcal{X}_\lambda(g(u)) \setminus \mathcal{X}_\mu u) < \varepsilon.$$

Remark 4.10 Let us give a very simple example of functions u et g where both cases (4.8) et (4.9) occur. Let us set for $x \in \mathbb{R}$, $u(x) = x$ and $g(x) = x$ if $x \leq 0$, $g(x) = x + 1$ if $x > 0$. Then the level set $\mathcal{X}_1(g(u))$ is not a level set of u but we have

$$\mathcal{X}_1(g(u)) =]0, +\infty[= \{\mathbf{x}, u(\mathbf{x}) > 0\}.$$

Exercise 4.8 *Compute in that case all level sets of u and $g(u)$ and compare them.*

Let us state a converse statement to Theorem 4.9 : it states roughly that if the level sets of v are level sets of u, then $v = g(u)$ for some contrast change.

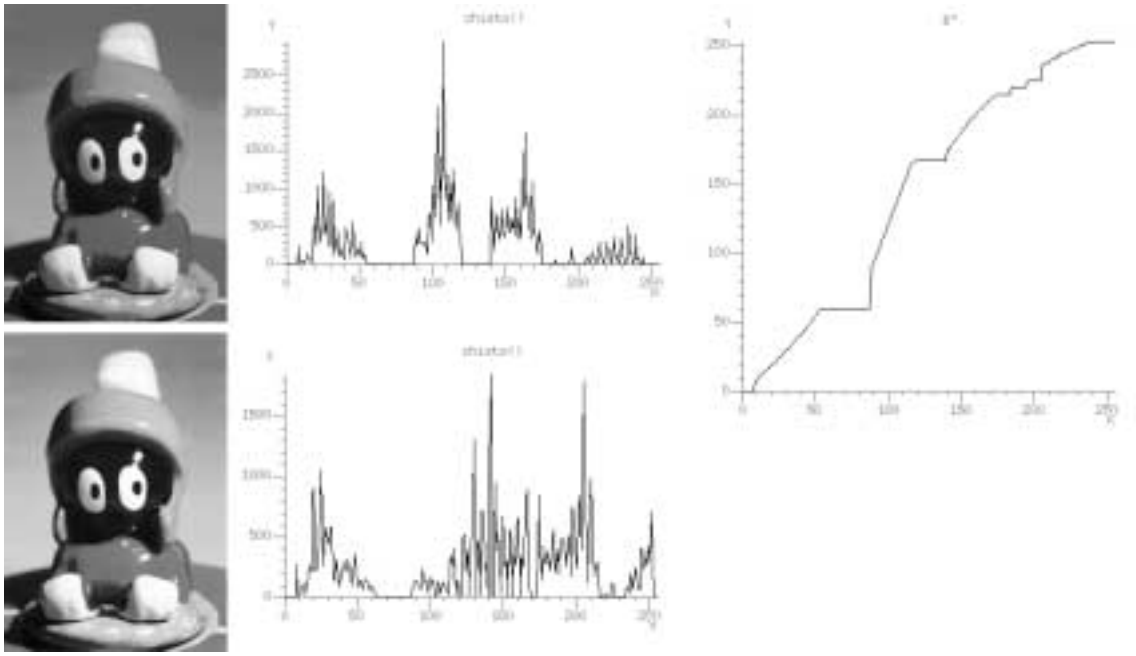

Figure 4.6: The two images (left) have the same set of level sets. The contrast change function that makes the upper image become the lower image is displayed on the right. It corresponds to one of the possible g functions whose existence is stated in Corollary 4.12. The function g may be locally constant on intervals where the histogram of the upper image is 0 (see top-middle graph). Indeed, on such grey level intervals, the level-sets are steady.

Theorem 4.11 *Let u and v be two real functions on $\mathbb{R}^N$ such that the following alternative holds for every $\mu \in \mathbb{R}$:*

$$\exists \mu,\ \mathcal{X}_\lambda v = \mathcal{X}_\mu u,\ \textit{or} \tag{4.10}$$

$$\exists \mu,\ \mathcal{X}_\lambda v = \{\mathbf{x}, u(\mathbf{x}) > \mu\}. \tag{4.11}$$

Then, there exists a real nondecreasing function $g : \overline{I\!R} \to \overline{I\!R}$ *such that* $v = g(u)$. *In addition,* g *can be defined by the simple formula*

$$g(\mu) = \sup\{\lambda, \mathcal{X}_\lambda v \supset \mathcal{X}_\mu u\}. \tag{4.12}$$

Proof

Step 1. We first show that $v(\mathbf{x}) \geq g(u(\mathbf{x}))$. By applying (4.1.ii) with μ replaced by $g(\mu)$, we see that the "sup" in the definition of g, (4.12) is in fact a "max", so that

$$\mathcal{X}_{g(\mu)} v \supset \mathcal{X}_\mu u. \tag{4.13}$$

Let us apply Relation (4.13) with $\mu = u(\mathbf{x})$. We obtain

$$\mathcal{X}_{g(u(\mathbf{x}))} v \supset \mathcal{X}_{u(\mathbf{x})} u \ni \mathbf{x}$$

and therefore $v(\mathbf{x}) \geq g(u(\mathbf{x}))$.

Step 2. Let us now show that $v(\mathbf{x}) \leq g(u(\mathbf{x}))$. We apply the assumption (4.10-4.11) to $\lambda = v(\mathbf{x})$ and call $\mu(\mathbf{x})$ the corresponding value of μ. Thus

$$\mathcal{X}_{v(\mathbf{x})} v = \mathcal{X}_{\mu(\mathbf{x})} u, \text{ or} \tag{4.14}$$

$$\mathcal{X}_{v(\mathbf{x})} v = \{\mathbf{x}, u(\mathbf{x}) > \mu(\mathbf{x})\}. \tag{4.15}$$

Assume first that (4.14) holds. Since $\mathcal{X}_{v(\mathbf{x})} v \ni \mathbf{x}$, we deduce from (4.14) that $\mathcal{X}_{\mu(\mathbf{x})} u \ni \mathbf{x}$ and therefore $u(\mathbf{x}) \geq \mu(\mathbf{x})$. Thus

$$\mathcal{X}_{v(\mathbf{x})} v = \mathcal{X}_{\mu(\mathbf{x})} u \supset \mathcal{X}_{u(\mathbf{x})} u,$$

which, by definition of g, yields $v(\mathbf{x}) \leq g(u(\mathbf{x}))$.

Assume now that (4.15) holds. Since $\mathcal{X}_{v(\mathbf{x})} v = \{\mathbf{x}, u(\mathbf{x}) > \mu(\mathbf{x})\} = \bigcup_{\mu > \mu(\mathbf{x})} \mathcal{X}_\mu u$ and $\mathbf{x} \in \mathcal{X}_{v(\mathbf{x})} v$, we deduce the existence of $\mu > \mu(\mathbf{x})$ such that $\mathbf{x} \in \mathcal{X}_\mu u$ and therefore $u(\mathbf{x}) \geq \mu$. Thus

$$\mathcal{X}_{v(\mathbf{x})} v \supset \mathcal{X}_\mu u \supset \mathcal{X}_{u(\mathbf{x})} u,$$

which, again by definition of g, implies that $v(\mathbf{x}) \leq g(u(\mathbf{x}))$. □

We conclude this section with a simple to remember statement : il all level sets of v are level sets of of u, then $v = g(u)$ for some g. The precise statement is

Corollary 4.12 *If* v *and* u *are two real functions such that all level sets of* v *are level sets of* u, *then there is a nondecreasing function* $g : \overline{I\!R} \to \overline{I\!R}$ *such that* $v = g(u)$. *An instance of such a function* g *is*

$$g(\mu) = \sup\{\lambda, \mathcal{X}_\lambda v \supset \mathcal{X}_\mu u\}. \tag{4.16}$$

4.3 Semicontinuous contrast changes

In this section, we adress the case where the considered contrast changes are upper semicontinuous (or lower semicontinuous), which makes sense for digital image processing because threshold functions and quantization functions usually are u.s.c. or l.s.c.. The results of this section are not used in the next chapters.

Definition 4.13 *We say that a real nondecreasing function $g : \overline{\mathbb{R}} \to \overline{\mathbb{R}}$ is upper semi-continuous if for all $r < +\infty$,*

$$g(r) = \lim_{\rho \to r^+} g(\rho) = \lim_{\rho \to r, \rho > r} g(\rho)$$

Corollary 4.14 *Assume g is a nondecreasing upper semi-continuous function. Then for every $\lambda \in \overline{\mathbb{R}}$ and every $s < +\infty$,*

$$g(s) \geq \lambda \Leftrightarrow s \geq g^{(-1)}(\lambda). \tag{4.17}$$

Exercise 4.9 *Show that the condition $s < +\infty$ is necessary : Take $g(s) = 0$ for all s and $\lambda = 1$. Then check that $g^{-1}(1) = +\infty$, while $g(s) \geq 1$ is impossible.*

Proof Assume first that $g^{(-1)}(\lambda) < +\infty$. If g is upper semi-continuous, then the "inf" defining $g^{(-1)}(\lambda)$ also is a minimum. Indeed, let $r_n \searrow g^{(-1)}(\lambda)$. Then by definition of $g^{(-1)}(\lambda)$, we have $g(r_n) \geq \lambda$. Since g is upper semi-continuous, we also have

$$g(g^{(-1)}(\lambda)) \geq \lim_n g(r_n) \geq \lambda.$$

By Lemma 4.8, we therefore have

$$g(s) \geq \lambda \Leftrightarrow s \geq g^{(-1)}(\lambda). \tag{4.18}$$

If now $g^{(-1)}(\lambda) = +\infty$, we have $g(s) \geq \lambda$ only if $s = +\infty$, which is excluded by the condition $s < +\infty$ we have put on (4.17). Thus $\lambda \leq g(g^{(-1)}(\lambda))$ and the conclusion of (4.6) holds. □

The next corollary shows that upper semicontinuous contrast changes preserve level sets.

Corollary 4.15 *Let g be a real nondecreasing upper semi-continuous function and $u(\mathbf{x})$ a real function defined on $\mathbb{R}^N$. Then every level set of $g(u)$, $\mathcal{X}_\lambda(g(u))$ satisfies*

$$\exists \mu,\ \mathcal{X}_\lambda(g(u)) = \mathcal{X}_\mu u, \tag{4.19}$$

where $\mu = g^{(-1)}(\lambda)$.

Proof Set, as in Lemma 4.8,

$$g^{(-1)}(\lambda) = \inf\{r,\ g(r) \geq \lambda\}.$$

According to Corollary 4.14, $g(u(\mathbf{x})) \geq \lambda \Leftrightarrow u(\mathbf{x}) \geq g^{(-1)}(\lambda)$, which proves (4.19). (Notice that $u(\mathbf{x}) < +\infty$ for all $\mathbf{x}$, so that (4.17) holds for $s = u(\mathbf{x})$.) □

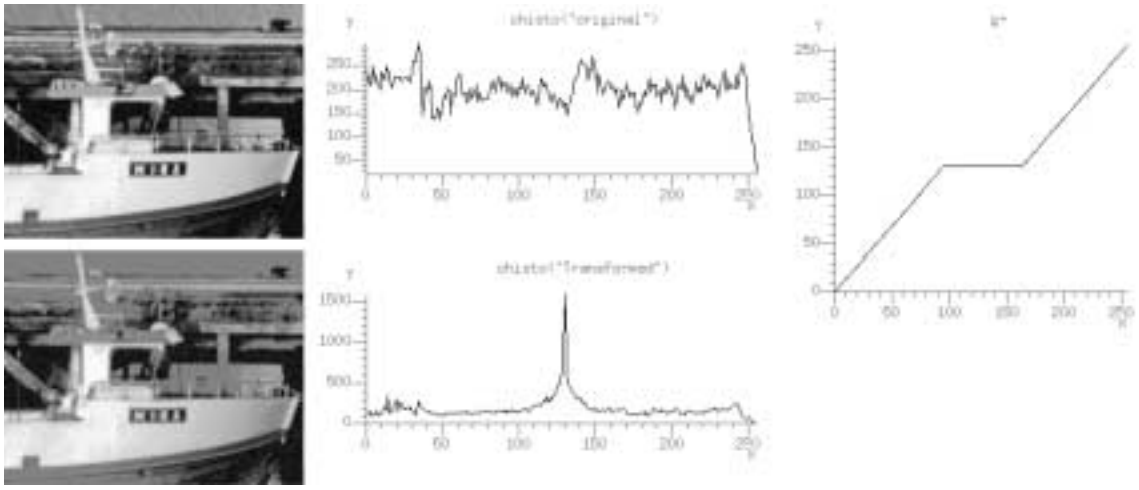

Figure 4.7: The original image (top-left) has a strictly positive histogram (all grey levels between 0 and 255 are represented). Therefore, if any non strictly increasing contrast change g is applied, some data will be lost. Every level set of the transformed image $g(u)$ is a level set of the original image. Now, the original image has more level sets than the transformed one.

References.

Section 4.1 : Wertheimer [435] noticed that the actual local values of grey levels could not be a relevant information to the human visual system. Contrast invariance is one of the fundamental model assumptions in Mathematical Morphology. The books of Matheron [290] and Serra [378], Serra[379] are seminal. In "flat morphology", images are decomposed into their level sets by the so-called threshold decomposition so that image analysis boils down to set analysis. Ballester et al. [45] defined an "image intersection" whose principle it is to keep all pieces of bilevel sets common to two images. Monasse et al. [300] defined recently a "fast level set transform" (FLST) whose principle it is to associate with every image the inclusion tree of connected components of level sets. They show that the inclusion tree of connected components of upper and lower level sets can be fused into a single inclusion tree and use it, among other applications, for image registration. See Monasse PhD [?].

Section 4.2 and ?? : Contrast changes are a very practical and usual tool for improving image visualization. They are implemented in any image analysis software. For level set-preserving contrast changes, see [88]. Many references on contrast invariant operators will be given at the end of Chapter 7.

Chapter 5

Level lines and level surfaces.

5.1 Curves, their normal, their curvature

In this section, we summarize what must be known about smooth curves of the plane, particularly Jordan curves, that is, curves without selfcrossing points. Such curves will appear in practice as boundaries of smoothed shapes in an image, and more generally in the next section, as level lines of a smooth enough image.

Definition 5.1 *We call planar curve and denote by C the range of a continuous map $\mathbf{x}(s) \ : s \in [a,b] \to I\!R^2$. If the restriction of $\mathbf{x}$ to $]a,b[$ is one to one, we say that C is a Jordan curve. If in addition $\mathbf{x}(a) = \mathbf{x}(b)$, we say that C is a closed Jordan curve. We call the map $\mathbf{x}(s)$ "a parameterization of the curve C". We say that a curve C is C^m ($m \in I\!N$, $m \geq 1$) if it admits a C^m parameterization $\mathbf{x}(s) \ : s \in [a,b] \to I\!R^2$ such that $\forall s, |\mathbf{x}'(s)| = 1$. We call such a parameterization euclidean.*

In shape analysis, the normal and curvature will play a central role and we define them now.

Definition 5.2 (and Proposition.)
Let C be a simple C^2 curve and $\mathbf{x}_0$ a point of C.
(i) C admits in a neighborhood of $\mathbf{x}_0$ exactly two euclidean parameterizations such that $\mathbf{x}(0) = \mathbf{x}_0$. If $\mathbf{x}(s) \ : [-a,a] \to I\!R^2$ is one such parameterization of C, then $\mathbf{x}(-s)$ is the other one. In addition, all other euclidean parameterizations of C have the form $\mathbf{x}(s+\sigma)$ or $\mathbf{x}(-s+\sigma)$ for some $\sigma \in I\!R$.
(ii) We call tangent vector at $\mathbf{x}$ to C (parameterized by an euclidean parameterization $\mathbf{x}(s)$) the unit vector $\tau(s) = \frac{\partial \mathbf{x}}{\partial s}$. As a consequence of (i), if τ_1 and τ_2 are two tangent vectors at $\mathbf{x}_0$, then either $\tau_1 = \tau_2$ or $\tau_1 = -\tau_2$.
(iii) The vector $\frac{\partial^2 \mathbf{x}}{\partial s^2}(0)$ is independent by (i) of the choice of the euclidean parameterization and it is orthogonal to $\tau(\mathbf{x}_0)$. It is called curvature vector *of the curve C at $\mathbf{x}_0$ and denoted by $Curv(\mathbf{x}_0)$.*
(iv) We call generalized normal *the vector $\mathbf{n}(\mathbf{x}_0) = \frac{Curv(\mathbf{x}_0)}{|Curv(\mathbf{x}_0)|}$ if $Curv(\mathbf{x}_0) \neq 0$ and equal to 0 otherwise.*

Proof. (i) Let $\mathbf{x}, p \in [a,b] \to \mathbf{x}(p) \in C$ be a C^1 parameterization of C such that $\forall p, \ \mathbf{x}'(p) \neq 0$ and $\mathbf{x}(p_0) = \mathbf{x}_0$. We seek for a new C^1 parameterization $\tilde{\mathbf{x}}(s) = \mathbf{x}(\varphi(p))$ (with φ C^1 and one to one) such that $|\tilde{\mathbf{x}}'(s)| = 1$, that is, $|\mathbf{x}'(\varphi(p))||\varphi'(p)| = 1$. Taking into account that this relation implies $\varphi'(p) \neq 0$, we have

two choices, namely

$$\varphi'(p) = \frac{1}{|\mathbf{x}'(\varphi(p))|} \text{ or } \varphi'(p) = -\frac{1}{|\mathbf{x}'(\varphi(p))|} \text{ and } \varphi(0) = p_0. \tag{5.1}$$

Since $\mathbf{x}'$ is a Lipschitz function which never vanishes, so is $\frac{1}{\mathbf{x}'}$. By Cauchy-Lipschitz Theorem, φ is uniquely defined on $\mathbb{R}$ by each one of both ordinary differential equations (5.1) and we get exactly two euclidean parameterizations $\tilde{\mathbf{x}}(s)$ such that $\tilde{\mathbf{x}}(0) = \mathbf{x}_0$. If $\mathbf{x}(s)$ is another euclidean parametrization, let $\mathbf{x}(\sigma) = \mathbf{x}_0$. Then $\mathbf{x}_1(s) = \mathbf{x}(s - \sigma)$ also is an euclidean parameterization and satisfies $\mathbf{x}(0) = \mathbf{x}_0$. Thus either $\mathbf{x}(s) = \tilde{\mathbf{x}}(s + \sigma)$ or $\mathbf{x}(s) = \tilde{\mathbf{x}}(-s + \sigma)$.

(ii) The statements of (ii) follow immediately from (i) by differentiation.

(iii) Since $|\tau(s)|^2 = 1$, we have by differentiation $\tau(s).\frac{\partial \tau(s)}{\partial s} = 0$, that is $\tau(s).\frac{\partial^2 \mathbf{X}}{\partial s^2} = 0$. Changing s into $-s$ does not alter the value of $\frac{\partial^2 \mathbf{X}}{\partial s^2}$ at $s = 0$. □

Exercise 5.1 *(i) Consider an euclidean parameterization for a circle with center 0 and radius R, $s \to \mathbf{x}(s) = (R\cos\frac{2\pi s}{R}, R\sin\frac{2\pi s}{R})$. Compute the normal and curvature and check that the curvature is $\frac{1}{R}$.*
(ii) Compute the curvature at 0 of the parabola of equation $y = \frac{a}{2}x^2$.
(iii) Give the general formula for the curvature of a C^2 graph $y = \varphi(x)$.

We must be able to compute the normal and curvature of a parameterized curve. Exercise **??** yields some useful formulas.

Exercise 5.2 *curvature of parametric curve* **Parametric curves**. *Let $\mathbf{c}(\sigma)$ be a parameterized C^2 curve and $\mathbf{x}(s)$ an euclidean parameterization of the same curve. We assume that the reparameterization $\varphi(s) = \sigma$ is monotonic and satisfies $|\varphi'(s)| > 0$.*
(i) Show that

$$\varphi'(s) = \frac{1}{|c'(\varphi(s))|} \tag{5.2}$$

(differentiate the relation $\mathbf{c}(\varphi(s)) = \mathbf{x}(s)$ and use $|\mathbf{x}'(s)| = 1$.)
(ii) Show that

$$\mathbf{x}''(s) = Curv(\mathbf{x})(s) = c''(\varphi(s))\varphi'(s)^2 + c'(\varphi(s))\varphi'(s). \tag{5.3}$$

Deduce from 5.2 that

$$\varphi''(s) = -\frac{c''(\varphi(s))\varphi'(s).c'(\varphi(s))}{|c'(\varphi(s))|^3} = -\frac{c''(\varphi(s)).c'(\varphi(s))}{|c'(\varphi(s))|^4}. \tag{5.4}$$

(iii) Using 5.4 and 17.20, show that the curvature vector of $\mathbf{c}$ is

$$Curv(\mathbf{c}(\sigma)) = x''(s) = \frac{1}{\mathbf{c}'^2(\sigma)}(\mathbf{c}''(\sigma) - (\mathbf{c}''(\sigma).\frac{\mathbf{c}'(\sigma)}{|\mathbf{c}'(\sigma)|})\frac{\mathbf{c}'(\sigma)}{|\mathbf{c}'(\sigma)|}) \tag{5.5}$$

(iv) Multiplying this formula by the vector normal to the curve, $\frac{\mathbf{c}(\sigma)^\perp}{|\mathbf{c}(\sigma)|}$, deduce a simple formula defining a "scalar curvature" :

$$curv(\mathbf{c}(\sigma)) = Curv(\mathbf{c}(\sigma)).\frac{\mathbf{c}(\sigma)^\perp}{|\mathbf{c}(\sigma)|} = \frac{\mathbf{c}''(\sigma).\mathbf{c}'^\perp(\sigma)}{|\mathbf{c}'(\sigma)|^3}. \tag{5.6}$$

5.2 The level line structure (topographic map) of an image.

In view of the applications to shape analysis (see Sections 3.4, 3.5), it is quite useful to have a description of an image in terms of Jordan curves. If the image is assumed to be C^1, the implicit function theorem yields such a description. We therefore begin with the case of dimension 2 and then will also define level surfaces of u in higher dimension.

Theorem 5.3 *Let $u(\mathbf{x}) = u(x,y)$ be a C^1 function in a neighborhood of a point $\mathbf{x}_0$ such that $Du(\mathbf{x}_0) \neq 0$. Set $\vec{i} = Du/|Du|(\mathbf{x}_0)$ and $\mathbf{x} = \mathbf{x}_0 + x\vec{i} + y\vec{j}$, where $\vec{j}$ is a unit vector orthogonal to $Du(\mathbf{x}_0)$. Then, there exists a disk $D(\mathbf{x}_0, \eta)$ and a C^1 function $\phi(x)$: $x \in [-\eta, \eta] \to \phi(x) \in \mathbb{R}$ such that if $\mathbf{x} = \mathbf{x}_0 + x\vec{i} + y\vec{j} \in D(\mathbf{x}_0, \eta)$, then*

$$u(x,y) = 0 \Leftrightarrow y = \phi(x).$$

The preceding theorem ensures that if u is C^1 and $Du(\mathbf{x}_0) \neq 0$, then around $\mathbf{x}_0$, the set $\{u(\mathbf{x}) = u(\mathbf{x}_0)\}$ is a C^1 graph. The next corollary permits to globalize this result.

Corollary 5.4 *Let u be a C^1 function on $\mathbb{R}^2$, and $\lambda \in \mathbb{R}$ a level such that:*

- *$u^{-1}(\lambda)$ is compact.*
- *$\forall \mathbf{x} \in u^{-1}(\lambda), Du(\mathbf{x}) \neq 0$.*

Then $u^{-1}(\lambda)$ is a finite union of C^1 Jordan curves.

Proof Indeed by the implicit function theorem, $u^{-1}(\lambda)$ is a C^1 graph around each of its points. Since $u^{-1}(\lambda)$ is compact, it can be covered by a finite number of disks inside each of which, by Theorem 5.3, $u^{-1}(\lambda)$ is a C^1 graph. It is easily deduced that $u^{-1}(\lambda)$ is a finite union of C^1 Jordan curves. □

In fact, Sard's Theorem ensures that the preceding situation is generic in λ.

Theorem 5.5 *(Sard [?]) Let u be a real C^1 function on a rectangle R of $\mathbb{R}^N$. Then for almost every λ in $\mathbb{R}$, the set $u^{-1}(\lambda)$ is nonsingular, that is $\forall \mathbf{x} \in u^{-1}(\lambda), Du(\mathbf{x}) \neq 0$.*

Corollary 5.6 *Let $u : \mathbb{R}^2 \to \mathbb{R}$ be a function of L^∞_C, i.e. $(2,2)$-periodic and C^1 $\mathbb{R}^N$. Then for almost every $\lambda \in \mathbb{R}$, the set $u^{-1}(\lambda)$ is a union of C^1 Jordan curves which are either closed or end on the boundary of R.*

Proof We consider a level λ for which $u^{-1}(\lambda)$ is nonsingular. By the implicit function theorem, this last set is a C^1 graph around each point and its restriction to C is compact. We can therefore cover $C \cap u^{-1}(\lambda)$ by a finite set of graphs and the statement follows. □

Figure 5.1: Level lines as a complete representation of the shapes present in an image. All level lines of the image of a sea bird for levels which are multiples of 12. Notice that we do not need a previous smoothing in order to visualize the shape structures in an image : a quantization of the displayed levels is enough. The extrema killer is also very efficient to that visualization aim.

5.3 The topographic map as a complete image representation.

Definition 5.7 *Let u be a C^1 function on a rectangle. We call topographic map of u the map which associates with every nonsingular λ (that is, every λ such that $Du \neq 0$ on $u^{-1}(\lambda)$), the finite set of the oriented Jordan curves of $u^{-1}(\lambda)$. By orientation, we mean that we keep the information of whether $u(\mathbf{x}) > \lambda$ or $u(\mathbf{x}) < \lambda$ inside or outside the Jordan curve. These are either closed, or meet the boundary of the rectangle at their endpoints.*

Theorem 5.8 *[360] The image of a Jordan curve c divides the plane in two connected components, a bounded one and an unbounded one. In other terms, denoting for commodity by $c \subset I\!R^2$ the range of c, the set $I\!R^2 \setminus c$ has exactly two connected components. We denote by $Int(c)$ the bounded one.*

Proposition 5.9 *Let u be a C^1 function on a rectangle. Then u can be recovered from its topographic map.*

Proof Indeed, u can be recovered from a set of level sets associated with a dense set of λ's. Now, the set of singular *lambda's* has zero measure. Thus, its complementary set of nonsingular values is dense. For each nonsingular value λ, the boundary of $\mathcal{X}_\lambda u$ is described by a finite set of oriented Jordan curves. We consider the filling operator *Interior* which associates with each Jordan curve c the set $Interior(c)$ the unique bounded connected component of $I\!R^2 \setminus c$ (Theorem5.8.) Then, calling c_i the positively oriented Jordan curves, understood as those which surround points of $\mathcal{X}_\lambda$, and $c_j, j \in J_i$ the negatively oriented curves which are contained in $Interior(c_i)$ and therefore surround a hole of $\mathcal{X}_\lambda u$, we have the reconstruction formula

$$\mathcal{X}_\lambda u = \bigcap_i (Int(c_i) - \bigcap_{j \in J_i} Int(c_{i,j})).$$

In computational terms, we "fill" the curves c_i and then remove their holes to reconstruct $\mathcal{X}_\lambda$. □

Remark 5.10 *In practice, u can be recovered in the following way: we keep the Jordan curves as arrays*

of vertices and we use a filling algorithm to reconstruct each $\mathcal{X}_\lambda u$. Then u is obtained by the classical

$$u(\mathbf{x}) = \sup\{\lambda, \mathbf{x} \in \mathcal{X}_\lambda u\},$$

which is a finite sup *to perform, since λ assumes a finite number of values in a digital image.*

5.4 Generalized level lines and topographic map

The preceding definition of level lines is slightly restrictive and may be generalized. In view of our definition of level sets, it may be advantageous to consider cases where the level set is smooth enough, so that its boundary is (e.g.) a union of Jordan curves. By boundary, we mean here the topological boundary. Let us give some examples. If c is a Jordan curve, then we may consider the function $u(\mathbf{x})$ which is equal to 1 if $\mathbf{x}$ is surrounded by the curve, 0 otherwise. We see that the implicit function theorem does not apply, and that the set $\{\mathbf{x}, u(\mathbf{x}) = 1\}$ is not a level line. The next definition, however, is adapted to this case :

Definition 5.11 *Let u be upper semicontinous, so that its level sets are closed. Consider the connected components of X^c, where X is a level set, and assume that the boundary of each is a finite union of non crossing Jordan curves meeting each other at a finite number of points. Then, we say that these Jordan curves are level lines of u.*

This definition may seem and is restrictive ; it has, however, the advantage of applying to a main example we will consider in the sequel : u.s.c. piecewise constant images, in particular images constant on pixels. In that case, each level set is a union of closed squares on a grid and it is easily checked that our definition applies and uniquely defines a set of Jordan curves which we call the level lines of the image. Conversely, these level lines uniquely define the associated level set and we can reconstruct from them the image.

As an example, let us consider a checkerboard image, where pixels have alternate values 255 (white) and 0 (black). If we decide that the function is u.s.c., then the white pixels include their own boundaries and are connected. The black pixels instead are disconnected. Thus, the level lines of u are simply all boundaries of all black pixels.

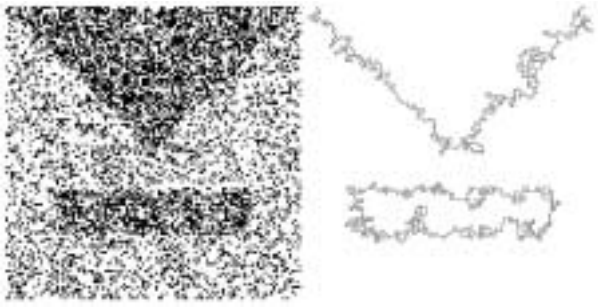

Figure 5.2: Level lines as representatives of the shapes present in an image. Left: noisy binary image with two apparent shapes, right: its two longest level lines.

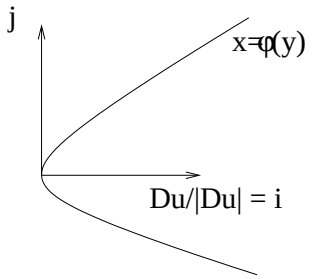

Figure 5.3: Intrinsic coordinates

5.5 The curvature of the level lines

In this section, we assume again that the dimension N is 2. Consider a real function u which is twice differentiable (C^2) in a neighborhood of $\mathbf{x} \in I\!R^2$. Without loss of generality, we choose the origin to stand at $\mathbf{x}$ so that in the following we set $\mathbf{x} = 0$ and we assume that $u(0) = 0$. In order to simplify the notation, we shall omit the mention that computations are done at 0 and write Du instead of $Du(0)$, etc.. We denote by $Du = (u_x, u_y)$ the gradient of u at 0, by $|Du| = (u_x^2 + u_y^2)^{\frac{1}{2}}$ its euclidean norm. If $Du \neq 0$, we call "orientation of the gradient" the normalized vector $\vec{i} = \frac{Du}{|Du|}$. We set $Du^\perp = (-u_y, u_x)$, a vector orthogonal to Du and $\vec{j} = \frac{Du^\perp}{|Du|}$. Taking into account that $\vec{j}$ and $\vec{i}$ are orthogonal and with norm equal to 1, we use them to define the Cartesian coordinates of points $\mathbf{x} = (x, y)$ in a neighborhood of 0. Thus, we set $\mathbf{x} = x\vec{i} + y\vec{j}$. If $Du = 0$, we simply choose for $\vec{i}$ and $\vec{j} = (\vec{i})^\perp$ two arbitrary orthogonal unit vectors.

Since u is C^2, we can by Taylor formula write

$$u(\mathbf{x}) = px + ax^2 + by^2 + cxy + O(|\mathbf{x}|^3), \tag{5.7}$$

where $p = |Du|(0) \geq 0$ and, if $p > 0$,

$$\begin{aligned} b &= \tfrac{1}{2}\tfrac{\partial^2 u}{\partial y^2}(0) = \tfrac{1}{2}D^2u(\tfrac{Du^\perp}{|Du|}, \tfrac{Du^\perp}{|Du|}) \\ a &= \tfrac{1}{2}\tfrac{\partial^2 u}{\partial x^2}(0) = \tfrac{1}{2}D^2u(\tfrac{Du}{|Du|}, \tfrac{Du}{|Du|}) \\ c &= \tfrac{\partial^2 u}{\partial x \partial y}(0) = D^2u(\tfrac{Du^\perp}{|Du|}, \tfrac{Du}{|Du|}) \end{aligned} \tag{5.8}$$

The Implicit Function Theorem 5.3 implies that in a neighborhood of 0, the set $\{\mathbf{x}, u(\mathbf{x})\}$ is a C^2 graph whose equation we can write $x = \varphi(y)$, where φ is a C^2 function in a neighborhood of 0. In order to see how φ is related to u, we rewrite (5.7) as

$$x = \varphi(y) = -\frac{1}{p}(ax^2 + by^2 + cxy) + o(x^2 + y^2).$$

¿From this equation we draw $x^2 = o(y^2)$ and $xy = o(y^2)$, so that finally

$$x = \varphi(y) = -\frac{b}{p}y^2 + o(y^2) \tag{5.9}$$

Equation (5.9) is the equation in intrinsic coordinates of the *level line* $\{\mathbf{x}, u(\mathbf{x}) = u(0)\}$ of u at 0. We deduce that $|\frac{2b}{p}| = \frac{1}{R}$ is the inverse of the radius of the osculatory circle to this curve, which we call absolute curvature. This and Relation (5.8) justify the next definition.

Definition 5.12 *Let u be a real C^2 function defined around a point $\mathbf{x} \in \mathbb{R}^2$ such that $Du(\mathbf{x}) \neq 0$. Then we call curvature of u at $\mathbf{x}$ the real number defined by*

$$\text{curv}(u)(\mathbf{x}) = \frac{1}{|Du|^3} D^2u(Du^\perp, Du^\perp)(\mathbf{x}) = \frac{u_{xx}u_y^2 - 2u_{xy}u_xu_y + u_{yy}u_x^2}{(u_x^2+u_y^2)^{\frac{3}{2}}}(\mathbf{x}). \tag{5.10}$$

Returning to the notations (5.8), we obtain

$$b = \frac{1}{2}|Du|\text{curv}(u)(0). \tag{5.11}$$

Exercise 5.3 *Show that Formula (5.10) can also be written in a compact way as*

$$\text{curv}(u)(\mathbf{x}) = \text{div}(\frac{Du}{|Du|})(\mathbf{x}), \tag{5.12}$$

where we define, as usual, $\text{div}(u) = \frac{\partial}{\partial x}u + \frac{\partial}{\partial y}u = u_x + u_y$. *Show that* $\text{curv}(u)$ *does not depend upon any C^2 increasing or decreasing contrast change. (Just substitute $g(u)$ to u in the formula (5.10) defining the curvature.)*

Exercise 5.4 *Let us make a useful verification. Our definition of curvature must match the intuitive notion that the curvature of a circle is the inverse of its radius. Define $u(\mathbf{x}) = |\mathbf{x}|^2$. Then level lines are circles and the normal vector to level lines is $\vec{n}(\mathbf{x}) = -\frac{\mathbf{x}}{|\mathbf{x}|}$. The level line passing by $\mathbf{x}$ is a circle with radius $|\mathbf{x}| = R$. We parameterize this circle by length : $C = \{(R\cos\frac{s}{R}, R\sin\frac{s}{R})\}$. Then the tangent vector to the level line is $\vec{\tau}(s) = (-\sin\frac{s}{R}, \cos\frac{s}{R})$ and $\frac{\partial\vec{\tau}(s)}{\partial s} = -\frac{1}{R}(\cos\frac{s}{R}, \sin\frac{s}{R}) = \frac{1}{R}\vec{n}(\mathbf{y}(s))$. Thus, the curvature of the level line (defined as the modulus of the acceleration along the curve) is $\frac{1}{R}$. Check that with Formula 5.10, one has $curv(u)(\mathbf{x}) = \frac{1}{R}$.*

Exercise 5.5 *Curvature and local comparaison of functions.*
We say that a function f is radial and increasing if there exists an increasing function $g : \mathbb{R}^+ \to \mathbb{R}$ and $\mathbf{x} \in \mathbb{R}^2$ such that $f(\mathbf{y}) = g((\mathbf{x}-\mathbf{y})^2)$. Let $u(\mathbf{x})$ be a C^2 function from $\mathbb{R}^2$ into $\mathbb{R}$. Assume that $Du(\mathbf{x}_0) \neq 0$. Our aim is to show that there exist for every $\epsilon > 0$ two C^2 radial and increasing functions f_ϵ^+ and f_ϵ^- such that

$$f_\epsilon^-(\mathbf{x}_0) = u(\mathbf{x}_0) = f_\epsilon^+(\mathbf{x}_0),$$
$$Df_\epsilon^-(\mathbf{x}_0) = Du(\mathbf{x}_0) = Df_\epsilon^+(\mathbf{x}_0),$$
$$curv(f_\epsilon^-)(\mathbf{x}_0) + \epsilon = curv(u)(\mathbf{x}_0) = curv(f_\epsilon^+)(\mathbf{x}_0) - \epsilon,$$
$$f_\epsilon^-(\mathbf{x}) + o((\mathbf{x}-\mathbf{x}_0)^2) \leq u(\mathbf{x}) \leq f_\epsilon^+(\mathbf{x}) + o((\mathbf{x}-\mathbf{x}_0)^2).$$

1. Choose, without loss of generality, $\mathbf{x}_0 = (0,0)$, $Du(\mathbf{x}_0) = (p,0)$, with $p > 0$. Then for some a, b, c we have by Taylor expansion

$$u(\mathbf{x}) = px + ax^2 + by^2 + cxy + o(x^2+y^2)$$

Show that for every $\epsilon > 0$,

$$px + (b-\epsilon)y^2 + (-\frac{|c|}{\epsilon} + a)x^2 + o(x^2+y^2) \leq u(x,y) \leq px + (b+\epsilon)y^2 + (\frac{|c|}{\epsilon} + a)x^2 + o(x^2+y^2).$$

2. Let $f(x,y) = g((x-x_0)^2 + y^2)$ be a radial function, show by an asymtotic expansion at $(0,0)$ that

$$f(x,y) = g(x_0^2) - 2x_0g'(x_0^2)x + \frac{1}{2}((4g''(x_0^2)x_0^2 + 2g'(x_0^2))x^2 + 2g'(x_0^2)y^2).$$

3. Conclude.

5.6 The principal curvatures of a level surface

Following the same line as in the preceding section, we shall now define differential operators of u invariant with respect to contrast changes, which we later on shall interpret as the "principal curvatures" of the level surface of u. In the following, when $\mathbf{z}$ is a vector of $\mathbb{R}^N$, we denote by $\mathbf{z}^\perp$ the hyperplane orthogonal to $\mathbf{z}$, $\mathbf{z}^\perp = \{\mathbf{y}, \mathbf{z}.\mathbf{y} = 0\}$.

Proposition 5.13 *Let $u(\mathbf{x})$ be a C^2 function in a neighborhood of a point $\mathbf{x}_0$. Assume that $Du(\mathbf{x}_0) \neq 0$ and consider the symmetric matrix $D^2u(\mathbf{x}_0)$, which we write for simplicity D^2u. Then the eigenvectors and eigenvalues of the restriction of D^2u to $Du^\perp$ are not altered by a C^2 contrast change $u \to g(u)$ satisfying $g'(s) > 0$.*

Proof If $\mathbf{y} \in \mathbb{R}^N$, we denote by $\mathbf{y} \otimes \mathbf{y}$ the linear map $\mathbf{x} \to (\mathbf{y} \otimes \mathbf{y})(\mathbf{x}) = (\mathbf{x}.\mathbf{y})\mathbf{y}$ whose range is $\mathbb{R}\mathbf{y}$. Its matrix on the canonical basis of $\mathbb{R}^N$ is $(y_i y_j)_{1 \leq i,j \leq N}$. An easy application of the chain rule shows that

$$D(g(u)) = g'(u)Du \text{ and}$$

$$D^2(g(u)) = g''(u)Du \otimes Du + g'(u)D^2u.$$

This implies first that $Du^\perp = D(g(u))^\perp$ (we have assumed $\forall s, g'(s) \neq 0$). If $\mathbf{y} \in Du^\perp$, then $(Du \otimes Du)(\mathbf{y}) = 0$ and therefore $D^2(g(u))(\mathbf{y}, \mathbf{y}) = g'(u)D^2u(\mathbf{y}, \mathbf{y})$. Thus, the matrices D^2u and $D^2(g(u))$ are proportional on $Du^\perp = D(g(u))^\perp$ and their eigenvectors and eigenvalues are equal. □

Exercise 5.6 *By computing explicitly the coordinates $\frac{\partial}{\partial x_i}(g(u))$ of $D(g(u))$, check that $D(g(u)) = g'(u)Du$. By computing explictly the values $\frac{\partial^2}{\partial x_i \partial x_j}(g(u))$, check the formula used in the preceding proof, $D^2(g(u)) = g''(u)Du \otimes Du + g'(u)D^2u$.*

We now proceed to define locally the level surface of a smooth function u.

Theorem 5.14 (Implicit function theorem) *[]*
Let $u(\mathbf{x}) = u(x_1, x_2, ..., x_N)$ be a C^1 function in a neighborhood a point $\mathbf{x}_0$. Assume that $Du(\mathbf{x}_0) \neq 0$. Set $\vec{i}_N = \frac{Du}{|Du|}(\mathbf{x}_0)$ and $\mathbf{x} = \mathbf{x}_0 + x_N \vec{i}_N + \mathbf{y}$, where $\mathbf{y} \in Du(\mathbf{x}_0)^\perp$. Then there exists a ball $B(\mathbf{x}_0, \eta)$ and a C^1 function $\varphi(\mathbf{y}) : \mathbf{y} \in B(\mathbf{x}_0, \eta) \cap \{\mathbf{x}, x_N = 0\} \to \varphi(\mathbf{y}) \in \mathbb{R}$ such that for every $\mathbf{x} = \mathbf{x}_0 + \mathbf{y} + x_N \vec{i}_N \in B(\mathbf{x}_0, \eta)$,

$$u(\mathbf{x}) = 0 \Leftrightarrow x_N = \varphi(\mathbf{y})$$

In other terms, the equation $x_N = \varphi(\mathbf{y})$ describes the set $\{u(\mathbf{x}) = 0\} \cap B(\mathbf{x}_0, \eta)$ as the graph of a C^2 function φ. Thus, we obtain a surface which we call "level surface of u around $\mathbf{x}_0$". If u is C^2, then so is φ.

As in Sections 5.2, 5.3, we can define a topographic map for C^1 images on a compact hyperrectangle. We begin by eliminating all λ's which are singular (i.e. $Du(\mathbf{x}) = 0$ for some $\mathbf{x} \in u^{-1}(\lambda)$), and then decompose $u^{-1}(\lambda)$ into a finite set of C^1 embedded oriented manifolds. The topographic map is thus uniquely defined and gives a complete representation of u, from which u can be reconstructed.

By combining Proposition 5.13 and Theorem 5.14, we can do two useful things : First, to give a simple intrinsic form to the level surface around $\mathbf{x}_0$ and second to interpret the eigenvalues introduced in Proposition 5.13 as curvatures of lines drawn the level surface of u.

Proposition 5.15 *Let u be a C^2 function around a point $\mathbf{x}_0 \in \mathbb{R}^N$. Assume that $p = |Du(\mathbf{x}_0)| \neq 0$ and denote by $\kappa_1, ..., \kappa_{N-1}$ the eigenvalues of the restriction of $D^2u(\mathbf{x}_0)$ to $Du(\mathbf{x}_0)^\perp$. Then, setting $\vec{i}_N = \frac{Du}{|Du|}(\mathbf{x}_0)$ and choosing $\vec{i}_1, ..., \vec{i}_{N-1}$ to be an orthonormal basis of eigenvectors of the restriction of $D^2u(\mathbf{x}_0)$ to $Du(\mathbf{x}_0)^\perp$, we have for $\mathbf{x} = \mathbf{x}_0 + x_1\vec{i}_1 + ... + x_N\vec{i}_N = \mathbf{x}_0 + \mathbf{y}$ small enough :*

$$u(\mathbf{x}) = 0 \Leftrightarrow x_N = -\frac{1}{2p}\sum_{i=1}^{N-1} \kappa_i x_i^2 + o(|\mathbf{y}|^2).$$

Proof Fixing the origin at $\mathbf{x}_0$, assuming without loss of generality that $u(0) = 0$ and using the Taylor expansion formula, we obtain

$$u(\mathbf{x}) = Du(0).\mathbf{x} + \frac{1}{2}D^2u(0)(\mathbf{x}, \mathbf{x}) + o(|\mathbf{x}|^2).$$

Setting $\mathbf{x} = x_N\vec{i}_N + \mathbf{y}$, where $\mathbf{y} \in Du(0)^\perp$, this can be written

$$u(\mathbf{x}) = px_N + \frac{1}{2}D^2u(0)(\mathbf{y}, \mathbf{y}) + x_N D^2u(0)(i_N, \mathbf{y}) + \frac{1}{2}x_N^2 D^2u(0)(i_N, i_N) + o(|\mathbf{x}|^2).$$

If $\mathbf{x}$ belongs to the level surface $\{u(\mathbf{x}) = 0\}$ and $|\mathbf{x}| \leq \eta$, we deduce that $x_N^2 = o(|\mathbf{x}|^2)$ and by Theorem 5.14 that

$$x_N = \varphi(\mathbf{y}) = -\frac{1}{p}(\frac{1}{2}D^2u(0)(\mathbf{y}, \mathbf{y}) + x_N D^2u(0)(i_N, \mathbf{y})) + o(|\mathbf{x}|^2).$$

This implies that $x_N = o(|\mathbf{y}|^2)$ and therefore by substitution $o(|\mathbf{x}|^2) = o(|\mathbf{y}|^2)$, so that

$$x_N = \frac{-1}{2p}D^2u(0)(\mathbf{y}, \mathbf{y}) + o(|\mathbf{y}|^2). \tag{5.13}$$

We have $\mathbf{y} \in \{x_N = 0\} \cap B(0, \eta) = B(0, \eta) \cap Du(0)^\perp$. Thus, using an orthonormal basis of eigenvectors of the restriction of $D^2u(0)$ to $Du(0)^\perp$, and calling x_i, $i = 1, ..., N-1$ the coordinates on this orthonormal basis of $Du(0)^\perp$,

$$x_N = \frac{-1}{2p}\sum_{i=1}^{N-1} \kappa_i x_i^2 + o(|\mathbf{y}|^2).$$

□

Let us now use our two-dimensional analysis and interpret the κ_i as curvatures.

Proposition 5.16 and definition of the principal curvatures. *Let u be a C^2 function around $\mathbf{x}_0$ such that $Du(\mathbf{x}_0) \neq 0$. Set again $\vec{i}_N = \frac{Du}{|Du|}(\mathbf{x}_0)$. Then for every unit vector $\vec{\nu}$ orthogonal to $Du(\mathbf{x}_0)$, the intersection of the level surface $\{u(\mathbf{x}) = 0\}$ with the plane $\{x_N\vec{i}_N + t\vec{\nu}, (x_N, t) \in \mathbb{R}^2\}$, is a C^2 curve for $|(x_N, t)| \leq \eta$. The curvature of this curve at $\mathbf{x}_0$ is*

$$\kappa_\nu = \frac{1}{|Du(\mathbf{x}_0)|}D^2u(\vec{\nu}, \vec{\nu}).$$

In the case where $\vec{\nu} = \vec{i}_j$ is an eigenvector of $D^2u(\mathbf{x}_0)$ restricted to $Du(\mathbf{x}_0)^\perp$, this curvature κ_j is called a principal curvature of the level surface of u at $\mathbf{x}_0$.

Proof We apply Formula (5.13) with $\mathbf{x} = x_N \vec{i}_N + t\vec{\nu} = \varphi(t\vec{\nu})\vec{i}_N + t\vec{\nu}$. Thus

$$x_N = \frac{-1}{2|Du(\mathbf{x}_0)|} D^2u(\mathbf{x}_0)(\vec{\nu},\vec{\nu})t^2 + o(t^2).$$

The curvature at 0 of this two-dimensional curve is

$$\kappa(u) = \frac{1}{|Du(\mathbf{x}_0)|} D^2u(\mathbf{x}_0)(\vec{\nu},\vec{\nu}).$$

□

Proposition 5.17 and Definition of The Mean Curvature in dimension N *The "Mean Curvature" is defined as the sum of the principal curvatures. It is also given by*

$$curv(u) = |Du| div(Du/|Du|)$$

Proof By the preceding proposition and since Δu is the sum of the eigenvalues of D^2u, it is easily checked that

$$curv(u) = |Du|\Delta u - D^2(Du, Du)/|Du|$$

Now, with $Du = (u_{x_0}, ..., u_{x_N})$, we have

$$div(Du/|Du|) = \sum_i \frac{\partial}{\partial x_i} \frac{u_{x_i}}{|Du|}$$

$$= \sum_i \frac{1}{|Du|^2} \frac{\partial^2}{\partial x_i^2} u - (\frac{\partial}{\partial x_i}|Du|) u_{x_i}/|Du|$$

Combining all terms, we obtain

$$div(Du/|Du|) = \frac{\Delta u}{|Du|} - \frac{D^2u(Du, Du)}{|Du|}$$

□

5.7 Visualization of the topographic map.

In pratice, the digital image is described by a matrix of values $u(i,j), 0 \leq i \leq n, 0 \leq j \leq m$. Those values are assumed to be a sampling of a continuous image $\tilde{u}(x,y)$ such that $\tilde{u}(i,j) = u(i,j)$, at each location. The function $\tilde{u}$ is generally given by Shannon theory ([277]). Now, since this Shannon interpretation can be heavy to implement and since we do not need such precision for a visualization, we shall in this book just display level lines as concatenations of vertical and horizontal lines bounding the pixels. This involves a zoom of factor 2 of the image size.

Defining a function $\phi_\lambda : I\!R \times I\!R \to \{0, 255\}$ by

$$\begin{array}{ll} \phi_\lambda(a,b) = 0 & \text{if } a < \lambda \text{ and } b \geq \lambda \\ \phi_\lambda(a,b) = 0 & \text{if } a \geq \lambda \text{ and } b < \lambda \\ \phi_\lambda(a,b) = 255 & \text{otherwise} \end{array},$$

and, given λ and a digital image $u(i,j)$, we define the image of the level lines at level λ of u by:

$$\begin{aligned} w(2i,2j) &= 255 \\ w(2i+1,2j) &= \phi_\lambda(u(i+1,j),u(i,j)) \\ w(2i+1,2j+1) &= \phi_\lambda(u(i+1,j+1),u(i,j)) \\ w(2i,2j+1) &= \phi_\lambda(u(i,j+1),u(i,j)). \end{aligned}$$

Using the code 0 =black and 255 =white, the level line λ of u will appear as a black line on a white background in image w. Of course, we can superpose several levels in order to obtain a joint map of several level lines, i.e. a topographic map. In order to be visually readable, this topographic map must not show too many levels ; in practice not more than about twenty. The removal of small level lines can be very useful in order to improve the readability of the topographic map and we shall discuss filters performing this task in the last section about the "extrema killer" in Chapter 7.

References.

Section 5.1 : For elementary presentations of the basic differential calculus used in this section, we refer to the books Cartan [76] and Spivak [391] where proofs of the Sard and implicit function theorems can be found. An introduction to the use of level lines in Computer Vision can be found in Caselles et al. [85].

Section ?? :

The Jordan curve theorem is proved (e.g.) in Rudin [360]. The reconstruction formula (**??**) permitting to rebuild a set from its boundary, when this boundary is made of a fine set of Jordan curves, is stated here in an elementary framework where the Jordan curve theorem is the only needed tool. A more general construction assumes that the set to be reconstructed from its boundary is a set with finite perimeter. Then it can be shown that its boundary can be defined in a unique way as a countable union of Jordan curves and the reconstruction formula is still valid. A complete definition and proof of this general reconstruction formula, is given in [24].

Sections ?? : and **??** : It is a classical mathematical trick to define implictly a set as the the zero set of its distance function. In the case where the set is a curve, one can compute the curvature of a curve at a point $\mathbf{x}$ by computing the curvature $curv(u)(\mathbf{x})$ at the same point of the distance function to the curve. This yields an intrinsic formula for the curvature, not linked to a parameterization of the curve. The same trick has been refreshed in recent years as a very useful numerical tool. This started with Barles report [**?**] on flame propagation and was extended by Sethian [381] and Osher-Sethian [331] in a series of papers on the numerical simulation of the motion of surface by its mean curvature.

Chapter 6

The main contrast invariant equations.

6.1 The normal and curvature of a shape as cues to recognition

In this section, we shall link shape recognition and scale space theory. Let us define a "shape" as a closed bounded subset of $\mathbb{R}^2$. Without loss of generality, we shall assume that this subset, X, has a boundary made of a finite set of simple closed curves (Jordan curves) with finite length. This finiteness assumption is computationally realistic and implies that X is a Cacciopoli set. The mathematician Renato Cacciopoli [?] proposed a theory for sets whose boundary has finite length, from which it can be deduced that the boundary of a Cacciopoli set is made of an enumerable set of Jordan curves. We restrict this assumption to finiteness for obvious quantization reasons. The main consequence of this model for shapes is that they are fully described by the finite list of their boundary Jordan curves. The shape recognition problem can therefore be reduced to the problem of recognition of the shape building elements : the Jordan curves. Thus, we shall in the following assume that the shape undergoing a recognition process is described by a single Jordan curve. In photographs of a natural environment, most observed shapes are partially occluded (hidden) and distorted by perspective deformations. The shape recognition technology has therefore focused on local methods, that is, methods which will work even if the shape is not fully in view and if the visible parts are distorted. When we observe a Jordan curve, we can therefore not be sure whether is belongs to one and the same object ; it may be the concatenation of several objects occluding each other. *Thus, recognition has to be based on local features of the Jordan curve and not on global one. If the shape boundary is in some degree smooth, those local features simply are derivatives of the curve, that is : orientation, curvature, etc..* Before starting with formal definitions of these geometric quantities, let us mention that most local recognition methods indeed involve the "salient" points of the shapes, identified with inflexion points (where the curvature is zero) and extrema of curvatures (the "corners" of the shape). In such local recognition methods, a shape is reduced to a finite code, namely the coordinates of a set of *characteristic points*, corners and inflexion point mainly. The shape recognition is then reduced to a comparison between finite shape codes by "voting methods" like *geometric hashing*. Let us now return to what will be our main concern here, the definition and effective computation of the normal and curvature to a shape.

6.2 Multiscale features and scale space

The above mentionned methods (and, as a matter of fact, all nonglobal computational shape recognition methods) make two basic assumptions, none of which is in practice true for the analyzed shapes (identified from now on with Jordan curves) :

- The shape has a smooth boundary
- The shape has a finite set of inflexion points and extrema of curvature.

The fact that this assumption can however become a reality has been mathematically proved in 1986-87 by Gage-Hamilton [193] and Grayson [180] who proved the possibility of smoothing a Jordan curve into a Jordan curve by the so called "intrinsic heat equation".

Definition 6.1 *Let $\mathbf{x}(t,s)$ be a family of Jordan C^2 curves and assume for each t that $s \to \mathbf{x}(t,s)$ is an euclidean parameterization. We say that $\mathbf{x}(t,s)$ satisfies the* intrinsic heat equation *if*

$$\frac{\partial \mathbf{x}}{\partial t}(t,s) = \frac{\partial^2 \mathbf{x}}{\partial s^2}(t,s) = Curv\ \mathbf{x}(t,s). \tag{6.1}$$

Theorem 6.2 *(Grayson) The intrinsic heat equation lets any Jordan curve $\mathbf{x}(s)$ with finite length evolve into a family of Jordan curves $\mathbf{x}(t,s)$ such that $\mathbf{x}(0,s) = \mathbf{x}_0(s)$, the curve $\mathbf{x}(t,s)$ being for every positive scale t a C^∞ Jordan curve, with a finite set of inflexion points and extrema curvature. The number of these "characteristic points" is nonincreasing with t. In addition, there is for every initial Jordan curve a scale t_0 such that the curve $s \to \mathbf{x}(t_0,s)$ becomes convex and another scale t_1 for which the curve shrinks to a point.*

Of course, the initial shape $\mathbf{x}_0(s)$ may well have infinitely many characteristic points and may even have unbounded curvature everywhere. The point is, that anyway $\mathbf{x}(t,s)$ has derivatives of any order for $t > 0$ and the curvature has finitely many zero-crossings and extrema for every fixed $t > 0$. Still better : the number of characteristic points cannot but decrease when the "time parameter" t increases, so that the scale space is *causal* according to the Vision theory terminology. The discovery of Gage-Hamilton and Grayson has found a significant response in the Vision Research, where several attempts to define *scale spaces* for shapes had come very close to the equation studied by Gage-Hamilton and Grayson. We call *shape scale space* any method allowing to smooth a shape, the degree of smoothing being measured by a positive real parameter t, the *scale*. Thus scale space associates with every initial Jordan curve $\mathbf{x}(0,s) = \mathbf{x}_0(s)$ a series of smoothed shapes, $\mathbf{x}(t,s)$. It is expected that the intrinsic heat equation acts as a selective way of eliminating spurious details of the shape, keeping a rougher but more reliable version of the shape and a shorter code as well. The matching of two instances $\mathbf{x}_0(s)$ and $\mathbf{x}_1(s)$ of a given shape should be done by comparing the characteristic points of $\mathbf{x}_0(t,s)$ and $\mathbf{x}_1(t,s)$ for a value of t large enough. Comparing the originals $\mathbf{x}_0$ and $\mathbf{x}_1$ is simply impossible : They are expected to depend highly upon the conditions of observation. Observation noise and distorsions of the initial Jordan curve make it impossible to single out directly on $\mathbf{x}_0(s)$ the significant corners and inflexion points.

6.3 From image motion to curve motion.

6.3.1 Level lines as curves

Let u be a real function which is twice differentiable (C^2) in a neighborhood of $\mathbf{x}$. We denote by $Du(\mathbf{x}) = (u_x, u_y)(\mathbf{x})$ the gradient of u at $\mathbf{x}$, by $|Du(\mathbf{x})| = (u_x^2 + u_y^2)^{\frac{1}{2}}(\mathbf{x})$ its euclidean norm. We consider the level curve of u passing by $\mathbf{x}$, defined as the set

$$\{\mathbf{y}, u(\mathbf{y}) = u(\mathbf{x})\}$$

If we assume that $Du(\mathbf{x}) \neq 0$, by the Implicit Function Theorem, we know that the level curve indeed is a curve, which inherits the smoothness of u in a neighborhood of $\mathbf{x}$. We set $Du^\perp = (-u_y, u_x)$, a vector orthogonal to Du. The next lemma links the intrinsic curvature of the level curve $Curv(\mathbf{x})$ introduced in Definition 5.2 and the curvature differential operator $\mathrm{curv}(u)(\mathbf{x})$ of Definition 5.12.

Lemma 6.3 *Let $u(\mathbf{x})$ be a C^2 real function : $\mathbb{R}^2 \to \mathbb{R}$ and $\mathbf{x}$ a point such that $Du(\mathbf{x}) \neq 0$. Call C the level line passing by $\mathbf{x}$. Then*

$$Curv(\mathbf{x}) = -\mathrm{curv}(u)(\mathbf{x})\frac{Du}{|Du|}(\mathbf{x}). \tag{6.2}$$

Proof Consider a level curve $\{\mathbf{y}, u(\mathbf{y}) = u(\mathbf{x})\}$. We notice that $\frac{Du^\perp}{|Du|}$ is a unit vector tangent to the curve at $\mathbf{x}$. Then there is by Lemma 5.2 a single euclidean parameterization $\mathbf{y}(s)$ of the level curve such that

$$\frac{\partial \mathbf{y}(s)}{\partial s}(0) = \frac{Du^\perp}{|Du|}(\mathbf{x}) \tag{6.3}$$

Differentiating the relation $u(\mathbf{y}(s)) = 0$ we have $Du(\mathbf{y}(s)).\frac{\partial \mathbf{y}(s)}{\partial s} = 0$ and differentiating again at $s = 0$ and using (6.3),

$$D^2u(\frac{Du^\perp}{|Du|}, \frac{Du^\perp}{|Du|}) + Du(\mathbf{x}).\frac{\partial^2 \mathbf{y}}{\partial s^2} = 0.$$

Using the definitions 5.12 and 5.2 of $\mathrm{curv}(u)$ and $Curv(\mathbf{x})$, this becomes

$$\mathrm{curv}(u)|Du|(\mathbf{x}) + Du(\mathbf{x}).Curv(\mathbf{x}) = 0.$$

Noticing that $Curv(\mathbf{x})$ is colinear to $Du(\mathbf{x})$, we deduce that

$$Curv(\mathbf{x}) = -\mathrm{curv}(u)\frac{Du}{|Du|}.$$

□

The intrinsic heat equation is only one instance of a wide range of nonlinear equations which move curves with a curvature-dependent speed. The only requirement for this dependence, which we shall justify in Chapter 22 is that the higher the curvature, the higher the speed:

Definition 6.4 *Let $\mathbf{x}(t, s)$ be a family of Jordan C^2 curves and assume for each t that $s \to \mathbf{x}(t, s)$ is an euclidean parameterization. We say that $\mathbf{x}(t, s)$ satisfies a* curvature equation *if it is C^2 and for some real function $g(\kappa, t)$, nondecreasing with respect to κ satisfies*

$$\frac{\partial \mathbf{x}}{\partial t}(t, s) = g(t, |Curv\mathbf{x}(t, s)|)\mathbf{n}(t, s). \tag{6.4}$$

Definition 6.5 *We say that $u(t,\mathbf{x})$ satisfies a* curvature equation *if it is C^2 and for some real function $g(\kappa,t)$, nondecreasing with respect to κ satisfies*

$$\frac{\partial u}{\partial t}u(t,\mathbf{x}) = g(\text{curv}u,t)|Du|(t,\mathbf{x}). \tag{6.5}$$

All of these equations are candidates to be curve or image scale spaces and one of our main tasks in this book is precisely to single out the most relevant ones. The definitions above are very restrictive in that they assume a C^2 regularity of the image or the curve. Now, as we shall see, this regularity requirement impossible to attain. A more general and applicable definition of solutions for the above image equations will be introduced in Chapter 17 on viscosity solutions. In the next section, we establish a formal link between curve motions and curvature motions for images.

6.3.2 Curvature equations for curves and real functions

Definition 6.6 *Let u(t,x) be C^2 function around a point $(t_0,\mathbf{x}_0)$ such that $Du(t_0,\mathbf{x}_0)\neq 0$. Then there exists a neighborhood V of $\mathbf{x}_0$, a time $t_1 > t_0$ and for $t\in[t_0,t_1]$ and $\tilde{\mathbf{x}}\in V$ a unique C^1 vector function $\mathbf{x}(t,\tilde{\mathbf{x}})$ satisfying $\mathbf{x}(0,\tilde{\mathbf{x}})=\tilde{\mathbf{x}}$, $u(t,\mathbf{x}(t,\tilde{\mathbf{x}})) = u(t_0,\tilde{\mathbf{x}})$ and such that $\frac{\partial \mathbf{x}(t,\tilde{\mathbf{x}})}{\partial t}$ is colinear to $Du(t,\mathbf{x}(t,\tilde{\mathbf{x}}))$. We call $\mathbf{x}(t,\tilde{\mathbf{x}})$ the "normal flow starting from $(t_0,\tilde{\mathbf{x}})$". The normal flow satisfies the ordinary differential equation*

$$\frac{\partial \mathbf{x}}{\partial t} = -(\frac{\partial u}{\partial t}\frac{Du}{|Du|^2})(t,\mathbf{x}(t)). \tag{6.6}$$

Proof Differentiating the relation $u(t,\mathbf{x}(t))=0$ with respect to t yields $\frac{\partial u}{\partial t} + Du.\frac{\partial \mathbf{x}}{\partial t}=0$. Thus $\frac{\partial \mathbf{x}}{\partial t}$ is colinear to Du if and only if

$$\frac{\partial \mathbf{x}}{\partial t} = -(\frac{\partial u}{\partial t}\frac{Du}{|Du|^2})(t,\mathbf{x}(t)). \tag{6.7}$$

This relation defines $\mathbf{x}(t)$ as the solution of an ordinary differential equation, with initial condition $\mathbf{x}(0)=\tilde{\mathbf{x}}$. Since u is C^2, the second member of (6.7) appears to be a Lipschitz function of $(t,\mathbf{x})$ provided $Du(t,\mathbf{x})\neq 0$, which is ensured for $(t,\mathbf{x})$ close enough to $(t_0,\mathbf{x}_0)$. Thus, by Cauchy-Lipschitz Theorem, there exists $t_1 > t_0$ such that the O.D.E. (6.7) has a unique solution $\mathbf{x}(t,\tilde{\mathbf{x}})$ for all $\tilde{\mathbf{x}}$ in a neighborhood of $\mathbf{x}_0$ and $t\in[t_0,t_1]$. □

Proposition 6.7 *Let $u(t,\mathbf{x})$ be a C^2 function such that $Du(t_0,\mathbf{x}_0)\neq 0$. Then u satisfies the curvature motion equation*

$$\frac{\partial u}{\partial t} = \text{curv}(u)|Du| \tag{6.8}$$

in a neighborhood of $(t_0,\mathbf{x}_0)$ if and only if the normal flow $\mathbf{x}(t)$ satisfies the intrinsic heat equation (6.1),

$$\frac{\partial \mathbf{x}}{\partial t}(t,s) = \frac{\partial^2 \mathbf{x}}{\partial s^2}(t,s) = Curv(\mathbf{x}(t,s)).$$

Proof If (6.1) holds, by (6.6), we get

$$-(\frac{\partial u}{\partial t}\frac{Du}{|Du|^2})(t,\mathbf{x}(t)) = Curv(\mathbf{x}(t)).$$

By (6.2), this yields

$$-(\frac{\partial u}{\partial t}\frac{Du}{|Du|^2})(t,\mathbf{x}(t)) = \mathrm{curv}(u)\frac{Du}{|Du|},$$

which implies the curvature motion equation (6.8). Conversely, substituting in (6.6) the value of $\frac{\partial u}{\partial t}$ given by (6.8) yields

$$\frac{\partial \mathbf{x}}{\partial t} = -(\mathrm{curv}(u)|Du|\frac{Du}{|Du|^2})$$

and using (6.2) we obtain the heat intrinsic equation (6.1), $\frac{\partial \mathbf{x}}{\partial t}(t,s) = Curv(\mathbf{x}(t))$. □

Corollary 6.8 *Let $u(t,\mathbf{x})$ be a C^2 function such that $Du(t_0,\mathbf{x}_0) \neq 0$ and $g(\kappa,t)$ a continuous function, nondecreasing with respect to κ. Then u satisfies the curvature motion equation (6.5)*

$$\frac{\partial u}{\partial t}(t,\mathbf{x}) = g(\mathrm{curv}(u),t)|Du|(t,\mathbf{x})$$

in a neighborhood of $(t_0,\mathbf{x}_0)$ if and only if the normal flow $\mathbf{x}(t)$ satisfies the intrinsic heat equation (6.4) (introduced in Definition 6.4).

$$\frac{\partial \mathbf{x}}{\partial t}(t,s) = g(t,|Curv\mathbf{x}(t,s)|)\mathbf{n}(t,s)$$

6.3.3 Introduction to affine curve and image equations

A particular instance of curvature equation is, as we shall see, affine invariant and therefore well suited to be used in shape recognition.

Definition 6.9 *We call "affine morphological scale space" (AMSS) the image evolution equation*

$$\frac{\partial u}{\partial t} = \mathrm{curv}(u)^{\frac{1}{3}}|Du|(t,\mathbf{x}), \tag{6.9}$$

where $k^{\frac{1}{3}}$ means $sgn(k)|k|^{\frac{1}{3}}$. We call "affine scale space" (ASS) the curve evolution equation

$$\frac{\partial \mathbf{x}}{\partial t}(t,s) = |Curv(\mathbf{x}(t,s))|^{\frac{1}{3}}\mathbf{n}(t,s) \quad \left(\quad = \frac{Curv(\mathbf{x}(t,s)}{|Curv(\mathbf{x}(t,s))^{\frac{2}{3}}} \quad \right). \tag{6.10}$$

It is easily seen that both considered equations are formally equivalent in the sense of Corollary 6.8. Let us now briefly give a first explanation for the "affine" denomination. A second explanation will be given in the next subsection.

Definition 6.10 *We say that an image evolution equation is* affine invariant *if for every linear map A with positive determinant, there is a constant $c(A)$ depending only on A such that $u(t,\mathbf{x})$ is a solution of the curvature equation if and only $u(c(A),A\mathbf{x})$ is.*

This definition is a bit puzzling but will be fully explained in Section 20.4. In Exercises 6.1, 6.2, we check that the AMSS and ASS equations indeed are affine invariant.

Exercise 6.1 *(i) Let u be a C^2 function, A a 2×2 matrix and set $v(\mathbf{x}) = u(A\mathbf{x})$. Prove that at each point $\mathbf{x}$ such that $Dv(\mathbf{x}) \neq 0$, one has* $\mathrm{curv}v(\mathbf{x})|Dv(\mathbf{x})| = |detA|^{\frac{1}{2}}curv(u)(A\mathbf{x})$.
(ii) Deduce that the AMSS equation 6.9 is affine invariant, namely u is solution of (AMSS) if and only if $v(\mathbf{x}) = u(A\mathbf{x})$ is solution of

$$\frac{\partial v}{\partial t} = (detA)^{\frac{2}{3}}\mathrm{curv}(v)^{\frac{1}{3}}|Dv|(t,\mathbf{x}).$$

Exercise 6.2 *Affine invariance of the affine scale space (ASS).*

From Exercise 5.2, we know that if $\mathbf{c}(\sigma)$ *is parametric* C^2 *curve with* $|\mathbf{c}'(\sigma)| > 0$*, then by Formula (5.5),*

$$Curv(\mathbf{c}(\sigma)) = x''(s) = \frac{1}{\mathbf{c}'^2(\sigma)}(\mathbf{c}''(\sigma) - (\mathbf{c}''(\sigma).\frac{\mathbf{c}'(\sigma)}{|\mathbf{c}'(\sigma)|})\frac{\mathbf{c}'(\sigma)}{|\mathbf{c}'(\sigma)|}). \tag{6.11}$$

Using this formula, and projecting both members of the intrinsic heat equation on the normal vector $\mathbf{c}'(\sigma)^\perp$*, we shall say that* $\mathbf{c}(\sigma)$ *satisfies a* curvature equation *if*

$$\frac{\partial \mathbf{c}}{\partial t}.\frac{\mathbf{c}'^\perp}{|\mathbf{c}'|} = \frac{\mathbf{c}''.\mathbf{c}'}{|\mathbf{c}'|^3}. \tag{6.12}$$

(i) Show that if $\mathbf{c}(\sigma)$ *is an euclidean parameterization, then this definition coincides with the usual one, (6.1). In the same way, we shall say that* $\mathbf{c}$ *satisfies an affine equation if for some positive constant* $\gamma > 0$*,*

$$\frac{\partial \mathbf{c}}{\partial t}.\mathbf{c}'^\perp = (\mathbf{c}''.\mathbf{c}')^{\frac{1}{3}}. \tag{6.13}$$

($a^{\frac{1}{3}}$ *denotes* $sgn(a)|a|^{\frac{1}{3}}$*). Let* A *be a* 2×2 *matrix with positive determinant. Our aim is to check that if* $\mathbf{c}$ *satisfies an affine motion, then so does* $\mathbf{y}(\sigma) = A\mathbf{c}(\sigma)$*.*

(ii) Let $\mathbf{x}, \mathbf{y} \in I\!R^2$*. Show that* $A\mathbf{x}.(A\mathbf{y})^\perp = (\det A)\mathbf{x}.\mathbf{y}^\perp$ *and* $A(\mathbf{x}^\perp).(A\mathbf{x})^\perp = (\det A)|\mathbf{x}|^2$*.*

(iii) Deduce the announced affine invariance. More precisely, show that if $\mathbf{c}(\sigma)$ *satisfies (6.13), then* $\mathbf{y}(\sigma) = A\mathbf{c}(\sigma)$ *satisfies*

$$\frac{\partial \mathbf{y}}{\partial t}.\mathbf{y}'^\perp = (detA)^{\frac{2}{3}}(\mathbf{y}''.\mathbf{y}')^{\frac{1}{3}}. \tag{6.14}$$

6.3.4 The Affine Scale Space as an intrinsic heat equation

Denote by $\mathbf{x}(t, \sigma)$ a curve at scale t, parametrized by a parameter σ. In this paragraph, we shall show a formal equivalence between the Affine Scale Space,

$$\frac{\partial \mathbf{x}}{\partial t} = |Curv(\mathbf{x})|^{\frac{1}{3}}\mathbf{n}(\mathbf{x}) = |k|^{\frac{1}{3}}\mathbf{n}(\mathbf{x}), \tag{6.15}$$

where $k = |Curv(\mathbf{x})|(\mathbf{x})$ is a useful abbreviation for the curvature, and an "intrinsic heat equation",

$$\frac{\partial \mathbf{x}}{\partial t} = \frac{\partial^2 \mathbf{x}}{\partial \sigma^2} \tag{6.16}$$

where σ is a special parameterization of the curve called affine length. Let us give some elements of affine differential geometry. We define an *affine length parameter*, or σ of a Jordan curve as a parameterization $\mathbf{x}(\sigma)$ satisfying

$$[\mathbf{x}_\sigma, \mathbf{x}_{\sigma\sigma}] = 1 \tag{6.17}$$

where $[,]$ stands for the determinant of two vectors. Here and in the following, we write for clarity of the formulas $\mathbf{x}_\sigma$ for $\frac{\partial \mathbf{x}}{\partial \sigma}$ and $\mathbf{x}_{\sigma\sigma}$ for $\frac{\partial^2 \mathbf{x}}{\partial \sigma^2}$. If we denote by s the arc length, and by $k = |Curv(\mathbf{x})|(\mathbf{x})$ the curvature, then a tangent vector and the intrinsic normal vectors to the curve are given by

$$\mathbf{t} = \mathbf{x}_s \qquad \mathbf{n} = k^{-1}\mathbf{x}_{ss}\ (= \frac{Curv(\mathbf{x})}{|Curv(\mathbf{x})|} \tag{6.18}$$

Moreover

$$\mathbf{x}_\sigma = \mathbf{x}_s\frac{\partial s}{\partial \sigma} \qquad \text{and} \tag{6.19}$$

$$\mathbf{x}_{\sigma\sigma} = \mathbf{x}_{ss}(\frac{\partial s}{\partial \sigma})^2 + \mathbf{x}_s \frac{\partial^2 s}{\partial \sigma^2} \tag{6.20}$$

Now (6.17) implies

$$[\mathbf{x}_s \frac{\partial s}{\partial \sigma}, \mathbf{x}_{ss}(\frac{\partial s}{\partial \sigma})^2 + \mathbf{x}_s \frac{\partial^2 s}{\partial \sigma^2}] = 1$$
$$\Rightarrow \quad [\mathbf{x}_s, \mathbf{x}_{ss}](\frac{\partial s}{\partial \sigma})^3 = 1$$

Since by (6.18), $[\mathbf{x}_s, \mathbf{x}_{ss}] = k$, we conclude that

$$\frac{\partial s}{\partial \sigma} = k^{-\frac{1}{3}} \tag{6.21}$$

Let us now deal with $\mathbf{x}_{\sigma\sigma}$. Using (6.21) and (6.18) in (6.20) we obtain

$$\mathbf{x}_{\sigma\sigma} = k^{\frac{1}{3}}\mathbf{n} + (\frac{\partial^2 s}{\partial \sigma^2})\mathbf{t}$$

We deduce that equation (6.16) is equivalent to

$$\frac{\partial \mathbf{x}}{\partial t} = k^{\frac{1}{3}}\mathbf{n} + (\frac{\partial^2 s}{\partial \sigma^2})\mathbf{t} \tag{6.22}$$

Now, Epstein and Gage [135] have proved that the tangential component of such an equation does not matter as far as the geometric evolution of the curve is concerned. Indeed, the tangential component produces a motion of the point along the curve itself and the whole curve evolution is characterized defined by the normal velocity. As a consequence, (6.16) is equivalent to Equation (6.15) on non-inflection points. Thus, we have the following equivalence in terms of curve evolution :

$\frac{\partial \mathbf{x}}{\partial t} = k^{\frac{1}{3}}\mathbf{n}$ if and only if

$$\frac{\partial \mathbf{x}}{\partial t} = \begin{cases} 0 & \mathbf{x} \text{ inflection point} \\ \mathbf{x}_{\sigma\sigma} & \mathbf{x} \text{ non-inflection point} \end{cases}$$

6.4 Curvature Motions in dimension N.

We consider a flow $u(t, \mathbf{x})$ where $\mathbf{x}$ is of dimension N and where $u(0, \mathbf{x}) = u_0(\mathbf{x})$, an initial N dimensional image. We denote by $\kappa_i(u)(t, \mathbf{x})$ the i^{th} principal cuvature of u the image at point $(t, \mathbf{x})$. (Principal curvatures have been introduced section 5.6). (i is between 1 and $N-1$). Let us now introduce major equations describing some interesting flows $u(t, \mathbf{x})$.

The mean Curvature Motion

$$\frac{\partial u}{\partial t}(t, \mathbf{x}) = |Du| \sum_{i=1}^{N-1} \kappa_i(\mathbf{x})$$

With this equation, the motion of a level-hypersurface of u in the normal direction is proportional to the mean value of its principal curvatures.

The Gauss Curvature Motion of convex functions. We call convex functions, functions that have all their principal curvature with same sign. Example of such a function is the signed distance function to a regular convex shape.

$$\frac{\partial u}{\partial t}(t, \mathbf{x}) = |Du| \prod_{i=1}^{N-1} \kappa_i(\mathbf{x})$$

The motion of a level-hypersurface of u is now proportional to the product of its principal curvatures. However, this equation can not be applied, as is, on non convex functions due to the maximum principle (see Chapter 21).

The Affine Invariant Curvature Motion.

$$\frac{\partial u}{\partial t}(t,\mathbf{x}) = |Du|| \prod_{i=1}^{N-1} \kappa_i(\mathbf{x})|^{\frac{1}{N+1}} H\big(\sum_{1\leq i\leq N-1} sgn(\kappa_i)\big)$$

where $H(N-1)=1$, $H(-N+1)=-1$ and H null otherwise.

This motion is similar to the Gauss Curvature Motion. The affine invariance implies to use the power $\frac{1}{N+1}$ of the gaussian curvature, the maximum principle implies to cancel the motion at point where principal curvatures have different signs.

References.

Role of curvature in shape analysis

We can mention [9] as one of the first references dealing with the role of curvature in the representation and recognition of objects in Computer Vision. The classical paper proposing to compute a "multiscale curvature" in order to recognize objects in an image is Asada-Brady's [37], "The Curvature Primal Sketch". The title is an allusion to the famous David Marr's "raw primal sketch", a set of geometric primitives extracted from and representing the image. This paper entailed a long series of more and more sophisticated attempts to represent Shape from Curvature [128, 132] and to compute the curvature correctly [306].

Affine invariant curve shortening

Affine invariant geometry seems to have been founded by Blaschke [**?**], which contains definitions of the affine length and affine curvature. Curves with constant affine curvature are discussed in [273]. The name Affine Shortening and the corresponding curve evolution equation was introduced by Sapiro-Tannenbaum in [370]. Some mathematical properties were developped by the same authors in [371, 372] In [29], Angenent, Sapiro and Tannenbaum gave a first existence and uniqueness proof for the affine shortening.

Curve shortening The mathematical study of the intrinsic heat equation (or curvature motion in dimension 2) was made in a series of brillant papers in differential geometry whose titles sufficiently indicate the progress made : Gage [165, 166] ("Curve shortening makes convex curves circular"), Epstein-Gage [135], Gage-Hamilton [193] "The heat equation shrinking convex plane curves", "Given a curve in the plane, let it deform in the direction of its curvature vector. It was conjectured that the curve would collapse in finite time to a point and that, as the curve collapsed, it would become asymptotically close to a shrinking standard circle. The authors prove this conjecture for a convex plane curve. " Finally, Grayson [] whose paper "The heat equation shrinks embedded plane curves to round points" is commented in laconic terms by his referee: "This paper contains the final solution of the long-standing "curve-shortening problem" for plane curves."

The first papers introducing the curve shortening and some variants in image processing as means to do

a multiscale analysis of curves (understood as shapes extracted from an image) were written by Kimia, Tannenbaum and Zucker [243] and Mackworth-Mokhtarian who propose the use of curve shortening as an efficient numerical tool for shape multiscale analysis [272].

Mean curvature motion In his famous paper "Shapes of worn stones", Firey [150] proposed a model for the erosion of stones on a beach. The surface of those stones evolves at a normal speed proportional to the speed of the Gauss curvature of the surface, and it was conjectured that the final shape of the stones is spherical. The first attempt to make a mathematical definition of the mean curvature motion is Brakke [62]. We shall discuss in the next chapters the clever numerical implementation of the same equation by Sethian ([382] "Curvature and the evolution of fronts") In [416], a more general formulation of the mean curvature motion is proposed, which adapts to crystal growth and in general to the evolution of anisotropic solids.

Chapter 7

Monotone and contrast invariant operators. The threshold decomposition principle

7.1 Contrast invariant function operators

In the following, we denote by $\mathcal{F}$ a space of functions defined on $\mathbb{R}^N$ with values in $\mathbb{R}$ and denote by $\mathcal{T}$ the set of all level sets of functions of $\mathcal{F}$. If (e.g.) $\mathcal{F}$ is the set of continuous real functions on $\mathbb{R}^N$, then $\mathcal{T}$ is the set of closed subsets of $\mathbb{R}^N$. If $\mathcal{F}$ is the set of Borel-Lebesgue measurable functions then $\mathcal{T}$ is the set of Borel subsets of $\mathbb{R}^N$, etc. We assume that $\mathcal{F}$ is stable by continuous nondecreasing contrast changes $u \to g(u)$, i.e. $g(u) \in \mathcal{F}$ whenever $u \in \mathcal{F}$. We consider function operators on $\mathcal{F}$,

$$T : u \in \mathcal{F} \to Tu,$$

where Tu is a real function, $Tu(\mathbf{x}) \in \mathbb{R}$, and set operators on $\mathcal{T}$,

$$\mathtt{T} : X \in \mathcal{T} \to \mathtt{T}X.$$

Our main interest here is in *monotone* operators, because they are nonlinear generalizations of the linear smoothing by a nonnegative convolution kernel .

Definition 7.1 *We say that a function operator T is <u>monotone</u> if*

$$u \geq v \Rightarrow Tu \geq Tv. \tag{7.1}$$

We say that a set operator $\mathtt{T}$ *is monotone if*

$$X \subset Y \Rightarrow \mathtt{T}(X) \subset \mathtt{T}(Y) \tag{7.2}$$

Definition 7.2 *We call <u>continuous contrast change</u> any nondecreasing continuous function $g : \mathbb{R} \to \mathbb{R}$.*

Definition 7.3 *We say that T is <u>contrast invariant</u> on $\mathcal{F}$ if for every continuous contrast change $g : \mathbb{R} \to \mathbb{R}$, one has $g(u) \in \mathcal{F}$ whenever $u \in \mathcal{F}$ and*

$$\forall u \in \mathcal{F},\ g(Tu) = T(g(u)). \tag{7.3}$$

The next proposition puts into evidence one of the main properties yielded by contrast invariance : the conservation of grey levels. This conservation property implies that, in contrast to linear operators, contrast invariant operators do not create new grey levels in an image. This means in practice that *images filtered by a contrast invariant operator remain well contrasted.* Fronts of u, which correspond to gaps in the range of u, are preserved. In particular, binary images remain binary.
We denote by $R(u) \subset \mathbb{R}$ the range of a function u, that is, $R(u) = \{s \in \mathbb{R}, \exists \mathbf{x}, u(\mathbf{x}) = s.\}$

Proposition 7.4 *Let T be a contrast invariant operator. Then for every function u, $R(Tu) \subset \overline{Ru}$. In particular, if u attains a finite number of values, then Tu attains a subset of them.*

Proof We consider a continuous increasing function g such that $g(s) = s$ when $s \in \overline{Ru}$, $g(s) > s$ otherwise. As an example of such function using the distance to the set Ru, we can take

$$g(s) = s + \frac{1}{2}d(s, Ru),$$

where $d(s, X)$ denotes the distance of s to X. We have $d(s, Ru) = 0$ if and only $s \in \overline{(Ru)}$. Thus $g(s) = s$ if and only $s \in \overline{(Ru)}$. In particular, $g(u) = u$. Using this and the contrast invariance of T we have

$$Tu = T(g(u)) = g(Tu).$$

Thus $(Tu)(\mathbf{x})$ is for every $\mathbf{x}$ a fixed point of g and therefore $(Tu)(\mathbf{x})$ belongs to $\overline{Ru}$, as announced.

□

Another interesting property of contrast invariant and translation invariant operators : they do not increase the Lipschitz constant of functions. In fact, the property is much more general, being true for all translation invariant operators commuting with the addition of constants : $T(u + C) = Tu + C$.

Lemma 7.5 *Let T be a monotone, translation invariant operator commuting with the addition of constants. If $u(\mathbf{x})$ is a Lipschitz function on $\mathbb{R}^2$ then the function $Tu(\mathbf{x})$ is Lipschitz, with a lower or equal Lipschitz constant.*

Proof of lemma 7.5. Assume u has a Lipschitz constant equal to K. For any $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$, we have

$$u(\mathbf{y} + \mathbf{z}) - K|\mathbf{x} - \mathbf{y}| \leq u(\mathbf{x} + \mathbf{z}) \leq u(\mathbf{y} + \mathbf{z}) + K|\mathbf{x} - \mathbf{y}| \tag{7.4}$$

Since T is monotone, considering the above functions as functions of $\mathbf{z}$ applying them T and taking the value at 0 of the obtained functions, we obtain

$$T(u(\mathbf{y} + \mathbf{z}) - K|\mathbf{x} - \mathbf{y}|)(0) \leq Tu(\mathbf{x} + \mathbf{z})(0) \leq T(u(\mathbf{y} + \mathbf{z}) + K|\mathbf{x} - \mathbf{y}|))(0).$$

Notice that by the translation invariance, $T(u(\mathbf{y} + \mathbf{z}))(0) = (Tu)(\mathbf{y})$, etc. Using the commutation of T with the additions of constants and the translation invariance, we therefore obtain

$$(Tu)(\mathbf{y}) - K|\mathbf{x} - \mathbf{y}| \leq (Tu)(\mathbf{x}) \leq (Tu)(\mathbf{y}) + K|\mathbf{x} - \mathbf{y}|.$$

□

The preceding regularity preserving property of contrast invariant operators is generalized in Exercise 7.1.

Exercise 7.1 *We say that a function $u(\mathbf{x})$ is uniformly continuous if there exists a real nonnegative function $\varepsilon(s)$ such that $\varepsilon(s) \to 0$ when $s \to 0$ and for all $\mathbf{x}, \mathbf{y} \in I\!R^N$,*

$$|u(\mathbf{x}+\mathbf{y}) - u(\mathbf{x})| \leq \varepsilon(|\mathbf{y}|).$$

The function $\varepsilon(s)$ is then called a uniform continuity modulus of u. Show that if u is uniformly continuous and T is translation invariant and commutes with the addition of constants, then Tu also is uniformly continuous, with the same modulus as u. (Adapt the proof of Lemma 7.5).

Exercise 7.2 *Let T be a monotone contrast invariant operator. Show that*
(i) T(C)=C for every constant function C.
(ii) $u \geq C \Rightarrow Tu \geq C$, $u \leq C \Rightarrow Tu \leq C$.
(iii) $\sup |Tu(\mathbf{x}) - Tv(\mathbf{x})| \leq \sup |u(\mathbf{x}) - v(\mathbf{x})|$.
Indication : write $-\sup |u(\mathbf{x}) - v(\mathbf{x})| \leq u(\mathbf{x}) - v(\mathbf{x}) \leq \sup |u(\mathbf{x}) - v(\mathbf{x})|$.
(iv) Notice that the above results still work if we only assume that T is monotone and commutes with the addition of constants.

7.2 From constrast invariant operators to set operators : the threshold superposition principle.

Assume that $\mathcal{F}$ contains the characteristic functions $\mathbb{1}_X$ of elements of X. We can associate with T a set operator $\mathtt{T}$, defined on the set $\mathcal{T}$ of all level sets of all functions in $\mathcal{F}$ by

$$\mathtt{T}(X) = \mathcal{X}_1(T(\mathbb{1}_X)). \tag{7.5}$$

where $\mathbb{1}_X(\mathbf{x}) = 1$ if $\mathbf{x} \in X$ and $\mathbb{1}_X(\mathbf{x}) = 0$ otherwise . Note that if T is monotone, then $\mathtt{T}$ is a set monotone operator. Indeed,

$$X \subset Y \Leftrightarrow \mathbb{1}_X \leq \mathbb{1}_Y.$$

The definition of $\mathtt{T}$ makes sense by Proposition 7.4 : according to this proposition, $T(\mathbb{1}_X)$ attains the values 0 and 1 like u and is therefore a characteristic function. See Exercise 7.3

Exercise 7.3 *In order to justify the definition of $\mathtt{T}$ when T is contrast invariant, we shall show that $T(\mathbb{1}_X)$ is a characteristic function, i.e. only attains the values 0 and 1. In order to do so, consider a contrast change g such that $g(0) = 0$, $g(1) = 1$ and $g(s) \neq s$ for $s \neq 0$ or 1. Show that $T(\mathbb{1}_X) = T(g(\mathbb{1}_X)) = g(T(\mathbb{1}_X))$. Deduce that $T(\mathbb{1}_X)$ has the only values 0 and 1 and conclude that $\mathbb{1}_{\mathtt{T}(X)} = T(\mathbb{1}_X)$. Notice that this result also follows from Proposition 7.4.*

In order to explore the relationships between T and $\mathtt{T}$, we shall first show roughly that T commutes with thresholds.

Lemma 7.6 *Let T be a contrast invariant monotone operator. Consider the threshold functions $\gamma_\lambda(s) = 1$ if $s \geq \lambda$ and $\gamma_\lambda(s) = 0$ otherwise. Then T commutes almost everywhere with almost every threshold, i.e.*

$$\gamma_\lambda(Tu) = T(\gamma_\lambda(u)) \text{ a.e. in } \lambda, \quad \text{a.e. in } \mathbf{x}.$$

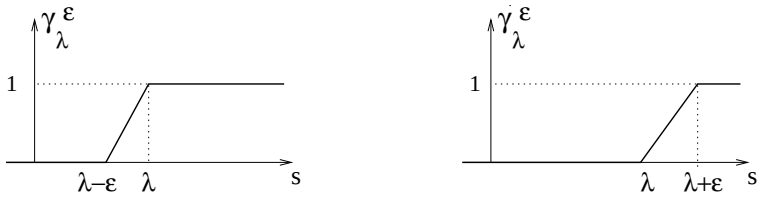

Figure 7.1: Approximation of the threshold function γ_λ from above and from below

Proof Let

$$\gamma_\lambda^\epsilon(s) = 0 \text{ if } s \leq \lambda - \epsilon,$$
$$\gamma_\lambda^\epsilon(s) = \frac{s - (\lambda - \epsilon)}{\epsilon} \text{ if } \lambda - \epsilon \leq s \leq \lambda,$$
$$\gamma_\lambda^\epsilon(s) = 1 \text{ otherwise .}$$

Then $\gamma_\lambda^\epsilon(s)$ is a contrast change and $\gamma_\lambda^\epsilon \geq \gamma_\lambda$. Thus

$$T(\gamma_\lambda(u)) \leq T(\gamma_\lambda^\epsilon(u)) = \gamma_\lambda^\epsilon(Tu) \to \gamma_\lambda(Tu), \text{ as } \epsilon \to 0.$$

Using in the same way continuous nondecreasing functions $\tilde{\gamma}_\lambda^\epsilon \leq \gamma_\lambda$, we also prove that $T(\gamma_\lambda(u)) \geq \gamma_\lambda^-(Tu)$, where $\gamma_\lambda^-(s) = 1$ if $s > \lambda$ and $\gamma(s) = 0$ otherwise. We therefore obtain

$$\gamma_\lambda^-(Tu) \leq T(\gamma_\lambda(u)) \leq \gamma_\lambda(Tu).$$

Let us consider the countable, and therefore negligible, subset $\Lambda \subset I\!R$ of all λ such that $meas(\{\mathbf{x}, Tu(\mathbf{x}) = \lambda\}) > 0$. For $\lambda \in I\!R \setminus \Lambda$, we have $\gamma_\lambda^-(Tu) = \gamma_\lambda(Tu)$ almost everywhere. Thus, for almost every λ, we obtain

$$T(\gamma_\lambda(u)) = \gamma_\lambda(Tu) \text{ almost everywhere.}$$

□

In a converse way to Relation (7.5) defining a set operator from a function operator, we can define a function operator T from a set operator $\mathtt{T}$ thanks to the *threshold superposition principle*.

Definition 7.7 *We say that T is obtained from $\mathtt{T}$ by the threshold superposition principle almost everywhere if*

$$Tu(\mathbf{x}) = \sup\{\lambda,\ \mathbf{x} \in \mathtt{T}(\mathcal{X}_\lambda u)\},\ \textit{a.e. in } \mathbf{x}. \tag{7.6}$$

If the preceding relation holds for every λ and $\mathbf{x}$, we say that the threshold superposition holds everywhere.

Proposition 7.8 *Let T be a monotone contrast invariant operator on a set of functions $\mathcal{F}$ containing the characteristic functions $\mathbb{1}_X$ of the elements X of $\mathcal{T}$. Define its associated set operator by $\mathtt{T}(X) = \mathcal{X}_1(T(\mathbb{1}_X))$. Then T is monotone, we have for every $u \in \mathcal{F}$*

$$\mathtt{T}(\mathcal{X}_\lambda u) = \mathcal{X}_\lambda(T(u)),\ \textit{a.e. in } \lambda,\ \ \textit{a.e. in } \mathbf{x} \tag{7.7}$$

and the threshold superposition principle holds almost everywhere :

$$Tu(\mathbf{x}) = \sup\{\lambda,\ \mathbf{x} \in \mathtt{T}(\mathcal{X}_\lambda u)\},\ \textit{a.e. in } \mathbf{x}. \tag{7.8}$$

In addition,

$$\mathtt{T}(\emptyset) = \emptyset \textit{ a.e.},\ \mathtt{T}(I\!R^N) = I\!R^N \textit{ a.e.}.$$

Proof Using the definition of $\mathtt{T}$, the obvious relations $\mathbb{1}_{\mathcal{X}_\lambda u} = \gamma_\lambda(u)$, $\mathcal{X}_1(\gamma_\lambda(v)) = \mathcal{X}_\lambda v$, and the commutation almost everywhere of T with γ_λ, obtained in Lemma 7.6, we get

$$\mathtt{T}(\mathcal{X}_\lambda u) = \mathcal{X}_1(T(\mathbb{1}_{\mathcal{X}_\lambda u})) = \mathcal{X}_1(T(\gamma_\lambda(u))) = \mathcal{X}_1(\gamma_\lambda(Tu)) = \mathcal{X}_\lambda(Tu), \quad \text{a.e. in } \lambda, \quad \text{a.e. in } \mathbf{x}.$$

Let us show the superposition principle (7.6). It follows immediately from (7.7) and Lemma 4.6, applied to $Y_\lambda = \mathtt{T}(\mathcal{X}_\lambda u)$ and $v = Tu$. We now prove the last statement. We take for u a constant function, say, equal to 0. Then $\mathcal{X}_\lambda u = \emptyset$ for $\lambda > 0$. By the relation (7.7) just proved, we deduce that $\mathcal{X}_\lambda(Tu) = \mathtt{T}(\mathcal{X}_\lambda u) = \mathtt{T}(\emptyset)$ almost everywhere and for almost every $\lambda > 0$. Thus the equality holds for many λ's. On the other hand, by contrast invariance, T commutes with the constant function 0. Thus $Tu = 0$ and therefore $\mathcal{X}_\lambda(Tu) = \emptyset$ for $\lambda > 0$. We finally obtain $\mathtt{T}(\emptyset) = \emptyset$ almost everywhere. The other announced relation follows in the same way. □

Relation (7.6) means that we can compute all level sets of Tu separately, by applying T to each characteristic function of the level set $\mathcal{X}_\lambda u$. We then have the following "stack filter" algorithm:

$$\begin{array}{ccccc} & & \mathcal{X}_\lambda u & \to & \mathtt{T}(\mathcal{X}_\lambda u) & & \\ u & \nearrow\searrow & \cdots & & & \searrow\nearrow & Tu(\mathbf{x}) = \sup\{\lambda, \mathbf{x} \in \mathtt{T}(\mathcal{X}_\lambda u)\}. \\ & & \mathcal{X}_\mu u & \to & \mathtt{T}(\mathcal{X}_\mu u) & & \end{array}$$

As developped in Chapter 24, the stack filter method will be used to maintain a very accurate contrast invariance in numerical schemes for several contrast invariant P.D.E.'s.

7.3 From set operators to contrast invariant function operators.

Can we extend a set monotone operator into a contrast invariant and monotone operator defined on an adequate set of functions ? We start from a set operator $\mathtt{T}$, defined on a subset $\mathcal{T}$ of $\mathcal{P}(I\!R^N)$ and with values in $\mathcal{T}$. We assume that $\mathtt{T}$ is a *monotone* operator, that is, $X \subset Y \Rightarrow \mathtt{T}(X) \subset \mathtt{T}(Y)$. We define $\mathcal{F}$ as the set of all real functions defined on $I\!R^N$ whose level sets $\mathcal{X}_\lambda u$ belong to $\mathcal{T}$. As a consequence, $\mathcal{F}$ contains the characteristic functions of elements of $\mathcal{T}$.

A natural definition for the function operator T associated with $\mathtt{T}$ is given by the threshold superposition principle,

$$Tu(\mathbf{x}) = \sup\{\lambda, \mathbf{x} \in \mathtt{T}(\mathcal{X}_\lambda u)\} \tag{7.9}$$

We have now to check whether T is contrast invariant.

Proposition 7.9 *Let $\mathtt{T}$, $\mathcal{T} \to \mathcal{T}$ be a monotone operator satisfying*

$$T(\emptyset) = \emptyset, \mathtt{T}(I\!R^N) = I\!R^N$$

Then the operator T defined on $\mathcal{F}$ by superposition principle,

$$Tu(\mathbf{x}) = \sup\{\lambda, \mathbf{x} \in \mathtt{T}(\mathcal{X}_\lambda u)\} \tag{7.10}$$

satisfies, for almost every $\lambda \in \mathbb{R}$,

$$\mathcal{X}_\lambda(Tu) = \mathtt{T}(\mathcal{X}_\lambda(u)) \text{ almost everywhere in } \mathbf{x} \tag{7.11}$$

and for any nondecreasing continuous contrast change g,

$$g(Tu) = T(g(u)). \tag{7.12}$$

Proof By Corollary 4.5, we have for almost every λ

$$\mathcal{X}_\lambda(Tu) = \mathtt{T}(\mathcal{X}_\lambda u) \text{ almost everywhere} \tag{7.13}$$

that is for all λ in a subset Λ of $\mathbb{R}$ such that $meas(\mathbb{R} \setminus \Lambda) = 0$. We notice that $u \leq v$ if and only if $\mathcal{X}_\lambda u \subset \mathcal{X}_\lambda v$ for all λ in a dense countable subset of $\mathbb{R}$. We deduce immediately that T is monotone, ie $u \geq v \ \Rightarrow Tu \geq Tv$.

Let us now show that T commutes with contrast changes. Assume first that g is strictly increasing and set $g(+\infty) = \lim_{s\to+\infty} g(s)$ and $g(-\infty) = \lim_{s\to-\infty} g(s)$.
For $\lambda > g(+\infty)$, we have $\mathcal{X}_\lambda g(u) = \emptyset$ and therefore $\mathtt{T}(\mathcal{X}_\lambda g(u)) = \emptyset$.
For $\lambda < g(-\infty)$, we have $\mathcal{X}_\lambda g(u) = \mathbb{R}^N$ and therefore $\mathtt{T}(\mathcal{X}_\lambda g(u)) = \mathbb{R}^N$.
As a consequence, using 7.10 we can restrict the range of λ in the definition of $T(g(u(\mathbf{x})))$:

$$T(g(u(\mathbf{x}))) = \sup\{\lambda,\ g(-\infty) \leq \lambda \leq g(+\infty),\ \mathbf{x} \in \mathtt{T}(\mathcal{X}_\lambda g(u))\}$$

$$= \sup\{g(\mu),\ \mathbf{x} \in \mathtt{T}(\mathcal{X}_{g(\mu)} g(u))\}$$

$$= \sup\{g(\mu),\ \mathbf{x} \in \mathtt{T}(\mathcal{X}_\mu u)\} = g(Tu(\mathbf{x})).$$

Let us now check that T commutes with general nondecreasing contrast changes g. We can find strictly increasing continuous functions g_n and h_n such that $g_n(s) \to g(s)$, $h_n(s) \to g(s)$ for all s and $g_n \leq g \leq h_n$. Thus, by using the just proven commutation of T with increasing contrast changes, we have

$$T(g(u)) \geq T(g_n(u)) = g_n(Tu) \to g(Tu) \text{ and}$$

$$T(g(u)) \leq T(h_n(u)) = h_n(Tu) \to g(Tu),$$

which yields $T(g(u)) = g(Tu)$. □

Exercise 7.4 *Let* g *be a continuous nondecreasing function. Construct increasing continuous functions* g_n *and* h_n *such that* $g_n(s) \to g(s)$, $h_n(s) \to g(s)$ *for all* s *and* $g_n \leq g \leq h_n$.

Exercise 7.5 *We explain the necessity of the assumptions* $\mathtt{T}(\emptyset) = \emptyset$ *in Proposition 7.9,* $\mathtt{T}(\mathbb{R}^N) = \mathbb{R}^N$ *in order to get a contrast invariant operator* T *from* $\mathtt{T}$.

1) Set $\mathtt{T}(X) = X_0$ *for all* $X \in \mathcal{T}$, *where* $X_0 \neq \emptyset$ *is a fixed set. Check that* $Tu(\mathbf{x}) = +\infty$ *if* $\mathbf{x} \in X_0$, $Tu(\mathbf{x}) = -\infty$ *otherwise. Let now* g *be a continous nondecreasing bounded function. Show that* $T(g(u)) \neq g(Tu)$.

2) Let $\emptyset \neq X_1 \subset X_0 \neq I\!R^N$ and set $\mathtt{T}(I\!R^N) = X_0$, $\mathtt{T}(X) = X_1$ if $X \neq I\!R^N$, $\mathtt{T}(\emptyset) = \emptyset$. Show that $Tu(\mathbf{x}) = -\infty$ on X_0^c and deduce again that $T(g(u)) \neq g(Tu)$ if g is bounded.

3) Let $\mathtt{T}$ be a monotone set operator, without further assumption. Show that its function operator T, associated by the threshold superposition principle, commutes with any continuous nondecreasing function g such that $g(+\infty) = +\infty$ and $g(-\infty) = -\infty$.

We shall now examine stronger conditions on $\mathtt{T}$, which ensure commutation of T with all thresholds $\mathcal{X}_\lambda$ (and not only with almost all of them).

Definition 7.10 *We say that a set operator $\mathtt{T}$ is upper semicontinuous if for every nonincreasing sequence of sets $X_n \in \mathcal{T}$, we have*

$$\mathtt{T}(\bigcap_n X_n) = \bigcap_n \mathtt{T}(X_n). \tag{7.14}$$

Exercise 7.6 *Show that a monotone operator $\mathtt{T}$ is upper semicontinuous if and only if it satisfies for every nonincreasing family of sets $(X_\lambda)_{\lambda \in I\!R}$, $\mathtt{T}(\bigcap_\lambda X_\lambda) = \bigcap_\lambda \mathtt{T}(X_\lambda)$.*

Theorem 7.11 *Let $\mathtt{T}$: $\mathcal{T} \to \mathcal{T}$ be a monotone upper semicontinuous set operator satisfying $\mathtt{T}(\emptyset) = \emptyset$ and $\mathtt{T}(I\!R^N) = I\!R^N$. Then the function operator defined by*

$$Tu(\mathbf{x}) = \sup\{\lambda, \mathbf{x} \in \mathtt{T}(\mathcal{X}_\lambda u)\}. \tag{7.15}$$

is contrast invariant and satisfies, for every λ,

$$\mathcal{X}_\lambda(Tu) = \mathtt{T}(\mathcal{X}_\lambda(u)). \tag{7.16}$$

Proof of Theorem 7.11 Let us check that (7.16) holds. By Proposition 4.1, this is true if and only if the family of sets $\mathtt{T}(\mathcal{X}_\lambda u)$ satisfies (4.1.i) and (4.1.ii). Now, (4.1.i) is true by monotonicity of $\mathtt{T}$ and the inclusion relations between the $X_\lambda u$. Relation (4.1.ii) follows immediately from (7.14) and $\mathcal{X}_\lambda u = \bigcap_{\mu<\lambda} \mathcal{X}_\mu u$, which is (4.1.ii) applied to u. The monotonicity of T is obvious : one has

$$u \leq v \Leftrightarrow (\forall \lambda,\ \mathcal{X}_\lambda u \subset \mathcal{X}_\lambda v).$$

Since $\mathtt{T}$ is set monotone, we deduce that

$$\forall \lambda,\ \mathtt{T}(\mathcal{X}_\lambda u) \subset \mathtt{T}(\mathcal{X}_\lambda v)$$

and therefore $\forall \lambda, \mathcal{X}_\lambda(Tu) \subset \mathcal{X}_\lambda(Tv)$, that is $Tu \leq Tv$. In order to show that T is contrast invariant, i.e. commutes with all continuous nondecreasing functions, we can now directly apply Proposition 7.9. □

Remark 7.12 *The upper semicontinuity is necessary to ensure that a monotone set operator defines a function operator such that the commutation with thresholds $\mathcal{X}_\lambda(Tu) = \mathtt{T}(\mathcal{X}_\lambda(u))$ holds for every λ. Let us choose for example the following set operator $\mathtt{T}$,*

$$\mathtt{T}(X) = X \quad \textit{if } meas(X) > a \textit{ and } \mathtt{T}(X) = \emptyset \textit{ otherwise .}$$

Let u be the function from $\mathbb{R}$ into $\mathbb{R}$ defined by $u(x) = -|x|$. Then

$$T(u)(x) = \sup\{\lambda, x \in \mathtt{T}(\mathcal{X}_\lambda u)\} = \min(-|x|, -a/2)$$

Therefore

$$\mathcal{X}_{-a/2}T(u) = [-a/2, a/2].$$

Now, $\mathcal{X}_{-a/2}u = [-a/2, a/2]$; its measure is a. Thus

$$T(\mathcal{X}_{-a/2}u) = \emptyset \neq \mathcal{X}_{-a/2}T(u),$$

which means that T is not contrast invariant.

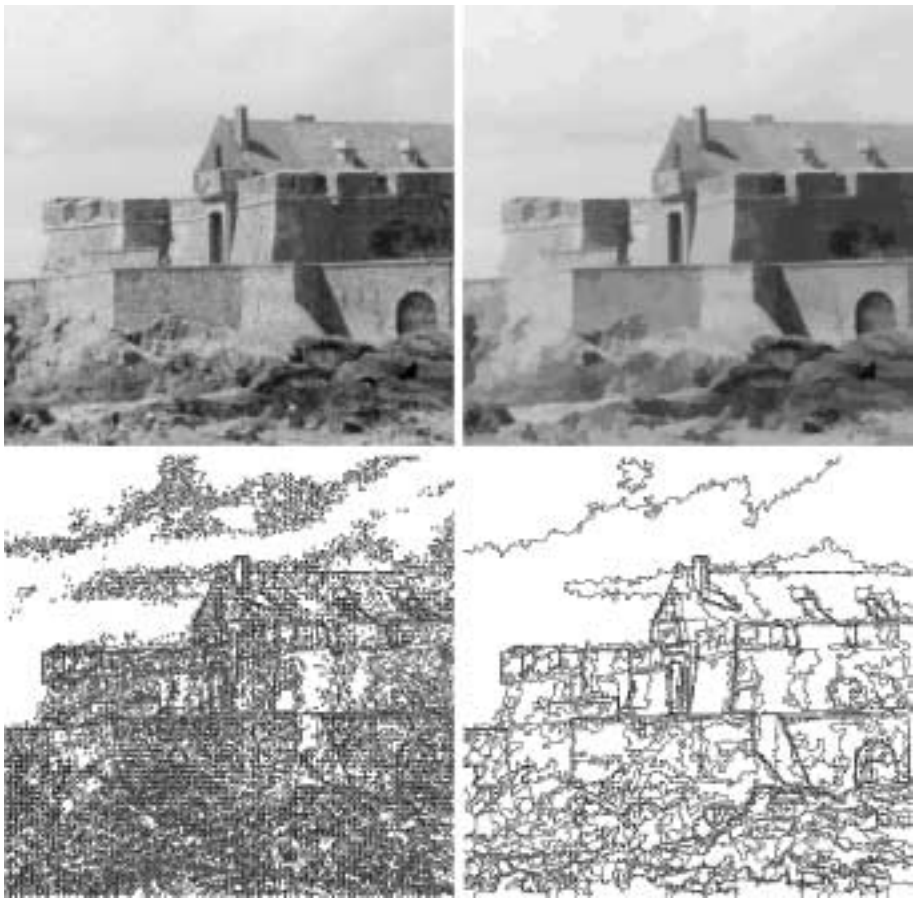

Figure 7.2: Extrema Killer : maxima killer followed by minima killer. The extrema killer removes all connected components of upper and lower level sets with area less than some threshold, here equal to 20 pixels.

7.4 Application : The "Extrema Killer".

We study in this section operators which remove peaks in an image. Such peaks are often created by impulse noise. The extrema killer operators show outstanding denoising properties for such kinds of noise. We shall use the method developped in Section 7.3 which permits to define these operators on sets and then extend them to images.

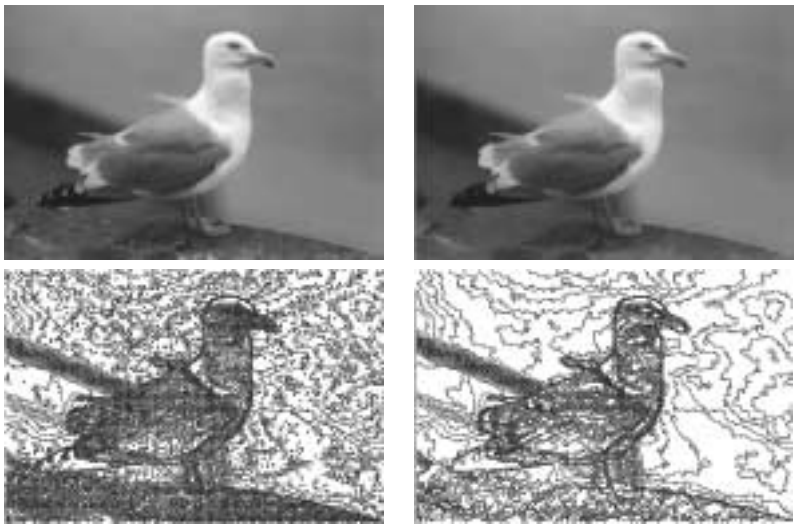

Figure 7.3: Extrema Killer : maxima killer followed by minima killer. Above, right: extrema killer of the above, left image, 20 pixels area. Below : level lines (for levels multiple of 4) of the image before and after the application of the extrema killer. Notice the strong simplification of the topographic map : most digital images have many small oscillations generating extrema.

Definition 7.13 *Let $X \subset I\!R^N$ be a closed set. We say that X is connected if it is not the union of two disjoint nonempty open sets. We call connected component of $\mathbf{x}$ in X, $cc(\mathbf{x}, X)$, the maximal connected subset of X containing $\mathbf{x}$. It is a closed set.*

Fix a scale parameter $a > 0$. We define a set denoising operator on $\mathcal{T}$, the set of all compact sets of $I\!R^N$ in the following way: Let $X \in \mathcal{T}$ be a compact set. Then X is the union of all of its connected components, $X = \bigcup_i X_i$ and this decomposition is uniquely defined. We then remove from X all connected components of measure strictly less than a : We therefore define a "small component killer"

$$\mathtt{T}_a(X) = \bigcup_{meas(X_i) \geq a} X_i. \tag{7.17}$$

Lemma 7.14 *Let Y_n be a nonincreasing sequence of subsets of $I\!R^N$ and set $Y = \bigcap_n Y_n$. Then, if Y_n are compact, so is Y. If Y_n are compact and connected, so is Y.*

Proof These are classical topological arguments and we just check the proof of the second statement. If Y were not connected, then $Y = Z_1 \cup Z_2$ where Z_1 and Z_2 are compact and disjoint. Thus, $d(Z_1, Z_2) = \inf_{z_1 \in Z_1, z_2 \in Z_2} ||z_1 - z_2|| > 0$. We then remark that if $Y_n \searrow Y$ and Y_n are compact, then the Hausdorff semi-distance of Y_n to Y tends to 0: $\delta(Y_n, Y) \to 0$. (We set $\delta(Y, Z) = \sup_{\mathbf{y} \in Y} \inf_{z \in \mathbf{z}} ||\mathbf{y} - \mathbf{z}||$.) We choose n_0 such that $\forall n > n_0$, $\epsilon = \delta(Y, Y_n^c) < \frac{1}{2} d(Z_1, Z_2)$. It is then easily checked that Y_n is split into two disjoint open nonempty sets:

$$Y_n = (Y_n \cap Z_1^\epsilon) \cup (Y_n \cap Z_2^\epsilon)$$

where we note $Z^\epsilon = \{\mathbf{z}, d(\mathbf{z}, Z) < \epsilon\}$. □

Lemma 7.15 *The small component killer $\mathtt{T}_a$ is monotone and upper semicontinuous in the sense of Definition 7.10 on the set $\mathcal{T}$ of all compact subsets of $\mathbb{R}^N$.*

Proof If $X \subset Y$, then for every $\mathbf{x}$, $cc(\mathbf{x}, X) \subset cc(\mathbf{x}, Y)$. Thus, if we assume that $meas(cc(\mathbf{x}, X)) \geq a$, then we also have $meas(cc(\mathbf{x}, Y)) \geq a$. We conclude that $\mathtt{T}_a(X) \subset \mathtt{T}_a(Y)$, that is, $\mathtt{T}_a$ is monotone. Let $(X_n)_{n\in\mathbb{N}}$ be a nonincreasing sequence of compact sets and let $X = \bigcap_n X_n$. We wish to show that $\mathtt{T}_a(X) = \bigcap_n \mathtt{T}_a(X_n)$. By Lemma 7.14, $\cap_n X_n$ is compact. Let now $Y = cc(\mathbf{x}, X)$ be a connected component of X and denote by $Y_n = cc(\mathbf{x}, X_n)$. Since $X \subset X_n$, we have $Y \subset Y_n$ for all n. Let us show that $Y = \bigcap_n Y_n$. Obviously, $Y \subset \bigcap_n Y_n$ and $\bigcap_n Y_n \subset \bigcap_n X_n = X$. We therefore just have to check that $\bigcap_n Y_n$ is connected, which follows from Lemma 7.14. □

Theorem and Definition 7.16 *One can associate with $\mathtt{T}_a$ a "maxima killer", defined on all continuous periodic functions $u \in L_C$. This operator T_a is defined by the threshold decomposition principle,*

$$T_a u(\mathbf{x}) = \sup\{\lambda, \mathbf{x} \in \mathtt{T}_a(\mathcal{X}_\lambda u)\},$$

we also have

$$\mathcal{X}(T_a u) = \mathtt{T}_a(\mathcal{X} u) \tag{7.18}$$

and, as a consequence, no connected maximum set of $T_a u$ has area less than a. In addition, $T_a u$ is continuous.

Proof By Theorem 7.11 and Lemma 7.15, we deduce immediately the announced commutation (7.18). Let us check that $T_a u$ is continuous. Since u is continuous and periodic, it is uniformly continuous. We can use the result of the (easy) exercise 7.1, whose proof mimics the proof of Lemma 7.4. Since T_a is monotone and translation invariant, $T_a u$ also is continuous. (And, by Lemma 7.4, if u is Lipschitz, so is $T_a u$, with the same Lipschitz constant.) □

We have defined a "maxima killer" and we could define in the same way a "minima killer". A faster way to define it is to simply set

$$T_a^- u = -T_a(-u)$$

A good denoiser can be obtained by alternating T_a and T_a^-. This alternance is licit since each one of both extrema killers maintains the continuity of the image. Note, however, that T_a and T_a^- do not necessarily commute as is shown by the next exercise.

Exercise 7.7 *Take $N = 1$, $u(x) = \sin(x)$. Let $a \in \mathbb{R}^+$. Compute $T_a u$ and $T_a^- u$ and show that they commute on u if $a \leq \pi$ and do not commute if $a > \pi$. Following the same kind of idea, construct an example in dimension 2 of function $u(x, y)$ such that $T_a T_a^- u \neq T_a^- T_a u$.*

Remark 7.17 *The operators T_a do not constitute a scale-space in the sense which we will consider in Chapter 20. In particular, they do not satisfy a local comparison principle. See Exercise 20.1.*

References.

Image operators commuting with thresholds have been very popular because of their simplicity of implementation on VLSI, which led to very simple patents in image or signal processing as late as 1987 [111]. These operators have received four different but equivalent names : "stack filters" [433, 434, 196, 70], "threshold decomposition" [204], "rank filters" [437, 99, 241] and "order filters" [399]. The most famous such operators are the sup, inf and median operators. The implementation of the last one has received a lot of attention because of his remarkable denoising properties [151, 444, 324].

The papers by Maragos and Shafer [284, 283] and Maragos and Ziff [285]introduce the functional notation in the debate and establish the link between stack filters and the Matheron formalism in "flat" mathematical morphology. Actually, Maragos and his collaborators prove the equivalence between stack filters and operators commuting with thresholds. The full equivalence between contrast invariant operators and stack filters, proved in this chapter, does not seem instead to be classical and we cannot give any other reference than the present book. A related classification of rank filters with beautiful and useful generalizations to the so-called "neighborhood filters" can be found in [241].

The extrema killer is probably the most efficient denoising filter for images degraded by impulse noise (small spots). In spite of its simplicity, it is a recent filter, probably because it involved nontrivial computations : the search of connected components of upper and lower level sets. A first attempt seems to be [102]. The filter in its generality was defined in Vicent [415]. Its definition fits in the general theory of connected filters developped by Salembier and Serra [364]. Simon Masnou [289] defined a variant which is invariant by contrast as well as by reverse contrast changes (the so-called "grain filter"). Monasse and Guichard developped a fast implementation of this filter based on the so called "fast level set transform" [301].

Chapter 8

Monotone contrast invariant operators as sup-inf operators.

In this chapter, we mainly state two theorems which give an analytic form to set monotone operators and to contrast invariant and monotone operators. Before starting with the statements, let us introduce a further useful invariance property. In practice, image analysis operators are translation invariant.

Definition 8.1 *We denote by $\tau_{\mathbf{x}} X = \mathbf{x} + X$ the result of an $\mathbf{x}$-translation. We also consider translations of a function u, which we define by $(\tau_{\mathbf{x}})u(\mathbf{y}) = u(\mathbf{y} - \mathbf{x})$.*

Exercise 8.1 *Show that the definitions of translations of sets and functions made in Definition 8.1 are made in such a way that*

$$\tau_{\mathbf{x}}(\mathcal{X}_\lambda u) = \mathcal{X}_\lambda(\tau_{\mathbf{x}} u).$$

Definition 8.2 *We say that a set operator $\mathtt{T}$ is translation invariant if*

$$\tau_{\mathbf{x}}(\mathtt{T}(X)) = \mathtt{T}(\tau_{\mathbf{x}} X).$$

We say that a function operator T is translation invariant if

$$\tau_{\mathbf{x}}(T(u)) = T(\tau_{\mathbf{x}} u).$$

8.1 Monotone set operators.

Theorem 8.3 (Matheron) *Let $\mathtt{T}$ be a translation invariant monotone operator acting on a set of subsets of $\mathbb{R}^N$. Then, there exists a family of sets $\mathcal{B} \subset \mathcal{P}(\mathbb{R}^N)$, which can be defined as $\mathcal{B} = \{X, 0 \in \mathtt{T}(X)\}$, such that*

$$\mathtt{T}(X) = \bigcup_{B \in \mathcal{B}} \bigcap_{\mathbf{y} \in B} (X - \mathbf{y}) = \{\mathbf{x},\ \exists B \in \mathcal{B},\ \mathbf{x} + B \subset X\}. \tag{8.1}$$

Conversely, (8.1) defines a monotone, translation invariant operator on $\mathcal{P}(\mathbb{R}^N)$.

Proof Let us first explain Formula (8.1). In fact,

$$\mathtt{T}(X) = \bigcup_{B\in I\!B} \bigcap_{\mathbf{y}\in B} (X-\mathbf{y}) = \{\mathbf{x},\ \exists B \in I\!B,\ \forall \mathbf{y}\in B, \mathbf{x}\in X-\mathbf{y}\} \tag{8.2}$$

This last set is obviously equal to $\{\mathbf{x},\ \exists B\in I\!B,\ \mathbf{x}+B\subset X\}$.

Using the monotonicity and the translation invariance, we have the following chain of equivalent properties :

$$\mathbf{x}\in \mathtt{T}(X) \Leftrightarrow 0\in \mathtt{T}(X)-\mathbf{x} \Leftrightarrow 0\in \mathtt{T}(X-\mathbf{x}) \Leftrightarrow X-\mathbf{x}\in I\!B \Leftrightarrow \exists B\in I\!B,\ X=\mathbf{x}+B \Leftrightarrow$$

$$\exists B\in I\!B,\ \mathbf{x}+B\subset X \Leftrightarrow \exists B\in I\!B,\ \forall \mathbf{y}\in B, \mathbf{x}\in X-\mathbf{y}.$$

The fifth equivalence comes from the obvious remark that if $B\subset X$ and $B\in I\!B$ then $X\in I\!B$. Thus $\mathtt{T}(X)=\bigcup_{B\in I\!B}\bigcap_{\mathbf{y}\in B}(X-\mathbf{y})$ and Relation (8.1) holds. Conversely, if an operator is defined by (8.1), it is obviously monotone and translation invariant. □

Exercise 8.2 *Show that the set $I\!B$ is not uniquely associated with* T. *Among those $I\!B$ which are associated with* T, *prove that there exists at least one with the following properties :*
(i) if $B, B'\in I\!B$, then B does not contain B' and B' does not contain B.
(ii) $I\!B$ is minimal for inclusion, i.e. does not contain another smaller $I\!B'$ also associated with T.

8.2 Sup-inf operators

Theorem 8.4 *Let $\mathcal{F}$ a set of functions, $\mathcal{T}$ the set of all level sets of functions of $\mathcal{F}$. Assume that $\mathcal{F}$ is stable under contrast changes and contains the characteristic functions $\mathbb{1}_X$ of elements of $\mathcal{T}$. Let* T *be the set operator associated with T, $\mathtt{T}(X)=\mathcal{X}_1(T(\mathbb{1}_X))$. Define for any $\mathbf{x}\in I\!R^N$ the family of sets $I\!B_{\mathbf{x}}=\{X, \mathbf{x}\in \mathtt{T}(X)\}$. Then for every u in $\mathcal{F}$,*

$$Tu(\mathbf{x}) = \sup_{B\in I\!B_{\mathbf{x}}} \inf_{\mathbf{y}\in B} u(\mathbf{y}),\ \textit{a.e. in } \mathbf{x} \tag{8.3}$$

If, in addition, T is translation invariant, then setting $I\!B=I\!B_0=\{X, 0\in \mathtt{T}(X)\}$ we have

$$Tu(\mathbf{x}) = \sup_{B\in \mathbf{x}+I\!B} \inf_{\mathbf{y}\in B} u(\mathbf{y}),\ \textit{a.e. in } \mathbf{x} \tag{8.4}$$

Conversely, if an operator is defined by (8.4) or (8.3), then it is monotone and contrast invariant, and is translation invariant in the first case.

Remark 8.5 *The set $I\!B$ is called in Mathematical Morphology* set of structuring elements. *It corresponds to an "impulse response" of the nonlinear operator T.*

Proof of Theorem 8.4. Set $\tilde{T}u(\mathbf{x})=\sup_{B\in I\!B_{\mathbf{x}}}\inf_{\mathbf{y}\in B}u(\mathbf{y})$, where $I\!B_{\mathbf{x}}=\{X, \mathbf{x}\in \mathtt{T}(X)\}$. Let us show that $\tilde{T}u(\mathbf{x})=Tu(\mathbf{x})$ almost everywhere. We argue as in Lemma 4.6 : We choose a countable dense set $\Lambda\subset I\!R$ such that for every $\lambda\in\Lambda$, $\mathcal{X}_\lambda Tu(\mathbf{x})=\mathtt{T}(\mathcal{X}_\lambda u)(\mathbf{x})$ for $\mathbf{x}\in I\!R^N\setminus N_\lambda$, where $meas(N_\lambda)=0$. We set

$N = \bigcup_\lambda N_\lambda$ and we still have $meas(N) = 0$. In order to prove that $Tu = \tilde{T}u$ almost everywhere, we show that for all $\lambda \in \Lambda$ and all $\mathbf{x} \in I\!R^N \setminus N$, we have

$$\tilde{T}u(\mathbf{x}) \geq \lambda \Leftrightarrow Tu(\mathbf{x}) \geq \lambda.$$

and we apply Lemma 4.6. In what follows, λ and μ denote elements of Λ. We have

$$Tu(\mathbf{x}) \geq \lambda \Leftrightarrow \forall \mu < \lambda,\ Tu(\mathbf{x}) \geq \mu \Leftrightarrow \forall \mu < \lambda,\ \mathbf{x} \in \mathcal{X}_\mu(Tu) \Leftrightarrow \forall \mu < \lambda,\ \mathbf{x} \in \mathtt{T}(\mathcal{X}_\mu u)$$

$$\Leftrightarrow \forall \mu < \lambda,\ \mathcal{X}_\mu u \in I\!B_{\mathbf{X}} \Leftrightarrow \forall \mu < \lambda,\ \exists B (= \mathcal{X}_\mu u) \in I\!B_{\mathbf{X}},\ \inf_{\mathbf{y} \in B} u(\mathbf{y}) \geq \mu$$

$$\Leftrightarrow \sup_{B \in I\!B_{\mathbf{X}}} \inf_{\mathbf{y} \in B} u(\mathbf{y}) \geq \lambda \Leftrightarrow \tilde{T}u(\mathbf{x}) \geq \lambda.$$

The fifth equivalence is true because if $B \subset X$ and $B \in I\!B_{\mathbf{X}}$, then $X \in I\!B_{\mathbf{X}}$. Thus, if for some $B \in I\!B_{\mathbf{X}}$, $\inf_{\mathbf{y} \in B} u(\mathbf{y}) \geq \mu$, then the set $\mathcal{X}_\mu u$ contains B and therefore also belongs to $I\!B_{\mathbf{X}}$. We conclude that $\tilde{T}u(\mathbf{x}) = Tu(\mathbf{x})$ almost everywhere. If, in addition, T is translation invariant, we obviously have $I\!B_{\mathbf{X}} = \mathbf{x} + I\!B_0$. By the preceding result, we obtain $Tu(\mathbf{x}) = \sup_{B \in \mathbf{x} + I\!B} \inf_{\mathbf{y} \in B} u(\mathbf{y})$ almost everywhere, which is (8.4).

Let us now show the converse statement of the theorem. If T is a "sup-inf" operator, the monotonicity of T is obvious and the commutation of T with continuous nondecreasing functions follows from the fact that if g is continuous nondecreasing and I a subset of $I\!R$, then $g(\inf I) = \inf g(I)$ and $\sup g(I) = g(\sup I)$. The translation invariance is easily checked. □

Corollary 8.6 *Let T and $\mathtt{T}$ be translation invariant function and set operators such that*

$$\mathcal{X}_\lambda T(u) = \mathtt{T}(\mathcal{X}_\lambda u), \tag{8.5}$$

the relation being true for every λ and $\mathbf{x}$. A simplified version of the proof of Theorem 8.4 shows that for all $u \in \mathcal{F}$ and $\mathbf{x}$, $Tu(\mathbf{x}) = \sup_{B \in \mathbf{X} + I\!B} \inf_{\mathbf{y} \in B} u(\mathbf{y})$, the relation being true everywhere.

It will be useful to state a converse statement to Theorem 8.4. It states that given a $sup-inf$ operator, we can keep the same set of structuring elements for its associated set operator $\mathtt{T}$.

Proposition 8.7 *Let $\mathcal{F}$ be a set of functions containing the characteristic functions of sets belonging to $\mathcal{T}$ and T a function operator having a sup-inf form (8.3), namely*

$$Tu(\mathbf{x}) = \sup_{B \in \mathbf{X} + I\!B} \inf_{\mathbf{y} \in B} u(\mathbf{y}).$$

Then its associated set operator $\mathtt{T}$ can be defined by (8.1), that is

$$\mathtt{T}(X) = \bigcup_{B \in I\!B} \bigcap_{\mathbf{y} \in B} (X - \mathbf{y}) = \{\mathbf{x},\ \exists B \in I\!B,\ \mathbf{x} + B \subset X\}.$$

Proof Notice that T, satisfying (8.4), is extended to all level sets of functions of $\mathcal{F}$. It is easy to deduce (8.1) from (8.4) : We apply this relation to the characteristic function of X, $\mathbb{1}_X$. Then,

$$\begin{aligned} \inf_{\mathbf{y}\in\mathbf{x}+B} \mathbb{1}_X &= 0 \text{ if } \mathbf{x}+B \not\subset X \\ &= 1 \text{ if } \mathbf{x}+B \subset X. \end{aligned}$$

Thus $\sup_{B\in\mathbb{B}} \inf_{\mathbf{y}\in\mathbf{x}+B} \mathbb{1}_X(\mathbf{y}) = 1$ if and only if there is B in $\mathbb{B}$ such that $\mathbf{x}+B \subset X$ and (8.1) follows. □

Exercise 8.3 *Apply Corollary 8.6 to the extrema killer T_a and deduce its canonically associated structuring set $\mathbb{B}$. Show that another set of structuring elements for T_a can be*

$$\mathbb{B}_1 = \{B \subset \mathbb{R}^N,\ \textit{compact, connected, } meas(B) > a\}.$$

Indication : show that if $\mathbb{B}$ is a set of structuring elements defining an operator T and if $\mathbb{B}_1$ is another structuring set satisfying

- *$\mathbb{B}_1 \subset \mathbb{B}$ and*
- *for every $B \in \mathbb{B}$, there is $B_1 \in \mathbb{B}_1$ such that $B_1 \subset B$, then $\mathbb{B}_1$ defines the same operator T.*

Exercise 8.4 *It is not true that the sup-inf forms (8.3) or (8.4) define in general an operator commuting with all thresholds $\mathcal{X}_\lambda$. This needs some restriction, either on $\mathbb{B}$, or on the domain of definition of T. As an easy counterexample, let us set $\mathbb{B} = \{\{\mathbf{x}\}, \mathbf{x} \in D(0,1)\}$ where $D(0,1)$ is the ball with center 0 and radius 1 and therefore $Tu(\mathbf{x}) = \sup_{\mathbf{y}\in\mathbf{x}+D} u(\mathbf{y})$. We consider a dense sequence of $\mathbb{R}^N$, $(q_n)_{n\in\mathbb{N}}$ and we set $u(\mathbf{x}) = 1 - 1/n$ if $\mathbf{x} = q_n$ for some n, $u(\mathbf{x}) = 0$ otherwise. Show that $T(\mathcal{X}_1 u) = \emptyset \neq \mathcal{X}_1 Tu = \mathbb{R}^N$. We shall see in Exercise 8.5 that T, which we define later on as a "dilation", yields an operator commuting with all thresholds when it is restricted to u.s.c. functions.*

Exercise 8.5 .

A first classical example for $\mathcal{F}$ is the set of all upper semi-continuous (u.s.c.) functions, that is, functions which satisfy at every $\mathbf{x}$,

$$u(\mathbf{x}) \geq \limsup_{\mathbf{y}\to\mathbf{x}} u(\mathbf{y}).$$

Check that $\mathcal{T}$ is the set of all closed subsets of $\mathbb{R}^N$. A classical operator on this set is the so-called dilation, $T_r u(\mathbf{x}) = \sup_{D(\mathbf{x},r)} u(\mathbf{y})$. Show that it corresponds to the set operator $\mathtt{T}X = \{\mathbf{x}, d(\mathbf{x}, X) \leq r\} = X^r$.

The Matheron-Maragos theorem 8.4 and its "everywhere" version (Corollary 8.6) yield a sup inf formulation to contrast invariant operators. It is easy to deduce that those operators also have an inf sup form. (See also Exercise 8.6)

Proposition 8.8 *Let T be a* sup inf *operator $Tu(\mathbf{x}) = \sup_{B\in\mathbb{B}} \inf_{\mathbf{y}\in\mathbf{x}+B} u(\mathbf{y})$. Then it also has an* inf sup *form for a new structuring set $\mathbb{B}'$*

$$Tu(\mathbf{x}) = \inf_{B\in\mathbb{B}'} \sup_{\mathbf{y}\in\mathbf{x}+B} u(\mathbf{y}). \tag{8.6}$$

Proof This follows from elementary Boolean algebra. Let us call " selection function" any map $\phi : I\!B \to \bigcup_{B \in I\!B} B$ such that $\forall B,\ \phi(B) \in B$. Let us call $\mathcal{S}_{I\!B}$ the set of all selection functions. We then have

$$\sup_{B \in I\!B} \inf_{\mathbf{y} \in B} u(\mathbf{y}) = \inf_{\phi \in \mathcal{S}_{I\!B}} \sup_{B \in I\!B} u(\phi(B))$$

$$= \inf_{\phi \in \mathcal{S}_{I\!B}} \sup_{\mathbf{y} \in \phi(I\!B)} u(\mathbf{y})$$

Let us note $\mathbf{D} = \{\phi(I\!B), \phi \in \mathcal{S}_{I\!B}\}$. We obtain

$$\sup_{B \in I\!B} \inf_{\mathbf{y} \in B} u(\mathbf{y}) = \inf_{\phi(I\!B) \in \mathbf{D}} \sup_{\mathbf{y} \in \phi(I\!B)} u(\mathbf{y})$$

which indeed is an inf-sup operator. Of course the structuring set $\mathbf{D}$ thus obtained for the inf sup form is not the same as in the sup inf form ! □

Exercise 8.6 *The result of Proposition 8.8 can be deduced in a slightly weaker form from Theorem 8.4. By applying this theorem to $-T(-u)$, show that if*

8.3 From contrast invariant function operators to set operators : the Evans-Spruck extension

In this section, we consider monotone contrast invariant function operators T which are only defined on a set of continuous functions. Thus, we cannot associate with them a set operator by the formerly used formula, $\mathtt{T}(X) = T(\mathcal{X}_1(\mathbb{1}_X))$. Indeed, T is no longer defined on characteristic functions of sets ! We shall see, however, that we can associate with T a set operator $\mathtt{T}$ in such a way that $\mathcal{X}_\lambda(Tu) = \mathtt{T}(\mathcal{X}_\lambda u)$ for all λ. The method, introduced by Osher and Sethian, has shown to be very powerful in the numerical analysis of front propagation. We shall need a slightly restrictive assumption, that all level sets of the processed functions u are compact. In image analysis, images are anyway defined in a compact set C (in general a square, a rectangle, a cube, etc.) so that this compactness assumption is no real restriction. Thus, in the following, we consider a space of bounded continuous functions $\mathcal{F}$ defined on a compact set C. The level sets of functions in $\mathcal{F}$ are compact and we can always assume that functions in $\mathcal{F}$ are defined in all of $I\!R^N$: we just make a continuous extension of u preserving compactness of level sets. A standard way to do that is the periodization proposed in Section 2.2, in which case the level sets become periodic sets and remain compact provided we endow them with the natural periodic topology.

Let us now explain the Osher-Sethian method. Let X be a compact subset of $I\!R^N$ and let us associate with X a continuous bounded function $u(\mathbf{x})$ such that X be its zero level set :

$$X = \{\mathbf{x},\ u(\mathbf{x}) \geq 0\} = \mathcal{X}_0 u \tag{8.7}$$

Using a distance function to X, that is $d(\mathbf{x}, X) = \inf\{|\mathbf{x} - \mathbf{y}|, \mathbf{y} \in X\}$, we can easily build many such functions, as for example $u(\mathbf{x}) = -\min(1, d(\mathbf{x}, X))$. We can define $\mathtt{T}(X)$, by using (8.7), in the following way.

Lemma 8.9 and definition (Evans-Spruck). *Let $\mathcal{F} \subset C^0(I\!R^N)$ be a set of real continuous functions. Assume that all level sets of all functions of $\mathcal{F}$ are compact. Let T be a monotone and contrast invariant*

operator defined on $\mathcal{F}$. Let X be any compact set of $\mathbb{R}^N$ and $u \in \mathcal{F}$ any continuous function having X as zero level set : $X = \{\mathbf{x}, u(\mathbf{x}) \geq 0\} = \mathcal{X}_0 u$. Define

$$\mathtt{T}(X) = \{\mathbf{x}, Tu(\mathbf{x}) \geq 0\} = \mathcal{X}_0(Tu). \tag{8.8}$$

Then the set $\mathtt{T}(X)$ does not depend upon the particular choice of u.

Corollary 8.10 *Let $\mathcal{F} \subset C^0(\mathbb{R}^N)$ be a set of real continuous functions. Assume that the level sets of functions of $\mathcal{F}$ are compact. Let T be a monotone and contrast invariant operator defined on $\mathcal{F}$, then the set operator $\mathtt{T}$ defined by (8.8) satisfies*

$$\mathtt{T}(\mathcal{X}_\lambda u) = \mathcal{X}_\lambda(Tu)$$

Proof We can rewrite Relation (8.8) as

$$\mathtt{T}(\mathcal{X}_0 u) = \mathcal{X}_0(Tu)$$

and, replacing u by $u - \lambda$ and using that $T(u - \lambda) = Tu - \lambda$, we in fact have

$$\mathtt{T}(\mathcal{X}_\lambda u) = \mathcal{X}_\lambda(Tu) \tag{8.9}$$

for every continuous bounded function with compact level sets. □

Remark 8.11 *If $u \geq 0$, then $Tu \geq 0$. Indeed, by commutation with the function $g(x) = 0, \forall x$, we have $T(0) = 0$ and by monotonicity, we obtain $Tu \geq 0$.*

Remark 8.12 *Without the compactness assumption on X, the result of Lemma 8.9 can be false. Take for instance $X = \emptyset$. Define $\mathcal{F} = C^0(\mathbb{R}^N)$ and $Tu(\mathbf{x}) = \sup_{\mathbb{R}^N} u(\mathbf{x})$. Take e.g. $v(\mathbf{x}) = -1$. Then $\mathcal{X}_0 v = \emptyset$ and $\mathcal{X}_0(Tv) = \emptyset$. On the other side, choosing $u(\mathbf{x}) = -\frac{1}{1+|\mathbf{x}|}$ we have $Tu(\mathbf{x}) = 0$, so that $\mathcal{X}_0 u = \emptyset$ and $\mathcal{X}_0(Tu) = \mathbb{R}^N$. Thus Lemma 8.9 does not work.*

Lemma 8.9 is easily deduced from the following comparison result.

Lemma 8.13 *Let X be a compact set and u and $\tilde{u}$ two non-positive bounded continuous functions satisfying $X = \{\mathbf{x}, u(\mathbf{x}) \geq 0\} = \{\mathbf{x}, \tilde{u}(\mathbf{x}) \geq 0\}$. Assume that the level sets of u and $\tilde{u}$ are compact. Then there exists a continuous nondecreasing function $\psi : \mathbb{R}^- \to \mathbb{R}^-$ such that $\psi(0) = 0$ and $u \geq \psi(\tilde{u})$.*

Proof of Lemma 8.9
Our aim is to prove, for all continuous functions with compact level sets u and $\tilde{u}$,

$$\mathcal{X}_0 u = \mathcal{X}_0 \tilde{u} \Rightarrow \mathcal{X}_0(Tu) = \mathcal{X}_0(T\tilde{u}). \tag{8.10}$$

Consider the contrast change $g(s) = \max(-1, min(0, u))$. Obviously, for every function u, $\mathcal{X}_0 u = \mathcal{X}_0 g(u)$ and therefore $\mathcal{X}_0 Tu = \mathcal{X}_0(g(Tu)) = \mathcal{X}_0 T(g(u))$. Thus, we may as well prove the implication (8.10) for $g(u)$ and $g(\tilde{u})$. In other terms, we can assume without loss of generality that u and $\tilde{u}$ are nonnegative and bounded. By Lemma 8.13, there exists a continuous nondecreasing function ψ such that $\psi(0) = 0$ and $\psi(\tilde{u}) \leq u$. Since T is monotone and contrast invariant, we deduce that $Tu \geq T(\psi(\tilde{u})) = \psi(T\tilde{u})$. Using $\psi(0) = 0$, we get $T\tilde{u}(\mathbf{x}) = 0 \Rightarrow Tu(\mathbf{x}) = 0$. The converse implication is also true by exchanging the roles of u and $\tilde{u}$ and we conclude that $\{\mathbf{x}, Tu(\mathbf{x}) = 0\} = \{\mathbf{x}, T\tilde{u}(\mathbf{x}) = 0\}$. □

Proof of Lemma 8.13 Since the level sets $\{\mathbf{x}, \tilde{u}(\mathbf{x}) \geq r\}, r \leq 0$ are compact, it is tempting to simply set

$$\tilde{\psi}(r) = \min_{\tilde{u}(\mathbf{x}) \geq r} u(\mathbf{x}).$$

By the definition of $\tilde{\psi}$,

$$u(\mathbf{x}) \geq \min_{\tilde{u}(\mathbf{y}) \geq \tilde{u}(\mathbf{x})} u(\mathbf{y}) = \tilde{\psi}(\tilde{u}(\mathbf{x})).$$

Notice that $\tilde{\psi}(0) = 0$. Let us show that $\tilde{\psi}$ is continuous at 0. Let $r_k \to 0$ be an increasing sequence and $\mathbf{x}_k$ such that $\tilde{u}(\mathbf{x}_k) \geq r_k$ and $\tilde{\psi}(r_k) = u(\mathbf{x}_k)$. Then $\mathbf{x}_k$ belongs to $\{\mathbf{x}, \tilde{u}(\mathbf{x}) \geq r_l\}$ for every $k \geq l$. By compactness of these level sets, a subsequence of $\mathbf{x}_k$ converges to some $\mathbf{x}$ such that $\tilde{u}(\mathbf{x}) \geq r_l$ for every l. Thus $\tilde{u}(\mathbf{x}) = 0$, $\mathbf{x} \in X$ and therefore $u(\mathbf{x}) = 0$. We conclude that $\tilde{\psi}(r_k) = u(\mathbf{x}_k) \to u(\mathbf{x}) = 0$, so that ψ is continuous at 0. Thus $\tilde{\psi}$ matches all requirements of Lemma 8.13 except one : it is not always continuous at all points ! This is easily fixed by choosing a continuous nondecreasing function ψ such that $\tilde{\psi} \geq \psi$ and $\psi(0) = 0$. As an example, $\psi(r) = \frac{1}{|r|}\int_{2r}^{r} \tilde{\psi}(s)ds$, for $r < 0$ answers the question, so that we finally have $u(\mathbf{x}) \geq \tilde{\psi}(\tilde{u}(\mathbf{x})) \geq \psi(\tilde{u}(\mathbf{x}))$. □

Exercise 8.7 *Show that the function ψ defined in the proof of Lemma 8.13 indeed is continuous. Find examples of functions u and $\tilde{u}$ such that $\tilde{\psi}$ is not continuous (such examples can be found in dimension $N = 1$.)*

Corollary 8.14 *(Evans-Spruck extension from periodic functions to periodic sets). Let $\mathcal{F} \subset L_C \cap C^0(\mathbb{R}^N)$ be a set of real continuous $(2, ..., 2)$-periodic functions, i.e. functions satisfying $u(\mathbf{x} + \mathbf{z}) = u(\mathbf{x})$ whenever $\mathbf{z} \in 2\mathbb{Z}^N$. Let T be a monotone and contrast invariant operator defined on $\mathcal{F}$. Then there exists a unique set operator* $\mathtt{T}$ *defined on (2,...,2) periodic sets by*

$$\mathtt{T}(\mathcal{X}_\lambda u) = \mathcal{X}_\lambda(Tu) \tag{8.11}$$

Proof We proceed exactly as the proof of Lemma 8.9. Indeed, the level sets of the considered functions being periodic, we can define their points modulo $\mathbb{Z}^N$, which ensures compactness of sequences of points in a level set. Thus, Corollary 8.10 also holds and T and $\mathtt{T}$ are associated by Relation (8.11). □

8.4 The extension to u.s.c. functions

Digital images should in principle be continuous, by Shannon's digitization theory. Now, Shannon's theory does not take into account the quantization of grey levels, which gives us only a finite set of level sets of the image (about 255 in all day technology). If we process each level separately, as is askable if we do contrast invariant processing, then we consider *de facto* the image as discontinuous along its quantized level lines. Assume that we have defined a contrast invariant monotone operator on continuous functions. We have just proved that such an operator has an extension to compact sets. It is therefore natural to extend, by the threshold superposition principle, this operator to upper semicontinuous functions. This extension will be in practice necessary for digital image processing and we shall use it to extend (e.g.) the curvature motion to u.s.c. functions.

Definition 8.15 *We say that a function $u : \mathbb{R}^N \to \mathbb{R}$ is upper semicontinuous (u.s.c.) if one of the following equivalent items is true.*
(i) the upper level sets $\mathcal{X}_\lambda^+ u = \{\mathbf{x}, u(\mathbf{x}) \geq \lambda\}$ are all closed.
(ii) at every $\mathbf{x} \in \mathbb{R}^N$, $u(\mathbf{x}) \geq \limsup_{\mathbf{y}\to\mathbf{x}} u(\mathbf{y})$.
We say that u is lower semicontinuous (l.s.c.) if $-u$ is u.s.c. A real function u is continuous if and only if it is both lower and upper semicontinuous.

Exercise 8.8 *Show that the equivalence of (i) and (ii) in Definition 8.15 is indeed true. Deduce that every u.s.c. function attains its maxima and every l.s.c. function attains its minima.*

Theorem 8.16 *Let $\mathcal{F} \subset C^0(\mathbb{R}^N)$ be a set of continuous functions whose level sets are compact (or $(2, ..., 2)$-periodic) and T a contrast invariant monotone operator on $\mathcal{F}$. Let $\mathtt{T}$ be the set operator associated with T by the Evans-Spruck method. Then T can be extended into a monotone contrast invariant operator, defined for all upper semi-continuous functions whose level sets are compact (periodic) and commuting with almost all thresholds :*

$$\mathcal{X}_\lambda(Tu) = \mathtt{T}(\mathcal{X}_\lambda u), \ \ a.e. \ in \ \lambda, \ \ a.e. \ in \ \mathbf{x}.$$

In the same way, T can be extended into a contrast invariant operator T^-, defined for all lower semi-continuous functions whose lower level sets are compact (or periodic) and commuting with almost all lower thresholds $\mathcal{X}_\lambda^-$:

$$\mathcal{X}_\lambda^-(T^- u) = \mathtt{T}(\mathcal{X}_\lambda^- u) \ a.e. \ in \ \lambda, \ \ a.e. \ in \ \mathbf{x}.$$

Remark 8.17 *The extensions T and T^- can be quite different, as shown by the checkerboard experiment (Figure 24.9).*

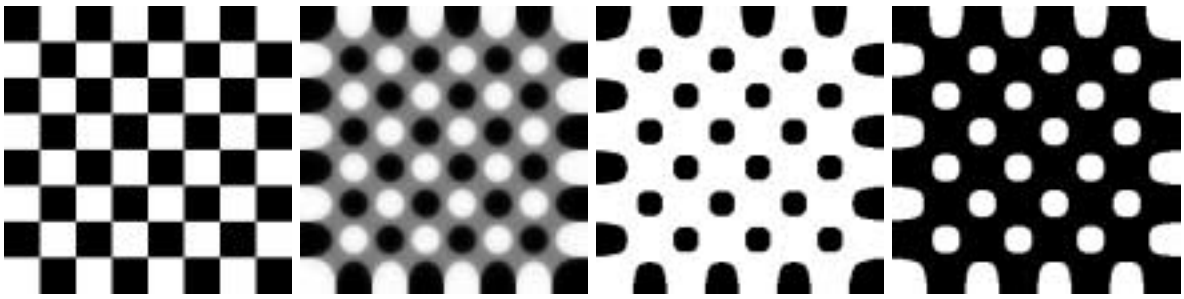

Figure 8.1: The chessboard dilemna. Left: chessboard image. Next: result with the finite difference scheme (FDS, Chapter 24) of the curvature motion, applied up to a fixed scale. The creation of a new gray level proves that the scheme is not fully contrast invariant. Indeed, by Proposition 7.4, a contrast invariant operator does not create new levels. The new observed gray level corresponds to an average of the existing ones, black and white. The next two images are obtained by the Evans-Spruck extension of the curvature motion, first under the assumption that the image is u.s.c. and second under the l.s.c. assumption. Thus, the schemes are in both cases fully contrast invariant and are extensions of the curvature motion as specified in Theorem 8.16.

Proof of Theorem 8.16 By Corollary 8.10, we can define $\mathtt{T}$ on compact sets X in such a way that for any $u \in \mathcal{F}$ and $\lambda \in I\!R$,

$$\mathtt{T}(\mathcal{X}_\lambda u) = \mathcal{X}_\lambda(T(u)) \tag{8.12}$$

For any u.s.c function (with compact level sets) u we define T' by the threshold superposition principle,

$$T'(u)(\mathbf{x}) = \sup\{\lambda, \mathbf{x} \in \mathtt{T}(\mathcal{X}_\lambda u)\}. \tag{8.13}$$

Let us now prove that $T'(u) = T(u)$ for any continuous function with compact level sets. This immediately follows from (8.12) and the definition of T':

$$T'(u)(\mathbf{x}) = \sup\{\lambda, \mathbf{x} \in \mathtt{T}(\mathcal{X}_\lambda u)\} = \sup\{\lambda, \mathbf{x} \in \mathcal{X}_\lambda T(u)\} = T(u).$$

In addition, it follows immediately from (8.13) and Proposition 7.9 that T' is monotone and contrast invariant on the u.s.c functions and commutes with almost all thresholds, $\mathcal{X}_\lambda(Tu) = \mathtt{T}(\mathcal{X}_\lambda u)$.

□

Exercise 8.9 *By Theorem 7.11, T' is contrast invariant on the u.s.c. functions if and only if* $\mathtt{T}$ *is upper semicontinuous. It is easily seen (see Exercise 8.11) that this is true if and only if* $\mathtt{T}$ *satisfies, for every nonincreasing family of compact sets X_λ, the implication*

$$(X_\lambda = \bigcap_{\mu<\lambda} X_\mu) \Rightarrow (\mathtt{T}(X_\lambda) = \bigcap_{\mu<\lambda} \mathtt{T}(X_\mu)).$$

Let us set $u(\mathbf{x}) = \sup\{\lambda, \mathbf{x} \in X_\lambda\}$ and assume that u is continuous. Then we obtain

$$\mathtt{T}(X_\lambda) = \mathtt{T}(\mathcal{X}_\lambda u) = \mathcal{X}_\lambda T(u) = \bigcap_{\mu<\lambda} \mathcal{X}_\mu T(u) = \bigcap_{\mu<\lambda} \mathtt{T}(\mathcal{X}_\mu u) = \bigcap_{\mu<\lambda} \mathtt{T}(X_\mu).$$

Now, u may well be discontinuous. In order to extend the preceding argument, we set, for some fixed λ and $\mu \leq \lambda$,

$$Y_\mu = \{\mathbf{x},\ dist(\mathbf{x}, X_\mu) \leq \lambda - \mu\}.$$

Consider the function v defined by $v(\mathbf{x}) = \sup\{\mu, \mathbf{x} \in Y_\mu\}$. Let us show that v is a Lipschitz function and therefore continuous. For any $\mathbf{x}_0$ and $\mathbf{y}_0$ such that $v(\mathbf{y}_0) > v(\mathbf{x}_0)$, we have $\mathbf{y}_0 \in Y_{u(\mathbf{y}_0)}$ and $\mathbf{x}_0 \notin Y_{(v(\mathbf{y}_0)+v(\mathbf{x}_0))/2}$. Thus, by the definition of Y_μ again,

$$\mathbf{x}_0 \notin \{\mathbf{x}, \exists \mathbf{y} \in B(\mathbf{x}, (v(\mathbf{y}_0) - v(\mathbf{x}_0))/2),\ u(\mathbf{y}) \geq (v(\mathbf{y}_0) + v(\mathbf{x}_0))/2\}$$

This implies that

$$dist(\mathbf{x}_0, Y_{u(\mathbf{y}_0)}) \geq (v(\mathbf{y}_0) - v(\mathbf{x}_0))/2$$

and therefore

$$|\mathbf{x}_0 - \mathbf{y}_0| \geq \frac{1}{2}|v(\mathbf{y}_0) - v(\mathbf{x}_0)|.$$

Thus v is continuous and we obtain

$$\mathtt{T}(Y_\lambda) = \bigcap_{\mu<\lambda} \mathtt{T}(Y_\mu).$$

Now, $\forall \mu \leq \lambda$, $X_\mu \subset Y_\mu$, and by the monotonicity of $\mathtt{T}$ *on compact sets,* $\mathtt{T}(X_\mu) \subset \mathtt{T}(Y_\mu)$. *We then have*

$$\bigcap_{\mu<\lambda} \mathtt{T}(X_\mu) \subset \bigcap_{\mu<\lambda} \mathtt{T}(Y_\mu) = \mathtt{T}(Y_\lambda) = \mathtt{T}(X_\lambda).$$

Conversely, $\mathtt{T}(X_\lambda) \subset \bigcap_{\mu<\lambda} \mathtt{T}(X_\mu)$ *by monotonicity of* $\mathtt{T}$. *We conclude that T' is contrast invariant on the u.s.c function with compact level sets. T' also is monotone by the monotonicity of* $\mathtt{T}$.

Remark 8.18 **$I\!B$ may change with the extension !** *Let us start with a set of structuring elements $I\!B$ and its associated operator*

$$Tu(\mathbf{x}) = \inf_{B \in I\!B} \sup_{\mathbf{y} \in B} u(\mathbf{x}+\mathbf{y}).$$

Then T is contrast invariant and can receive the contrast invariant extension of Theorem 8.16. Now, contrast invariant operators have an inf sup *form, so that we can assert the existence of a set $I\!B'$ such that*

$$Tu(\mathbf{x}) = \inf_{B \in I\!B'} \sup_{\mathbf{y} \in B} u(\mathbf{x}+\mathbf{y}).$$

It is easily seen that $I\!B'$ must be in some cases different from $I\!B$. As an example, if $Tu(\mathbf{x}) = \sup_{\mathbf{y} \in B(0,1)} u(\mathbf{x}+\mathbf{y})$ is a dilation by an open ball $B(0,1)$, then the extension T' to u.s.c. functions is associated with (e.g.) $I\!B' = \{\overline{B}(0,1)\}$, the closure of the ball. The original $I\!B$ cannot work, since T' must transform a closed level set into a closed level set, while the dilation by an open ball transforms any set into an open set.

8.5 The who's who of monotone and contrast invariant operators.

Let us summarize what we have proved in the preceding section.
First case : $\mathcal{F}$ is a set of functions whose level sets are $\mathcal{T}$ and $\mathcal{F}$ contains all characteristic functions of elements of $\mathcal{T}$.

Consider monotone set operators $\mathtt{T}$ (with $T(\emptyset) = \emptyset$, $T(I\!R^N) = I\!R^N$) and monotone contrast invariant function operators T. Then, we can either define T from $\mathtt{T}$, by the threshold superposition principle,

$$Tu(\mathbf{x}) = \sup\{\lambda,\ \mathbf{x} \in \mathtt{T}(\mathcal{X}_\lambda u)\} \tag{8.14}$$

or define $\mathtt{T}$ from T, provided the domain $\mathcal{F}$ of T contains the characteristic functions, by thresholding :

$$\mathtt{T}(X) = \mathcal{X}_1(T(\mathbb{1}_X)). \tag{8.15}$$

The relation between T and $\mathtt{T}$ is also characterized by the commutation with thresholds,

$$\mathtt{T}(\mathcal{X}_\lambda u) = \mathcal{X}_\lambda(Tu). \tag{8.16}$$

This relation being true almost everywhere in λ and $\mathbf{x}$. Every contrast invariant operator T has a sup inf formulation,

$$Tu(\mathbf{x}) = \sup_{B \in I\!B_{\mathbf{X}}} \inf_{\mathbf{y} \in B} u(\mathbf{y}). \tag{8.17}$$

where $\mathbb{B}$ is defined from $\mathtt{T}$ by the "impulse response" formula,

$$\mathbb{B} = \{B,\ 0 \in T(B)\}. \tag{8.18}$$

For the same $\mathbb{B}$, we also have, by Proposition 8.7, the analytic form (8.1), that is

$$\mathtt{T}(X) = \bigcup_{B \in \mathbb{B}} \bigcap_{\mathbf{y} \in B} (X - \mathbf{y}).$$

We can summarize the full equivalences proved in one sentence.

It is equivalent to define : a set monotone operator $\mathtt{T}$ *with* $\mathtt{T}(\emptyset) = \emptyset$*, a contrast invariant operator* T *or an* inf sup *operator. Each one of the forms of the operator is deduced from each other one with the help of Formulas (8.14-8.18) above.* This equivalence between the three modes of definition is summarized in the first column of Figure 8.2. It is sometimes easier and more intuitive to define an image operator as a set operator: we have seen or shall see several examples : the "extrema killer", the median filter and the classical mathematical morphology operators (erosions, dilations, etc.).

The case of contrast invariant operators defined on a set of continous functions.

Now, we shall also have cases, namely the case of solutions of partial differential equations, where the operator is much easier to define as an operator on a set of continuous functions. This will be the case for the solutions of the curvature motion and the affine morphological scale space. In such cases, the Evans-Spruck method applies and we directly define an extension of the operator T as a compact set operator by Relation (8.16), the relation being then true everywhere (for all λ and all $\mathbf{x}$). Then, the operator T receives an extension $\tilde{T}$ to upper semicontinuous functions, which still satisfies (8.16). In one sentence,

Given a contrast invariant operator T *defined on continuous functions, we can define an extension* $\tilde{T}$ *which satisfies all of equivalence properties of the first column of Figure 8.2.* This situation is again summarized in the second and third column of the mentionned figure : from the contrast invariance on continuous functions, we go up to define a set operator on compact sets by thresholding and then a function operator on u.s.c. functions.

Exercise 8.10 *Let* T *and* $\mathtt{T}$ *be respectively a contrast invariant monotone operator and a set operator such that for all* λ*,*

$$\mathcal{X}_\lambda(Tu) = \mathtt{T}(\mathcal{X}_\lambda(u)).$$

Show that $\mathtt{T}$ *is upper semicontinuous. Show that* T *commutes with all u.s.c. nondecreasing functions* g*, i.e.* $T(g(u)) = g(Tu)$*.*

Exercise 8.11 *Let* $\mathtt{T}$ *be set operator such that* $\mathtt{T}(\emptyset) = \emptyset$*,* $\mathtt{T}(\mathbb{R}^N) = \mathbb{R}^N$*. Show that* $\mathtt{T}$ *is upper semicontinuous if and only if, for any family of sets* $(X_\lambda)_{\lambda \in \mathbb{R}}$ *such that* $X_\lambda = \bigcap_{\mu<\lambda} X_\mu$*, one has* $\mathtt{T}(X_\lambda) = \bigcap_{\mu<\lambda} \mathtt{T}(X_\mu)$*.*

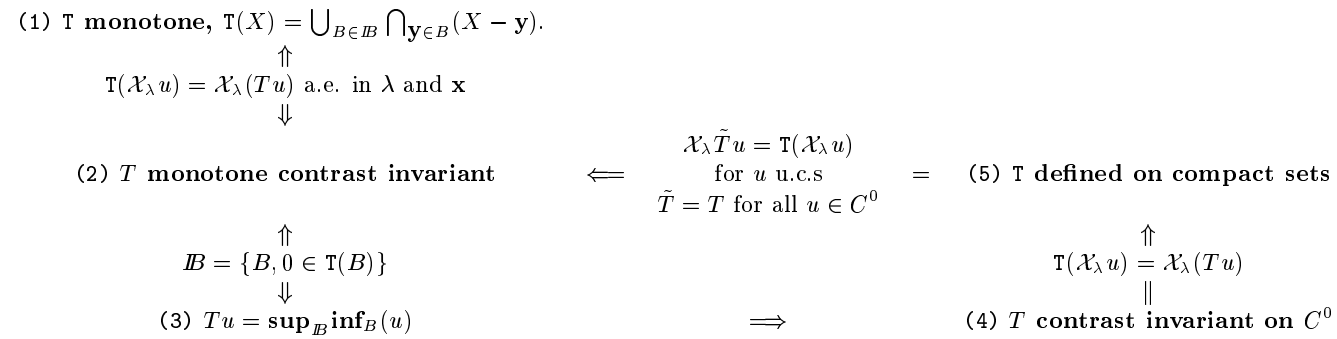

Figure 8.2: Relations between set and function operators. (All operators are monotone and translation invariant - See text for exact assumptions).

(1) $\Leftrightarrow$ (2): Propositions 7.8 and 7.9, (2) $\Rightarrow$ (3): Theorems 8.3 and 8.4, (3) $\Rightarrow$ (4): Theorem 8.4, (4) $\Rightarrow$ (5): Lemma 8.9, Corollary 8.10, (5) $\Rightarrow$ (2): Theorem 8.16.

References.

The formalism presented in this chapter is due to Matheron [290] in the case of the set operators and to Serra [378] and Maragos [279] in the case of function operators. Actually, the Serra formalism is more general than the one presented here and will be developped in Chapter 16 on "non-flat" morphology. Maragos proved the sup-inf formula given here for operators commuting with thresholds. Our presentation is original in relating directly the sup-inf form to contrast invariance and establishing the full equivalence between sup-inf operators and contrast invariant monotone operators. The mysterious "set of structuring elements" has retained a lot of attention in the literature, where the ways to find the right sets of structuring elements [363, 397], or to simplify them, [374], or to decompose them into simpler ones as one does with linear filters [334, 451, 452], or to reduce their number [351]

Chapter 9

Erosions and dilations

9.1 Erosions and dilations as multiscale contrast invariant operators

In this chapter, we study with some detail the simplest operators of mathematical morphology. We have seen in the previous chapter that we can associate with every contrast invariant and monotone operator an "inf-sup" or a "sup-inf" form. It is natural to study first the operators whose set of structuring elements, $\mathbb{B}$, is a singleton. In that case, the operators, if they are translation invariant, are simply written as

$$Tu(\mathbf{x}) = \sup_{\mathbf{y} \in B + \mathbf{x}} u(\mathbf{y}) \quad \text{or} \quad Tu(\mathbf{x}) = \inf_{\mathbf{y} \in B + \mathbf{x}} u(\mathbf{y}).$$

In the first case, we call the operator a "dilation" and in the second one, an "erosion". These operators can be defined on subsets of $\mathbb{R}^N$ as well. Since we do not wish to be dependent upon a particular choice as for the size of B, we shall introduce a scale parameter t and consider the family of erosions or dilations by tB.

Definition 9.1 *Let B be a subset of $\mathbb{R}^N$. Let $t \geq 0$ be a scale parameter. We call dilation with structuring element B and scale t of a subset X of $\mathbb{R}^N$ the set*

$$D_t X = X + tB = \{\mathbf{x} + \mathbf{y}, \mathbf{x} \in X,\ \mathbf{y} \in tB\} = \bigcup_{\mathbf{y} \in tB} (X + \mathbf{y}) \tag{9.1}$$

We call erosion with structuring element B and scale t of a set X the set

$$E_t X = \{\mathbf{x}, \mathbf{x} + tB \subset X\} = \bigcap_{\mathbf{y} \in -tB} (X + \mathbf{y}). \tag{9.2}$$

This definition is made in such a way that E_t and D_t be somehow inverse of each other. This is the case if, e.g., $B = \{\mathbf{x}_0\}$ is reduced to a point. Then D_t is the translation by $t\mathbf{x}_0$ and E_t by $-t\mathbf{x}_0$. When B is an open ball centered at 0, we notice that the dilation of X at scale t is nothing but its $t-$neighborhood, the set of points which lie at a distance less than t from X. When B is a symmetric set with respect to 0, one also defines an operator called "opening at scale t", by composing D_t and E_t and a "closing at scale t" operator by $E_t D_t$. These names have the following origin. If B is (e.g.) an open ball centered at 0 with radius 1, then the opening at scale t of a set X is the union of all balls with radius t contained in

X. Now, the classical, topological, opening of X is the union of all open balls contained in X. Thus, the opening at scale t appears as a quantified opening. Besides, the classical statement that "the closing of the complementary is equal to the complementary of the opening" remains true, as shows the next exercise.

Exercise 9.1 *1) Let $B = B(0,1)$ the unit open ball. Show that $D_t E_t$ is the union of all open balls with radius t contained in X.*
2) We note $X^c = \mathbb{R}^N \setminus X$. Show that for any structuring element B, symmetric with respect to 0, $D_t(X^c) = (E_t X)^c$.

Figure 9.1: Dilation of a set. A set X (in black), its dilation by a ball of radius 20, and the difference set.

Figure 9.2: Erosion of a set. A set X (in black), its erosion by a ball of radius 20, and the difference set.

Definition 9.2 *The dilation at scale t with structuring element B of a function u_0 is defined by*

$$D_t u_0(\mathbf{x}) = \sup_{\mathbf{y} \in tB} u_0(\mathbf{x} - \mathbf{y}),$$

Similarly, the erosion at scale t with structuring element B of a function f is defined by

$$E_t u_0(\mathbf{x}) = \inf_{\mathbf{y} \in -tB} u_0(\mathbf{x} - \mathbf{y})$$

Exercise 9.2 *Show that if B is symmetric with respect to 0, then $E_t(-u) = D_t(u)$.*

We have defined independently set or function erosions and dilations. Let us now see under which conditions the definitions coincide. We can use the equivalence scheme described in Figure 8.2 : for the

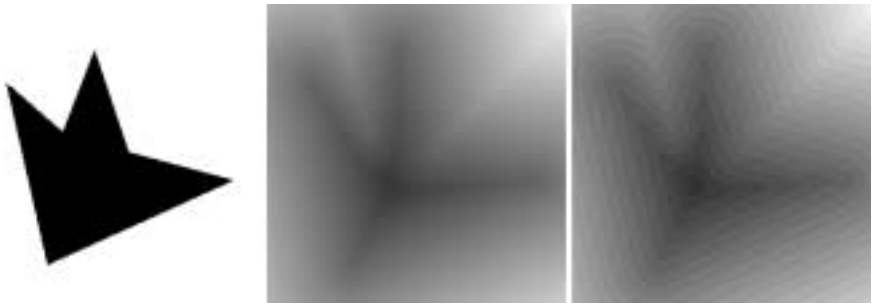

Figure 9.3: Distance image.
Dilations and erosions as level sets of the signed distance function. Left : a set X, middle : the signed distance to X, $u(\mathbf{x}) = d(\mathbf{x}, X)$ if $\mathbf{x} \in X^c$, $u(\mathbf{x}) = -d(\mathbf{x}, X)$, if $\mathbf{x} \in X$, right : by quantizing the grey levels of u, as done on the right, one can easily check that the level sets of u are dilated or eroded of X. Compare with Figures 9.1 and 9.2.

dilation, we obviously take $I\!B = \{-tb,\ b \in B\}$. We get

$$D_t u(\mathbf{x}) = \sup_{\mathbf{y} \in -tB} u(\mathbf{x} + \mathbf{y}) \leftrightarrow D_t X = \bigcup_{\mathbf{y} \in -tB} (X - \mathbf{y}) = \bigcup_{\mathbf{y} \in tB} X + \mathbf{y},$$

where the double arrow $\leftrightarrow$ means that the set operator on the right and the function operator on the left are in a T-T relation, i.e. the threshold superposition principle applies and we have

$$D_t(\mathcal{X}_\lambda u) = \mathcal{X}_\lambda(D_t u) \tag{9.3}$$

for every real function u, this relation being true $\mathbf{x}$-almost everywhere, for almost every $\lambda \in I\!R$. In the same way, we take for the erosion $I\!B = \{-tB\}$, a single structuring element, and we obtain

$$E_t u(\mathbf{x}) = \inf_{\mathbf{y} \in tB} u(\mathbf{x} + \mathbf{y}) \leftrightarrow E_t X = \bigcap_{\mathbf{y} \in tB} (X - \mathbf{y})$$

by applying the equivalence scheme. Thus, we have proved the following theorem.

Theorem 9.3 *The pair : set dilation, function dilation and the pair set erosion, function erosion, denoted respectively by D_t and E_t, satisfy the threshold superposition principle and we have*

$$D_t(\mathcal{X}_\lambda u) = \mathcal{X}_\lambda(D_t u) \text{ and } E_t(\mathcal{X}_\lambda u) = \mathcal{X}_\lambda(E_t u),\ \textit{a.e. in } \mathbf{x},\ \textit{a.e. in } \lambda. \tag{9.4}$$

Exercise 9.3 *Let u be a real continuous function on $I\!R^N$ and consider the dilation D_t associated with open unit ball. From Formula 9.4, we have $D_t(\mathcal{X}_\lambda u) = \mathcal{X}_\lambda(D_t u)$ for almost every λ and almost everywhere in $\mathbf{x}$. Prove that the set on the left is open and that the set on the right is closed. Deduce that the relation will be true for every $\mathbf{x}$ only if both sets are empty, or equal to $I\!R^N$.*

The preceding exercise shows that we cannot expect perfect commutation of dilations with thresholds. The next proposition gives, however, a framework where that property is true.

Proposition 9.4 *Let B be a compact (bounded and closed) subset of $I\!R^N$. Let $\mathcal{T}$ be the set of closed parts of $I\!R^N$ and $\mathcal{F}^+$ be the set of upper semi-continuous functions, that is the set of functions whose upper level sets $\mathcal{X}_\lambda u = \{\mathbf{x}, u(\mathbf{x}) \geq \lambda\}$ are in $\mathcal{T}$. Let $\mathcal{F}$ be the set of lower semi-continuous functions, that is,*

functions whose lower level sets $\mathcal{X}_\lambda^- u = \{\mathbf{x}, u(\mathbf{x}) \leq \lambda\}$ *are in* $\mathcal{T}$. *Then for any function* u *in* $\mathcal{F}^+$ *one has* $D_t(\mathcal{X}_\lambda u) = \mathcal{X}_\lambda(D_t u)$ *and for any function in* $\mathcal{F}^-$, $E_t(\mathcal{X}_\lambda^- u) = \mathcal{X}_\lambda^-(E_t u)$. *In other terms, it is equivalent to directly dilate (resp. erode)* u *or to dilate (resp. erode) first each level set of* u *and then define as dilation (resp. erosion) of* u *the function which has these level sets.*

Proof of Proposition 9.4. We have $\mathbf{x} \in \mathcal{X}_\lambda(D_t(u))$ if and only if $\sup_{\mathbf{y} \in tB} u(\mathbf{x} - \mathbf{y}) \geq \lambda$. Since u is u.s.c and therefore attains its maxima on compact sets, this is equivalent to $\exists \mathbf{y} \in tB, u(\mathbf{x} - \mathbf{y}) \geq \lambda$ and, setting $\mathbf{z} = \mathbf{x} - \mathbf{y}$, to $\exists \mathbf{z} \in \mathcal{X}_\lambda(u)$ and $\mathbf{y} \in tB,\ \mathbf{x} = \mathbf{z} + \mathbf{y}$, which means $\mathbf{x} \in \mathcal{X}_\lambda(u) + tB$ and is therefore equivalent to $\mathbf{x} \in D_t(\mathcal{X}_\lambda(u))$. □

We now look for a property which is desirable for a family of scaled operators, the recursivity. We say that the erosions (or the dilations) associated with a structuring element B are a recursive family of operators if $E_{t+s} = E_t \circ E_s$ (resp. $D_{t+s} = D_t \circ D_s$.) This property is also called "semigroup" property. It implies that $D_t = (D_{\frac{t}{n}})^n$, a very useful computational property.

Proposition 9.5 *The erosions and dilations are recursive if and only if their structuring element* B *is convex.*

Lemma 9.6 *One has* $(t+s)B = tB + sB$ *for any* $s \geq 0$ *and* $t \geq 0$ *if and only if* B *is convex.*

Proof of Lemma 9.6 : If B is convex, then by definition of the convexity, for any s and t, $sB + tB$ is included in $(s+t)B$, and the reverse inclusion is obvious. Conversely, assume that $sB + tB$ is contained in $(s+t)B$. Then, for any $\mathbf{x}$ and $\mathbf{y}$ in B one can find $\mathbf{z}$ in B such that $(s+t)\mathbf{z} = s\mathbf{x} + t\mathbf{y}$. This means that the barycenter of $\mathbf{x}$ and $\mathbf{y}$ with weights $\frac{s}{s+t}$ and $\frac{t}{s+t}$ also is in B. Since this is true for any positive s and t, we deduce that B is convex. □

Proof of Proposition 9.5 : We do the proof for the dilation. We have

$$D_t D_s X = (X + sB) + tB = X + sB + tB$$

and

$$D_{s+t} X = X + (s+t)B$$

Since we can take $X = \{0\}$, we deduce that the dilation is recursive if and only if $(t+s)B = tB + sB$. By Lemma 9.6, this is true if and only if B is convex. □

9.2 P.D.E.'s associated with erosions and dilations

Multiscale dilations and erosions are canonically associated with a partial differential equation. Let us denote by $||\mathbf{x}||_B$ the gauge associated with a convex set B containing 0, that is

$$||\mathbf{y}||_B = \sup_{\mathbf{z} \in B} \mathbf{y}.\mathbf{z},$$

Figure 9.4: Opening of a set as curvature threshold from above. A set X, its opening by a ball of radius 20, and the difference between original and processed. This opening transforms X into the union of all balls of radius 20 contained in it. The resulting operation can be understood as a threshold from above of the curvature of the set boundary.

Figure 9.5: Closing of a set as a curvature threshold, from below. A set X, its closing by a ball of radius 20, and the difference between original and processed. The closing of X is nothing but the opening of X^c. It can be viewed as a threshold from below of the curvature of the set boundary.

where $\mathbf{y}.\mathbf{z}$ denotes the Euclidean scalar product of $\mathbf{y}$ and $\mathbf{z}$. When B is a ball centered at 0 and with radius 1, $||.||_B$ is the usual Euclidean norm.

Proposition 9.7 *(Hopf-Lax formula [258][141]) Assume that B is a bounded convex set containing 0. Set $u(t,\mathbf{x}) = D_t u_0(\mathbf{x})$ (resp. $E_t u_0(\mathbf{x})$). Then $u(t,\mathbf{x})$ satisfies*

$$\partial u/\partial t = ||Du||_{-B}.$$

(resp. $\partial u/\partial t = -||Du||_{-B}$) at each point $(t,\mathbf{x})$ where u is C^1 with respect to t and $\mathbf{x}$.

Proof. Let us first show the property at $t = 0$. Assume that u_0 is C^1 at $\mathbf{x}$. We have $u(t,\mathbf{x}) = D_t u_0(\mathbf{x})$ and $u(0,\mathbf{x}) = u_0(\mathbf{x})$. Thus

$$u(h,\mathbf{x}) - u(0,\mathbf{x}) = \sup_{\mathbf{y}\in hB} (u_0(\mathbf{x}-\mathbf{y}) - u_0(\mathbf{x}))$$

Since $u_0(\mathbf{x})$ is C^1 at $\mathbf{x}$ and B is bounded, we get

$$\begin{aligned} u(h,\mathbf{x}) - u(0,\mathbf{x}) &= \sup_{\mathbf{y}\in hB} (-Du_0(\mathbf{x}).\mathbf{y}) + o(h) \\ &= h \sup_{\mathbf{z}\in -B} (Du_0(\mathbf{x}).\mathbf{z}) + o(h). \end{aligned}$$

By dividing by h and passing to the limit as h tends to zero, we get

$$\frac{\partial u}{\partial t}(0,\mathbf{x}) = ||Du_0(\mathbf{x})||_{-B}, \tag{9.5}$$

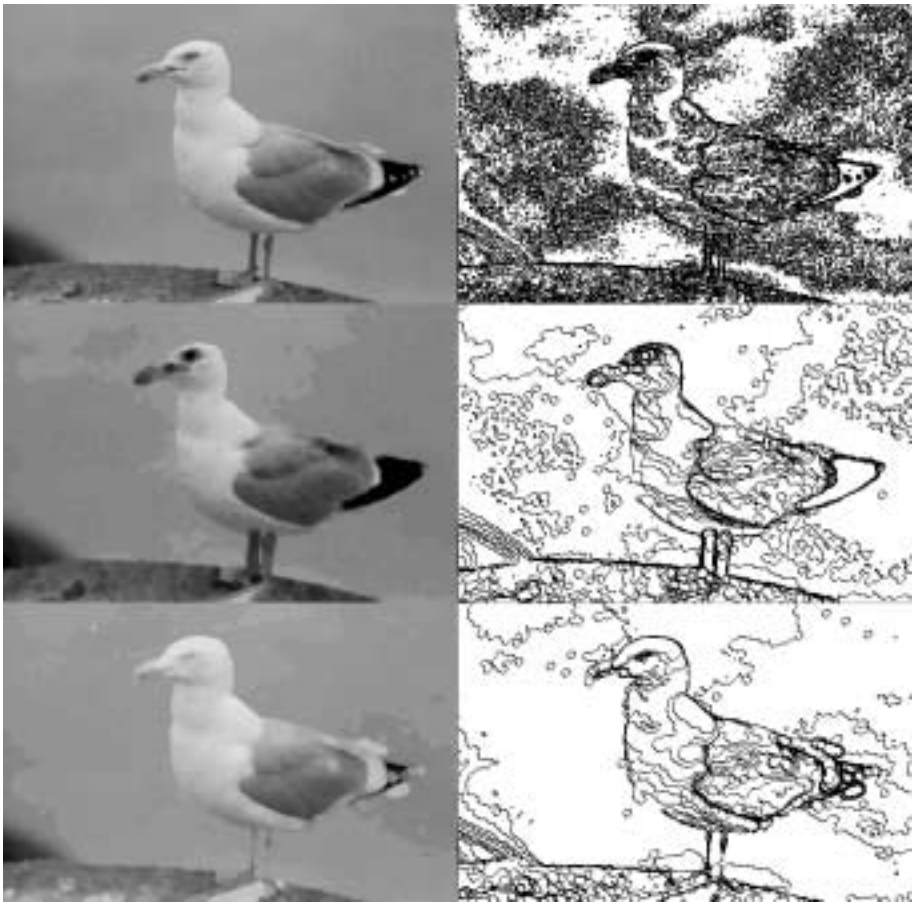

Figure 9.6: Erosion and dilation of a real image. On the first row, a sea bird image, and its level lines for all levels multiple of 12. On the second row, an erosion with radius 4 has been applied. On the right, the resulting level lines where the circular shape of the structuring element (a disk with radius 4) appears around each local minimum of the original image. Erosion removes local maxima (in particular, all small white spots), but expands minima. Thus, all dark spots, like here the eye of the bird, are expanded. The third row displays the effect of a dilation with radius 4, and the resulting level lines. We see how local minima are removed (see e.g. the eye of the bird) and how white spots on the tail expand. Here, in turn, circular level lines appear around all local maxima of the original image.

which is the announced equation in the particular case where $t = 0$. Let us prove the same relation at an arbitrary scale t. Since $D_{t+h} = D_t D_h = D_h D_t$, we can write

$$u(t+h, \mathbf{x}) - u(t, \mathbf{x}) = D_h(u(t))(\mathbf{x}) - u(t)(\mathbf{x}).$$

We divide by h, let h tend to zero and apply the preceding result with $u(t)$ instead of u_0. This yields the general equation. □

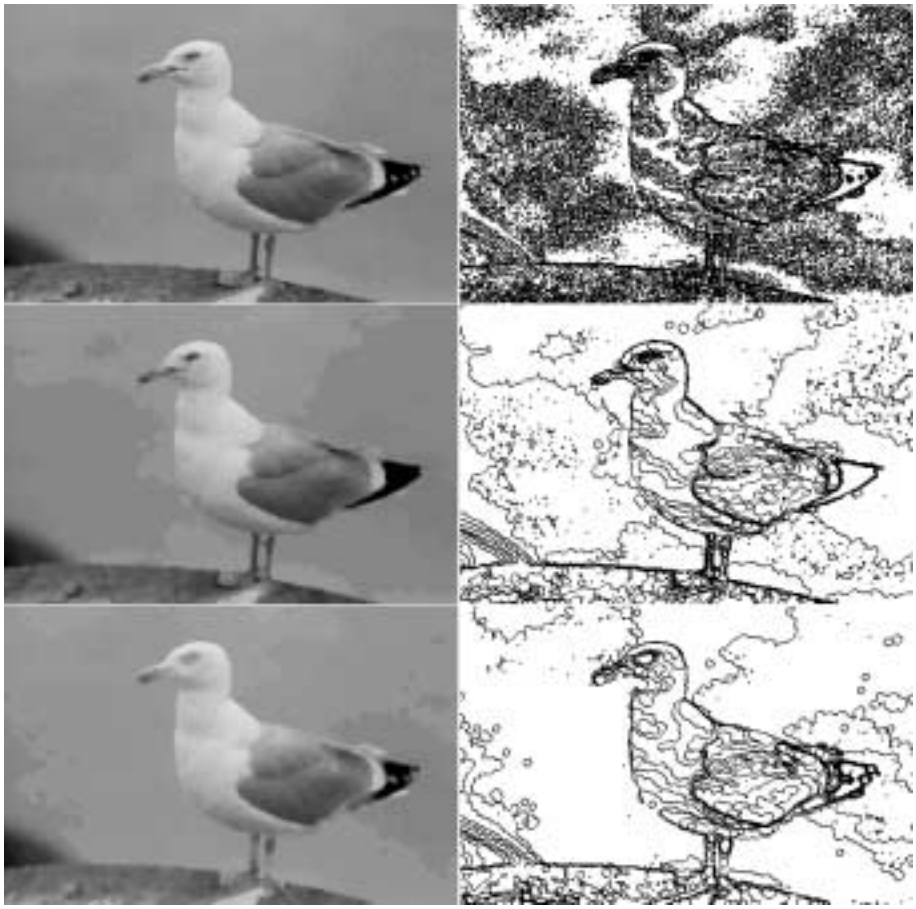

Figure 9.7: Opening and closing of a real image. On the first row, the original image and its level lines for all levels multiple of 12. On the second row (resp. third row), an opening (resp. closing) with radius 4 has been applied. In the level-lines displayed on the right, we can recognize the circular shape of the structuring element.

Figure 9.8: Opening and closing based denoising. On the first row : scanned picture of the word "operator" with black dots and a black line added, a dilation with a 2×5 rectangle (middle), then an erosion with the same structuring element. The resulting operator is a closing. Small black structures are removed by a such process. In the second row : the word "operator" with a white line and white dots inside the letters, erosion with a rectangle 2×5, followed by a dilation. The resulting operator is a opening. Small white structures are removed.

References.

Matheron introduced dilations and erosions as a useful toof for set and shape analysis in his seminal book [290]. A full account of the set and function properties of dilations, erosions, openings and closings can be found in the classical books of Serra [378, 380]. See also the introductive paper by Haralick and

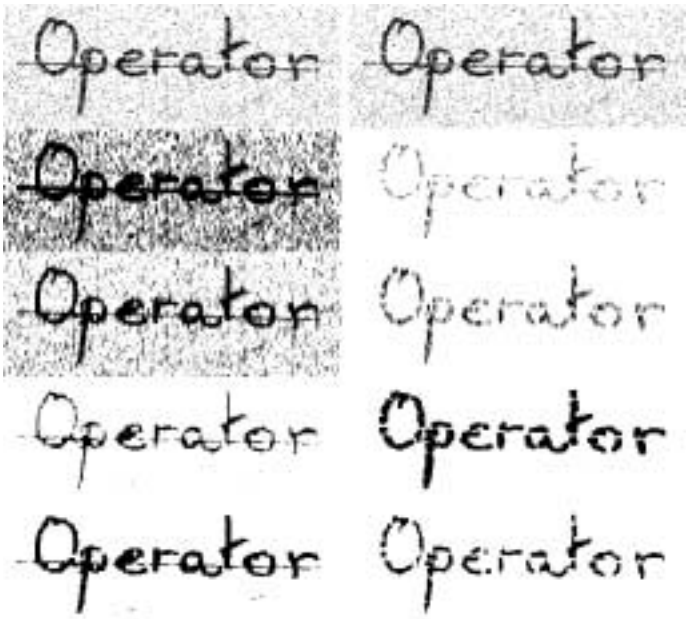

Figure 9.9: Examples of denoising based on opening or closing, as in Figure 9.8. Here are added on the "operator" image some perturbations made of both black and white lines or dots. In the first column, up-down : original picture, erosion with a 1×3 rectangle, then dilation with the same structuring element. In other terms opening with this rectangle. In continuation, a dilation is applied with a rectangle 3×1, and finally an erosion with the same rectangle. In the second column, the same process is applied, but with erosions and dilations exchanging their roles. It does not work so well because closing expands white perturbations and opening expands black perturbations : those operators do not commute. See Figure **??**, where the median filter applies with more success.

al. [195] and the early Nakagawa and Rosenfeld [314]. An algebraic axiomatics for erosions, dilations, openings and closings has been developped by Ronse and Heijmans [201, 352]. We did not expose here this algebraic point of view. The obvious relationships between dilations, erosions of a set and the distance function have been numerically exploited in [211], [386] and [245]. The skeleton of a shape can be defined as the set of points where the distance function to the shape is singular. A numerical procedure computing the skeleton in this way is proposed in [247]. The relationship between multiscale dilations or erosions and the Partial Differential Equations $\frac{\partial u}{\partial t} = \pm |Du|$ is known by the work of Lax [258] where it is used to give efficient stable numerical schemes for the P.D.E.. Rouy and Tourin have shown that the distance function to a shape is a viscosity solution of $1 - |Du| = 0$, with null boundary condition (Dirichlet condition) on the boundary of the shape. In order to define efficient numerical schemes for computing the distance function, they actually implement the evolution equation $\frac{\partial u}{\partial t} = 1 - |Du|$, starting from 0 and with null boundary condition on the boundary of the shape. The fact that the multiscale dilations and erosions can be computed by the partial differential equation $\frac{\partial u}{\partial t} = \pm |Du|$ has been rediscovered or revived , thirty years after Lax, in several papers, namely Alvarez et al. [15], Boomgard and Smeulders [411], Maragos [281, 282]. As a curve evolution algorithm, see [366]. As a strange matter of fact, the link between erosions, dilations and their P.D.E.'s seems to have remained unknown or unexploited until 1992. The erosion and dilation P.D.E.'s can used for shape thinning, a popular way to compute the skeleton. Pasquignon [335] developped an erosion P.D.E. with adaptive stopping time, permitting to computed directly a skeleton which does not look like barbed wire.

Chapter 10

Median filters and mathematical morphology

In the whole chapter, we consider real functions denoted by $u(\mathbf{x})$ and defined for $\mathbf{x} \in I\!R^N$. We make no assumption at all about their regularity, except for one thing : we shall use the Lebesgue measure of the level sets of u. Thus, we need to assume that the level sets of u are measurable, which is equivalent to say that u is Lebesgue measurable. For the mathematical treatment of measurability, we refer to classic books like (e.g.) ([**?**]). The Lebesgue measure is the simplest mathematically correct definition of the intuitive concept of length in dimension 1, area in dimension 2 or volume in dimension 3. Thus all that will be said will remain thoroughly understandable. We refer to Appendix **??** for the statement of the Lebesgue dominated convergence theorem.
In the whole chapter, we consider a "weight function" $k(\mathbf{y})$ defined on $I\!R^N$ with values in $[0, +\infty[$ and satisfying

$$\int_{I\!R^N} k(\mathbf{y}) d\mathbf{y} = 1.$$

We call k-measure of a subset B of $I\!R^N$ and denote by $|B|_k$ the integral

$$|B|_k = \int_B k(\mathbf{y}) d\mathbf{y}.$$

Of course, $0 \leq |B|_k \leq 1$. We shall also write, when there is no risk of ambiguity, $|B|$ instead of $|B|_k$. As a main example for k, let us mention $k(\mathbf{x}) = c_N^{-1} \mathcal{X}_{B(0,r)}$, where c_N is the Lebesgue measure of the ball $B(0,r)$ and $\mathcal{X}_{B(0,r)}$ the characteristic function of the ball $B(0,r)$. The median value (weighted by k) of a function u is the result of an attempt to define a smoothed version of u which does not depend upon contrast changes. This is done by "averaging" the level sets of u, so that our final definition will be better understood if we start by defining a median operator on sets.

Definition 10.1 *Let X be a measurable subset of $I\!R^N$ and k a weight function. We call median set of X (weighted by k) and denote by* $\mathrm{med}_k X$ *(or* $\mathrm{med} X$*, since there is no risk of ambiguity) the set*

$$\mathrm{med} X = \{\mathbf{x}, |X - \mathbf{x}|_k \geq \frac{1}{2}\} \tag{10.1}$$

Example 10.2 *As a first example, let $k(\mathbf{x}) = \frac{1}{\pi r^2} \mathcal{X}_{B(0,r)}(\mathbf{x})$ in $I\!R^2$. Then $\mathbf{x}$ belongs to* $\mathrm{med}_k X$ *if and only if the Lebesgue measure of $X \cap B(0,r)$ is larger or equal to the half measure of $B(0,r)$. In intuitive terms, $\mathbf{x}$ belongs to* $\mathrm{med}_k X$ *if points of X are in majority around $\mathbf{x}$.*

Example 10.3 *The Koenderink-Van Dorn " dynamic shape" (Section 3.4). Take* $k(\mathbf{x}) = G_t(\mathbf{x}) = \frac{1}{4\pi t} e^{\frac{-|\mathbf{X}|^2}{4t}}$. *Then, since* $\int_{\mathbb{R}^2} G_t(\mathbf{x})d\mathbf{x} = 1$, *we immediately see that the dynamic shape is nothing but a gaussian weighted median filter.*

Let us now extend the set median operator med into a contrast invariant operator acting on measurable functions.

Lemma 10.4 *The operator* $\mathrm{med}_k : \mathcal{T} \to \mathcal{T}$ *is a monotone operator satisfying the set continuity property (7.14), that is : If* $(X_\lambda)_{\lambda\in\mathbb{R}}$ *is a non-increasing family of measurable sets such that* $\forall\lambda \in \mathbb{R}$, $X_\lambda = \cap_{\mu<\lambda} X_\mu$, *then*

$$\mathrm{med}_k(X_\lambda) = \cap_{\mu<\lambda}\mathrm{med}_k(X_\mu). \tag{10.2}$$

Proof Since med_k is monotone, we already have $\mathrm{med}_k(X_\lambda) \subset \cap_{\mu<\lambda}\mathrm{med}_k(X_\mu)$. Conversely, let $\mathbf{x} \in \cap_{\mu<\lambda}\mathrm{med}_k X_\mu$. Then, by (10.1), we have for every μ, $|X_\mu - \mathbf{x}|_k \geq \frac{1}{2}$. Since X_μ is a nondecreasing sequence of sets, all with finite measure, we have by Lebesgue theorem (Appendix **??**) $|X_\mu - \mathbf{x}|_k \to |X_\lambda - \mathbf{x}|_k$. Thus $|X_\lambda - \mathbf{x}|_k \geq \frac{1}{2}$ and, by (10.1) again, $\mathbf{x} \in \mathrm{med}_k(X_\lambda)$. □

Definition 10.5 and Proposition *The median set operator defined on measurable sets has a unique contrast invariant extension to measurable functions, obtained by the threshold superposition principle. This extension has the* sup inf *form*

$$\mathrm{med}_k(u)(\mathbf{x}) = \sup_{|B|_k\geq\frac{1}{2}} \inf_{\mathbf{y}\in B+\mathbf{x}} u(\mathbf{y}) \tag{10.3}$$

and satisfies

$$\mathcal{X}_\lambda(\mathrm{med}_k(u)) = \mathrm{med}_k(\mathcal{X}_\lambda(u)).$$

Proof By Lemma 10.4 and Theorem 7.11, the median operator med_k, defined on measurable sets, has a unique contrast invariant extension to measurable functions such that

$$\mathrm{med}_k(\mathcal{X}_\lambda u) = \mathcal{X}_\lambda(\mathrm{med}_k u)$$

for every measurable function u. In addition, by Corollary 8.6, the median operator thus defined on functions has the "sup inf" form

$$(\mathrm{med}_k u)(\mathbf{x}) = \sup_{B\in\mathcal{B}} \inf_{\mathbf{y}\in\mathbf{x}+B} u(\mathbf{y}),$$

where $\mathcal{B} = \{B, 0 \in \mathrm{med}_k(B).\}$ This immediately yields (10.3).

□

Example 10.6 *In order to understand the consequences of the definition of the median operator in case of functions which have "flat" parts and jumps, it is useful to consider the following examples of functions defined on* $\mathbb{R}$ *with* $k = \frac{1}{2}\mathbb{1}_{[-1,1]}$ *:*

• $u(x) = x$ *if* $x \leq 0$, $u(x) = 0$ *if* $x > 0$: $\mathrm{med}_k u(0) = 0$. *In that case, the median value coincides with the supremum.*

• $u(x) = -1$ *if* $x \leq 0$, $u(x) = 1$ *if* $x > 0$ *:* $\mathrm{med}_k u(0) = 1$. *Notice the symmetry breaking : +1 and -1 are equally good candidates to define a median value, but our definition of the median operator gives a privilege to the larger value.*
• $u(x) = 0$ *if* $x \leq -\frac{1}{4}$, $u(x) = 1$ *if* $-\frac{1}{4} < x < \frac{1}{3}$, $u(x) = 2$ *if* $x \geq \frac{1}{3}$ *:* $\mathrm{med}_k u(0) = 1$.
Notice that the median value is, in all of the three cases, different from the mean value on the same neighborhood.

Exercise 10.1 *Prove that if* u *attains a finite number of values, then* $\mathrm{med}(u)$ *takes its values in the range of* u.

Remark 10.7 Upper and lower median operators *The median operator, as we have defined it, is not invariant by reverse contrast : we do not have* $\mathrm{med}(-u) = -\mathrm{med}(u)$ *as is clear from the second example above. Thus, we can consider an alternative definition of for the median operator, as*

$$\mathrm{med}_k^-(u) = \inf_{|B|_k \geq \frac{1}{2}} \sup_{B+\mathbf{x}} u(\mathbf{y}) \tag{10.4}$$

Obviously,

$$\mathrm{med}_k^- u = -\mathrm{med}_k(-u),$$

so that both definitions yield the same formalism and properties. In particular, we have

$$\mathrm{med}_k^- X = \{\mathbf{x}, |X - \mathbf{x}|_k > \frac{1}{2}\}, \tag{10.5}$$

which is easily deduced from the remark that

$$\mathrm{med}_k^- X = (\mathrm{med}_k X^c)^c,$$

where X^c *is the complementary set of* X, $\mathbb{R}^N \setminus X$. *The choice of* med_k *is more adapted to closed sets and u.s.c. functions, since, as we shall now see,* $\mathrm{med}_k X$ *is closed if* X *is closed and* $\mathrm{med}_k u$ *is u.s.c. if* u *is. In the same way,* med_k^- *is adapted to open sets and l.s.c. functions.*

Theorem 10.8 *Let* u *be a measurable function. Then* $\mathrm{med}_k u$ *is an upper semi-continuous function (and* $\mathrm{med}_k^- u$ *a lower semi-continuous function).*

Lemma 10.9 *For every subset* X *of* $\mathbb{R}^N$, $\mathrm{med}_k X$ *is a closed set (and* $\mathrm{med}_k^- X$ *is an open set.)*

Proof If $\mathbf{x}_n \in \mathrm{med}_k X$, then $|X - \mathbf{x}_n|_k \geq \frac{1}{2}$, that is $\int_X k(\mathbf{y} - \mathbf{x}_n) d\mathbf{y} \geq \frac{1}{2}$. If $\mathbf{x}_n \to \mathbf{x}$, the last integral converges to $\int_X k(\mathbf{y} - \mathbf{x}) d\mathbf{y}$ by Proposition 2.3. We obtain $|X - \mathbf{x}|_k \geq \frac{1}{2}$, that is, $\mathbf{x} \in \mathrm{med}_k(X)$. □

Proof of Theorem 10.8 Let u be a measurable function on $\mathbb{R}^N$. By Definition-Proposition 10.5, the level sets of med$u$ are obtained by applying med to the level sets of u. By Lemma 10.9, the level sets of $\mathrm{med}(u)$ are therefore closed, so that $\mathrm{med}(u)$ is upper semi-continuous. One also deduces that $\mathrm{med}^-(u)$ is a lower semi-continuous function from the relation $\mathrm{med}^-(u) = -\mathrm{med}(-u)$. □

We shall now give a very general sufficient condition ensuring first that $\mathrm{med}_k u$ is continuous if u is and second that both possible definitions of the median operator, med and med^-, coincide on continuous functions.

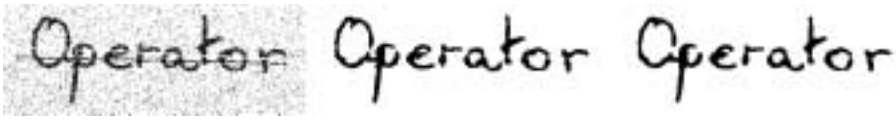

Figure 10.1: Example of denoising with the median filter : scanned picture of the word "operator" with perturbations and noise made of black or white lines and dots. Middle : first iteration of the median filter with a circular neighborhood of radius 2, right second iteration. Compare with the denoising by opening-closing experiment (Figure 9.9).

Proposition 10.10 *(i) For every measurable function u,*

$$\mathrm{med}_k(u) \geq \mathrm{med}_k^-(u) \tag{10.6}$$

(ii) Assume that k satisfies the following property :

$$\forall B, B' \text{ such that } meas_k(B) \geq \frac{1}{2} \text{ and } meas_k(B') \geq \frac{1}{2}, \text{ then } \overline{B} \cap \overline{B'} \neq \emptyset,$$

where $\overline{B}$ denotes the topological closing of B. Then, for every continuous function u, $med_k(u)$ is continuous and

$$\mathrm{med}_k(u) = \mathrm{med}_k^-(u) \tag{10.7}$$

Exercise 10.2 *Condition (ii) in Proposition 10.10 is very weak and grossly corresponds to the assumption that the support of k cannot be split into two connected components with measure $\frac{1}{2}$. Check that if (e.g.) k is continuous and its support connected, then Condition (ii) holds.*

Proof (i) Set $\lambda = \mathrm{med}_k(u)(\mathbf{x}) = \sup_{|B|_k \geq \frac{1}{2}} \inf_{\mathbf{y} \in B+\mathbf{x}} u(\mathbf{y})$. For every $\epsilon > 0$, we then have $meas_k(\mathcal{X}_{\lambda+\epsilon} u) < \frac{1}{2}$. Thus $meas_k((\mathcal{X}_{\lambda+\epsilon} u)^c) \geq \frac{1}{2}$ and therefore $\sup_{(\mathcal{X}_{\lambda+\epsilon} u)^c} u \leq \lambda + \epsilon$. Hence

$$\forall \epsilon > 0, \inf_{|B|_k \geq \frac{1}{2}} \sup_{B+\mathbf{x}} u(\mathbf{y}) \leq \lambda + \epsilon$$

Letting ϵ tends to zero yields (i).

(ii) The assumption on k implies that for all B and B' with k-measure larger or equal than $1/2$ we have

$$\inf_{\overline{B}+\mathbf{x}} (u) \leq \sup_{\overline{B'}+\mathbf{x}} (u)$$

and since u is continuous,

$$\inf_{B+\mathbf{x}} (u) \leq \sup_{B'+\mathbf{x}} (u).$$

Taking the inf over all B' and the sup over all B yields

$$\forall \mathbf{x}, \sup_{|B|_k \geq \frac{1}{2}} \inf_{B+\mathbf{x}} u(\mathbf{y}) \leq \inf_{|B'|_k \geq \frac{1}{2}} \sup_{B'+\mathbf{x}} u(\mathbf{y})$$

¿From this last relation and (i), we deduce the equality of the med and med$^-$ operators on continuous functions. Now, by Theorem 10.8, the operator *med* transforms any measurable function into an upper semi-continuous one, and *med*$^-$ into a lower semi-continuous one. Thus *med*(u) is continuous if u is. □

Exercise 10.3 *Show that*

$$\sup_{|B|_k \geq \frac{1}{2}} \inf_{B+\mathbf{X}} u(\mathbf{y}) \geq \sup_{|B|_k > \frac{1}{2}} \inf_{B+\mathbf{X}} u(\mathbf{y}) \geq \inf_{|B|_k \geq \frac{1}{2}} \sup_{B+\mathbf{X}} u(\mathbf{y}),$$

$$\sup_{|B|_k \geq \frac{1}{2}} \inf_{B+\mathbf{X}} u(\mathbf{y}) \geq \inf_{|B|_k > \frac{1}{2}} \sup_{B+\mathbf{X}} u(\mathbf{y}) \geq \inf_{|B|_k \geq \frac{1}{2}} \sup_{B+\mathbf{X}} u(\mathbf{y}).$$

Remark 10.11 Discrete median filters and the "usual" median value.

We define a discrete median filter by considering a measure made of a finite number of Dirac masses δ_i, $i \in \{1, ..N\}$, centered at points $\mathbf{x}_i$. The locations of the Dirac masses represent the discrete neighborhood on which the median value will be computed around each point. (Notice that such a measure is not associated with an integrable function k, so that Theorem 10.8 does not hold anymore.)
The formula of the median filter can be simplified in that case. We have

$$med\, u(\mathbf{x}) = \inf_{P \in I\!P(N),\, card(P) \geq N/2} \sup_{i \in P} u(\mathbf{x} - \mathbf{x}_i),$$

$$med^- u(\mathbf{x}) = \sup_{P \in I\!P(N),\, card(P) \geq N/2} \inf_{i \in P} u(\mathbf{x} - \mathbf{x}_i)$$

where $I\!P(N)$ denotes the set of all parts of {1, ..., N}. Let us now denote by $M = E(N/2)+1$, the smallest integer larger or equal than $N/2$. For any set P containing strictly more than M elements, we can construct a smaller set P', with still a number of elements larger than M, by simply removing an element from P. The sup *on P' is therefore smaller than the* sup *on P, and since the* med *operator chooses the smallest value over $I\!P(N)$, we can remove such P's from $I\!P(N)$ without changing the value of* med. *We therefore have, with $M = E(N/2) + 1$,*

$$\text{med}\, u(\mathbf{x}) = \inf_{P \in I\!P(N),\, card(P) = M} \sup_{i \in P} u(\mathbf{x} - \mathbf{x}_i)$$

$$\text{med}^- u(\mathbf{x}) = \sup_{P \in I\!P(N),\, card(P) = M} \inf_{i \in P} u(\mathbf{x} - \mathbf{x}_i)$$

Is it now easy to see that $med u(\mathbf{x})$ (resp. $med^- u(\mathbf{x})$) corresponds to the M **smallest** *(resp. M largest) value out of the real numbers $u(\mathbf{x} - \mathbf{x}_1), ...u(\mathbf{x} - \mathbf{x}_N)$ which are the values of u in the discrete neighborhood of $\mathbf{x}$.*

The operators med *and* med^- *are identical if and only if N is odd. Indeed, in that case, $M = (N+1)/2$ and the M smallest value out of N real numbers is also the M largest value. This is compatible with Proposition 10.10 : indeed, if N is odd, then any two sets of cardinality $M = (N+1)/2$ made out of N points must intersect. (Take e.g. N=7, therefore M=4. Two sets of 4 pixels out of a set of 7 pixels have necessarily at least one pixel in common.) This is not true anymore if N is even, and therefore in that case* med *and* med^- *can differ. This is why in general a median value on an odd number of pixels is preferred.*

Now, the discrete median filter can be also defined by putting a weight on each Dirac mass, which is equivalent to say that neighboring pixels will weigh differently in the computation of the median value. A simple example is given in Figure 10.2, where the weight is set according to the area of the intersection between each pixel and a 2×2 mask.

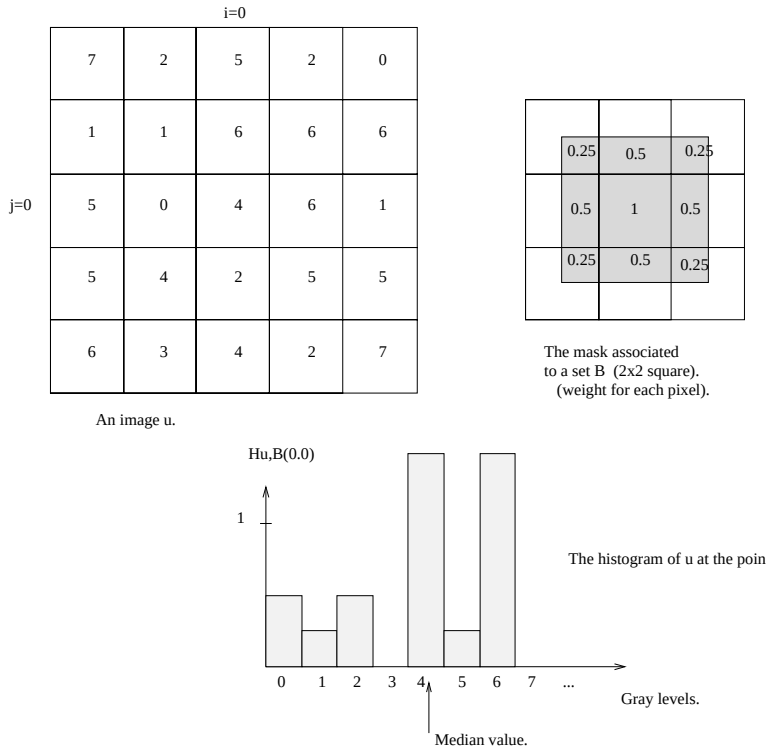

Figure 10.2: Discretization of a median filter. The digital image (above,left) is considered as constant on each square pixel. The support of the illustrated median filter is a square centered at zero. The subfigure above, left, shows the intersections of the nine neighboring pixels with the mask. The values correspond to the areas of those intersections. These values are used as weights to the pixel values around zero. They permit to build the weighted histogram of the neighborhood of the central pixel. In this histogram (below), the values 0 and 2 appear once with weight 0.5, the value 1 appears once with weight 0.25, the value 2 does not appear, the value 4 appears once with weight 1 and once with weight 0.25, resulting in a histogram value of 1.25, etc. The obtained median value is 4.

Exercise 10.4 *By* $\operatorname{arginf}_m f(m)$*, we denote a value of* m *(if any) at which* f *attains its infimum. Consider* N *real numbers* x_i*.*
(i) Show that their mean value satisfies

$$\frac{1}{N}\sum_{i=1}^{N} x_i = \operatorname{arginf}_m \sum_{i=1}^{N} (x_i - m)^2.$$

Show that

$$\operatorname{med}^-(x_i) \leq \operatorname{arginf}_m \sum_{i=1}^{N} |x_i - m| \leq \operatorname{med}(x_i).$$

(ii) Set $k = \mathbb{1}_B$*, where* B *is a set with Lebesgue measure equal to 1. Denote by* $\operatorname{med}_B u$ *the "median value of* u *in* B*" defined by* $\operatorname{med}_B u = \operatorname{med}_k u(0)$*. Consider a bounded and measurable function* u *on* B*. Show that*

$$\int_B u = \operatorname{arginf}_m \int_{\mathbf{x}\in B} (m - u(\mathbf{x}))^2 d\mathbf{x}$$

and

$$\operatorname{med}_B^-(u) \leq \operatorname{arginf}_m \int_{\mathbf{x}\in B} |m - u(\mathbf{x})| d\mathbf{x} = \frac{\operatorname{med}_B^- u + \operatorname{med}_B u}{2} \leq \operatorname{med}_B(u).$$

(iii) Conclude that the mean (resp. median) value is the "best" constant approximation with respect to the L^2 *(resp.* L^1*) norm.*

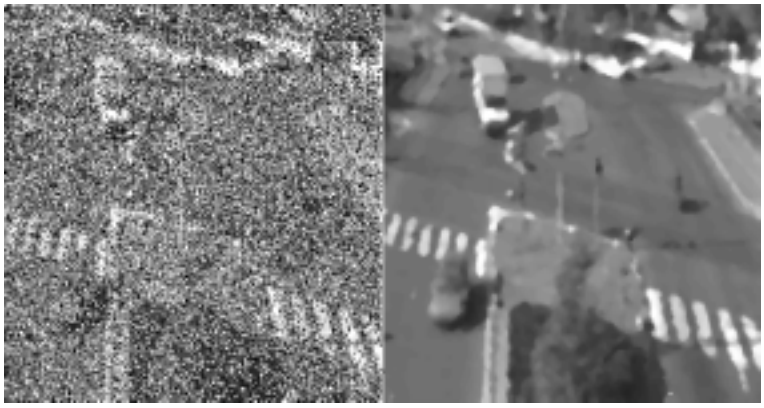

Figure 10.3: Denoising properties of the median filter. Left : image altered by a 40% salt and pepper noise. Right : Three iterations of the median filter on a 3×3 square mask.

Remark 10.12 *The Discrete Median Filter* M *is a cyclic operator on discrete images. Consider an image* u *having* N *pixels. The number of differents gray levels can be obviously bounded by the number of pixels that is again* N*. Since the Median Filter is contrast invariant, it preserves the discrete set of gray levels. That is to say that* $M^k(u)$ *is an image with* N *pixels and with level among the different levels of* u *(*N *at max). Since the maximal number of different realizations is* N^N *then after* N^N *iteration of the median* M*, there will be two identical images among the images* $M^k(u)$*,* $k \in \{1, ..., N^N\}$*. As consequence the Median is cyclic.*

Simple example to illustrate the cyclicity is as follow. Consider a pixel set made of one pixel and its four neighbor and the associated median filter taking the median value among the five pixels. Consider also

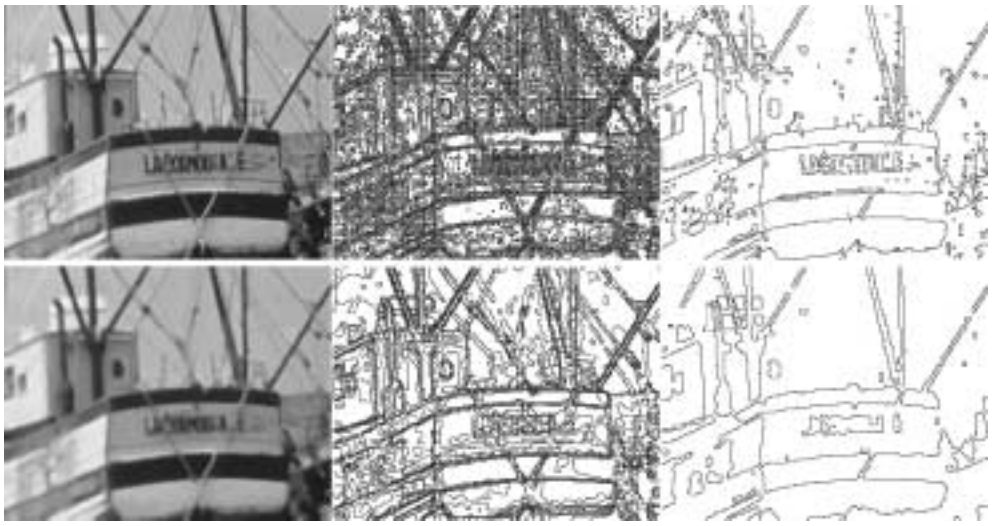

Figure 10.4: Smoothing effect of the median filter on level lines. Above : original image, then all of its level lines (boundaries of level sets) with levels multiple of 12, then level lines of level 100. Below : result of two iterations of a median filter on a disk with radius 2, then corresponding level lines (levels multiple of 12), then level lines with level 100.

the "chessboard" image, that is $u(i,j) = 255$ *if* $i + j$ *is even, and* $u(i,j) = 0$ *otherwise. When we apply the median filter the chessboard is "reversed". Indeed, any white pixel (value=255) is surrounded by 4 black pixels, so that the median filter transforms it into a black pixel. Conversely, a black pixel is transformed into a white one.*

References.

The remarkable denoising properties and numerical efficiency of median filters are well-known and acclaimed for the removal of all kinds of impulse noise in digital images, movies and video signals [315, 224, 339, 117, 349]. The last mentionned paper proposes an efficient, low complexity implementation, as well as [115, 212, 39]. An introduction to the weighted median filter can be found in [67, 444] and some generalizations (conditional median filters, etc.) in [260, 396, 34] and Unser-Yaroslavski [?]. Max, Min and median filters are just particular instances of rank order filters. See [112] for a general presentation of these filters. There are few studies on iterated median filters. The use of iterated median fiters as a scale space is, however, proposed in [48]. The extension of median filtering to multichannel (colour) images is problematic. Let some interesting attempts be mentionned [350], [91].

Part III

LOCAL ASYMPTOTIC ANALYSIS OF OPERATORS

Chapter 11

Asymptotic behavior of contrast invariant isotropic operators (Dimension 2).

In this chapter, we consider the "inf-sup" operators introduced in Section 8.2,

$$Tu = \inf_{B \in I\!B} \sup_{\mathbf{y} \in \mathbf{X}+B} u(\mathbf{y}). \tag{11.1}$$

Since $T'(u) = -T(-u)$, every statement about T will be easily adapted to "sup-inf" operators

$$T'u(\mathbf{x}) = \sup_{B \in I\!B} \inf_{\mathbf{y} \in \mathbf{X}+B} u(\mathbf{y}) \tag{11.2}$$

By Theorems 8.4 and Theorem 8.16, we know that the monotone contrast invariant and translation invariant operators simply are operators in the "inf sup" (or sup inf) form. Thus, our asymptotic analysis of contrast invariant operators will be fully general if we adopt the "inf sup" formalism. In order to simplify the analysis, we shall, in addition, assume that the considered operators are isotropic, that is, $I\!B$ is invariant under all linear isometries R of $I\!R^2$: $RI\!B = I\!B$. In this chapter, we also assume that $I\!B$ is bounded, that is, $B \subset D(0, M)$ for every B in $I\!B$ and some disk $D(0, M)$. The action of such operators seems *a priori* to open a wide range of possibilities, as many as possible sets of structuring elements. We shall see that this freedom of choice is rather an illusion. Let us introduce a scale parameter $h > 0$. We associate with $I\!B$ and T the scaled family $I\!B_h = hI\!B$ and the scaled operator T_h defined by

$$T_h u(\mathbf{x}) = \sup_{B \in hI\!B} \inf_{\mathbf{y} \in \mathbf{X}+B} u(\mathbf{y}).$$

In fact, we shall prove that when h tends to zero, the action of T_h on smooth functions u is described by a fairly small set of possibilities.

We shall prove a theorem which plays the same role for isotropic contrast invariant filters as Theorem 3.2 played for linear filters. As a main example, we shall prove that if T_h is a shrinked median operator, then

$$T_h u - u = Ch^2 curv(u)|Du| + o(h^2).$$

Thus, the operator $|Du|curv(u) = D^2u(\frac{Du^\perp}{|Du|}, \frac{Du^\perp}{|Du|})$ plays the same role for contrast invariant operators as the laplacian Δu for linear ones.

11.1 Asymptotic behavior theorem.

In the following, we set

$$H(\alpha) = T[x + \alpha y^2](0), \tag{11.3}$$

where $T[x + \alpha y^2]$ stands for "$T(u)$ with $u(x, y) = x + \alpha y^2$". Since T is monotone, H is a nondecreasing function, which is, in addition, Lipschitz. Indeed, by the monotonicity of T and the boundedness of $I\!B$,

$$T[x + \alpha_1 y^2](0) - |\alpha_2 - \alpha_1| M^2 \leq H(\alpha_2) \leq T[x + \alpha_1 y^2](0) + |\alpha_2 - \alpha_1| M^2$$

and therefore

$$|H(\alpha_2) - H(\alpha_1)| \leq M^2 |\alpha_2 - \alpha_1|, \tag{11.4}$$

where we have used the fact that all elements of $I\!B$ are contained in a disk $D(0, M)$.

Let us also note the following relations :

$$T_h[x] = hT[x] = hH(0) \tag{11.5}$$

$$T_h[x + \alpha y^2](0) = hT[x + h\alpha y^2](0) = hH(\alpha h) \tag{11.6}$$

Theorem 11.1 *Let $I\!B$ be a family of structuring elements in $I\!R^2$ which is bounded ($\forall B \in I\!B$, $B \subset D(0, M)$) and isotropic (if $B \in I\!B$, then $RB \in I\!B$ for every linear isometry R of $I\!R^2$). Let $Tu(\mathbf{x}) = \inf_{B \in I\!B} \sup_{\mathbf{y} \in \mathbf{x}+B} u(\mathbf{y})$ (or $Tu(\mathbf{x}) = \sup_{B \in I\!B} \inf_{\mathbf{y} \in \mathbf{x}+B} u(\mathbf{y})$) and define the rescaled operator*

$$T_h u(\mathbf{x}) = \inf_{B \in hI\!B} \sup_{\mathbf{y} \in \mathbf{x}+B} u(\mathbf{y}) \quad (\textit{ resp. } T_h u(\mathbf{x}) = \sup_{B \in hI\!B} \inf_{\mathbf{y} \in \mathbf{x}+B} u(\mathbf{y}) \textit{ }).$$

Then for every C^2 function u on $I\!R^2$,

$$(T_h u)(\mathbf{x}) = u(\mathbf{x}) + hT[x](0)|Du|(\mathbf{x}) + O(h^2).$$

Proof. Since $T(u - u(\mathbf{x})) = Tu - u(\mathbf{x})$, we can impose without loss of generality $u(\mathbf{x}) = 0$. T being translation and isometry invariant, we can choose the origin 0 at $\mathbf{x}$, and orthogonal axes $(\vec{i}, \vec{j})$ satisfying $\vec{i} = \frac{Du}{|Du|}(0)$ and $\vec{j} = (Du^{\perp}/|Du|)(0)$ when $Du \neq 0$. Since u is C^2, we can by Taylor formula write for $\mathbf{y} = (x, y)$ in a neighborhood of 0,

$$u(\mathbf{y}) = px + O(|\mathbf{y}|^2), \tag{11.7}$$

where $p = |Du|(0) \geq 0$. ¿From (11.7) we deduce that for every $\mathbf{y} = (x, y)$ in $hD(0, M)$, we have

$$px - O(h^2) \leq u(\mathbf{y}) \leq px + O(h^2).$$

Let us apply T_h to both members of this relation. By the monotonicity of T_h, and since every hB is contained in $D(0, Mh)$, we deduce that

$$T_h u(0) = T_h[px](0) + O(h^2). \tag{11.8}$$

Using (11.5), we have $T_h u(0) = hT[px](0) + O(h^2) = ph(T[x](0)) + O(h^2)$, which proves the theorem, since $p = |Du(0)|$. □

Let us state an informal but meaningful conclusion yielded by the preceding theorem. Let T be a contrast

invariant and monotone operator. The first "test" to which T must be submitted is simply to compute its action on $u(\mathbf{x}) = u(x,y) = x$. If $T[x](0) \neq 0$, then when the scale h of a scaled operator T_h tends to zero, T_h behaves like a dilation if $T[x](0) > 0$ and like an erosion otherwise (See Proposition 9.7). Thus, if $T[x](0) \neq 0$, there is no use to define it with complicated sets of structuring elements : they all yield asymptotically an erosion or a dilation which can be performed with a single and simple structuring element : the disk.

Exercise 11.1 *An analogous statement when T is no more isotropic : show that under the same assumptions as in Theorem11.1, except the isotropy of T, one has*

$$(T_h u)(0) - u(0) = hT[Du(0).\mathbf{x}](0) + O(h^2).$$

We now consider the case where $T[x](0) = 0$. In such a case, the operator T has a desirable symmetry : If we understand $Tu(\mathbf{x})$ as a kind of average value of u in a neighborhood of $\mathbf{x}$, the assumption $T[x](0) = 0$ is of very sound. It means that all isotropic averages of $u(x,y) = x$ around $(0,0)$ should be 0.

Theorem 11.2 *Let $\mathbb{B}$ be a family of structuring elements in $\mathbb{R}^2$ which is bounded ($\forall B \in \mathbb{B}$, $B \subset D(0,M)$) and isotropic (if $B \in \mathbb{B}$, then $RB \in \mathbb{B}$ for every linear isometry R of $\mathbb{R}^2$).*

Let $Tu(\mathbf{x}) = \inf_{B\in\mathbb{B}} \sup_{\mathbf{y}\in\mathbf{x}+B} u(\mathbf{y})$ (or $Tu(\mathbf{x}) = \sup_{B\in\mathbb{B}} \inf_{\mathbf{y}\in\mathbf{x}+B} u(\mathbf{y})$) and define the rescaled operator

$$T_h u(\mathbf{x}) = \inf_{B\in h\mathbb{B}} \sup_{\mathbf{y}\in\mathbf{x}+B} u(\mathbf{y}) \quad (\text{ resp. } T_h u(\mathbf{x}) = \sup_{B\in h\mathbb{B}} \inf_{\mathbf{y}\in\mathbf{x}+B} u(\mathbf{y}) \text{ }).$$

Set $H(h) = T(x+hy^2)(0)$ and assume that $H(0) = T[x](0) = 0$. Then for every C^3 function u on $\mathbb{R}^2$:

(i) On every compact set K contained in $\{\mathbf{x}, Du(\mathbf{x}) \neq 0\}$,

$$T_h u(\mathbf{x}) = u(\mathbf{x}) + h|Du(\mathbf{x})|H(\frac{1}{2}h \operatorname{curv}(u)) + O_{\mathbf{x}}(h^3),$$

where $|O_{\mathbf{x}}(h^3)| \leq C_K h^3$ for some constant C_K depending only on u and K.

(ii) On every compact set K of $\mathbb{R}^2$,

$$|T_h u(\mathbf{x}) - u(\mathbf{x})| \leq M^2 h^2 ||D^2 u(\mathbf{x})|| + O_{\mathbf{x}}(h^3),$$

where $|O_{\mathbf{x}}(h^3)| \leq C_K h^3$ for some constant C_K depending only on u and K.

In (ii), we notice that $D^2u(\mathbf{x})$ is a 2×2 matrix and we take $||A|| = |a| + |b| + |c| + |d|$ as a norm for such a matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. As a canonical application of the preceding theorem which we will develop later in this chapter, the reader may think of the median filter. Indeed, the median filter does not alter linear functions and we therefore have $H(h) = \operatorname{med}(px)(0) = 0$.

Proof. We adopt the same notation as in the proof of Theorem 11.1, take again and without loss of generality $\mathbf{x} = 0$, $u(0) = 0$ and set again, for (x,y) in a neighborhood of 0,

$$u(x,y) = px + ax^2 + by^2 + cxy + O(|x^2+y^2|^{\frac{3}{2}}). \tag{11.9}$$

We refer to Section 5.5 for a detailed account of the notation and the differential interpretation of p, a, b, c. Note that by Taylor formula,

$$O(|x^2+y^2|^{\frac{3}{2}}) \leq ||D^3u(0)|||x^2+y^2|^{\frac{3}{2}}. \tag{11.10}$$

The same reasoning as in Theorem 11.1 and an obvious rescaling yield

$$T_h u(0) = hT[px + ahx^2 + bhy^2 + chxy](0) + O(h^3) = hT[u_h(x,y)](0) + O(h^3), \tag{11.11}$$

where we set $u_h(x,y) = px + ahx^2 + bhy^2 + chxy$. In order to prove the theorem, we have to show that c and a play no role in the asymptotic behavior of $Tu_h(0) = \inf_{B\in I\!B} \sup_B u_h$.

Step 1 : An estimate. If $(x,y) \in B \in I\!B$, then

$$px - h(|a|+|b|+|c|)M^2 \leq u_h(x,y) \leq px + h(|a|+|b|+|c|)M^2.$$

By assumption, $T(px) = 0$, and we obtain, by applying T_h to the inequalities,

$$|Tu_h(0)| \leq h(|a|+|b|+|c|)M^2. \tag{11.12}$$

Returning to the definitions of a, b, c, using Relations (11.9) to (11.11) and remarking that $D^3u(\mathbf{x})$ is continuous and therefore bounded on each compact set K, we deduce the assertion (ii) from (11.12).
¿From now on, we focus on (i) and assume that $p \neq 0$. We set

$$C = (|a|+|b|+|c|)M^2.$$

Step 2 : First reduction. By Step 1, for every $B \in I\!B$, we have

$$\sup_B u_h \geq \inf_{B\in I\!B} \sup_B u_h = Tu_h(0) \geq -Ch.$$

Now, if $(x,y) \in B$ and $x \leq -2Ch/p$, then

$$u_h(x,y) = px + ahx^2 + bhy^2 + chxy \leq -2Ch + h(|a|+|b|+|c|)M^2 = -Ch.$$

Let us set $C' = \frac{2C}{p}$. We obtain

$$\forall B \in I\!B \ \sup_B u_h = \sup_{B\cap\{(x,y), x\geq -C'h\}} u_h.$$

Step 3 : Second reduction. Since $Tu_h(0) \leq Ch$ (Step 1), one does not need to consider the sets B for which $\sup_B u_h \geq Ch$. If $\sup_B u_h \leq Ch$, then

$$\forall (x,y) \in B, \ \ px + ahx^2 + bhy^2 + chxy \leq Ch$$

and therefore

$$x \leq \frac{1}{p}(Ch + (|a|+|b|+|c|)M^2h) \leq \frac{2Ch}{p} = C'h.$$

Hence,

$$Tu_h(0) = \inf_{B\in I\!B, B\subset\{(x,y),\ x\leq C'h\}} \sup_B u_h \tag{11.13}$$

and by Step 2,

$$Tu_h(0) = \inf_{B\in I\!B, B\subset\{(x,y),\ x\leq C'h\}} \sup_{B\cap\{(x,y),\ x\geq -C'h\}} u_h(x,y). \tag{11.14}$$

Now, for those (x, y), belonging to $B \subset \{(x, y), x \leq C'h\}$ and satisfying $x \geq -C'h$, we have $|ahx^2 + chxy| \leq C''h^2$ and therefore $u_h(x, y) = px + bhy^2 + O(h^2)$. Thus (11.14) implies

$$T(u_h)(0) = T[px + bhy^2](0) + O(h^2). \tag{11.15}$$

Conclusion. Using the definition of H, $H(h) = T[x + hy^2](0)$, we obtain

$$T_h u(0) = hT(u_h)(0) = phH(\frac{bh}{p}) + O(h^3). \tag{11.16}$$

Since, by Formula (5.11),

$$b = \frac{1}{2}\text{curv}(u)|Du|(0)$$

and $p = |Du|(0)$, we obtain the relation announced in (i), at $\mathbf{x} = 0$. Now, all computations have been done with the origin fixed at $\mathbf{x} = 0$. If we let vary $\mathbf{x}$ in a compact set K on which $p = |Du(\mathbf{x})| \neq 0$ (so that $p = |Du(\mathbf{x})|$, a, b, c now depend on $\mathbf{x}$), then we have for some positive constants c_K and C_K only depending on K and u,

$$p \geq c_K, \;\; |a|, \; |b|, \; |c| \leq C_K.$$

Indeed, $\mathbf{x} \to D^2u(\mathbf{x})$ is continuous and therefore a, b, c are bounded on each compact set K. By compactness again, the lower bound c_K of p on K is attained and therefore positive. It then easy to check that the constants $C(\mathbf{x})$ and $C'(\mathbf{x})$ involved in the preceding proof also are bounded independently of $\mathbf{x}$ on K. Thus the $O(h^3)$ involved in (11.16) is uniform on K. The same argument applies to the statement (ii), which we have proven above at $\mathbf{x} = 0$. □

11.2 Median filters and curvature motion.

We recall that the median filter, med, defined in Chapter 10, can be written, as in Formula (10.3) :

$$\text{med}_k u(\mathbf{x}) = \sup_{|B|_k \geq \frac{1}{2}} \inf_{\mathbf{y} \in B + \mathbf{x}} u(\mathbf{y}). \tag{11.17}$$

In the following, we take as a first example for k the uniform measure on a disk $D(0, 1)$, so that

$$|D(0,1)|_k = 1, \;\; k_h = \frac{\mathbb{1}_{D(0,h)}}{\pi h^2}$$

and we shall examine only continuous functions. This entails by Proposition 10.10 that med and med^- are simply equal. So we shall simply write "med" and talk about "the" median filter. The "infsup" form of the median filter on continuous function is given by

$$\text{med}_{D(0,h)} u(\mathbf{x}) = \inf_{meas B \geq \frac{1}{2}, B \subset D(0,h)} \sup_{\mathbf{y} \in \mathbf{x} + B} u(\mathbf{y}).$$

The main result of this section gives an infinitesimal interpretation of the median filter. This theorem will be generalized in the next chapter to much more general weighted median filters (Theorem 12.1).

Theorem 11.3 *Let u be a C^3 function in $\mathbb{R}^2$. Then*

$$(i)\quad \mathrm{med}_{D(0,h)}u(\mathbf{x}) = u(\mathbf{x}) + \frac{1}{6}\mathrm{curv}(u)|Du|(\mathbf{x})h^2 + O_{\mathbf{x}}(h^3),$$

where $O_{\mathbf{x}}(h^3) \leq c_K h^3$ on every compact subset of $\{\mathbf{x}, Du(\mathbf{x}) \neq 0\}$ and

$$(ii)\quad |\mathrm{med}_{D(0,h)}u(\mathbf{x}) - u(\mathbf{x})| \leq ||D^2u(\mathbf{x})||h^2 + O_{\mathbf{x}}(h^3)$$

where $O_{\mathbf{x}}(h^3) \leq c_K h^3$ on every compact subset of $\mathbb{R}^2$.

Figure 11.1: The erosion does a smoothing independent of the curvature of the level lines. Left : image of a simple shape. Right : difference of this image with its eroded by a ball with radius 6. We see in black the points which have changed. The width of the difference is constant.

Figure 11.2: Median filter and the curvature of level lines. The median filter does a smoothing linked to the curvature of the level lines. Left : image of a simple shape. Right : difference of this image with itself after it has been smoothed by one iteration of the median filter. We see, in black, the points which have changed. The width of the difference is proportional to the curvature, according to Theorem 11.3.

Lemma 11.4

$$\mathrm{med}_{D(0,1)}[x + hy^2](0) = \frac{h}{3} + O(h^3)$$

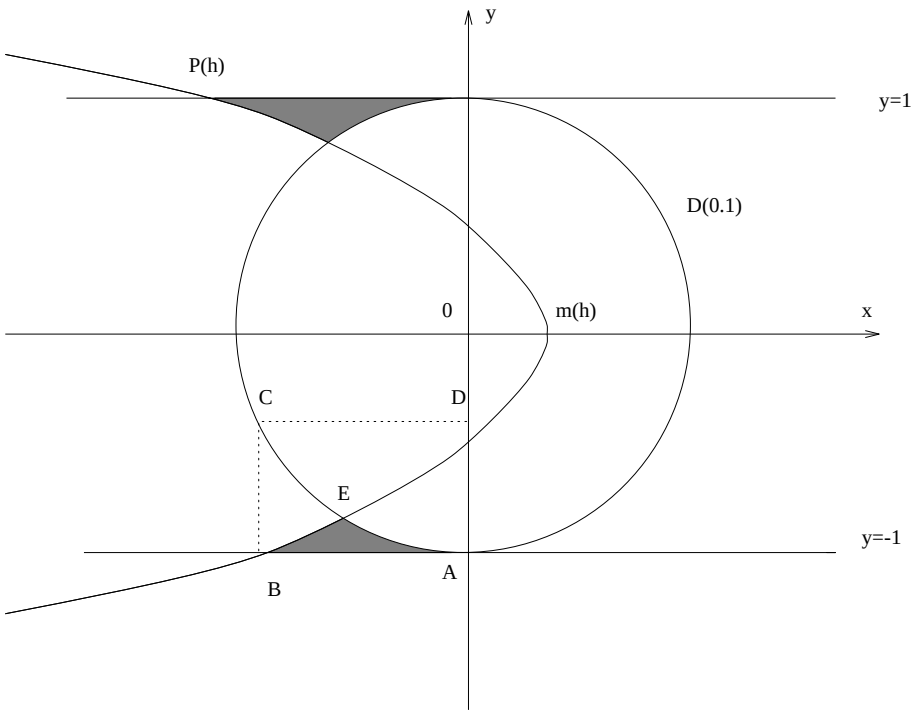

Figure 11.3: When h is small, the parabola $P(h)$ with equation $x+hy^2 = m$ divides $D(0,1)$ into two connected components. The median value $m(h)$ of $x + hy^2$ on $D(0,1)$ simply is the value m for which these two connected components have equal area.

.

Proof. When h is small, the parabola $P(h)$ with equation $x+hy^2 = m$ divides $D(0,1)$ into two connected components. By Definition 10.1, the median value $m(h)$ of $x + hy^2$ on $D(0,1)$ is defined as the value m for which these two connected components have equal area. The whole geometric situation is shown in Figure 11.3. The algebraic area between the Oy axis and the parabola is (for $|y| \leq 1$)

$$\int_{-1}^{1} (m(h) - hy^2)dy = 2m(h) - \frac{2h}{3}$$

Thus $m(h)$ is the median value if and only if

$$2m(h) - 2h/3 = 2area(ABE) \tag{11.18}$$

where ABE denotes the curved triangle bounded by the parabola, the circle and the line $y = 1$. This area could be explicitly computed, but we prefer to give it the simple bound : $\frac{\text{area}(ABCD)}{2}$. Now, the length of (AB) is $|m(h) - h|$ and the length of (BC) is less than $(m(h) - h)^2$, so that (11.18) implies

$$|2m(h) - \frac{2h}{3}| \leq |m(h) - h|^3.$$

We conclude that $m(h) = h/3 + O(h^3)$, as announced. □

Figure 11.4: Fixed point property of the median filter, proving its grid-dependence. On the left, original image. On the right, result of 46 iterations of the median filter with a radius of 2. The resulting image turns out to be a fixed point of the median filter with radius 2. This is not in agreement with Theorem 11.3 showing the median filter moving images by their curvature : the image on the right clearly has non zero curvatures ! Now, the discrete, pixelized median filter which we have applied here is grid-dependent and blind to small curvatures.

Proof of Theorem 11.3. The operator

$$Tu(\mathbf{x}) = \text{med}_{D(0,1)} u(\mathbf{x})$$

satisfies the conditions of Theorem 11.2. By Lemma 11.4, the function H associated with T satisfies $H(0) = 0$, so that the conclusions (i) and (ii) of Theorem 11.2 are satisfied. From Theorem 11.2 and Lemma 11.4 we obtain

$$H(h) = \frac{h}{3} + O(h^3)$$

Thus Relation (i) in Theorem 11.2 yields

$$\text{med}_{D(0,h)} u(\mathbf{x}) = u(\mathbf{x}) + h|Du(\mathbf{x})|(\frac{1}{6}h\text{curv}(u)) + O(h^3)$$

on compact subsets of $\{\mathbf{x}, |Du(\mathbf{x})| \neq 0\}$. □

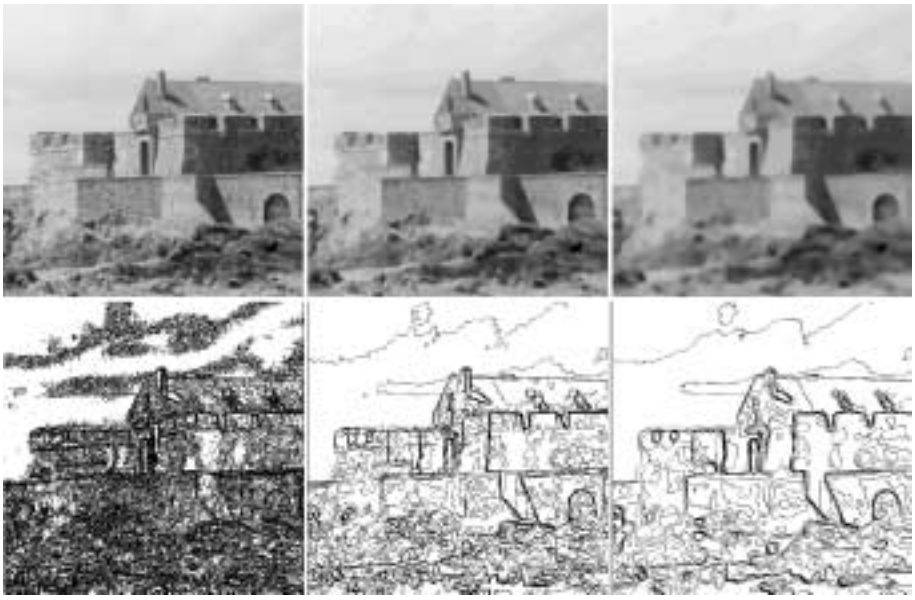

Figure 11.5: Comparison between iterated median filter and median filter : Top-middle : 16 iterations of the median filter with a radius 2, top-right : one iteration of the same median filter with a radius 8. Below : for each image, the level-lines for grey levels multiple of 16. To iterate a small size median filter results in more accuracy and less shape-mixing than to apply a large size median one. See the comparison between the Koenderink-Van Dorn shape smoothing and the Bence-Merriman-Osher iterated filter in Chapter 3, in particular Figures 3.10, 3.9, 3.12 .

Second application : the Catté-Dibos scheme

Theorem 11.5 *Let $\mathbb{B}$ be the set of all segments of the plane with length 2 and centered at zero. Set*

$$SI_h u(\mathbf{x}) = \sup_{B \mathbb{B}} \inf_{\mathbf{y} \in \mathbf{x}+hB} u(\mathbf{y})$$

and, similarly,

$$IS_h u(\mathbf{x}) = \inf_{B \mathbb{B}} \sup_{\mathbf{y} \in \mathbf{x}+hB} u(\mathbf{y}).$$

Let u be a C^3 function in $\mathbb{R}^2$. Then

$$\frac{1}{2}(IS_h + SI_h)u(\mathbf{x}_0) = u(\mathbf{x}_0) + h^2 \frac{1}{4} \mathrm{curv}(u)|Du|(\mathbf{x}_0) + O(h^3)$$

if $Du(\mathbf{x}_0) \neq 0$.

Proof. Let us compute the function $H_1(h) = IS_h[x + hy^2](0)$ associated with IS_h. Writing $(x, y) = (r\cos\theta, r\sin\theta)$ in polar coordinates, we have

$$H_1(h) = \inf_{-\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}} \sup_{-1 \leq r \leq 1} (rcos\theta + hr^2 sin^2\theta).$$

Since, for $h \geq 0$, the function $r \to rcos\theta + hr^2 sin^2\theta$ is increasing when $r \geq 0$, we obtain

$$H_1(h) = \inf_{-\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}} (cos\theta + hsin^2\theta) = h$$

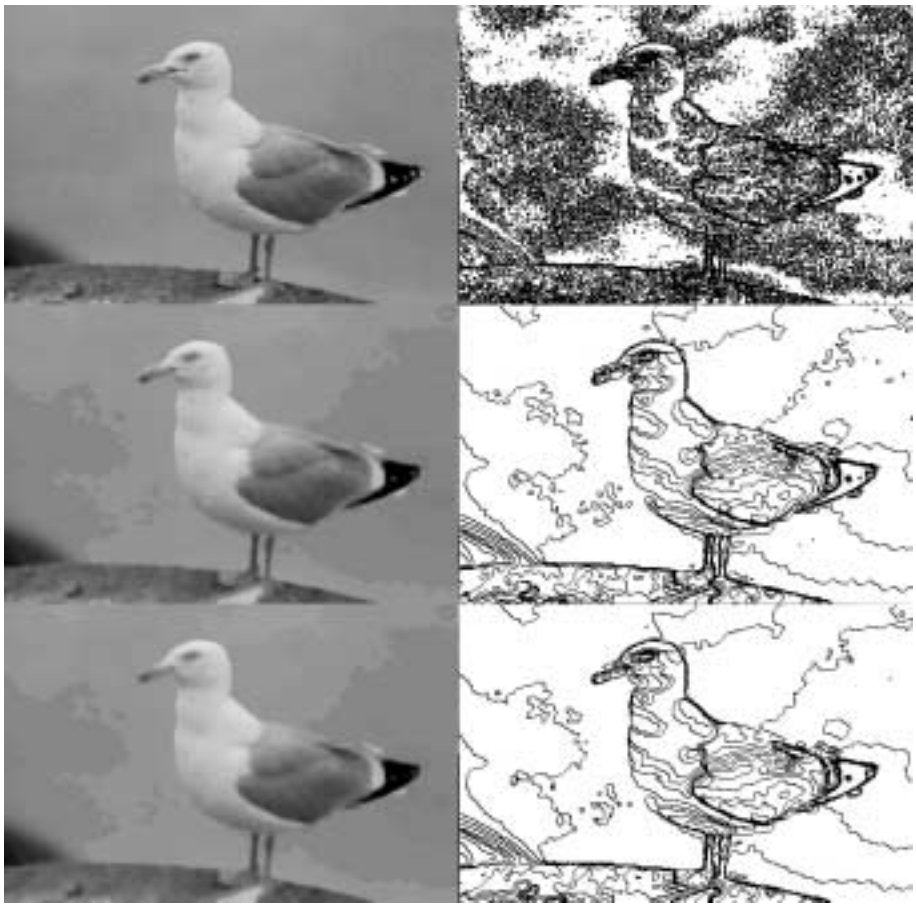

Figure 11.6: Consistency of the median filter and of the Catté-Dibos scheme with curvature motion. On the first row, the sea bird image and its level lines for all levels multiple of 12. On the second row, a median filter on a disk with radius 2 has been iterated twice. On the third row, an inf-sup and then a sup-inf filter based on segments have been applied. On the right : the corresponding level lines of the results, which, according to the theoretical results, must have moved at a normal speed proportional to their curvature, (Theorems 11.3 and **??**). The results are very close. This yields a cross-validation of two very different numerical schemes implementing the curvature motion.

if $h \geq 0$ is small enough (e.g. $h < \frac{1}{2}$). If $h \leq 0$, we have $H_1(h) = 0$ because $rcos\theta + hr^2 sin^2\theta \leq cos\theta$. By doing exactly the same computations for SI_h, we get in that second case for the associated function $H_2(h) = h^-$. We notice that in both cases, $H_i(0) = 0$, so that the conclusions (i) and (ii) of Theorem 11.2 apply. Adding the relations (i) thus obtained for IS_h and SI_h yields the announced result. □

References.

Bence, Merriman and Osher [291] discovered and gave some heuristic arguments proving that a convolution of a shape with a gaussian followed by a $\frac{1}{2}$ thresholding simulated the motion by mean curvature $\frac{\partial u}{\partial t} = |Du| curv(u)$. The consistency of their arguments was checked by Mascarenhas [288]. Barles-Georgelin [50] and Evans [136] also give a consistency proof and in addition prove that the iterated weighted gaussian median filtering converges to the mean curvature motion. An extension of this result to any it-

erated weighted median filter is given in [225]. An interesting attempt to generalize this result to vector median filters is made in Caselles et al. [91]. Catté, Dibos and Koepfler [94] linked the mean curvature motion with classical morphological filters whose structuring elements are oriented one-dimensional sets in all directions [311, 390] A question arises from the main asymptotic convergence theorem : are there sets of structuring elements permitting to obtain any function H ? As is shown in this chapter, $H(s) = s$ is attained by the median filter. Pasquignon [336] has extensively studied this question and gives a positive answer for all power functions $H(s) = s^\alpha$ with simple families of structuring elements. The presentation of the main results of this chapter is mainly original and was announced in the tutorials [191, 187] with an early version in [190].

Chapter 12

Asymptotic behavior of contrast invariant operators (Dimension N)

We now consider a generalization of the asymptotic results of the preceding chapter to arbitrary dimensions. Our aim is to show that all inf-sup filters, when rescaled, are equivalent to a motion of the volumic image by its principal curvatures. We refer to Section 5.6 for the definition of principal curvatures of a level surface. We shall then, as in the preceding chapter, analyse several multidimensional filters. We shall in particular relate the median filter to the mean curvature of the level surface.

12.1 Asymptotic behavior theorem in arbitrary dimension.

Our main asymptotic theorem will be a simple adaptation to arbitrary dimensions of Theorems 11.1 and 11.2 ; the proof will be basically the same. We just need to fix some notations. We consider the Euclidean space $\mathbb{R}^N$, and if $\mathbf{x} \in \mathbb{R}^N$ we set $\mathbf{x} = (x, y_2, ..., y_N) = (x, y)$, so that $y \in \mathbb{R}^{N-1}$. We also set $b = (b_2, ..., b_N)$ and for any $h > 0$,

$$H(b) = T(x_1 + b_2 y_2^2 + ... + b_N y_N^2)(0) \tag{12.1}$$

where

$$Tu(\mathbf{x}) = \sup_{B \in \mathbb{B}} \inf_{\mathbf{y} \in \mathbf{x}+B} u(\mathbf{y}).$$

In this chapter, $\mathbb{B}$ is assumed to be a set of parts of $\mathbb{R}^N$ which is invariant by linear isometries of $\mathbb{R}^N$ and such that all elements of $\mathbb{B}$ are contained in a ball $B(0, M)$. As in the preceding section,

$$T[x_1 + h(b_2 y_2^2 + ... + b_N y_N^2)]$$

is a short notation for
"$T(u)$ with $u(\mathbf{x}) = u(x, y) = x_1 + h(b_2 y_2^2 + ... + b_N y_N^2)$".
Since T is monotone, H is a nondecreasing function with respect to each b_i. It is, in addition, Lipschitz, by the same argument yielding (11.4) :

$$|H(b) - H(b')| \leq M^2 |b - b'|. \tag{12.2}$$

If $u(\mathbf{x})$ is a C^3 function, we denote, as in Definition 5.16, by $\kappa(u) = (\kappa_2, ..., \kappa_N)$ the $N-1$ eigenvalues of the restriction of $D^2u(\mathbf{x})$ to the hyperplane $(Du)^\perp$ orthogonal to Du (provided $Du \neq 0$.) The κ_i are the principal curvatures of the level surface of u, $\{\mathbf{y},\ u(\mathbf{y}) = u(\mathbf{x})\}$ passing by $\mathbf{x}$.

Theorem 12.1 *Let $\mathcal{B}$ be a family of structuring elements in $\mathbb{R}^N$ which is bounded ($\forall B \in \mathcal{B}$, $B \subset B(0,M)$) and isotropic (if $B \in \mathcal{B}$, then $RB \in \mathcal{B}$ for every isometry R of $\mathbb{R}^N$).*

Let $Tu(\mathbf{x}) = \inf_{B\in\mathcal{B}} \sup_{\mathbf{y}\in\mathbf{x}+B} u(\mathbf{y})$ or $Tu(\mathbf{x}) = \sup_{B\in\mathcal{B}} \inf_{\mathbf{y}\in\mathbf{x}+B} u(\mathbf{y})$ and define the rescaled operator

$$T_h u(\mathbf{x}) = \inf_{B\in h\mathcal{B}} \sup_{\mathbf{y}\in\mathbf{x}+B} u(\mathbf{y}) \quad (\text{ resp. } T_h u(\mathbf{x}) = \sup_{B\in h\mathcal{B}} \inf_{\mathbf{y}\in\mathbf{x}+B} u(\mathbf{y}) \ )$$

Then, setting $H(b) = T[x_1 + b_2 y_2^2 + ... + b_N y_N^2](0)$ and assuming that u is a C^3 function on $\mathbb{R}^N$,

(i) if $H(0) \neq 0$, then

$$(T_h u)(\mathbf{x}) = u(\mathbf{x}) + hH(0)|Du|(\mathbf{x}) + O(h^2);$$

(ii) if $H(0) = 0$, then on every compact set K contained in $\{\mathbf{x}, Du(\mathbf{x}) \neq 0\}$,

$$T_h u(\mathbf{x}) = u(\mathbf{x}) + h|Du(\mathbf{x})|H(h\frac{1}{2}\kappa(u)) + O_{\mathbf{x}}(h^3);$$

where $|O_{\mathbf{x}}(h^3)| \leq C_K h^3$ for some constant C_K depending only on u and K;

(iii) if $H(0) = 0$, then on every compact K of $\mathbb{R}^N$,

$$|T_h u(\mathbf{x}) - u(\mathbf{x})| \leq h^2 M^2 ||D^2 u(\mathbf{x})|| + O_{\mathbf{x}}(h^3),$$

where $|O_{\mathbf{x}}(h^3)| \leq C_K h^3$ for some constant C_K depending only on u and K.

Proof. This proof follows exactly the proofs of Theorems 11.1 and 11.2 ; we simply have to adjust the preliminaries of the proof of Theorem 11.1. We notice that $T(u - u(\mathbf{x})) = Tu - u(\mathbf{x})$ and that, T being translation and isometry invariant, we can choose the origin 0 at $\mathbf{x}$, and the orthogonal axes $(\vec{i}_1, ..., \vec{i}_N)$ so that $\vec{i}_1 = \frac{Du}{|Du|}(0)$ if $Du \neq 0$ and $\vec{i}_2, ..., \vec{i}_N$ are the eigenvectors of the restriction of $D^2u(\mathbf{x})$ to the hyperplane $Du(\mathbf{x})^\perp$.

Since u is C^3, we can therefore write

$$u(\mathbf{x}) = px + ax^2 + b_2 y_1^2 + ... + b_N y_N^2 + (c.y)x + O(|\mathbf{x}|^3), \tag{12.3}$$

where $p = |Du|(0) \geq 0$, $c = (c_2, ..., c_N)$, $c.y$ denotes the scalar product in $\mathbb{R}^{N-1}$ and for $l = 2, ..., N$,

$$\begin{aligned} b_l &= \tfrac{1}{2}\tfrac{\partial^2 u}{\partial y_l^2}(0) = \tfrac{1}{2}D^2u(\vec{i}_l, \vec{i}_l) \\ a &= \tfrac{1}{2}\tfrac{\partial^2 u}{\partial x^2}(0) = \tfrac{1}{2}D^2u(\vec{i}_1, \vec{i}_1) \\ c_l &= \tfrac{\partial^2 u}{\partial x \partial y_l}(0) = D^2u(\vec{i}_1, \vec{i}_l). \end{aligned} \tag{12.4}$$

In addition, if $p \neq 0$, we have

$$b_l = \frac{1}{2}|Du|\kappa_i(u)(0). \tag{12.5}$$

¿From this point, the proof of Theorem 12.1 is identical to the proof of Theorem 11.2. We simply have to replace in formulas (11.9) to (11.15) the expressions "cxy" by "$x(c.y)$", "by^2" by "$b_2y_1^2 + ... + b_Ny_N^2$" and "curv$(u)$" by "$\kappa(u)$". Of course, "$|c|$" denotes the Euclidean norm of c in $\mathbb{R}^{N-1}$. □

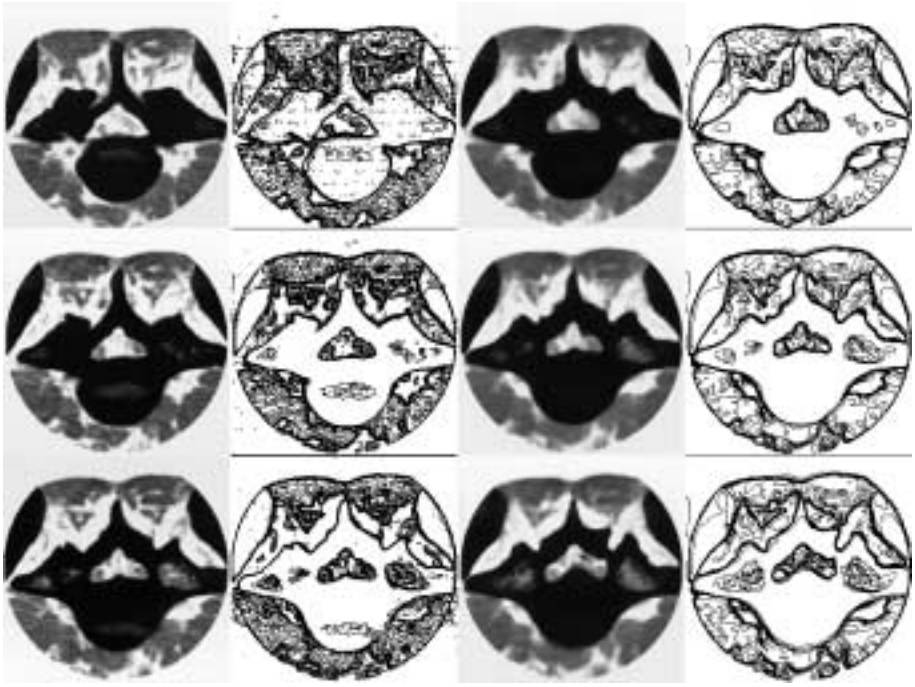

Figure 12.1: Three dimensional median filter. The original three dimensional image is made of 20 cuts of a vertebra. We display on the left column three succesives cuts, and on the next column their level-lines with levels multiple of 20. The next column shows these three cuts after one iteration of the median filter done with a 3D ball of radius 2, and the resulting level-lines.

12.2 Asymptotic behavior of weighted median filters in arbitrary dimension.

In this section, we consider C^3 real functions u defined in $\mathbb{R}^N$ and a radial, nonnegative, continuous function $k(\mathbf{x}) = k(|\mathbf{x}|)$ satisfying $\int_{\mathbb{R}^N} k(\mathbf{x})d\mathbf{x} = 1$. We have defined the measure with density k by

$$|E|_k = \int_E k(\mathbf{x})d\mathbf{x}$$

and the median operator, weighted by k by Formula (10.3), that is

$$\mathrm{med}_k(u)(\mathbf{x}) = \sup_{|B|_k \geq \frac{1}{2}} \inf_{\mathbf{y} \in B+\mathbf{x}} u(\mathbf{y}) \tag{12.6}$$

We assume also that k satisfies the assumption of Proposition 10.10, so that for continuous functions u, $\mathrm{med}_k^- u$ and $\mathrm{med}_k u$ coincide. This is true, e.g., if k is continuous and if the set $\{\mathbf{x}, k(\mathbf{x}) > 0\}$ is connected. We now give an infinitesimal interpretation of the median filters. In order to do so, we "shrink" $k(\mathbf{x})$ into $k_h(\mathbf{x}) = h^{-N}k(\frac{\mathbf{x}}{h})$, which corresponds to the scaling of the structuring elements by a factor h. Without risk of confusion, we write med_h for med_{k_h}. We have analysed in the preceding chapter the asymptotic behaviour of the median filter in two dimensions when k is the characteristic function of a disk. The aim of the next theorem is to adress, in arbitrary dimension, more general forms of k. We shall not make the analysis in full generality, because this would be uselessly heavy. We shall just treat the case where k is smooth and compactly supported. We shall also adress the case where k is no more compactly supported,

but in that case we won't be in a position to apply Theorem 12.1 because the set of structuring elements is then no more bounded. We shall adress and solve later this difficulty, in Chapter **??**. As a canonical example of application of the next theorem we can think of the case where k is the Gauss function.

Theorem 12.2 *Let k be a radial nonnegative function with integral 1 in the Schwartz class $\mathcal{S}$. We assume that the support of k is compact, connected, and contains 0. Then, there exists a constant $c(k)$ such that if u is a C^3 Lipschitz function in $\mathbb{R}^N$, then*

(i) On every compact set K contained in $\{\mathbf{x}, Du(\mathbf{x}) \neq 0\}$,

$$(\mathrm{med}_h u)(\mathbf{x}) = u(\mathbf{x}) + \frac{1}{2}h^2|Du(\mathbf{x})|c(k)\Sigma_{i=2}^N \kappa_i(u)) + O_{\mathbf{x}}(h^3);$$

where $|O_{\mathbf{x}}(h^3)| \leq C_K h^3$ for some constant C_K depending only on u and K.

(ii) On every compact K of $\mathbb{R}^N$,

$$|\mathrm{med}_h u(\mathbf{x}) - u(\mathbf{x})| \leq h^2 M^2 ||D^2 u(\mathbf{x})|| + O_{\mathbf{x}}(h^3),$$

where $|O_{\mathbf{x}}(h^3)| \leq C_K h^3$ for some constant C_K depending only on u and K.

Lemma 12.3 *Let k be a radial nonnegative function with integral 1 in the Schwartz class $\mathcal{S}$. We assume that the support of k is connected, and contains 0. (We do not assume that the support of k is compact, so that the Gauss function is a valid example). Then the function $H(hb) = \mathrm{med}_k[x + h(b_2 y_2^2 + ... + b_N y_N^2)](0)$ associated with the weighted median filter by Relation (12.1) satisfies*

$$H(hb) = c(k)(\Sigma_{i=2}^N b_i)h + o(h),$$

where

$$c(k) = \frac{\int_{\mathbb{R}^{N-1}} y_2^2 k(y)dy}{\int_{\mathbb{R}^{N-1}} k(y)dy}.$$

and $y = (y_2, ..., y_N) \in \mathbb{R}^{N-1}$, $b = (b_2, ..., b_N) \in \mathbb{R}^{N-1}$.

Proof. Repeating the analysis of Lemma 11.4, we notice that the median value $m(h) = \mathrm{med}_k[x + h(b_2 y_2^2 + ... + b_N y_N^2)](0)$ is defined by

$$\int_{\mathbb{R}^{N-1}} dy \int_0^{m(h) - h(b_2 y_2^2 + ... + b_N y_N^2)} k((x^2 + y^2)^{\frac{1}{2}})dx = 0. \tag{12.7}$$

This formula uniquely defines $m(h)$ because its first member is a strictly increasing, continuous function of $m(h)$ in a neighborhood of 0. Formally differentiating (12.7) with respect to h yields

$$m'(h) \int_{\mathbb{R}^{N-1}} k((y^2 + (m(h) - h(b_2 y_2^2 + ... + b_N y_N^2))^2)^{\frac{1}{2}})dy =$$

$$\int_{\mathbb{R}^{N-1}} (b_2 y_2^2 + ... + b_N y_N^2) k((y^2 + (m(h) - h(b_2 y_2^2 + ... + b_N y_N^2))^2)^{\frac{1}{2}})dy.$$

In order to reduce the size of formulas, we set

$$\varphi(a, h) = \int_{\mathbb{R}^{N-1}} (b_2 y_2^2 + ... + b_N y_N^2) k((y^2 + (a - h(b_2 y_2^2 + ... + b_N y_N^2))^2)^{\frac{1}{2}})dy, \tag{12.8}$$

$$\psi(a,h) = \int_{I\!R^{N-1}} k((y^2 + (a - h(b_2 y_2^2 + ... + b_N y_N^2))^2)^{\frac{1}{2}})dy. \tag{12.9}$$

These functions are well defined, C^∞ and positive for small a and h because of the assumptions on k. So we obtain

$$m'(h)\psi(m(h),h) - \varphi(m(h),h) = 0,$$

and formally again,

$$m'(0) = \frac{\varphi(0,0)}{\psi(0,0)} = (\Sigma_{i=2}^N b_i)\frac{\int_{I\!R^{N-1}} y_2^2 k(y)dy}{\int_{I\!R^{N-1}} k(y)dy} = c(k)(\Sigma_{i=2}^N b_i). \tag{12.10}$$

This formal computation is easy to justify. We simply introduce the ordinary differential equation

$$\tilde{m}'(h) = \frac{\psi(\tilde{m}(h),h)}{\varphi(\tilde{m}(h),h)}, \quad \tilde{m}(0) = 0, \tag{12.11}$$

which has, by Peano Theorem ([?]), a solution on some interval $[0, h_0[$. Indeed, the second member of (12.11) is C^∞ with respect to h and $\tilde{m}(h)$. Multiplying both members of (12.11) by $\varphi(\tilde{m}(h),h)$ and integrating between 0 and h yields

$$\int_{I\!R^{N-1}} dy \int_0^{\tilde{m}(h) - h(b_2 y_2^2 + ... + b_N y_N^2)} k((x^2 + y^2)^{\frac{1}{2}})dx = 0. \tag{12.12}$$

Thus $\tilde{m}(h)$ satisfies (12.7) and therefore $m(h) = \tilde{m}(h)$ on $[0, h_0]$. As a consequence, (12.10) is true. We can also differentiate (12.11) as many times as we wish with respect to h and we therefore obtain by Taylor formula

$$H(hb) = m(h) = c(k)(\Sigma_{i=2}^N b_i)h + O(h^2).$$

In addition, since ϕ and ψ and their derivatives with respect to a and h are continuous functions of b at any order of derivation, the obtained $O(h^2)$ is obviously uniform if $b = (b_2, ..., b_N)$ stays in an arbitrary compact set of $I\!R^{N-1}$. □

Proof of Theorem 12.2. The operator $Tu(\mathbf{x}) = \mathrm{med}_k u$ has a "supinf" form and his set of structuring elements

$$I\!B = \{B,\ |B|_k \geq \frac{1}{2},\ B \subset \mathrm{Support}(k)\}$$

is isotropic and bounded. Thus, T satisfies the conditions of Theorem 12.1. By Lemma 12.3, the function H associated with T satisfies $H(0) = 0$, and we obtain the conclusions (ii) and (iii) of Theorem 12.1. Using Lemma 12.3 yields

$$H(hb) = c(k)(\Sigma_{i=2}^N b_i)h + O(h^2)$$

and therefore, by Theorem 12.1(ii),

$$(\mathrm{med}_h u)(\mathbf{x}) = u(\mathbf{x}) + h^2 |Du(\mathbf{x})| c(k)(\frac{1}{2}\Sigma_{i=2}^N \kappa_i(u))) + O_{\mathbf{x}}(h^3)$$

with $|O_{\mathbf{x}}(h^3)| \leq O_K(h^3)$ on compact sets K where $|Du(\mathbf{x})| \neq 0$. The assertion (ii) follows immediately from Theorem 12.1(ii). □

Exercise 12.1 *Adapt the proof of Theorem 12.2 and of Lemma 12.3 to the case where k only is C^0 and compactly supported, with connected support containing 0 in its interior. (As an example, we can take k proportional to the characteristic function of the unit ball).*

12.2.1 The three-dimensional case

The third dimension opens many degrees of freedom in the synthesis of contrast invariant operators with respect to dimension 2 ; indeed, we can now play with two variables instead of one, the principal curvatures of the level surface, which we order so that $\kappa_1 \leq \kappa_2$. We list a series of examples which are easy applications of Theorem 12.1. We shall in the following formulas consider a two-dimensional set B and we set $I\!B = \{RB, \quad R \in SO(3)\}$. The set of structuring elements $I\!B$ is obtained by rotating in $I\!R^3$ a single subset of $I\!R^2$ in all possible ways. We set, as usual,

$$SI_h u(\mathbf{x}) = \sup_{B \in h I\!B} \inf_{\mathbf{y} \in \mathbf{x}+B} u(\mathbf{y})$$

and

$$IS_h u(\mathbf{x}) = \inf_{B \in h I\!B} \sup_{\mathbf{y} \in \mathbf{x}+B} u(\mathbf{y}).$$

We then have the following formulas.

- If B is a segment with length 2 centered at 0:

$$IS_h u = u + \frac{1}{2} h^2 \kappa_1^+ |Du| + O(h^3),$$

$$IS_h(SI_h)u = u + \frac{1}{2} h^2 (sign(\kappa_1(u)) + sign(\kappa_2(u)) min(|\kappa_1|, |\kappa_2|) + O(h^3).$$

- If B is made of two symmetric points, (1,0,0) and (-1, 0, 0),

$$IS_h u = u + \frac{1}{2} h^2 |Du| min(\kappa_1(u), \kappa_2(u)) + O(h^3),$$

$$SI_h u = u + \frac{1}{2} h^2 |Du| max(\kappa_1(u), \kappa_2(u)) + O(h^3),$$

$$IS_h(SI_h)u = u + \frac{1}{2} h^2 |Du| (\kappa_1(u) + \kappa_2(u)) + O(h^3).$$

The last formula yields the mean curvature of the level surface of u at $\mathbf{x}$.

- If B is made of two orthogonal segments with length 2 each and centered at 0,

$$IS_h u = u + h^2 |Du| \frac{1}{2} (\kappa_1(u) + \kappa_2(u))^+ + O(h^3),$$

$$SI_h u = u + h^2 |Du| \frac{1}{2} (\kappa_1(u) + \kappa_2(u))^- + O(h^3).$$

Of course, one can get directly the mean curvature by simply taking for B the endpoints of both orthogonal segments. Another possibility to obtain the mean curvature is to alternate the preceding schemes, or to add them.

Sketch of proof. In order to prove the preceding formulas, we simply compute for all considered operators T the function

$$H(b) = T[x_1 + b_2 y_2^2 + b_3 y_3^2](0),$$

where $\mathbf{x} = (x_1, y_2, y_3)$. In all cases, we have $H(0) = 0$ so that, by Theorem 12.1,

$$T_h u(\mathbf{x}) = u(\mathbf{x}) + h|Du(\mathbf{x})|H(h\frac{1}{2}\kappa_1, h\frac{1}{2}\kappa_2) + O(h^3). \tag{12.13}$$

Let us for instance compute $H(b)$ in the first case, when $T = IS$ and B is a segment with length 2 centered at 0. By a symmetry argument, it is easily seen that the "inf-sup" is attained for a segment contained in the plane $x_1 = 0$, so that

$$H(b) = \inf_{I\!B} \sup_{B \in I\!B} b_2 y_2^2 + b_2 y_3^2 = \inf_{\theta} \sup_{0 \le r \le 1} r^2 (b_2 \cos^2\theta + b_3 \sin^2\theta).$$

If $b_2 \le 0$ or $b_3 \le 0$, this yields $H(b) = 0$. If $0 \le b_2 \le b_3$, we get $H(b) = b_2$. Thus $H(b) = b_2^+$ and we obtain from (12.13) :

$$T_h u(\mathbf{x}) = u(\mathbf{x}) + h^2 |Du(\mathbf{x})| \frac{1}{2}\kappa_1^+ + O(h^3)$$

as announced. The other formulas are computed in a totally similar way and are left as exercises.

□

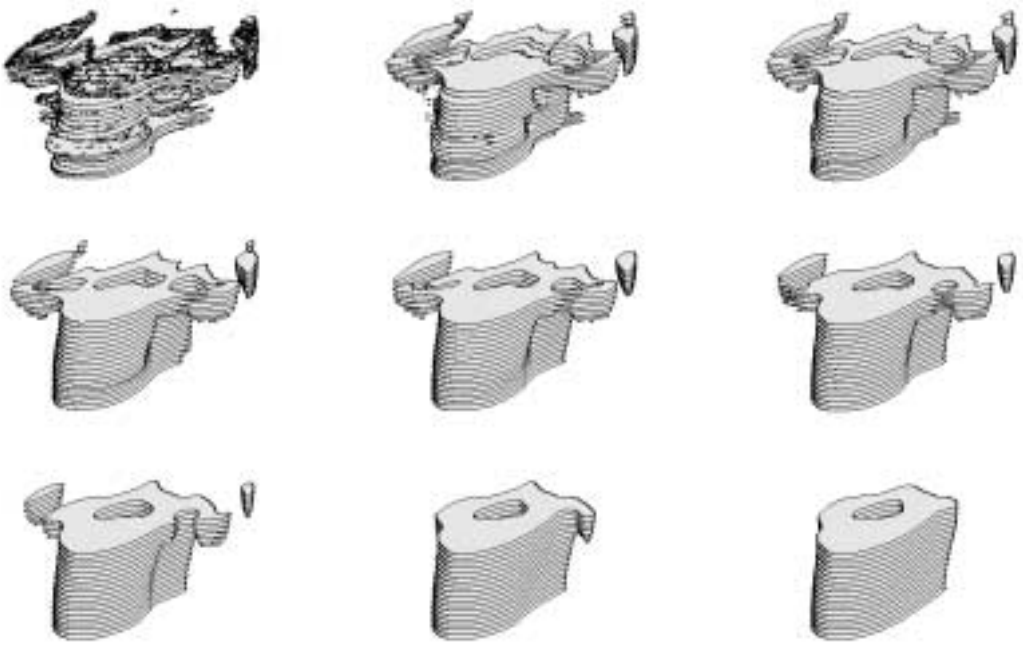

Figure 12.2: Median filtering of a three-dimensional image. First image : representation of the horizontal slices of a 3-D level-surface of the 3-D image of a vertebra. Right-left and up-down : 1, 2, 5, 10, 20, 30, 60, 100 iterations of a three-dimensional median filter, supported by a ball with radius 3. We have just proved that this scheme is a possible implementation of the mean curvature motion, originally proposed as such by Bence, Merriman and Osher.

References.

The references for this chapter are essentially the same as in the preceding one, devoted to the dimension 2. The main theorem on the asymptotic behaviour of morphological filters was first stated and proved in [191, 187]. The examples, given in the last section, of possible morphological filtering with their subjacent surface motion equations are not published elsewhere. The consistency of gaussian smoothing followed by thresholding with the Mean Curvature Motion was proved in increasing mathematical sophistication and generality by Bence, Merriman and Osher [291], Mascarenhas [288], Barles-Georgelin [50] and Evans [136]. Our presentation here is slightly more general than the one in these authors (we allow any weight function in the Schwartz class). The best general result is given by Ishii [225].

Chapter 13

Affine invariant mathematical morphology

In this chapter, we consider ways to erode or dilate a shape in an affine invariant way. The affine invariance requirement is a further extension of the obvious translation invariance and the isotropy (rotation invariance) which we already considered for convolutions, erosions and dilations. Let us mention that a translation is characterized as a transform which preserves distances between points and the angle of each vector with a fixed direction. Isometries (rotations-translations) preserve distances between points and angles between vectors. Similarities (i.e. isometries followed by a zoom) preserve angles between points. Then, affine transforms are characterized as transforms of an image which preserve parallelism and projective transforms as transforms which preserve alignments. This is the classical hierarchy of "anamorphoses", or image deformations, we can face in image processing.

When we take a photograph of a plane image, like a painting, we perform a transform which preserves alignments, that is, a projective transform. When we take a photograph of a plane image from a long enough distance, however, the transform tends to preserve parallelism. We then perform an affine transform and a rectangle looks like a parallelogram. This rule is used in the paintings of the traditional chinese painting, which tends to always display scenes seen from some distance. (See Exercise 13.3 at the end of this chapter.) To take a more up to date example, most photocopy machines perform an affine transform and so do fax machines and even scanners. Thus, the analysis of scanned or copied plane documents must be affine invariant in order to get rid of this artefactual deformation. We shall here explore with much detail set operators which are affine invariant. We shall prove further in this book (Chapter 21 that it is impossible to perform a shape smoothing with more invariance than the affine one.

Definition 13.1 *Let* $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ *be an arbitrary matrix such that* $\det \mathrm{A} = \mathrm{ad} - \mathrm{cb} = 1$. *The set of such matrices is the so called "special linear group",* $SL(\mathbb{R}^2)$. *We say that an operator* T *is special affine invariant if* T *commutes with* A *for every* A *in* $SL(\mathbb{R}^2)$ *:* $AT = TA$.

We first define an "affine invariant distance" of a point $\mathbf{x}$ to a set X which will be a substitute to the classical euclidean one. We consider shapes X, that is, in whole generality, subsets of $\mathbb{R}^2$. Let $\mathbf{x} \in \mathbb{R}^2$ and Δ an arbitrary straight line passing by $\mathbf{x}$. We consider all connected components of $\mathbb{R}^2 \setminus (\bar{(}X) \cup \Delta)$. If $\mathbf{x} \notin \bar{X}$, two and only two of them contain $\mathbf{x}$ in their boundary. We denote them by $\mathrm{CA}_1(\mathbf{x}, \Delta, \mathrm{X})$,

$\mathrm{CA}_2(\mathbf{x}, \Delta, \mathrm{X})$, we call them the "chord-arc sets" defined by $\mathbf{x}$, Δ and X, and we order them so that $\mathrm{area}(\mathrm{CA}_1(\mathbf{x}, \Delta, \mathrm{X})) \leq \mathrm{area}(\mathrm{CA}_2(\mathbf{x}, \Delta, \mathrm{X}))$.

Definition 13.2 *Let X be a subset of $\mathbb{R}^2$ and $\mathbf{x} \in \mathbb{R}^2, \mathbf{x} \notin \bar{X}$. We call affine distance of $\mathbf{x}$ to X the real number $\delta(\mathbf{x}, X) = \inf_\Delta \mathrm{area}(\mathrm{CA}_1(\mathbf{x}, \Delta, \mathrm{X}))^{1/2}$ if $\mathbf{x} \notin \bar{(}X)$, $\delta(\mathbf{x}, X) = 0$ if $\mathbf{x} \in \bar{(}X)$.*

Remark 13.3 *Obviously, we take the power $1/2$ in order that the affine distance be homogeneous to a length. The affine distance can be infinite :*

Take e.g. a closed convex set X and $\mathbf{x}$ outside X. Then it is easily seen that $\delta(\mathbf{x}, X) = +\infty$ because all chord-arc sets defined by all straight lines Δ are infinite.

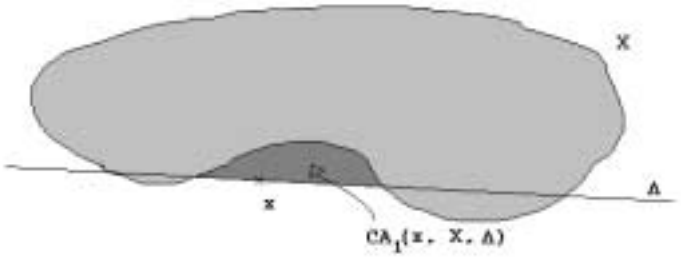

Figure 13.1: Affine distance to a set.

Definition 13.4 *Let X be a shape, i.e. a subset of $\mathbb{R}^2$. We call affine a-dilate of a set X the set $\tilde{D}_a X = \{\mathbf{x}, \delta(\mathbf{x}, X) \leq a^{1/2}\}$. We call affine a-eroded of set X the set $\tilde{E}_a X = \{\mathbf{x}, \delta(\mathbf{x}, X^c) > a^{1/2}\} = (\tilde{D}_a X^c)^c$.*

Exercise 13.1 *Show that $\tilde{E}_a X = (\tilde{D}_a (X^c))^c$. This relation shows that it is equivalent to erode a set or to dilate its complementary set : a useful symmetry, since the same shape may appear as an upper level set or as the complementary set of an upper level set, depending on whether it is darker or brighter than the background.*

Remark 13.5 *The denominations "erosion" and "dilation" for the preceding operators do not correspond to the standard terminology in Mathematical Morphology. Indeed, a dilation must commute with the set union and an erosion with the set intersection. It is easily checked that the affine erosions and dilations as defined above do not satisfy this requirement. [1]. All the same, we shall maintain these names, because affine erosions and dilations as we have just defined are clearly sound generalizations of the euclidean ones.*

Proposition 13.6 *The affine invariant erosions and dilations $\tilde{E}_a$ and $\tilde{D}_a$ are special affine invariant monotone operators.*

[1] We thank Michel Schmitt for pointing us out this fact

Proof It is easily seen that if $X \subset Y$, then for every $\mathbf{x}$, $\delta(\mathbf{x}, X) \geq \delta(\mathbf{x}, Y)$. From this, we deduce that $X \subset Y \Rightarrow \tilde{D}_a X \subset \tilde{D}_a Y$. The monotonicity of $\tilde{E}_a$ follows by the duality relation $\tilde{E}_a X = (\tilde{D}_a X^c)^c$. The special affine invariance of $\tilde{D}_a$ and $\tilde{E}_a$ follows from the fact that if detA = 1, then area(X) = area(AX). □

Exercise 13.2 *Show that $\tilde{E}_a$ and $\tilde{D}_a$ are affine invariant in a sense similar to Definition 20.19, namely that for every linear map $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with $\det A > 0$, $A\tilde{E}_{(\det A)^{-1/2}a} = \tilde{E}_a A$.*

We shall now use Matheron Theorem 8.3 in order to give a standard form to $\tilde{E}_a$ and $\tilde{D}_a$. According to this theorem, we can associate withh any translation invariant monotone operator acting on a set of subsets of $\mathbb{R}^N$ a family of sets $\mathcal{B} \subset \mathcal{P}(\mathbb{R}^N)$, defined by $\mathcal{B} = \{X, 0 \in T(X)\}$ and such that

$$T(X) = \bigcup_{B \in \mathcal{B}} \bigcap_{\mathbf{y} \in B} X - \mathbf{y} = \{\mathbf{x}, \exists B \in \mathcal{B}, \mathbf{x} + B \subset X\}.$$

We can apply this theorem to $\tilde{E}_a$, provided we identify its associated structuring elements.

Definition 13.7 *We say that B is an affine structuring element if B is a set whose interior contains 0, and if there is some $b > 1$ such that for every line Δ passing by 0, both connected components of $B \setminus \Delta$ containing 0 in their boundary have an area larger or equal to b. We denote the set of affine structuring elements by $\mathcal{B}_{\text{aff}}$.*

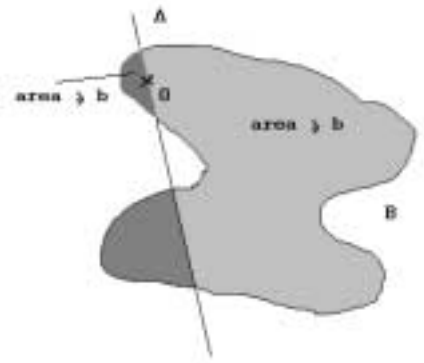

Figure 13.2: An affine structuring element : all lines passing by 0 divide B into several connected components. The two of them which contain 0 in their boundary have area larger or equal to B.

Proposition 13.8 *For every set X,*

$$\tilde{E}_a X = \bigcup_{B \in \mathcal{B}_{\text{aff}}} \bigcap_{\mathbf{y} \in a^{1/2} B} X - \mathbf{y} = \{\mathbf{x}, \exists B \in \mathcal{B}_{\text{aff}}, \mathbf{x} + a^{1/2} B \subset X\}$$

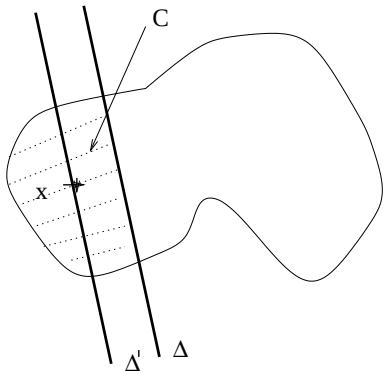

Figure 13.3: Illustration of the Corollary 13.10.

Proof We simply apply Matheron Theorem 8.3. The set of structuring elements associated with $\tilde{E}_a$ is $\mathbb{B} = \{X, \tilde{E}_a X \ni 0\}$. Now,

$$\tilde{E}_a X \ni 0 \Leftrightarrow \delta(0, X^c) > a^{1/2} \Leftrightarrow \inf_{\Delta} \text{area}(\text{CA}_1(0, \Delta, X))^{1/2} > a^{1/2}.$$

Let us call b the value of the preceding infimum. This las relation means that for every Δ, both connected components of $X \setminus \Delta$ containing 0 have area larger than $b > a$. Thus, X belongs to $a^{1/2}\mathbb{B}_{\text{aff}}$ by definition of $\mathbb{B}_{\text{aff}}$. □

By Proposition 13.8, $\mathbf{x}$ belongs to $\tilde{E}_a X$ if and only if for every straight line Δ, chord-arc sets containing $\mathbf{x}$ have an area strictly larger than a. Let us now characterize the points which belong to X but not to $\tilde{E}_a X$.

Definition 13.9 *We call chord-arc set of X any connected component of $X \setminus \Delta$, where Δ is an arbitrary straight line.*

Corollary 13.10 *$\tilde{E}_a X$ can be obtained from X by removing, for every straight line Δ, all chord-arc sets contained in X which have an area smaller or equal than a.*

Proof ATTENTION DEMONSTRATION A CONTROLER Let C be a chord-arc set of X, with area less or equal to a, and bounded by a straightline Δ. Then, we claim that for every $\mathbf{x} \in C$, $\delta(\mathbf{x}, X^c) \leq a^{\frac{1}{2}}$.

Indeed, consider Δ', the straight line parallel to Δ and passing by $\mathbf{x}$. One of $CA_1(\mathbf{x}, \Delta', X^c)$ or $CA_2(\mathbf{x}, \Delta', X^c)$ is contained in C and therefore has area less or equal to a. Thus $\delta(\mathbf{x}, X^c) \leq a^{\frac{1}{2}}$ and $\mathbf{x}$ does not belong to $\tilde{E}_a X$. Conversely, if $\mathbf{x} \in X \setminus \tilde{E}_a X$, then by definition there is some Δ passing by $\mathbf{x}$ such that $area(CA_1(\mathbf{x}, \Delta, X^c)) \leq a$. Thus $C = CA_1(\mathbf{x}, \Delta, X^c)$ is a chord-arc set with area less or equal to a and $\mathbf{x}$ belongs to C.

□

Corollary 13.10 has a huge numerical relevance : it gives an easy way to perform affine dilations and affine erosions. In the next section, we shall make a firt hint on how those operators are related to the (AMSS) model.

13.1 Application to affine alternate curve filters

One of the main difficulties of mathematical morphology is the non commutation of erosions and dilations, which entails the non commutation of opening and closing. Why is it a problem ? By performing an opening (with a ball) of a shape, we simply remove its external peaks. More precisely, if we do an opening by a ball with radius h, then the radius of curvature of the external peaks becomes larger than h and if we perform an opening of the shape, the radius of curvature of the peaks pointing inside becomes larger than h. Thus, the aim of opening and closing is to smooth out the peaks in exactly the same way and it would be quite desirable to have both operations made simultaneously. In order terms, they should commute, and we could talk about a "curvature thresholding" operator. This type of commutation is roughly asymptotically attained by performing alternate filters [378]. The idea is to chose a very small scale h, and to alternate the openings O and closings F in the following way :

$$O_t F_t O_{\frac{t}{2}} F_{\frac{t}{2}} ... O_{\frac{t}{2^n}} F_{\frac{t}{2^n}} ...$$

Since O_t and F_t are idempotent, the growth in scale is necessary in order to perform a progressive smoothing, up to the scale t.

In the same way, affine erosions and dilations do not commute and should alternate. Now, they are not idempotent and the situation is simpler : we only need to choose an incremental scale h small enough and to alternate $\tilde{E}_h$ and $\tilde{D}_h$, that is, to compute $(\tilde{E}_h \tilde{D}_h)^n$. In Chapter 24 we describe how to implement such an alternate filter. Its relevance will be demonstrated in Chapter **??**, where we prove that there is only one way to smooth out shapes in an affine invariant way, the "affine shortening" or AMSS model. In Section 18.5, we shall prove that when $h \to 0$, the iterated alternate affine erosion $(\tilde{E}_h \tilde{D}_h)^n$ converges, when $n \to +\infty$ and $h \to 0$ in an adequate way, to the solution of the affine shortening equation. More precisely, let c_0 be a Jordan curve, which is the boundary of a simply connected set X. Iterating affine erosions and dilations on X gives a numerical scheme that computes the affine shortening c_T of c_0 at a given scale T. If c_T is the curve represented by the function $s \mapsto \mathbf{C}(s,T)$, then

$$\frac{\partial \mathbf{C}}{\partial t}(s,t) = |\mathrm{Curv(s,t)}|^{1/3} \vec{\mathrm{n}}(\mathrm{s,t}) \tag{13.1}$$

where Curv(s,t) and $\vec{n}(s,t)$ are the curvature and the normal vector at the point with abscissa s of the curve $c_t = \mathbf{C}(\cdot, T)$.

Exercise 13.3 *The aim of this exercise is to prove that if an application $\tilde{A} : \mathbb{R}^2 \to \mathbb{R}^2$ preserves parallelism between pairs of points, then there is a linear map A and a vector $\mathbf{b}$ such that $\tilde{A}(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$. The preservation of parallelism will be stated in the following way. If any four points $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$ and $\mathbf{x}_4$ satisfy $\mathbf{x}_1 - \mathbf{x}_2 = \lambda(\mathbf{x}_3 - \mathbf{x}_4)$ for some real number λ, then there exists a real number μ such that $\tilde{A}\mathbf{x}_1 - \tilde{A}\mathbf{x}_2 = \mu(\tilde{A}\mathbf{x}_3 - \tilde{A}\mathbf{x}_4)$.*
1) Consider a basis $(\vec{i}, \vec{j})$ of the plane, set $0 = (0,0)$ and $\mathbf{x} = x\vec{i} + y\vec{j}$. Set $A\mathbf{x} = \tilde{A}\mathbf{x} - \tilde{A}0$ and prove that there exist real functions $\mu_1(x)$ and $\mu_2(y)$ such that $A(x\vec{i}) = \mu_1(x)\vec{i}$ and $A(y\vec{j}) = \mu_2(y)\vec{j}$.
2) Notice that A preserves parallelism and satisfies $A0 = 0$.
3) Show that $A(\mathbf{x}) = \mu_1(x)\vec{i} + \mu_2(y)\vec{j}$.
4) By using Thales' Theorem, show that $\mu_1(x) = \alpha\mu_2(x)$ for some real constant α.
5) Replace $\vec{i}$ by $\beta\vec{i}$ and apply the preceding results : deduce that $\beta\mu_1(\beta^{-1}x) = c$ is a constant and conclude.

References.

Clearly, planar shape recognition algorithms are be more efficient when they are affine invariant, since most optical duplication devices (scanners, photocopiers, faxes) perform slight affine distortion on planar images. Also, all diffeomorphisms are locally affine distortions. Shape affine invariant matching techniques are described in [222] and discussion on the role of affine and projective invariance for object recognition in [61, 431, 239]. Clearly, corners and T-junctions can be seen in projective scenes with arbitrary angles. The detection of angles between two straight lines must be affine invariant. Algorithms are proposed in [60, 17, 413, 121]. See Merriman, Bence, Osher [292] for a very original numerical view of the filtering of multiple junctions. Because of its relevance in Computer Vision, affine invariant definitions of classical geometric measure or integral geometry concepts have been actively sought for. [123] contains an interesting attempt to define an "affine invariant length" and an "affine invariant dimension", by analogy with Hausdorff lengths and dimensions. The diameters of the sets of a Hausdorff covering are simply replaced by their area. Several attempts to define affine invariant analyses of discrete set points are described in [169, 365]. [171] defines an affine invariant symetry set (skeleton) for shapes and [345] relates to affine invariance the 1/3 power law of planar motion perception and generation. Some of the techniques on affine erosions and dilations presented in this chapter were announced in [268]. We used liberally the Matheron formalism for set monotone operators [290].

Chapter 14

Localizing a family of structuring elements

When the set of structuring elements $\mathbb{B}$ is affine invariant, the associated rescaled operators

$$IS_h u(\mathbf{x}) = \inf_{B \in h\mathbb{B}} \sup_{\mathbf{y} \in B} u(\mathbf{x} + \mathbf{y})$$

commute with a lot of plane transforms : with space translations, $u(\mathbf{x}) \to u(\mathbf{x} + \mathbf{y}_0)$, with grey level translations $u(\mathbf{x}) \to u(\mathbf{x}) + C$, with linear maps with determinant 1, $u(\mathbf{x}) \to u(\mathrm{A}\mathbf{x})$ where $\mathrm{A} \in SL(\mathbb{R}^2)$ and with contrast changes $u \to g \circ u$ when g is continuous and nondecreasing. Now, the affine invariance entails a loss of locality. The value at $\mathbf{x}$ of $IS_h u$ may depend upon values of $u(\mathbf{y})$ at points which are arbitrarily far from $\mathbf{x}$, no matter how small h is. Indeed, the linear invariance permits to stretch an element of $\mathbb{B}$ in any direction. Indeed, if $B \in \mathbb{B}$, then $A_\varepsilon B \in \mathbb{B}$, where $A_\varepsilon = \begin{bmatrix} \varepsilon & 0 \\ 0 & \frac{1}{\varepsilon} \end{bmatrix}$.

Thus IS_h looks like an *a priori* non local operator. We shall see that this is not the case for a very general class of families $\mathbb{B}$, for which IS_h behaves like a local operator, though involving arbitrarily elongated sets. The locality property which we wish for IS_h (in fact, a local comparison principle) will be deduced from a corresponding "localizability" property for $\mathbb{B}$. The same problematic applies to median filters when their weight function is not compactly supported. So we shall also treat the case of gaussian weighted median filters and show that they in fact like local filters, their set of structuring elements $\mathbb{B}$ being "localizable".

14.1 Localizable families of structuring elements

Proposition 14.1 and definition *Let $\alpha > 0$ be a positive constant which will play the role of an "exponent of localizability". We shall say that a set of structuring elements $\mathbb{B}$ is α-localizable if there exists a constant $c > 0$ such that for every $\rho > c^{-1}$ we can assert that*

$$\forall B \in \mathbb{B}, \exists B' \in \mathbb{B}, B' \subset D(0, \rho)$$

and $B' \subset D_{\frac{c}{\rho^\alpha}}(B) = \{x, d(x, B) \leq \frac{c}{\rho^\alpha}\}$, where d denotes the euclidean distance, $d(x, B) = \inf_{y \in B} d(x, y)$. As a consequence, by rescaling by a scale factor h, we have the equivalent property :

$$\exists c > 0,\ \forall r > 0,\ \forall h \leq cr,\ \forall B \in \mathbb{B}_h,\ \exists B' \in \mathbb{B}_h,\ B' \subset D(0, r)$$

and $B' \subset D_{\frac{ch^{\alpha+1}}{r^\alpha}}(B)$.

Proof In order to deduce the second relation from the first, we simply set $r = \rho h$. We have $B \in I\!B$ if and only if $hB \in I\!B_h$. We therefore replace B by hB and B' by hB' and we get for the new B and B' in $I\!B_h$:

$$B' \subset D_{\frac{ch}{\rho^\alpha}}(B) = D_{\frac{h^{\alpha+1}}{r^\alpha}}(B),$$

provided $\rho > c^{-1}$, i.e. $r > c^{-1}h$, i.e. $h < cr$. □

Let us give a first criterion for the 1-localizability, which will apply to affine invariant families of structuring elements.

Proposition 14.2 *Let $I\!B$ be made of subsets of $I\!R^2$ containing 0. Assume that there exists $c > 0$ such that if $B \in I\!B$ and $r > c$, then the connected component of $D_{\frac{c}{r}}(B) \cap D(0,r)$ containing 0 is in $I\!B$ (resp. contains an element of $I\!B$). Then $I\!B$ is 1-localizable. (We denote by $D_{\frac{c}{r}}(B)$ the dilate of B, $\{x, d(x,B) \leq \frac{c}{r}\}$).*

Proof For any B in $I\!B$, we consider B', the connected component of $D_{\frac{c}{r}}(B) \cap D(0,r)$ containing 0. In the second case, we consider an element B' of $I\!B$ contained in this connected component. □

Proposition 14.3 *If $I\!B = I\!B_{\text{aff}}$ is the set of all affine structuring elements, then $I\!B$ is localizable.*

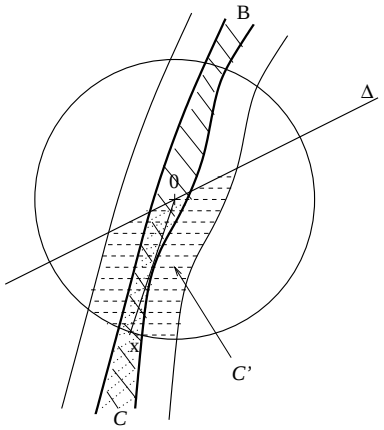

Figure 14.1: Proof of Proposition 14.3.

Proof We want to apply proposition 14.2. Let $B \in I\!B_{\text{aff}}$ and $b_B > 1$ such that for every Δ passing by 0, the areas of the connected components of $X \setminus \Delta$ containing 0 are larger than b_B. Let B' be the connected component of $D_{\frac{3}{r}}(B) \cap D(0,\rho)$ containing 0. We shall show that B' belongs to $I\!B_{\text{aff}}$. Let us consider $\mathcal{C}'$, one of the two connected components of $B' \setminus \Delta$ containing 0 in its boundary. Consider also $\mathcal{C}$, the connected component of $(B \setminus \Delta) \cap D(0,\rho)$ containing 0 in its boundary and on the same side of Δ as $\mathcal{C}'$. Notice that $\mathcal{C} \subset \mathcal{C}'$.

Two cases : If $\mathcal{C}$ does not meet $\partial D(0,\rho)$, then $\mathcal{C}$ is the one of the connected components of $B \setminus \Delta$ containing 0 and, in addition, $\mathcal{C}'$ contains $\mathcal{C}$. Thus area($\mathcal{C}' \geq b_B$ and we are done.

Let Δ be a line passing by 0. We consider one of the two connected components $\mathcal{C}'$ of $B' \setminus \Delta$ containing 0 in their boundary. We also consider $\mathcal{C}$ the connected component of $B \setminus \Delta$ containing 0 in their boundary and such that $\mathcal{C} \cap D(0,r) \subset B'$.
Two cases : if $\mathcal{C}$ is contained in $D(0,r)$, then by definition of chord-arc sets, the area of $\mathcal{C}$ is larger than b. Therefore, the area of $\mathcal{C}'$ is larger than b.
Second case : if $\mathcal{C}$ is not contained in $D(0,r)$, we consider the connected component $\mathcal{C}_1$ of $\mathcal{C} \setminus \partial D(0, r - \frac{2b}{r})$ which contains 0. $\mathcal{C}_1 \subset \mathcal{C}$ meets $\partial D(0, r - \frac{2b}{r})$ at some points, then, by connectedness, each line orthogonal to $[0,x]$ and passing by tx, with $0 \leq t \leq 1$, meets $\mathcal{C}_1$ at at least one point $x(t) \in D(0, r - \frac{2b}{r})$. Noting $\nu = \frac{x^\perp}{\|x\|}$, a unit vector orthogonal to x, we notice that the interval $[x(t) - \frac{2\nu}{r}, x(t) + \frac{2\nu}{r}]$ is contained in $D_{\frac{2b}{r}}(\mathcal{C}_1) \subset D_{\frac{2b}{r}}(\mathcal{C}) \subset B'$. In addition, one half at least of this interval is contained in $\mathcal{C}'$. Thus, provided $r \geq 2b$, area$(\mathcal{C}') \geq (r - \frac{2b}{r})\frac{2b}{r} = 2b - \frac{4b^2}{r^2} \geq b$. □
We can roughly say that all affine invariant families made of connected sets containing 0 are localizable, provided the shape of the elements is not too much distorted. The next example gives another example, where the "not too much distorted" condition is given by convexity.

Exercise 14.1 *Examples of localizable families.*
1) Let $I\!B$ be an affine invariant family of convex sets containing 0, all with area less than 1. Show that $I\!B$ is localizable. Here is a way to prove it : Let $B \in I\!B$. If $B \subset D(0,\rho)$, we are done. Otherwise, let $\mathbf{x}$ be the element of B with largest norm. Consider the special linear symmetric map A with eigenvectors $\mathbf{x}$ and $\mathbf{x}^\perp$ and respective eigenvalues $\frac{\rho}{|\mathbf{x}|}$ and $\frac{|\mathbf{x}|}{\rho}$. Then check that if ρ is larger than some universal constant C, then $AB \subset D(0,\rho)$. Set $B' = AB$, which belongs by assumption to $I\!B$, and prove that for some other universal constant c, $B' \subset D_{\frac{c}{\rho}}(B)$. Conclude.
2) Let B be a bounded connected set whose interior contains 0 and define an affine invariant family $I\!B = \{AB,\ A \in SL(I\!R^2)\}$, where $SL(I\!R^2)$ denotes the special linear group of all linear maps with determinant 1. Show, by using the same method as in 1), that $I\!B$ is localizable.

14.2 Localizability of the gaussian median filter

We shall now give another relevant instance of localizable family, associated with the weighted median filter. Although one can state much more general results, we shall restrict ourselves to the case where the weight function k of the median filter is the Gauss function $G(\mathbf{x}) = \frac{1}{(4\pi)^{\frac{N}{2}}} e^{-\frac{|\mathbf{x}|^2}{4}}$. We then have $\int_{I\!R^N} G(\mathbf{x})d\mathbf{x} = 1$ and we can consider in the following the family of structuring elements $I\!B$ associated with G. For brievety, we write $|B|_G$ for $\int_B G(\mathbf{x})d\mathbf{x}$. Then

$$I\!B = \{B \subset I\!R^N,\ |B|_G \geq \frac{1}{2}\}.$$

Proposition 14.4 *The set $I\!B$ associated with the gaussian median filter is α-localizable for every $\alpha > 0$. More precisely,*

$$\forall \alpha > 0,\ \exists c = c(\alpha),\ \forall \rho \geq c,\ \forall B \in I\!B,\ \exists B' \in I\!B,\ B' \subset B(0,\rho) \text{ such that } B' \subset D_{\frac{1}{\rho^\alpha}}(B).$$

Corollary 14.5 *As an immediate consequence of Definition-proposition 14.1, we obtain by scaling $\mathbb{B}$ into $\mathbb{B}_h = h\mathbb{B}$:*

$$\forall r > 0,\ \forall h \leq cr,\ \forall B \in \mathbb{B}_h,\ \exists B' \in \mathbb{B}_h,\ B' \subset B(0,r) \ \ \textit{and}\ B' \subset D_{\frac{h^\alpha+1}{r^\alpha}}(B).$$

Proof of Proposition 14.4 We begin by choosing ρ_0 such that

$$|B(0,\rho_0)|_G = \frac{3}{4}.$$

We then call ρ_1 the value $\rho_1 \geq \rho_0$ for which

$$|B(0,\rho_1)|_G = \frac{1}{8}$$

and we assume in the following $\rho \geq \rho_1$.
Two cases :

• If $B \in \mathbb{B}$ satisfies

$$|B|_G \geq \frac{1}{2} + |B(0,\rho)^c|_G,$$

we simply set $B' = B \cap B(0,\rho)$ and we get

$$|B'|_G \geq |B|_G - |B(0,\rho)^c|_G \geq \frac{1}{2},$$

which proves that $B' \in \mathbb{B}$, $B' \subset B$ answers the question.

• If $B \in \mathbb{B}$ satisfies

$$|B|_G \leq \frac{1}{2} + |B(0,\rho)^c|_G \leq \frac{1}{2} + \frac{1}{8},$$

then

$$|B \cap B(0,\rho_0)|_G \geq |B|_G - |B(0,\rho_0)^c|_G \geq \frac{1}{2} - \frac{1}{4} = \frac{1}{4}, \tag{14.1}$$

$$|B^c \cap B(0,\rho_0)|_G \geq |B(0,\rho_0)|_G - |B|_G \geq \frac{3}{4} - (\frac{1}{2} + \frac{1}{8}. \tag{14.2}$$

Since $G \leq 1$, we have for every measurable set A, $|A|_G \leq meas(A)$. Thus, from (14.1-14.2),

$$meas(B \cap B(0,\rho_0)) \geq \frac{1}{4} \text{ and } meas(B^c \cap B(0,\rho_0)) \geq \frac{1}{8}. \tag{14.3}$$

By Lemma 14.6, which we state in continuation, and Relation (14.3), we obtain, for a universal constant C,

$$meas(B(0,\rho_0) \cap D_{\frac{1}{\rho^\alpha}}(B \cap B(0,\rho_0)) \setminus B)) \geq \frac{C}{\rho^\alpha}$$

and therefore

$$|B(0,\rho) \cap D_{\frac{1}{\rho^\alpha}}(B \cap B(0,\rho)) \setminus B|_G \geq \frac{C}{\rho^\alpha} \frac{e^{-\frac{\rho_0^2}{4}}}{(4\pi)^{\frac{N}{2}}} \geq \frac{C}{\rho^\alpha} \tag{14.4}$$

for some universal constant C. We finally set

$$B' = B(0,\rho) \cap D_{\frac{1}{\rho^\alpha}}(B \cap B(0,\rho),$$

so that $d(B', B) \leq \frac{1}{\rho^\alpha}$, $B' \subset B(0, \rho)$. We then have by (14.4)

$$|B'|_G \geq |B|_G - |B(0,\rho)^c|_G + \frac{C}{\rho^\alpha},$$

that is

$$|B'|_G \geq |B|_G + \frac{C}{\rho^\alpha} - \int_{|\mathbf{x}|\geq\rho} G \geq |B|_G \geq \frac{1}{2},$$

for $\rho \geq c(\alpha)$, where $c(\alpha)$ only depends on α. Indeed,

$$\int_{|\mathbf{x}|\geq\rho} G \leq c_N \rho^{N-1} e^{-\frac{\rho^2}{4}}$$

for a universal constant c_N. □

Lemma 14.6 *Let $B(0,\rho_0$ be a ball and B a subset of $\mathbb{R}^N$ satisfying for some constant $\delta > 0$,*

$$meas(B \cap B(0,\rho_0)) \geq \delta \text{ and } meas(B^c \cap B(0,\rho_0)) \geq \delta.$$

Then there are constants $C(\rho_0,\delta)$ and $\eta_0(\rho_0,\delta)$ such that

$$\forall \eta \leq \eta_0, \quad meas(B(0,\rho_0) \cap D_eta(B \cap B(0,\rho_0)) \setminus B) \geq C\eta.$$

Proof This is an easy consequence of the isoperimetric inequality ([?]). □

14.3 A local comparison principle.

We shall now prove that our definition of "localizable family" for $\mathcal{B}$ indeed yields locality properties for the associated operator

$$IS_h u(\mathbf{x}) = \inf_{B\in h\mathcal{B}} \sup_{\mathbf{y}\in B} u(\mathbf{x}+\mathbf{y}).$$

Intuitively, an operator T_h depending on a scale parameter h is local if its value $(T_h u)(\mathbf{x})$ at a point $\mathbf{x}$ primarily depends upon the values of $u(\mathbf{y})$ in a neighborhood of $\mathbf{x}$ whose size tend to zero as h tends to zero. This can be stated in many ways. Thanks to the monotonicity of the considered operators, we can formulate it as a local comparison principle.

Lemma 14.7 [local maximum principle] *Let $\mathcal{B}$ be an α-localizable set of plane closed nonempty sets. Let u and v be two Lipschitz functions in a disk $D(\mathbf{x}_0, r)$ satisfying $u(\mathbf{x}) \leq v(\mathbf{x})$ in $D(\mathbf{x}_0, r)$. Let K be the Lipschitz constant of u. Then if $h \leq cr$,*

$$(IS_h u)(\mathbf{x}_0) \leq (IS_h v)(\mathbf{x}_0) + Kc\frac{h^{\alpha+1}}{r^\alpha}.$$

Proof of Lemma 14.7.

$$(IS_h v)(\mathbf{x}_0) = \inf_{B \in I\!B_h + \mathbf{x}_0} \sup_{\mathbf{y} \in B} v(\mathbf{y}) \geq \inf_{B \in I\!B_h + \mathbf{x}_0} \sup_{\mathbf{y} \in B \cap D(\mathbf{x}_0, r)} v(\mathbf{y})$$

$$\geq \inf_{B \in I\!B_h + \mathbf{x}_0} \sup_{\mathbf{y} \in B \cap D(\mathbf{x}_0, r)} u(\mathbf{y})$$

By Definition 14.1, for all $B \in I\!B_h + \mathbf{x}_0$, there exists B' in $I\!B_h + \mathbf{x}_0$, $B' \subset D(\mathbf{x}_0, r)$, such that $\delta(B, B') \leq c\frac{h^{\alpha+1}}{r^\alpha}$. Thus, since u is K-Lipschitz,

$$\sup_{\mathbf{y} \in B \cap D(\mathbf{x}_0, r)} u(\mathbf{y}) \geq \sup_{\mathbf{y} \in B'} u(\mathbf{y}) - Kc\frac{h^{\alpha+1}}{r^\alpha}.$$

The two preceding relations yield

$$(IS_h v)(\mathbf{x}_0) \geq \inf_{B \in I\!B_h + \mathbf{x}_0} \sup_{\mathbf{y} \in B \cap D(\mathbf{x}_0, r)} u(\mathbf{y}) \tag{14.5}$$

$$\geq \inf_{B' \in I\!B_h + \mathbf{x}_0,\ B' \subset D(\mathbf{x}_0, r)} \sup_{\mathbf{y} \in B'} u(\mathbf{y}) - Kc\frac{h^{\alpha+1}}{r^\alpha} \geq \inf_{B \in I\!B_h + \mathbf{x}_0} \sup_{\mathbf{y} \in B} u(\mathbf{y}) - Kc\frac{h^{\alpha+1}}{r^\alpha}.$$

We obtain

$$(IS_h u)(\mathbf{x}_0) = \inf_{B \in I\!B_h + \mathbf{x}_0} \sup_{\mathbf{y} \in B} u(\mathbf{y}) \leq (IS_h v)(\mathbf{x}_0) + Kc\frac{h^{\alpha+1}}{r^\alpha}.$$

□

The next lemma permits to fix an optimal relationship between the localization scale r and the operator's scale h.

Lemma 14.8 *(Locality) Let $I\!B$ be an α-localizable set of subsets of $I\!R^N$. Let u, v be two K-Lipschitz functions such that*

$$|u(\mathbf{x}) - v(\mathbf{x})| \leq C|\mathbf{x}|^3$$

in a neighborhood of 0. Then

$$|IS_h u(0) - IS_h v(0)| \leq (C + Kc)h^{\frac{3(\alpha+1)}{3+\alpha}}.$$

Proof. Applying the local maximum principle (Lemma 14.7), we deduce from the relation

$$v(\mathbf{x}) - Cr^3 \leq u(\mathbf{x}) \leq v(\mathbf{x}) + Cr^3$$

on $D(\mathbf{x}, r)$ that

$$(IS_h v)(0) - Cr^3 - Kc\frac{h^{\alpha+1}}{r^\alpha} \leq (IS_h u)(0) \leq (IS_h^r v)(0) + Cr^3 + Kc\frac{h^{\alpha+1}}{r^\alpha}$$

We choose r in such a way that both infinitesimals appearing in the second member of the former inequality be of the same order : $r^3 = \frac{h^{\alpha+1}}{r^\alpha}$, which yields

$$r = h^{\frac{\alpha+1}{3+\alpha}}\ r^3 = \frac{h^{\alpha+1}}{r^\alpha} = h^{\frac{3(\alpha+1)}{3+\alpha}}.$$

We obtain

$$(IS_h v)(0) - Ch^{\frac{3(\alpha+1)}{3+\alpha}} - Kch^{\frac{3(\alpha+1)}{3+\alpha}} \leq (IS_h u)(0) \leq (IS_h v)(0) + Ch^{\frac{3(\alpha+1)}{3+\alpha}} + Kch^{\frac{3(\alpha+1)}{3+\alpha}}.$$

□

The main interest of lemma 14.8 is to reduce the asymptotic analysis of the operator IS_h as $h \to 0$ to the case where it is applied to second order polynomials. In all that follows, we adopt the rule of scale $r = h^{\frac{\alpha+1}{3+\alpha}}$. This choice will be justified in Lemma14.8 to follow.
In the following, it will be useful to associate with the operators IS_h new operators

$$(IS_h^r u)(\mathbf{x}_0) = \inf_{B \in I\!B_h + \mathbf{x}_0} \sup_{\mathbf{y} \in B \cap D(\mathbf{x}_0, r)} u(\mathbf{y}),$$

whose locality has been enforced : we truncate all elements in $I\!B$ by removing their parts outside $D(0,r)$. We shall now estimate the asymptotic difference between IS_h and IS_h^r.

Lemma 14.9 *(Localization Lemma)*
Let $I\!B$ be an α-localizable set of structuring elements in $I\!R^N$ and u a K-lipschitz function on $D(\mathbf{x}_0, 1)$. Let us set $r = h^{\frac{\alpha+1}{3+\alpha}}$. Then

$$(i) \qquad IS_h^r u(\mathbf{x}_0) \leq IS_h u(\mathbf{x}_0) \leq IS_r^r u(\mathbf{x}_0) + Kch^{\frac{3(\alpha+1)}{3+\alpha}}.$$

for r small enough. As a consequence,

$$(ii) \qquad |SI_h^r u(\mathbf{x}_0) - SI_h u(\mathbf{x}_0)| \leq Kch^{\frac{3(\alpha+1)}{3+\alpha}}.$$

$$(iii) \qquad |IS_h^r u(\mathbf{x}_0) - IS_h u(\mathbf{x}_0)| \leq Kch^{\frac{3(\alpha+1)}{3+\alpha}},$$

$$(iv) \qquad |SI_h^r IS_h^r u(\mathbf{x}_0) - SI_h IS_h u(\mathbf{x}_0)| \leq 2Kch^{\frac{3(\alpha+1)}{3+\alpha}}$$

Proof : The inequality (i) follows from the inequalities (14.5), in the proof of Lemma 14.7, by taking $u = v$. The inequalities (ii) and (iii) are deduced by using $IS_s(-u) = -SI_s(u)$. In order to obtain (iv), we notice that if u is globally K-Lipschitz, then so is $IS_h u$ by Lemma 7.5. We have by (ii) and (iii)

$$SI_h^r u \leq SI_h u \leq Kch^{\frac{3(\alpha+1)}{3+\alpha}} + SI_h^r u \tag{14.6}$$

$$IS_h^r u \leq IS_h u \leq Kch^{\frac{3(\alpha+1)}{3+\alpha}} + IS_h^r u \tag{14.7}$$

Taking in (14.6) $u = IS_h u$ and using (14.7), we obtain

$$SI_h^r IS_h^r u \leq SI_h IS_h u \leq 2Kch^{\frac{3(\alpha+1)}{3+\alpha}} + SI_h^r IS_h^r u,$$

which proves (iv). □

References.

The mathematical techniques for localizing a set of structuring elements developped in this chapter were first explained in [191, 187, 190]. The version presented here is much simpler. Frédéric Cao's and Denis Pasquignon PhD's contain related localization techniques [73, 74, 336].

Chapter 15

Asymptotic behavior of affine and contrast invariant filters.

In this chapter, we analyze, in dimension 2, the asymptotic behaviour of contrast invariant affine invariant monotone operators. By *affine invariance*, we mean that the operators commute with all linear maps of the plane with determinant 1. We know that such operators have an inf sup (or sup inf) form,

$$Tu(\mathbf{x}) = \inf_{B \in I\!B} \sup_{\mathbf{y} \in \mathbf{x}+B} u(\mathbf{y})$$

We denote by $SL(I\!R^2)$ the special linear group, that is, the group of all linear transforms $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ such that $|det(\mathrm{A})| = |ad - bc| = 1$. If the operator T is affine invariant, its set $I\!B$ of structuring elements is obviously also invariant under special linear transforms of the plane : $AB \in I\!B$ if $B \in I\!B$ and $det(A) = 1$. Conversely, if $\mathrm{A}I\!B = \{AB, B \in I\!B\}$ is equal to $I\!B$, then the associated inf sup or sup inf operators T are affine invariant : Denoting by $\mathrm{A}u(\mathbf{x}) = u(\mathrm{A}\mathbf{x})$, we obviously have $T(\mathrm{A}u) = \mathrm{A}T(u)$. Since, in addition, T commutes with translations, it indeed commutes with all affine maps of the plane. By abuse, we shall sometimes say that $I\!B$ is affine invariant. An obvious example of affine invariant set $I\!B$ is the family of ellipses with area 1 centered at 0. More generally, we can consider the set $\{\mathrm{A}B, \mathrm{A} \in SL(I\!R^2)\}$ where B is an arbitrary bounded set of the plane.

In what follows, we write as usual $\mathbf{x}, \mathbf{y}, \mathbf{z}$ for elements of $I\!R^2$ and $\mathbf{x} = (x, y)$. We consider a scaling parameter $h > 0$, and we set $I\!B_h = \{hB, B \in I\!B\}$.

We set, for every real function $u(\mathbf{x})$ defined on the plane,

$$SI_h u(\mathbf{x}) = \sup_{B \in I\!B_h} \inf_{\mathbf{y} \in \mathbf{x}+B} u(\mathbf{y}),$$

$$IS_h u(\mathbf{x}) = \inf_{B \in I\!B_h} \sup_{\mathbf{y} \in \mathbf{x}+B} u(\mathbf{y}).$$

$SI_h u$ is understood as an "affine erosion" of u and $IS_h u$ as an "affine dilation". Since $SI_h u = -IS_h(-u)$, we choose to study in the following one of these operators, namely IS_h. All theorems will be trivially adapted to SI_h. Our main concern is the behavior of $IS_h u(x)$ when $u(x)$ is a smooth (C^3) function and $h \to 0$. We prove in this chapter that if $I\!B$ is affine invariant and 1-localizable (as shown in the preceding chapter, affine invariance entails easily the the 1- localizability),

$$\lim_{h \to 0} \frac{(IS_h SI_h u)(\mathbf{x}_0) - u(\mathbf{x}_0)}{h^{\frac{4}{3}}} = c_{I\!B} |Du| \mathrm{curv}(u)^{\frac{1}{3}},$$

where $c_{I\!B}$ is a suitable constant and $r^{\frac{1}{3}}$ is meant for $\frac{r}{|r|}|r|^{\frac{1}{3}}$. This main result extends to affine invariant Matheron operators the asymptotic analysis performed in the preceding chapter for euclidean invariant local operators.

Lemma 14.8 shows that if $I\!B$ is localizable, then the behaviour of a C^3 function is very close to the behaviour of its Taylor expansion of order 2. If u is C^2, let us write

$$u(\mathbf{x}) = u(x,y) = u(0,0) + ax^2 + by^2 + cxy + O(|\mathbf{x}|^3) = v(\mathbf{x}) + O(|\mathbf{x}|^3).$$

If $I\!B$ is 1-localizable, then by Lemma14.8 we have

$$|IS_h u(0) - IS_h v(0)| \leq (C + Kc)h^{\frac{3}{2}}.$$

Thus we have to compute the action of IS_h on polynomials of degree 2. We shall consider in the series of lemmas to follow more and more general such polynomials. In the following, because of the translation invariance of the operators IS_h, we can assume without loss of generality for the asymptotic analysis of $IS_h u(\mathbf{x})$ that $\mathbf{x} = 0$ and $u(0) = 0$. Thus, we take "$IS_h(x + ax^2 + by^2 + cxy)$" as an abbreviation for "$IS_h(x + ax^2 + by^2 + cxy)(0)$". As we shall see, the cases $a = c = 0, b = +1$ and $b = -1$ play a canonical role, so that we set

$$c^+_{I\!B} = \inf_{B \in I\!B} \sup_{\mathbf{y} \in B} (x + y^2), \quad c^-_{I\!B} = \inf_{B \in I\!B} \sup_{\mathbf{y} \in B} (x - y^2).$$

Since our main results involve these constants, it is worth noticing that they can be different from zero.

Lemma 15.1 *(i) If $I\!B$ is a set of convex sets, invariant by $SL^2(I\!R^2)$ (or a set of boundaries of convex sets), all with area 1 and symmetric with respect to 0, then $c^+_{I\!B} > 0$ and $c^-_{I\!B} = 0$.*
(ii) If $I\!B_{aff}$ is the set of all affine structuring elements defined in Definition 13.7, then $c^+_{I\!B} \geq 1$, $c^-_{I\!B} = 0$.

Proof. Proof of (i). Consider the disk $D = D(0, \frac{1}{\sqrt{\pi}})$, with area 1. All elements B in $I\!B$ have area 1 and therefore meet the boundary ∂D of D. Thus, taking into account that $x \geq x^2$ for $\mathbf{x} = (x,y) \in D$,

$$c^+_{I\!B} \geq \inf_{B \in I\!B} \sup_{(x,y) \in B \cap D} (x + y^2) \geq \inf_{B \in I\!B} \sup_{(x,y) \in B \cap D} (x^2 + y^2) \geq \frac{1}{\pi}.$$

Indeed, since the sets B are symmetric with respect to 0, $\sup_{B \cap D}(x + y^2)$ is always attained for some (x,y) with $x \geq 0$. In order to prove that $c^-_{I\!B} = 0$, we first remark that since every $B \in I\!B$ surrounds 0, it contains at least one point where $x > 0$ and $y = 0$. Thus $c^-_{I\!B} \geq 0$. We then fix a set B in $I\!B$ and consider the new sets obtained by squeezing B onto the line $x = 0$:

$$B_\varepsilon = \{(\varepsilon x, \frac{y}{\varepsilon}) \quad , (x,y) \in B\}.$$

Then $B_\varepsilon \in I\!B$ and therefore

$$c^-_{I\!B} \leq \sup_{B_\varepsilon}(x - y^2) \leq \sup_{B_\varepsilon}(x) \leq C\varepsilon.$$

Thus, $c^-_{I\!B} = 0$.

Proof of (ii). Let B be an affine structuring element. Then its intersection with every half plane whose

boundary contains 0 has measure larger than 1. Let us choose as half plane $H = \{(x,y),\ x \geq 0.\}$ We deduce that

$$(\sup_{B\cap H} x)(\sup_{B\cap H} |y|) \geq 1.$$

Let us estimate

$$c_{I\!B} \geq \sup_B (x+y^2) \geq \sup_{B\cap H} (x+y^2) \geq \inf(\sup_{B\cap H} x, \sup_{B\cap H} y^2).$$

Setting $\xi = \sup_{B\cap H} x$, we deduce that

$$c_{I\!B} \geq \inf_{\xi\geq 0} (\xi, \frac{1}{\xi^2}) = 1.$$

□

Exercise 15.1 *As shown in this exercice, on can have $c_{I\!B} < 0$ for an even simple affine invariant family of structuring elements.*
1) Let $I\!B = \{AC, A \in SL(I\!R^2)\}$, where C is a truncated square with center 0, half side 2 and one of the sides missing, (e.g. $(x,y) \in C$ if $x = -2$ and $-2 \leq y \leq 2$ or $-2 \leq x \leq 2$ and $y = +2$ or $y = -2$) Show that $c^+_{I\!B} > 0$ and $c^-_{I\!B} < 0$.
2) Show that this family of structuring elements is localizable.

There is one case of simple function, which turns out to be canonic, where we can compute by invariance argument the action of an affine invariant operator IS_h.

Lemma 15.2 *Let $I\!B$ be a set of structuring elements invariant by the special linear group and let IS_h its associated* inf sup *operator. Let $p > 0$. Then*

$$IS_h(px+by^2) = c^+_{I\!B}(\frac{b}{p})^{\frac{1}{3}} p h^{\frac{4}{3}} \quad \text{if} \quad b > 0,$$

$$IS_h(px+by^2) = c^-_{I\!B}(-\frac{b}{p})^{\frac{1}{3}} p h^{\frac{4}{3}} \quad \text{if} \quad b \leq 0.$$

Proof. We note that if $b \neq 0$,

$$B \in I\!B \Leftrightarrow \begin{pmatrix} h^{\frac{4}{3}}|b|^{\frac{1}{3}} & 0 \\ 0 & h^{\frac{2}{3}}|b|^{-\frac{1}{3}} \end{pmatrix} B \in I\!B_h$$

Thus

$$\inf_{B\in I\!B_h} \sup_{(x,y)\in B} (x+by^2) = \inf_{B\in I\!B} \sup_{(x,y)\in B} (|b|^{\frac{1}{3}} h^{\frac{4}{3}} x + b(|b|^{-\frac{2}{3}} h^{\frac{4}{3}} y^2))$$

$$= |b|^{\frac{1}{3}} h^{\frac{4}{3}} \inf_{B\in I\!B} \sup_{(x,y)\in B} (x + \frac{b}{|b|} y^2)$$

$$= c^+_{I\!B} b^{\frac{1}{3}} h^{\frac{4}{3}} \quad \text{if} \quad b > 0,$$

$$= c^-_{I\!B} (-b)^{\frac{1}{3}} h^{\frac{4}{3}} \quad \text{if} \quad b < 0.$$

Since $p > 0$, we deduce that

$$IS_h(px+by^2) = pIS_h(x + \frac{b}{p} y^2)$$

$$= c^+_{I\!B}(\frac{b}{p})^{\frac{1}{3}} p h^{\frac{4}{3}} \quad \text{if} \quad b > 0,$$

$$= c^-_{I\!B}(-\frac{b}{p})^{\frac{1}{3}} p h^{\frac{4}{3}} \quad \text{if} \quad b < 0.$$

Let us now consider the case where $b = 0$. We fix a set B in $I\!B$ and denote by R its diameter. Then B is contained in $[-R, R] \times [-R, R]$ and therefore the set $\begin{pmatrix} h^2\epsilon & 0 \\ 0 & \epsilon^{-1} \end{pmatrix} B$, which belongs to $I\!B_h$, is contained in $[-R\epsilon h^2, R\epsilon h^2] \times [-\frac{R}{\epsilon}, \frac{R}{\epsilon}]$. Thus

$$0 \leq IS_h(px) \leq \sup_{[-R\epsilon h^2, R\epsilon h^2] \times [-\frac{R}{\epsilon}, \frac{R}{\epsilon}]} px \leq pR\epsilon h^2$$

We can take ϵ arbitrarily small and conclude that $IS_h(px) = 0$. □

The next lemma deals with the case where the function to which we apply IS_h has a zero gradient at zero.

Lemma 15.3 *Let $I\!B$ be an affine invariant set of structuring elements, one of which is bounded and all of which contain O. Then there is a continuous function $G(Du, D^2u)$ satisfying $G(0,0) = 0$ such that for every C^3 function u, the scaled operators IS_h and SI_h associated with $I\!B$ satisfy*

$$0 \leq |IS_h u(\mathbf{x}_0) - u(\mathbf{x}_0)| \leq G(D^2u(\mathbf{x}_0), Du(\mathbf{x}_0))h^{\frac{4}{3}} + o_{\mathbf{x}_0}(h^{\frac{4}{3}}),$$

where $o_{\mathbf{x}_0}(h^{\frac{4}{3}})$ is uniform on compact sets of $I\!R^2$. In the same way, by the relation $IS_h(u) = -SI_h(-u)$,

$$0 \geq |SI_h u(\mathbf{x}_0) - u(\mathbf{x}_0)| \geq -G(D^2u(\mathbf{x}_0), Du(\mathbf{x}_0))h^{\frac{4}{3}} + o_{\mathbf{x}_0}(h^{\frac{4}{3}}).$$

Proof. Since all elements of $I\!B$ contain 0, we always have $\sup_{\mathbf{y} \in \mathbf{X}+B} u(\mathbf{y}) \geq u(\mathbf{x})$. Let B be a bounded element of $I\!B$ and let $R = \sup_{\mathbf{X} \in B} |\mathbf{x}|$. Then B is contained in $[-R, R] \times [-R, R]$ and therefore the set $\begin{pmatrix} h^{\frac{4}{3}} & 0 \\ 0 & h^{\frac{2}{3}} \end{pmatrix} B$,

which belongs to $I\!B_h$, is contained in $[-Rh^{\frac{4}{3}}, Rh^{\frac{4}{3}}] \times [-Rh^{\frac{2}{3}}, Rh^{\frac{2}{3}}]$. By making a Taylor expansion of u in a neighborhood of $\mathbf{x}_0$, we have for the usual local coordinates (x, y),

$$u(\mathbf{x}) - u(\mathbf{x}_0) = px + ax^2 + by^2 + cxy + o_{\mathbf{x}_0}(|\mathbf{x} - \mathbf{x}_0|^3).$$

Thus

$$0 \leq IS_h u(\mathbf{x}_0) - u(\mathbf{x}_0) \leq \sup_{(x,y) \in [-Rh^{\frac{4}{3}}, Rh^{\frac{4}{3}}] \times [-Rh^{\frac{2}{3}}, Rh^{\frac{2}{3}}]} px + ax^2 + by^2 + cxy + O_{\mathbf{x}_0}(|\mathbf{x}|^3) \leq (|p| + |a| + |b| + |c|)R^2 h^{\frac{4}{3}} + o_{\mathbf{x}_0}(h^2).$$

We obtain

$$0 \leq IS_h u(\mathbf{x}_0) - u(\mathbf{x}_0) \leq h^{\frac{4}{3}}(||Du(\mathbf{x}_0)|| + ||D^2u(\mathbf{x}_0)||)R^2 + o_{\mathbf{x}_0}(h^{\frac{4}{3}}).$$

□

We are now in a position to state and prove the first main result of this chapter. We consider affine erosions and affine dilations.

Theorem 15.4 *Let $I\!B$ be a 1-localizable set of plane sets which is invariant by the special linear group $SL(I\!R^2)$. Assume $u(\mathbf{x})$ is a Lipschitz function and C^3 in a neighborhood of $\mathbf{x}_0$. Then*

$$\frac{(IS_h u)(\mathbf{x}_0) - u(\mathbf{x}_0)}{h^{4/3}} = |Du| g(\mathrm{curv}(u))(\mathbf{x}_0) + o_{\mathbf{x}}(h^{\frac{4}{3}}),$$

where $\o_{\mathbf{x}}(h^{\frac{4}{3}}) \leq o_K(h^{\frac{4}{3}})$ on every compact set K of $I\!R^2$ where $Du(\mathbf{x})$ does not vanish and

$$\begin{array}{rcll} \textit{where } g(r) & = & c^+_{I\!B}(\frac{r}{2})^{\frac{1}{3}} & \textit{if} \quad r > 0 \\ & = & c^-_{I\!B}(-\frac{r}{2})^{\frac{1}{3}} & \textit{if} \quad r < 0. \end{array}$$

Proof of Theorem 15.4. SI_h being translation and isometry invariant, we can choose the origin 0 at $\mathbf{x}_0$, and orthogonal axes $(\vec{i}, \vec{j})$ so that $\vec{i} = \frac{Du}{|Du|}(0)$ if $Du \neq 0$. In addition,since $IS_h(u - u(0)) = IS_h u - u(0)$, we assume without loss of generality that $u(0) = 0$. With these conventions, since u is C^3, we can write again the relation (5.7) :

$$u(\mathbf{x}) = px + ax^2 + by^2 + cxy + O(|\mathbf{x}|^3), \tag{15.1}$$

where $p = |Du|(0) \geq 0$, $\mathbf{x} = (x, y) = x\vec{i} + y\vec{j}$ and, if $p > 0$,

$$\begin{aligned} b &= \tfrac{1}{2}\tfrac{\partial^2 u}{\partial y^2}(0) = \tfrac{1}{2}D^2u(\vec{j}, \vec{j}) = \tfrac{1}{2}|Du|\mathrm{curv}(u)(0) \\ a &= \tfrac{1}{2}\tfrac{\partial^2 u}{\partial x^2}(0) = \tfrac{1}{2}D^2u(\vec{i}, \vec{i}) \\ c &= \tfrac{\partial^2 u}{\partial x \partial y}(0) = D^2u(\vec{j}, \vec{i}). \end{aligned} \tag{15.2}$$

In the following, we consider a radius $r = h^{\frac{1}{2}}$ and the truncated operator IS^r_h already considered in Localization Lemma 14.9. From (15.1) we deduce that for every $\mathbf{x}$ in $D(0, r)$,

$$px + ax^2 + by^2 + cxy - O(r^3) \leq u(\mathbf{x}) \leq px + ax^2 + by^2 + cxy + O(r^3).$$

Therefore,

$$(IS^r_h u)(0) = IS^r_h(px + ax^2 + by^2 + cxy) + O(r^3) \tag{15.3}$$

(Recall that, by an anterior convention, $IS^r_h(px + ax^2 + by^2 + cxy)$ means $IS^r_h(px + ax^2 + by^2 + cxy)(0)$.)

• **The case $p = |Du(0)| = 0$.**

In this case we have, by the preceding lemma 15.3, $IS_h(ax^2+by^2+cxy) = O(h^2)$ and therefore $h^{-\frac{4}{3}}(IS_h u)(0) = O(h^{\frac{2}{3}})$, which proves the assertion in the case $|Du(x_0)| = 0$.

• **The case $p = |Du(0)| \neq 0$.**

Setting $\varepsilon = h^\theta$, where $\theta > 0$ is a very small real number, we have

$$-|c|\varepsilon y^2 - \frac{|c|}{\varepsilon}x^2 \leq cxy \leq |c|\varepsilon y^2 + \frac{|c|}{\varepsilon}x^2.$$

Thus

$$(IS^r_h)u(0) \leq IS^r_h(px + (a + |c|\varepsilon^{-1})x^2 + (b + \varepsilon|c|)y^2) + O(r^3), \tag{15.4}$$

$$(IS^r_h)u(0) \geq IS^r_h(px + (a - |c|\varepsilon^{-1})x^2 + (b - \varepsilon|c|)y^2) + O(r^3) \tag{15.5}$$

We are led to study expressions of the type

$$IS_h^r(px + a(\varepsilon)x^2 + b(\varepsilon)y^2),$$

where $a(\varepsilon) = O(h^{-\theta})$ and $b(\varepsilon) = O(1)$. (We shall only examine the case where $b \neq 0$, the case $b = 0$ leading to the same calculations and conclusion.) We now use the contrast invariance property of T : $g \circ T = T \circ g$, where we fix $g(s) = s - \frac{a}{p^2}s^2$. This function is increasing in a neighborhood of zero. Thus, since $IS_h^r(px + ax^2 + by^2)$ tends to zero as r and h tend to zero, using $g \circ T = T \circ g$,

$$IS_h^r(px + ax^2 + by^2) =$$

$$g^{-1}(IS_h^r(px + by^2 - \frac{a}{p^2}((ax^2 + by^2)^2 + 2px(ax^2 + by^2))))$$

We then use the following relations (Lemma 15.2 and Localization Lemma 14.9 (iii))

$$g^{-1}(t) = t + O(\varepsilon^{-1}t^2), \quad IS_h(px + by^2) = O(h^{\frac{4}{3}}),$$

$$IS_h^r(px + by^2) = IS_h(px + by^2) + O(h^{\frac{3}{2}}) = O(h^{\frac{4}{3}}),$$

$$a(\varepsilon) = O(\varepsilon^{-1}), \quad b(\varepsilon) = O(1), \quad r = h^{\frac{1}{2}}$$

to deduce that

$$IS_h^r(px + ax^2 + by^2) = g^{-1}(IS_h^r(px + by^2) + O(\varepsilon^{-2}h^{\frac{3}{2}})) = g^{-1}(IS_h(px + by^2) + O(\varepsilon^{-2}h^{\frac{3}{2}}))$$

$$= IS_h(px + by^2) + O(\varepsilon^{-2}h^{\frac{3}{2}} + \varepsilon^{-1}h^{\frac{8}{3}}),$$

and finally

$$IS_h^r(px + ax^2 + by^2) = IS_h(px + by^2) + O(\varepsilon^{-2}h^{\frac{3}{2}}).$$

Using Relations 15.4 and 15.5 and Localization Lemma 14.9, we obtain

$$IS_h(px + b_{-\varepsilon}y^2) - O(h^{\frac{3}{2}}\varepsilon^{-2}) \leq IS_h(u)(0) \leq IS_h(px + b_\varepsilon y^2) + O(h^{\frac{3}{2}}\varepsilon^{-2}),$$

where $b_\varepsilon = (b + |c|\varepsilon)$ and $b_{-\varepsilon} = (b - |c|\varepsilon)$. Using $\varepsilon = h^\theta$, Lemma 15.2 and the inequalities

$$(\alpha + \beta)^{\frac{1}{3}} \leq \alpha^{\frac{1}{3}} + \beta^{\frac{1}{3}}, \quad \alpha^{\frac{1}{3}} - \beta^{\frac{1}{3}} \leq (\alpha - \beta)^{\frac{1}{3}}$$

for $\alpha, \beta \geq 0$, we get

$$IS_h(u)(0) = c_{I\!B}^+(\frac{b}{p})^{\frac{1}{3}}ph^{\frac{4}{3}} + O(h^{\frac{3}{2}} - 2\theta) + O(h^{\frac{4}{3}+\frac{\theta}{3}}) \quad \text{if} \quad b > 0,$$

$$= c_{I\!B}^-(-\frac{b}{p})^{\frac{1}{3}}ph^{\frac{4}{3}} + O(h^{\frac{3}{2}-2\theta}) + O(h^{\frac{4}{3}+\frac{\theta}{3}}) \quad \text{if} \quad b < 0.$$

In order to conclude the proof, we simply replace p by $|Du|$ and b by $\frac{1}{2}|Du|\text{curv}(u)$ in the above relations and choose (e.g.) θ such that $\frac{3}{2} - 2\theta = \frac{4}{3} + \frac{\theta}{3}$, i.e. $\theta = \frac{1}{14}$. We finally argue that all of the asymptotic behaviours involved in the proof are uniform in a, b, c, p, provided p does not approach 0 and a, b, c remain bounded. Thus, we can write

$$\frac{(IS_h u)(\mathbf{x}_0) - u(\mathbf{x}_0)}{h^{4/3}} = |Du|g(\text{curv}(u))(\mathbf{x}_0) + o_{\mathbf{X}}(h^{\frac{4}{3}}),$$

where $\o_{\mathbf{X}}(h^{\frac{4}{3}}) \leq o_K(h^{\frac{4}{3}})$ on every compact set K of $I\!R^2$ where $Du(\mathbf{x})$ does not vanish. Indeed, on such a compact set, Du, D^2u and D^3u remain bounded and so does $\frac{1}{p} = \frac{1}{|Du(\mathbf{X})|}$, which is present in the estimates of the proof. □

15.1 Alternate schemes

In this section, we extend the preceding results to "alternate schemes", that is, products of the kind IS_hSI_h. We shall obtain for such alternate schemes a convergence theorem extending Theorem 15.4. The alternate schemes are easier to implement and numerically more efficient if we want to get the following natural property for a contrast invariant operator T : $T(-u) = -T(u)$. Recall that, precisely because this property is not satisfied by erosion and dilation operators, it has been proposed with some experimental success to build alternate operators like $T = IS_hSI_h$ (those alternate operators could be called "openings" or "closings"). It is, in that case, not true that $T(-u) = -T(u)$. We shall prove, however, that if we let $h \to 0$ and consider iterates $(IS_hSI_h)^n$ with $n \to \infty$, then the limit operators do satisfy this property. It is not possible to obtain directly an asymptotic result for IS_hSI_h by applying twice Theorem 15.4. Indeed, there is no guarantee (and it is in general false) that IS_hu is C^3 whenever u is C^3. The next lemma explains how we can, however, extend to alternate operators convergence results like the ones given in Theorem 15.4.

Lemma 15.5 *Let T_h and S_h be two* inf sup *operators. Let $F_1(A, p, \mathbf{x})$ and $F_2(A, p, \mathbf{x})$ be two functions which are continuous for $(A, p, \mathbf{x})$ in every compact set on which $p \neq 0$. Assume that for a suitable exponent β we have*

$$T_hu(\mathbf{x}) - u(\mathbf{x}) = h^\beta F_1(D^2u, Du, \mathbf{x}) + o_{\mathbf{X}}(h^\beta),$$
$$S_hu(\mathbf{x}) - u(\mathbf{x}) = h^\beta F_2(D^2u, Du, \mathbf{x}) + o_{\mathbf{X}}(h^\beta).$$

Assume that S_h is localisable, i.e. for a suitable exponent γ,

$$|S_hu(\mathbf{x}) - S_hu^{h^\gamma}u(\mathbf{x})| \leq o_{\mathbf{X}}(h^\beta)$$

for every Lipschitz function u. In these assumptions, $o_{\mathbf{X}}(h^\beta)$ denotes any function of $\mathbf{x}$ and h^β such that $o_{\mathbf{X}}(h^\beta) \leq o_K(h^\beta)$ for $\mathbf{x}$ in a compact set K on which $Du(\mathbf{x}) \neq 0$. Then the alternate operator satisfies

$$T_hS_hu(\mathbf{x}) - u(\mathbf{x}) = h^\beta(F_1(D^2u, Du, \mathbf{x}) + F_2(D^2u, Du, \mathbf{x})) + o_{\mathbf{X}}(h^\beta)$$

and $o_{\mathbf{X}}(h^\beta) \leq o_K(h^\beta)$ on every compact set K on which $Du(\mathbf{x}) \neq 0$.

Proof Fix a compact K on which $Du \neq 0$. Let $\mathbf{x}_0 \in K$ and $\overline{B}(\mathbf{x}_0, r)$ a ball with r small enough to ensure that we still have $Du(\mathbf{x}) \neq 0$ on $\overline{B}(\mathbf{x}_0, r)$. Let $f^+(r) = \sup_{\mathbf{X}\in B(\mathbf{x}_0,r)} |F_1(D^2u, Du, \mathbf{x})|$ and $f^-(r) = \inf_{\mathbf{X}\in B(\mathbf{x}_0,r)} |F_1(D^2u, Du, \mathbf{x})|$. We have

$$T_hu(\mathbf{x}) - u(\mathbf{x}) = h^\beta F_1(D^2u, Du, \mathbf{x}) + o_{\mathbf{X}}(h^\beta).$$

Let $o_{\mathbf{X}_0,r}(h) = \sup_{\mathbf{X}\in\overline{B}(\mathbf{x}_0,r)} o_{\mathbf{X}}(h^\beta)$. Thus, for $\mathbf{x} \in \overline{B}(\mathbf{x}_0, r)$,

$$u(\mathbf{x}) + h^\beta f^-(r) - o_{\mathbf{x}_0,r}(h^\beta) \leq T_hu(\mathbf{x}) \leq u(\mathbf{x}) + h^\beta f^+(r) + o_{\mathbf{x}_0,r}(h^\beta).$$

By applying $S_h^{h^\gamma}$ at $\mathbf{x}_0$ to the members of this inequality, using the comparison principle, provided we set $h^\gamma = r$, and the commutation of S_h with the addition of constants :

$$S_h^{h^\gamma}u(\mathbf{x}) + h^\beta f^-(r) - o_{\mathbf{x}_0,r}(h^\beta) \leq S_h^{h^\gamma}T_hu(\mathbf{x}) \leq S_h^{h^\gamma}u(\mathbf{x}) + h^\beta f^+(r) + o_{\mathbf{x}_0,r}(h^\beta) \tag{15.6}$$

Now, F being continuous,

$$f^{\pm} = F_1(D^2u(\mathbf{x}_0), Du(\mathbf{x}_0), \mathbf{x}_0) + o(1) \tag{15.7}$$

and, from the assumptions,

$$S_h^{h^\gamma} u(\mathbf{x}_0) - u(\mathbf{x}_0) = h^\beta F_2(D^2u, Du, \mathbf{x}_0) + o_{\mathbf{X}_0}(h^\beta), \tag{15.8}$$

$$(S_h^{h^\gamma} T_h u)(\mathbf{x}_0) = S_h T_h u(\mathbf{x}_0) + o_{\mathbf{X}_0}(h^\beta). \tag{15.9}$$

¿From (15.6)- (15.9), we obtain

$$S_h T_h u(\mathbf{x}_0) = h^\beta (F_1(D^2u, Du, \mathbf{x}_0) + F_2(D^2u, Du, \mathbf{x}_0)) + o_{f\,xx_0}(h^\beta).$$

It is noticeable that we can fix r independtly of $\mathbf{x}_0$ on K, in such a way that $Du(\mathbf{x})$ does not vanish on the dilated compact K^r. Thus, all convergences in the preceding proof are in fact uniform for $\mathbf{x}_0$ on K and we can bound $o_{\mathbf{X}_0}(h^\beta)$ by some $o_K(h^\beta)$. □

By repeating word for word the above proof, we can also prove the following, which will prove of use.

Lemma 15.6 *Let T_h and S_h be two* inf sup *operators. Let $G_1(A, p, \mathbf{x})$ and $G_2(A, p, \mathbf{x})$ be two continuous functions. Assume that for suitable exponents γ and β we have*

$$0 \leq T_h u(\mathbf{x}) - u(\mathbf{x}) \leq h^\beta G_1(D^2u, Du, \mathbf{x}) + o_{\mathbf{X}}(h^\beta),$$

$$0 \geq S_h u(\mathbf{x}) - u(\mathbf{x}) \geq h^\beta G_2(D^2u, Du, \mathbf{x}) + o_{\mathbf{X}}(h^\beta).$$

Assume that S_h is localisable, i.e.

$$|S_h u(\mathbf{x}) - S_h u^{h^\gamma} u(\mathbf{x})| \leq o_{\mathbf{X}}(h^\beta)$$

for every Lipschitz function u. In these assumptions, $o_{\mathbf{X}}(h^\beta)$ denotes any function of $\mathbf{x}$ and h^β such that $o_{\mathbf{X}}(h^\beta) \leq o_K(h^\beta)$ for $\mathbf{x}$ in a compact set K. Then the alternate operator satisfies

$$h^\beta (G_1(D^2u, Du, \mathbf{x}) + o_{\mathbf{X}}(h^\beta) \leq T_h S_h u(\mathbf{x}) - u(\mathbf{x}) \leq h^\beta G_2(D^2u, Du, \mathbf{x})) + o_{\mathbf{X}}(h^\beta)$$

and $o_{\mathbf{X}}(h^\beta) \leq o_K(h^\beta)$ on every compact set K.

We are now in a position to state an asymptotic behaviour theorem for the main affine invariant alternate filter, obtained by alternating affine erosions and dilations.

Theorem 15.7 *Let $\mathbb{B}_{aff}$ the set of "affine structuring elements" and respectively IS_h and SI_h the associated affine erosions and dilations as defined in Definition ??. Then for every Lipschitz function u which is C^3 in a neighborhood of $\mathbf{x}_0$, we have*

$$\lim_{h \to 0} \frac{(T_h u)(\mathbf{x}_0) - u(\mathbf{x}_0)}{h^{4/3}} = c_{\mathbb{B}} |Du| (\mathrm{curv}(u))(\mathbf{x}_0) \tag{15.10}$$

$$\begin{array}{rcll} \textit{where } g(r) & = & (r^-)^{\frac{1}{3}} & \textit{if} \quad T_h = SI_h \\ & = & (r^+)^{\frac{1}{3}} & \textit{if} \quad T_h = IS_h \\ & = & (r)^{\frac{1}{3}} & \textit{if} \quad T_h = SI_h\ IS_h \end{array}$$

This convergence is uniform on all compact sets of $\mathbb{R}^2$ *on which* $Du(\mathbf{x})$ *does not vanish. In addition, there are continuous functions* $G^{\pm}(D^2u, Du)$ *such that* $G^{\pm}(0,0) = 0$ *and*

$$h^{\frac{4}{3}}G^{-}(D^2u, Du) + o_{\mathbf{X}}(h^{\frac{4}{3}}) \leq T_h u(\mathbf{x}) - u(\mathbf{x}) \leq h^{\frac{4}{3}}G_2(D^2u, Du)) + o_{\mathbf{X}}(h^{\frac{4}{3}}) \tag{15.11}$$

and $o_{\mathbf{X}}(h^{\frac{4}{3}}) \leq o_K(h^{\frac{4}{3}})$ *on every compact set* K *of* $\mathbb{R}^2$.

Proof By Theorem 15.4, we know that (15.10) holds for $T_h = IS_h$ or $T_h = SI_h$. Indeed, $\mathbb{B}_{aff}$ is a 1-localizable affine invariant family of structuring elements, as requested by Theorem 15.4. By Lemma 15.1, we know in addition that $c^{+}_{\mathbb{B}_{aff}} \geq 1$ and $c_{\mathbb{B}_{aff}} = 0$. We then apply Lemma 15.5, with obviously $\beta = \frac{4}{3}$ and $\gamma = \frac{1}{2}$, to obtain (15.10) for the alternate operator.

In order to prove (15.11), we simply apply Lemma **??** to IS_h and SI_h, which holds because $\mathbb{B}_{aff}$ is affine invariant and all of its elements contain 0 and some are bounded. We apply in continuation Lemma 15.6 to extend (15.11) to the alternate operator IS_hSI_h. □

By way of principle, all affine invariant families of connected sets which contain all 0, some of which are bounded, and whose area is bounded from above, are localizable and permit to define alternate schemes with the same properties as above. The next exercise examines some examples. Now, we have seen that affine invariant dilations and erosions have magnificent numerical and structural properties and are the natural best candidates for affine invariant shape analysis.

Exercise 15.2 *Check that the preceding proof and results apply to the affine invariant localizable families considered in Exercice 14.1.*

References.

Early versions of the results and proofs contained in this chapter are given in [191, 187, 190]. Our version here is easier and shorter. Alternate sequential filters are well introduced in Serra [379]. A general axiomatic theory of Self-dual morphological operators is given by Heijmans [200]. Applications of alternating sequential filters are discussed in [338].

Chapter 16

Monotone image operators: "nonflat" morphology

16.1 General form of monotone operator.

Theorem 16.1 *Let T be a monotone function operator defined of $\mathcal{F}$, invariant by translation and commuting with the addition of constant. There exists a family $I\!F$ of functions of $\mathcal{F}$ such that*

$$Tu(\mathbf{x}) = \sup_{f \in \mathcal{F}} \inf_{\mathbf{y} \in I\!R} u(\mathbf{y}) - f(\mathbf{x} - \mathbf{y})$$

Proof We choose $I\!F = \{f \in \mathcal{F}, Tf(0) \geq 0\}$ Then,

$$Tu(\mathbf{x}) \geq \lambda \Leftrightarrow \forall \epsilon > 0,\ Tu(\mathbf{x}) \geq \lambda - \epsilon$$

$$\Leftrightarrow \forall \epsilon > 0,\ \tau_{-\mathbf{x}}(T(u - \lambda + \epsilon))(0) \geq 0$$

$$\Leftrightarrow \forall \epsilon > 0,\ T(\tau_{-\mathbf{x}}(u - \lambda + \epsilon))(0) \geq 0$$

$$\Leftrightarrow \forall \epsilon > 0,\ \tau_{-\mathbf{x}}(u - \lambda + \epsilon) \in I\!F$$

$$\Leftrightarrow \forall \epsilon > 0, \exists v \in I\!F,\ \inf_{\mathbf{y}} u(\mathbf{y}) - \lambda + \epsilon - v(\mathbf{y} - \mathbf{x}) \geq 0$$

($\Rightarrow$ is true by simply choosing $v = u - \lambda + \epsilon$. The converse implication is true due to the monotony of the operator T and definition of $I\!F$ which imply that if $u \geq v$ and $v \in I\!F$ then $u \in I\!F$.)

$$\Leftrightarrow \forall \epsilon > 0,\ \sup_{v \in I\!F} \inf_{\mathbf{y}} u(\mathbf{y}) - \lambda + \epsilon - v(\mathbf{y} - \mathbf{x}) \geq 0$$

$$\Leftrightarrow \sup_{v \in I\!F} \inf_{\mathbf{y}} u(\mathbf{y}) - v(\mathbf{y} - \mathbf{x}) \geq \lambda$$

□

16.2 Asymptotic behavior of monotone operators

The aim of this section is to study the asymptotic behavior of a monotone operator. More precisely we assume to have a base of functions $\mathbb{F}$ and an operator T defined by

$$T(u)(\mathbf{x}) = \inf_{f \in \mathbb{F}} \sup_{\mathbf{y} \in I\!R^N} u(\mathbf{y}+\mathbf{x}) - f(\mathbf{y}).$$

We want first to define a local version of it T_h and then to estimate $T_h(u) - u$ when h tends to 0.

16.2.1 The rescaling issue

As we have seen until now, the scale is related to the space by the following consideration: assume that u and v are two functions such that $v(\mathbf{x}) = u(2\mathbf{x})$. ($u$ corresponds somehow to a zoom of v). If we want to smooth the two images similarly we have to change the scale of the filter. For contrast invariant filter, this is quite straightforward, the scale is directly and uniquely linked to the size of the structuring elements. E.g. if the filter is the median filter on a disk. The size of the disk (the scale) has to be chosen two times bigger for u than for v. For such filters, the down-scaling corresponds to a spatial shrinkage of the structuring elements.

For linear filter, (think the mean value to be simpler) the scaling was also straightforward. Indeed, the mean value on u has to performed on a neighborhood two times larger than for v. But in that case, this does not only mean a spatial shrinkage ! Indeed the kernel of the mean value on a disk of radius h centered in 0 is given by

$$\begin{array}{ll} g_h(\mathbf{x}) & = \frac{1}{\pi h^2} \quad \text{for } |\mathbf{x}| \leq h \\ & = 0 \quad \text{otherwise} \end{array}$$

That is that the structuring element is scaled also in amplitude. Here the amplitude-scaling factor h^{-2} is so that $\int_{I\!R} g_h = 1$ which was a assumption made for a linear smoothing.

As for the linear filter, at this point we can guess that an amplitude-scaling factor might be needed for a general monotone filter. So that the structuring elements, that is the functions of $\mathbb{F}$ will be scaled as $f(\mathbf{x}) \to h^\beta f(\mathbf{x})$, where β is a real number which will be discussed later. (To be noted that is all that follow h^β could be replace by a function of β).

We therefore define the scaled operator T_h associated to T by

$$T_h(u)(\mathbf{x}) = \inf_{f \in \mathbb{F}} \sup_{\mathbf{y} \in I\!R^N} u(\mathbf{x}+\mathbf{y}) - h^\beta f(\mathbf{y}/h). \tag{16.1}$$

16.2.2 Legendre Fenchel transform

Definition 16.2 *Let f be a function from $I\!R^N$ into $\bar{I\!R}$, we denote the Legendre conjugate of f by f^* : $I\!R^N \to \bar{I\!R}$ defined by*

$$f^*(p) = \sup_{\mathbf{x} \in I\!R} (p.\mathbf{x} - f(\mathbf{x}))$$

Let us note that if f is convex then the legendre transform is finite for every p.

16.2.3 Asymptotic theorem, first order case

Lemma 16.3 *Let f be a function satisfying the following conditions:*

$$\exists C > 0 \text{ and } \alpha > \max(\beta, 1) \text{ such that } \liminf_{|\mathbf{x}|\to\infty} \frac{f(\mathbf{x})}{|\mathbf{x}|^\alpha} \geq C \text{ and } f(0) \leq 0 \tag{16.2}$$

Then, for any C^1 and bounded function u, if $\beta < 2$:

$$\sup_{\mathbf{y}\in I\!R^N} (u(\mathbf{x}+\mathbf{y}) - h^\beta f(\mathbf{y}/h)) - u(\mathbf{x}) = h^\beta f^*(h^{1-\beta} Du(\mathbf{x})) + O(h^{2(1-\frac{\beta-1}{\alpha-1})})$$

A interesting particular case is when $\beta = 1$:

$$\sup_{\mathbf{y}\in I\!R^N} (u(\mathbf{x}+\mathbf{y}) - hf(\mathbf{y}/h)) - u(\mathbf{x}) = hf^*(Du(\mathbf{x})) + O(h^2)$$

Proof Without loss of generality we can choose $\mathbf{x} = 0$ and $u(\mathbf{x}) = 0$ so that we are looking for an estimate of

$$\sup_{\mathbf{z}\in I\!R^N} (u(\mathbf{z}) - h^\beta f(\mathbf{z}/h))$$

when h tends to 0. Setting $\mathbf{y} = \mathbf{z}/h$, we have,

$$\sup_{\mathbf{z}\in I\!R^N} (u(\mathbf{z}) - h^\beta f(\mathbf{z}/h)) = \sup_{\mathbf{y}\in I\!R^N} (u(h\mathbf{y}) - h^\beta f(\mathbf{y}))$$

Let us first prove that we can discard from the preceding sup the $\mathbf{y}$ that goes too fast toward ∞ as h tends to 0. We consider the subset S_h of $I\!R^N$ of the $\mathbf{y}$ such that

$$u(h\mathbf{y}) - h^\beta f(\mathbf{y}) \geq u(0) - h^\beta f(0) \geq 0.$$

We obviously have

$$\sup_{\mathbf{y}\in I\!R^N} (u(h\mathbf{y}) - h^\beta f(\mathbf{y})) = \sup_{\mathbf{y}\in S_h} (u(h\mathbf{y}) - h^\beta f(\mathbf{y})).$$

Since u is bounded, we have $\forall \mathbf{y} \in S_h$, $f(\mathbf{y}) \leq C_1 h^{-\beta}$ for some constant C_1 depending only on $||u||_\infty$. Assume that there exists $\mathbf{y}_h \in S_h$ tending to ∞ as h tends to zero. For h small enough, condition (16.2) gives $f(\mathbf{y}_h) \geq C|\mathbf{y}_h|^\alpha$, which combined with the preceding inequality yields $|\mathbf{y}_h| \leq C_2 h^{-\beta/\alpha}$. Such a bound holds if $\mathbf{y}_h \in S_h$ is bounded, so that we have

$$\forall y \in S_h, |\mathbf{y}| \leq C_2 h^{-\beta/\alpha}$$

As consequence, $\forall y \in S_h$ we have $|h\mathbf{y}| = o(1)$ and we can do an expansion of u around 0, so that

$$\sup_{\mathbf{y}\in I\!R^N} (u(h\mathbf{y}) - h^\beta f(\mathbf{y})) = \sup_{\mathbf{y}\in S_h} (hDu(0).\mathbf{y} - h^\beta f(\mathbf{y}) + O(h^2|\mathbf{y}|^2))$$

We can now find finer bound for the set S_h repeating the same argument. $\forall \mathbf{y} \in S_h$ we have,

$$hp.\mathbf{y} - h^\beta f(\mathbf{y}) + O(h^2\mathbf{y}^2) \geq 0$$

which yields

$$|p| \geq h^{\beta-1} f(\mathbf{y})/|\mathbf{y}| + O(h|\mathbf{y}|)$$

Assume that $\mathbf{y}_h \in S_h$, satisfying the preceding inequation, tends to ∞ when h tends to 0, then by (16.2), we obtain $|\mathbf{y}_h| = O(h^{-\frac{\beta-1}{\alpha-1}})$. Once again, if $\mathbf{y}_h$ is bounded this estimate holds. So we have

$$\sup_{\mathbf{y}\in I\!R^N}(u(h\mathbf{y}) - h^\beta f(\mathbf{y})) = h^\beta(\sup_{\mathbf{y}\in S_h}(h^{1-\beta}p.\mathbf{y} - f(\mathbf{y}) + O(h^{2(1-\frac{\beta-1}{\alpha-1})-\beta})))$$

$$= h^\beta(\sup_{\mathbf{y}\in I\!R^N}(h^{1-\beta}p.\mathbf{y} - f(\mathbf{y})) + O(h^{2(1-\frac{\beta-1}{\alpha-1})})) = h^\beta(f^*(h^{1-\beta}p)) + O(h^{2(1-\frac{\beta-1}{\alpha-1})})$$

It is easily checked that $O(h^{2(1-\frac{\beta-1}{\alpha-1})}) = o(h^\beta)$ for all $\beta < 2$. □

Theorem 16.4 *Let $I\!F$ be a family of functions, all satisfying the condition (16.2) with a constant C non dependant on the choice of a function within the family. Let T_h be the rescaled operator associated with the family $I\!F$ and with a rescaling parameter β equal to 1. Then for all C^1 and bounded function u we have:*

$$\frac{(T_h(u) - u)(\mathbf{x})}{h} = H_1(Du(\mathbf{x})) + o(1)$$

where

$$H_1(p) = \inf_{f\in I\!F} f^*(p)$$

16.2.4 Second order case - some heuristics.

Theorem 16.4 gives the first order possible behavior of a non-flat monotone operator. Question occurs on what happens if this first order term is 0, that is if $H_1(p) = 0$ for all p. In that case, it is necessary to push the expansion to the second order:

We have with $p = Du(0)$ and $A = D^2u(0)/2$,

$$\sup_{\mathbf{y}\in I\!R^N} u(h\mathbf{y}) - h^\beta f(\mathbf{y}) = \sup_{\mathbf{y}\in I\!R^N} hp.\mathbf{y} + h^2 A\mathbf{y}.\mathbf{y} - h^\beta f(\mathbf{y}) + O(|h\mathbf{y}|^3)$$

Since this last expression is increasing with respect to A it is then expected that the left side of the equality converges when h tends to 0, to some function $F(A, p)$ where F is non decreasing with respect to A. As consequence, among second order operator only elliptic operator can be obtained as the asymptotical limit of a general monotone operator.

16.3 Application to image enhancement: Kramer's operators and the Rudin-Osher shock filter

In [256], Kramer defines a filter for sharpening blurred images. The filter replaces the gray level value at a point by either the minimum or the maximum of the gray level values in a neighborhood. This choice depending on which is the closiest to the current value.

In [?], Rudin and Osher proposes to shapen blurred images by applying the following equation:

$$\frac{\partial u}{\partial t} = sgn(\Delta u)|Du|$$

As, we will see in the following section, this two filters are asymptotically the same in 1D, but differs in 2D. The first one yields to the Canny differential operator for edge detection (sign of $D^u(Du, Du)$), while the second explicitely uses the sign of the laplacian.

16.3.1 The Kramer operator.

This filter can be seen as a conditional erosion or dilation and an easy link can be made with the "shock filters" [?]. A finer version of it, is proposed in [373] and proceed as follow: Let $q(\mathbf{x}) = \mathbf{x}^2/2$, and $I\!F^+ = \{q\}$. Set T_h^+ the rescaled, (with $\beta = 1$), non-flat operator associated with the structuring elements set $I\!F^+$ and T_h^- its dual operator. We have

$$(T_h^+ u)(\mathbf{x}) = \sup_{\mathbf{y}\in I\!R^N} u(\mathbf{y}) - hq((\mathbf{x}-\mathbf{y})/h) = \sup_{\mathbf{y}\in I\!R^N} u(\mathbf{y}) - \frac{(\mathbf{x}-\mathbf{y})^2}{2h}$$

$$(T_h^- u)(\mathbf{x}) = \inf_{\mathbf{y}\in I\!R^N} u(\mathbf{y}) - hq((\mathbf{x}-\mathbf{y})/h) = \inf_{\mathbf{y}\in I\!R^N} u(\mathbf{y}) + \frac{(\mathbf{x}-\mathbf{y})^2}{2h}$$

The Shock filter T_h is then defined by

$$(T_h u)(\mathbf{x}) = \begin{cases} (T_h^+ u)(\mathbf{x}) & \text{if } (T_h^+ u)(\mathbf{x}) - u(\mathbf{x}) < u(\mathbf{x}) - (T_h^- u)(\mathbf{x}) \\ (T_h^- u)(\mathbf{x}) & \text{if } (T_h^+ u)(\mathbf{x}) - u(\mathbf{x}) > u(\mathbf{x}) - (T_h^- u)(\mathbf{x}) \\ u(\mathbf{x}) & \text{otherwise} \end{cases} \tag{16.3}$$

The figure **??** illustrates the action of such an operator. In order to understand mathematically the action of T_h, let us examine its asymptotical behaviour. The following exercise proposes to apply Theorem 16.4 to get the asymptotic of T_h^+ and T_h^-. It will however not permit to conclude for T_h, this is done in the next proposition.

Exercise 16.1 *1. Check that $\forall u$ and $\forall \mathbf{x}$:*

$$T_h^- u(\mathbf{x}) \leq u(\mathbf{x}) \leq T_h^+ u(\mathbf{x})$$

2. Using Lemma 16.3 Show that $q^*(p) = q(p)$ and that $\forall \mathbf{x}$ where u is C^2:

$$(T_h^+ u)(\mathbf{x}) - u(\mathbf{x}) = h|Du(\mathbf{x})|^2/2 + O(h^2) \text{ and}$$

$$(T_h^- u)(\mathbf{x}) - u(\mathbf{x}) = -h|Du(\mathbf{x})|^2/2 + O(h^2)$$

So that

$$\lim_{h\to 0} \frac{(T_h u)(\mathbf{x}) - u(\mathbf{x})}{h} = \pm|Du(\mathbf{x})|^2/2$$

At this step, we remark that the differences $(T_h^+ u)(\mathbf{x}) - u(\mathbf{x})$ and $u(\mathbf{x}) - (T_h^- u)$ are equal at the first order, and therefore the choice will be made based on second order estimates on u.

Proposition 16.5 *Let T_h be the "Kramer" operator (given by 16.3), one has for any function $u \in C^3$,*

$$\lim_{h\to 0} \frac{(T_h u) - u}{h} = \frac{1}{2} sgn(D^2 u(Du, Du))\ |Du(\mathbf{x})|^2$$

Figure 16.1: Shock Filter implemented by using non flat morphogical filters. Top, left :original image, right: blurred image using Heat Equation, Middle-left: two iterations of the kramer filter, Middle-right: two iterations of the Rudin-Osher filter. The scale parameter is chosen such that the parabola passes the range of the image at a distance of 6 pixels. Down: zoom version of a detail, left: original image, middle: kramer filter, right: Rudin-Osher filter. We see a tendancy of this last to smooth shapes toward circles.

Proof According to Exercise 16.1, one has to push the asymptotic of T_h^+ and T_h^- to the second order. We have

$$T_h^+(u)(\mathbf{x}) = \sup_{\mathbf{y}\in I\!R^N} u(\mathbf{y}) - \frac{(\mathbf{x}-\mathbf{y})^2}{2h} \qquad \text{and } T_h^-(u)(\mathbf{x}) = \inf_{\mathbf{y}\in I\!R^N} u(\mathbf{y}) + \frac{(\mathbf{x}-\mathbf{y})^2}{2h}$$

Since T_h^+ and T_h^- are translation invariant, we can limit our study at $\mathbf{x} = 0$. Moreover, since u is bounded, we can limit the sup to the $\mathbf{y} \in B(0,h)$. If u is C^3 at point 0, we can set $u(\mathbf{y}) = u(0) + \mathbf{p}.\mathbf{y} + A(\mathbf{y},\mathbf{y}) + o(\mathbf{y})^2$

So that,

$$T_h^+(u)(0) - u(0) = \sup_{\mathbf{y}\in B(0,h)} u(\mathbf{y}) - \frac{|\mathbf{y}|^2}{2h} - u(0) = \sup_{\mathbf{y}\in B(0,h)} (\mathbf{p}.\mathbf{y} + A(\mathbf{y},\mathbf{y}) - \frac{|\mathbf{y}|^2}{2h} + o(h)^2$$

Set $Q_h(\mathbf{y}) = 2h\mathbf{p}.\mathbf{y} + (2hA - Id)(\mathbf{y},\mathbf{y})$, so that we have

$$T_h = \sup_{\mathbf{y}\in B(0,h)} (Q_h(\mathbf{y})/(2h)) + o(h)^2$$

For h small enough $B_h = Id - 2hA$ is positive and inversible. Therefore, the sup of Q_h over the $\mathbf{y}$ exists, and is achieved for $\mathbf{y}_h$ such that

$$2h\mathbf{p} + 2B\mathbf{y}_h = 0 \Rightarrow \mathbf{y}_h = -hB^{-1}(\mathbf{p})$$

Thus,

$$T_h^+(u)(0) - u(0) = \frac{h}{2}(Id - 2hA)^{-1}(\mathbf{p},\mathbf{p}) + o(h^2) = \frac{h}{2}(Id + 2hA)(\mathbf{p},\mathbf{p}) + o(h^2)$$

We conclude that

$$T_h^+(u)(0) - u(0) = \frac{h}{2}|\mathbf{p}|^2 + h^2 A(\mathbf{p},\mathbf{p}) + o(h^2) \tag{16.4}$$

Similarly,

$$T_h^-(u)(0) - u(0) = \frac{h}{2}|\mathbf{p}|^2 - h^2 A(\mathbf{p},\mathbf{p}) + o(h^2) \tag{16.5}$$

From these two last equalities we deduce that

$$((T_h^+ u)(\mathbf{x}) - u(\mathbf{x})) - (u(\mathbf{x}) - (T_h^- u)(\mathbf{x})) = h^2 (D^2 u(\mathbf{x}))(Du(\mathbf{x}), Du(\mathbf{x})) + o(h^2) \tag{16.6}$$

We therefore have

$$T_h(u)(\mathbf{x}) - u(\mathbf{x}) = |Du(\mathbf{x})|^2 \; sgn(\; D^2 u(\mathbf{x}) \; (Du(\mathbf{x}), Du(\mathbf{x}))\;) + o(h)$$

□

Let us remark that if u is a 1D function, then $sgn(D^u(Du, Du))$ coincides with the sign of the laplacian. That is that the Kramer operator corresponds, in 1D, asymptoticaly the Rudin Osher shock filter.

16.3.2 The Rudin Osher Shock Filter.

Let us simply define a scheme that yields asymptoticaly the Rudin Osher shock filter equation.

Let B_h be a disk of radius h centered at 0. Let $Mean$ be the mean value on the disk B_h. We define the operator T_h by:

$$\begin{array}{rcll} T_h u(\mathbf{x}) & = & min_{\mathbf{y}\in B_h} u(\mathbf{x}+\mathbf{y}) & \text{if } Mean(u)(\mathbf{x}) > u(\mathbf{x}) \\ & = & max_{\mathbf{y}\in B_h} u(\mathbf{x}+\mathbf{y}) & \text{if } Mean(u)(\mathbf{x}) < u(\mathbf{x}) \\ & = & u(\mathbf{x}) & \text{otherwise} \end{array}$$

Exercise 16.2 *Prove that*

$$\lim_{h\to 0} T_h u - u = sgn(\Delta u)|Du|$$

16.4 Can we approximate a parabolic PDE by the iterations of a monotone image operator ?

16.4.1 Approximation of first order equation.

Let us address the converse of theorem 16.4: being given the function G is it possible to construct a scaled familly of structuring elements such that the associated scale space T_h satisfies

$$T_h u - u = hG(Du) + O(h^2)?$$

As we shall see, the main difficulty stands in the localization of the structuring elements when the scale tends to 0. In all the following, we work with the scaling parameter β equal to 1.

Theorem 16.6 *Let G be a convex function, such that G^* satisfies condition 16.2, then choosing $\mathbb{F}_h = \{hG^*(\mathbf{x}/h)\}$ one has for the operator T_h associated to $\mathbb{F}_h$ and for any function $u \in C^3$, $(T_h u - u)(\mathbf{x}) = hG(Du(\mathbf{x})) + O(h^2)$*

Proof This is a imediat consequence of Lemma 16.3 and of the fact that if a function G is convex then $G^{**} = G$. An example of such function G is $G(\mathbf{x}) = |\mathbf{x}|^2$. □

When G is non convex, then exhibiting a function M such that $M^* = G$ is non straighforward. It is better to consider G as the infimum of a familly of convex functions $\{g_q\}_q$.

Theorem 16.7 *Let G be a function being the infimum of a familly of convex functions $\{g_q\}_q$, such that for all q, g_q^* satisfies the condition 16.2, then choosing $\mathbb{F}_h = \{hg_q^*(\mathbf{x}/h)\}$ one has for the operator T_h associated to $\mathbb{F}_h$ and for any function $u \in C^3$, $(T_h u - u)(\mathbf{x}) = hG(Du(\mathbf{x})) + O(h^2)$*

Note also that for negative function G, the same result work by switching the sup and the inf in the definition of the operator T_h.

Proof The proof of Theorem 16.7 is a straighforward consequence of Theorem 16.4. □

Examples of functions G that fit the hypothesis of the theorem 16.7 are the positive and Lipschitz functions. Indeed, if G is K-Lipschitz then setting for $q \in \mathbb{R}^N$,

$$g_q(\mathbf{x}) = G(q) + K|\mathbf{x} - q|$$

We obviously have $G(\mathbf{x}) = inf_{q \in \mathbb{R}^N} g_q(\mathbf{x})$. And,

$$g_q^*(p) = \begin{cases} pq - G(q) & if\, |p| \leq K \\ +\infty & otherwise \end{cases}$$

So that $g_q^*(p)$ satisfies the condition 16.2.

Remark 16.8 *However, the hypotheses of Theorem 16.7 do not permit to construct any function G. The main issue is in fact the condition 16.2, which localizes the filter when $h > 0$ tends to 0, in the theorem 16.4.*

Frédéric Cao proposes in [73] a way to avoid such an issue for any positive l.s.c function G. His idea is to define a two scales family of structuring elements. He first set

$$g_q(p) = \begin{cases} G(q) & if p = q \\ +\infty & otherwise \end{cases}$$

It is then obvious that $G(p) = \inf_{q\in I\!R^N} g_q(p)$. He then set $f_q(\mathbf{x}) = (g_q^*)(\mathbf{x}) = -G(q) + q\mathbf{x}$ and $I\!F_h = \{f_{q,h}, q \in I\!R^N\}$ where, for a $\alpha \in]1/2, 1[$,

$$f_{q,h}(\mathbf{x}) = \begin{cases} -hG(q) + q\mathbf{x} & if \ \mathbf{x} \in B(0, h^\alpha) \\ +\infty & elsewhere \end{cases}$$

The familly $I\!F_h$ is not a rescaling of the familly $I\!F_1$. There is indeed, two scales: the explicit one h, and an implicit one, h^α since the functions of $I\!F_h$ are truncated outside a ball of radius h^α. This truncature localizes the corresponding operator T_h and makes the result of theorem 16.4 true, even if the functions of $I\!F_h$ do not satisfy the condition 16.2.

16.4.2 Approximation of some second order equation.

Let us start with a simple remark. Set $f_q(\mathbf{x}) = q\mathbf{x}$, $\forall \mathbf{x}$ in $B(0,h)$ and $f_q(\mathbf{x}) = +\infty$ otherwise. By an imediat consequence of the Taylor expansion we have

$$q = Du(0) \Leftrightarrow \sup_{\mathbf{x}\in I\!R^N} u(\mathbf{x}) - f_q(\mathbf{x}) = O(h^2)$$

$$q \neq Du(0) \Leftrightarrow \sup_{\mathbf{x}\in I\!R^N} u(\mathbf{x}) - f_q(\mathbf{x}) > C(q,u)h$$

This indicates that a way to get second order operator is to choose the familly of functions $I\!F$ so that $\forall f \in I\!F$ and $\forall q \in I\!R^N$ one has $f + q\mathbf{x} \in I\!F$.

The Heat Equation as the asymptotic of a non-flat morphological operator.

Lemma 16.9 *Let A be in $SM(I\!R^N)$ (set of the $N \times N$ symmetric matrices). Then,*

$$Tr(A) = N \inf_{Q\in SM(I\!R^N), Tr(Q)=0} \sup_{\mathbf{x},|\mathbf{x}|=1} (A-Q)(\mathbf{x},\mathbf{x}) \tag{16.7}$$

Proof We know that, since A and Q are symmetric, $\sup_{\mathbf{x},|\mathbf{x}|=1}(A-Q)(\mathbf{x},\mathbf{x})$ is the largest eigenvalue of $A-Q$. As consequence $\forall Q \in SM(I\!R^N)$, $\mathrm{Nsup}_{\mathbf{x},|\mathbf{x}|=1}(A-Q)(\mathbf{x},\mathbf{x}) \geq Tr(A-Q) = Tr(A)$. Thus

$$N \inf_{Q\in SM(I\!R^2), Tr(Q)=0} \sup_{\mathbf{x},|\mathbf{x}|=1} (A-Q)(\mathbf{x},\mathbf{x}) \geq Tr(A).$$

Choosing Q diagonalizable in the same base that diagonalizes A, and denoting by $\lambda_1 \leq ... \leq \lambda_N$ (resp. q_1, ..., q_N) the eigenvalues of A, (resp. of Q), we have

$$\sup_{\mathbf{x},|\mathbf{x}|=1} (A-Q)(\mathbf{x},\mathbf{x}) = \max\{\lambda_1 + q_1, ..., \lambda_N + q_N\}$$

So that

$$\inf_{Q\in SM(I\!R^2), Tr(Q)=0} \sup_{\mathbf{x},|\mathbf{x}|=1} (A-Q)(\mathbf{x},\mathbf{x}) \leq \inf_{\{q_1,...,q_N\}, q_1+...+q_N=0} \max\{\lambda_1+q_1, ..., \lambda_N+q_N\} = (\lambda_1+...+\lambda_N)/N$$

□

Lemma 16.10 *We set for $p \in I\!R^N$, $Q \in SM(I\!R^N)$, and $h > 0$,*

$$\begin{aligned} f_{p,Q,h}(\mathbf{x}) &= p\mathbf{x} + Q(\mathbf{x},\mathbf{x}) \quad \text{if } \mathbf{x} \in B(0,h) \\ &= -\infty \quad \text{otherwise} \end{aligned}$$

We then set $I\!F_h = \{f_{p,Q,h}; \text{with } Q \in SM(I\!R^N); Tr(Q) = 0 \text{ and } p \in I\!R^N\}$ which is to say that $I\!F_h$ is made of the truncature around zero of all quadratic forms whose trace is zero. With $T_h(u)(\mathbf{x}) = \inf_{f \in I\!F_h} \sup_{\mathbf{y} \in I\!R^N} u(\mathbf{x}+\mathbf{y}) - f(\mathbf{y})$, one has for any $u \in C^3$,

$$T_h(u)(\mathbf{x}) - u(\mathbf{x}) = \frac{1}{2N} h^2 \Delta u(\mathbf{x}) + o(h^2)$$

Proof We make the proof at point $\mathbf{x} = 0$, we set $A = \frac{1}{2} D^2 u(0)$. We have

$$T_h(u)(0) - u(0) = \inf_{p \in I\!R^N, Q \in SM(I\!R^N); Tr(Q)=0} \sup_{\mathbf{y} \in B(0,h)} u(\mathbf{y}) - u(0) - p\mathbf{y} - Q(\mathbf{y},\mathbf{y})$$

$$= \inf_{p,Q} \sup_{\mathbf{y} \in B(0,1)} u(h\mathbf{y}) - u(0) - hp\mathbf{y} - h^2 Q(\mathbf{y},\mathbf{y})$$

$$= \inf_{p,Q} \sup_{\mathbf{y} \in B(0,1)} h(Du(0) - p)\mathbf{y} - h^2 (A - Q)(\mathbf{y},\mathbf{y}) + o(h^2)$$

$$= h^2 \inf_{Q \in SM(I\!R^N); Tr(Q)=0} \sup_{\mathbf{y} \in B(0,1)} (A - Q)(\mathbf{y},\mathbf{y}) = \frac{1}{N} h^2 Tr(A)$$

□

References.

Nonflat mathematical morphology

The basic theorem of the "nonflat", or "grey level" mathematical morphology expresses any monotone translation invariant operator as an inf-convolution with a family of structuring functions. We proved it at the very beginning of this chapter. Its consequences are developped in Serra [379], Sternberg, [395], Maragos [279]. An algebraic general framework is proposed by Heijmans [199]. The problematics of simplifying the structuring set of functions is adressed in [164, 230, 97]. Jackway [227] develops the notion of a scale space based on the inf-convolution of the image with quadratic forms at different scales.

Hamilton Jacobi evolution equations associated with nonflat Mathematical Morphology

The discovery of the asymptotics of morphological filters is relatively recent. See [65] and [282], who reinterpret in image analysis results like the Hopf-Lax formula, which were classical in the numerical analysis of P.D.E.'s and in Optimal Control. See the books Evans [141] and Barles [53]. The use of the Fenchel-Legendre of f transform for studying the equation $\frac{\partial u}{\partial t} = f(|Du|)$ when f is convex. The "slope transform" used in [198, 280] is nothing but a new name for the Fenchel-Legendre transform. For a general study of the approximation of parabolic equations by sup-inf schemes see [74].

Deblurring and inverse heat equation, shock filters

Gabor seems to be the first to have introduced the problem of deblurring an image by inverting the heat equation [266]. For more and more sophisticated attempts to perform this inverse heat equation, see Hummel et al. [218, 219]. Nonlinear local filters for image enhancement are proposed as early as 1970 [376, 375]. Kramer and Bruckner introduced in 1975 the nonlinear enhancement filter which we called Kramer filter [256]. In [373], the Kramer filter and its asymptotics is discussed in one dimension. We extended this discussion to two dimensions in the present chapter and showed that the Kramer filter is a shock filter where the Laplacian is replaced by the Canny operator $D^2u(Du, Du)$. Aubert and Kornprobst independently proposed the use of Canny operator for shock filters [**?**]. The term of shock filter itself was framed by Rudin in his PhD dissertation [358], inspired from the use of nonlinear filters in shock simulation for P.D.E.'s [134]. The shock filter discussed in the text uses Osher-Rudin [359] Alvarez and Mazorra [23] proposed to combine in restoration an anisotropic diffusion with a shock filter. Price et al. [346] propose instead closely related reaction-diffusion equations to the same aim. See also [12] for the use of reaction-diffusion equations in image quantization : in some extent, quantization acts as a sharpening. As noticed in [445], the Perona-Malik equation leads to the formation of step images, in the same way as the studied shock filters.

Part IV

Viscosity Solutions and Convergence of Iterated Filters

Chapter 17

Viscosity solutions.

17.1 Definition and main properties.

In what follows, we always consider functions $u(t,\mathbf{x})$ which are continuous on $[0,\infty[\times I\!R^N$. We shall also consider first and second derivatives of real functions. We recall that if u is C^2, then we denote by Du and D^2u the first and second partial derivatives of u with respect to $\mathbf{x}$. $D^2u(t,\mathbf{x})$ is a bilinear form, or, if we choose an euclidean basis, a symmetric matrix. Symmetric matrices (or bilinear forms) can be ordered in the following way. We say that a symmetric matrix $A=(A_{ij})_{1\leq i,j\leq N}$ is nonnegative if for all $p\in I\!R^N$,

$$A(p,p) = {}^t pAp = \sum_{i,j=1}^{N} A_{ij}p_ip_j \geq 0.$$

We say that $A\geq B$ if $A-B\geq 0$. The parabolic equations we shall consider in this chapter are associated with a differential operator $F(D^2u,Du,\mathbf{x},t)$ where F is assumed to be continuous with respect to all of its arguments (except, in some case, at $Du=0$) and nondecreasing with respect to the first argument :

$$\forall\ A,B,p,t,\quad A\geq B \Rightarrow F(A,p,\mathbf{x},t)\geq F(B,p,\mathbf{x},t). \tag{17.1}$$

It will be convenient to consider the case where $F(A,p,\mathbf{x},t)$ is not continuous at $p=0$. This occurs for one of the most relevant equations considered in this book, the curvature equation $\frac{\partial u}{\partial t}=|Du|\mathrm{curv}(u)$. This equation corresponds to $F(A,p,\mathbf{x},t)=A(\frac{p^\perp}{|p|},\frac{p^\perp}{|p|})$ which is not continuous at $p=0$, but admits an continuous bound.

Definition 17.1 and assumption. *We call admissible function F a function F satisfying (17.1), which is continuous for all $A,p\neq 0,\mathbf{x},t$ and such that there exists two continuous functions $G^+(A,p,\mathbf{x},t)$ and $G^-(A,p,\mathbf{x},t)$, with*

$$G^+(0,0,\mathbf{x},t)=G^-(0,0,\mathbf{x},t)=0,$$

$$\forall A\geq 0, G^+(A,0,\mathbf{x},t)\geq 0 \text{ and } G^-(-A,0,\mathbf{x},t)\leq 0 \text{ and}$$

$$\forall A,p,\mathbf{x},t, \text{ we have } G^-(A,p,\mathbf{x},t)\leq F(A,p,\mathbf{x},t)\leq G^+(A,p,\mathbf{x},t). \tag{17.2}$$

Remark 17.2 *Let us first note that if F is continuous everywhere and $F(0,0,\mathbf{x},t)=0$, then F is admissible. (Choose, e.g. $G^\pm=F$).*

Remark 17.3 (Main example : the curvature equation) *We remark that in the case of the curvature equation that F is admissible, with e.g. $G^+(A,p,\mathbf{x},t)$ (resp. G^-) equal to the largest (resp. smallest) of the eigenvalues of A.*

Returning to some examples which are of main interest for us, let us list
• $F(A,p) = |p|$ or $-|p|$, which is the case of dilation and erosion with scale t, associated with $\frac{\partial u}{\partial t} = |Du|$ or $-|Du|$,
• $F(A,p) = \text{trace}(A)$, which is the case of the heat equation $\frac{\partial u}{\partial t} = \Delta u$,
• $F(A,p) = \frac{p_1^2 A_{2,2} - 2p_1 p_2 A_{1,2} + p_2^2 A_{1,1}}{|p|^2}$, which corresponds to the curvature equation in dimension 2,
• $F(A,p) = (p_1^2 A_{2,2} - 2p_1 p_2 A_{1,2} + p_2^2 A_{1,1})^{\frac{1}{3}}$ which corresponds to an affine invariant and contrast invariant smoothing which we call later on "Affine Morphological Scale Space (AMSS)". The associated equation is

$$\frac{\partial u}{\partial t} = |Du|(\text{curv}(u))^{\frac{1}{3}}. \tag{17.3}$$

In this later case, notice that F is continuous at all points (A,p).

Remark 17.4 *It would be comfortable to define a solution of the preceding equations by stating that u is C^2 in $\mathbf{x}$, C^1 in t and satisfies the equation $\frac{\partial u}{\partial t} = F(D^2u, Du, \mathbf{x}, t)$ at all points $(t,\mathbf{x})$. In this case, we say that u is a "classical" solution of the equation. Such a definition works (e.g.) for the heat equation, for which we have shown the existence of classical solutions. There are however, equations among the ones considered above for which a C^2, or even a C^1 solution is not to be expected. Let us give an example where a seemingly "classical" solution is not the right one and violates the comparison principle. Set $u_0(\mathbf{x}) = -|\mathbf{x}|$. Then one is tempted to propose, as a solution for $\frac{\partial u}{\partial t} = |Du|$, the function $\tilde{u}(t,\mathbf{x}) = t - |\mathbf{x}|$. For all t and all $\mathbf{x} \neq 0$, we indeed have $\frac{\partial \tilde{u}}{\partial t} = 1 = |D\tilde{u}(\mathbf{x})|$. Now, this "solution" violates the comparison principle. This can be checked by comparing this solution with the C^∞ solution of the same equation with initial datum $\varphi_0(\mathbf{x}) \equiv 0$, which obviously is $\varphi(t,\mathbf{x}) \equiv 0$. The right, comparison preserving, solution is in fact $u(t,\mathbf{x}) = \min(0, t - |\mathbf{x}|)$. Notice that this last solution is not C^1, so that it cannot be defined as a classical solution. The difficulty of defining a right concept of solution is still more challenging in the case of equations like the mean curvature equation. A fast way to capt the difficulty is to use the contrast invariance : let $g(r)$ be a nondecreasing continuous, but not C^1 real function and $u_0(\mathbf{x})$ an initial datum in $\mathbb{R}^N$. Assume that we have been able to define a "classical" solution u for a curvature equation like (17.3) or the mean curvature motion with initial datum u_0. By contrast invariance, it is to be expected that $g(u)$ will be a solution for the same equation with initial datum $g(u_0)$. Since g is not C^1, it cannot, however, be expected that this solution will be even C^1. We only can assert that it is continuous.*

Although we cannot write properly the equation for an expected solution because of its lack of regularity, we can instead compare it with smooth, classical solutions. To this aim, we denote by $C_b^\infty(\mathbb{R}^N)$ the set of continuous, infinitely differentiable bounded functions on $\mathbb{R}^N$.

Definition 17.5 *Let us first assume that F is admissible. We shall say that u is a viscosity subsolution at point $\mathbf{x}_0$ and scale t_0 of*

$$\frac{\partial u}{\partial t}(t,\mathbf{x}) - F(D^2u(t,\mathbf{x}), Du(t,\mathbf{x}), \mathbf{x}, t) = 0 \tag{17.4}$$

if u is a continuous function and if for all φ in $C_b^\infty([0,T]\times I\!R^N)$ such that $(t_0,\mathbf{x}_0)$ is a local strict maximum point of $u-\varphi$, we have if $D\varphi(t_0,\mathbf{x}_0)\neq 0$

$$\frac{\partial\varphi}{\partial t}(t_0,\mathbf{x}_0) - F(D^2\varphi(t_0,\mathbf{x}_0), D\varphi(t_0,\mathbf{x}_0),\mathbf{x}_0,t_0)\leq 0, \tag{17.5}$$

and if $D\varphi(t_0,\mathbf{x}_0)=0$ and $D^2\varphi(t_0,\mathbf{x}_0)=0$

$$\frac{\partial\varphi}{\partial t}(t_0,\mathbf{x}_0)\leq 0. \tag{17.6}$$

Similarly, u is a viscosity supersolution at point $\mathbf{x}_0$ and scale t_0 if for all φ in $C_b^\infty([0,T]\times I\!R^N)$ such that $(t_0,\mathbf{x}_0)$ is a local strict maximum point of $\varphi-u$, we have if $D\varphi(t_0,\mathbf{x}_0)\neq 0$

$$\frac{\partial\varphi}{\partial t}(t_0,\mathbf{x}_0) - F(D^2\varphi(t_0,\mathbf{x}_0), D\varphi(t_0,\mathbf{x}_0),\mathbf{x}_0,t_0)\geq 0. \tag{17.7}$$

and if $D\varphi(t_0,\mathbf{x}_0)=0$ and $D^2\varphi(t_0,\mathbf{x}_0)=0$

$$\frac{\partial\varphi}{\partial t}(t_0,\mathbf{x}_0)\geq 0. \tag{17.8}$$

We call u a viscosity solution at point $\mathbf{x}_0$ and scale t_0 if it is both a viscosity subsolution and a supersolution.
If $u(t,\mathbf{x})$ is a viscosity solution of (17.4) at each point of $I\!R^{+,\star}\times I\!R^N$ and if $u(0,\mathbf{x})=u_0(\mathbf{x})$, we say that u is a viscosity solution of equation (17.4) with initial condition u_0.

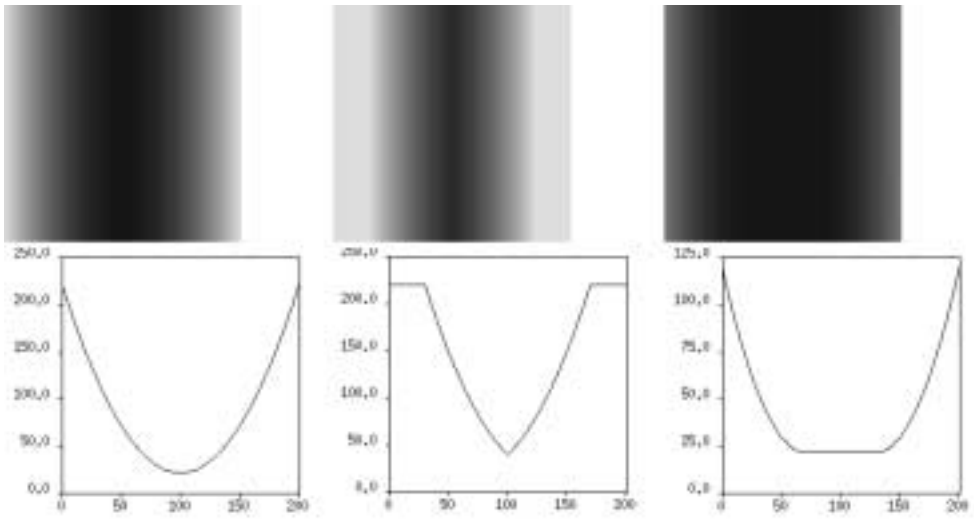

Figure 17.1: Erosions and dilations can create singularities. Top-left : original C^∞ image, and below : representation of the intensity along the horizontal axis. Middle : dilation with a circle of 30 pixels : at the central point, the image is no more C^∞ or even C^1. Right : erosion with the same circle ; we also see some loss of regularity : the solution is no more C^2.

Of course, this definition makes sense only if we prove that "classical" solutions of (17.4) also are viscosity solutions ! This will be done in Proposition 17.10 below. The next lemmas yield a significant simplification of the calculations when we check that a function is a viscosity solution. Of course, all statements on subsolutions are also valid for supersolutions with the adequate changes.

Lemma 17.6 *If u is a viscosity subsolution and $u-\varphi$ has a (not necessarily strict) local maximum at $(t_0,\mathbf{x}_0)$, then the same conclusion holds for (17.5 or 17.6).*

Figure 17.2: Erosion and dilation can generate singularities within the level-lines. Left : original image where the only level-line, corresponding to the boundary of the black shape, is C^2. Right : erosion with a circle of radius 10 : the resulting level-lines have angles.

Proof Just replace $\varphi(t,\mathbf{x})$ by $\psi(t,\mathbf{x}) = \varphi(t,\mathbf{x}) + (t-t_0)^4 + (\mathbf{x}-\mathbf{x}_0)^4$. Then the first and second derivatives of ψ are equal to those of φ and $(t_0,\mathbf{x}_0)$ is now a strict local maximum of $u-\psi$. □

Remark 17.7 *Let us note that the relations (17.5 - 17.8) do not give any bound of the partial derivaty of φ with respect to t, in the case where $D\varphi = 0$ and $D^2\varphi \neq 0$. The next lemma proves that in fact (17.5) and resp. (17.7) hold even in this case, with F replaced by any function G^+ (resp. G^-), where $G^\pm$ satisfies 17.2. In particular if F is continuous everywhere then (17.5) and (17.7) are satisfied everywhere.*

Lemma 17.8 *Let us assume that F is admissible, and G^+ and G^- two continuous functions satisfying 17.2. A continuous function u is a viscosity subsolution (resp. supper-solution) at point $(t_0,\mathbf{x}_0)$ of (17.4) if and only if it satisfies (17.5) and*

$$\frac{\partial\varphi}{\partial t}(t_0,\mathbf{x}_0) - G^+(D^2\varphi(t_0,\mathbf{x}_0), D\varphi(t_0,\mathbf{x}_0), \mathbf{x}_0, t_0) \leq 0 \tag{17.9}$$

(resp. it satisfies (17.7) and

$$\frac{\partial\varphi}{\partial t}(t_0,\mathbf{x}_0) - G^-(D^2\varphi(t_0,\mathbf{x}_0), D\varphi(t_0,\mathbf{x}_0), \mathbf{x}_0, t_0) \geq 0 \tag{17.10}$$

An immediat consequence of this lemma is that if F is admissible and continous, and u a viscosity subsolution (resp. supper-solution), then (17.5) (resp. (17.7)) holds even if the gradient is null.

Proof It is clear that if (17.9) holds then (17.6) holds. Let us prove the reciproque: Let u be a continuous function satisfying (17.5) and (17.6). Let also φ be a C^2 function such that $(t_0,\mathbf{x}_0)$ is astrict local maximum of $u-\varphi$. We have to prove that for any continuous G^+ satisfying condition 17.2, if $D\varphi(t_0,\mathbf{x}_0) = 0$ and $D^2\varphi(t_0,\mathbf{x}_0) \neq 0$, (17.9) holds. (The others cases are immediat).

Let us consider the function $\psi_\epsilon(t,\mathbf{x},\mathbf{y}) = u(t,\mathbf{x}) - \varphi(t,\mathbf{y}) - \frac{|\mathbf{x}-\mathbf{y}|^4}{\epsilon}$. Since $(t_0,\mathbf{x}_0)$ is a maximum point of $u-\varphi$, one proves easily that there exists a sequence $(t_\epsilon,\mathbf{x}_\epsilon,\mathbf{y}_\epsilon)$ of local maximum points of ψ_ϵ converging to $(t_0,\mathbf{x}_0,\mathbf{x}_0)$. $\mathbf{x}_\epsilon$ fixed, ψ is twice differentiable with respect to $\mathbf{y}$, and has a local maximum for $\mathbf{y} = \mathbf{y}_\epsilon$, therefore one has

$$D\varphi_\epsilon(t_\epsilon,\mathbf{y}_\epsilon) = \frac{4(\mathbf{x}_\epsilon - \mathbf{y}_\epsilon)|\mathbf{x}_\epsilon - \mathbf{y}_\epsilon|^2}{\epsilon},$$

and

$$D^2\varphi(t_\epsilon,\mathbf{y}_\epsilon) \geq -\frac{1}{\epsilon}(4|\mathbf{x}_\epsilon - \mathbf{y}_\epsilon|^2 Id - 8(\mathbf{x}_\epsilon - \mathbf{y}_\epsilon) \oplus (\mathbf{x}_\epsilon - \mathbf{y}_\epsilon)).$$

Two cases:

1. $D\varphi(t_\epsilon, \mathbf{y}_\epsilon) = 0$, then $\mathbf{x}_\epsilon = \mathbf{y}_\epsilon$.

   Now, fixing $\mathbf{y} = \mathbf{y}_\epsilon$, ψ_ϵ has a local maximum at $\mathbf{x}_\epsilon$ with respect to the $\mathbf{x}$ variable, using (17.6) we get

   $$\frac{\partial \varphi}{\partial t}(t_\epsilon, \mathbf{y}_\epsilon) \leq 0$$

   and since $D^2\varphi(t_\epsilon, \mathbf{y}_\epsilon) \geq 0$, for any continuous G^+ satisfying 17.2, $G^+(D^2\varphi(t_\epsilon, \mathbf{y}_\epsilon), 0, \mathbf{y}_\epsilon, t_\epsilon) \geq 0$, thus

   $$\frac{\partial \varphi}{\partial t}(t_\epsilon, \mathbf{y}_\epsilon) \leq G^+(D^2\varphi(t_\epsilon, \mathbf{y}_\epsilon), 0, \mathbf{y}_\epsilon, t_\epsilon)$$

   and letting ϵ tends to zero, by continuity of G^+, we deduce that (17.9) holds at point $(t_0, \mathbf{x}_0)$.

2. $D\varphi(t_\epsilon, \mathbf{y}_\epsilon) \neq 0$. We remark that $(t_\epsilon, \mathbf{x}_\epsilon)$ is a maximum point of

   $$(t, \mathbf{x}) \to u(t, \mathbf{x}) - \varphi(t, \mathbf{x} - (\mathbf{x}_\epsilon - \mathbf{y})) - \frac{|\mathbf{x} - \mathbf{y}|^4}{\epsilon}$$

   Then using (17.5), we have

   $$\frac{\partial \varphi}{\partial t} \leq F(D^2\varphi(t_\epsilon, \mathbf{y}_\epsilon), D\varphi(t_\epsilon, \mathbf{y}_\epsilon), \mathbf{y}_\epsilon, t_\epsilon) \leq G^+(D^2\varphi(t_\epsilon, \mathbf{y}_\epsilon), D\varphi(t_\epsilon, \mathbf{y}_\epsilon), \mathbf{y}_\epsilon, t_\epsilon)$$

   Letting ϵ tends to zero, and using the continuity of G^+ we conclude that (17.9) holds at point $(t_0, \mathbf{x}_0)$.

□

Lemma 17.9 *If u satisfies the subsolution (resp. the supersolution) definition 17.5 for any C_b^∞ function φ of the form $\varphi(\mathbf{x}, t) = f(\mathbf{x}) + g(t)$, then it is a subsolution (resp. a supersolution).*

Proof Let $\varphi(\mathbf{x}, t)$ be a C_b^∞ function such that $u - \varphi$ attains its maximum at $(\mathbf{x}_0, t_0)$. We wish to show that (17.5) is satisfied. Without loss of generality, we may assume that the origin is at $(\mathbf{x}_0, t_0)$, so that $(\mathbf{x}_0, t_0) = (0, 0)$. By Taylor formula, we can write

$$\varphi(\mathbf{x}, t) = a + bt + \prec p, \mathbf{x} \succ + ct^2 + \prec Q\mathbf{x}, \mathbf{x} \succ + t \prec q, \mathbf{x} \succ + o(|\mathbf{x}^2| + t^2),$$

where $a = \varphi(0)$, $q = \frac{1}{2}(\frac{\partial^2 \varphi}{\partial x_1 \partial t}(0), ..., \frac{\partial^2 \varphi}{\partial x_N \partial t}(0))$, etc.. We then set

$$f(\mathbf{x}) = a + \prec p, \mathbf{x} \succ + \prec Q\mathbf{x}, \mathbf{x} \succ + \varepsilon |\mathbf{x}|^2 + \varepsilon |q||\mathbf{x}|^2$$

and

$$g(t) = bt + \frac{|q|}{\varepsilon} t^2 + \varepsilon t^2 + ct^2,$$

so that

$$\varphi(\mathbf{x}, t) = f(\mathbf{x}) + g(t) - (\varepsilon |q||\mathbf{x}|^2 + \frac{|q|}{\varepsilon} t^2 + t \prec q, \mathbf{x} \succ + \varepsilon(|\mathbf{x}|^2 + t^2)) + o(|\mathbf{x}|^2 + t^2).$$

Since, by Cauchy-Schwartz inequality, $\varepsilon |q||\mathbf{x}|^2 + \frac{|q|}{\varepsilon} t^2 + t \prec q, \mathbf{x} \succ \geq 0$, we have $\varphi(\mathbf{x}, t) \leq f(\mathbf{x}) + g(t)$ for $(\mathbf{x}, t)$ small enough. Thus, in a neighborhood of 0, $u(t, \mathbf{x}) - \varphi(\mathbf{x}, t) \geq u(t, \mathbf{x}) - f(\mathbf{x}) - g(t)$. This means that $u - f - g$ attains its maximum at 0 and therefore, by assumption,

1. if $D\varphi(0) = D(f+g)(0) \neq 0$ then by (17.5)

$$\frac{\partial(f+g)}{\partial t}(0) \leq F(D^2(f+g)(0), D(f+g)(0), 0, t).$$

Noting that $\frac{\partial(f+g)}{\partial t}(0) = \frac{\partial\varphi}{\partial t}(0)$, $D^2(f+g)(0) = D^2\varphi(0) - 2\varepsilon|q|Id$ and $D(f+g)(0) = D\varphi(0)$, we finally obtain

$$\frac{\partial\varphi}{\partial t}(0) \leq F(D^2\varphi(0) - 2\varepsilon|q|Id, D\varphi(0), 0, t).$$

We let $\varepsilon \to 0$, use the continuity of F with respect to X and conclude.

2. if $D\varphi(0) = D(f+g)(0) = 0$ then by (17.9),

$$\frac{\partial(f+g)}{\partial t}(0) \leq G^+(D^2(f+g)(0), D(f+g)(0), 0, t).$$

Then letting $\varepsilon \to 0$, use the continuity of G^+, yields (17.9) with φ.

The same proof applies with obvious adaptations to the suppersolution property. □

Exercise 17.1 *Let F be admissible. Show that, in the definition of the viscosity solutions (associated with $\frac{\partial u}{\partial t} = F(D^2u, Du, \mathbf{x}, t)$), we can enforce further the test function f to belong to any class $\mathcal{C}$ of C^2 functions having the following property: For any $\mathbf{x} \in \mathbb{R}^N$ and any quadratic form Q, there exists $\varphi \in \mathcal{C}$ such that*

$$f(\mathbf{y}) = Q(\mathbf{y}) + o((\mathbf{y}-\mathbf{x})^2).$$

(Indication: use the order 2 Taylor expansion of φ.)

The next two propositions show that the notion of viscosity solution is a generalization of the concept of "classical" solution.

Proposition 17.10 *Let $F(A, p, \mathbf{x}, t)$ be admissible and continuous everywhere, and u be a classical, (i.e. C^2 with respect to $\mathbf{x}$ and C^1 with respect to t), solution at point $(\mathbf{x}_0, t_0)$ of*

$$\frac{\partial u}{\partial t}(t, \mathbf{x}) = F(D^2u, Du, \mathbf{x}, t).$$

Then u is a viscosity solution at this point.

Proof: We do the case where $Du \neq 0$ at point $(t_0, \mathbf{x}_0)$. (The other case, $Du(t_0, \mathbf{x}_0) = 0$ and $D^2u(t_0, \mathbf{x}_0) = 0$, is imediat). Let φ be a $C^\infty(\mathbb{R}^N \times \mathbb{R}^+ \to \mathbb{R})$ function such that $(t_0, \mathbf{x}_0)$ is a global maximum point of $u - \varphi$, that is

$$u(t_0, \mathbf{x}_0) - \varphi(t_0, \mathbf{x}_0) \geq u(t, \mathbf{x}) - \varphi(t, \mathbf{x}).$$

We want to prove that

$$\frac{\partial\varphi}{\partial t} \leq F(D^2\varphi, D\varphi, \mathbf{x}, t). \tag{17.11}$$

We have

$$u(t, \mathbf{x}) - u(t_0, \mathbf{x}_0) \leq \varphi(t, \mathbf{x}) - \varphi(t_0, \mathbf{x}_0).$$

Thus, by Taylor expansion of order 1,

$$(t-t_0)\frac{\partial u}{\partial t}(t_0,\mathbf{x}_0)+ \prec Du(t_0,\mathbf{x}_0),\mathbf{x}-\mathbf{x}_0\succ +o((\mathbf{x}-\mathbf{x}_0))+o(t-t_0))\leq \tag{17.12}$$

$$(t-t_0)\frac{\partial \varphi}{\partial t}(t_0,\mathbf{x}_0)+ \prec D\varphi(t_0,\mathbf{x}_0),\mathbf{x}-\mathbf{x}_0\succ .$$

Setting $\mathbf{x}=\mathbf{x}_0$, taking $t>t_0$ and $t<t_0$ and letting $t\to t_0$, we get

$$\frac{\partial u}{\partial t}(t_0,\mathbf{x}_0)=\frac{\partial \varphi}{\partial t}(t_0,\mathbf{x}_0).$$

Setting $\mathbf{x}-\mathbf{x}_0=\varepsilon\mathbf{y}$, and letting $\varepsilon\to 0$, we get

$$\forall \mathbf{y},\ \prec Du(t_0,\mathbf{x}_0),\mathbf{y}\succ\leq\prec D\varphi(t_0,\mathbf{x}_0),\mathbf{y}\succ .$$

Thus $Du(t_0,\mathbf{x}_0)=D\varphi(t_0,\mathbf{x}_0)$. We now take $t=t_0$ and do a Taylor expansion in $\mathbf{x}$ of order 2 of the inequality

$$u(t_0,\mathbf{x})-u(t_0,\mathbf{x}_0)\leq\varphi(t_0,\mathbf{x})-\varphi(t_0,\mathbf{x}_0).$$

We obtain

$$\prec D^2u(t_0,\mathbf{x}_0)(\mathbf{x}-\mathbf{x}_0),(\mathbf{x}-\mathbf{x}_0)\succ\leq\prec D^2\varphi(t_0,\mathbf{x}_0)(\mathbf{x}-\mathbf{x}_0),(\mathbf{x}-\mathbf{x}_0)\succ +o((\mathbf{x}-\mathbf{x}_0)^2),$$

which yields

$$D^2u(t_0,\mathbf{x}_0)\leq D^2\varphi(t_0,\mathbf{x}_0).$$

Thus,

$$\frac{\partial \varphi}{\partial t}(t_0,\mathbf{x}_0)=\frac{\partial u}{\partial t}(t_0,\mathbf{x}_0)=F(D^2u(t_0,\mathbf{x}_0),Du(t_0,\mathbf{x}_0),\mathbf{x}_0,t_0)\leq F(D^2\varphi(t_0,\mathbf{x}_0),D\varphi(t_0,\mathbf{x}_0),\mathbf{x}_0,t_0).$$

We deduce that u is a subsolution and we prove in the same way that u also is a supersolution. □

Proposition 17.11 *Let $F(A,p,\mathbf{x},t)$ be admissible and continuous everywhere. Let $u(t,\mathbf{x})$ be C^2 in $\mathbf{x}$ and C^1 in t and assume that u is a viscosity solution of $\frac{\partial u}{\partial t}=F(D^2u,Du,\mathbf{x},t)$. Then u is a classical solution of the same equation.*

Proof : We assume that u is a viscosity solution, and is C^2 at point $(t_0,\mathbf{x}_0)$. We note $\tilde{D}u=(\frac{\partial u}{\partial t},Du)\in I\!R^{N+1}$. By Taylor formula,

$$u(t,\mathbf{x})=u(t_0,\mathbf{x}_0)+\prec\tilde{D}u(t_0,\mathbf{x}_0),(t-t_0,\mathbf{x}-\mathbf{x}_0)\succ+\prec\tilde{D}^2u(t_0,\mathbf{x}_0)(t-t_0,\mathbf{x}-\mathbf{x}_0),(t-t_0,\mathbf{x}-\mathbf{x}_0)\succ+o(|\mathbf{x}-\mathbf{x}_0|^2+|t-t_0|^2).$$

For $\varepsilon\in I\!R$ we set

$$\varphi_\varepsilon=u(t_0,\mathbf{x}_0)+\prec\tilde{D}u,(t-t_0,\mathbf{x}-\mathbf{x}_0)\succ+\prec(\tilde{D}^2u(t_0,\mathbf{x}_0)+\varepsilon Id)((t-t_0,\mathbf{x}-\mathbf{x}_0),(t-t_0,\mathbf{x}-\mathbf{x}_0)\succ .$$

Then, $(t_0,\mathbf{x}_0)$ is a local maximum of $u-\varphi_\varepsilon$, and a local maximum of $\varphi_{-\varepsilon}-u$. Since F is continuous and u is a viscosity solution at point $(t_0,\mathbf{x}_0)$, we have by lemma 17.8, at this point, (with $G^+=F$),

$$\frac{\partial u}{\partial t}=\frac{\partial \varphi_\varepsilon}{\partial t}\leq F(D^2\varphi_\varepsilon,D\varphi_\varepsilon,\mathbf{x}_0,t_0)=F(D^2u+\varepsilon Id,Du,\mathbf{x}_0,t_0)$$

$$\frac{\partial u}{\partial t} = \frac{\partial \varphi_{-\varepsilon}}{\partial t} \geq F(D^2\varphi_{-\varepsilon}, D\varphi_{-\varepsilon}, \mathbf{x}_0, t_0) = F(D^2u - \varepsilon Id, Du, \mathbf{x}_0, t_0).$$

Letting $\varepsilon \to 0$, we obtain

$$\frac{\partial u}{\partial t}(t_0, \mathbf{x}_0) = F(D^2u, Du, \mathbf{x}_0, t_0).$$

□

17.2 Application to mathematical morphology.

Consider the dilation of a function $u_0(\mathbf{x})$ by balls with radii t, that is

$$u(t, \mathbf{x}) = \sup_{\mathbf{y} \in B(0,t)} u_0(\mathbf{x} + \mathbf{y}),\ u(0, \mathbf{x}) = u_0(\mathbf{x}).$$

In that case, we have shown that at points $(t, \mathbf{x})$ where $u(t, \mathbf{x})$ is C^1, we have $\frac{\partial u}{\partial t} = |Du|$. Let us now show that this same equation is satisfied in the viscosity sense by u at *all* points. The following theorem and proofs are immediately generalizable to dilations or erosions by an arbitrary convex set.

Theorem 17.12 *The function u defined by $u(t, \mathbf{x}) = \sup_{\mathbf{y} \in B(0,t)} u_0(\mathbf{x} + \mathbf{y})$ is a viscosity solution of*

$$\frac{\partial u}{\partial t} = |Du|, \qquad u(0, \mathbf{x}) = u_0(\mathbf{x}) \tag{17.13}$$

Proof : We shall use the fact that u satisfies the recursive property

$$u(t, \mathbf{x}) = \sup_{|\mathbf{y}| \leq h} u(t - h, \mathbf{x} + \mathbf{y}). \tag{17.14}$$

Let $\varphi \in C^\infty(\mathbb{R}^N \times \mathbb{R}^+)$. Let $(t_0, \mathbf{x}_0)$ a local maximum point of $u - \varphi$. In order to prove that u is a subsolution of the equation (17.13) we have to prove that

$$\frac{\partial \varphi}{\partial t}(t_0, \mathbf{x}_0) - |D\varphi|(t_0, \mathbf{x}_0) \leq 0.$$

We have, for h and y small enough,

$$u(t_0 - h, \mathbf{x}_0 + \mathbf{y}) - \varphi(t_0 - h, \mathbf{x}_0 + \mathbf{y}) \leq u(t_0, \mathbf{x}_0) - \varphi(t_0, \mathbf{x}_0).$$

We deduce that

$$\sup_{|y| \leq h} u(t_0 - h, \mathbf{x}_0 + \mathbf{y}) \leq u(t_0, \mathbf{x}_0) - \varphi(t_0, \mathbf{x}_0) + \sup_{|y| \leq h} \varphi(t_0 - h, \mathbf{x}_0 + \mathbf{y})$$

and, using (17.14),

$$u(t_0, \mathbf{x}_0) \leq u(t_0, \mathbf{x}_0) - \varphi(t_0, \mathbf{x}_0) + \sup_{|y| \leq h} \varphi(t_0 - h, \mathbf{x}_0 + \mathbf{y}).$$

Thus

$$\varphi(t_0, \mathbf{x}_0) \leq \sup_{|y| \leq h} \varphi(t_0 - h, \mathbf{x}_0 + \mathbf{y}).$$

Subtracting $\varphi(t_0 - h, \mathbf{x}_0)$ on both sides, we get

$$\varphi(t_0, \mathbf{x}_0) - \varphi(t_0 - h, \mathbf{x}_0) \leq \sup_{|y| \leq h} \big(\varphi(t_0 - h, \mathbf{x}_0 + \mathbf{y}) - \varphi(t_0 - h, \mathbf{x}_0)\big).$$

Dividing by h and letting h tends to 0 yields

$$\frac{\partial \varphi}{\partial t}(t_0, \mathbf{x}_0) - |D\varphi|(t_0, \mathbf{x}_0) \leq 0.$$

Thus u is a subsolution and we prove in the same way that it is a supersolution of (17.13). □

17.3 Approximation theory of viscosity solutions

In the following, we consider for simplicity a slightly less general kind of equation for which the second member is independent of t, that is

$$\frac{\partial u}{\partial t}(t, \mathbf{x}) - F(D^2u(t, \mathbf{x}), Du(t, \mathbf{x}), \mathbf{x}) = 0. \tag{17.15}$$

$F(A, p, \mathbf{x})$ is assumed to be admissible (Definition 17.1). For such equations, we can expect that the operator $S_t : u_0 \to u(\mathbf{x}, t)$ can be approximated by the iteration of a single operator T_h, that is, $(T_h)^n \to S_t$ in some sense as $nh \to t$. Of course, we have *exactly* $S_t = (S_h)^n$ if $t = nh$. Now, we are concerned with operators T_h with a definition plainer than S_h. To be precise, we have in mind precisely the same scaled "inf sup" operators which have been considered in the preceding chapter of this book (Chapters 11-15). We shall now list some very reasonable properties which T_h must have if it is asked to have the same structure as S_h when $h \to 0$. It is immediately checked that those properties are true for the scaled "inf sup" operators we have just mentioned.

Definition 17.13 *We say that a family of operators $T_h, h > 0$ is uniformly consistent with Equation 17.15 if for every C^3, Lipschitz function u we can assert that*

$$\text{if } Du(\mathbf{x}) \neq 0, (T_h u)(\mathbf{x}) - u(\mathbf{x}) = hF(D^2u, Du, \mathbf{x}) + o_{\mathbf{x}}(h), \tag{17.16}$$

where the convergence of $o_{\mathbf{x}}(h)$ is uniform for $\mathbf{x}$ in every compact set K contained in the set $\{\mathbf{x}, Du(\mathbf{x}) \neq 0\}$. (That is $o_{\mathbf{x}}(h) \leq C_{u,K}|h|$, for a constant $C_{u,K}$ depending only of u and K.)

$$\text{if } Du(\mathbf{x}) = 0, |(T_h u)(\mathbf{x}) - u(\mathbf{x})| \leq hG(D^2u, 0, \mathbf{x}) + o_{\mathbf{x}}(h) \tag{17.17}$$

for a continuous functions G, with $G(0, 0, \mathbf{x}) = 0$, and, where the convergence of $o(h)$ is uniform for $\mathbf{x}$ in every compact set.

Definition 17.14 *We say that a family of operators $T_h, h > 0$ satisfies a uniform local comparison principle if for every L and all L-Lipschitz functions u and v such that $u(\mathbf{y}) > v(\mathbf{y})$ on a disk $D(\mathbf{x}, r) \setminus \{\mathbf{x}\}$ deprived of its center, we can assert that*

$$(T_h u)(\mathbf{x}) \geq (T_h v)(\mathbf{x}) - o(h), \tag{17.18}$$

where the function $o(h)$ only depends upon the Lipschitz constant L and r.

Definition 17.15 *If T_h, $h > 0$, is a family of operators consistent with Equation 17.15, We call* **approximate solutions** *of (17.4) with initial condition $u_0(\mathbf{x})$ the functions $u_h(t, \mathbf{x})$ defined for every $h > 0$ by*

$$\forall n \in I\!N, \quad u_h(nh, \mathbf{x}) = (T_h^n u_0)(\mathbf{x}),$$

$$\forall t \in [(n-1)h, nh[, \ u_h(t, \mathbf{x}) = u_h((n-1)h, \mathbf{x}).$$

Proposition 17.16 (Convergence) *Let $(T_h)_{h \geq 0}$ be a family of operators uniformly consistent with (17.15), satisfying a uniform comparison principle and commuting with the addition of constants ($T_h(u + C) = T_h u + C$). Assume that a sequence of approximate L-Lipschitz solutions u_{h_k} converges uniformly on every compact set to a function u. Then u is a viscosity solution of (17.4).*

Before starting with the proof, let us state an obvious but useful lemma.

Lemma 17.17 *Let v_k be a sequence of continuous functions converging uniformly on a disk $D(\mathbf{x}, r)$ to a function v. Assume that x is a strict maximum point for v on $D(\mathbf{x}, r)$. Let $\mathbf{x}_k$ be a maximum point of v_k on $D(\mathbf{x}, r)$. Then $\mathbf{x}_k$ tends to $\mathbf{x}$ as $k \to +\infty$.*

Proof of Proposition 17.16 Without risk of ambiguity, we shall write u_h instead of u_{h_k}. Let $D = D((\mathbf{x}, t), r)$ be a disk and $\varphi(s, \mathbf{y})$ a C^∞ function such that $(u - \varphi)(s, \mathbf{y})$ attains its strict maximum on D at $(t, \mathbf{x})$. Without loss of generality, we can assume by Lemma 17.9 that $\varphi(t, \mathbf{y}) = f(\mathbf{y}) + g(t)$. Since $u_h - \varphi \to u - \varphi$ uniformly on D, we know by Lemma 17.17 that a sequence $(\mathbf{x}_h, t_h)$ of maxima of $u_h - \varphi$ on D converges to $(\mathbf{x}, t)$.

Thus we have in particular

$$u_h((n_h - 1)h, \mathbf{y}) - \varphi((n_h - 1)h, \mathbf{y}) \leq u_h(t_h, \mathbf{x}_h) - \varphi(t_h, \mathbf{x}_h). \tag{17.19}$$

Assume that $n_h h \leq t_h < (n_h + 1)h$. Assume first that $t_h \neq n_h h$. Then, by Definition 17.15, we know that $\frac{\partial \varphi}{\partial t}(t_h, \mathbf{x}_h) = 0$. Since φ is C^∞ then we can write

$$\varphi(t_h, \mathbf{x}_h) = \varphi(n_h h, \mathbf{x}_h) + o(h).$$

This relation still holds if $t_h = n_h h$.

In addition, by Definition 17.15 again, we have $u_h(t_h, \mathbf{x}_h) = u_h(n_h h, \mathbf{x}_h)$. Therefore we deduce from the inequality (17.19) that

$$u_h((n_h - 1)h, \mathbf{y}) - \varphi((n_h - 1)h, \mathbf{y}) \leq u_h(n_h h, \mathbf{x}_h) - \varphi(n_h h, \mathbf{x}_h) + o(h).$$

for every $\mathbf{y}$ such that $(\mathbf{y}, (n_h - 1)h) \in D$ and therefore

$$u_h((n_h - 1)h, \mathbf{y}) \leq u_h(n_h h, \mathbf{x}_h) - \varphi(n_h h, \mathbf{x}_h) + \varphi((n_h - 1)h, \mathbf{y}) + o(h)$$

for h small enough (i.e. k large enough) and every $\mathbf{y} \in D(\mathbf{x}, \frac{r}{2})$. Applying on both sides T_h and using the local comparison principle and the commutation of T_h with the addition of constants,

$$T_h(u_h((n_h - 1)h, .))(\mathbf{x}_h) \leq u_h(n_h h, \mathbf{x}_h) - \varphi(n_h h, \mathbf{x}_h) + (T_h \varphi((n_h - 1)h), .)(\mathbf{x}_h) + o(h).$$

Since $\varphi(t,\mathbf{y}) = f(\mathbf{y}) + g(t)$ and $T_h(u((n_h-1)h),.)(\mathbf{x}) = u(n_h h, \mathbf{x})$, we get

$$0 \leq -f(\mathbf{x}_h) - g(n_h h) + T_h f(\mathbf{x}_h) + g((n_h-1)h) + o(h),$$

where we have used again the commutation of T_h with the addition of constants.

Let us first assume that $Df(\mathbf{x}) \neq 0$. By the uniform consistency assumption (17.16), since for h small enough $Df(\mathbf{x}_h) \neq 0$,

$$(T_h f)(\mathbf{x}_h) = f(\mathbf{x}_h) + hF(D^2 f(\mathbf{x}_h), Df(\mathbf{x}_h), \mathbf{x}_h) + o(h).$$

Thus

$$g(n_h h) - g((n_h-1)h) \leq hF(D^2 f(\mathbf{x}_h), Df(\mathbf{x}_h), \mathbf{x}_h) + o(h).$$

Dividing by h, letting $h \to 0$ so that $(\mathbf{x}_h, n_h h) \to (\mathbf{x}, t)$ and using the continuity of F, we get

$$\frac{\partial g}{\partial t}(t) \leq F(D^2 f(\mathbf{x}), Df(\mathbf{x}), \mathbf{x}),$$

that is to say

$$\frac{\partial \varphi}{\partial t}(t) \leq F(D^2\varphi(\mathbf{x}), D\varphi(\mathbf{x}), \mathbf{x}).$$

We treat now the case where $Df(\mathbf{x}) = 0$ and $D^2 f(\mathbf{x}) = 0$. The uniform consistency yields

$$\left|\frac{(T_h f)(\mathbf{x}_h) - f(\mathbf{x}_h)}{h}\right| \leq G(D^2 f(\mathbf{x}_h), Df(\mathbf{x}_h), \mathbf{x}_h) + o(1).$$

The right term, by continuity of G, tends to zero, when h tends to 0. Thus

$$\frac{\partial \varphi}{\partial t}(t) \leq 0$$

Thus u is a subsolution of Equation (17.15) and we prove in exactly the same way that it is a supersolution and therefore a viscosity solution. □

17.4 Uniqueness of viscosity solutions.

Theorem 17.18 *Uniqueness.*

We consider an admissible function F, u_0 and v_0 two continuous and bounded functions of $\mathbb{R}^N$. Assume that u_0 and v_0 are defined on the hypercube $[0,1]^N$, and subsequently extended to $C = [-1,1]^N$ by symmetry across the coordinate hyper-planes and then periodized. Assume that so is $\mathbf{x} \to F(A,p,\mathbf{x},t)$.

If u and v are continuous and viscosity solutions of

$$\frac{\partial u}{\partial t} = F(D^2 u, Du, \mathbf{x}, t) \tag{17.20}$$

with initial conditions respectively u_0 and v_0. Then,

$$\sup_{\mathbf{x}\in\mathbb{R}^N, t\in\mathbb{R}^+} u(t,\mathbf{x}) - v(t,\mathbf{x}) \leq \sup_{\mathbf{x}\in\mathbb{R}^N} u_0(\mathbf{x}) - v_0(\mathbf{x}) \tag{17.21}$$

As consequence, if $u_0 = v_0$ then $u(t,\mathbf{x}) = v(t,\mathbf{x})$ for all $\mathbf{x}$ and t.

Theorem 17.19 *"Theorem on Sums". (revisited). [Crandal-Ishii-Lions Theorem 8.3]*

Let u and v be continuous functions of $\mathbb{R}^+ \times \mathbb{R}^N$, and ϕ be twice differentiable function of $\mathbb{R}^+ \times \mathbb{R}^N$. Set $w(t,\mathbf{x},\mathbf{y}) = u(t,\mathbf{x}) - v(t,\mathbf{y}) - \phi(t,\mathbf{x}-\mathbf{y})$, suppose that $(t_0,\mathbf{x}_0,\mathbf{y}_0)$ is a local maximum of w. Then, there exist two sequences of points $(t_n,\mathbf{x}_n)$ and $(s_n,\mathbf{y}_n)$ converging respectively towards $(t,\mathbf{x}_0)$ and $(t,\mathbf{y}_0)$, and two sequences of twice differentiable functions of $(\mathbb{R}^+ \times \mathbb{R}^N)$: u_n and v_n, such that

$$(t_n,\mathbf{x}_n) \text{ is a local maxima of } u-u_n, \text{ and}$$

$$(\frac{\partial u_n}{\partial t}, Du_n, D^2u_n) \to (a, D\phi(t_0,\mathbf{x}_0,\mathbf{y}_0), X) \tag{17.22}$$

$$(s_n,\mathbf{y}_n) \text{ is a local maxima of } v_n - v, \text{ and}$$

$$(\frac{\partial v_n}{\partial t}, Dv_n, D^2v_n) \to (b, D\phi(t_0,\mathbf{x}_0,\mathbf{y}_0), Y) \tag{17.23}$$

with X and $Y \in S(\mathbb{R}^N)$, such that

$$\text{if } D^2\phi(t_0,\mathbf{x}_0,\mathbf{y}_0) \neq 0 : X \leq Y \tag{17.24}$$

$$\text{if } D^2\phi(t_0,\mathbf{x}_0,\mathbf{y}_0) = 0 : X \leq 0 \text{ and } Y \geq 0 \tag{17.25}$$

$$\text{and with} \qquad a - b = \frac{\partial \phi}{\partial t}(t_0,\mathbf{x}_0,\mathbf{y}_0). \tag{17.26}$$

Lemma 17.20 *Set $\phi(t,\mathbf{x}-\mathbf{y}) = (4\epsilon)^{-1}|\mathbf{x}-\mathbf{y}|^4 + \lambda t$, with ϵ and λ two positive numbers. We consider equation (17.20), where F is assumed to be admissible. Let u and v be continuous and some viscosity solutions of (17.20) with initial condition u_0 and v_0. Set $w(t,\mathbf{x},\mathbf{y}) = u(t,\mathbf{x}) - v(t,\mathbf{y}) - \phi(t,\mathbf{x}-\mathbf{y})$. If $(t_0,\mathbf{x},\mathbf{y})$ is a local maxima of w then $t_0 = 0$.*

proof Let us assume $t_0 > 0$, we want to obtain a contradiction by using the theorem of the sums, and the fact that u and v are viscosity solutions. We have for some $\frac{\partial \phi}{\partial t}(t_0,\mathbf{x}_0-\mathbf{y}_0) = \lambda > 0$, $D\phi(t_0,\mathbf{x}_0-\mathbf{y}_0) = \epsilon^{-1}|\mathbf{x}_0-\mathbf{y}_0|^2(\mathbf{x}_0-\mathbf{y}_0) = p$. There exists two sequences of points $(t_n,\mathbf{x}_n)$ and $(s_n,\mathbf{y}_n)$ converging respectively towards $(t,\mathbf{x}_0)$ and $(t,\mathbf{y}_0)$, and two sequences of twice differentiable functions of $(\mathbb{R}^+ \times \mathbb{R}^N)$: u_n and v_n, and X, Y, $a - b = \lambda$ such that

$$(t_n,\mathbf{x}_n) \text{ is a local maxima of } u-u_n, \text{ and } (\frac{\partial u_n}{\partial t}, Du_n, D^2u_n) \to (a,p,X) \tag{17.27}$$

$$(s_n,\mathbf{y}_n) \text{ is a local maxima of } v_n - v, \text{ and } (\frac{\partial v_n}{\partial t}, Dv_n, D^2v_n) \to (b,p,Y) \tag{17.28}$$

Two cases :

1. if $p \neq 0$, then for n large enough, $Du_n \neq 0$. Then, since u is a viscosity solution of (17.20), (17.27) implies
$$\frac{\partial u_n}{\partial t}(t_n,\mathbf{x}_n) \leq F(D^2u_n(t_n,\mathbf{x}_n), Du_n(t_n,\mathbf{x}_n), \mathbf{x}_n, t_n)$$
Letting n tends to ∞, and using the continuity of F when its second argument is not zero, we obtain
$$a \leq F(X,p,\mathbf{x}_0,t_0) \tag{17.29}$$

Similarly with v_n, (17.28) implies

$$\frac{\partial v_n}{\partial t}(t_n, \mathbf{x}_n) \leq F(D^2 v_n(t_n, \mathbf{x}_n), Dv_n(t_n, \mathbf{x}_n), \mathbf{x}_n, t_n)$$

And at the limit

$$b \geq F(Y, p, \mathbf{x}_0, t_0) \tag{17.30}$$

(17.29) and (17.30) gives $\lambda = a - b \leq 0$, which is a contradiction.

2. if $p = 0$, then $\mathbf{x}_0 = \mathbf{y}_0$. In that case $D^2\phi = 0$, then $X \leq 0$ and $Y \geq 0$. We define $\tilde{u}_n(t, \mathbf{x}) = u_n(t, \mathbf{x}) + \prec X(\mathbf{x}_n - \mathbf{x}), \mathbf{x}_n - \mathbf{x}\succ$, and $\tilde{v}_n(t, \mathbf{x}) = u_n(t, \mathbf{x}) + \prec Y(\mathbf{x}_n - \mathbf{x}), \mathbf{x}_n - \mathbf{x}\succ$. We then have

$$(t_n, \mathbf{x}_n) \text{ is a local maxima of } u - \tilde{u}_n, \text{ and } (\frac{\partial \tilde{u}_n}{\partial t}, D\tilde{u}_n, D^2\tilde{u}_n) \to (a, 0, 0) \tag{17.31}$$

$$(s_n, \mathbf{y}_n) \text{ is a local maxima of } \tilde{v}_n - v, \text{ and } (\frac{\partial \tilde{v}_n}{\partial t}, D\tilde{v}_n, D^2\tilde{v}_n) \to (b, 0, 0) \tag{17.32}$$

By (17.9) we have, since F is admissible, for a continuous function G satisfying 17.2:

$$\frac{\partial \tilde{u}_n}{\partial t}(t_n, \mathbf{x}_n) \leq G(D^2\tilde{u}_n(t_n, \mathbf{x}_n), D\tilde{u}_n(t_n, \mathbf{x}_n), \mathbf{x}_n, t_n)$$

and

$$\frac{\partial \tilde{v}_n}{\partial t}(t_n, \mathbf{x}_n) \geq -G(D^2\tilde{v}_n(t_n, \mathbf{x}_n), D\tilde{v}_n(t_n, \mathbf{x}_n), \mathbf{x}_n, t_n)$$

Letting n tends to ∞, one obtains $a \leq 0$ and $b \geq 0$, which is again in contradiction with $a - b = \lambda > 0$.

□

Proof of Theorem 17.18 We consider $w(t, \mathbf{x}, \mathbf{y}) = u(t, \mathbf{x}) - v(t, \mathbf{y}) - \phi(t, \mathbf{x} - \mathbf{y})$, with ϕ given in the preceding lemma. Since u and v are assumed to be bounded, w tends to $-\infty$ when t increases. The periodicity of u, and v with respect to $\mathbf{x}$ insures existence of a maxima of w for a point $(\mathbf{x}, \mathbf{y}, t) \in I\!R^N \times I\!R^N \times I\!R^+$. According to lemma 17.20, $t = 0$, so we have

$$u(t, \mathbf{x}) - v(t, \mathbf{y}) - (4\varepsilon^{-1}|\mathbf{x} - \mathbf{y}|^4 - \lambda t) \leq \sup_{\mathbf{x}, \mathbf{y} \in I\!R^N} \left((u_0(\mathbf{x}) - v_0(\mathbf{y})) - (4\varepsilon)^{-1}|\mathbf{x} - \mathbf{y}|^4\right)$$

Letting λ tend to 0, we finally get

$$u(t, \mathbf{x}) - v(t, \mathbf{x}) \leq \sup_{\mathbf{x} \in I\!R^N} (u_0(\mathbf{x}) - v_0(\mathbf{x}))$$

□

References.

The simple definition of viscosity solution given in this chapter was proposed by Crandall and Lions [?], originally for first order equations arising in Optimal Control. It was then proved to apply to second order equations as well, in particular the so called geometric equations like the mean curvature motion.

The first complete treatise is Crandall, Ishii, Lions *User's guide to viscosity solutions* [113]. The first order equations are extensively treated in Barles [53]. The Evans textbook on Partial Differential Equations Evans [141] gives an elementary presentation for the first order equations. Crandall's later presentation of the theory for both first and second order equations, published in [**?**], is a masterpiece of simplicity and brievety. This volume contains a pretty complete overview of the techniques, results and applications of viscosity solutions, not including the applications to image processing, though.

The approximation theory of viscosity solutions presented here is based on the seminal Barles and Souganidis paper [52].

The proof of uniqueness for viscosity solutions of second order parabolic or elliptic solutions is the technically difficult part of the theory. Let us mention Jensen [229] and [450, 408] for recent general uniqueness proofs for the parabolic case. Some alternative or related existence and uniqueness theories, namely the nonlinear semigroup theory and the barrier theory are respectively discussed in [175] and [137].

Chapter 18

Curvature equations and iterated contrast invariant operators

18.1 Main curvature equations for image processing

In this chapter, we study the convergence of contrast invariant operators to equations of the kind

$$\frac{\partial u}{\partial t} = |Du|\beta(\text{curv}(u)), \tag{18.1}$$

or, in dimension N, of the kind

$$\frac{\partial u}{\partial t} = |Du|\beta(\kappa_1(u), \kappa_2(u), ...\kappa_{N-1}(u)), \tag{18.2}$$

where β is a continuous real function nondecreasing with respect to its variables and $\kappa_1, ..., \kappa_{N-1}$ denote the principal curvatures of the level surface of u. In other terms, they are the eigenvalues of the restriction of D^2u to the plane orthogonal to Du (Proposition 5.16).

We shall deduce from the convergence results for approximate solutions the existence of viscosity solutions to these equations, wherever approximate solutions u_h have been constructed. There are other methods for proving existence of viscosity solutions (the so called "vanishing viscosity" method, and the Perron method). Now, we shall obtain, besides existence, a proof that the main iterated contrast invariant operators considered in the preceding chapters converge to equations of the kind (18.2). Let us list the main examples we have in mind : this will orient us as for the assumptions we shall make on β. In dimension 2, we have put in evidence as relevant for image processing the equations

$$\frac{\partial u}{\partial t} = |Du|\text{curv}(u) \tag{18.3}$$

and

$$\frac{\partial u}{\partial t} = |Du|\text{curv}(u)^{\frac{1}{3}}, \tag{18.4}$$

as well as variants like

$$\frac{\partial u}{\partial t} = |Du|(\text{curv}(u)^+)^{\frac{1}{3}} \tag{18.5}$$

which correspond to affine erosions. In dimension 3, we shall be concerned with

$$\frac{\partial u}{\partial t} = |Du|\text{curv}(u) = |Du|(\kappa_1(u) + \kappa_2(u)), \tag{18.6}$$

which is the classical "mean curvature motion", important because it arises as a limit of iterated median filters (Theorem 12.2). What we'll state also applies to variants, like

$$\frac{\partial u}{\partial t} = |Du|\text{curv}(u) = |Du| \min(\kappa_1(u), \kappa_2(u)), \tag{18.7}$$

an equation which performs a less destructive smoothing of the 3D images than the mean curvature motion. Finally, let us mention the Gaussian curvature motion, as a particularly important equation in 3D : it performs an affine invariant smoothing (see Theorem 21.8), and can also be used in movie smoothing because of its galilean invariance (Chapter 23) :

$$\frac{\partial u}{\partial t} = |Du| sgn(\kappa_1(u))((\kappa_1(u)\kappa_2(u))^+)^{\frac{1}{4}}. \tag{18.8}$$

In the next section, we shall fix a common formal framework for these equations.

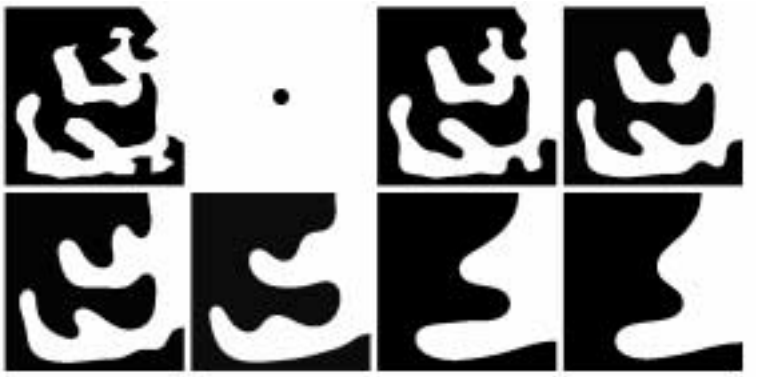

Figure 18.1: Scale-space based on iterations of the median filter. From left to right and up to down : original shape, the size of the disk used for the median, and the result of the iterated median filter for an increasing number of iterations.

Figure 18.2: Comparison of the iterated median filter and of the curvature motion. Numerically, the iterated median filter and the curvature motion are very close, when the curvatures of the level lines are not too small. Indeed, the iterated median filter converges towards the viscosity solution of the curvature motion. Left : simple shape smoothed by a finite difference scheme of the curvature motion, middle : smoothing by a median filter a the same scale, right : difference between left and middle images. The difference is not larger that one pixel width.

18.2 Contrast invariance and viscosity solutions

In this section, we check that the contrast invariance requirement for the solutions of the equations considered in the preceding chapters is compatible with the concept of viscosity solution.

Lemma 18.1 *Equations (18.4-18.5-18.8-18.3-18.6-18.7) can be written as* $\frac{\partial u}{\partial t} = F(D^2u, Du)$ *where* $F(A,p)$ *is an admissible function (see definition 17.1) of* A *and* p *and satisfies*

$$\text{where } Du \neq 0 \qquad F(D^2(g(u)), D(g(u))) = g'(u)F(D^2u, Du) \tag{18.9}$$

for any C^2 *nondecreasing function* g *and any* C^2 *function* u*, and where*

$$\text{where } Du = 0 \text{ and } D^2u = 0, \qquad F(D^2(g(u)), D(g(u))) = F(0,0) = 0 \tag{18.10}$$

Proof If $Du \neq 0$, we know that $\text{curv}(g(u)) = \text{curv}(u)$ and $\kappa_i(g(u)) = \kappa_i(u)$ (Proposition 5.13). Thus

$$F(D^2(g(u)), D(g(u)) = |D(g(u)|\beta(\kappa_1(g(u)), ...\kappa_{N-1}(g(u))) =$$

$$g'(u)|Du|\beta(\kappa_1(u), ...\kappa_{N-1}(u)) = g'(u)F(D^2u, Du),$$

as announced. Let us now consider the case of the mean curvature equations when $Du = 0$. We only treat the case of the mean curvature motions (18.3) and (18.6), the case of (18.7) being similar. In that case, $F(A,p) = \text{Tr}(A) - A(\frac{p}{|p|}, \frac{p}{|p|})p$ and therefore $F(0,0) = 0$.

In addition, if we take $A = D^2(g(u)) = g'(u)D^2u + g''(u)Du \otimes Du$ and $D^2u = Du = 0$, we have $D(g(u)) = 0$ and $D^2(g(u)) = g'(u)D^2u = 0$, so that we immediately obtain $F(D^2(g(u)), D(g(u))) = 0$, when $Du = 0$ and $D^2u = 0$. □

Proposition 18.2 *If* u *is a viscosity solution of one of the equations (18.4-18.5-18.8-18.3-18.6-18.7) and* g *a continuous nondecreasing function, then* $g(u)$ *also is a viscosity solution of the same equation.*

Proof Let us first assume that g is C^∞ and strictly increasing and set for commodity $f = g^{-1}$. Let $(t, \mathbf{x})$ be a local strict maximum of $g(u) - \varphi$. Then $(t, \mathbf{x})$ also is a local strict maximum of $u - g^{-1}(\varphi) = u - f(\varphi)$. Since $f(\varphi)$ is C^∞, and u a viscosity subsolution, we deduce that if $D(f(\varphi))(t, \mathbf{x}) \neq 0$:

$$\frac{\partial(f(\varphi))}{\partial t}(t, \mathbf{x}) \leq F(D^2(f(\varphi))(t, \mathbf{x}), D(f(\varphi))(t, \mathbf{x})),$$

which yields by Lemma 18.1

$$f'(\varphi)\frac{\partial \varphi}{\partial t}(t, \mathbf{x}) \leq f'(\varphi)F(D^2\varphi, D\varphi)(t, \mathbf{x}),$$

and, taking into account that $f' > 0$,

$$\frac{\partial \varphi}{\partial t}(t, \mathbf{x}) \leq F(D^2\varphi, D\varphi)(t, \mathbf{x}).$$

And if $D(f(\varphi)) = 0$, and $D^2(f(\varphi)) = 0$

$$\frac{\partial(f(\varphi))}{\partial t} \leq 0 \text{ and so } \frac{\partial \varphi}{\partial t}(t, \mathbf{x}) \leq 0$$

Thus, $g(u)$ is a viscosity subsolution and, in the same way, a viscosity supersolution. Let us now consider the case where g is not C^∞ and increasing. In this case, we replace $g(s)$ by $g_\varepsilon(s) = \psi_\varepsilon(s) * (1 + \varepsilon s)g(s)$, where ψ is a compactly supported C^∞ real function such that $\int_{I\!R} \psi = 1$, $\psi \geq 0$. It is easily seen that g_ε

converges uniformly on every compact subset of $\mathbb{R}$ to g, is C^∞ and satisfies $g'_\varepsilon \geq \varepsilon$. If $(t, \mathbf{x})$ is a local strict maximum of $g(u) - \varphi$, we deduce from Lemma 17.17 that we can find a sequence $(t_\varepsilon, \mathbf{x}_\varepsilon) \to (t, \mathbf{x})$ of local maxima of $g_\varepsilon(u) - \varphi$. Thus when $D\varphi(t, \mathbf{x}) \neq 0$, so is $D\varphi(t_\varepsilon, \mathbf{x}_\varepsilon)$ for ϵ small enough, and then

$$\frac{\partial \varphi}{\partial t}(t_\varepsilon, \mathbf{x}_\varepsilon) \leq F(D^2\varphi(t_\varepsilon, \mathbf{x}_\varepsilon), D\varphi(t_\varepsilon, \mathbf{x}_\varepsilon)),$$

and if $D\varphi(t, \mathbf{x}) = 0$, and $D^2\varphi(t, \mathbf{x}) = 0$, using lemma (17.8) for a continuous function G^+ such that $G^+(0, 0) = 0$,

$$\frac{\partial \varphi}{\partial t}(t_\varepsilon, \mathbf{x}_\varepsilon) \leq G^+(D^2\varphi(t_\varepsilon, \mathbf{x}_\varepsilon), D\varphi(t_\varepsilon, \mathbf{x}_\varepsilon))$$

and in both cases by passing to the limit as $\varepsilon \to 0$ and using the continuity of F in the first case, and of G^+ in the second case, we conclude that $g(u)$ is a viscosity subsolution. The same arguments apply of course for proving that $g(u)$ is a viscosity supersolution. □

18.3 Uniform continuity of approximate solutions

Lemma 18.3 *Let $\mathbb{B}$ be a family of structuring elements in $\mathbb{R}^N$ and T its associated operator $Tu(\mathbf{x}) = \inf_{B \in \mathbb{B}} \sup_{\mathbf{y} \in \mathbf{x}+B} u(\mathbf{y})$. If u is a Lipschitz function with Lipschitz constant L, then so is Tu.*

Proof Same as Lemma 7.5 and the joint Remark 7.4 □

Lemma 18.4 *Assume that there exist a continuous real function, $k(t)$ satisfying $k(0) = 0$ and such that for $nh \leq t$, $((T_h)^n(L|x|))(0) \leq Lk(t)$ and $((T_h)^n(-L|x|))(0) \geq -Lk(t)$. Assume that the operators T_h are monotone, and commute with the addition of constants and with translations. Then for every L-Lipschitz function u_0, one has $-Lk(t) \leq ((T_h)^n u_0)(\mathbf{x}) - u_0(\mathbf{x}) \leq Lk(t)$.*

Proof Since the operators T_h commute with translations, we can prove the statements in the case of $\mathbf{x} = 0$ without loss of generality. Since u_0 is L-Lipschitz, we have

$$-L|\mathbf{x}| \leq u_0(\mathbf{x}) - u_0(0) \leq L|\mathbf{x}|$$

Applying $(T_h)^n$, using its monotonicity and its commutation with the addition of constants and taking the value at 0,

$$((T_h)^n(-L\mathbf{x}))(0) \leq ((T_h)^n u_0)(0) - u_0(0) \leq ((T_h)^n(L\mathbf{x}))(0),$$

that is, by assumption if $nh \leq t$,

$$-Lk(t) \leq ((T_h)^n u_0)(0) - u_0(0) \leq Lk(t).$$

□

Lemma 18.5 *Let $u_0(\mathbf{x})$ be a lipschitz function on $\mathbb{R}^N$. Let T_h be a family of operators satisfying the assumptions of Lemmas 18.3 and 18.4 with a function $k(t)$ satisfying $k(t) \leq t^\alpha$ for some $\alpha > 0$ if t is small*

enough. Then the approximate solutions $u_h(\mathbf{x},t)$ associated with T_h are uniformly equicontinuous when we restrict t to the set nh. More precisely, for all $n,m \in \mathbb{N}$ and all $\mathbf{x},\mathbf{y}$ in $\mathbb{R}^N$,

$$|u_h(\mathbf{x},nh) - u_h(\mathbf{y},mh)| \leq L|\mathbf{x}-\mathbf{y}| + k(|n-m|h). \tag{18.11}$$

As a consequence, we can extract sequences u_{h_n}, with $h_n \to 0$, which converge uniformly on every compact subset of $\mathbb{R}^N \times \mathbb{R}^+$.

Proof Since by definition $u_h(\mathbf{x}) = ((T_h)^n u_0)(\mathbf{x})$, the result is a direct consequence of Lemmas 18.3 and 18.4 : By the first mentioned lemma, $|u_h(\mathbf{x},nh) - u_h(\mathbf{y},nh)| \leq L|\mathbf{x}-\mathbf{y}|$ and by the second one applied with $(T_h)^{n-m}$,

$$|u_h(\mathbf{x},nh) - u_h(\mathbf{x},mh)| = |((T_h)^{n-m}u_h(.,mh))(\mathbf{x}) - u_h(\mathbf{x},mh)| \leq k(|n-m|h).$$

In order to end with the argument, we simply notice that

$$|u_h(\mathbf{x},nh) - u_h(\mathbf{y},mh)| \leq |u_h(\mathbf{x},nh) - u_h(\mathbf{y},nh)| + |u_h(\mathbf{y},nh) - u_h(\mathbf{y},mh)|.$$

Consider the linear interpolation of u_h,

$$\tilde{u}_h(\mathbf{x},t) = \frac{t-nh}{h}u_h(\mathbf{x},(n+1)h) + \frac{(n+1)h-t}{h}u_h(\mathbf{x},nh).$$

It is easily checked that

$$\tilde{u}_h(\mathbf{y},t) - \tilde{u}_h(\mathbf{y},s)| \leq \frac{k(h)}{h}|t-s| \leq |t-s|^\alpha \text{ for } |t-s| \leq h$$

and

$$|\tilde{u}(\mathbf{y},t) - \tilde{u}(\mathbf{y},s)| \leq k(2|t-s|) \text{ for } |t-s| \geq h.$$

Thus, by the same argument as above, we conclude that the family of functions $\tilde{u}_h$ is uniformly equicontinuous on all of $\mathbb{R}^N \times [0,+\infty]$. In addition, $\tilde{u}_h(\mathbf{x},0) = u_0(\mathbf{x})$ is fixed. We can therefore apply Ascoli-Arzela Theorem ([]) which asserts that under such conditions, the family of functions $\tilde{u}_h(\mathbf{x},t)$ has subsequences converging uniformly on every compact of $\mathbb{R}^N \times [0,+\infty]$ towards a uniformly continuous function $u(\mathbf{x},t)$. Same conclusion holds for $u_h(\mathbf{x},t)$. □

18.4 Convergence of iterated median filters to mean curvature motion

We prove here one of the main practical and theoretical results of this book : the iterated median filters converge to the mean curvature motion equation.

Lemma 18.6 *(median filter) Let k be a radial, nonnegative, compactly supported bounded function and $k_h(\mathbf{y}) = \frac{1}{h^N}h(\frac{\mathbf{x}}{h})$ the associated scaled function. Assume, without loss of generality, that the support of k_h is $B(0,h)$ and consider the weighted median filter associated with k_h, $T_h u(\mathbf{x}) = \text{med}_{k_h} u(\mathbf{x})$. Set $v_0(\mathbf{x}) = v_0(|\mathbf{x}|) = v_0(r) = Lr$. Then, if $nh^2 \leq t$,*

$$(T_h^n v)(0) \leq L\sqrt{2t}.$$

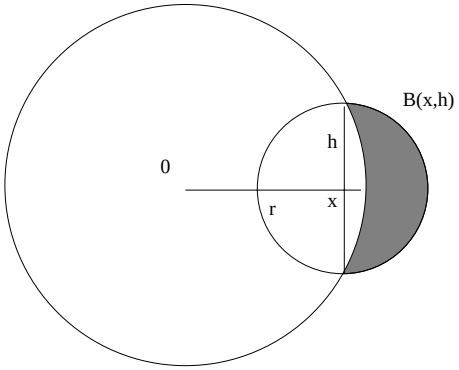

Figure 18.3: Illustrating the inequality (18.12).

Proof Let us first estimate $T_h v(r)$ when $v(\mathbf{x}) = v(|\mathbf{x}|) = v(r)$ is any radial nondecreasing function. To this aim, let $\mathbf{x}$ be such that $|\mathbf{x}| = r$. By the triangular inequality, the sphere with center 0 and radius $\sqrt{r^2+h^2}$ divides the ball $B = B(\mathbf{x}, h)$ into two parts such that

$$\text{meas}_{k_h}(\{\mathbf{y}, |\mathbf{y}| \geq \sqrt{r^2+h^2}\} \cap B - \mathbf{x}) \leq \text{meas}_{k_h}(\{\mathbf{y}, |\mathbf{y}| \leq \sqrt{r^2+h^2}\} \cap B - \mathbf{x}). \tag{18.12}$$

As a consequence, v being nondecreasing, we have

$$\text{med}_{k_h} v \leq v(\sqrt{r^2+h^2}). \tag{18.13}$$

Let us set for brievety $f_h(r) = \sqrt{r^2+h^2}$ and $r_{n+1}(r) = f_h(r_n)$, $r_0 = r$. Then we obviously have from (18.13) and the monotonicity of T_h

$$(T_h^n v)(r) \leq v(r_n(r)). \tag{18.14}$$

In addition, since $\sqrt{r^2+h^2} \leq r + \frac{1}{2r}h^2$ and r_n is an increasing sequence, we obtain $r_{n+1} \leq r_n + \frac{h^2}{2r_n} \leq r_n + \frac{h^2}{2r_0}$ and therefore

$$r_n \leq r + \frac{nh^2}{2r}. \tag{18.15}$$

Let us assume that $nh^2 \leq t$. Taking into account that v is a nondecreasing function, (18.14-18.15) yield

$$(T_h^n v)(r) \leq v(r + \frac{t}{2r}). \tag{18.16}$$

Since $(T_h^n v)(r)$ is a nondecreasing function of r, we deduce that $(T_h^n v)(r) \leq v(\sqrt{2t})$ if $r \leq \sqrt{\frac{t}{2}}$. Thus, if $v(r) = Lr$, we have for $nh^2 \leq t$

$$(T_h^n v)(0) \leq L\sqrt{2t}$$

as announced.

□

Theorem 18.7 Convergence of iterated weighted median filter. *Let k_h be as in Lemma 18.6 and set again $(T_h u)(\mathbf{x}) = \text{med}_{k_h} u$. Let u_0 be a Lipschitz function defined on $\mathbb{R}^N$. Then the approximate solutions u_h converge to a viscosity solution u of the associated equation $\frac{\partial u}{\partial t} = F(D^2u, Du, \mathbf{x}) = c_k |Du| \text{curv}(u)$.*

Proof We know (Theorems 11.3-12.2) that the weighted median is consistent in any dimension with the equation $\frac{\partial u}{\partial t} = F(D^2u, Du) = c_k|Du|\text{curv}(u)$. Bounds for the result of the iterated filter $(T_h)^n$ applied to $+L|\mathbf{x}|$ and $-L|\mathbf{x}|$ have been computed in Lemma 18.6, so that the assumption of Lemma 18.4 is true. In addition, we know that med_h is monotone, satisfies a local comparison principle (Relation 17.18), commutes with translations and the addition of constants. Thus, we can apply Lemma 18.5 which asserts that if u_0 is a Lipschitz function, then a subsequence of u_h converges uniformly on compact sets of $\mathbb{R}^N * \mathbb{R}^+$ to a function u. In addition, by Proposition 17.16, u is a viscosity solution of the equation $\frac{\partial u}{\partial t} = F(D^2u, Du, \mathbf{x}) = c_k|Du|\text{curv}(u)$. Since by Theorem 17.19, this solution is unique, we deduce that all of $(T_h)^n u_0$ converges to u. We have thus proved both existence of a viscosity solution and convergence of the iterated median filter. □

18.5 Convergence of iterated affine invariant operators to affine invariant curvature motion

The most important case in dimension 2 for which we analyze convergence of iterated contrast invariant filters is the equation $\frac{\partial u}{\partial t} = |Du|(\text{curv}(u))^{\frac{1}{3}}$, which is associated with affine invariant inf-sup operators $T_h u(\mathbf{x}) = \inf_{B\in\mathcal{B}} \sup_{\mathbf{y}\in\mathbf{x}+h^{\frac{2}{3}}B} u(\mathbf{y})$, where $\mathcal{B}$ is a family of structuring elements invariant by every linear transform of $\mathbb{R}^2$ with determinant 1. We also set

$$SI_h u(\mathbf{x}) = \sup_{B\in\mathbb{B}_s} \inf_{\mathbf{y}\in B} u(\mathbf{y})$$

$$IS_h u(\mathbf{x}) = \inf_{B\in\mathbb{B}_s+\mathbf{x}_0} \sup_{\mathbf{y}\in B} u(\mathbf{y})$$

where, here again, the relation $s = h^{\frac{3}{2}}$ is imposed. SI_h is understood as an "affine erosion" of u and IS_h as an "affine dilation" and we also consider the alternate scheme : $SI_h(IS_h u)$ Denoting by T_h one of the schemes IS_h, SI_h or $SI_h IS_h$, we set

$$u_h(\mathbf{x}, (n+1)h) = T_h u_h(\mathbf{x}, nh)$$

$$u_h(\mathbf{x}, 0) = u_0(\mathbf{x})$$

Theorem 18.8 *There exists a constant $c_{\mathbb{B}} > 0$ such that if $u_0(\mathbf{x})$ is any Lipschitz function on $\mathbb{R}^2$, then $u_h(\mathbf{x}, nh)$ tends uniformly on every compact set to $u(\mathbf{x}, t)$, where $u(\mathbf{x}, t)$ is the unique viscosity solution of*

$$\frac{\partial u}{\partial t} = |Du|g(\text{curv}(u)) \tag{18.17}$$

$$\begin{array}{rcll} \textit{where } g(r) & = & c_{\mathbb{B}}(r^-)^{\frac{1}{3}} & \textit{if} \quad T_h = SI_h \\ & = & c_{\mathbb{B}}(r^+)^{\frac{1}{3}} & \textit{if} \quad T_h = SI_h \\ & = & c_{\mathbb{B}}(r)^{\frac{1}{3}} & \textit{if} \quad T_h = SI_h\ IS_h \end{array}$$

Theorem 18.8 is somehow a consequence of theorem 15.4, which states a consistency result for the schemes SI_h, IS_h, $SI_h IS_h$. In order to achieve the proof of Theorem 18.8, we need to check that the assumptions of Lemma 18.4 are satisfied.

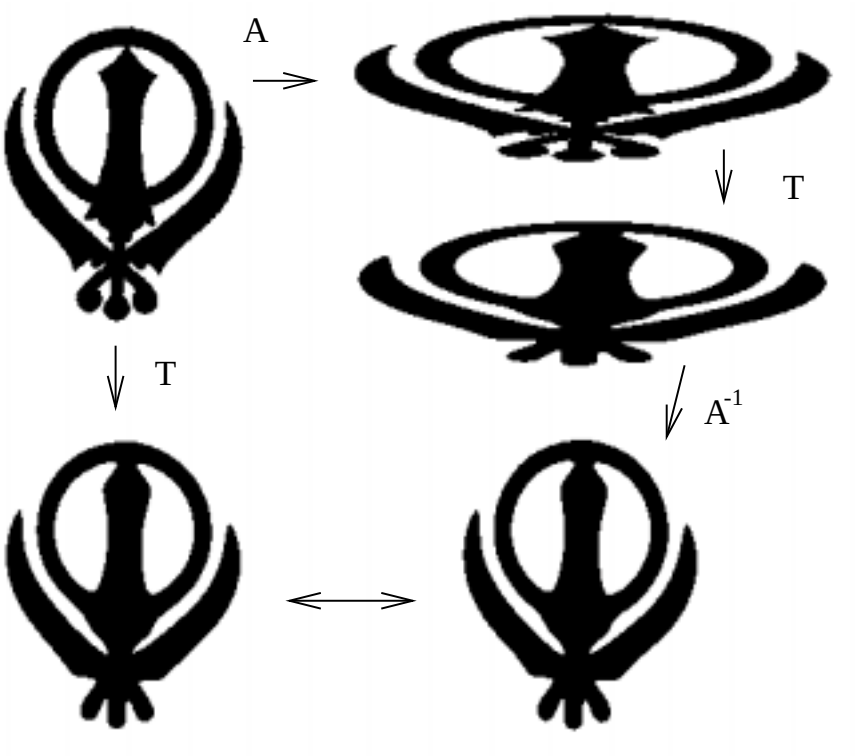

Figure 18.4: Affine invariance of (AMSS). We check the affine invariance of the affine and morphological scale space (AMSS). A simple shape (top-left) is smoothed using a finite differences discretization of (AMSS) followed by thresholding (bottom-left). We apply on the same shape an affine transform, with determinant equal to 1, (top-right), then the same smoothing process (middle-right), and finally the inverse of the affine transform (down-right). The final results of both processes are experimentally equal.

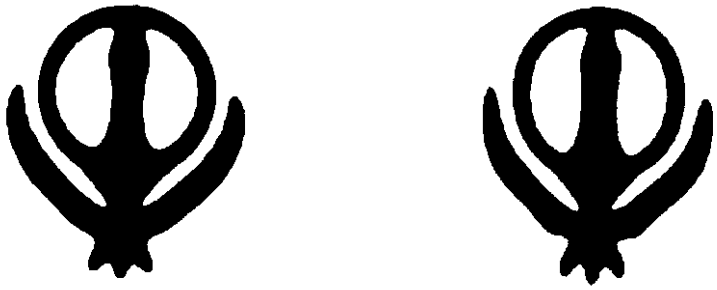

Figure 18.5: Checking the affine invariance of an affine "inf-sup". We display the final outcomes of the same comparison process as in Figure 18.4, with an affine "infsup". The chosen structuring set is an approximately affine invariant set $I\!B$ of 49 ellipses, all with same area. This implementation is heavy and proves to be less affine invariant as the one obtained by a finite difference scheme. This is due to grid effects.

Lemma 18.9 *Let $I\!B$ be a family of subsets of $I\!R^2$ invariant by $SL(I\!R^2)$. Assume that the convex hull of each element B of $I\!B$ is contained in a rectangle with area a^2. Set for any real function v defined in $I\!R^2$.*

$$T_h v(\mathbf{x}) = \inf_{B \in I\!B} \sup_{\mathbf{y} \in h^{\frac{3}{2}} B + \mathbf{x}} v(y).$$

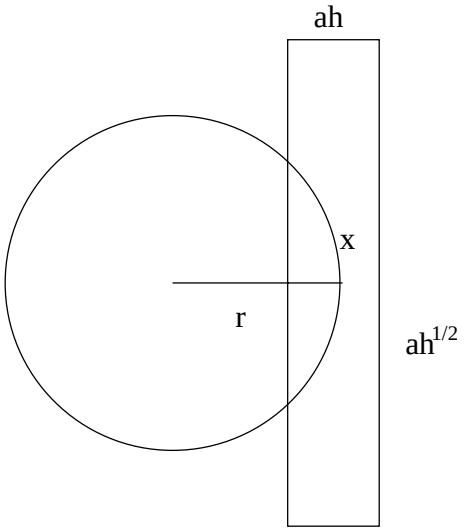

Figure 18.6: Illustration of the proof of the inequality (??).

Consider the radial function $v(\mathbf{x}) = v(|\mathbf{x}|) = v(r) = Lr$. *Then for* $nh \leq t$,

$$(T_h^n v)(0) \leq L(at + 2a\sqrt{t}).$$

Proof We first assume that $v(\mathbf{x}) = v(|\mathbf{x}|) = v(r)$ is an arbitrary radial nondecreasing function and estimate $(T_h^n v)(r)$. Let $\mathbf{x} = (r, 0)$. By assumption, we can enclose any of the elements B of $h^{\frac{3}{2}}\mathcal{B}$ by a rectangle R_h with area $h^{\frac{3}{2}}a^2$. By the Euclidean invariance of $\mathcal{B}$, we can choose this rectangle to have its sides parallel to the axes, and by the $SL^2(\mathbb{R}^2)$ invariance of $\mathcal{B}$ we can further choose B in such a way that its side parallel to the x-axis has length ah and the other one $a^2h^{\frac{1}{2}}$. Since $B \subset R_h$, we have

$$T_h v(\mathbf{x}) \leq \sup_{\mathbf{x}+B} v(\mathbf{y}) \leq \sup_{\mathbf{x}+R_h} v(\mathbf{y}).$$

Thus

$$T_h v(\mathbf{x}) \leq v((r + \frac{ah}{2})^2 + a^2h)^{\frac{1}{2}}) \tag{18.18}$$

We set for conciseness $f_h(r) = ((r + \frac{ah}{2})^2 + a^2h)^{\frac{1}{2}}$ and $r_{n+1}(r) = f_h(r_n)$, $r_0 = r$. Notice that $T_h v$ is a radial nondecreasing function, so that we can replace v by $T_h v$ in (18.18). By the monotonicity of T_h, we obtain

$$(T_h^n v)(r) \leq v(r_n(r)) \tag{18.19}$$

In addition, since $(r^2 + \varepsilon)^{\frac{1}{2}} \leq r + \frac{1}{2r}\varepsilon$ for all $r, \varepsilon > 0$, we have for $h \leq 1$

$$f_h(r) \leq (r^2 + ahr + a^2h + \frac{1}{4}a^2h^2)^{\frac{1}{2}} \leq (r^2 + 2a^2h + ahr)^{\frac{1}{2}} \leq r + \frac{1}{2r}(2a^2h + ahr).$$

Finally,

$$f_h(r) \leq r + ah + \frac{a^2h}{r}. \tag{18.20}$$

Thus $r_{n+1} = f_h(r_n) \leq r_n + ah + \frac{a^2h}{r_n} \leq r_n + ah + \frac{a^2h}{r}$, because r_n is an increasing sequence. Finally, $r_n \leq r + n(ah + \frac{a^2h}{r})$ and, by (18.19),

$$(T_h^n v)(r) \leq v(r + n(ah + \frac{a^2h}{r})).$$

Let us assume that $nh \leq t$. Then

$$(T_h^n v)(r) \leq v(r + (a + \frac{a^2}{r})t).$$

Considering that the minimum value of $r \to r + (a + \frac{a^2}{r})$ is attained at $\sqrt{t}$, we finally obtain

$$(T_h^n v)(r) \leq v(2a\sqrt{t} + at)$$

if $r \leq a\sqrt{t}$. Thus if $v(r) = Lr$, we obtain

$$(T_h^n v)(0) \leq L(at + 2a\sqrt{t})$$

for $nh \leq t$, as announced. □

Proof of Theorem 18.8 By Theorems 15.4 and 15.7, the operators T_h are consistent with their corresponding partial differential equations $\frac{\partial u}{\partial t} = F(D^2u, Du)$ and satisfy a uniform local maximum principle. Being contrast invariant, they commute with the addition of constants. Thus, by Proposition 17.16, if a sequence of approximate L-Lipschitz solutions u_{h_n} converges uniformly on every compact set to a function u, then u is a viscosity solution of (17.4).

Now, since by Lemmas 18.5 and 18.9 the approximate solutions u_h are equicontinuous on every compact set of $\mathbb{R}^N \times \mathbb{R}^+$ and therefore have subsequences which indeed converge to a function u on every compact subset of $\mathbb{R}^N \times \mathbb{R}^+$. Thus, u is a viscosity solution. In addition, we know that a viscosity solution is unique (Theorem 17.18). Thus the limit u does not depend on the particular considered subsequence. Thus the whole sequence u_h converges to u. So we have fully proven both the existence of a viscosity solution and the convergence of the iterated $T_h u_0$ to it, under the assumption that the initial function u_0 is Lipschitz. □

18.6 Solutions of curvature equations for nonsmooth initial images

Preceeding sections give existency and uniqueness of solution (in sense of viscosity) to the curvature equations when the initial image is continuous. Curvature equations yield contrast invariant operator defined on continuous function. Then, using the theorem 8.16 yields to two unique contrast invariant extensions to semi-continuous images: one in case of upper semi continuous initial images, the other in case of lower semi-continuous images.

Existence and uniqueness in the case of upper or lower semi-continuous images is then a immediat consequence of this Theorem.

References.

The existence and uniqueness theory of viscosity solutions for the Motion by Mean Curvature and the relationship of these solutions with other types of solutions (classical, variational) was developped

independently by Evans, Spruck [138, 392, 393, 394] and Chen, Giga, Goto. [100, 101]. We do not follow their existence proofs, but rather the elegant numerical approximation (the iterated gaussian median filter !) invented by Bence, Merriman and Osher [291] and the subsequent convergence proof to the viscosity solution by Barles and Georgelin [50]. Other proofs of the gaussian iterated median filter towards the Mean Curvature Equation are given by [136], using semigroup theory and Ishii [225]. Let it be recalled how relevant the iterated median filter is for denoising applications [35, 224]

Chapter 19

A snake from A to Z...

19.1 An active contour model

Let us consider an edge as a curve located mostly on high gradient points. Indeed, in some cases, boundaries of objects induce some discontinuities in the gray level, resulting in high gradient. The aim of the active contour methods is to find such edge around an initial curve given e.g. by hand. The curve is moved from its original location until it maximizes the amount of gradient along its path.

This can be formulated as an optimization problem. We choose a function g from $\mathbb{R}^2$ into $\mathbb{R}$ representing for each point $\mathbf{x}$ a penalty for the curve to pass by the point $\mathbf{x}$. Ideally, g has to be chosen small when the magnitude of the image gradient is large. E.g. one could choose

$$g = \frac{1}{|DI| + \epsilon}$$

Given an image I, and an initial curve $C_0 = \mathbf{x}_0(s)$, we want to find a curve $C = \mathbf{x}(s)$ that minimizes the following energy:

$$E(C) = \int_0^L g(\mathbf{x}(s))ds \tag{19.1}$$

around $\mathbf{x}_0$, where s is the arc length and L the total length of the curve C.

In all the following, we will assume g to be differentiable with respect to $\mathbf{x}$.

Proposition 19.1 *Let $C(t) = \mathbf{x}(t,s)$ the curve resulting from a gradient descent of the energy (19.1). While $C(t)$ is regular and denoting by $\mathbf{n}$ its normal, $C(t)$ satisfies the following equation*

$$\frac{\partial \mathbf{x}(t,s)}{\partial t} = g(\mathbf{x}(t,s))Curv(\mathbf{x}(t,s))\mathbf{n}(t,s) - (Dg(\mathbf{x}(t,s)).\mathbf{n})\mathbf{n} \tag{19.2}$$

$$\textit{and } \mathbf{x}(0,s) = \mathbf{x}_0(s) \textit{ that is } (C(0) = C_0)$$

Proof Let us first change the parameterization of the curve C so that its length is not a parameter of the energy. We parameterize the curve with $z \in [0,1]$. We have $ds = |\mathbf{x}'(z)|dz$, where $'$ denotes the derivate with respect to z. Thus

$$E(C) = \int_0^1 g(\mathbf{x}(z))|\mathbf{x}'(z)|dz$$

Then,

$$E(C+dC)-E(C)=\int_0^1 Dg(\mathbf{x}(z))|\mathbf{x}'(z)|d\mathbf{x}(z)dz+\int_0^1 g(\mathbf{x}(z))\frac{\mathbf{x}'(z)}{|\mathbf{x}'(z)|}d\mathbf{x}'(z)dz+o(|dC|)$$

Integrating by parts the last integral, we have

$$E(C+dC)-E(C)=\int_0^1 Dg(\mathbf{x}(z))|\mathbf{x}'(z)|d\mathbf{x}(z)dz-\int_0^1 (Dg(\mathbf{x}(z))\mathbf{x}'(z))\frac{\mathbf{x}'(z)}{|\mathbf{x}'(z)|}d\mathbf{x}(z)dz$$
$$-\int_0^1 g(\mathbf{x}(z))Curv(\mathbf{x}(z))|\mathbf{x}'(z)|\mathbf{n}d\mathbf{x}(z)dz+o(|dC|)$$

The two first integrals can be merged so that $E(C+dC)-E(C)=$

$$\int_0^1 (Dg(\mathbf{x}(z)).\mathbf{n})|\mathbf{x}'(z)|\mathbf{n}d\mathbf{x}(z)dz-\int_0^1 g(\mathbf{x}(z))Curv(\mathbf{x}(z))|\mathbf{x}'(z)|\mathbf{n}d\mathbf{x}(z)dz+o(|dC|)$$

The intrinsic scalar product between two vectorial functions f and h defined on the curve $\mathbf{x}(z)$ is given by

$$f.g=\int_0^1 f(\mathbf{x}(z))g(\mathbf{x}(z))|\mathbf{x}'(z)|dz$$

So that

$$E(C+dC)-E(C)=dC\ .\ ((Dg(\mathbf{x}(z)).\mathbf{n})\mathbf{n}-g(\mathbf{x}(z))Curv(\mathbf{x}(z))\mathbf{n})+o(|dC|)$$

We therefore have

$$\nabla E(C)=(Dg(\mathbf{x}(z)).\mathbf{n})\mathbf{n}-g(\mathbf{x}(z))Curv(\mathbf{x}(z))\mathbf{n}$$

As consequence the gradient descent is described by the following equation

$$\frac{\partial\mathbf{x}(t,z)}{\partial t}=-(Dg(\mathbf{x}(t,z)).\mathbf{n})\mathbf{n}+g(\mathbf{x}(z))Curv(\mathbf{x}(z))\mathbf{n}$$

□

Unfortunately, there is no guaranty that such an evolution yields a regular curve for all t. In fact, it is in general false, since topological changes can appear.

19.2 Study of the Active Contour Equation

We study in this section the following equation

$$\frac{\partial u}{\partial t}=g|Du|curv(u)+Dg.Du \tag{19.3}$$

Admissibility of the equation and uniqueness of solutions. Given a vector p of $\mathbb{R}^2$, we denote by $p^{\pm}$ a vector orthogonal to p and with norm equal to 1. We define the function F from $SM(\mathbb{R}^2)\times\mathbb{R}^2\times\mathbb{R}^2\times\mathbb{R}^+$ into $\mathbb{R}$ by

$$F(A,p,\mathbf{x},t)=g(\mathbf{x})A(p^{\pm},p^{\pm})+Dg(\mathbf{x})p$$

Equation 19.3 can be obviously write as

$$\frac{\partial u}{\partial t}=F(D^2u,Du,\mathbf{x},t)$$

It is easy to check that F is admissible (see Definition 17.1). As consequence Theorem 17.18 ensures uniqueness of viscosity solutions of 19.3 for any Lipschitz initial condition.

Existence of solutions by approximation. Let us now construct a approximation scheme to the solution of equation (19.3). Even if it is possible to construct a family of structuring elements having as asymptotic behavior the two terms in the right part of the equation (19.3), it is simpler to consider for each term one family of structuring elements and to alternate their corresponding filters. To be noted that, due to g, the equation is not invariant by translation. As consequence, the families of structuring elements are depending on $\mathbf{x}$.

Approximation of $-DgDu$. We consider the family made of a single element:

$$\mathcal{B}_h(\mathbf{x}) = \{\{h^2 Dg\}\}$$

We then have, for each point where u is C^2:

$$(S_h u)(\mathbf{x}) = \inf_{B \in \mathcal{B}_h(\mathbf{x})} \sup_{\mathbf{y} \in B} u(\mathbf{x}+\mathbf{y}) = u(\mathbf{x} + h^2 Dg) = u(\mathbf{x}) + h^2 Dg(\mathbf{x}) Du(\mathbf{x}) + O(h^4)|D^2 u(\mathbf{x})|$$

The consistency is in fact uniform on any compacts K where u is C^2 ; there exists a $O(h^4, K)$ that does not depend on $\mathbf{x}$ where

$$(S_h u)(\mathbf{x}) - u(\mathbf{x}) = -h^2 Dg(\mathbf{x}) Du(\mathbf{x}) + O_K(h^4)$$

Approximation of $g|Du|curv(u)$. We consider the structuring elements of the median filter (See Chapter 10):

$$\mathcal{B}'_h(\mathbf{x}) = \mathcal{B}'_{6g(\mathbf{x})h} = \{B; B \subset B(0, 6g(\mathbf{x})h) \text{ and } meas(B) \geq \pi(6g(\mathbf{x})h)^2/2\}$$

We set

$$(S'_h u)(\mathbf{x}) = \sup_{B \in \mathcal{B}'_h(\mathbf{x})} \inf_{\mathbf{y} \in B} u(\mathbf{x}+\mathbf{y})$$

Thanks to Theorem 11.3, we immediately have, on any compact K where u is C^2 and where $|Du| \neq 0$,

$$(S'_h u)(\mathbf{x}) - u(\mathbf{x}) = h^2 g(\mathbf{x})|Du(\mathbf{x})|curv(u)(\mathbf{x}) + O_K(h^3)$$

and on any compact K of $\mathbb{R}^2$,

$$(S'_h u)(\mathbf{x}) - u(\mathbf{x}) = O_K(h^2)|D^2 u(\mathbf{x})|$$

Alternating the two filters. We set $T_h = S_h S'_h$. S'_h is an inf-sup operator, with uniform consistency. In addition, its asymptotic is continuous, that is that we have

$$(S'_h u)(\mathbf{x}) - u(\mathbf{x}) = h^2 g(\mathbf{x})|Du(\mathbf{x})|curv(u)(\mathbf{x}) = h^2 F(D^2 u, Du, \mathbf{x})$$

for a continuous function F with respect to its arguments. (g is assumed to be derivable). So that, the Lemma 15.5 insures that we have for any compact set K where $|Du(\mathbf{x})| \neq 0$

$$(T_h u)(\mathbf{x}) - u(\mathbf{x}) = h^2(g(\mathbf{x})Du(\mathbf{x})curv(u)(\mathbf{x}) + Dg(\mathbf{x})Du(\mathbf{x})) + O_K(h^3)$$

And on any compact K:

$$(T_h u)(\mathbf{x}) - u(\mathbf{x}) = O_K(h^2)|D^2 u(\mathbf{x})|$$

As consequence the filter T_h is uniformly consistent (see Definition 17.16) with equation 19.3.

Construction of the approximate solutions. We consider an K-Lipschitz initial function u_0. We then define $u_h(t,\mathbf{x})$ for every $h>0$ by

$$\forall n \in \mathbb{N}, \quad u_h(nh,\mathbf{x}) = (T_h^n u_0)(\mathbf{x}),$$

$$\forall t \in [(n-1)h, nh[,\ u_h(t,\mathbf{x}) = u_h((n-1)h,\mathbf{x}).$$

Since S_h and $S'(h)$ are "inf-sup" operator they transform any K-Lipschitz function into a K-Lipschitz function (See Remark 7.4). We therefore have that for all t, $u_h(t,\mathbf{x})$ is K-Lipschitz.

Uniform continuity in t of the approximate solutions. Let us bound operator T_h by two isotropic and translation invariant operators. Let C_1 be the upper bound of g and C_2 the upper bound of $|Dg|$. For any K-Lipschitz function u, one has

$$u(\mathbf{x}) - Kh^2|Dg| \leq (S_h u)(\mathbf{x}) = u(\mathbf{x} + h^2 Dg) \leq u(\mathbf{x}) + Kh^2|Dg|$$

$$u(\mathbf{x}) - Kh^2 C_2 \leq (S_h u)(\mathbf{x}) \leq u(\mathbf{x}) + Kh^2 C_2$$

Then due to the fact that $S'_h(u+c) = S'_h(u) + c$ for any constant c, we also have

$$(S'_h u) - Kh^2 C_2 \leq T_h u = S'_h S_h u \leq (S'_h u) + Kh^2 C_2 \tag{19.4}$$

Let us consider $v(\mathbf{y}) = K|\mathbf{x}-\mathbf{y}|$. The family $\mathcal{B}'_h(\mathbf{x})$ of structuring elements of the filter S'_h is made of the subsets of the disk of center 0 and radius $6g(\mathbf{x})h \leq 6C_1 h$. For any $B \in \mathcal{B}'_{6g(\mathbf{x})h}$ there exists $B' \in \mathcal{B}'_{6C_1 h}$, such that

$$\inf_{\mathbf{y}\in B} v(\mathbf{x}-\mathbf{y}) = \inf_{\mathbf{y}\in B'} v(\mathbf{x}-\mathbf{y})$$

So that

$$(S'_h v)(\mathbf{x}) = \sup_{B\in\mathcal{B}'_{6g(\mathbf{x})h}} \inf_{\mathbf{y}\in B} v(\mathbf{x}-\mathbf{y}) \leq \sup_{B\in\mathcal{B}'_{6C_1 h}} \inf_{\mathbf{y}\in B'} v(\mathbf{x}-\mathbf{y})$$

That is

$$\forall \mathbf{x},\ (S'_h v)(\mathbf{x}) \leq (M_{6C_1 h} v)(\mathbf{x}) \tag{19.5}$$

with M the median filter, as defined in Chapter 10. Similarly, for $w(\mathbf{y}) = -K|\mathbf{x}-\mathbf{y}|$, we have

$$\forall \mathbf{x},\ (M_{6C_1 h} w)(\mathbf{x}) \leq (S'_h w)(\mathbf{x}) \tag{19.6}$$

We deduce from (19.5), (19.6) and (19.4) the following inequalities

$$(T_h v)(\mathbf{x}) \leq (M_{6C_1 h} v)(\mathbf{x}) + Kh^2 C_2 \qquad (M_{6C_1 h} w)(\mathbf{x}) - Kh^2 C_2 \leq (T_h w)(\mathbf{x}) \tag{19.7}$$

By monotony of T_h and of the median operator, we thus have for all $n \in \mathbb{N}$, and for all $\mathbf{x} \in \mathbb{R}^2$,

$$(T_h^n v)(\mathbf{x}) \leq (M^n_{6C_1 h} v)(\mathbf{x}) + nKh^2 C_2 \qquad (M^n_{6C_1 h} w)(\mathbf{x}) - nKh^2 C_2 \leq (T_h^n w)(\mathbf{x}) \tag{19.8}$$

Now, since u_0 is Lipschitz, one has

$$u(\mathbf{x}) - K|\mathbf{x}-\mathbf{y}| \leq u(\mathbf{y}) \leq u(\mathbf{x}) - K|\mathbf{x}-\mathbf{y}|$$

So

$$u(\mathbf{x}) + (T_h^n w)(\mathbf{x}) \leq (T_h^n u)(\mathbf{x}) \leq u(\mathbf{x}) + (T_h^n v)(\mathbf{x})$$

Using (19.8), we obtain

$$(M_{6C_1h}^n w)(\mathbf{x}) - nKh^2C_2 \leq (T_h^n u)(\mathbf{x}) - u(\mathbf{x}) \leq (M_{6C_1h}^n v)(\mathbf{x}) + nKh^2C_2$$

Lemma 18.6 tells us that for h small enough and for $nh^2 \leq t$, one has

$$(M_{6C_1h}^n v)(\mathbf{x}) \leq L\sqrt{t}$$

for some constant L depending only on C_1. Similar inequality obviously holds for w, so that for h small enough

$$\forall t, \forall n; nh^2 \leq t \qquad -(L\sqrt{t} + KC_2t) \leq T_h^n u - u \leq +(L\sqrt{t} + KC_2t)$$

This yields the uniform continuity in t of the approximate solutions.

Convergence of the approximate solutions toward viscosity solutions. The operator T_h is monotone (and therefore satisfies the uniform local maximum principle), is uniformly consistent to (19.3) and commutes with the addition of constant. Its associated approximate solutions $h \to u_h(t, \mathbf{x})$ is uniformly in h Lipschitz in $\mathbf{x}$ and Holderian in t for any initial Lipschitz function. So a sub-sequence of the sequence $h \to u_h$ is uniformly converging on every compact set towards a function $u(t, \mathbf{x})$. Thus, by Proposition (17.16), we obtain that u is a viscosity solution of (19.3). In others words, we get the existence of a viscosity solution for any initial Lipschitz function, and that the iteration of the operator T_h converges toward this solution. We therefore have prove:

Theorem 19.2 *We consider a derivable function g from $\mathbb{R}^2$ into $\mathbb{R}^+$ satisfying*

$$\exists C_1, C_2, \text{such that } \forall \mathbf{x}, 0 \leq g(\mathbf{x}) \leq C_1 \text{ and } |Dg(\mathbf{x})| \leq C_2$$

Then for any Lipschitz function u_0, there exists an unique viscosity solution $u(t, \mathbf{x})$ of

$$\frac{\partial u}{\partial t} = F(D^2u, Du, \mathbf{x}, t) = g|Du|curv(u) + Dg.Du \qquad u(0, \mathbf{x}) = u_0(\mathbf{x})$$

In addition, $u(t, \mathbf{x})$ is Lipschitz in $\mathbf{x}$ and holderian in t.

Moreover, when h tends to 0 and $nh^2 \to t$, $(T_h^n u_0)(\mathbf{x})$ converges towards $u(t, \mathbf{x})$.

19.3 Curve evolution and Image evolution

Assume that $u(t, \mathbf{x})$ viscosity solution of (19.3) is C^2 around point $(t_0, \mathbf{x}_0)$ and $Du(t_0, \mathbf{x}_0) \neq 0$. Then, for some $t_1 > t_0$, there exists $\mathbf{x}(t)$ a unique C^1 function satisfying for $t \in [t_0, t_1]$, $u(t, \mathbf{x}(t)) = u(t_0, \mathbf{x}_0)$ and such that $\frac{\partial \mathbf{x}(t)}{\partial t}$ is colinear to $Du(t, \mathbf{x})$. $\mathbf{x}(t)$ is the normal flow (see Definition (6.6)).

$$\frac{\partial \mathbf{x}}{\partial t} = -(\frac{\partial u}{\partial t}\frac{Du}{|Du|^2})(t, \mathbf{x}(t)) = g\text{curv}(u)\mathbf{n} - (Dg\mathbf{n})\mathbf{n} \tag{19.9}$$

where $\mathbf{n}$ is the gradient of u direction, that is the normal of the level line passing through the considered point.

Parametrizing the level line of level $u(t_0, \mathbf{x}_0)$ around point $(t_0, \mathbf{x}_0)$ by $\mathbf{x}(t,s)$ so that $\mathbf{x}(t,0) = \mathbf{x}(t)$. We deduce that $\mathbf{x}(t,s)$ satisfies the Active contour evolution equation (19.2). Therefore the following proposition holds:

Proposition 19.3 *Let $u(t,\mathbf{x})$ a viscosity solution of*

$$\frac{\partial u}{\partial t} = F(D^2u, Du, \mathbf{x}, t) = g|Du|curv(u) + Dg.Du$$

Assume that u is C^2 around point $(t_0, \mathbf{x}_0)$, and that $Du(t_0, \mathbf{x}) \neq 0$. Then, the level line $\mathbf{x}(t,s)$ passing through this point satisfies locally around the point:

$$\frac{\partial \mathbf{x}(t,s)}{\partial t} = g(\mathbf{x}(t,s))Curv(\mathbf{x}(t,s))\mathbf{n}(t,s) - (Dg(\mathbf{x}(t,s)).\mathbf{n})\mathbf{n}$$

The curve evolution equation and the image evolution equation are so strongly related.

We now consider the operator T_t which associates to any Lipschitz function u_0, $u(t,.)$ the viscosity solution of (19.3) with initial condition u_0. It is clearly a monotone operator. According to Lemma 18.2, T_t is also constrast invariant.

Proposition 19.4 *The monotone and contrast invariant image operator T_t, defined on Lipschitz function, defines a unique set operator $\mathtt{T}_t$ on defined on the set of the compact sets. We have in addition, for any Lipschitz function u:*

$$\mathcal{X}_\lambda T_t(u) = \mathtt{T}_t(\mathcal{X}_\lambda u)$$

Proof Let us use the Evans-Spruck extension (see Section 8.3). We consider the family of functions $\mathcal{F}$ made of the Lipschitz functions on which we apply any continuous and increasing change of contrast. The family $\mathcal{F}$ is obviously stable under contrast change. The levelsets of the functions of $\mathcal{F}$ is the set of the compacts. Since T_t is contrast invariant and is defined on Lipschitz functions, it is also defined and a monotone contrast invariant operator on $\mathcal{F}$. According to the Evans-Spruck Theorem (see Section 8.3), we know that it defines an unique monotone set operator $\mathtt{T}$ such that

$$\mathcal{X}_\lambda T_t(u) = \mathtt{T}_t(\mathcal{X}_\lambda(u))$$

This means that T_t modifies the level sets of u independantly from each others. □

The snake algorithm...

Given a closed curve $C = \mathbf{c}(s)$. The curve C surrounds a compact set X of $\mathbb{R}^2$. We define the generalized "curve" evolution of C by the following algorithm:

Step 1 We construct a function u_0 so that:

- $\mathcal{X}_0 u_0 = X$
- u_0 is Lipschitz or differ from a Lipschitz function by a continuous and increasing contrast change.

Such a function u can e.g. be obtained by considering the signed distance function to the set X. That is $|u(\mathbf{x})| = \pm dist(\mathbf{x}, C)$.

Step 2 We construct the viscotity solution $u(t,\mathbf{x})$ of equation (19.3) with initial condition u_0.

Step 3 We set $X(t) = \mathcal{X}_0 u(t,.)$ and $C(t) = X(t) \cap \bar{X(t)}^c$. (That is $C(t)$ is the boundary of $X(t)$.)

According to all preceding considerations, the algorithm is valid. That is it defines for any curve C and for any $t \leq 0$ an unique set of points $C(t)$. $C(t)$ is independant from the choice of the intermediate function u. And it corresponds to a generalization of the equation (19.2). However, the so defined $C(t)$ is not necessarly a curve of $\mathbb{R}^2$. It is therefore difficult to check if the energy estimated on $C(t)$ is decreasing.

Conjecture

We conjecture that $\forall u$ Lipschitz,

$$E_u(t) = \limsup_{\epsilon \to 0} \int_{\mathbf{X} \in \mathbb{R}^2, u(t,\mathbf{X}) \in]-\epsilon,\epsilon[} g(\mathbf{x}) d\mathbf{x}$$

is a decreasing function of t. We conjecture also that for any analytic functions g,

$$E(t) = \int_{\mathbf{X} \in C(t)} g(\mathbf{x}) d\mathcal{H}^1$$

is finite and is a decreasing function of t.

19.4 Implementation.

We consider an image I. We choose the function g decreasing with respect to the magnitude of the gradient of I. That is g large where the gradient of I is small, and conversely. If one expects to have large gradient at the edges of I, then one can consider that g is small one the edges of I.

Let us now first make a simple heuristical study of the equation.

$$\frac{\partial u}{\partial t} = g|Du|curv(u) + Dg.Du$$

The first term is the well known mean curvature motion. As we have seen, its tends to shrink the level line towards points. The speed of this motion is related to the amplitude of g. At edge, g is small and thus the motion is slow down, but do not stop.

The second term is the erosion term. It tends to move the level lines towards the edges (see figure **??**), creating, by that, shocks around edges. However, in contrary to the first term, on flat zones it is inactive. Even worse, due to noise, little gradient of I will induces non negligeable variation of amplitude of g, resulting in non negligeable Dg with random like direction. In others words, on flat zone, one can expect to have creation of small shocks, with random shapes.

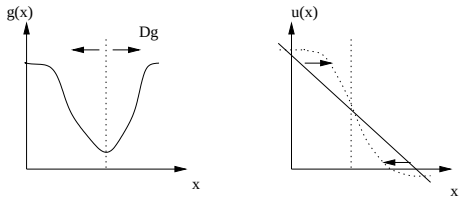

Figure 19.1: Convection term of the active contour equation. The convection term of the active contour equation tends to create shocks aroung values where g is small. Indeed, the level lines of u are moved in the direction opposite to the gradient of g.

More precisely, we have:

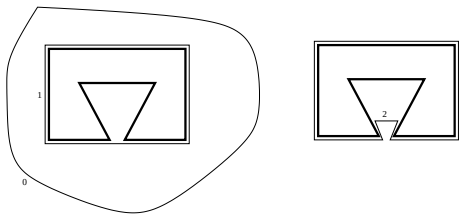

Figure 19.2: A difficulty : the local minima of the active contour energy. Left image: assume that g is null on the shape (drawn in bold) and that the initial contour is the line #0. From the initial line to the contour of the shape, the intermediate state #1 consisting of the convex hull of the polygon shows a smaller energy than the intermediate state #2 (drawn on the right). This illustrates the difficulties arising with such an equation when we wish to land the active contour onto concave parts of the desired contour.

near an edge The first term moves the level line slowly across the edge. The second term moves the level line towards the edge. Their effect are opposit and therefore a balance between the two terms needs to be found so that the second term wins.

far from an edge The first term moves the level line fastly. The second term attract the level line towards little oscillation of the image creating little shocks. Here again, a balance between the two terms needs to be found so that the first term wins.

Even if the equation, does not show any scale parameter between the two terms, a weight between the two terms is in fact hidden in the choice of the function g. Finding a function g that makes the correct balance both near an edge and far an edge is somehow complex and in anycase non-generic.

Assuming such a balance could be find, relying on the single (weighted) mean curvature motion to shrink the level line is not good. Indeed, a weighted mean curvature motion will never modify a convex level line to a general non-convex one. That is starting with a circle, it is impossible to recover e.g. a star, see Figure (19.2).

To cope this problem, we have added, in our experiments, an extra term to the equation, yielding the balance between now three terms more complex to fix. This term is a classical erosion with the same weight than than the curvature motion. So that we have applied

$$\frac{\partial u}{\partial t} = g|Du|(curv(u) + 1) + Dg.Du$$

We have chosen for the function g: $g(\mathbf{x}) = \sigma^2/(\sigma^2 + |DI(\mathbf{x})|^2)$, with σ^2 the estimated variance of the noise and texture around the object. So that g is between 0 and 1. Our scheme is the following:

$$u^{n+1}(\mathbf{x}) = u^n(\mathbf{x}) + dtg(\mathbf{x})(E_1(u^n)(\mathbf{x}) + M_2(u^n)(\mathbf{x}) - u^n(\mathbf{x})) + dt|Dg(\mathbf{x})|(E_1(u^n)(\mathbf{x} - Dg(\mathbf{x})) - u^n(\mathbf{x}))$$

Where E denotes the erosion, and M the median filter. The scheme is not contrast invariant, but satisfies the maximum principle, provided $dt|Dg(\mathbf{x})| \leq 1$ and $dtg(\mathbf{x}) \leq 1$. Figure **??** illustrates the extraction of the bird shape on a textured background.

Exercise 19.1 *Construction of another inf-sup scheme converging towards viscosity solution of the equation (19.3). We consider the following family of structuring elements:*

$$I\!B_h = \{B; B \subset B(+Dgh^2, h) \text{ and } meas(B) \geq \pi h^2/2\}$$

Figure 19.3: Silhouette of a bird by active contour. Left: original image, middle: initial contour, right: final contour (steady state of the snake equation).

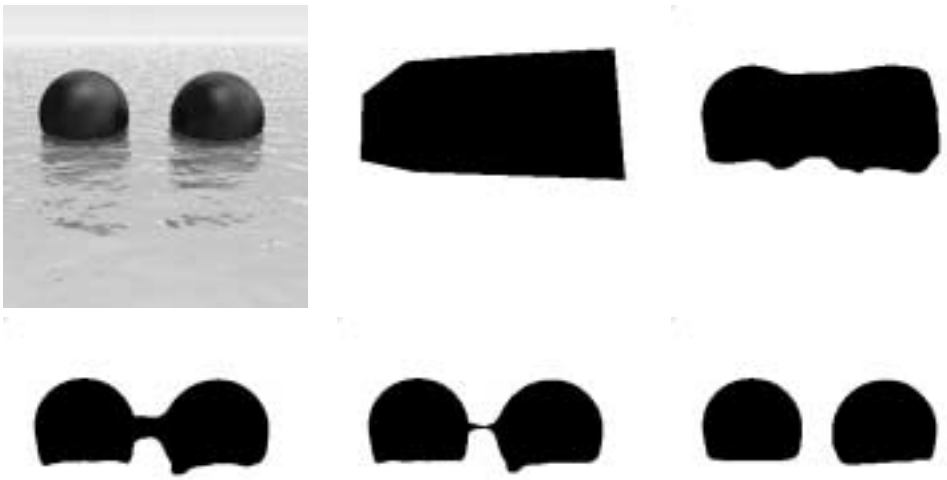

Figure 19.4: Active contour with topological change. Top, left: original image, middle: initial contour, right: intermediate state. Down, left and middle: successive intermediate states, down-right: final contour (steady state). Original image is "Vue d'esprit 3", by courtesy of *e-on software*.

And we define

$$T_h u(\mathbf{x}) = \inf_{B \in \mathcal{B}} \sup_{\mathbf{y} \in B} u(\mathbf{x} + \mathbf{y})$$

1. *Interpret the operator T_h as a shifted median filter.*
2. *Show that T_h is uniformly consistent with equation (19.3).*
3. *Show that the iteration of T_h yields towards viscosity solutions of (19.3).*

References.

The variational definition of "snakes", or "active contours" is due to the seminal paper of Kass, Witkin and Terzopoulos [**?**]. Many papers have proposed improvements and attempts to overcome the very numerous numerical difficulties of this first method. See (e.g.) [**?**]. Caselles, Catté, Coll and Dibos pointed out these difficulties and their mathematical explanations [78, 125]. They were the first to propose a (nonvariational) definition permitting to model an active contour by mean curvature motion. [276] proposed some numerical improvements. A review is presented in [77]. The fusion of both numerical models is achieved in [87]. For this last definition, the so called "geodesic snake" moves, by curve shortening, towards a geodesic of an image dependent surface. For a general mathematical study of curve shortening on surfaces, see the Angenent papers [31, 32]. Kimmel [244] proposes an image scale space on surfaces based on the same principle. Many extensions of the snakes method to deal with images in motion [81] or medical images are currently being developped.

Part V

Axiomatic Scale-Space Theory

Chapter 20

Scale spaces and Partial Differential Equations

20.1 What basic principles must obey a scale space?

In this chapter, we introduce an abstract framework, the "scale space", which at the end boils down, from the algorithmic viewpoint, to iterated filtering. This framework will make it easier to classify and model the possible asymptotic behaviors of iterated filtering. We define a "scale space" as an abstract family of image smoothing operators T_t, depending on a scale parameter t. Given an image $u_0(\mathbf{x})$, $(T_t u_0)(\mathbf{x}) = u(t, \mathbf{x})$ is the "image u_0 analyzed (in fact : smoothed) at scale t". Formal, but natural and classical, assumptions on the smoothing process follow.

The pyramidal structure. We assume that the output at scale t can be computed from the output at a scale $t - h$ for very small h. This is natural, since a coarser view of the original picture is likely to be deduced from a finer one without any dependence upon the original picture. By that way the finest picture smoothing is the identity. T_t is obtained by composition of "transition filters", which we denote by $T_{t+h,t}$.

Definition 20.1 *We shall say that a scale space is* **pyramidal** *if*

$$T_{t+h} = T_{t+h,t} T_t, \quad T_0 = Id. \tag{20.1}$$

A strong version of the pyramidal structure assumption is the semigroup property

Definition 20.2 *We shall say that a scale space is* **recursive** *if*

$$T_0 = Id, \qquad T_s \circ T_t(u) = T_{s+t}(u) \text{ on } \mathbb{R}^N, \text{ for all } s, t \geq 0 \text{ and } u \text{ in } \mathcal{F}. \tag{20.2}$$

If the recursivity is satisfied, T_t can be deduced from the n-times iteration of $T_{t/n}$. Let us continue with an intuitive requirement which is called in image processing "causality". Since the visual pyramid is assumed to yield more and more global information about the image and its features, it is clear that when the scale increases, no new feature should be created by the scale space : the image at scale t'>t must be simpler than the image at scale t. Furthermore, the operators (or transition filters) $T_{t+h,t}$ are assumed

Figure 20.1: A multiscaled world ... This series of images is an experiment in the relative stability of perception of objects seen at different distances. Each photograph has been taken in a park by stepping forward and taking each snapshot at a much closer distance than the former one. We display by a rectangle in each image the part of the object which has been photographed in the next image. As one can appreciate, when getting closer, the visual aspect of objects changes and new structures arise. Thus, the computing of primitives in an image is always a scale-dependent task, quite dependent on the distance to objects. When we look at a certain distance at an object, we do not perceive the too much finer structure : for instance, leaves cannot be seen in the two first photographs. We do not see them either in the two last ones, since we got too close ! Multiscale smoothing tries to emulate this physical and perceptual fact by defining a smoothing of a digital image at different scales which gets rid of the finer structures, but minimally modifies the image at scales above the desired scale.

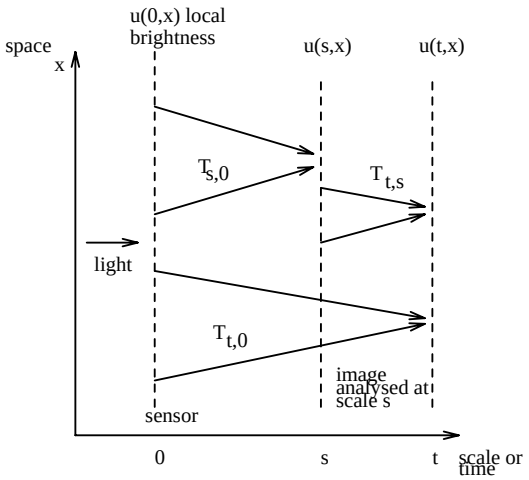

Figure 20.2: The scale-space visual pyramid : perception is thought of as a flow of images passing through transition operators $T_{t,s}$. These operators receive an image previously analysed at scale s and deliver an image analysed at a larger scale t. The scale $t = 0$ corresponds to the original percept. In this simple model, the perception process is irreversible (no feedback from coarse scales to fine scales).

to act "locally", that is, to look at a small part of the processed image. In other terms, $(T_{t+h,t}u_0)(\mathbf{x})$ must essentially depend upon the values of $u_0(\mathbf{y})$ when $\mathbf{y}$ lies in a small neighborhood of $\mathbf{x}$. We condense the locality and the simplification assumptions in a *local comparison principle* : if an image u is locally brighter than another v, then this order must be conserved some time by the analysis (prevalence of local order on global order).

Definition 20.3 *A scale space satisfies the* **Local Comparison Principle** *, if for all u and v such that $u(\mathbf{y}) > v(\mathbf{y})$ for $\mathbf{y}$ in a neighborhood of $\mathbf{x}$ and $\mathbf{y} \neq \mathbf{x}$, then for h small enough we have*

$$(T_{t+h,t}u)(\mathbf{x}) \geq (T_{t+h,t}v)(\mathbf{x})$$

and if for all u and v such that $\forall \mathbf{y} \in I\!R^N$, $u(\mathbf{y}) \geq v(\mathbf{y})$, then

$$\forall \mathbf{x}, \forall h > 0, \qquad (T_{t+h,t}u)(\mathbf{x}) \geq (T_{t+h,t}v)(\mathbf{x}). \tag{20.3}$$

In order to propose a classification of scale spaces, we finally need some assumption stating that a very smooth image must evolve in a smooth way with the scale space. As we shall prove, it is enough to make this assumption for quadratic functions. From the mathematical viewpoint, the next assumption implies the existence of an infinitesimal generator for the semigroup T_t (see [**?**], [**?**]). In this chapter, we shall denote for clarity the scalar product between two vectors of $I\!R^N$ by $\prec \mathbf{x}, \mathbf{y} \succ = \mathbf{x}.\mathbf{y} = \sum_1^N x_i y_i$.

Definition 20.4 *Let $u(\mathbf{y}) = \frac{1}{2} \prec A(\mathbf{y} - \mathbf{x}), \mathbf{y} - \mathbf{x} \succ + \prec p, \mathbf{y} - \mathbf{x} \succ + c$ be a quadratic form of $I\!R^N$. ($A = D^2u(\mathbf{x})$ is a $n * n$ matrix, $p = Du(\mathbf{x})$ a vector of $I\!R^N$, $c = u(\mathbf{x})$ a constant.)*

We shall say that a scale space is **regular** *if there exists a function $F(A, p, \mathbf{x}, c, t)$, continuous with respect to A, such that*

$$\frac{(T_{t+h,t}u - u)(\mathbf{x})}{h} \to F(A, p, \mathbf{x}, c, t) \quad \textit{when} \quad h \to 0. \tag{20.4}$$

Since we shall in the following constantly assume that the considered scale spaces are pyramidal, regular, and satisfy the local comparison principle, it is opportune for this set of basic properties to be given a single name.

Definition 20.5 *We call* **causal** *any scale space which satisfies the* **Local Comparison Principle** *and is pyramidal and regular.*

Exercise 20.1 *Consider the extrema killer operator defined in Section 7.4, T_t where t denotes the area threshold. Show that the family $(T_t)_{t\in\mathbb{R}^+}$ is pyramidal, satisfies the global comparison principle, but not the local comparison principle. Show that the family is, however, regular at $t = 0$ and more precisely,*

$$F(A, p, 0) = 0, \text{ if } p \neq 0.$$

As we shall prove in the next section, the causality assumption is enough to imply that the scale space is governed by a partial differential equation.

20.2 Why Scale Spaces are governed by PDE's

Theorem 20.6 *If a scale space T_t is causal, (that is, pyramidal, regular and satisfying the local comparison principle), then there exists a function F such that*

$$((T_{t+h,t}u - u)/h)(\mathbf{x}) \to F(D^2u(\mathbf{x}), Du(\mathbf{x}), u(\mathbf{x}), \mathbf{x}, t) \tag{20.5}$$

as h tends to 0^+ for all u and $\mathbf{x}$ where u is C^2. In addition, F is a continuous and non-decreasing function with respect to its first argument, that is

$$\text{If } A \geq \tilde{A}, \quad \text{for the ordering of symmetric matrices }, \tag{20.6}$$

$$\text{then,} \quad F(A, p, c, \mathbf{x}, t) \geq F(\tilde{A}, p, c, \mathbf{x}, t)).$$

Once this theorem is proved, the classification of all iterated nonlinear low pass filtering (or, in other words, of all scale spaces) will be reduced to the classification of all interesting functions F. In dimension 2, these real functions have nine arguments. This number, however, will be drastically reduced when we impose obvious and rather necessary and useful invariance properties to the associated scale space. After the proof, we shall list these properties and give the resulting classification of "interesting" functions F.

Proof of Theorem 20.6 Let $w(\mathbf{x}) = \frac{1}{2} \prec A\mathbf{x}, \mathbf{x}\succ + \prec p, \mathbf{x}\succ + c$ be a quadratic form of $\mathbb{R}^N$. We have assumed, by regularity, that $(T_{t+h,t}w - w)(\mathbf{x})/h$ converges towards $F(A, p, \mathbf{x}, c, t)$ when h tends to zero, for a function F, continuous with respect to its first argument. Let u be a function of $\mathbb{R}^N$ into $\mathbb{R}$ such that u is C^2 at point $\mathbf{x}$. We shall prove that (20.5) is true for u at point $\mathbf{x}$ for the same function F introduced in the regularity assumption. Since u is C^2 at $\mathbf{x}$, we write

$$u(\mathbf{y}) = u(\mathbf{x}) + \prec Du, \mathbf{y} - \mathbf{x}\succ + \frac{1}{2} \prec (D^2u)(\mathbf{y} - \mathbf{x}), \mathbf{y} - \mathbf{x}\succ + o(|\mathbf{x} - \mathbf{y}|^2),$$

where Du and D^2u mean $Du(\mathbf{x})$ and $D^2u(\mathbf{x})$ respectively. We define, for $\varepsilon > 0$, the quadratic forms

$$Q^+(\mathbf{y}) = u(\mathbf{x}) + \prec Du, \mathbf{y}-\mathbf{x}\succ + \frac{1}{2}\prec (D^2u)(\mathbf{y}-\mathbf{x}), \mathbf{y}-\mathbf{x}\succ + \frac{1}{2}\epsilon(\mathbf{x}-\mathbf{y})^2$$

$$Q^-(\mathbf{y}) = u(\mathbf{x}) + \prec Du, \mathbf{y}-\mathbf{x}\succ + \frac{1}{2}\prec (D^2u)(\mathbf{y}-\mathbf{x}), \mathbf{y}-\mathbf{x}\succ - \frac{1}{2}\epsilon(\mathbf{x}-\mathbf{y})^2$$

When $\mathbf{y} \neq \mathbf{x}$ and $|\mathbf{y}-\mathbf{x}|$ is small enough,

$$Q^-(\mathbf{y}) < u(\mathbf{y}) < Q^+(\mathbf{y})$$

Then using the local comparison principle (Definition 17.18),

$$T_{t+h,t}Q^-(\mathbf{x}) < T_{t+h,t}u(\mathbf{x}) < T_{t+h,t}Q^+(\mathbf{x}).$$

Therefore, using $Q^+(\mathbf{x}) = u(\mathbf{x}) = Q^-(\mathbf{x})$,

$$(T_{t+h,t}Q^-)(\mathbf{x}) - Q^-(\mathbf{x}) < (T_{t+h,t}u)(\mathbf{x}) - u(\mathbf{x}) < (T_{t+h,t}Q^+)(\mathbf{x}) - Q^+(\mathbf{x}).$$

We divide by h, and let h tend to zero. Since T_t is regular, we obtain

$$\lim_{h\to 0}\frac{(T_{t+h,t}Q^-)(\mathbf{x}) - Q^-(\mathbf{x})}{h} \leq \liminf_{h\to 0}\frac{(T_{t+h,t}u)(\mathbf{x}) - u(\mathbf{x})}{h} \leq$$

$$\limsup_{h\to 0}\frac{(T_{t+h,t}u)(\mathbf{x}) - u(\mathbf{x})}{h} \leq \lim_{h\to 0}\frac{(T_{t+h,t}Q^+)(\mathbf{y}) - Q^+(\mathbf{y})}{h}$$

Thus

$$F(D^2u - \epsilon Id, Du, u, \mathbf{x}, t) \leq \liminf_{h\to 0}\frac{(T_{t+h,t}u)(\mathbf{x}) - u(\mathbf{x})}{h} \leq$$

$$\limsup_{h\to 0}\frac{T_{t+h,t}(u(.))(\mathbf{x}) - u(\mathbf{x})}{h} \leq F(D^2u + \epsilon Id, Du, u, \mathbf{x}, t)$$

When ϵ tends to zero, using again the fact that T_t is regular (continuity of F with respect to its first argument), we obtain

$$\lim_{h\to 0}\frac{(T_{t+h,t}u)(\mathbf{x}) - u(\mathbf{x})}{h} = F(D^2u, Du, u, \mathbf{x}, t)$$

We now prove that F is nondecreasing with respect to its first argument. Setting now for any vector p and any matrices $A \geq \tilde{A}$,

$$Q^+(\mathbf{y}) = c + \prec p, \mathbf{y}-\mathbf{x}\succ + \frac{1}{2}\prec A(\mathbf{y}-\mathbf{x}), \mathbf{y}-\mathbf{x}\succ + \frac{1}{2}\epsilon(\mathbf{x}-\mathbf{y})^2,$$

$$Q^-(\mathbf{y}) = c + \prec p, \mathbf{y}-\mathbf{x}\succ + \frac{1}{2}\prec \tilde{A}(\mathbf{y}-\mathbf{x}), \mathbf{y}-\mathbf{x}\succ - \frac{1}{2}\epsilon(\mathbf{x}-\mathbf{y})^2,$$

and applying, as above, the local comparison principle to Q^+ and Q^- around $\mathbf{x}$, the regularity and letting $\varepsilon \to 0$, we obtain $F(A, p, c, \mathbf{x}, t) \geq F(\tilde{A}, p, c, \mathbf{x}, t)$. □

In the next section, we shall start drawing consequences from askable invariance properties for a scale space. We shall immediately impose, and always assume in the following, a basic one. Any smoothing process should leave untouched the constant functions and, a little more generally, should not be altered by the addition of a constant.

Definition 20.7 *We shall say that a scale space $T_{t,t+h}$ is* **invariant by grey level translation** *if*

$$T_{t,t+h}(0) = 0, T_{t,t+h}(u + C) = T_{t,t+h}(u) + C \text{ for any } u \text{ and any constant } C. \tag{20.7}$$

This axiom is equivalent, in the case where $T_{t,t+h}$ is a linear filter, defined by $T_{t,t+h}u = u * \phi$, to

$$\int \phi(x)dx = 1.$$

Let us state the consequence on F.

Proposition 20.8 *Let T_t be a causal scale space, invariant by grey level translation. Then its associated function $F(A, p, \mathbf{x}, c, t)$ does not depend on c.*

Proof Let C be an arbitrary real number. Consider the quadratic form $u(\mathbf{y}) = \frac{1}{2} \prec A(\mathbf{y} - \mathbf{x}), \mathbf{y} - \mathbf{x} \succ + \prec p, \mathbf{y} - \mathbf{x} \succ + c$. By the regularity assumption (20.4),

$$\frac{(T_{t+h,t}(u + C) - (u + C))(\mathbf{x})}{h} \to F(A, p, \mathbf{x}, c + C, t) \quad \text{when} \quad h \to 0. \tag{20.8}$$

Using the grey level translation invariance (20.7) and again (20.4),

$$\frac{(T_{t+h,t}(u + C) - (u + C))(\mathbf{x})}{h} = \frac{(T_{t+h,t}(u) + C - u - C)(\mathbf{x})}{h} \to F(A, p, \mathbf{x}, c, t) \tag{20.9}$$

when $h \to 0$. Combining (20.8) and (20.9), we obtain

$$F(A, p, \mathbf{x}, c + C, t) = F(A, p, \mathbf{x}, c, t).$$

□

Because of the assumed commutation of T_t with the addition of constants, we shall remove in the following the c-dependence of F.

20.3 Why multiscale analyses compute viscosity solutions.

In this section, we prove that the concept of viscosity solution is the right concept for defining solutions of partial differential equations associated with a scale space. In one sentence, if T_t is a causal scale space then $T_t u_0$ is the viscosity solution of the equation

$$\frac{\partial u}{\partial t} = F(D^2 u, Du, t) \tag{20.10}$$

which is canonically associated with the scale space. To be precise, we shall prove the following theorem.

Theorem 20.9 *Let T_t be a scale space which is causal and commutes with grey level tranlations $u \to u + C$. We also assume that $u(t, \mathbf{x}) = (T_t u_0)(\mathbf{x})$ is a bounded uniformly continuous function on $[0, +\infty] \times I\!R^N$. Assume finally that the function $F(A, p, \mathbf{x}, t)$ associated with T_t by the regularity assumption is continuous with respect to t. Then $u(t, \mathbf{x}) = T_t u_0(\mathbf{x})$ is a viscosity solution of Equation (20.10).*

We have taken the assumption that $u(t, \mathbf{x}) = (T_t u_0)(\mathbf{x})$ is uniformly continuous, i.e.

$$|u(t, \mathbf{x}) - u(s, \mathbf{y})| \leq f((t - t')^2 + (\mathbf{x} - \mathbf{y})^2) \tag{20.11}$$

for some nondecreasing nonnegative continuous real function f.

Exercise 20.2 *Show if (20.11) holds, then there exists another nonnegative nondecreasing continuous function h such that*

$$|u(t,\mathbf{x}) - u(s,\mathbf{y})| \leq h(|t-t'|) + h(|\mathbf{x}-\mathbf{y}|) \tag{20.12}$$

Show that if u is bounded, then we can assume that h is itself uniformly continuous and bounded.

Lemma 20.10 *Let $\tilde{f}(\mathbf{x})$, $\tilde{g}(t)$ be two C^∞ functions and $u(t,\mathbf{x})$ a uniformly continuous function such that, for some $r>0$,*

$$\tilde{f}(\mathbf{x}) + \tilde{g}(t) \leq u(t,\mathbf{x}) \tag{20.13}$$

on $B((t_0,\mathbf{x}_0),r)$. Then there exist two continuous functions $f(\mathbf{x})$ and $g(t)$ which coincide with $\tilde{f}$ and $\tilde{g}$ on $B((t_0,\mathbf{x}_0),r)$ and satisfy a global comparison relation with u,

$$f(\mathbf{x}) + g(t) \leq u(t,\mathbf{x}). \tag{20.14}$$

Proof Without loss of generality, we can assume that $B((t_0,\mathbf{x}_0),r) = B(t_0,r)\times B(\mathbf{x}_0,r)$. By the result of Exercise 20.2, we can assume that (20.12) holds. We then associate with every $t\in I\!R$ its projection $p(t)$ on the interval $B(t_0,r)$ and its distance to the interval, $|t-p(t)|$. In the same way, we associate with every $\mathbf{x}\in I\!R^N$ its projection $p(\mathbf{x})$ on the ball $B(\mathbf{x}_0,r)$ and its distance to the ball, $|\mathbf{x}-p(\mathbf{x})|$. We then simply set

$$g(t) = \tilde{g}(p(t)) - h(|t-p(t)|)$$

and

$$f(\mathbf{x}) = \tilde{f}(p(\mathbf{x})) - h(|\mathbf{x}-p(\mathbf{x})|).$$

Then $f(\mathbf{x})+g(t)$ simply coincides with $\tilde{f}(\mathbf{x})+\tilde{g}(t)$ on $B(t_0,r)\times B(\mathbf{x}_0,r)$. Let us show that (20.14) holds. By the uniform continuity assumption on u,

$$u(t,\mathbf{x}) \geq u(p(t),p(\mathbf{x})) - h(|t-p(t)|) - h(|\mathbf{x}-p(\mathbf{x})|).$$

By (20.13) and the definition of f and g we then obtain

$$u(t\mathbf{x}) \geq \tilde{f}(p(\mathbf{x})) + \tilde{g}(p(t)) - h(|t-p(t)| - h(|\mathbf{x}-p(\mathbf{x})|) = g(t) + f(\mathbf{x}),$$

which yields (20.14). □

Proof of Theorem 20.9 In order to prove that $u(t,\mathbf{x}) = T_t(u_0)(\mathbf{x})$ is a viscosity solution of (20.10), we only to check that u is a viscosity subsolution of (20.10), the supersolution property being shown in the same way. Let $(t_0,\mathbf{x}_0)$ in $[0,\infty]\times I\!R^N$ be a strict local maximum point of $u-\phi$ where ϕ is in $C_b^\infty(I\!R^N\times[0,T])$ for any $T<\infty$. (We denote by C_b^∞ the set of C^∞ bounded functions). We need to show that

$$\frac{\partial\phi}{\partial t}(t_0,\mathbf{x}_0) - F(D^2\phi(t_0,\mathbf{x}_0), D\phi(t_0,\mathbf{x}_0),\mathbf{x}_0,t) \leq 0. \tag{20.15}$$

Without loss of generality, we may assume that $u(t_0,\mathbf{x}_0) = \phi(t_0,\mathbf{x}_0)$. Without loss of generality again, we may assume by Lemma 17.9 that ϕ is of the form $\phi(t,\mathbf{x}) = f(\mathbf{x})+g(t)$ where $g(t_0)=0$, $f(\mathbf{x}_0) = u(t_0,\mathbf{x}_0)$

and f is in $C_b^\infty(I\!R^N)$, g is in $C^\infty([0,\infty[)$. By Lemma 20.10 we can assume (w.l.o.g.) that the inequality between u and ϕ is global, i.e. $u \leq \phi$ on $[0,\infty] \times I\!R^N$. Let $h > 0$. We set $u(t) = u(t,.)$ for all functions u on $I\!R^N * [0,\infty[$. We have $u(t) \leq \phi(t)$, then $u(t_0 - h, \mathbf{x}) \leq \phi(t_0 - h, \mathbf{x})$ for all $h < t_0$ and $\mathbf{x}$. Using the global comparison principle (20.3), we obtain

$$T_{t_0,t_0-h}(u(t_0-h))(\mathbf{x}_0) \leq T_{t_0,t_0-h}(\phi(t_0-h))(\mathbf{x}_0)$$

Now, by the commutation of T_t with the addition of constants and since $\phi(t,\mathbf{x}) = f(\mathbf{x}) + g(t)$ we have

$$T_{t_0,t_0-h}(u(t_0-h))(\mathbf{x}_0) \leq T_{t_0,t_0-h}(\phi(t_0-h))(\mathbf{x}_0) \leq T_{t_0,t_0-h}(f)(\mathbf{x}_0) + g(t_0-h).$$

We deduce that

$$u(t_0,\mathbf{x}_0) - (T_{t_0,t_0-h}(f)(\mathbf{x}_0) + g(t_0-h)) \leq 0$$

Since $u(t_0,\mathbf{x}_0) = \phi(t_0,\mathbf{x}_0)$ and then

$$g(t_0) - g(t_0-h) \leq T_{t_0,t_0-h}(f)(\mathbf{x}_0) - f(\mathbf{x}_0)$$

Then,

$$(g(t_0) - g(t_0-h))/h \leq (T_{t_0,t_0-h}(f)(\mathbf{x}_0) - f(\mathbf{x}_0))/h$$

and letting h go to 0, using (i), we recover

$$\frac{\partial g}{\partial t}(t_0-h) - F(D^2 f(\mathbf{x}_0), Df(\mathbf{x}_0), \mathbf{x}_0, t_0-h) \leq 0$$

Using the facts that ϕ and g are in C_b^∞ in a neighborhood of $(t_0,\mathbf{x}_0)$ and that F is continuous with respect to t, letting h tends to zero, we finally deduce

$$\frac{\partial g}{\partial t}(t_0) - F(D^2 f(\mathbf{x}_0), Df(\mathbf{x}_0), \mathbf{x}_0, t) \leq 0$$

and (iii) is shown since $\partial\phi/\partial t(t_0,\mathbf{x}_0) = g'(t_0)$ and $D^\alpha\phi(t_0,\mathbf{x}_0) = D^\alpha f(\mathbf{x}_0)$ for $|\alpha| = 1,\ 2$. We conclude that $T_t u$ is the unique viscosity solution of (20.10). □

20.4 Scale space and invariance properties

In this section, we shall do a first classification of the admissible functions F, depending on the invariance assumptions made on the associated scale space.

20.4.1 Geometric invariance axioms

We shall first list a series of axioms which traduce the invariance of scale space with respect to the respective positions of the *percipiens* and *perceptum* in the image generation process. This leads us to define more and more precisely which functions F are admissible for a general purpose, and therefore invariant, scale space. We define the translation operator $\tau_{\mathbf{z}}$ by :

$$(\tau_{\mathbf{z}}.u)(\mathbf{x}) = u(\mathbf{x}-\mathbf{z}) \tag{20.16}$$

where $\mathbf{z}$ is in $I\!R^N$.

Definition 20.11 *We shall say that the scale space is* **translation invariant** *if*

$$T_{t+h,t}(\tau_{\mathbf{y}}.u) = \tau_{\mathbf{y}}.(T_{t+h,t}u) \text{ for all } \mathbf{y} \text{ in } I\!R^N,\ t \geq 0,\ h \geq 0. \tag{20.17}$$

Proposition 20.12 *Let T_t be a causal and translation invariant scale space. Then its associated function $F(A, p, \mathbf{x}, c, t)$ does not depend on $\mathbf{x}$.*

Proof Consider the quadratic forms $u(\mathbf{y}) = \frac{1}{2} \prec A\mathbf{y}), \mathbf{y}\succ + \prec p, \mathbf{y}\succ + c$ and $\tau_{\mathbf{x}}u(\mathbf{y}) = \frac{1}{2} \prec A(\mathbf{y}-\mathbf{x}), \mathbf{y}-\mathbf{x}\succ + \prec p, \mathbf{y}-\mathbf{x}\succ + c$. By the regularity assumption (20.4),

$$\frac{(T_{t+h,t}u - u)(0)}{h} \to F(A, p, 0, c, t) \quad \text{when} \quad h \to 0. \tag{20.18}$$

and

$$\frac{(T_{t+h,t}(\tau_{\mathbf{x}}u) - \tau_{\mathbf{x}}u)(\mathbf{x})}{h} \to F(A, p, \mathbf{x}, c, t) \quad \text{when} \quad h \to 0. \tag{20.19}$$

On the other hand, by the translation invariance (20.17),

$$\frac{(T_{t+h,t}(\tau_{\mathbf{x}}u) - \tau_{\mathbf{x}}u)(\mathbf{x})}{h} = \tau_{\mathbf{x}}(\frac{(T_{t+h,t}u - u)(\mathbf{x})}{h}) = \frac{(T_{t+h,t}u - u)(0)}{h} \tag{20.20}$$

Combining (20.18, 20.19, 20.20) we obtain

$$F(A, p, \mathbf{x}, c, t) = F(A, p, 0, c, t).$$

□

If R is an isometry of $I\!R^N$, we denote by Ru the function $Ru(\mathbf{x}) = u(R\mathbf{x})$.

Definition 20.13 *We shall say that a scale space $T_{s,t}$ is* **Euclidean invariant** *if for every isometry R of $I\!R^N$,*

$$RT_{t+h,t} = T_{t+h,t}R \tag{20.21}$$

Lemma 20.14 *If a translation invariant causal scale space is Euclidean invariant (isotropic), then its associated function F satisfies for every linear isometry $R \in O^N$,*

$$F(RA^tR, Rp, t) = F(A, p, t) \tag{20.22}$$

Proof Let us first recall the chain rules :

$$D^2(u \circ R) =^t R(D^2u \circ R)R, \quad D(u \circ R) =^t RDu \circ R. \tag{20.23}$$

We have $T_{t+h,t}R = RT_{t+h,t}$. We apply this relation to a quadratic form u such that $D^2u(\mathbf{x}) = A$, $Du(\mathbf{x}) = p$ and deduce

$$T_{t+h,t}(Ru) = R(T_{t+h,t}u), \text{ that is, } T_{t+h,t}(u \circ R) = (T_{t+h,t}u) \circ R$$

Differentiating this relation with respect to h at $h = 0$, which is licit by the regularity of the scale space, we get

$$\frac{\partial}{\partial h}_{|h=O}(T_{t+h,t}(u \circ R)(x)) = \frac{\partial}{\partial h}_{|h=O}(((T_{t+h,t}u) \circ R)(x)).$$

Using on both sides the regularity formula (20.5) and taking into account that, by the chain rules (20.23), $D^2(Ru) = {}^t RAR$ and $D(Ru) = {}^t Ru$, we obtain

$$F({}^t R(D^2u \circ R)R, {}^t RDu \circ R, t) = F(D^2u, Du, t) \circ R,$$

which yields

$$F({}^t RAR, {}^t Rp, t) = F(A, p, t).$$

Relation (20.22) follows by changing R into ${}^t R$. □

20.4.2 Contrast invariance

Definition 20.15 *We shall say that a scale space is contrast invariant if*

$$g \circ T_{t+h,t} = T_{t+h,t} \circ g, \tag{20.24}$$

for any nondecreasing continuous function g from $\mathbb{R}$ into $\mathbb{R}$.

We now seek for a relation traducing for F the "contrast invariance".

Lemma 20.16 *If a causal scale space is contrast invariant, then its associated function F satisfies*

$$F(\mu A + \lambda p \otimes p, \mu p, t) = \mu F(A, p, t), \tag{20.25}$$

for every real values λ, μ, every symmetric matrix A and every two-dimensional vector p.

By $p \otimes p$ we mean the tensorial product, or in a more intuitive way, the matrix product $p^t p$, where p is thought of as a column matrix, that is

$$p \otimes p = \begin{bmatrix} p_1^2 \cdots & p_1 p_N \\ p_1 p_N \cdots & p_N^2 \end{bmatrix}.$$

Proof By taking $g(s) = 0$, and $g(s) = s + C$, we see that the scale space is invariant by grey level translations. Thus, by Proposition 20.8, $F(A, p, \mathbf{x}, c, t)$ does not depend on c. We have by contrast invariance $T_{t+h,t} \circ g = g \circ T_{t+h,t}$. We apply this relation to an arbitrary quadratic form u such that $D^2u(\mathbf{x}) = A$, $Du(\mathbf{x}) = p$ and to an arbitrary C^2 function g. We deduce that

$$T_{t+h,t}(g \circ u) = g \circ (T_{t+h,t}u).$$

Differentiating this relation with respect to h at $h = 0$ (which is licit by Theorem 20.6), we get

$$\frac{\partial}{\partial h}_{|h=O}(T_{t+h,t}(g \circ u)(\mathbf{x})) = \frac{\partial}{\partial h}_{|h=O}(g \circ (T_{t+h,t}u)(\mathbf{x})). \tag{20.26}$$

Using Formula (20.5) on both sides of (20.26), we deduce that

$$F(D^2(g(u)), D(g(u)), \mathbf{x}, t) = g'(u)F(D^2u, Du, \mathbf{x}, t).$$

By the chain rule,

$$D^2(g(u)) = g'(u)D^2u + g''(u)Du \otimes Du, \quad D(g(u)) = g'(u)Du.$$

Thus, considering that $Du = p$, $D^2u = A$, $h(u)$, $h'(u)$, $h''(u)$ can take arbitrary values, we obtain (20.25). □

20.4.3 Scale invariance : How to fix a relation between scale and space

This subsection is devoted to a study of the effect of the scale invariance axiom, and more generally of the affine invariance on the scale space, in order to establish a normalized link between scale and space. We shall assume, for avoiding spurious cases, that the scale space is not cyclic, that is, $T_t = T_s$ implies $t = s$. This is no serious restriction for a scale space ! In fact, we shall need no other assumption on the family T_t, except for the scale invariance assumptions which follow. So all statements to come will be true for any scale invariant one to one family of operators T_t, which depend upon a scale parameter. The consequences of the scale/space normalization results are therefore wide and can be stated as a general principle, according to which, "the relation between scale and space can be normalized", and more precisely normalized so that $t' = \lambda t$ if we denote by t the scale parameter before zooming with a factor λ and t' the smoothing scale after zooming.

In order to perform such a scale/space normalization, we need to state the scale invariance in a more precise way : we need some technical regularity assumptions on the scale relation. Define $H_\lambda(\mathbf{x}) = \lambda\mathbf{x}$ for any $\mathbf{x} \in I\!R^N$.

Definition 20.17 *We shall say that a family of operators T_t is scale invariant if for any positive λ and t, there exists $t' > 0$ such that $H_\lambda T_{t'} = T_t H_\lambda$ and if, in addition, $t'(t,\lambda)$ is differentiable with respect to λ at $\lambda = 1$, and the function $\phi(t) = \partial t'/\partial\lambda(t,1)$ is continuous and positive for $t > 0$.*

Definition 20.18 *We say that a family of operators T_t is* **scale invariant** *if there exists a rescaling function $t'(t,\lambda)$, defined for any $\lambda > 0$ and $t \geq 0$, such that*

$$H_\lambda T_{t'} = T_t H_\lambda \tag{20.27}$$

and, more generally :

$$H_\lambda T_{t',s'} = T_{t,s} H_\lambda. \tag{20.28}$$

where we note $t' = t'(t,\lambda)$ and $s' = t'(s,\lambda)$. In addition, $t'(t,\lambda)$ is assumed to be differentiable with respect to λ at $\lambda = 1$, and the function $\phi(t) = \partial t'/\partial\lambda(t,1)$ is continuous and positive for $t > 0$.

Remarks. In fact, (20.28) implies (20.27) and must be considered as a slightly stronger form which will make easier our classification of scale invariant scale spaces in the following chapters. Now, as we shall see in Section 20.4.3, (20.27) is enough to ensure a normalization of the function $t'(t,\lambda)$. We cannot fix *a priori* the form of t' because it can vary in concrete examples of scale spaces. We shall prove in Section 20.4.3 that it can be "normalized", which means that we always can, by an adequate rescaling of any scale space, set $t' = \lambda t$.

The scale invariance means that the result of the scale space T_t is independent of the size of the analyzed features : this is very important in "natural images", since the same object can be photographed at very different distances and therefore at very different scales. Thus, it is essential for the stability of shape analysis that the result of an analysis of this object should not yield a different "shapes" at different distances. Thus, the sequence of the shapes obtained by scale space must be independent of the (*a priori* unknown) size of the object in the picture. Of course, we cannot impose $t' = t$ because scale of smoothing and scale of the image are covariant, as can be appreciated by considering the examples of linear filtering and the heat equation.

The assumption $g(t) = \partial t'/\partial\lambda(t,1) > 0$ can be interpreted by looking at the relation $H_\lambda T_{t'} = T_t H_\lambda$ when the scale λ increases, i.e. when the size of the image is reduced before analysis by T_t. Then, the corresponding scale before reduction is increased. In more informal terms, we can say that the scale of analysis increases with the size of the picture. This is a natural assumption, and satisfied by all classical models. The continuity and differentiability assumptions on t' are also satisfied by all classical models and seem natural. Notice that no condition has been imposed on the relation between t' and (t,λ).
In order to fix ideas, let us examine which function t' is associated with several classical scale spaces. In the case of the basic morphological operators, dilation and erosion, it is easily seen that $t'(t,\lambda) = \lambda t$. In the case of the heat equation and of the mean curvature motion equation, one has $t'(t,\lambda) = \lambda^2 t$.

Finally, we state a general invariance axiom which implies the euclidean and the scale invariances, and also the invariance of the scale space under any orthographic projection of a planar shape. Combining those transformations leads to an arbitrary linear transform A of the plane. Set, for any such transform, $Af(\mathbf{x}) = f(A\mathbf{x})$. With the same formalism as for scale invariance, we get :

Definition 20.19 *We shall say that a family of operators T_t is* **affine invariant** *if it is scale invariant and if the associated function $t'(t,\lambda)$ can be extended into a function $t'(t,A)$, defined for every $t > 0$ and any linear application A of $\mathbb{R}^N$ with $det(A) \neq 0$, and satisfying*

$$t'(t,\lambda) = t'(t,\lambda Id)$$

and, for every A, $s > t > 0$,

$$AT_{t',s'} = T_{t,s}A, \tag{20.29}$$

where we note $t' = t'(t,A)$ and $s' = t'(t,A)$.

This relation means that the result of the scale space T_t is independent of the distance and orientation in space of the analyzed planar image. Indeed, any affine map $u \to Au$ can be interpreted as the anamorphosis of a planar image u when it is presented to the eye at any distance large enough with respect to its size and with an arbitrary orientation in space.

The fact that the scale-space function $t'(t,\lambda)$ can be *a priori* different for each scale space looks mysterious. We shall fix it in the next lemma by proving that we can, without loss of generality, **impose** all scale spaces to have the same scale-space function. It may be anyway convenient to keep a different one from the normalized one we propose here. Now, as we shall see in the chapters devoted to the identification of **all** scale spaces of a certain kind, the next lemma is an essential tool to perform this identification, which will then be made "up to a rescaling".

Lemma 20.20 (Normalization of scale.)
(i) Assume that $t \to T_t$ is a one to one, scale invariant, family of operators acting on real functions defined in $\mathbb{R}^N$ and satisfying $T_0 = Id$. Then there exists an increasing differentiable rescaling function $\sigma : [0, \infty] \to [0, \infty]$, such that $t'(t, \lambda) = \sigma^{-1}(\sigma(t)\lambda)$. If we set $S_t = T_{\sigma^{-1}(t)}$, we then have $t'(t, \lambda) = t\lambda$ for the rescaled analysis.

(ii) Assume that the family of operators T_t is affine invariant. Then the function $t'(t, B)$ only depends on t and $|detB|$: $t'(t, B) = t'(t, |detB|^{1/N})$ and is increasing with respect to t. Moreover, there exists an increasing differentiable rescaling function $\sigma : [0, \infty] \to [0, \infty]$, such that $t'(t, B) = \sigma^{-1}(\sigma(t)|detB|^{1/N})$ and if we set $S_t = T_{\sigma^{-1}(t)}$ we have $t'(t, B) = t|detB|^{1/N}$ for the rescaled analysis.

Before proving this lemma, we shall state a very useful consequence : the characterization of affine invariant scale spaces by a simple relation regarding their associated function F.

Lemma 20.21 *If a causal scale space is affine invariant, then, after the adequate renormalization yielded by Lemma 20.20, its associated function F satisfies*

$$F(BA^tB, Bp, t) = |detB|^{\frac{1}{N}} F(A, p, |detB|^{\frac{1}{N}} t). \tag{20.30}$$

for any linear map B. If a causal scale space is scale invariant, then, after the adequate scale renormalization (Lemma 20.20), its associated function F satisfies

$$F(\mu^2 A, \mu p, t) = \mu F(A, p, \mu t). \tag{20.31}$$

Proof We have $T_{t+h,t}B = BT_{\lambda(t+h),\lambda t}$, where $\lambda = |detB|^{\frac{1}{N}}$. We apply this relation to a quadratic form u such that $D^2u(\mathbf{x}) = A$, $Du(\mathbf{x}) = p$:

$$T_{t+h,t}(Bu) = B(T_{\lambda(t+h),\lambda t}u), \text{ that is, } T_{t+h,t}(u \circ B) = (T_{\lambda(t+h),\lambda t}u) \circ B.$$

Differentiating with respect to h at $h = 0$ on both sides yields

$$\frac{\partial}{\partial h}_{|h=O}(T_{t+h,t}(u \circ B)) = \frac{\partial}{\partial h}_{|h=O}(T_{\lambda(t+h),\lambda t}u) \circ B).$$

Taking into account that $D^2(u \circ B) =^t B(D^2u \circ B)B$ and $D(Bu) =^t B(Du \circ B)$ and using Formula (20.5), we get

$$F(^tB(D^2u \circ B)B, ^tB(Du \circ B), t) = \lambda F(D^2u \circ B, Du \circ B, \lambda t),$$

which yields

$$F(^tBAB, ^tBp, t) = |detB|^{\frac{1}{N}} F(A, p, |detB|^{\frac{1}{N}} t).$$

The relation (20.30) follows by changing B into ^tB. The second part of the lemma is obtained by replacing in the preceding proof B by μI_N, where I_N is the identity of $\mathbb{R}^N$. □

Proof of Lemma 20.20. We directly prove (ii), and will then say how to simplify the proof of (ii) in order to get (i). First we notice that for any linear transforms B and C and any t one has the semigroup property

$$t'(t, BC) = t'(t'(t, B), C). \tag{20.32}$$

Indeed, we have $BCT_{t'(t,BC)} = T_t BC = BT_{t'(t,B)}C = BCT_{t'(t'(t,B),C)}$. The map which associates T_t with t being one to one, we deduce (20.32). Next, we show that

$$t'(t, A) \text{ is increasing with respect to } t. \tag{20.33}$$

Let us prove that $t'(t, A)$ is one to one with respect to t for any A. Indeed, if not, there would be some A and some (s, t) such that $t'(t, A) = t'(s, A)$. Thus $T_t A = AT_{t'(t,A)} = AT_{t'(s,A)} = T_s A$ and therefore $t = s$ because T_t is one to one. Notice that this implies, in particular, that

$$t'(0, A) = 0 \tag{20.34}$$

Indeed, $AT_{t'(0,A)} = T_0 A = A$. Thus $T_{t'(0,A)} = Id = T_0$ and therefore $t'(0, A) = 0$. As a consequence, since $t'(t, A)$ is nonnegative (by definition), one to one, and continuous with respect to t, we can deduce that it is increasing with respect to t.

Let us now show that $t'(t, A)$ satisfies

$$t'(t, R) = t \quad \text{for any orthogonal transform} \quad R. \tag{20.35}$$

Iterating the formula of (20.32) we have

$$t'(t'(t'(t'(...t'(t, R)..., R), R), R) = t'(t, R^n)$$

Remark that there is a subsequence of $(R^n)_{n\in\mathbb{N}}$ tending to Id. (Indeed, there is a subsequence R^{n_k} which converges to some H, orthogonal, because the orthogonal group is compact. Therefore, the subsequence $R^{n_{k+1}-n_k}$ converges to Id.) Since there exists a subsequence of R^n tending to Id and since t' is continuous we have for this subsequence $\lim t'(t, R^n) = t'(t, Id) = t$. Assume by contradiction that $t'(t, R) = t$" with $t" < t$. Then $t'(t'(t, R), R) = t'(t", R) \leq t'(t, R) = t"$ and by recursion, $t'(t, R^n) = t'(t'(t'(t'(...t'(t, R)..., R), R), R) \leq t" < t$. This yields a contradiction. Thus $t'(t, R) \geq t$. We prove the converse inequality in the same way and we obtain $t'(t, R) = t$.

We note that any linear transform B of $\mathbb{R}^N$ can be obtained as a product of orthogonal transforms and of linear transforms of the kind $A(\lambda)$: $(x_1, x_2..., x_N) \to (\lambda x_1, x_2, ..., x_N)$ where λ is nonnegative. We only need to make a singular value decomposition [?] of B: $B = R_1 D R_2$, where R_1 and R_2 both are orthogonal transforms and D is a transform of the kind $(x_1, x_2..., x_N) \to (\lambda_1 x_1, \lambda_2 x_2, ..., \lambda_N x_N)$ where λ_i are non negative. Now, it is clear that D itself can be decomposed as $D = A(\lambda_1)R_2 A(\lambda_2)R_2^{-1}...R_N A(\lambda_N)R_N^{-1}$ where R_i is the orthogonal transform: $(x_1, ..., x_i..., x_N) \to (x_i, ..., x_1..., x_N)$ which exchanges x_1 and x_i. Using Relation (20.32), $t'(t, R_i) = t$ (which comes from (20.35)), the singular value decomposition and the obvious relation $A(\lambda_1)A(\lambda_2) = A(\lambda_1\lambda_2)$, we obtain that t' only depends on t and $|\det(B)| = \lambda_1...\lambda_N$. So we can write

$$t'(t, B) = t'(t, |detB|^{1/N}).$$

Using (20.32) and (20.33), we have

$$t'(t,\lambda\mu) = t'(t'(t,\mu),\lambda) \tag{20.36}$$

for any positive λ and μ. Differentiating this relation with respect to μ at $\mu = 1$ yields

$$\lambda\frac{\partial t'}{\partial\lambda}(t,\lambda) = \frac{\partial t'}{\partial t}(t,\lambda)\frac{\partial t'}{\partial\lambda}(t,1) \tag{20.37}$$

We consider the function $\phi(t) = \frac{\partial t'}{\partial\lambda}(t,1)$, which is positive by assumption. We then choose σ such that

$$\phi\sigma' = \sigma \tag{20.38}$$

in such a way that

$$\sigma(t) = exp(\int_1^t ds/\phi(s)).$$

Let us finally set

$$G(x,y) = t'(x,\frac{y}{\sigma(x)})$$

, so that

$$t'(t,\lambda) = G(t,\sigma(t)\lambda).$$

By (20.37) and (20.38) and the differentiability assumption on t' (Definition 20.18),

$$\frac{\partial G}{\partial x}(x,y) = \frac{\partial t'}{\partial t}((x,\frac{y}{\sigma(x)}) - \frac{\partial t'}{\partial\lambda}(x,\frac{y}{\sigma(x)})\frac{y\sigma'(x)}{\sigma(x)^2} = 0.$$

Thus $G(x,y) = \beta(y)$ for some differentiable nondecreasing function β. We obtain

$$t'(t,\lambda) = \beta(\sigma(t)\lambda). \tag{20.39}$$

Returning to the definition of $\phi(t)$ and using (20.38), we have $\phi(t) = \frac{\partial t'}{\partial\lambda}(t,1) = \sigma(t)\beta'(\sigma(t)) = \phi(t)\sigma'(t)\beta'(\sigma(t))$. Dividing both sides by $\phi(t)$, we obtain

$$\frac{\partial(\beta(\sigma(t)))}{\partial t} = 1.$$

Integrating this last relation between 0 and t yields $\beta(\sigma(t)) = t + \beta(\sigma(0))$. Now, by (20.34), $t'(0,\lambda) = 0$. This implies by (20.39) that $\beta(\sigma(0)) = 0$ and therefore $\beta = \sigma^{-1}$, so that

$$t'(t,\lambda) = \sigma^{-1}(\lambda\sigma(t)).$$

In order to complete the proof, we set $S_t = T_{\sigma^{-1}(t)}$ and we prove that the affine invariance is true for S_t with $t'(t,\lambda) = \lambda t$. We have

$$S_tB = T_{\sigma^{-1}(t)}B = BT_{t'(\sigma^{-1}(t),\lambda)} = BT_{\sigma^{-1}(\lambda\sigma(\sigma^{-1}(t)))} =$$

$$BT_{\sigma^{-1}(\lambda t)} = BS_{\lambda t}.$$

Proof of (i). The proof of the first item of the scale normalization lemma is identical to the proof of the second, which we have just performed. We only need to replace "B" and "C" by λ, μ everywhere. □

20.5 First application : axiomatization of linear scale space

We shall now deduce from Theorem 20.6 a characterization of the heat equation $\partial u/\partial t = \Delta u$ as the unique linear and isotropic scale space. As a consequence for image smoothing models, linearity and contrast invariance are incompatible, and we obtain an explanation for the coexistence of (at least) two different schools in image processing : contrast invariant mathematical morphology on the one side and classical linear scale spaces (convolution with gaussians) on the other side.

Theorem 20.22 *Let T_t be a causal, translation invariant scale space defined on a space of functions $\mathcal{F}$ having the following having the following property $\mathcal{P}$:*
For all $\mathbf{x} \in I\!R^N, p \in I\!R^N$ and A symmetric $N \times N$ matrix, there is $u \in \mathcal{F}$ such that $Du(\mathbf{x}) = p$, $D^2u(\mathbf{x}) = A$.
If the operators $T_{t,s}$ are **linear** *and* **Euclidean invariant**, *then (up to a rescaling $t' = h(t)$), the function $u(t, \mathbf{x}) = (T_t u_0)(\mathbf{x})$ is the solution of the heat equation*

$$\partial u/\partial t - \Delta u = 0 \text{ in } I\!R^N \times [0, \infty[, \quad u(0, .) = u_0(.) \text{ in } I\!R^N. \tag{20.40}$$

Exercise 20.3 *Our axiomatic for causal scale spaces assumes that $T_t u_0$ is well defined when u_0 is a quadratic form. Now, we have defined the solution of the heat equation in Chapter 2 only when u_0 is bounded. One can show that this solution is easily extended to quadratic forms with the same convolution formula, $u(t) = G_t * u_0$. Show, by using the fact that $G_t(\mathbf{x})$ has exponential decay at infinity, that $u(t, \mathbf{x}) = G_t * u_0$ is well defined and C^∞ when there exist C and $k > 0$ such that*

$$|u_0(\mathbf{x})| \leq C(1 + |\mathbf{x}|^k).$$

Show that $u(t, \mathbf{x})$ is then a classical solution of the heat equation.

Proof. Since the scale space is translation invariant in space, we know that $F(A, p, \mathbf{x}, c, t) = F(A, p, c, t)$ does not depend on $\mathbf{x}$ and since it is invariant with respect to grey level translations, $F(A, p, c, t) = F(A, p, t)$ does not depend upon c (Propositions 20.8 and 20.12). Since, if u is any C^2 function, by Theorem 20.6, $F(D^2u, Du, t) = \lim_{h \to 0}(T_{t+h,t}u - u)/h$, we deduce from the linearity of $T_{t+h,t}$ that F is linear in u and therefore satisfies

$$F(rD^2u + sD^2v, rDu + sDv, t) = rF(D^2u, Du, t) + sF(D^2v, Dv, t).$$

for any real numbers r and s and any C^2 functions u and v. In the arguments which follow, we fix t and we therefore omit to mention it and write $F(A, p)$ instead of $F(A, p, t)$, etc.. Since the values of Du, Dv, D^2u, D^2v are by Property $\mathcal{P}$ arbitrary and can be independently taken to be 0, we obtain for any vectors p, p' and symmetric matrices A, A' that

$$F(rA + sA', rp + sp') = rF(A, p) + sF(A', p')$$

$$\text{and} \quad F(A, p) = F(A, 0) + F(0, p).$$

Thus we can set $F(A, p) = F_1(p) + F_2(A)$ and F_1 and F_2 are linear. ¿From the invariance of the scale space by linear isometry, we also obtain $F({}^tRAR, {}^tRp) = F(A, p)$ for any isometry R of $I\!R^N$.
Taking $A = 0$, we deduce from the preceding relations that $F_1(Rp) = F_1(p)$ and therefore $F_1(Rp) = F_1(p)$

for any isometry R. Since F_1 is linear, this is only possible if $F_1(p) = 0$ for any p. Thus $F(A,p) = F_2(A)$. By the isometry invariance again, we have $F_2({}^tRAR) = F_2(A)$ for any isometry R and any symmetric matrix A. Since every symmetric matrix can be diagonalized in an orthonormal basis and every pair of orthonormal bases can be exchanged by an isometry, we see that F_2 only depends on the eigenvalues $\lambda_1, ..., \lambda_n$ of A. Thus we can write $F_2(A) = F_2(\lambda_1, ..., \lambda_n)$. Since the eigenspaces can also be exchanged by isometries, $F_2(\lambda_1, ..., \lambda_n)$ must be invariant under any permutation of the eigenvalues. Thus F only depends on the symmetric functions of the eigenvalues.
Now, the only linear symmetric function of n variables is, up to multiplicative constant, the sum. Thus $F_2(A)$ only depends (linearly) on the trace of A and therefore $F_2(A) = c\ trace(A)$ for some constant c. We conclude that $F(D^2u, Du) = c\Delta u$. Since F must be increasing in A, the constant c is nonnegative. We now remember that all of this argument has been made with the omission of the t-dependence of F. Thus our real conclusion is $F(D^2u, Du, t) = c(t)\Delta u$ for some continuous nonnegative function $c(t)$. Doing the rescaling $t'(t)$ defined by $\partial t'(t)/\partial t = c(t)$, we again obtain a heat equation $\partial u/\partial t' - \Delta u = 0$. □

20.5.1 Further invariance properties of the linear scale space.

We now list some invariance properties of the operator $T_t u_0 = G_t * u_0$. These properties follow immediately from the properties of G_t, so that we essentially leave them as exercises to the reader.

Exercise 20.4 *Prove, by using the convolution formula, that The linear scale space is scale invariant. More precisely, for any homothety H_λ, $\lambda > 0$, one has $H_\lambda T'_t = T_t H_\lambda$, with*

$$t' = t\lambda^2. \tag{20.41}$$

Prove that the scale space also is isotropic, i.e.

$$T_t(Ru_0) = R(T_t u_0)$$

for any isometry R of $\mathbb{R}^N$. (Use the fact that if R is a linear isometry, then $G_t(R\mathbf{x}) = G_t(\mathbf{x})$.

Lemma 20.23 *The scale space defined by the heat equation (20.40) is invariant by grey level translations, but not contrast invariant.*

Proof : Let $\partial u/\partial t = \Delta u$ and $u = h(v)$, where h is a C^2 nondecreasing function. Then

$$h'(v)\partial v/\partial t = h'(v)\Delta v + h''(v)|Dv|^2$$

If the equation were contrast invariant, then v would satisfy $\partial v/\partial t = \Delta v$. Combining these equations yields $h''(v)|Dv|^2 = 0$, and since v is arbitrary, $h'' = 0$. Thus, the equation is only invariant under affine transforms of the grey level scale. □

20.5.2 Non causal linear scale-spaces

It is important to notice that the axioms we have considered can be weakened and that, in particular, other linear scale-spaces can be derived from slightly different principles. As an instance of axiomatics leading to the heat equation but also to other convolution kernels let us mention that E.J. Pauwels, P. Fiddelaers, T. Moons, and L.J. Van Gool have explored linear operators T_t satisfying the recursivity assumption as well as the scale invariance. They proved in [337] the following theorem

Theorem 20.24 [Pauwels and al. 94] N=2. *Assume that T_t is recursive, and that $T_t u_0 = k_t \circ u_0$, for a continuous function k_t from $\mathbb{R}^2$ into $\mathbb{R}$. If T_t satisfies also the scale invariance, commutes with grey level translations and rotations, then k_t is of the form*

$$k_t(\mathbf{x}) = \int_0^\infty J_0(r\rho)\exp(-\alpha\rho^\alpha t)\rho d\rho \qquad \textit{with } r = |\mathbf{x}|,$$

where $\alpha > 0$ and J_0 is the zeroth order Bessel function,

$$J_0(x) = \frac{1}{2\pi}\int_0^{2\pi} \exp(-ix\cos(\theta))d\theta$$

We obtain that scale-invariance reduces the possible linear and recursive scale-spaces to a family with only one free parameter α.

Choosing $\alpha = 2$ leads to the gaussian kernel which is, as we have seen, the unique linear scale-space satisfying also the local maximum principle.

For $\alpha > 2$, k_t is sometimes negative, then contredicts the comparison principle.

For $\alpha < 2$, k_t violates the locality condition included in the local maximum principle. Roughly speaking, this is due to the fact that the function k_t does not decrease fast enough to zero when $|\mathbf{x}|$ tends to infinity.

References.

The presentation of this chapter and of the next one mainly follows the Alvarez et al. papers [21, 20, 15] whose claim was "We describe all multiscale causal, local, stable and shape preserving smoothing operators. This classification contains the classical "morphological" operators, and some new ones." Several survey papers present this axiomatics with increasingly simple sets of axioms, Lions [267], Alvarez [18], [185] and the PhD [190] which actually is an early version of the present book.

Linear scale space

The scale space theory was founded (in a linear framework) by Witkin [440], Marr [287] Koenderink. An earlier development of the linear scale space has been traced in Japan [424, 425]. The work of Lindeberg [264] represents a high state of sophistication in the use of Linear Scale Space. Many works of Florack, ter Haar Romeny, Koenderink and Viergever focus on the computation of partial derivatives of any order in an image and their use in image analysis [154, 155, 156, 158]. The concept of *causality*, used by all of these authors is crucial and has reinterpreted in this chapter as the combination of two requirements: pyramidal structure and a comparison principle. De Giorgi [174] founded on similar principles his mathematical theory of barriers for geometric motions. There are many axiomatic characterizations of the linear scale space by its invariance, conservativeness and causal properties. A particular mention is due to the early Babaud, Witkin, Baudin and Duda [42] and to Hummel [216]. A slight relaxation of the initial axioms led

Pauwels et al. [337] to discover other possibilities for the linear scale space, less local though. There have also been several attempts to define non linear scale spaces, understood as nonlinear invariant families of scaled smoothing operators. In Mathematical Morphology, let us mention Chen [98] and Toet [406, 407] and Jackway [228] who emphasizes the scale space properties of multiscale erosions and dilations. After [15], several different axiomatics have been proposed for nonlinear scale-spaces. Weikert[421, 423] insists on grey level conservation (which excludes all mathematical morphology operators) and proposes a line of conservative parabolic nonlinear P.D.E.'s. The axiomatic presentation of Olver et al. [326] deduces the various scale spaces as invariant heat flows. See also [319] and [405] for a book with miscellaneous contributions to the subject of geometric diffusion.

Extensions

Caselles et al. [303] have questionned the very soundness of applying any of the proposed scale spaces to natural images : they argue following the Kanisza psychophysical theory that occlusion generates T-junctions in images and that these T-junctions should be detected before any smoothing is performed. In [84], the same authors propose the set of level lines of the image, the so called topographic map, as an alternative multiscale structure for describing images. Geraets et al. propose a generalization of scale space to discrete pointsets [169].

Contrast invariance

The invariance of image analysis with respect to changes of illumination is stated in perception theory by [435], as early as 1923. In Mathematical Morphology, the contrast invariance is commented and used in Serra [378]. Koenderink [251] insist a lot on this requirement and devise photometric invariants. A study of contrats invariant differential operators is made in Florack [157]. See also [348] and Romeny et al. [404], where third order contrast invariant operators are constructed for T-junction detection.

Scale invariance

[56] is one of the first discussions about rotation invariant image operators. See also [261] [43] discusses scale invariant shape representation. Alvarez and Gousseau [14] confirm by numerical experiments on natural images their scale invariance.

Affine invariance

Affine invariance is viewed as an approximate projective invariance [103], an obvious requirement in 3D object recognition [61, 453]. The work of Forsyth, Mundy and Zissermann [159, 307, 308] has been seminal and lunched a wide discussion on this theme and many contributions on the computation and use of affine and projective differential invariants in image processing [362, 60, 412, 431].

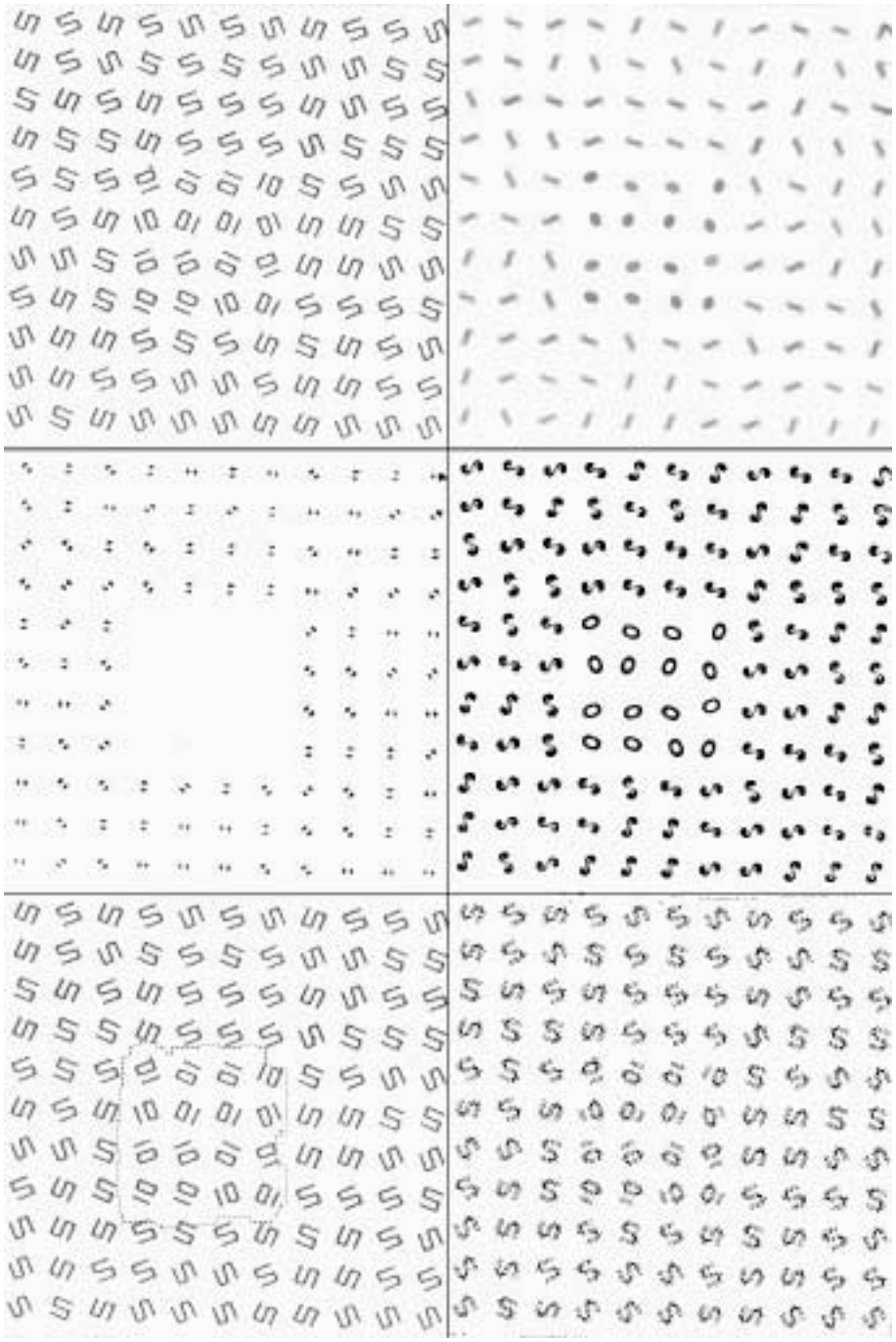

Figure 20.3: Hyperdiscrimination of textures by nonlinear scale space. According to Julesz' theory of textons, textures are discriminable to the human perception if their average behaviour in terms of "texton" density is different. As shown in its mathematical formalization, proposed by C. Lopez, some of the texton densities can be interpreted as densities of the positive and negative parts of the image curvature at different scales. In this remarkable experiment, C. Lopez proved that one of the simplest contrast invariant scale spaces beats by far the human discrimination performance. From left to right and up-down :
1-an original preattentively undiscriminable texture pair. The central square of the image is made of rotated "10's" and the rest of rotated "S's". Those patterns are different, but have the same number of bars, angles, etc.
2-curvature motion applied to the original up to some scale 3-negative part of the curvature at the same scale 4-positive part of the curvature at the same scale 5-multichannel segmentation of the multiimage made of the curvatures 6-negative part of the curvature at scale 0. As seen in 2, 3, 4 and 5, this nonlinear scale space permits to discriminate easily the undiscriminable textures.

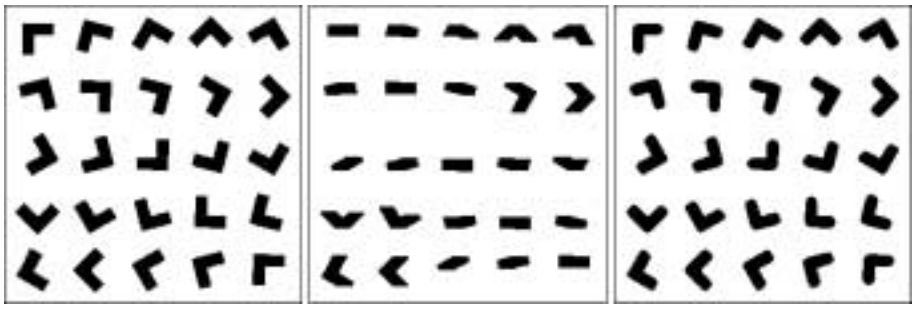

Figure 20.4: An euclidean invariant filter and a non-euclidean one. The left frame contains simple shapes that can be deduced from each other by rotations. The middle image is the closing of the left image by a horizontal rectangle of size 6*2 pixels. This non euclidean filter produces different results, depending on the shapes orientations. The right image is the closing of the left image by a circle of radius 4 pixels (same area up to the pixel precision as the rectangle used in the middle image). This filter is euclidean, therefore the resulting shapes can also be deduced from each other by rotations.

Figure 20.5: The heat equation is not contrast invariant. This experiment shows that the linear scale-space is not contrast invariant. First row : original image. Second row : two different contrast changes have been applied to this image. Third row : a convolution by a gaussian is applied to both images of the second row. Fourth row : the inverse contrast change is applied to the images of the third row. If the linear scale-space were contrast invariant, these images should be equal : this is not the case, since the difference (displayed in the fifth row) is not null.

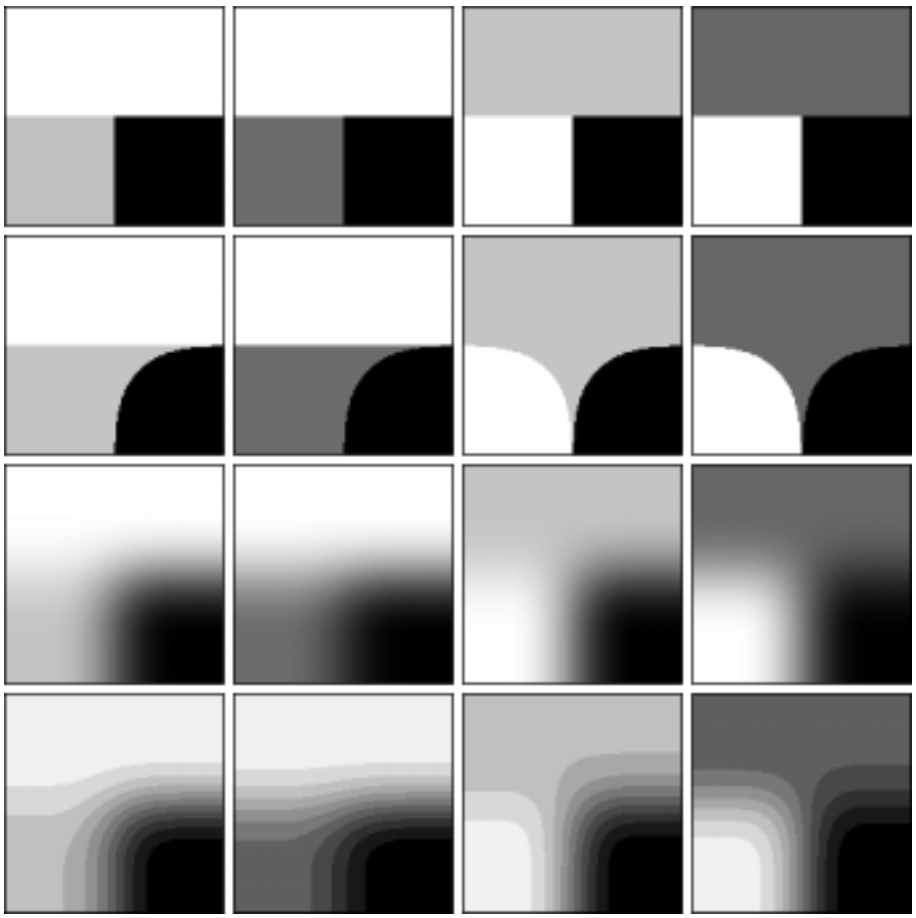

Figure 20.6: Same geometric figures, different evolutions under smoothing. 1st row : four T-junctions that only differ by the grey-level order between the three regions.
2nd row : result of a smoothing by the AMSS model. We see that two different evolutions are possible : if the regions of the image keep the same order, then the geometric evolution is identical. If, instead, a non monotone contrast change has been applied, the evolutions differ.
3rd row : result of a smoothing by the linear scale-space : all of the T-junctions give different evolutions. The evolution depends on the gray-level values of the three level sets, and not only on their order.
4rd row : quantization of the 3rd row in order to display the shapes of some level lines.

Figure 20.7: Contrast invariance of the affine morphological scale space (AMSS). First row : original image. Second row : two contrast changes applied to the original. Third row : AMSS applied to both images of the second row, by the finite difference scheme (FDS) explained in Chapter 24. Fourth row : inverse contrast change applied to the filtered images. A visual check shows that they are almost identical. Bottom image : numerical check by taking the difference of the images in the fourth row. Compare with the same experiment performed with the linear scale space, Figure 20.5.

Chapter 21

All contrast invariant and affine scale spaces

21.1 The two dimensional case

In this section, we give a characterization of two-dimensional contrast invariant scale spaces as "curvature evolution equations". We prove that if a two-dimensional scale space T_t is causal and invariant by isometries and contrast changes (contrast invariance), then it obeys an equation of the kind

$$\frac{\partial u}{\partial t} = |Du| G(\mathrm{curv}(u), t). \tag{21.1}$$

How to choose G ? If we impose affine invariance and reverse contrast invariance for T_t, that is, $T_t(-u) = -T_t u$, then there is a single equation left : the so called Affine Morphological Scale Space (AMSS),

$$\frac{\partial u}{\partial t} = (\mathrm{curv}(u))^{\frac{1}{3}} |Du|.$$

This study can be generalized to any dimension. For simplicity, we first treat "à la main" the case of dimension 2, because the computations are in that case more intuitive.

By Theorem 20.6, we were led to study scale space models defined by a partial differential equation

$$\frac{\partial u}{\partial t} = F(D^2 u, Du, u, x, t), \quad u(0) = u_0$$

where u_0 is the image to analyze (the datum), $u(t)$ is the image analyzed at scale t and $F(A, p, c, x, t)$ depends on a symmetric 2*2 matrix A, a two-dimensional vector p, a constant c, a point of the plane x and a real positive scale t. If we assume :

• The translation invariance, then F does not depend upon x (Proposition 20.12).

• The invariance by translation of grey level $u \to u + C$, then F does not depend upon c (Proposition 20.8).

• The isotropy, that is, the euclidean invariance assumption, then we obtain a further constraint on F, which is given by Lemma 20.14: F satisfies, for every linear isometry $R \in O^N$,

$$(20.22) \qquad F(RA^t R, Rp, t) = F(A, p, t)$$

• The contrast invariance : we know by Lemma 20.16 that it implies

$$F(\mu A + \lambda p \otimes p, \mu p, t) = \mu F(A, p, t), \tag{20.25}$$

for every real values λ, μ, every symmetric matrix A and every two-dimensional vector p. By $p \otimes p$ we mean the tensorial product, or in a more intuitive way, the matrix product $p^t p$, where p is thought of as a column matrix. In dimension 2, this yields

$$p \otimes p = \begin{bmatrix} p_1^2 & p_1 p_2 \\ p_1 p_2 & p_2^2 \end{bmatrix}.$$

We shall now draw from Relations (20.22) and (20.25) a huge reduction of F, which will appear to depend upon only two real parameters $\tilde{a}_{1,1}(A,p)$ and $\tilde{a}_{1,2}(A,p)$. These parameters are defined by considering the rotation matrix

$$R_p = \frac{1}{|p|} \begin{bmatrix} p_1 & p_2 \\ -p_2 & p_1 \end{bmatrix}.$$

R_p has been chosen so that it sends $\frac{p}{|p|}$ onto the unit vector $e_1 =^t (1,0)$, that is

$$R_p p = |p| e_1. \tag{21.2}$$

Then we set

$$\frac{1}{|p|} R_p A^t R_p = \begin{bmatrix} \tilde{a}_{1,1} & \tilde{a}_{1,2} \\ \tilde{a}_{1,2} & \tilde{a}_{2,2} \end{bmatrix}. \tag{21.3}$$

An easy computation yields (setting $p^\perp = (-p_2, p_1)$, a vector orthogonal to p)

$$\tilde{a}_{1,2} = \frac{1}{|p|^3}((p_1^2 - p_2^2)a_{1,2} + p_1 p_2 (a_{2,2} - a_{1,1})) = \frac{A(p, p^\perp)}{|p|^3}, \tag{21.4}$$

$$\tilde{a}_{2,2} = \frac{1}{|p|^3}(a_{1,1} p_2^2 - 2a_{1,2} p_1 p_2 + a_{2,2} p_1^2)) = \frac{A(p^\perp, p^\perp)}{|p|^3}, \tag{21.5}$$

Lemma 21.1 *If $F(A,p,t)$ satisfies (20.22) and (20.25), then there is a function G, depending on three real variables only, such that*

$$F(A,p,t) = |p| G(\tilde{a}_{1,1}, \tilde{a}_{1,2}, t). \tag{21.6}$$

Of course, we must always keep in mind that A represents D^2u and p represents Du so that the result above and the calculations below, though purely algebraic, have a differential geometry interpretation. In particular, we have

$$\tilde{a}_{2,2}(D^2u, Du) = \operatorname{div}\left(\frac{Du}{|Du|}\right) = \operatorname{curv}(u) \quad \text{and}$$

$$\tilde{a}_{1,2}(D^2u, Du) = \operatorname{div}\left(\frac{Du^\perp}{|Du|}\right) = anticurv(u),$$

Exercise 21.1 *Interpretation of $anticurv(u)$. Show that $anticurv(u)(\mathbf{x})$ can be interpreted as the curvature of the gradient lines of u. Those lines are the lines tangent to the gradient of u at every point. They form a bunch of lines orthogonal to the isolevel lines of u.*

Proof of Lemma 21.1 We first use (20.25), setting $\mu = \frac{1}{|p|}$. Thus

$$F(A,p,t) = |p|F(\frac{A}{|p|} + \lambda p \otimes p, \frac{p}{|p|}, t)$$

for any real value λ. Then, we apply (20.22) with $R =^t R_p$:

$$F(A,p,t) = |p|F(R_p(\frac{A}{|p|})^t R_p + \lambda R_p(p \otimes p)^t R_p, e_1, t),$$

for any real number λ. It is easily checked that

$$R_p(p \otimes p)^t R_p = \begin{bmatrix} |p|^2 & 0 \\ 0 & 0 \end{bmatrix}.$$

Thus, by (21.3),

$$F(A,p,t) = |p|F(\begin{bmatrix} \tilde{a}_{1,1} + \lambda|p|^2 & \tilde{a}_{1,2} \\ \tilde{a}_{1,2} & \tilde{a}_{2,2} \end{bmatrix}, e_1, t)$$

for any λ. Consequently, we see that F only depends upon $\tilde{a}_{2,2}$ and $\tilde{a}_{1,2}$, which proves the assertion (21.6). □

Lemma 21.2 *The function G only depends upon $\tilde{a}_{2,2}$ and t.*

Proof Here we use the fact that $F(A,p,t)$ is a nondecreasing continuous function with respect to A (one of the conclusions of Theorem 20.6). Recall that by Relations (21.4) and (21.5),

$$\tilde{a}_{1,2} = \frac{A(p,p^\perp)}{|p|^3},$$

$$\tilde{a}_{2,2} = \frac{A(p^\perp,p^\perp)}{|p|^3}.$$

A vector $p \neq 0$ being fixed, let us consider two matrices A and B such that $A \geq B$. We have

$$F(A,p,t) = G(\frac{A(p,p^\perp)}{|p|^3}, \frac{A(p^\perp,p^\perp)}{|p|^3}, t)$$

In order to simplify notations, we set as above $\tilde{a}_{2,2} = \frac{1}{|p|^3}A(p^\perp,p^\perp)$, $\tilde{b}_{2,2} = \frac{1}{|p|^3}B(p^\perp,p^\perp)$, $\tilde{a}_{1,2} = \frac{1}{|p|^3}A(p^\perp,p)$, $\tilde{b}_{1,2} = \frac{1}{|p|^3}B(p^\perp,p)$, and analogs for $\tilde{a}_{1,1}$ and $\tilde{b}_{1,1}$. Since the system $(p,p^\perp)$ is orthonormal, one has $A \geq B$ if and only if for every pair of real numbers (x,y),

$$(\tilde{a}_{1,1} - \tilde{b}_{1,1})x^2 + 2(\tilde{a}_{1,2} - \tilde{b}_{1,2})xy + (\tilde{a}_{2,2} - \tilde{b}_{2,2})y^2 \geq 0 \tag{21.7}$$

We fix $\tilde{a}_{2,2} - \tilde{b}_{2,2} = \varepsilon > 0$ and we notice that if $-\tilde{b}_{1,1}$ is chosen large enough in correspondence of arbitrarily fixed values of $\tilde{a}_{1,1}$, $\tilde{a}_{1,2}$, $\tilde{b}_{1,2}$ and ε, then (21.7) is satisfied for every (x,y). Thus $F(A,p,t) \geq F(B,p,t)$ and therefore

$$G(\tilde{a}_{1,2}, \tilde{a}_{2,2}, t) \geq G(\tilde{b}_{1,2}, \tilde{a}_{2,2} - \varepsilon, t),$$

this relation being true for every $\tilde{a}_{1,2}$, $\tilde{b}_{1,2}$, $\tilde{a}_{2,2}$ and ε. We can let $\varepsilon \to 0$ in the last inequality, since F is continuous with respect to A. It follows that

$$G(\tilde{a}_{1,2}, \tilde{a}_{2,2}, t) \geq G(\tilde{b}_{1,2}, \tilde{a}_{2,2}, t),$$

for any value of the three variables. Thus G only depends on its second and third argument. □
We can summarize what we have proved in the following theorem.

Theorem 21.3 *If a two-dimensional scale space T_t is causal and invariant by isometries and contrast changes, then it obeys an equation of the kind*

$$\frac{\partial u}{\partial t} = |Du| G(\text{curv}(u), t), \tag{21.8}$$

where G is nondecreasing and continuous with respect to its first argument.

We have not yet used the scale and affine invariances. This is to come now.

Theorem 21.4 *Let T_t be a causal, isotropic constrast invariant scale space. Assume that it is, in addition, scale invariant (that is, invariant with respect to zooms) and normalized according to Lemma 20.20. Then its equation is*

$$\frac{\partial u}{\partial t} = |Du| \beta(t\text{curv}(u)) \tag{21.9}$$

where β is a nondecreasing continuous function.
If we impose a full affine invariance to the scale space and assume again that the scale is normalized (Lemma 20.20), then its equation is

$$\frac{\partial u}{\partial t} = |Du| \gamma(t\text{curv}(u), \tag{21.10}$$

where for some positive constants C and D, $\gamma(s) = Cs^{\frac{1}{3}}$ if $s > 0$, $\gamma(s) = -D|s|^{\frac{1}{3}}$ if $s < 0$. Conversely, this equation defines an affine invariant scale space.

Remark 21.5 *In order to have $C = D$, it is enough to impose the "reverse contrast invariance", which extends the contrast invariance $T_{t+h,t} \circ g = g \circ T_{t+h,t}$ to nondecreasing contrast functions g.*

Proof. By Lemma 20.21, if a causal scale space is affine invariant, then, after the adequate renormalization (Lemma 20.20), its associated function F satisfies

$$F(BA^tB, Bp, t) = |detB|^{\frac{1}{2}} F(A, p, |detB|^{\frac{1}{2}} t). \tag{21.11}$$

for any linear map B. We first apply Relation (21.11) to the case of a zoom $B = cId$. Then

$$\tilde{a}_{2,2}(BA^tB, Bp) = \tilde{a}_{2,2}(c^2 A, cp) = \tilde{a}_{2,2}(\frac{c^2 A(cp^\perp, cp^\perp)}{|cp|^3}) = c\tilde{a}_{2,2}(A, p).$$

Thus, by relation (21.11),

$$c|p| G(c\tilde{a}_{2,2}(A, p), t) = c|p| G(\tilde{a}_{2,2}(A, p), ct).$$

Since this relation holds for arbitrary values of A, p, c, t, we have for any c, s and $t \geq 0$, $G(cs,t) = G(s,ct)$. Choosing $c = t^{-1}$, we get

$$G(s,t) = G(st,1) = \beta(st)$$

for a continuous nondecreasing function β. This proves the first statement of the theorem. Let us now pass to the general affine invariance. In order to identify the power $\frac{1}{3}$, we only need to express the affine invariance in the particular case of orthogonal affinities with determinant 1,

$$B(\lambda) = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{bmatrix}.$$

Indeed, every linear map of the plane is a product of isometries and such affinities and we already have fully used the invariance by isometries. We first compute

$$BA^tB = \begin{bmatrix} \lambda^2 a_{1,1} & a_{1,2} \\ a_{1,2} & \lambda^{-2} a_{2,2} \end{bmatrix},$$

$$Bp = (\lambda p_1, \frac{p_2}{\lambda}).$$

Therefore, using (21.5),

$$\tilde{a}_{2,2}({}^tBAB, Bp) = \frac{a_{1,1}p_2^2 - 2a_{1,2}p_1p_2 + a_{2,2}p_1^2}{(\lambda^2 p_1^2 + \lambda^{-2}p_2^2)^{\frac{3}{2}}}. \tag{21.12}$$

We know that $F(A,p,t) = \beta(t\tilde{a}_{2,2}(A,p,t))|p|$. Plugging this relation into the affine invariance relation (21.11) yields

$$|Bp|\beta(t\tilde{a}_{2,2}(BA^tB, Bp)) = |p|\beta(t\tilde{a}_{2,2}(A,p)). \tag{21.13}$$

We impose $p_2 = 0$, $p_1 = 1$, $a_{2,2} = 1$, so that from (21.12) and (21.13) follows

$$|\lambda|\beta(\frac{t}{\lambda^3}) = \beta(t),$$

which achieves the proof. □

21.2 General form of contrast invariant scale-space equations in dimension N.

We do the same general assumptions on the considered scale spaces as in the preceding section : we consider a scale space defined by an equation $\partial u/\partial t = F(D^2u, Du, t)$ and we assume that $F(A,p,t)$ is nondecreasing with respect to A, that is $F(A,p,t) \geq F(B,p,t)$ if $A \geq B$ (comparison principle). The considered F's will always be continuous with respect to A, at least for $p \neq 0$. Our aim is to deduce from the contrast invariance the general form of the function F in dimension N. We shall prove that if we assume also the isometry invariance, then the scale space is described by an equation

$$\frac{\partial u}{\partial t} = G(\lambda_1, ..., \lambda_{N-1}, |Du|, t)$$

where $\lambda_1, ..., \lambda_{N-1}$ are the principal curvatures of the level hypersurface of u and G is a real function defined on $\mathbb{R}^{N-1} \times [0, +\infty[\times [0, +\infty[$ which is symmetric with respect to $(\lambda_1, ..., \lambda_{N-1})$, positively homogenous with respect to $(\lambda_1, ..., \lambda_{N-1}, |Du|)$ and nondecreasing with respect to each $\lambda_i (1 \leq i \leq N-1)$ for all t in $[0, +\infty[$. The main axiom which we introduce in this section is the contrast invariance, $T_t \circ h = h \circ T_t$ for all $t \geq 0$, where h is any nondecreasing real function. In the calculations below, we shall forget the t-dependence of F because it does not change calculation and results and write $F(A, p)$ instead of $F(A, p, t)$. By Lemma 20.16, the function F associated with a contrast invariant scale space satisfies (Relation (20.25))

$$F(\mu A + \lambda p \otimes p, \mu p, t) = \mu F(A, p, t), \tag{21.14}$$

for all real values λ, $\mu \geq 0$, every symmetric matrix $A \in S^N$ and every vector $p \in \mathbb{R}^N$. Taking $\lambda = 0$, this yields in particular that F is positively homogeneous,

$$F(\mu A, \mu p) = \mu F(A, p) \quad \text{for all } \mu \geq 0, A \in S^N,\ p \in \mathbb{R}^N,\ p \neq 0. \tag{21.15}$$

(In particular, $F(0, 0) = 0$). Now, provided (21.15) holds, the contrast invariance reduces to

$$F(A + \lambda p \otimes p, p) = F(A, p) \text{ for all } \lambda \in \mathbb{R}, A \in S^N, p \in \mathbb{R}^N, p \neq 0. \tag{21.16}$$

If $N = 1$, this implies that F depends only on p and in view of (21.15), F is necessarily given by

$$F(A, p) = ap^+ + bp^- \quad \text{for some} \quad a, b \in \mathbb{R}, \tag{21.17}$$

where $p^+ = max(p, 0)$ and $p^- = min(p, 0)$. Less trivial situations occur if $N \geq 2$, an assumption which we make in all that follows. If $p \in \mathbb{R}^N$, we denote by Q_p the matrix of the projection onto the hyperplane $p^\perp$ orthogonal to p.

Theorem 21.6 *(Giga, Goto [?]) Let T_t be a scale space satisfying the contrast invariance. Then the associated F satisfies*

$$F(A, p) = F(Q_p A Q_p, p) \ \textit{for all}\ \ A \in S^N,\ p \textit{ in } \mathbb{R}^N,\ p \neq 0, \tag{21.18}$$

where Q_p is the projection matrix, $Q_p = I_N - p \otimes p / |p|^2$.

Proof of Theorem 21.6. In order to show (21.18), we fix p in $\mathbb{R}^N$, $p \neq 0$ and we select a coordinate system such that $p = |p|(0, ..., 0, 1)$, in which case $Q_p A Q_p$ becomes $A' = (a'_{ij})_{1 \leq i,j \leq N}$ where $a'_{ij} = a_{ij}$ if $1 \leq i, j \leq N-1$, $a'_{ij} = 0$ otherwise and $p \otimes p = |p|^2 (\delta_{Ni} \delta_{Nj})_{1 \leq i,j \leq N}$. (As usual, we set $\delta_{k,l} = 0$ if $k \neq l$ and $\delta_{k,l} = 1$ otherwise.) Relation (21.16) implies that $F(A, p)$ does not depend on a_{NN}. Set $M = a_{N,1}^2 + ... + a_{N,N-1}^2$ and $I_\epsilon = \epsilon I_N + (M/\epsilon - \epsilon)(\delta_{Ni} \delta_{Nj})_{1 \leq i,j \leq N}$. One easily checks that $Q_p A Q_p \leq A + I_\epsilon$ and $A \leq Q_p A Q_p + I_\epsilon$. Using $F(A, p) \geq F(B, p)$ if $A \geq B$, and letting ϵ tend to zero, we obtain, since F does not depend on a_{NN} and is continuous for $p \neq 0$, $F(A, p) = F(Q_p A Q_p, p)$. □

We need to introduce some notations. Since $Q_p A Q_p$ leaves invariant $(\mathbb{R}p)^\perp$ (for $p \neq 0$ fixed) and $Q_p A Q_p$ vanishes on $\mathbb{R}p$, $Q_p A Q_p$ admits N real eigenvalues which include 0 and $(\alpha_1, ..., \alpha_{N-1})$ which are the real eigenvalues of the restriction of $Q_p A Q_p$ to $(\mathbb{R}p)^\perp$. We set, where $|p| \neq 0$, $\lambda_i = \alpha_i / |p|$ for each i between 0 and $N-1$. When $A = D^2 u$ and $p = Du$, the λ_i are by Proposition 5.16 the principal curvatures of the level hypersurface of u. If $N = 2$, by Definition 5.12, $\lambda_1 = tr(Q_p A Q_p)/|p| = \text{curv}(u)$.

Corollary 21.7 *Let $N \geq 2$. Let T_t be a scale space satisfying the contrast and the euclidean invariance, then*

$$F(A,p) = |p| G(\lambda_1, ..., \lambda_{N-1}), \tag{21.19}$$
$$\text{for all} \quad A \in S^N,\ p \in I\!R^N, p \neq 0,$$

for some continuous function G defined on $I\!R^{N-1} \times [0, +\infty[$ which is symmetric with respect to $(\lambda_1, ..., \lambda_{N-1})$ and nondecreasing with respect to each λ_i $(1 \leq i \leq N-1)$.

Proof of Corollary 21.7. By Lemma 20.14 the function F associated with the scale space satisfies

$$F({}^tRAR, {}^tRp) = F(A,p) \quad \text{for all} \ \ A \in S^N,\ p \in I\!R^N,\ p \neq 0,\ R \in O^N, \tag{21.20}$$

where O^N denotes the group of orthogonal transforms of $I\!R^N$, that is the linear transforms preserving the euclidean and therefore satisfying $R^{-1} = {}^t R$.

The first step consists in showing that for $p \neq 0$ fixed, (21.20) implies that F is a function of $(\lambda_1, ..., \lambda_{N-1})$. In order to do so, we consider the subgroup of O^N defined by those transforms R that leave p invariant i.e. $Rp = p$ so that ${}^tRpR^{-1} = p$. Then, clearly Q_p and R commute and thus (21.20) implies that $F({}^tRQ_pAQ_pR, {}^tRp) = F(Q_pAQ_p, {}^tRp) = F(Q_pAQ_p, p)$. There therefore is a function G such that $F(A,p) = G_1(\lambda_1, ..., \lambda_{N-1}, p)$, where $(\lambda_1, ..., \lambda_{N-1})$ are the eigenvalues of Q_pAQ_p. Let R in O^N and $q = {}^t Rp$. We notice that $Q_q^t RARQ_q = {}^t RQ_pAQ_pR$. We deduce that $Q_q^t RARQ_q$ and Q_pAQ_p have the same eigenvalues. Thus, by using the euclidean invariance (21.20) again, $F(A,p) = F({}^tRAR, {}^tRp) = G_1(\alpha_1, ..., \alpha_{N-1}, q)$. Thus F only depends on the modulus of p and therefore we can write $F(A,p) = G_1(\alpha_1, ..., \alpha_{N-1}, |p|)$. We use again (21.15), which yields

$$G_1(\mu\alpha_1, ..., \mu\alpha_{N-1}, \mu|p|) = \mu G_1(\alpha_1, ..., \alpha_{N-1}, |p|).$$

Choosing $\mu = |p|^{-1}$, we obtain $F(A,p) = |p| G_1(\alpha_1/|p|, ..., \alpha_{N-1}/|p|, 1) = |p| G(\lambda_1, ..., \lambda_{N-1})$. In addition, G is a symmetric function of $(\lambda_1, ..., \lambda_{N-1})$, and, since F is monotone, G is clearly nondecreasing with respect to each λ_i for $1 \leq i \leq N-1$. □

21.3 Affine morphological scale spaces ($N \geq 3$).

Theorem 21.8 $(N \geq 2)$. *Let T_t be an affine and contrast invariant scale space satisfying $T_t(-u) = T_t(u)$. Assume that it is renormalized according to Lemma 20.20. Then its associated equation is*

$$\frac{\partial u}{\partial t} = |Du| \Big| \prod_{1 \leq i \leq N-1} t\lambda_i \Big|^{\frac{1}{N+1}} \ H\Big(\sum_{1 \leq i \leq N-1} sgn(\lambda_i) \Big) \tag{21.21}$$

where the λ_i are the mean curvatures of the level hypersurface, $sgn(\lambda_i)$ denotes the sign of λ_i and H is defined by

$$H(N-1) = -H(-(N-1)) = 1 \qquad \text{and } \forall i \in Z\!\!\!Z,\ |i| \neq N-1, \quad H(i) = 0$$

In other terms, H is equal to zero if all the λ_i have not the same sign.

Proof We begin by using the result of Corollary 21.7. The contrast and euclidean invariances imply that the scale space is associated an equation

$$\frac{\partial u}{\partial t} = F(D^2u, Du, t)$$

with

$$F(A, p, t) == |p|G(\lambda_1, ..., \lambda_{N-1}, t),$$

and G is symmetric with respect to the λ_i. If $p \neq 0$, we have set $\lambda_i = \frac{\alpha_i}{|p|}$, where α_i are the eigenvalues of the restriction of A to the hyperplane orthogonal to p, $p^\perp$ (see Corollary 21.7). In order to simplify the proof, we prefer to use here a more general form for the function F

$$F(A, p, t) = G_1(\alpha_1, ..., \alpha_{N-1}, |p|, t) \tag{21.22}$$

Of course, $|p|G(\lambda_1, ..., \lambda_{N-1}, t)$ is a particular instance of such a function G_1. Since the restriction of A to the hyperplane $p^\perp$ is a symmetric matrix, we can choose a set of orthogonal vectors $e_1, ..., e_{N-1}$ such that $\alpha_i = A(e_i, e_i)$. $(0 \leq i \leq N-1)$. Notice that each e_i is orthogonal to p. Then, in order to have a complete basis of $\mathbb{R}^N$, we set $e_N = p/|p|$. Let B_i, (for $1 \leq i \leq N-1$), be the linear transform from $\mathbb{R}^N$ into $\mathbb{R}^N$ defined by

$$B_i(e_1, ..., e_i, ..., e_{N-1}, e_N) = (e_1, ..., \beta e_i, ..., e_{N-1}, \beta^{-1}e_N),$$

for $\beta \in \mathbb{R}$, so that $det(B_i) = 1$. By (20.30) in Lemma 20.21, applied with $B = B_1$,

$$F(A, p, t) = G(\alpha_1\beta^2, \alpha_2, ..., \alpha_{N-1}, \beta^{-1}|p|, t).$$

By choosing $\beta = |\alpha_1|^{-1/2}$ we obtain

$$F(A, p, t) = G(sgn(\alpha_1), \alpha_2, ..., \alpha_{N-1}, |\alpha_1|^{\frac{1}{2}}|p|, t).$$

Iterating the same argument with $B = B_i$ for $i = 2$ to $i = N-1$ yields

$$F(A, p, t) = G\big(sgn(\alpha_1), ..., sgn(\alpha_{N-1}), |p| \mid \prod_{0 \leq i \leq N-1} \alpha_i|^{\frac{1}{2}}, t\big). \tag{21.23}$$

We now use again the scale invariance relation (20.30) in Lemma 20.21 and lemma 20.20, which yields

$$F(\beta^2 A, \beta p, t) = \beta F(A, p, \beta t)$$

and therefore

$$G\big(sgn(\alpha_1), ..., sgn(\alpha_{N-1}), \beta|p|\beta^{N-1}|\prod_{1 \leq i \leq N-1} \alpha_i|^{\frac{1}{2}}, t\big) = \beta G\big(sgn(\alpha_1), ..., sgn(\alpha_{N-1}), |p| \prod_{1 \leq i \leq N-1} |\alpha_i|^{\frac{1}{2}}, \beta t\big).$$

We deduce that

$$\beta^{-1}G\big(sgn(\alpha_1), ..., sgn(\alpha_{N-1}), \beta^N|p| \prod_{1 \leq i \leq N-1} |\alpha_i|^{\frac{1}{2}}, t\beta^{-1}\big) = G\big(sgn(\alpha_1), ..., sgn(\alpha_{N-1}), |p| \prod_{1 \leq i \leq N-1} |\alpha_i|^{\frac{1}{2}}, t\big).$$

Setting $\beta = t$, we obtain

$$F(A, p, t) = t^{-1}G\big(sgn(\alpha_1), ..., sgn(\alpha_{N-1}), t^N|p| \prod_{1 \leq i \leq N-1} |\alpha_i|^{\frac{1}{2}}, 1\big)$$

$$= t^{-1} G_1\big(t^N |p| \prod_{1 \leq i \leq N-1} |\alpha_i|^{\frac{1}{2}}, sgn(\alpha_1), ..., sgn(\alpha_{N-1})\big).$$

Now, by the contrast invariance relation (20.25) in Lemma 20.16, we also have

$$F(\beta^{-1} A, \beta^{-1} p, t) = \beta^{-1} F(A, p, t),$$

so that

$$\beta G_1\big(\beta^{-1} t^N |p| \prod_{1 \leq i \leq N-1} |\beta^{-1} \alpha_i|^{\frac{1}{2}}, ...\big) = G_1\big(t^N |p| | \prod_{1 \leq i \leq N-1} \alpha_i|^{\frac{1}{2}}, ...\big).$$

Choosing $\beta = \big(t^N |p| \prod_{1 \leq i \leq N-1} |\alpha_i|^{\frac{1}{2}}\big)^{\frac{2}{N+1}}$, the function F is reduced to a power function

$$F(A, p, t) = t^{\frac{N-1}{N+1}} |p|^{\frac{2}{N+1}} \ | \prod_{1 \leq i \leq N-1} \alpha_i|^{\frac{1}{N+1}} \quad G_1\big(1, sgn(\alpha_1), ..., sgn(\alpha_{N-1})\big)$$

$$= t^{\frac{N-1}{N+1}} |p|^{\frac{2}{N+1}} | \prod_{1 \leq i \leq N-1} \alpha_i|^{\frac{1}{N+1}} \quad H_1\big(sgn(\alpha_1), ..., sgn(\alpha_{N-1})\big)$$

for some function H_1. Since λ_i $(= \alpha_i / |p|)$, we finally obtain

$$F(A, p, t) = t^{\frac{N-1}{N+1}} |p|^{\frac{2}{N+1}} |p|^{\frac{N-1}{N+1}} | \prod_{1 \leq i \leq N-1} \lambda_i|^{\frac{1}{N+1}} \quad H_1\big(sgn(\lambda_1), ..., sgn(\lambda_{N-1})\big),$$

which yields

$$F(A, p, t) = |p| t^{\frac{N-1}{N+1}} \prod_{1 \leq i \leq N-1} |\lambda_i|^{\frac{1}{N+1}} \quad H_1\big(sgn(\lambda_1), ..., sgn(\lambda_{N-1})\big).$$

Now, H_1 must satisfy some properties : Since T_t is invariant by rotation, H_1 must be symmetric with respect to its variables. Because of the term $(\prod_{1 \leq i \leq N-1} |\lambda_i|^{\frac{1}{N+1}}$, we are allowed to set $H(sgn(\lambda_1), ..., sgn(\lambda_{N-1})) = 0$ if some λ_i vanishes. If all λ_i are nonzero, $H_1(sgn(\lambda_1), ..., sgn(\lambda_{N-1}))$ therefore only depends on the number of λ_i's which are negative and therefore upon the sum : $\sum_{1 \leq i \leq N-1} sgn(\lambda_i))$. Thus

$$H_1\big(sgn(\lambda_1), ..., sgn(\lambda_{N-1})\big) = H\big(\sum_{1 \leq i \leq N-1} sgn(\lambda_i)\big).$$

It is finally easy to see, by changing u into $-u$ and using the assumption $T_t(-u) = -T_t(u)$ and the regularity relation (20.4), that H satisfies $H(i) = -H(-i)$. At last, we know that H is a nondecreasing function, since F is nondecreasing with respect to A.

To summarize, we have

$$F(A, p, t) = |p| t^{\frac{N-1}{N+1}} \prod_{1 \leq i \leq N-1} |\lambda_i|^{\frac{1}{N+1}} \quad H\big(\sum_{1 \leq i \leq N-1} sgn(\lambda_i)\big), \tag{21.24}$$

where H is a nondecreasing real function such that $H(i) = -H(-i)$.

the case $N = 2$.

In that case, $N - 1 = 1$ and we deduce from (21.24),

$$F(A, p, t) = t^{\frac{1}{3}} |p| \big(\frac{1}{|p|} A\big(\frac{p^\perp}{|p|}, \frac{p^\perp}{|p|}\big)\big)^{\frac{1}{3}}.$$

Returning to the scale space equation, we have $p = Du$, $A = D^2u$ and $\mathrm{curv}(u) = \frac{1}{|Du|} D^2u(\frac{Du^\perp}{|Du|}, \frac{Du^\perp}{|Du|})$, so that the associated scale space equation is

$$\frac{\partial u}{\partial t} = F(D^2u, Du, t) = |Du|(\mathrm{curv}(u))^{\frac{1}{3}}.$$

We retrieve, as was to be expected, the equation (18.4).

The case $N = 3$.
The equation we obtain is

$$\frac{\partial u}{\partial t} = |Du| t^{\frac{1}{2}} |\lambda_1 \lambda_2|^{\frac{1}{4}} H\big(sgn(\lambda_1) + sgn(\lambda_2)\big)$$

where λ_1 and λ_2 are the principal curvatures of the level surface. Their product is nothing but the Gauss curvature $\mathrm{G}(u)$ of the level surface of u, so that we can write

$$\frac{\partial u}{\partial t} = |Du| t^{\frac{1}{2}} |\mathrm{G}(u)|^{\frac{1}{4}} H\big(sgn(\lambda_1) + sgn(\lambda_2)\big)$$

Since $sgn(\lambda_1) + sgn(\lambda_2)$ can only take three values : -2, 0, 2, and using $H(-2) = -H(2)$, we can write

$$H(2) = -H(-2) = b \geq 0 \qquad H(0) = a \qquad \text{and} \qquad |a| \leq b$$

Now, $F(A, Du, t)$ must be nondecreasing with respect to A and therefore with respect to the curvatures λ_1 and λ_2. It is easy to check, by taking particular values for λ_1 and λ_2 that a must be equal to 0. Indeed, if we choose two pairs of λ_i : $(-1, \alpha)$ and (α, α), for some α real and positive, then F must be larger for the second pair than for the first one, so that

$$\alpha H(0) \leq \alpha^2 H(2)$$

Letting α tend to 0, we obtain $H(0) = a \leq 0$. By choosing the two pairs $(-\alpha, \alpha)$ and $(-\alpha, 1)$ we obtain in the same way $H(0) = a \geq 0$. Thus $a = 0$. The consequence is that when the principal curvatures have opposite signs, then $F(A, Du, t)$ is equal to zero. At last, we obtain, up to a multiplicative constant for the second member, the equation

$$\frac{\partial u}{\partial t} = sgn(\lambda_1)\ |Du| t^{\frac{1}{2}} \big((\mathrm{G}(u))^+\big)^{\frac{1}{4}} \tag{21.25}$$

where x^+ stands for $\sup(0, x)$. This equation describes the unique multiscale analysis in dimension 3 which is both affine and contrast invariant and satisfies $T_t(u) = -T_t(-u)$.

The case of arbitrary dimension N. The same argument as dimension 3 is easily extended in any dimension and yields Equation (21.21), where H is defined by

$$H(N-1) = -H(-(N-1)) = 1 \qquad \text{and } \forall i \in \mathbb{Z},\ |i| \neq N-1, \quad H(i) = 0$$

In other terms, H is equal to zero if all the λ_i have not the same sign. □

References.

In this chapter, we followed an axiomatic presentation of scale spaces developped in [15], which is a simplified version of the original [15, 190] Other axiomatics for the affine scale spaces have been proposed by Olver et al. [327, 329]. The first complete mathematical study on the motion of a surface by its mean curvature is Brakke [62]. The main concern of scale space theory is : how regular is a surface which is smoothed by the mean curvature motion ? Huisken[213] proved that the smooth motion of a convex surface ended into a sphere before vanishing to a point, generalizing to more dimensions a Gage result for curve evolution. The question of whether a curvature motion ended or not into a sphere had been first considered by Firey for the Gauss curvature motion. Osher and Sethian developped the first numerical codes for the motion by mean curvature, where topological changes of the surface could be dealt with efficiently. Yuille [446] notices that the Koenderink-Van Dorn Dynamic Shape algorithm can create singularities in a dumb-bell shapes surface, the thin central part evolving into two separate pinches. This contradicts the main causality axiom in Scale Space : no creation of new features in the shape. See the comments by Koenderink [250]. The corresponding mathematical study of the phenomenon is due to Grayson [181]. Gradient bounds for the mean curvature equation are given in Barles [51]. A general survey of singularity formation by mean curvature is made in Angenent [33] and [11] prove the smoothness of the evolution of rotationally symmetric hypersurfaces, estimating the number of singularity points. On the theme of regularity and singularities by mean curvature flow, see also [137, 30, 214, 223]. Ishii and Souganidis [226] made a general theory of viscosity solutions for curvature equations of any type, including any power functions of the curvature and the gaussian curvature. A particular mention must be made of the Caselles et al. [92] on the properties of scale spaces in three dimension. They prove that the neckpinches do not occur in the dumb-bell by the affine scale space but show other examples of surfaces where anyway singularities may appear. Chow [104, 105] proved that the affine scale space deforms strictly convex surfaces into spheres, a result equivalent to the above mentionned Huisken result for the mean curvature motion.

Extensions of the affine scale space in two dimensions Several authors have made attempts to get in two dimensions a scale space still more invariant than the affine scale space, namely a projectively invariant scale space Clearly, the results of this chapter show that no local and causal multiscale analysis can be projective invariant : we pointed out that asking affine invariance exhausted all degrees of freedom in the choice of the partial differential equation. Thus, one of the requirements must be given up. Faugeras and Keriven [147, 144, 146, 145] and Bruckstein and Shaked [68] give up the maximum principle, but get higher order partial differential equations which can hardly be considered as smoothing equations. Mathematical existence proofs and numerical simulations of this projective curve scale space are open problems. See also Olver et al. [328]. Dibos [124, 122] does not give up locality or causality and is able to simulate numerically her scale space. This scale space does not depend anymore upon a single scale parameter, but on two. Geraets et al. [169, 170] propose affine invariant scale spaces for discrete sets and applications to object recognition. One of the first attempts to use the AMSS model for affine invariant shape recognition is Cohignac et al. [106]. A more complete and sophisticate attempt is [268] which performs image comparison by applying the affine scale space to all level lines of each image. Alvarez and Morales [17] used the affine scale space for corner and T-junction detection in digital images.

Chapter 22

Scale spaces of shapes

22.1 All shape scale spaces.

In this section, we shall list four principles which a shape scale space (identified with a curve scale space) should satisfy. We shall prove that a shape scale space must, according to these principles, be a curvature motion equation whose form is slightly more general than the intrinsic heat equation. For some of the principles discussed here, it will be useful to consider both a Jordan curve and the bounded set $\mathbf{X}$ surrounded by the Jordan curve. So we call *shape or silhouette* any bounded closed set $\mathbf{X}$ whose boundary is a Jordan curve of $I\!R^2$. We denote by $T_t(\mathbf{X})$ the *shape of* $\mathbf{X}$ *analysed at scale* t. We call shape scale space any family of operators $(T_t)_{t\geq 0}$ acting on shapes and set $\mathbf{X}(t) = T_t(\mathbf{X})$.

As in Chapter 20, we suppose that the operators T_t satisfy a pyramidal assumption, $T_t = T_{t,s}T_s$. We now consider a translation in the shape analysis framework of the other basic assumptions made for image scale space. The next assumption is an adaptation to shape analysis of the local comparison principle.

We denote by $B(\mathbf{x}, r)$ the open disk with center $\mathbf{x}$ and radius r, and for a shape $\mathbf{X}$, by $\partial\mathbf{X}$ the boundary of $\mathbf{X}$. Assume that $\mathbf{X}$ and $\mathbf{Y}$ are two compact shapes and that for some $\mathbf{x} \in \partial\mathbf{Y}$ and some $r > 0$, one has $\mathbf{X} \cap B(\mathbf{x}, r) \subset \mathbf{Y} \cap B(\mathbf{x}, r)$. Assume further that the inclusion is strict in the sense that $\partial\mathbf{X}$ and $\partial\mathbf{Y}$ only meet possibly at $\mathbf{x}$. Then we shall say that *the shape* $\mathbf{X}$ *is included in shape* $\mathbf{Y}$ *around* $\mathbf{x}$.

Definition 22.1 *We say that* T_t *satisfies the* `[Shape local inclusion principle]` *, if for all* $\mathbf{X}$ *and* $\mathbf{Y}$ *subsets of* $I\!R^2$ *such that* $\mathbf{X}$ *is included in* $\mathbf{Y}$ *around* $\mathbf{x}$*, then for* h *small enough,* $T_{t+h,t}(\mathbf{X}) \cap B(\mathbf{x}, r) \subset T_{t+h,t}(\mathbf{Y}) \cap B(\mathbf{x}, r)$.

This axiom has two aspects : First, it implies that the value of $T_{t+h,t}(\mathbf{X})$ for h small, at any point $\mathbf{x}$, is determined by the behaviour of $\mathbf{X}$ near $\mathbf{x}$. Second, taking two shapes $\mathbf{X} \subset \mathbf{Y}$, it implies for r large enough that $T_{t+h,t}(\mathbf{X}) \subset T_{t+h,t}(\mathbf{Y})$, that is, a `Global inclusion principle`. Both preceding principles allow, as we shall see, to localize the shape analysis process in space and scale. In order to totally specify a scale space, we only need, as we shall prove, to say what the scale space makes of very simple shapes. We add two principles which state what happens to disks. Thanks to the local inclusion principle, disks will appear as the "basis" on which every more complex shape can be decomposed : When we know what happens to disks in scale space, we can deduce the destiny of every other shape !

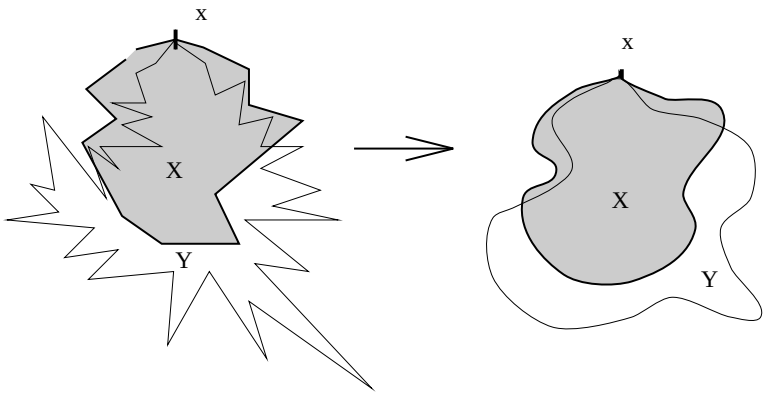

Figure 22.1: Local Inclusion Principle (Definition 22.6).

[Isotropy] Let $D = D(\mathbf{x}, 1/r)$ be a disk with curvature $1/r$ and center $\mathbf{x}$. Then $T_{t+h,t}(D)$ is a disk with radius $\rho(t, h, 1/r)$ and center $\mathbf{x}$.

[Regularity] For all $t, r, h > 0$, function $h \to \rho(t, h, 1/r)$ is differentiable with respect to h at $h = 0$ and the derivative $g(t, 1/r) = \frac{\partial \rho}{\partial h}(t, 0, 1/r)$ is continuous with respect to $1/r$ and has a limit $g(t, 0)$ as $r \to \infty$.

The last assumption, that $g(t, 0)$ exists, is no serious restriction in view of the next lemma.

Lemma 22.2 *The function $g(t, \kappa) = \frac{\partial \rho}{\partial h}(t, 0, \kappa)$ is nondecreasing with respect to κ.*

Proof Let $r > r'$, set $0 = (0, 0)$ and $a = (r - r', 0)$ and consider the disks $D = D(0, r)$ and $D' = D(a, r')$, which are tangent at $\mathbf{x} = (r, 0)$. By the isotropy principle, $T_{t,t+h}D$ is a disk with center 0 and radius $\rho(t, h, \frac{1}{r})$ and $T_{t,t+h}D'$ a disk with center a and radius $\rho(t, h, \frac{1}{r'})$. By the shape local inclusion principle, the second disk is contained is the first, which implies $r' - \rho(t, h, \frac{1}{r'}) \geq r - \rho(t, h, \frac{1}{r})$. Dividing this relation by h and letting $h \to 0$ yields by the regularity assumption

$$g(t, \frac{1}{r'}) = \frac{\partial \rho}{\partial h}(t, 0, \frac{1}{r'}) \geq \frac{\partial \rho}{\partial h}(t, 0, \frac{1}{r}) = g(t, \frac{1}{r}).$$

Setting $\kappa' = \frac{1}{r'}$ and $\kappa = \frac{1}{r}$ we obtain

$$\kappa' \geq \kappa \Rightarrow g(t, \kappa') \geq g(t, \kappa).$$

□

The last shape preserving principles which we shall consider here are the scale invariance and the affine invariance of the shape scale space. Define $\mathrm{A}\mathbf{X} = \{ \mathrm{A}\mathbf{x}, \mathbf{x} \in \mathbf{X} \}$ for any linear map A.

[Affine invariance]. There exists a C^1 function $t'(t, \mathrm{A}) \geq 0$ defined for any A and $t \geq 0$, such that $\mathrm{A}T_{t'(t, \mathrm{A}), t'(s, \mathrm{A})} = T_{t,s}\,\mathrm{A}$. In case where we restrict this last relation to homotheties, we shall say that the scale space is *scale invariant*. By using Lemma 20.20, we can always take $t'(t, \mathrm{A}) = t.|det\ \mathrm{A}|^{1/2}$ up to

a rescaling, so that we finally write.

$$\mathrm{A}T_{t.|\det\ \mathrm{A}|^{1/2},s.|\det\ \mathrm{A}|^{1/2}} = T_{t,s}\ \mathrm{A}. \tag{22.1}$$

Let $\mathbf{x}(s)$ be a curve parameterized by length. We recall that in Definition 5.2 we have defined an intrinsic curvature vector $Curv\mathbf{x}(s))$. The modulus of the curvature vector can be interpreted as the inverse of the radius of the osculatory circle to the curve at $\mathbf{x}(s)$. The curvature is zero if the radius is infinite. When $Curv\mathbf{x}(s) \neq 0$, the intrinsic normal vector $\mathbf{n}(\mathbf{x})(s) = \frac{Curv\mathbf{X}(s)}{|Curv\mathbf{X}(s)|}$ has norm one and can be defined in more geometric terms as the unit vector normal to the curve and pointing towards the concavity of the curve.

Theorem 22.3 *(i) Let T_t be a shape scale space satisfying the four shape analysis principles (pyramidal architecture, shape local inclusion, isotropy and regularity). Consider an initial Jordan curve $\mathbf{x}_0(s)$. Then at each point (t,s) around which $\mathbf{x}(t,s) = (T_t\mathbf{x}_0)(s)$ is differentiable with respect to t and C^2 with respect to an euclidean parameter s, $\mathbf{x}(t,s)$ satisfies a curvature equation*

$$\frac{\partial \mathbf{x}}{\partial t}(t,s) = g(t, |Curv\mathbf{x}(t,s)|)\mathbf{n}(t,s). \tag{22.2}$$

where $\kappa \to g(t,\kappa)$ is a nondecreasing function.
(ii) If in addition the scale space is affine invariant, then the equation of the scale space is

$$\frac{\partial \mathbf{x}}{\partial t}(t,s) = (t|Curv\mathbf{x}(t,s)|)^{\frac{1}{3}}\mathbf{n}(t,s). \tag{22.3}$$

With obvious abbreviations, we shall write this last equation

$$\frac{\partial \mathbf{x}}{\partial t} = (t.Curv(\mathbf{x}))^{\frac{1}{3}}\mathbf{n}(\mathbf{x}) \tag{22.4}$$

and call it Affine Scale Space (A.S.S.)

Proof The proof of (i) is essentially identical to the proof of Theorem 20.6 : Instead of inserting locally the image function between two quadratic forms, we shall insert locally the curve $s \to \mathbf{x}(t,s)$ between two circles (surosculator and subosculator) for which we know the infinitesimal evolution by the isotropy and regularity principles. Denote by $X(t)$ the compact plane set whose boundary is the curve $\mathbf{x}(t,s)$. Let $\mathbf{x}$ be a point of this curve. We shall also when convenient denote this point by $\mathbf{x}(t)$. Let us consider a subosculatory and a surosculatory disk to the curve at $\mathbf{x}$. One is the closed disk D' with radius $r = \frac{1}{\kappa - \epsilon}$ and the other one the closed disk D with radius $r = \frac{1}{\kappa + \epsilon}$ and both are tangent to the curve $\mathbf{x}(t,s)$ at $\mathbf{x}$. (In the case where $\kappa = 0$, we simply take for D' the half plane tangent to the shape at $\mathbf{x}$ containing D . This does not alter what follows). Let $B(\mathbf{x},r)$ be a disk with r small enough so that $D \cap B(\mathbf{x},r) \subset X(t) \cap B(\mathbf{x},r) \subset D' \cap B(\mathbf{x},r)$ and ∂D, $\partial D'$ and the curve $\mathbf{x}(t,s)$ only meet at $\mathbf{x}$ inside $B(\mathbf{x},r)$. Applying the local inclusion principle, we deduce that

$$T_{t+h,t}(D) \cap B(\mathbf{x},r) \subset T_{t+h,t}(X) \cap B(\mathbf{x},r) \subset T_{t+h,t}(D') \cap B(\mathbf{x},r).$$

Thus, denoting by $\mathbf{x}(t+h)$ the point of $\partial T_{t+h}(X)$ such that $\mathbf{x}(t+h) - \mathbf{x}(t,s)$ is parallel to $\mathbf{n}(\mathbf{x}(t,s))$, we obtain

$$\rho(t,h,\kappa-\epsilon) - \rho(t,0,\kappa-\epsilon) \leq (\mathbf{x}(t+h) - \mathbf{x}(t)).\mathbf{n}(\mathbf{x}(t)) \leq \rho(t,h,\kappa+\epsilon) - \rho(t,0,\kappa+\epsilon)$$

Dividing by h and passing to the inf- and sup-limits when h tends to 0 yields

$$\frac{\partial \rho}{\partial h}(t,0,\kappa-\epsilon) \leq \liminf_{h\to 0} \frac{\mathbf{x}(t+h)-\mathbf{x}(t)}{h}.\mathbf{n}(\mathbf{x}(t)) \leq$$

$$\limsup_{h\to 0} \frac{\mathbf{x}(t+h)-\mathbf{x}(t)}{h}.\mathbf{n}(\mathbf{x}(t)) \leq \frac{\partial \rho}{\partial h}(t,0,\kappa+\epsilon)$$

By the regularity principle, $\kappa \to \frac{\partial \rho}{\partial h}(t,0,\kappa)$ is continuous. Passing to the limit when ϵ tends to 0,

$$\frac{\partial \mathbf{x}(t)}{\partial t}(.\mathbf{n}(\mathbf{x})(t) = g(t,|Curv\mathbf{x}(t)|), \tag{22.5}$$

which yields Equation 22.2 because by assumption $\frac{\partial \mathbf{x}(t)}{\partial t}$ is colinear to $\mathbf{n}(\mathbf{x}(t))$.
(ii) After renormalization, we can use the identity

$$T_{t+h,t}D_\lambda = D_\lambda T_{(t+h)\lambda,t\lambda}$$

, where $D_\lambda = \lambda Id$. We deduce that the function ρ of the basic principle must satisfy $\rho(t,h,\frac{\lambda}{r}) = \frac{1}{\lambda}\rho(\lambda t,\lambda h,1/r)$ After differentiation with respect to h at 0, we obtain

$$g(t,\lambda s) = g(\lambda t,s)$$

for any $t>0$, $\lambda>0$ and $s\in I\!R$. Changing t in t/λ and taking $\lambda = 1/t$ we get $g(t,s) = g(1,ts) = \beta(ts)$, β being defined as $\beta(x) = g(1,x)$. Let us now use the full affine invariance. After renormalization, we can use the identity $T_{t+h,t}A = AT_{t+h,t}$, where A is the linear transform with determinant equal to 1,

$$(x,y) \to (\lambda x, \frac{y}{\lambda}), \quad \lambda > 0$$

Let us apply the identity $T_{t+h,t}AD = AT_{t+h,t}D$ to the unit disk $D = D(0,1)$.Consider the point $x_0 = (1,0)$ on the boundary of D. The velocity by T_t of x_0 is $(-\beta(t),0)$, and this velocity is transformed by A into $-\lambda\beta(t)$. Since AD is an ellipse with curvature λ^3 at point Ax_0, the velocity of Ax_0 is $(-\beta(t.\lambda^3),0)$. So we obtain $\beta(t.\lambda^3) = \lambda\beta(t)$. Taking $t=1$, we get $\beta(x) = a.x^{1/3}$ with$(a=\beta(1))$. □

22.2 From curve motion to image motion and viscosity solutions.

Theorem 22.4 *If u is a viscosity solution of*

$$\frac{\partial u}{\partial t} = |Du|F(curv(u))$$

and if $\mathbf{x}(t)$ is a C^∞ (in time and space) level line of u, then this level line also is a classical solution of

$$\frac{\partial \mathbf{x}}{\partial t} = F(Curv(\mathbf{x}))\mathbf{n}$$

Lemma 22.5 *Let $u(t,\mathbf{x})$ be a continuous function whose isolevelset $C(t) = \partial\{\mathbf{x}, u(\mathbf{x}) \geq \lambda\} = \partial\{\mathbf{x}, u(\mathbf{x}) < \lambda\}$, is, for some level λ and some interval of time around t_0, a finite union of disjoint closed C^∞ Jordan curves. Then there exists $\epsilon > 0$, a C^∞ function $\varphi(t,\mathbf{x})$ and a continuous non decreasing contrast change g, such that*

- *$\varphi \geq g(u)$, for all $[t_0-\epsilon, t_0+\epsilon]$ and $\mathbf{x}\in I\!R^2$,*
- *$\varphi(t_0,\mathbf{x}_0) = u(t_0,\mathbf{x}_0)$,*
- *$C(t)$ is a isolevelset of both $g(u)$ and φ for $t\in[t_0-\epsilon,t_0+\epsilon]$.*

Proof of lemma 22.5 We consider for $\mathbf{x} \in I\!R^2$, $\psi(t,\mathbf{x}) = d((t,\mathbf{x}), C(t))$, where d denotes the signed Hausdorff distance, i.e.

- $d((t,\mathbf{x}), C(t)) > 0$ if $\mathbf{x}$ is surrounded by $C(t)$
- $d((t,\mathbf{x}), C(t)) < 0$ otherwise

Then, there exists ϵ such that for all $t \in [t_0 - \epsilon, t_0 + \epsilon]$ and for all $\mathbf{x}$ such that $|d((t,\mathbf{x}), C(t))| < \epsilon$, $\psi(t,\mathbf{x})$ is a C^∞ function. We can extend ψ into a C^∞ function on $I\!R^+ \times I\!R^2$, φ. By Lemma 8.13, we can find a continuous and nondecreasing function g such that $\varphi \geq g(u)$ and $\varphi = g(u)$ on $C(t)$ for all $t \in [t_0 - \epsilon, t_0 + \epsilon]$. (Notice that Lemma 8.13 is stated for functions defined on $I\!R^N$, but still holds for functions defined in a closed subset of $I\!R^N$, without any change.) □

Proof of theorem 22.4 We apply lemma 22.5 and the fact that u is a viscosity solution. We deduce that

$$\frac{\partial \varphi}{\partial t} \leq F(curv(\varphi))|D\varphi|$$

Let us consider the normal flow $\mathbf{x}(t)$ associated with $\varphi(t,\mathbf{x})$, satisfying therefore the relation (6.6),

$$\frac{\partial \mathbf{x}}{\partial t} = -(\frac{\partial \varphi}{\partial t}\frac{D\varphi}{|D\varphi|^2})(t,\mathbf{x}(t)).$$

Since the level lines of $g(u)$ and φ coincide for $t \in [t_0 - \epsilon, t_0 + \epsilon]$, we have $curv(\varphi) = curv(u) = Curv(\mathbf{x})$. Thus,

$$\frac{\partial \mathbf{x}}{\partial t}.D\varphi \leq F(Curv(\mathbf{x}))|D\varphi|,$$

which yields,

$$\frac{\partial \mathbf{x}}{\partial t}.\mathbf{n} \leq F(Curv(\mathbf{x})).$$

We get the converse inequality by similar arguments.

In the following, our purpose is to define an image evolution by handling separately all level lines of the initial image. The level lines evolution is defined by e.g. the intrinsic heat equation, see chapter 6 or any curve scale space ensuring the existence of a smooth curve evolution. We need to define a kind of image u for which all iso level sets $\{\mathbf{x}, u(\mathbf{x}) = \lambda\}$ are uniquely described as finite unions of Jordan curves.

Definition 22.6 *We say that a curve scale-space satisfies a uniform local comparison principle, if there is a function $h(r,\epsilon) > 0$ satisfying the following condition:*

Let c_1 and c_2 any two Jordan curves which can be locally parameterized as graphs $y = f_1(x)$, $y = f_2(x)$, $-r \leq x \leq r$ and such that

$$r \geq f_1(x) \geq f_2(x) + \epsilon x^2 \geq -r$$

for every $-r \leq x \leq r$. Then the evolution of c_1 and c_2 by the curve scale-space T_t defines locally two graphs for $t \leq h(r,\epsilon)$ and $-r/2 \leq x \leq r/2$: $y = f_1(t,x)$, $y = f_2(t,x)$ such that

$$f_1(t,x) \geq f_2(t,x).$$

Remark 22.7 *The Grayson evolution and the affine shortening (respectively $G(s) = s$ and $G(s) = s^{\frac{1}{3}}$, satisfy this property. See [180], lemma ??, page ??, [29] lemma ??, page ??*

Definition 22.8 Curve evolution equation and Curve Scale Space.

We consider a curve evolution equation

$$\frac{\partial \mathbf{x}}{\partial t} = G(Curv(\mathbf{x}))\mathbf{n}, \tag{22.6}$$

and its associated function evolution equation

$$\frac{\partial u}{\partial t} = |Du|G(curv(u)) = F(D^2u, Du) \tag{22.7}$$

where G is a non-decreasing continuous function and F is admissible (Definition 17.1). (Note that the admissibility implies that $r(G(1/r)$ is bounded for r around 0).

We say that (22.6) is a smoothing scale-space, if we can assert that

1. *Every closed Jordan curve with finite length $\mathbf{x}_0(s)$ evolves by (22.6) into a C^2 closed Jordan curve $\mathbf{x}(t,s)$ such that $\mathbf{x}(0,s) = \mathbf{x}_0(s)$.*
2. *The curve map $\mathbf{x}_0(.) \to \mathbf{x}(t,.)$ is continuous for the Hausdorff distance,*
3. *satisfies the global shape inclusion principle*
4. *and the uniform shape local inclusion principle.*

Definition 22.9 *We define a class $\mathcal{T}$ of subsets of $\mathbb{R}^2$ which we call "regular" sets. A regular set is a closed subset of $\mathbb{R}^2$ whose boundary is a locally finite union of disjoint closed C^2 Jordan curves. If $\mathbf{c}$ is a Jordan curve we denote by $\mathrm{Int}(\mathbf{c})$ the unique closed bounded subset of $\mathbb{R}^2$ whose boundary is $\mathbf{c}$. If X belongs to $\mathcal{T}$ has no unbounded connected component, we can write*

$$X = \cup_{i=1}^{k} \left(\mathrm{Int}(\mathbf{c}_i) \setminus \cup_{j=1}^{l_i} \mathrm{interiorInt}(\mathbf{c}_{i,j}) \right), \tag{22.8}$$

where the index i runs over all connected components of X, the index $j \in \{1, ..., l_i\}$ runs over all bounded connected components of the complementary set to the i-th connected component (its holes). If X has an unbounded connected component, then X^c has no unbounded connected component and satisfies (22.8).

We associate with $\mathcal{T}$ the class $\mathcal{F}$ of continuous functions u such that $\mathcal{X}_\lambda u$ belongs to $\mathcal{T}$ for every λ except a finite number of them which we call critical levels. The functions of $\mathcal{F}$ will be called "regular".

Example 22.10 *It is easily seen that a periodic Morse function, that is, a C^2 function such that $detD^2u(\mathbf{x}) \neq 0$ for all $\mathbf{x}$, is an instance of regular function.*

Definition 22.11 *(Curve evolution as a set operator $\mathtt{T}_t$).*
We consider a smoothing curve scale-space $\mathbf{c} \to \mathbf{c}(t)$, as defined in Definition (22.8). For every set in $\mathcal{T}$, if X satisfies:

$$X = \cup_{i=1}^{k} \left(\mathrm{Int}(\mathbf{c}_i) \setminus \cup_{j=1}^{l_i} \mathrm{interiorInt}(\mathbf{c}_{i,j}) \right),$$

we set

$$\mathtt{T}_t(X) = \cup_{i=1}^{k} \left(\mathrm{Int}(\mathbf{c}_i(t)) \setminus \cup_{j=1}^{l_i} \mathrm{interiorInt}(\mathbf{c}_{i,j}(t)) \right),$$

where $\mathbf{c}_i(t)$ and $\mathbf{c}_{i,j}(t)$ denote the curves evolved at time t by the curve scale-space.

If $X \in \mathcal{T}$ has an unbounded component, we set $\mathtt{T}_t(X) = (\mathtt{T}_t(X^c))^c$.

We also set $\mathtt{T}(\emptyset) = \emptyset$ and therefore $\mathtt{T}(\mathbb{R}^2) = \mathbb{R}^2$.

Lemma 22.12 *$\mathtt{T}_t$ is monotone on $\mathcal{T}$ and transforms a set of $\mathcal{T}$ into a set pf $\mathcal{T}$.*

Proof If $X \subset Y$, then each external Jordan curve of a connected component of Y is enclosed by an exterior Jordan curve of a connected component of X and each internal Jordan curve of X is enclosed by an internal Jordan curve of Y. Since the evolution by a smoothing scale-space preserves the inclusion of curves, we easily deduce the announced statement. □

Definition 22.13 *Let $u \in \mathcal{F}$ be a regular function. We call level line evolution of u the unique monotone functional extension T to $\mathcal{F}$ of the set evolution $\mathtt{T}_t$ given by Proposition 7.9. We then have*
(i) $T(u)$ is regular.
(ii) If u is $K-$Lipschitz, then $T(u)$ is equal almost everywhere to a $K-$Lipschitz function.
(iii) $\mathtt{T}(\mathcal{X}_\lambda u) = \mathcal{X}_\lambda(T(u))$, a.e and for almost every $\lambda \in \mathbb{R}$ such that λ is not critical for u.
(iv) As a consequence, T_t is recursive: $T_{t+t'} = T_t T_{t'}$ for all $t, t' \geq 0$.

Proof (i) By Exercise 8.6, we know that $T_t(u)$ is obtained from u by a sup-inf operator, so that the Lipschitz constant of u is preserved.
(ii) Let $\lambda_0 \notin \{\lambda_1, ..., \lambda_l\}$ critical levels of u, that is levels for which $\mathcal{X}_\lambda u$ is not in $\mathcal{T}$. As λ tends to λ_0, $\mathcal{X}_\lambda u$ tends to $\mathcal{X}_{\lambda_0} u$ locally for the Hausdorff distance. (Each Jordan curve of the boundary of $\mathcal{X}_\lambda u$ tends to a corresponding curve of the boundary of $\mathcal{X}_{\lambda_0} u$). By Definition 22.8.2, we then have

$$\mathtt{T}_t(\mathcal{X}_\lambda u) \to \mathtt{T}_t(\mathcal{X}_{\lambda_0} u)$$

locally for the Hausdorff distance. Since we can choose an increasing sequence $\lambda_k \to \lambda_0$ of non critical values for u, we then have by (22.13.ii) $\mathtt{T}_t(\mathcal{X}_{\lambda_k} u) = \mathcal{X}_{\lambda_k}(T_t u)$, a.e. By passing to the limit and using (4.1.ii) we obtain

$$\mathtt{T}_t(\mathcal{X}_{\lambda_0} u = \lim_{\lambda_k \to \lambda_0} \mathcal{X}_{\lambda_k}(T_t u) = \bigcap_{\lambda_k \to \lambda_0} \mathcal{X}_{\lambda_k}(T_t u) = \mathcal{X}_{\lambda_0}(T_t u) \text{ a.e.}$$

(iii) Obvious consequence of (ii) and the pyramidality of the curve scale-space. □

Lemma 22.14 *Let u and $v \in \mathcal{F}$ be two regular functions such that $u(\mathbf{y}) > v(\mathbf{y})$ for every $\mathbf{y} \in D(\mathbf{x}, r) \backslash \{\mathbf{x}\}$. Then there is some $h_0 > 0$ such that for all $h < h_0$, $(T_h u)(\mathbf{x}) \geq (T_h v)(\mathbf{x})$. In other terms the level line evolution T_t satisfies the local comparison principle.*

Proof The set operator $\mathtt{T}_t$ satisfies by definition the shape local inclusion principle (Definition 22.8.3). By, comparing the level lines of u and v in the disk $D(\mathbf{x}, r)$, it is easily checked that the function operator T_t inherits a local comparison principle. □

Lemma 22.15 *(i) Let G be a function associated with a smoothing curve scale-space (Definition 22.8). Then, for every C^2 increasing radial function Q we have for $t \to 0$ and when $DQ(\mathbf{x}) \neq 0$*

$$\frac{T_t(Q)(\mathbf{x}) - Q(\mathbf{x})}{t} \to |DQ(\mathbf{x})| G(curv(Q)(\mathbf{x})),$$

(ii) The same result is true for any $C^\infty \cap \mathcal{F}$ function f.
(iii) Let f be a $C^\infty \cap \mathcal{F}$ function such that $f(\mathbf{x}) = 0$ for all $\mathbf{x} \in D(\mathbf{x}_0, r)$. Then

$$\lim_{t \to 0} |\frac{T_t(f)(\mathbf{x}_0) - f(\mathbf{x}_0)}{t}| \to 0.$$

Remark 22.16 *We choose the above functions in $\mathcal{F}$, so that their evolution by the curve scale-space is imediately described by the evolution of their level lines.*

Proof of Lemma 22.15 (i) This is an easy computation which we leave as exercise.
(ii) Consider a C^∞ function f, such that $Df(\mathbf{x}_0) \neq 0$. It is easily seen that we can find a radial function $v_{\pm\epsilon}(\mathbf{x})$ such that

$$v_{\pm\epsilon}(\mathbf{x}_0) = f(\mathbf{x}_0), \qquad Dv_{\pm\epsilon}(\mathbf{x}_0) = Df(\mathbf{x}_0), \qquad curv(v_{\pm\epsilon})(\mathbf{x}_0) = curv(f(\mathbf{x}_0) \pm \epsilon)$$

and, in addition,

$$v_{-\epsilon}(\mathbf{x}) \leq f(\mathbf{x}) \leq v_\epsilon(\mathbf{x})$$

in a neighborhood $D(\mathbf{x}_0, r)$. By lowering r if necessary we can assume by the implicit function theorem that for every $\lambda \in [v(\mathbf{x}_0) - \eta, v(\mathbf{x}_0) + \eta]$ and η small enough, the level lines c_1, c_2 with level λ of v_ϵ and f respectively satisfy the assumptions of Definition (22.6). (The same applies to the level lines of f and $v_{-\epsilon}$.

Thus, for $t \leq h(r, \epsilon)$ the inclusion of level lines of $v_{\pm\epsilon}$ and f is preserved for all levels close to $\lambda_0 = f(\mathbf{x}_0)$. We deduce that in a neighborhood of $\mathbf{x}_0$, say $D(\mathbf{x}_0, r/4)$ we have

$$T_h(v_{-\epsilon})(\mathbf{x}_0) \leq T_h(f)(\mathbf{x}_0) \leq T_h(v_\epsilon)(\mathbf{x}_0).$$

Substracting $f(\mathbf{x}_0) = v_{\pm\epsilon}(\mathbf{x}_0)$, dividing by h and letting $h \to 0$ we deduce from (i) that

$$G((curv(f))(\mathbf{x}_0) - \epsilon) \leq \liminf_{h\to 0} \frac{T_h(f)(\mathbf{x}_0) - f(\mathbf{x}_0)}{h}$$

$$\leq \limsup_{h\to 0} \frac{T_h(f)(\mathbf{x}_0) - f(\mathbf{x}_0)}{h} \leq G((curv(f))(\mathbf{x}_0) + \epsilon).$$

We conclude by letting $\epsilon \to 0$ and using the continuity of G.
(iii) No level line of f meets $D(\mathbf{x}_0, r)$. Since the evolution by the curve scale-space is continuous for the Hausdorff distance, we deduce that for $t \leq t_0(r)$, no level line of $f(t) = T_t(f)$ meets $D(\mathbf{x}_0, r/2)$. Thus $f(t)(\mathbf{x}_0) = 0$ for $t \leq t_0(r)$. □

Theorem 22.17 *We consider a smoothing curve scale-space $\frac{\partial \mathbf{c}}{\partial t} = G(Curv(\mathbf{c}))\mathbf{n}$, as defined in Definition (22.8) and its unique extension T_t to the set of regular function $\mathcal{F}$. Let $u_0 \in \mathcal{F}$. Then $u(t, \mathbf{x}) = T_t(u_0)(\mathbf{x})$ is viscosity solution of*

$$\frac{\partial u}{\partial t} = |Du| G(curv(u)).$$

Proof In order to prove that $u(t, \mathbf{x}) = T_t(u_0)(\mathbf{x})$ is a viscosity solution of (20.10), we only to check that u is a viscosity subsolution of (20.10), the supersolution property being shown in the same way. Let $(t_0, \mathbf{x}_0)$ in $[0, \infty] \times I\!R^N$ be a strict local maximum point of $u - \phi$ where $\phi(t, \mathbf{x}) = f(\mathbf{x}) + g(t)$ where f is regular, $C^\infty(I\!R^2)$ and equal to a quadratic form in a neighborhood of $\mathbf{x}_0$, and g is $C^\infty(I\!R)$. By Exercise 17.1, we are allowed to do this restriction on test functions. By Lemma 17.8, it is enough to show that when $Df(\mathbf{x}_0) \neq 0$

$$\frac{\partial g}{\partial t}(t_0, \mathbf{x}_0) - |Df(\mathbf{x}_0)| G(curv(f)(\mathbf{x}_0)) \leq 0. \tag{22.9}$$

and when $Df(\mathbf{x}_0) = 0$ and $D^2 f(\mathbf{x}_0) = 0$,

$$\frac{\partial g}{\partial t}(t_0, \mathbf{x}_0) \leq 0. \tag{22.10}$$

By Lemma 20.10 we can assume (w.l.o.g.) that the inequality between u and ϕ is global, i.e. $u \leq \phi$ on $[0, \infty] \times I\!R^N$. Let $h > 0$. We set $u(t) = u(t, .)$ for all functions u on $I\!R^N * [0, \infty[$. We have $u(t) \leq \phi(t)$, then $u(t_0 - h, \mathbf{x}) \leq \phi(t_0 - h, \mathbf{x})$ for all $h < t_0$ and $\mathbf{x}$. Using the monotonicity of T_h (Definition 22.13), we obtain

$$T_h(u(t_0 - h))(\mathbf{x}_0) \leq T_h(\phi(t_0 - h))(\mathbf{x}_0)$$

Now, by the commutation of T_t with the addition of constants and since $\phi(t, \mathbf{x}) = f(\mathbf{x}) + g(t)$ we have

$$T_h(u(t_0 - h))(\mathbf{x}_0) \leq T_h(\phi(t_0 - h))(\mathbf{x}_0) \leq T_h(f)(\mathbf{x}_0) + g(t_0 - h).$$

We deduce that

$$u(t_0, \mathbf{x}_0) - (T_h(f)(\mathbf{x}_0) + g(t_0 - h)) \leq 0$$

Since $u(t_0, \mathbf{x}_0) = \phi(t_0, \mathbf{x}_0)$ and then

$$g(t_0) - g(t_0 - h) \leq T_h(f)(\mathbf{x}_0) - f(\mathbf{x}_0)$$

Then,

$$(g(t_0) - g(t_0 - h))/h \leq (T_h(f)(\mathbf{x}_0) - f(\mathbf{x}_0))/h$$

If $Df(\mathbf{x}_0) \neq 0$ then letting h go to 0, using (i), we recover

$$\frac{\partial g}{\partial t}(t_0 - h) - |Df(\mathbf{x}_0)| G(curv(f)(\mathbf{x}_0)) \leq 0$$

By letting h tends to zero, we conclude.

If $Df(\mathbf{x}_0) = 0$ and $D^2 f(\mathbf{x}_0) = 0$ then by the assumption on f we know that f is zero on a neighborhood of $\mathbf{x}_0$, and by Lemma 22.15.(iii) we deduce that for h small enough, $|T_h(f)(\mathbf{x}_0) - f(\mathbf{x}_0)| = 0$. □

22.3 Motion by curvature and Total Variation

Lemma 22.18 *Let $\mathbf{x}(t, s)$ satistying the curve equation evolution*

$$\frac{\partial \mathbf{x}}{\partial t} = F(Curv(\mathbf{x}))\mathbf{n}$$

. Assume that for some $(t, s) \in [t_1, t_2] \times I\!R$, $\mathbf{x}$ is C^∞. As we have seen, this holds for e.g. $F(x) = x$.

Let $L(t)$ be the perimeter of $\mathbf{x}(t, .)$. Then $\forall t \in]t_1, t_2[$, we have

$$\frac{\partial L}{\partial t}(t) = -\int_0^L (t) Curv(\mathbf{x}(t, s)) F(Curv(\mathbf{x}(t, s)))ds$$

where s is the arc length parametrization. As consequence the curve equation evolution makes the perimeter of the curves decrease.

Proof We choose to parametrize the curve with z between 0 and 1. We thus have

$$L(t) = \int_0^1 |\frac{\partial \mathbf{x}}{\partial z}(t,z)|dz$$

Thus

$$\frac{\partial L}{partialt}(t) = \int_0^1 \frac{\partial}{\partial t}||\frac{\partial \mathbf{x}}{\partial z}(t,z)|dz$$

Denoting by $'$ the partial derivation with respect to z:

$$= \int_0^1 \frac{\mathbf{x}'(z,t)}{|fxx'(z,t)|}\frac{\partial}{\partial t}(\mathbf{x}')(t,z)dz$$

By integration by parts

$$= -\int_0^1 Curv(\mathbf{x}(t,s))\mathbf{n}\frac{\partial \mathbf{x}}{\partial t}(z,t)\mathbf{x}'(t,z)dz$$

Using the curve equation evolution, we obtain:

$$\frac{\partial L}{partialt}(t) = -\int_0^1 Curv(\mathbf{x}(t,s))F(Curv(\mathbf{x}(t,s))dz$$

$$= -\int_0^L (t)Curv(\mathbf{x}(t,s))F(Curv(\mathbf{x}(t,s)))ds$$

□

Theorem 22.19 Total Variation of a Lipschitz function. *[?] Let u be a K-Lipschitz function defined on a square C. Then Du is integrable. And the total variation of u satisfies:*

$$TV(u) = \int_C |Du(\mathbf{x})|d\mathbf{x} \leq Karea(C)$$

Definition 22.20 Total Variation by the Coarea formula *Let u be a Lipschitz and regular (see definition in the preceding section) function defined on a square $C = [0,1[^2$, then we define the Total Variation of the image u by:*

$$\int_C |Du(\mathbf{x})|d\mathbf{x} = \int_{\lambda \in I\!R} Perimeter(\partial \mathcal{X}_\lambda(u)) \cap C d\lambda$$

This last definition makes sense since the regularity of u insures that the Perimeter of the boundaries of the λ level set of u is finite except for a set of λ whose measure is zero.

We consider a regular and K-Lipschitz function u_0 and its evolution $u(t)$ by a regular curve scale space. By 22.13 we know that $u(t)$ is K-Lipschitz and regular, so that we can estimate its total variation.

Theorem 22.21 *We consider a regular and K-Lipschitz function u_0 and its evolution $u(t)$ by a regular curve scale space. By 22.13 we know that $u(t)$ is K-Lipschitz and regular, so that we can estimate its total variation.*

One has

$$\frac{\partial TV(u(t))}{\partial t} \leq 0$$

or in other terms, a regular curve scale space decreases the total variation.

Proof This is an imediat consequence of 22.18 and the coarea formula. □

.

Remark 22.22 *Let us give a heuristic that illustrates (but not proves) the same result. Consider the total variation as an energy, and compute the gradient of the energy. (Assume that $u(t,\mathbf{x})$ is C^∞ and $Du \neq 0$). Denoting by $G_{TV}(u)$ the gradient of the total variation at function u. It is easily proved (integration by parts) that $G(u) = -div(Du/|Du|) = curv(u)$.*

$$TV(u(t+dt)) = TV(u(t)) + \int_C (u(t+dt) - u(t))(G_{TV}u)(\mathbf{x}))d\mathbf{x}$$

$$\frac{\partial TV(u(t))}{\partial t} = -dt \int_C \frac{\partial u}{\partial t}(t,\mathbf{x}) curv(u) d\mathbf{x}$$

References.

The need for a multiscale curvature in shape analysis

Global shape recognition is relatively easy by normalization algorithm, which singles out in each equivalent class of shapes a unique representative. Invariant comparison of two shapes boils then down to a comparison of both invariant representatives. Affine normalization, and even projective normalization can be made for invariant shape recognition [40, 354, 355]. See also [239]. A relatively efficient comparison of clouds of points can also be performed by geometric hashing [222] or by Hough transform [210] Now, most shapes under recognition undergo distorsions and occlusions which reduce the efficiency of global comparison methods. Such methods assume that the shapes undergoing a recognition process are in good shape. This is seldom the case with shapes extracted by a computer program from a digital image. Thus, the smoothing of shapes and their comparisons parts by parts are necessary. This is the main reason of the development of shape scale spaces. The problem of defining an efficient shape scale space is adressed in the early works of Asada and Brady [37], Horn et Weldon [208], Lowe [270] and Mackworth-Mockhtarian [271]. After these papers, the search for a good definition of a multiscale curvature has been aknowledged as the key to efficient shape recognition : Dudek, [131, 129, 130, 133], Mackworth and Mokhtarian[272], Cohignac [106], Williams [438], Mokhtarian [142], Lisani et al. [268].

Morphology and structural decompositions of shapes We do not treat here two other ways to analyse a shape, one called *structural analysis* whose aim it is to decompose the shape into its significant parts and the other one using the shape's skeleton. In both cases, however, scale space is called to help since skeletonization is made by multiscale erosions and shape structural decompositions by openings, closings, erosions, dilations, and their differences at different scales [310, 325, 343, 449]. See also Maragos definition of a pattern spectrum [278]. The work of Kimia, Tannenbaum and Zucker [243, 242] can be viewed as a synthesis of

Mathematical proofs of scale space properties As we already mentionned, the proof that curvature motion in dimension 2 is a perfect scale space was made from the mathematical point view by Gage[166], Gage and Hamilton [167] and Grayson [179] The mathematical arguments for existence and uniqueness of

the solutions of the Affine Morphological Scale Space were given in Alvarez et al. Alvarez:1993:AFE in the viscosity framework and by Angenent et al. [29] for the affine curve shortening. The evolution of convex curves had been studied in Sapiro & Tannenbaum [371, 372].

Axiomatics of shape scale spaces

Our presentation of the requirements for the shape scale space follows a simplified version first proposed in [?], which emphasizes the shape inclusion principle as the main axiom, following the mathematical view of De Giorgi and his disciples [367]. Olver, Sapiro and Tannenbaum have proposed an alternative axiomatics based on the concept of invariant heat flow [328].

Chapter 23

Movie Scale-spaces.

This chapter is concerned with the axiomatic characterization of the multiscale analyses $\{T_t\}_{t\geq 0}$ of movies. We shall formalize a movie as a bounded function $u_0(x, y, \theta)$ defined on $\mathbb{R}^3$, where x and y are the spatial variables and θ the time variable. We note $\mathbf{x} = (x, y, \theta)$.

As in the preceding chapters, we assume that T_t is **causal** (Definition 20.5), **Translation invariant** (Definition **??**) and **invariant by grey level translation** (Definition 20.7). Therefore, as shown in Chapter 20, there exists $T_{t,s}$ such that $T_t = T_{t,s}T_s$, for all $t \geq s \geq 0$. And,

$$((T_{t+h,t}u - u)/h)(\mathbf{x}) \to F(D^2u(\mathbf{x}), Du(\mathbf{x}), t)$$

as h tends to 0^+ for all u and $\mathbf{x}$ where u is C^2. The properties of F are the same as in chapter 20, that is, $F(A, \mathbf{p}, t)$ is nondecreasing with respect to its first argument, $F(A, \mathbf{p}, t)$ is continuous at all points where $\mathbf{p} \neq 0$. But, now F has ten scalar arguments.

Finally, we assume that the equation

$$\frac{\partial u}{\partial t} = F(D^2u, Du, t)$$

a unique viscosity solution $u(x, y, \theta, t)$, (this will of course be checked a posteriori for the models we derive).

23.1 Geometrical axioms for the movie scale-space.

Let us first define the geometrical axioms for the multiscale analysis of movies. All axioms considered in chapter 20 make sense, but we need to specify them in order to take into account the special role of time (θ). (For example, we shall not consider invariance by spatio-temporal rotations as an essential property...) This will change a little the assumptions on geometrical invariance. As usual we will denote for any affine operator C of $\mathbb{R}^3$, by Cu the function $Cu(\mathbf{x}) = u(C\mathbf{x})$.

The first property states that the analysis be invariant under all linear transforms of the spatial plane $\mathbb{R}^2 * \{0\}$. That is, when we apply the same affine transform on each image of the movie.

Definition 23.1 *We shall say that a movie scale-space T_t is* **affine invariant** *if, for any linear map B of the form*

$$\begin{pmatrix} a & b & 0 \\ c & d & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

there exists $t'(t,B)$ *such that* $B(T_{t'(t,B)}u) = T_t(Bu)$, *and* $B(T_{t'(t,B),t'(s,B)}u) = T_{t,s}(Bu)$.

We also state a weaker property than the affine invariance, by restricting the invariance to the rotations of the two first coordinates, and the homotheties.

Definition 23.2 *We shall say that a movie scale-space* T_t *is euclidean invariant if for any linear map*

$$A = \begin{pmatrix} a\cos(b) & -a\sin(b) & 0 \\ a\sin(b) & a\cos(b) & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

there exists a scale $t'(t,A)$ *such that* $A(T_{t'(t,A)}u) = T_t(Au)$ *and* $A(T_{t'(t,A),t'(s,A)}u) = T_{t,s}(Au)$

Note that the t' is the same for the two definitions 23.1 and 23.2. It establishs the link between the space dimension and the scale. Since in the following either the affine or the Euclidean invariance will be considered, we shall always have this link. We now establish the link between time and scale, by considering the homotheties with respect to time θ. (We accelerate or decelerate uniformly the movie.)

Definition 23.3 *For any* e *in* $\mathbb{R}^+$ *we define by* S_e *the linear map* $S_e(x,y,\theta) = (x,y,e\theta)$ *We shall say that a movie scale-space* T_t *is* **time scale invariant** *if there exists* $t''(t,e)$ *such that*

$$S_e(T_{t''(t,e)}u) = T_t(S_e u) \qquad \text{and } S_e(T_{t''(t,e),t''(s,e)}u) = T_{t,s}(S_e u)$$

Of course, the function t'' *can be different from the function* t' *of definitions 23.1 and 23.2.*

Now, we want to state the scale invariance, as done in chapter 20.4. We begin by noticing that the combination of the affine (or Euclidean) invariance and the time scale invariance implies invariance with respect to homotheties of $\mathbb{R}^3$. That is, setting $H_\lambda = \lambda Id$, we have for some function $\tau(t,\lambda)$:

$$H_\lambda(T_{\tau(t,\lambda)}u) = T_t(H_\lambda u)$$

So, for scale invariance we could impose that the function τ is differentiable with respect to λ and that $\partial\tau/\partial\lambda(t,1)$ is continuous and positive. Now, we prefer to obtain the scale-invariance assumption by using the affine and time scale invariances.

Lemma 20.20 implies that t' is a function only of t and of the determinant of B. Then, setting $\lambda = det(B)$, we assume that $t'(t,\lambda)$ is differentiable with respect to λ at $\lambda = 1$, and that the function $g(t) = \frac{\partial t'}{\partial \lambda}(t,1)$ is continuous for $t > 0$. We assume the same thing for the time: We assume that $t''(t,e)$ is differentiable with respect to e at $e = 1$, and that $h(t) = \frac{\partial t''}{\partial e}(t,1)$ is continuous. For the scale normalization we must impose in addition that at least one of $g(t)$ or $h(t)$ is positive for $t > 0$. If we assume $g(t) > 0$, then the scale normalization is established with respect to spatial variables. And, by an easy adaptation of Lemma 20.20, we deduce that we can normalize the relation between t, B and t' so that

$$t' = (det(B))^{\frac{1}{2}} t \tag{23.1}$$

Thus the affine invariance is reduced to the property :

$$F(BA^tB, Bp, t) = |det(B)|^{\frac{1}{2}} F(A, p, t|det(B)|^{\frac{1}{2}}) \tag{23.2}$$

If now we assume $h(t) > 0$, then the scale normalization is established with respect to time. And then time scale invariance is reduced to

$$F(S_e A S_e, S_e p, t) = eF(A, p, et) \tag{23.3}$$

Of course, since these assumptions imply a re-normalisation, we can not assume both. In the following, we shall assume that at least one of the two conditions is achieved. We then state the regular scale invariance axiom :

Definition 23.4 *We shall say that a scale-space T_t satisfying the Affine or Euclidean invariance and the time-scale invariance is* **scale-invariant** *if*

(i) $t'(t,\lambda)$ is differentiable with respect to λ at $\lambda = 1$, and $g(t) = \frac{\partial t'}{\partial \lambda}(t,1)$ is continuous for $t > 0$

(ii) $t''(t,e)$ is differentiable with respect to e at $e = 1$, and $h(t) = \frac{\partial t''}{\partial e}(t,1)$ is continuous for $t > 0$.

(iii) One of the function g or h is positive, and the other one is continuous at $t = 0$.

(iv) $t \to T_t$ is injective.

(where t' and t'' are these defined in 23.1 or 23.2 and 23.3).

For the last "geometrical axiom" we assume that the analysis is invariant under "travelling" : a motion of a whole single picture with constant velocity v does not alter the analysis. We denote by B_v the galilean translation operator,

$$B_{v=(v_x,v_y)} u(x,y,\theta) = u(x - v_x\theta, y - v_y\theta, \theta)$$

In fact B_v is an affine operator,

$$B_{v=(v_x,v_y)} = \begin{pmatrix} 1 & 0 & -v_x \\ 0 & 1 & -v_y \\ 0 & 0 & 1 \end{pmatrix}$$

Definition 23.5 *We shall say that a movie scale-space is* **Galilean invariant** *if for any v and t, there exists $t^*(t, B_v)$ such that*

$$B_v(T_{t^*} u) = T_t(B_v u), \text{ and } B_v(T_{t^*(t,v),t^*(s,v)} u) = T_{t,s}(B_v u)$$

$t^*(t, B_{-v}) = t^*(t, B_v)$, and t^* is nondecreasing with respect to t.

The second part means that reversing time should not alter the analysis. Let us simplify the definition. By using Lemma 20.20(i), we have

$$t^*(t^*(t, B_v), B_v) = t^*(t^*(t, B_v), B_{-v}) = t^*(t, B_v B_{-v}) = t^*(t, Id) = t.$$

Repeating the argument of the step (ii) of the proof of the Lemma 20.20, we deduce from this relation that $t^*(t, B(v)) = t$. Thus the Galilean invariance reduces to the simpler relation (to which we give the same name)

$$B_v(T_t u) = T_t(B_v u) \Leftrightarrow F({}^t B_v A B_v, {}^t B_v \mathbf{p}, t) = F(A, \mathbf{p}, t) \qquad \forall A \text{ in } S^3, \mathbf{p} \in I\!R^3 \tag{23.4}$$

Finally, we state the morphological property, (as in definition 20.24):

Definition 23.6 *We shall say that a movie scale-space is* **contrast invariant** *if for any monotone and continuous function h from $I\!R$ into $I\!R$, $T_t h(u) = h(T_t u)$*

We have seen in lemma 20.16 that this implies

$$F(\mu A + \lambda \mathbf{p} \otimes \mathbf{p}, \mu \mathbf{p}, t) = \mu F(A, \mathbf{p}, t), \tag{23.5}$$

for every real values λ, μ, every symmetric matrix A and every three-dimensional vector $\mathbf{p}$.

23.2 Optical flow and properties for a movie scale-space.

The aim of this section is not to do a exhaustive list of the techniques for optical flow estimation, but from general considerations we will remark that lot of methods involve a step of smoothing, which could be modelized by a scale-space. In parallel, we will notice that the contrast and the Galilean invariances are not only compatible but somehow justified by the aim of estimating an optical flow. This will make more clear what motivated the choice of the properties stated in the preceding section.

The notion of optical flow has been introduced in the studies of human preattentive perception of motion. The optical flow associates with each point of the movie, a vector representing the optical velocity of this point. We shall denote by v the optical flow vector ($v = (v_x, v_y)$ is in $\mathbb{R}^2$), and by $\mathbf{v}$ the vector $(v_x, v_y, 1)$. So that if $\Delta\theta$ is the time interval between two frames, $\mathbf{x} + \mathbf{v}(\mathbf{x})\Delta\theta$ denotes the point $\mathbf{x}$ shifted by $v(\mathbf{x})$ in the next frame.

The classical definition involves a conservation assumption, which generally is that the points move with a constant gray level (u : the gray level value). From a discrete point of view, we are looking for $\mathbf{v}(\mathbf{x})$ such that ([149, 172, ?, 207, 414],...)

$$u(\mathbf{x} + \mathbf{v}(\mathbf{x})\Delta\theta) = u(\mathbf{x}) + o(\Delta\theta) \Leftrightarrow Du.\mathbf{v} = 0 \tag{23.6}$$

This leads us to compare the gray level value from one frame to the next and to associate the points which have the same intensity. Considering that the single value $u(\mathbf{x})$ is not a reliable information because of the many perturbation in capturing the image, the images are often smoothed before doing this matching. Of course, it would be possible to use an image scale-space, that is to smooth each frame independently. But, we might probably do better by smoothing the whole movie, with interactions between the different frames. Following the idea of Marr, Hildreth, Koenderink, and Witkin many authors proposed to use the convolution by the 3D Gaussian function G_t (the 3D heat equation). And, then they check :

$$(G_t * u)(\mathbf{x} + \mathbf{v}(\mathbf{x})\Delta\theta) = (G_t * u)(\mathbf{x}) \tag{23.7}$$

where $*$ denotes the convolution operator. The main problem of this formulation is that it is not equivalent for two movies u and $\tilde{u}$ representing the same object with different constant velocity. For example, consider that the movie $\tilde{u}$ is an accelerated version of u, $\tilde{u}(x, y, \theta) = u(x, y, 2\theta) = u(A\mathbf{x})$. Set $\mathbf{v}_1$ (resp. $\mathbf{v}_2$) the velocity at the point $\mathbf{x}$ in the movie u (resp. at the point $A\mathbf{x}$ in the movie $\tilde{u}$). We have $\mathbf{v}_2 = 2\mathbf{v}_1$. Now, after the smoothing, using the formula (23.7), $\mathbf{v}_2$ must satisfy

$$(G_t * u(A.))(\mathbf{x} + \mathbf{v}_2\Delta\theta) = (G_t * u(A.))(\mathbf{x}) \tag{23.8}$$

And, we easily see that since in general $(G_t * u(A.)) \neq (G_t * u)(A.)$, after a such smoothing we shall not always obtain with formula (23.7), $\mathbf{v}_2 = 2\mathbf{v}_1$. Indeed, in the two cases, the smoothing is not done in the same way : because this linear smoothing is not Galilean invariant. Therefore a such smoothing implies some perturbation into the estimation of the velocities.

Adelson and Bergen [8], and Heeger [197] propose in order to avoid such problem, to design "oriented smoothing". Such an approach yields more Galilean invariance, even if, of course, we cannot exactly recover all the directions. (It would involve an infinite number of filters !)

Let us note also that the equation (23.6) is contrast invariant. Indeed one can apply a change of contrast for the entire movie : change u into $\tilde{u} = g(u)$, where g is strictly monotonous function from $\mathbb{R}$

into $\mathbb{R}$, then the equation (23.6) with $\tilde{u}$ is strictly equivalent to the equation with u :

$$u(\mathbf{x}+\mathbf{v}(\mathbf{x})\Delta\theta) = u(\mathbf{x}) \Leftrightarrow (g(u))(\mathbf{x}+\mathbf{v}(\mathbf{x})\Delta\theta) = (g(u))(\mathbf{x})$$

for any strictly monotonous change of contrast g.

It is important that this property be conserved after a smoothing of the movie u. Once more if we apply the linear smoothing defined by the convolution by the 3D Gaussian kernel, we lost this property. Indeed

$$(G_t * u)(\mathbf{x}+\mathbf{v}(\mathbf{x})\Delta\theta) = (G_t * u)(\mathbf{x}) \qquad \text{is not equivalent to}$$

$$(G_t * (g(u)))(\mathbf{x}+\mathbf{v}(\mathbf{x})\Delta\theta) = (G_t * (g(u)))(\mathbf{x})$$

except for some specific change of contrast, or kind of motion. In order to keep the equivalence after smoothing it is necessary that the scale-space be contrast invariant as it has been defined in the preceding section.

As well known, the conservation law (23.6) only gives the component of the optical flow in the direction of the spatial gradient. The other component remains indeterminated. The usual approach to determine the optical flow then involves balance between the conservation law and some smoothing constraint on the flow. Since it is not our subject here, we refer to the papers of Barron and al [55], Snyder [387], Nagel [312], Nagel and Enkelmann [?]...

First, we can remark that most of the approaches involve derivatives of the intensity of the movie, that by itself can justify the fact to smooth the movie before.

Secondly, the question occurs to know whether of not it is possible to smooth the movie so that resulting trajectories (this needs to be defined, but at least say the level surfaces, since due to conservation law trajectories are embedded within them) will be smoothed as well.

In conclusion, optical flow approaches often lead back to the problem of the definition of a smoothing. And we do not know a priori how much we have to smooth : the degree of smoothing is a free scale parameter. This indicates that a multi-scale analysis must be applied. In addition we have seen that the conservation law justifies the contrast and the Galilean invariances for the scale-space.

23.3 The axioms lead to an equation.

We are now going to introduce some useful notation.

1. We denote by $\nabla u = (\frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}, 0)$ the spatial gradient of the movie $u(x,y,\theta)$. When $\nabla u \neq 0$, we associate with $Du = (\frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}, \frac{\partial u}{\partial \theta})$ the two normal vectors $\mathbf{e}^{\perp}$ and $\mathbf{e}^{\pm}$ defined by

$$\mathbf{e}^{\perp} = \frac{1}{|\nabla u|}(-\frac{\partial u}{\partial y}, \frac{\partial u}{\partial x}, 0) \qquad \mathbf{e}^{\pm} = \frac{1}{|\nabla u||Du|}(\frac{\partial u}{\partial x}\frac{\partial u}{\partial \theta}, \frac{\partial u}{\partial y}\frac{\partial u}{\partial \theta}, -((\frac{\partial u}{\partial x})^2 + (\frac{\partial u}{\partial x})^2))$$

   When ∇u is not equal to zero, $\{Du, \mathbf{e}^{\perp}, \mathbf{e}^{\pm}\}$ is an orthonormal basis of $\mathbb{R}^3$. To be noted that $\mathbf{e}^{\perp}$ is spatial, that is it does not have a temporal component.

2. Again when $\nabla u \neq 0$, we then define

$$\Gamma_1 = (D^2u)(\mathbf{e}^{\perp}, \mathbf{e}^{\perp}), \qquad \Gamma_2 = (D^2u)(\mathbf{e}^{\perp}, \mathbf{e}^{\pm}), \qquad \Gamma_3 = (D^2u)(\mathbf{e}^{\pm}, \mathbf{e}^{\pm}).$$

Then Γ_1 is the second derivative of u in the direction $Du^{\perp}$, Γ_3 in the direction of $Du^{\pm}$, and Γ_2 the cross derivative in both directions.

3. Then, the spatial curvature $curv(u)$ is given by

$$curv(u) = \frac{\Gamma_1}{|\nabla u|}.$$

4. The gaussian curvature $\mathrm{G}(u)$ is given by

$$\mathrm{G}(u)\frac{\Gamma_1\Gamma_3 - \Gamma_2^2}{|Du|^2}$$

At last, we introduce the "apparent acceleration", as a normalized ratius between the gaussian curvature and the spatial curvature : given by

$$accel(u) = \frac{\mathrm{G}(u)}{curv(u)}\frac{|Du|^4}{|\nabla u|^4} = (\frac{|Du|}{|\nabla u|})^2(\Gamma_3 - \frac{\Gamma_2^2}{\Gamma_1})/|\nabla u|$$

Theorem 23.7 *Let a multiscale analysis T_t be causal (as defined in theorem 20.6), translation, Euclidean, Galilean, and constrast invariant. Then, there exists a function F such that T_t is governed by the equation*

$$\frac{\partial u}{\partial t} = |\nabla u|\ F(curv(u), accel(u), t) \tag{23.9}$$

(for the exact meaning of "governed by", we refer to the theorem 20.10.)

If in addition, T_t is affine, time-scale and time invariant then the only possible scale-space equations are

$$\textit{(AMG)} \qquad \frac{\partial u}{\partial t} = |\nabla u|\ curv(u)^{\frac{1-q}{3}}\ (sgn(curv(u))accel(u)^q)^+ \tag{23.10}$$

for some $q \in]0,1[$, or

$$(q=0) \qquad \frac{\partial u}{\partial t} = |\nabla u| curv(u)^{\frac{1}{3}} \tag{23.11}$$

$$(q=0) \qquad \frac{\partial u}{\partial t} = |\nabla u| curv(u)^{\frac{1}{3}} (sgn(accel(u)curv(u))^+ \tag{23.12}$$

$$(q=1) \qquad \frac{\partial u}{\partial t} = |\nabla u| sgn(curv(u))(sgn(curv(u))accel(u))^+ \tag{23.13}$$

In the above formulae, we use the convention that the power preserves the sign, that is $a^q = |a|^q \mathrm{sgn}(a)$. And we set $x^+ = sup(0,x)$.

Remark. Before begining with the proof of the theorem, let us notice that the terms appearing in equation (23.10) are not defined everywhere. Indeed, we can write $curv(u)$ only when $|\nabla u| \neq 0$, and $accel(u)$ only when $\nabla u \neq 0$ and $\Gamma_1 \neq 0$ (then $curv(u) \neq 0$). So, we must specify what happens when one of these conditions does not hold. Equation (23.10) is equivalent to

$$\frac{\partial u}{\partial t} = |\nabla u|^{\frac{2-8q}{3}} \Gamma_1^{\frac{1-4q}{3}} (\Gamma_1\Gamma_3 - \Gamma_2^2)^{q+} |Du|^{2q}$$

By continuity, when Γ_1 tends to zero, we set $\frac{\partial u}{\partial t} = 0$.

The case $\nabla u = 0$ is more problematic. We distinguish three cases :

- If $q < 1/4$, the right hand side ot the equation is continuous and we obtain, when ∇u tends to zero, $\frac{\partial u}{\partial t} = 0$.
- In the case $q = 1/4$, which is a limit case, ∇u does not appear in the equation. Now, the definitions of Γ_1, Γ_2,... depend on the direction of ∇u. We have in this case

$$\frac{\partial u}{\partial t} = |Du|^{\frac{1}{2}}(\Gamma_1\Gamma_3 - \Gamma_2^2)^{\frac{1}{4}+}$$

  where, $(\Gamma_1\Gamma_3 - \Gamma_2^2)$ is the determinant of D^2u restricted to the orthogonal plan to Du. If $|Du| \neq 0$, this determinant is defined independently of the Γ_i, and the formulation makes sense. Now, if $|Du|$ tends to 0, by continuity we have $\frac{\partial u}{\partial t} = 0$.
- At last, if $q > 1/4$, Equation (23.10) has singularities since the right hand side of this equation may tend to infinity when ∇u tends to zero.

Let us now set the obtained relation between space, time and scale.

Corollary 23.8 *Let A be an affine transform of the coordinates*

$$\begin{pmatrix} a & b & 0 \\ c & d & 0 \\ 0 & 0 & e \end{pmatrix} \text{ for any } a,\ b,\ c,\ d,\ e \in I\!R$$

and let $p = \sqrt{ad - bc}$. Then, the multiscale analysis defined by equation (23.10) satifies $A(T_\tau u) = T_t(Au)$ with

$$\tau(A, t) = (p^{4(\frac{1-q}{3})}e^{2q})t \tag{23.14}$$

We see in relation (23.14), that q is a parameter which represents the respective weights between space variables and time variables in the equation. For example, by taking $q = 0$, we remove the time dependance in the equation and we obtain the purely spatial affine and constrast invariant scale-space (or a slight variant). On the other side by taking $q = 1$, we remove the space dependance of the scale : we obtain the equation (23.13). At last, by taking $q = \frac{1}{4}$, we impose an homogeneous dependance in time and space. $\tau = pe^{\frac{1}{2}}\ t = (det(A)^{\frac{1}{2}})\ t$ In that case, by formulating the equation with $\mathrm{G}(u)$ the gaussian curvature of u, we obtain

$$\frac{\partial u}{\partial t} = |Du|(\mathrm{G}(u)^+)^{\frac{1}{4}} \tag{23.15}$$

which is the unique contrast and 3D affine invariant scale-space as described in chapter 21.8.

Let us before begining the proof of the theorem give a hint on the kind of smoothing the equation (23.10) should do on a movie. Let us decompose this equation into two parts

$$\frac{\partial u}{\partial t} = \qquad |\nabla u|\ curv(u)^{power...} \qquad (\mathrm{sgn}(curv(u))accel(u)^{power...})^+$$

The first term $curv(u)^{power...}$ is roughly a term of spatial diffusion, and then tends to remove objects when $t \to \infty$. It's quite close from the diffusion term of affine and contrast invariant scale-space of static images.

The second term $accel(u)...$ can be seen as the speed of this spatial diffusion. The bigger is *accel*, faster the spatial diffusion is executed. As we shall see in the following the differential operator *accel* can be interpreted as some kind of acceleration of objects in the movie. So, we can conclude that the

equation will smooth (and then remove) faster the object with big acceleration, than object with low acceleration. Therefore we can expect that this will produce a discrimination between trajectories (smooth and unsmooth).

Proof of Theorem 23.7 The proof is essentially based on algebraic calculations. Its main ingredient is that the terms $|\nabla u|^3 curv(u)$ and $|Du|^4 G(u) = |\nabla u|^4 curv(u) accel(u)$ are affine covariant of degree 2,2,0 and 2,2,2, with respect to the coordinates (x, y, θ).

Since the proof is quite long and technical, we refer to [15]. □

.

23.4 Optical flow and apparent acceleration.

In this section, we shall give to $accel(u)$ a cinematic interpretation as an "apparent acceleration". As pointed before, the conservation law related to the optical flow fixes only the component of the flow in the direction of the spatial gradient.

First, we shall see that the model (23.10) and the definition of $accel(u)$ can be associated with a special choice for the other component the apparent velocity. This choice corresponds to the a priori assumption that only objects in translation are observed. In other terms, $accel(u)$ gives the correct estimate of the acceleration of objects when they are in translation motion. Secondly, we will establish a formula that provides an estimation of *accel* without any calculating of the apparent velocity.

In all this section, we work only at points where $\nabla u \neq 0$.

What are the possible velocities ? We define the optical flow $\vec{v}(x, y, \theta)$ as a function from $\mathbb{R}^3$ into $\mathbb{R}^2$ representing the velocity of the point (x, y) at time θ. As before, we add a third component to the flow, which will always be equal to 1 : $\mathbf{v}(x, y, \theta) = (\vec{v}(x, y, \theta), 1)$. We denote by $\mathcal{W}$ the set of "possible" velocity vectors

$$\mathcal{W} = \{ \mathbf{v} = (\vec{v}, 1) \text{ for all } \vec{v} \text{ in } \mathbb{R}^2 \} \tag{23.16}$$

Assuming the conservation law, the optical flow is a vector of $\mathcal{W}$ which is orthogonal to Du, therefore when $Du \neq 0$, it belongs to the set $\mathcal{V}$:

$$\mathcal{V} = \{ \mathbf{v}_\mu = \frac{|Du|}{|\nabla u|}(\mu \mathbf{e}^\perp - \mathbf{e}^\pm), \text{ for all } \mu \in \mathbb{R} \} \tag{23.17}$$

All $\mathbf{v}_\mu$ have their component in the direction of ∇u fixed to $-\frac{u_\theta}{|\nabla u|}$. We have one free parameter μ left. It corresponds to the component of the velocity vector in the spatial direction orthogonal to ∇u, that is by definition : $\mathbf{e}^\perp$. In the next paragraph, we define μ so that $accel(u)$ is an apparent acceleration.

Definition 23.9 *Definition of the "velocity vector". When ∇u and $curv(u) \neq 0$, we define the "velocity vector": $\mathbf{V}$ by*

$$\mathbf{V} = \frac{|Du|}{|\nabla u|}(\frac{\Gamma_2}{\Gamma_1}\mathbf{e}^\perp - \mathbf{e}^\pm) \tag{23.18}$$

Then, if we set $v_1 = (\mathbf{V}.\nabla u)/|\nabla u|$ (resp. $v_2 = (\mathbf{V}.\mathbf{e}^{\perp})/|\mathbf{e}^{\perp}|$), the component of $\mathbf{V}$ in the direction (resp. orthogonal direction) of the spatial gradient ∇u, we have:

$$v_1 = -\frac{u_\theta}{|\nabla u|} \qquad v_2 = \frac{|Du|}{|\nabla u|}\frac{\Gamma_2}{\Gamma_1} \tag{23.19}$$

Proposition 23.10 *Let $\vec{i}, \vec{j}$ be an orthonormal basis of the image plane. Consider a picture in translation motion with velocity $\vec{v} = (v^x, v^y)$: $u(x, y, \theta) = w(x - \int_0^\theta v^x(\theta)d\theta, y - \int_0^\theta v^y(\theta)d\theta)$. Then, at every points such that $\nabla u \neq 0$ and $curv(u) \neq 0$, $\vec{v}$ satisfies the explicit formula*

$$(\vec{v}, 1) = \mathbf{V}$$

In other terms, the definition (23.9) of the flow $\mathbf{V}$ is exact for any translation motion.

The definition of the optical flow that fixes one component of the flow corresponds to say that points move on their space-time level surface (gray-level does not change). Fixing the other component as we do with the definition 23.9 is to make the choice of a travelling direction on the space-time level surface. With the definition 23.9, we choose the direction which does not change the orientation of the spatial gradient.

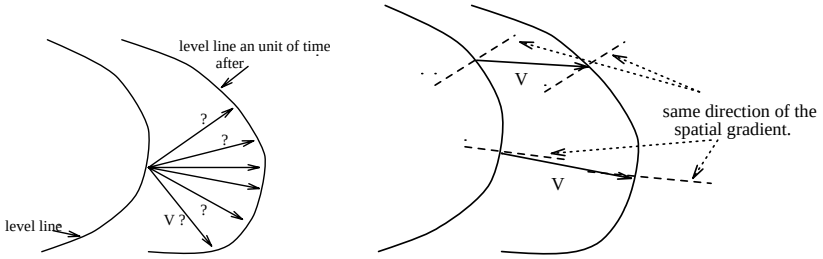

Figure 23.1: According to the optical flow definition, all above drawn velocity vectors are possible, since they allow the moving point to remain on the same level surface. One possibility to get rid of this ambiguity is choose as velocity the direction which does not change the orientation of the spatial gradient.

Of course, in general, the velocity vector $\mathbf{V}$ is not equal to the real velocity for others motions than the translations, but we shall consider it, for any type of movement. In others words we make for a point a choice of trajectory along the the iso-surface it belongs.

We shall now look for simpler expressions and interpretation of $accel(u)$. The next proposition shows that first, *accel* can be seen as an apparent acceleration and second as a curvature in space-time of our choice of trajectories along iso-surface.

Proposition 23.11 *1. accel as an apparent acceleration. For all points such that $\nabla u \neq 0$ and $curv(u) \neq 0$, let $\mathbf{V} = (v_x, v_y, 1)$ be the velocity vector defined as above (23.9), and v_1 its component in the direction of the spatial gradient.*

$$\begin{aligned} accel(u) &= -\frac{\mathrm{D}v_1}{\mathrm{D}\theta} = -(v^x\frac{\partial v_1}{\partial x} + v^y\frac{\partial v_1}{\partial y} + \frac{\partial v_1}{\partial \theta}) \\ &= -((Dv_1).\mathbf{V}) = -(D(\mathbf{V}.\nabla u).\mathbf{V}) \end{aligned} \tag{23.20}$$

This formula[1] *shows that $accel(u)$ is the acceleration in the direction of $-\nabla u$. As v_1 the component*

[1]We denote by $\frac{\mathrm{D}f}{\mathrm{D}\theta}$ the variation of f along the trajectory of the considered point ($= ((Df).\mathbf{V})$ where $\mathbf{V}$ is the velocity of the point). This is generally different from $\frac{\partial f}{\partial \theta}$ which is the partial variation of f with respect to θ.

of the velocity in the spatial gradient direction is called the "apparent" velocity, $accel(u)$ can be called the "apparent acceleration".

2. Let $\mathbf{V}$ be the "velocity vector" defined in Definition 23.18, then

$$accel(u) = \frac{(D^2u)(\mathbf{V},\mathbf{V})}{|\nabla u|} \tag{23.21}$$

Proof of proposition 23.11 The proof is just some simple calculations. □

Discretization of the apparent acceleration. We shall prove some equalities allowing a robust computation of the term $accel(u)$. As we have seen before, the "possible velocity" vectors are in $\mathcal{W}$. They also must be orthogonal to the gradient of the movie Du, and therefore lie in $\mathcal{V}$ We will first obtain a formula for $accel(u)$ that involves a minization over the vectors of $\mathcal{V}$, and secondly we will extend this minimization over the vectors of $\mathcal{W}$.

Lemma 23.12 *Whenever the spatial gradient ∇u and the spatial curvature $curv(u)$ are not equal to zero,*

$$|\nabla u|(sgn(curv(u))\ accel(u))^+ = min_{\mathbf{v}\in\mathcal{V}}|(D^2u)(\mathbf{v},\mathbf{v})| \tag{23.22}$$

Proof Let us recall that the set $\mathcal{V}$ is the set of the vectors

$$\mathbf{v}_\mu = \frac{|Du|}{\nabla u}(\mu\mathbf{e}^\perp - \mathbf{e}^\pm)$$

We have

$$((D^2u)\mathbf{v}_\mu.\mathbf{v}_\mu) = \frac{|Du|^2}{|\nabla u|^2}(\Gamma_1\mu^2 - 2\Gamma_2\mu + \Gamma_3) = P(\mu),$$

where $P(\mu)$ is a polynomial of degree 2 in μ. When $|\nabla u|$ and $curv(u)$, (and therefore Γ_1) are not equal to zero the extremum of $P(\mu)$ is reached when $\mu = \Gamma_2/\Gamma_1$, that is when $\mathbf{v}_\mu = \mathbf{V}$. Thus the extremum value of $P(\mu)$ is $|\nabla u|accel(u)$, by proposition (23.11). We obtain

$$ext_{\mathbf{v}\in\mathcal{V}}(D^2u)(\mathbf{v},\mathbf{v}) = |\nabla u|accel(u),$$

where by $ext_{\mathbf{v}\in\mathcal{V}}$ we denote the finite extremal value in $\mathcal{V}$.

Assume first that $curv(u)$ and $accel(u)$ have the same sign. This implies that the second order coefficient and the extremum of the polynomial have the same sign. Thus the expression $(D^2u)(\mathbf{v},\mathbf{v})$ has the same sign for all $\mathbf{v}\in\mathcal{V}$. This yields $|\nabla u|(sgn(curv(u))\ accel(u)) = min_{\mathbf{v}\in\mathcal{V}}|(D^2u)(\mathbf{v},\mathbf{v})|$.

If now, $curv(u)$ and $accel(u)$ have opposite signs then $|\nabla u|(sgn(curv(u))\ accel(u))^+ = 0$. And $P(\mu)$ is equal to zero for at least one vector $\mathbf{v}$ of $\mathcal{V}$. Thus, for this vector, $|(D^2u)(\mathbf{v},\mathbf{v})| = 0$, and $min_{\mathbf{v}\in\mathcal{V}}|(D^2u)(\mathbf{v},\mathbf{v})| = 0$. So (**??**) is still satisfied. □

From a numerical viewpoint, the minimization on the set of vectors $\mathcal{V}$ is not easy. Indeed, first, the direction of the gradient of the movie is quite unstable because $\Delta\theta$, the time interval between two images, can be large.

We will restrict $\mathcal{W}$ to the vectors that stand in a ball $B(0,R)$ for an arbitrary R that can be chosen large enough. In others words, we will only consider bounded possible velocities, which is not a real restriction in pratice.

Lemma 23.13 *Let ∇u and $curv(u)$ be not equal to zero, and u be C^2, then the expression*

$$min_{\mathbf{v}\in\mathcal{W}}(\frac{1}{\Delta\theta^2}(|u(\mathbf{x}-\mathbf{v}\Delta\theta)-u(\mathbf{x})|+|u(\mathbf{x}+\mathbf{v}\Delta\theta)-u(\mathbf{x})|)) \tag{23.23}$$

converges towards $|\nabla u|(sgn(curv(u))\ accel(u))^+$ when $\Delta\theta$ tends to zero.

Proof Due to the fact that $\mathbf{v}\in\mathcal{W}$ are assumed to be bounded, we have that $\mathbf{v}\Delta\theta$ tends to 0 as $\Delta\theta$ tends to 0. As consequence, we can restrict the proof to the case where u is a quadratic form without loss of generality.

So, let u be a quadratic form : $u(\mathbf{x})=\frac{1}{2}A(\mathbf{x},\mathbf{x})+\mathbf{p}.\mathbf{x}+c$, and define

$$F(\mathbf{v},h)=(|u(\mathbf{x}-\mathbf{v}h)-u(\mathbf{x})|+|u(\mathbf{x}+\mathbf{v}h)-u(\mathbf{x})|)/h^2$$

We have

$$F(\mathbf{v},h)=|-\frac{\mathbf{p}.\mathbf{v}}{h}+\frac{1}{2}A(\mathbf{v},\mathbf{v})|+|\frac{\mathbf{p}.\mathbf{v}}{h}+\frac{1}{2}A(\mathbf{v},\mathbf{v})| \tag{23.24}$$

Let $\mathbf{w}\in\mathcal{V}$ be a vector which minimizes the min in (23.22), $\mathbf{w}\in\mathcal{V}$ then $\mathbf{w}.\mathbf{p}=0$), thus (23.24) becomes

$$F(\mathbf{w},h)=|A(\mathbf{w},\mathbf{w})|$$

Therefore

$$\lim_{h\to 0}(min_{\mathbf{v}\in\mathcal{W}}F(\mathbf{v},h))\leq F(\mathbf{w},h)=|\nabla u|(sgn(curv(u))\ accel(u))^+ \tag{23.25}$$

Moreover $min_{\mathbf{v}\in\mathcal{W}}F(\mathbf{v},h)$ exists for every h and is bounded. We denote by $\mathbf{v}_h$ a vector of $\mathcal{W}$ such that $F(\mathbf{v}_h,h)=min_{\mathbf{v}\in\mathcal{W}}F(\mathbf{v},h)$. Since $F(\mathbf{v}_h,h)$ is bounded and $F(\mathbf{v}_h,h)\geq 2|(\mathbf{p}.\mathbf{v}_h)/h|$, we necessarly have

$$|(\mathbf{p}.\mathbf{v}_h)|=O(h) \tag{23.26}$$

Let decompose $\mathbf{v}_h$ into two vectors : $\mathbf{v}_h=\mathbf{v}_h^\perp+h\mathbf{v}_h^\pm$ such that $\mathbf{v}_h^\perp$ is orthogonal to $\mathbf{p}$, and (23.26) leads that $|\mathbf{v}_h^\pm|$ is bounded when h tends to zero. As before, we have

$$F(\mathbf{v}_h,h)\geq|A(\mathbf{v}_h,\mathbf{v}_h)|\geq|A((\mathbf{v}_h^\perp+\mathbf{v}_h^\pm),(\mathbf{v}_h^\perp+\mathbf{v}_h^\pm))|\geq$$

$$|A(\mathbf{v}_h^\perp,\mathbf{v}_h^\perp)+2hA(\mathbf{v}_h^\perp,\mathbf{v}_h^\pm)+h^2A(\mathbf{v}_h^\pm,a\mathbf{v}_h^\pm)|$$

Since $|\mathbf{v}_h^\pm|$ is bounded, we get $\lim_{h\to 0}F(\mathbf{v}_h,h)\geq|A(\mathbf{v}_h^\perp,\mathbf{v}_h^\perp)|$ Now, $\mathbf{v}_h^\perp$ is in $\mathcal{V}$ then $|A(\mathbf{v}_h^\perp,\mathbf{v}_h^\perp)|\geq min_{\mathbf{v}\in\mathcal{V}}|A(\mathbf{v},\mathbf{v})|$, so

$$\lim_{h\to 0}(min_{\mathbf{v}\in\mathcal{W}}F(\mathbf{v},h))=\lim_{h\to 0}F(\mathbf{v}_h,h)$$

$$\geq min_{\mathbf{v}\in\mathcal{V}}|A(\mathbf{v},\mathbf{v})|=|\nabla u|(sgn(curv(u))\ accel(u))^+ \tag{23.27}$$

(23.25) and (23.27) conclude the proof of the proposition. □

In addition to a quantization problem, if we wish to recover an "acceleration" interpretation of the term "accel" we need somehow to make appearing in the formulation of accel the velocities before and after the considered point.

Lemma 23.14 *Let u be C^2, ∇u and $curv(u)$ not zero, then*

$$min_{\mathbf{v}\in\mathcal{W}}(|u(\mathbf{x}-\mathbf{v}\Delta\theta)-u(\mathbf{x})|+|u(\mathbf{x}+\mathbf{v}\Delta\theta)-u(\mathbf{x})|)= \tag{23.28}$$

$$min_{\mathbf{v}_b,\mathbf{v}_a\in\mathcal{W}}(|u(\mathbf{x}-\mathbf{v}_b\Delta\theta)-u(\mathbf{x})|+|u(\mathbf{x}+\mathbf{v}_a\Delta\theta)-u(\mathbf{x})|+\Delta\theta|\nabla u.(\mathbf{v}_b-\mathbf{v}_a)|)+o(\Delta\theta^2)$$

Proof First, we remark by taking $\mathbf{v}_b = \mathbf{v}_a$ that the first part is larger than the second part of the expression.

$$\begin{aligned}
& (|u(\mathbf{x}-\mathbf{v}_b h)-u(\mathbf{x})|+|u(\mathbf{x}+\mathbf{v}_a h)-u(\mathbf{x})|+h|\nabla u.(\mathbf{v}_b-\mathbf{v}_a)|) \\
= \quad & |-h(Du.\mathbf{v}_b)+\frac{h^2}{2}(D^2u)(\mathbf{v}_b,\mathbf{v}_b)|+|h(Du.\mathbf{v}_a)+\frac{h^2}{2}(D^2u)(\mathbf{v}_a,\mathbf{v}_a)| \\
& +h|Du.(\mathbf{v}_b-\mathbf{v}_a)|)+o(h^2) \\
\geq \quad & \frac{h^2}{2}(|(D^2u)(\mathbf{v}_b,\mathbf{v}_b)|+|(D^2u)(\mathbf{v}_a,\mathbf{v}_a)|)+o(h^2) \\
\geq \quad & min_{\mathbf{v}\in\mathcal{W}}(|(D^2u)(\mathbf{v},\mathbf{v})|)+o(h^2) \\
= \quad & min_{\mathbf{v}\in\mathcal{W}}(|u(\mathbf{x}-\mathbf{v}h)-u(\mathbf{x})|+|u(\mathbf{x}+\mathbf{v}h)-u(\mathbf{x})|)+o(h^2)
\end{aligned}$$

by Proposition 23.13. □

Interpretation. We deduce from all of these propositions an explicit formula for the apparent acceleration

$$|\nabla u|(sgn(curv(u))\ accel(u))^+ = \tag{23.29}$$
$$min_{\mathbf{v}_b,\mathbf{v}_a\in\mathcal{W}}\frac{1}{\Delta\theta^2}(|u(\mathbf{x}-\mathbf{v}_b\Delta\theta)-u(\mathbf{x})|+|u(\mathbf{x}+\mathbf{v}_a\Delta\theta)-u(\mathbf{x})|+\Delta\theta|\nabla u.(\mathbf{v}_b-\mathbf{v}_a)|)+o(1)$$

Of course for numerical experiments, we shall not compute the minimum for all vectors in $\mathcal{W}$, but only for the vectors on the grid. We have two differents parts in the second term : The first part is the variations of the grey level value of the point $\mathbf{x}$, for candidate velocity vectors : $\mathbf{v}_b$ between $\theta-\Delta\theta$ and θ (velocity *b*efore θ), and $\mathbf{v}_a$ between θ and $\theta+\Delta\theta$ (velocity *a*fter θ). These variations must be as small as possible, because a point is not supposed to change its grey level value during its motion. The second part is nothing but the "acceleration", or the difference between $\mathbf{v}_b$ and $\mathbf{v}_a$ in the direction of the spatial gradient $|\nabla u|$.

23.5 Destruction of the non-smooth trajectories.

Since trajectories are included into the spatio-temporal gray-level surfaces (level surfaces), it is interesting to look at the evolution of such surfaces. According to the equation, the surfaces move (in scale) at each point with a speed in the direction of ∇u given by $curv(u)^{\frac{1-q}{3}}$ $(\text{sgn}(curv(u))accel(u)^q)^+$. (We do not consider the case where $q=0$ that corresponds to a pure spatial smoothing).

Therefore any level surfaces that corresponds to an uniform motion does not move in scale (it is a steady state for the equation AMG). Such surfaces are straight in one direction of the space-time.

We see also that parts of the surfaces where the curvature and the operator *accel* have opposite signs do not move as well. Then if we take example of a uniform circle under acceleration, the level surface corresponding to the circle moves only in one of its side.

More geometrically the smoothing can only occur at points where the level surface is strictly convex or strictly concave. We can give an intuitive hint of why the smoothing is stopped on saddle points. This property of the model AMG, comes directly from the contrast invariance and the causality. They imply

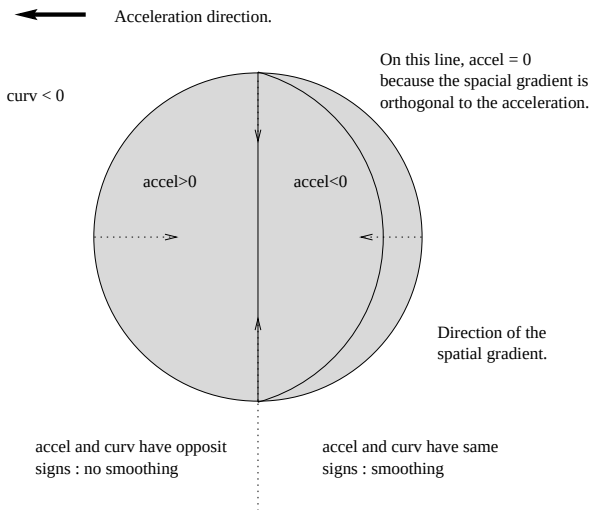

Figure 23.2: The AMG model erodes a circle in acceleration only on one side. Indeed, when the curvature and the acceleration have opposite signs, the evolution in scale is zero. (see the AMG equation).

a independent and continuous motion of level surfaces that makes that two level surfaces can not cross them-selves. Now as shown in the picture 23.5, we can bound non-convex and non-concave part of surfaces by straight surfaces that have no evolution, and then easily see why such parts does not move.

As a consequence, we can not expect from a such modelization to obtain a smoothing of the trajectories. Non-smooth trajectories are not really smoothed by the model but are simply destroyed. Let us take an example. In figure 23.5, we display a oscillatory trajectory (in gray). The limit of a smoothing of this trajectory should be a straight trajectory. Now using the same argument as in the preceding paragraph the gray surface can not cross the white surface which has no evolution. Therefore the gray surface can not become straight, because it should have to cross the white one. A such trajectory is shrunk by the AMG model and disappears at a finite scale of smoothing (see figure 23.5).

We conclude that the assumptions we made for our model are incompatible with the notion of smoothing trajectories. Indeed non-straight trajectories are not more and more smoothed, but are more and more removed. And by consequence a small perturbation in a straight trajectory might imply a destruction of this trajectory although it would have been kept without the perturbation.

23.6 Conclusion.

We have seen that there exists an unique affine, contrast and Galilean invariant scale-space for movies, the AMG. This model does a spatial smoothing with a speed depending on the spatial curvature and an apparent acceleration. The larger is the acceleration the larger is the speed of smoothing. Therefore, as shown on the experiments it has a strong denoising property since the noise does not generally generate regular trajectories.

Now we have seen that the properties asked to the scale-space are compatible with the definition of the optical flow. In the sense that the definition of the optical flow satisfies as well the contrast, the affine, and the Galilean invariance. But, the contrast invariance added to the causality (that defines the scale-space) is incompatible with the notion of smoothing trajectories. In others terms, non-smooth level-surfaces (on

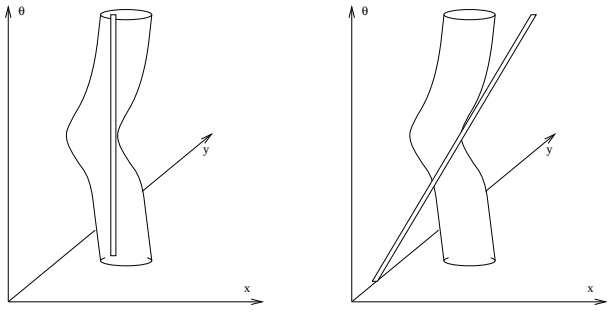

Figure 23.3: Saddle points of level surfaces remain steady by the AMG model. Indeed, our scale-space can be seen as a motion in scale of gray level-surfaces (isophotes). The level-surfaces that are straight in time correspond to a uniform translation and are not changed by the smoothing. Therefore, the two thin cylindric level-surfaces drawn left and right in the figures above do not move in scale. Now, by the inclusion principle, two level surfaces can never cross during the evolution in scale. Since, as displayed in the picture, it is possible to squeeze any surface saddle point between two such steady cylinders, it follows that saddle points do not move in scale as well. This property is readable in the scale space equation : at saddle points, the positive part of the product of the curvature and of the acceleration is zero.

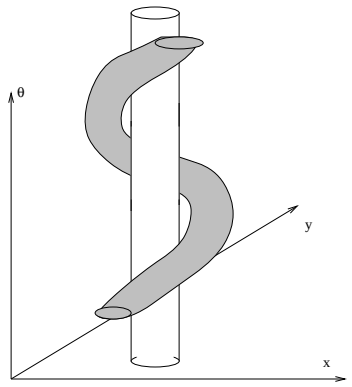

Figure 23.4: The level-surface in gray cannot become straight : it would have to cross the white level-surface which is invariant by the scale space. Now, during the smoothing process, the level-surface in gray will be eroded on its convex part, and will eventually disappear at a fixed scale : it cannot converge to any steady surface since all of them are straight in time. Thus, trajectories that are contained in the grey level surface end being removed from the movie.

which are contained the trajectories by definition of the optical flow) are more shrunk than smoothed. In fact the AMG model as to be seen as a riddle that progressively remove non-smooth trajectories.

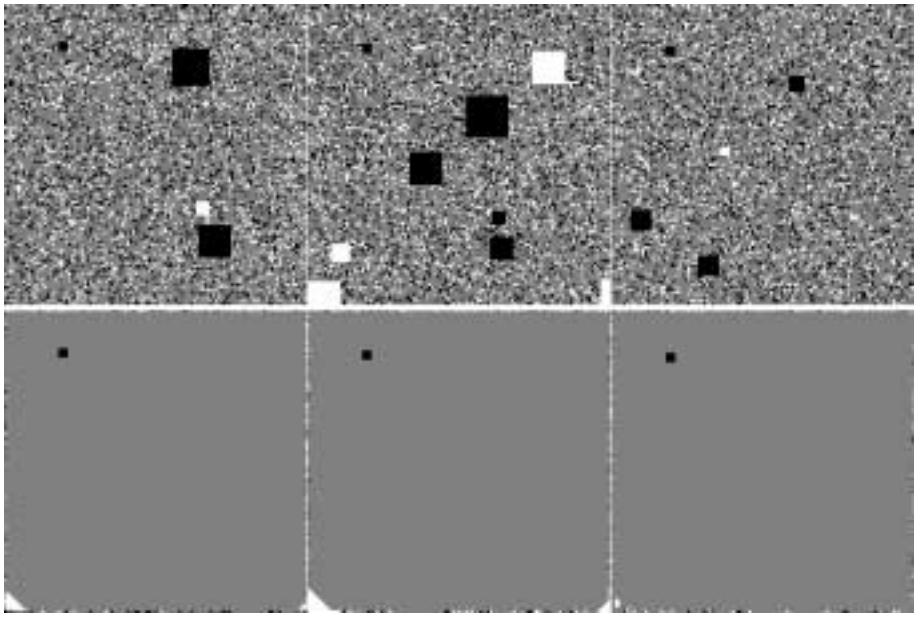

Figure 23.5: The affine, morphological, galilean (AMG) model used for image sequence restoration (extraction of coherent trajectories). Above : three successive images extracted from a synthetic sequence, made of salt and pepper noise, plus some squares placed at random locations. In addition, a little black square in uniform motion has been added in the whole sequence. Bottom : resulting images at calibrated scale 500p (scale at which a spatio-temporal sphere of 500 pixels disapears by AMG). Only the little black square remains, as it has a coherent motion.

References.

The Optical Flow: The problem of estimating dense velocities field from image sequence is a entire research topic by itself. Since it is not the main point of this book we refer to some articles dealing with that subject: [149, 207, 206, 8, 197, 414, 312, 26, 55, 402]... The aperture problem of the optical flow - that is its non uniqueness- has appeared very early and has been often adressed by e.g. some smoothness constraint on the flow it self, see e.g. [207, 313, 387, 55] or in some cases by an implicit smoothing of image sequences see e.g. [8]...

Smoothing images sequences: Explicit smoothing of image sequences, for the purpose of estimating the optical flow or for other purposes has first appeared as a direct extension of the 2D smoothing to the 3D. That is no specific rule was given to the time. In that sense most all 2D filters can be adapted to N-dimensional data, and in particular the images sequences.

In [8], it is implicitely proposed to tune the sequence filtering to few different orientations in space-time. All designed filters give different answers, answers that were used as basis of the optical flow estimates. Even if it was impossible to use a filter for all spatio-temporal directions, the idea to orient the filtering in the direction of the (unknown) motion was there.

In [15] the basic principles explained in this chapter were proposed. In particular the "Galilean Invariance". Surprisingly, these formal principles yield an anisotropic diffusion oriented, for each point, in the

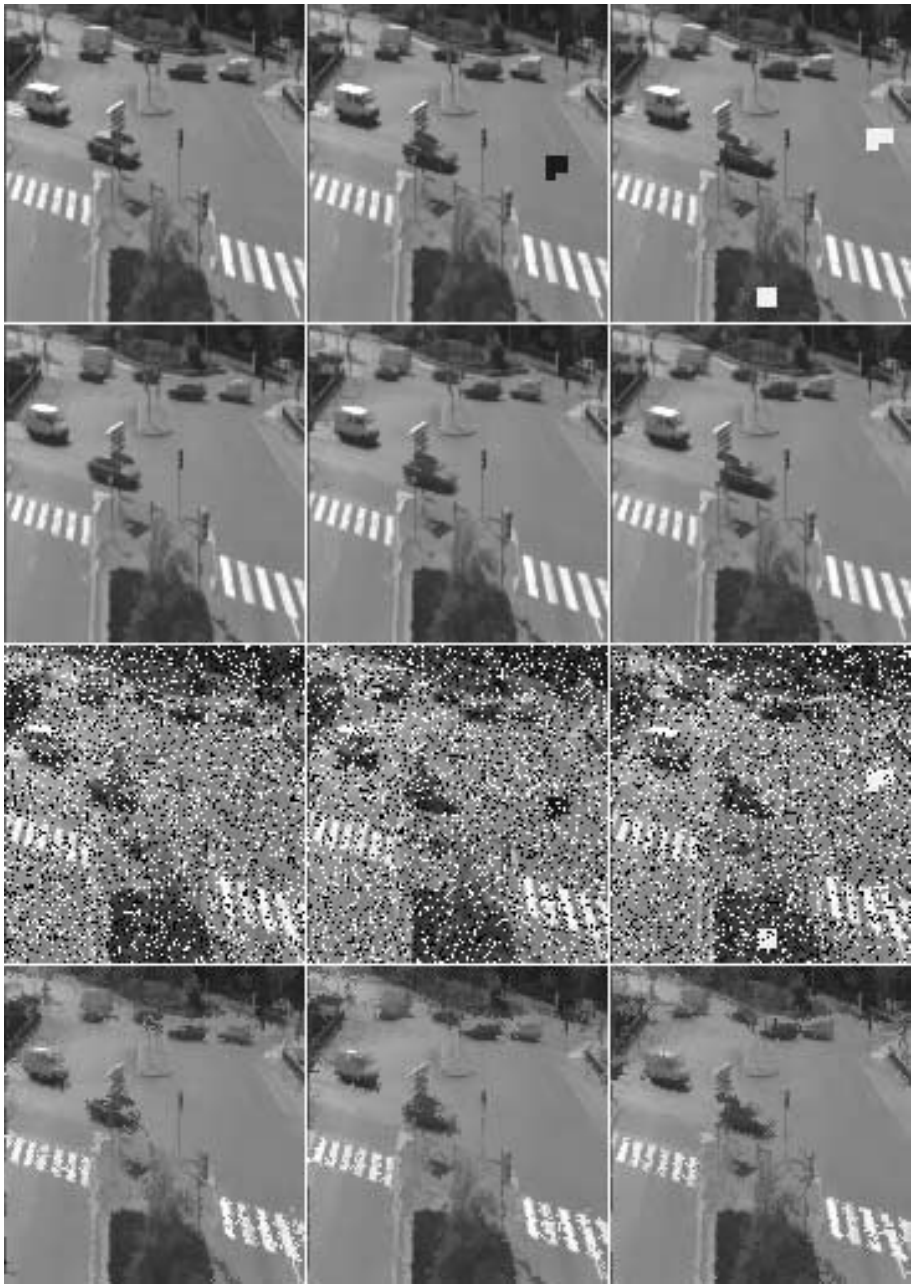

Figure 23.6: AMG model (Affine, Morphological, Galilean) used for image sequence "denoising". Above : three successive images extracted from a sequence. Second row : resulting images at calibrated scale 100 pixels (scale at which a spatiotemporal sphere of 100 pixels disapears). Third row : Some noise has been added to the original sequence (25% of the pixels are corrupted). Bottom : resulting images at scale 100.

direction of the (unknown) optical flow [15, 184]... Several other works have introduced other smoothings depending on its aim and where the time plays a specific rule. In [295] the author formalizes a smoothing compatible with the aim of estimating depth from an image sequence. In [318] and one could find adaptations of the 2D linear smoothing theory to an anisotropic diffusion in the direction of an estimated optical flow.

Chapter 24

Numerical Implementations.

24.1 Digital level sets and curves.

Depending on the expected precision of the level set or level curve representation, there exists different strategies concerning their extraction. We address here the simplest, by considering a pixel precision. For a higher precision, a model of interpolation between pixel values can be chosen. Candidates of such models stand e.g. in Shannon interpolaton, Lipschitz interpolation, bi-linear interpolation, etc...

We consider the digital image as a array which associates to each pixel a gray level value.

The level sets. The upper or lower level set of level λ can be simply obtained by thresholding the image at the level λ. The upper level set is the set of the pixels that have a gray level higher or equal to λ. (And conversely for the lower).

The Connected components of level sets To extract the connected components of the level sets, we have first to define what is meant by connected. As well known, there exist two kinds of connectivity in a square-based grid : The 4-connectivity and the 8-connectivity (see figure 24.1).

One can arbitrary choose one or the other, being aware that this choice violates the symmetry $u \to -u$.

Choosing the 8-connectivity for the upper level sets, means in fact considering the digital image as an upper semi-continuous function. This function is constructed by putting the gray level value of each pixel to all the point enclosed in it and by setting the pixel boundaries to the sup of the adjacent pixels values. Similarly, choosing the 4-connectivity for the upper level sets, yields to a lower semi-continuous representation of the image.

Once the choice of connectivity is done, extracting the connected composants of the level sets is strait-

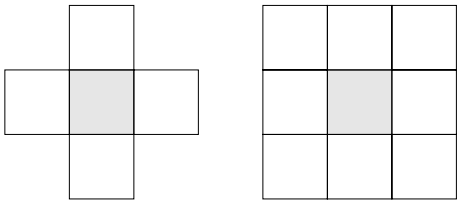

Figure 24.1: Connectivity on a grid. Left, 4-connectivity that is each pixel is connected to the ones directly up, left, right and down. Right, 8-connectivity.

ghforward.

The level lines extraction. The level lines are defined as the connected components of the boundaries of upper the level sets. Assuming that the connected components of the level sets have been extracted, one has then just to follow their boundaries.

Note that the boundaries of connected components of a level sets are of two kinds. There is in one hand the exterior envelop, in the other hand there are the boundaries that envelop the holes of the level set (if it has some).

Assume that we follow one of boundary (exterior or a hole) of the set, in a way so that the set is located on the left when facing the path direction. The path can be followed without ambiguity until a junction. In case of 4-connectivity (resp. 8), the most left (resp. right) direction of the junction should be chosen.

24.2 Median Filter

General median filter

Many different digital implementations of the median filter have been proposed. If we forget problem of time computation, the simplest implementation is as follow:

Given a mask (support of the measure) with a odd number of pixels: $2N+1$, the median value at pixel p is the "N" largest value among the pixel in the mask centered at p. Note that partial sorting, using e.g. quick-sort, can provide such a value. More clever ways to extract the median value, with a limited number of comparaisons, have been proposed depending on the chosen mask.

median filter on a disk

The problem is to approximate adequatly a disk by a mask made of pixels. One strategy is to use a weighted median filter. We define the weight of a pixel as the area of the intersection of the pixel itseld and the disk. Then the median value is the smallest value such that the sum of the weight of the pixels with smaller value passes 50% of the total weight.

In case of binary images, (or shapes), as remarked in Chapters 2 and 3, the weigthed median filter is equivalent to a linear convolution with the weighted masked followed by a threshold. Figure 24.2 illustrated this fact with a Gaussian weighted median filter.

Implementing the median filter using the mean curvature motion...

In any cases, the median filter suffers two drawbacks:

First, it is too local to catch small curvature. Indeed, it is impossible to make the distinction between a straight line and a slightly curved line if one looks locally and with a pixel precision.

Second, it generates quantized evolutions due to the quantization of space (pixels) and of the gray levels.

As we have seen, iterated median filter converges towards viscosity solution of the curvature motion. One can therefore consider to use the curvature motion equation as an implementation of an "ideal" median filter.

24.3 Extrema Killer : Maxima and Minima Killer.

The "Extrema Killer" has been formaly defined in section 7.4. Its formal definition involve an infinite set of structuring elements, which makes its formal definition not in practical use. We have shown that it

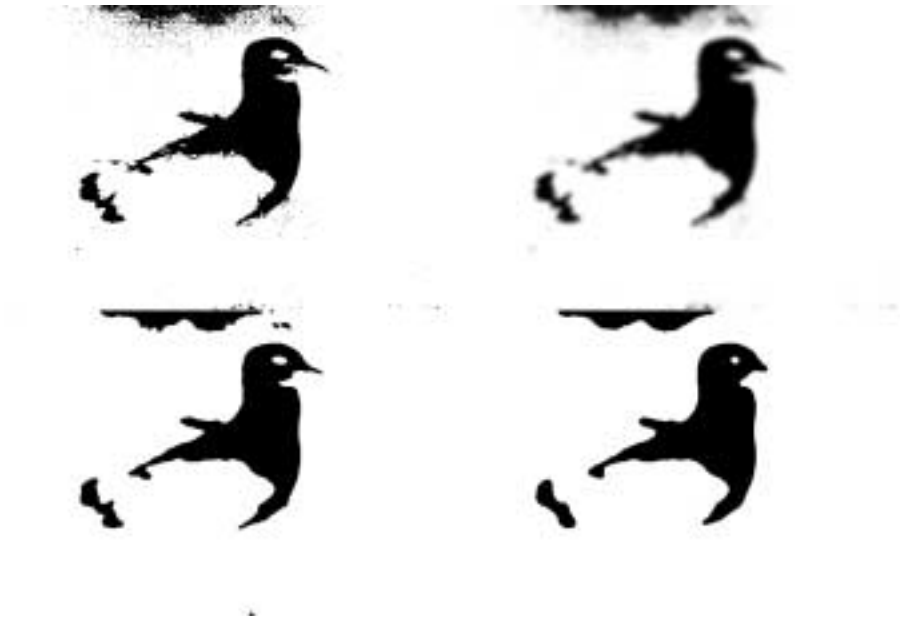

Figure 24.2: "Gauss convolution+threshold=median". A convolution with a gaussian followed by a threshold yields a weighted median filter. The alternate iterations on a binary shape of the linear smoothing and of a 1/2 threshold emulate the mean curvature motion. Left : original picture, then, result of the convolution with a small gaussian kernel, in continuation threshold at the mid-level of the blurred image, right : result after few iterations of this alternate process.

effect is just to remove the small (in area) connected components of the upper level sets for the maxima killer (and lower level sets for the minima killer).

Therefore an idea to implement the maxima killer it is then to start at local maxima and to go down until the area is achieved. Going down means to add adjacent pixels to a stack starting a the local maximum. The stack represents then a region of pixels which could be interpreted as a local "mountain" if the gray levels are interpreted as altitudes... Now, problem occurs when the stack meets an saddle points. An saddle point can be where two "mountains" meets. And the stack of the two "mountains" should be added, which would mean that the stack corresponding to each local maxima should be computed in parallel.

A strategy to avoid that is then to stop the increase of the stack when it meets an saddle point, that is in other words, when an adjacent pixel to the stack has a larger value than the min value of the pixels in the stack. The pixels in the stack are then set to this min value, correponding to the value of the saddle point. As consequence the saddle point is no more a saddle point, and these pixels will continue to be treated later, see Figure 24.3

An implementation of the maxima killer is then as follow:

1. Choose a connectivity 4 or 8 and an area A (number of pixels).

2. Start from a local maximum pixel. (The value of the pixel is larger or equal to its neighbors, and at least strictly larger than one of them). Let λ be its gray level. Initialize a stack with this pixel.

3. Among the neighbor pixels of all pixels of the stack, choose the one that has the largest value. Add it to the stack.

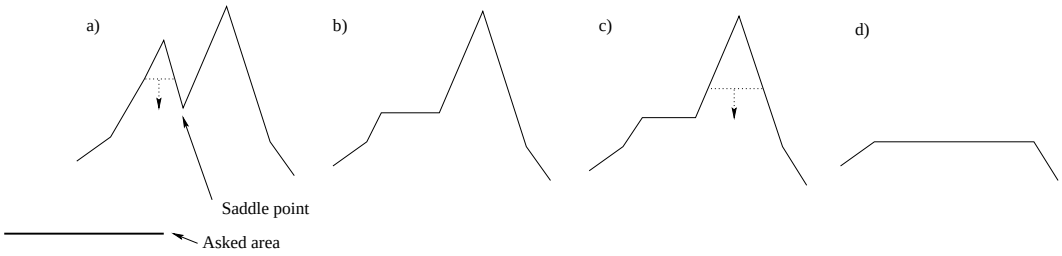

Figure 24.3: Extrema killer implementation. This figure illustrates the implementation of the maxima killer for one dimensional image. (As described in the text, the algorithm is similar in dimension two). The size asked is displayed in bottom. The algorithm starts with a local maxima, a). And, it goes down in gray level until either the size is achieved or a saddle point is touched. In both case, the pixels are set to the achieved gray level value, b). In the shown case, the area is not achieved, but it will be so, when the next maxima will be treated, c) and d). The minima killer works symmetrically.

4. If this pixel has a value less than λ, set λ to its value, else goto 6

5. If the stack is smaller than A goto 3.

6. Set all pixels of the stack to the value λ. Empty the stack.

7. If it remains a local maxima pixel not treated, find it and goto 2.

The symmetrical implementation will perform the minima killer. The extrema killer made of an maxima and a minima killer defines in fact two possible filters depending on which of the maxima or the minima killer is done first. A symmetrical version of the extrema killer can be found in [289, ?]

24.4 Finite Difference Scheme (FDS) for the Curvature Motion and AMSS.

We shall consider the classical discrete representation of an image u on a grid $u_{i,j} = u(i,j)$, with $1 \leq i \leq N$, $1 \leq j \leq N$. The image is the union of the squares centered at the points (i,j), and the brightness in each square is constant : $u_{i,j}$. Each one of the squares is called *pixel* (for "picture element").

24.4.1 Case of Mean curvature motion.

We start with the "Mean curvature motion" equation (M.C.M.) given by

$$\frac{\partial u}{\partial t} = |\nabla u| curv(u) = \frac{u_y^2 u_{xx} - 2u_x u_y u_{xy} + u_x^2 u_{yy}}{u_x^2 + u_y^2}$$

In order to discretize this equation by finite differences we shall introduce an explicit scheme which uses a fixed stencil of $3x3$ points to discretize the differential operators. For simplicity, we assume that the spatial increment Δx is the same in the x-axis and the y-axis. We approach the first derivatives u_x and u_y in a point (i,j) of the lattice by using the following linear scheme:

$$(u_x)_{i,j} = \frac{2(u_{i+1,j} - u_{i-1,j}) + u_{i+1,j+1} - u_{i-1,j+1} + u_{i+1,j-1} - u_{i-1,j-1}}{4\Delta x} + O(\Delta x^2)$$

λ4	λ2	λ3
λ1	–4 λ0	λ1
λ3	λ2	λ4

Figure 24.4: A 3x3 stencil

$$(u_y)_{i,j} = \frac{2(u_{i,j+1} - u_{i,j-1}) + u_{i+1,j+1} - u_{i+1,j-1} + u_{i-1,j+1} - u_{i-1,j-1}}{4\Delta x} + O(\Delta x^2)$$

$$|\nabla u_{i,j}| = ((u_x)_{i,j}^2 + (u_y)_{i,j}^2)^{\frac{1}{2}}$$

When $|\nabla u| \neq 0$, we can denote by ξ the direction orthogonal to the gradient of u, one easily sees that $|\nabla u| curv(u)$ is equal to $u_{\xi\xi}$. Defining θ as the angle between the horizontal axe and the gradient, we have

$$\xi = (-\sin\theta, \cos\theta) = (\frac{-u_y}{\sqrt{u_x^2 + u_y^2}}, \frac{u_x}{\sqrt{u_x^2 + u_y^2}}), \text{and}$$

$$u_{\xi\xi} = \sin^2(\theta) u_{xx} - 2\sin(\theta)\cos(\theta) u_{xy} + \cos^2(\theta) u_{yy}. \tag{24.1}$$

We want to write $u_{\xi\xi}$ as a linear combination of the values of u on the fixed stencil 3*3. Of course, the coefficients of the linear combination may depend on ξ. Because the direction of the gradient (and then ξ) is defined modulo π, by symmetry we must assume that the coefficients of points symmetrical with respect to the central point are the same (see figure 24.4).

In order to be consistant, we must find $\lambda_0, \lambda_1, \lambda_2, \lambda_3, \lambda_4$, such that

$$(u_{\xi\xi})_{i,j} = \frac{1}{\Delta x^2}(-4\lambda_0 u_{i,j} + \lambda_1(u_{i+1,j} + u_{i-1,j}) + \lambda_2(u_{i,j+1} + u_{i,j-1})$$

$$+\lambda_3(u_{i-1,j-1} + u_{i+1,j+1}) + \lambda_4(u_{i-1,j+1} + u_{i+1,j-1})) + O\Delta x^2 \tag{24.2}$$

We write

$$u_{i+1,j} = u_{i,j} + \Delta x (u_x)_{i,j} + \frac{\Delta x^2}{2}(u_{xx})_{i,j} + O\Delta x^3,$$

and the same relation for the other points of the stencil. By feeding (24.2) with these relations and by using relation (24.1), we obtain four relations between our five coefficients

$$\begin{cases} \lambda_1(\theta) = 2\lambda_0(\theta) - \sin^2\theta \\ \lambda_2(\theta) = 2\lambda_0(\theta) - \cos^2\theta \\ \lambda_3(\theta) = -\lambda_0(\theta) + 0.5(\sin\theta\cos\theta + 1) \\ \lambda_4(\theta) = -\lambda_0(\theta) + 0.5(-\sin\theta\cos\theta + 1) \end{cases} \tag{24.3}$$

There remains one degree of freedom for our coefficients given by the choice of $\lambda_0(\theta)$. We shall choose $\lambda_0(\theta)$ following the stability and geometric invariance criteria. Denoting by $u_{i,j}^n$ an approximation of $u(i\Delta x, j\Delta x, n\Delta t)$ we can write our explicit scheme as

$$u_{i,j}^{n+1} = u_{i,j}^n + \Delta t (u_{\xi\xi}^n)_{i,j} \tag{24.4}$$

Note that this scheme can be rewritten as $u_{i,j}^{n+1} = \sum_{k,l=-1}^{1} \alpha_{k,l} u_{i+k,j+l}^{n}$ where the $\alpha_{k,l}$ satisfy $\sum_{k,l=-1}^{1} \alpha_{k,l} = 1$

The following obvious lemma shows a general condition to have $\mathbf{L}^\infty$ **stability** in this kind of schemes:

Lemma 24.1 *Let a finite difference scheme given by*

$$T(u)_{i,j} = \sum_{k,l=-1}^{1} \alpha_{k,l} u_{i+k,j+l}$$

where $\alpha_{k,l}$ *satisfy* $\sum_{k,l=-1}^{1} \alpha_{k,l} = 1$. *Then the scheme satisfies* $\mathbf{L}^\infty$ **stability** *if and only if* $\alpha_{k,l} \geq 0$ *for any* k, l.

Proof. If $\alpha_{k,l} \geq 0$ for any k, l, set $min = inf_{i,j}\{u_{i,j}\}$, $max = sup_{i,j}\{u_{i,j}\}$ and take a point (i, j). Then $\mathbf{L}^\infty$ **stability** follows from the inequality:

$$min = \sum_{k,l=-1}^{1} \alpha_{k,l} min \leq \sum_{k,l=-1}^{1} \alpha_{k,l} u_{i+k,j+l} = (Tu)_{i,j} \leq \sum_{k,l=-1}^{1} \alpha_{k,l} max = max$$

On the other hand, if there exists $\alpha_{k_0,l_0} < 0$ then choosing u and (i, j) such that $u_{i+k_0,j+l_0} = min$ and $u_{i+k,j+l} = max$ for any other k, l, we obtain

$$(Tu)_{i,j} = \sum_{k \neq k_0, l \neq l_0}^{1} \alpha_{k,l} max + \alpha_{k_0,l_0} min = max + \alpha_{k_0,l_0}(min - max) > max$$

And therefore $\mathbf{L}^\infty$ **stability** is violated.

□

Following this lemma, in order to have $\mathbf{L}^\infty$ **stability** in the scheme (24.4) we must seek for λ_0 such that $\lambda_1, \lambda_2, \lambda_3, \lambda_4 \geq 0$ and $(1 - \frac{4\lambda_0}{\Delta x^2}) \geq 0$. Unfortunately, because of the relations between our coefficients, it is impossible to obtain these relations, except for particular values of $\theta = (0, \frac{\pi}{4}, \frac{\pi}{2}, ...)$. Indeed, We remark that for θ in $[0, \frac{\pi}{4}]$,

$$\lambda_1 \geq \lambda_2 \qquad \text{and} \qquad \lambda_3 \geq \lambda_4$$

But

$$\lambda_2(\theta) \geq 0 \Rightarrow \lambda_0(\theta) \geq \frac{\cos^2(\theta)}{2}$$

$$\lambda_4(\theta) \geq 0 \Rightarrow \lambda_0(\theta) \leq \frac{1 - \sin(\theta)\cos(\theta)}{2}$$

So, we cannot find $\lambda_0(\theta)$ satisfying both inequalities, since

$$\frac{\cos^2(\theta)}{2} \geq \frac{1 - \sin(\theta)\cos(\theta)}{2}$$

Then, if we choose $\lambda_0(\theta) \geq \frac{\cos^2(\theta)}{2}$ we have $\lambda_4(\theta)$ very negative. If we take $\lambda_0(\theta) \leq \frac{1-\sin(\theta)\cos(\theta)}{2}$ we obtain $\lambda_2(\theta)$ very negative. We prefer to choose λ_0 between both functions, and then to have λ_2 and λ_4 negative, but slightly. (see figure 24.5)

On the other hand, we impose to λ_0 the following geometrical requirements

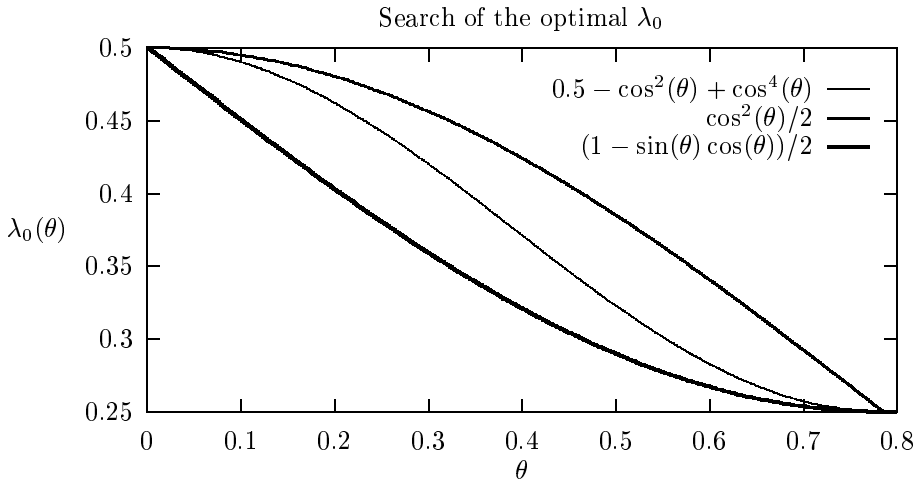

Figure 24.5: The middle curve represents the choice of the function λ_0 of formula 24.6. The upper function represents the smallest possibility for achieving for all angles $\lambda_2 \geq 0$, the lower one represent the largest possibility for achieving $\lambda_4 \geq 0$. Therefore, it is not possible to satisfy both conditions. We then have chosen the simplest trigonometric function which is in between these two constraints.

(i). Invariance by rotation of angle $\frac{\pi}{2}$

$$\lambda_0(\theta + \frac{\pi}{2}) = \lambda_0(\theta)$$

(ii). Pure diffusion in the case $\theta = 0, \frac{\pi}{2}, ...$

$$\lambda_0(0) = 0.5$$

This condition implies that $\lambda_2(0) = \lambda_3(0) = \lambda_4(0) = 0$

(iii). Pure diffusion in the case $\theta = \frac{\pi}{4}, \frac{3\pi}{4}, ...$

$$\lambda_0(\frac{\pi}{4}) = 0.25$$

This condition implies that $\lambda_1(\frac{\pi}{4}) = \lambda_2(\frac{\pi}{4}) = \lambda_4(\frac{\pi}{4}) = 0$

(iv). Symmetry with respect to the axes i+j and i-j,

$$\lambda_0(\frac{\pi}{2} - \theta) = \lambda_0(\theta)$$

We remark that by the above conditions it is enough to define the function $\lambda_0(\theta)$ in the interval $[0, \frac{\pi}{4}]$ because it can be extended by periodicity elsewhere.

We have tested two choices for the function $\lambda_0(\theta)$ using as basis the trigonometric polynomials. The first one corresponds to an average of the boundary functions:

$$\lambda_0(\theta) = \frac{\cos^2(\theta) + 1 - \sin(\theta)cos(\theta)}{4} \tag{24.5}$$

As we shall see this choice is well-adapted to the "affine curvature motion" equation. However, if we extend this function by periodicity, the extended function is not smooth at $\frac{\pi}{4}$. If we seek for a smooth function for $\lambda_0(\theta)$, we must impose $\lambda_0'(0) = \lambda_0'(\frac{\pi}{4}) = 0$. The simplest polynomial, of degree as small as possible, satisfying the above conditions, and between both boundary functions is

$$\lambda_0(\theta)) = 0.5 - \cos^2(\theta)\sin^2(\theta) \tag{24.6}$$

We deduce the other λ's values using (24.3). For instance with the above choice of $\lambda_0(\theta)$ we have

$$\begin{cases} \lambda_1(\theta) = \cos^2(\theta)(\cos^2(\theta) - \sin^2(\theta)) \\ \lambda_2(\theta) = \sin^2(\theta)(\sin^2(\theta) - \cos^2(\theta)) \\ \lambda_3(\theta) = \cos^2(\theta)\sin^2(\theta) + 0.5\sin(\theta)\cos(\theta) \\ \lambda_4(\theta) = \cos^2(\theta)\sin^2(\theta) - 0.5\sin(\theta)\cos(\theta) \end{cases}$$

When $|\nabla u| = 0$, the direction of the gradient is not known. Therefore the diffusion term $u_{\xi\xi}$ is not defined. We choosed to replace this term by half the laplacian. (The laplacian corresponds to the sum of the two second derivative in orthogonal directions, whereas the diffusion term $u_{\xi\xi}$ is the second derivaty in one). However others possibilities will be considered in section 24.4.5. So the FDS scheme for the mean curvature motion is, (iterations start with u^0 as initial function)

Where $|\nabla u| \leq T_g$

$$u^{n+1} = u^n + \frac{\Delta t}{\Delta x^2}(-4\lambda_0 u_{i,j} + \lambda_1(u_{i+1,j} + u_{i-1,j}) + \lambda_2(u_{i,j+1} + u_{i,j-1})$$
$$+\eta_3(u_{i-1,j-1} + u_{i+1,j+1}) + \eta_4(u_{i-1,j+1} + u_{i+1,j-1}))$$

Otherwise,

$$u^{n+1} = u^n + \frac{1}{2}\frac{\Delta t}{\Delta x^2}(-4\lambda_0 u_{i,j} + u_{i+1,j} + u_{i-1,j} + u_{i,j+1} + u_{i,j-1})$$

The iteration step scale $(\frac{\Delta t}{\Delta x^2})$ has to be chosen as large as possible in order to reduce the number of iterations. However, there is a natural upper bound of $\frac{1}{2}$. Indeed, denoting s this step, and considering the following image $u^0_{i,j} = 0$ for all i, j, except for $i = j = 0$ where we set $u^0_{0,0} = 1$. Then second formula yields $u^1_{0,0} = 1 - 2 * s$. The point (0,0) should not become smaller than its neighbors, and so $u^1_{0,0} \geq 0$ which yields $s \geq 1/2$.

Experimentaly we have noticed that if we impose

$$\frac{\Delta t}{\Delta x^2} \leq \frac{1}{2}, \tag{24.7}$$

then the algorithm has a good behaviour and remains stable in the sense that there exists experimentally a (small with respect to 255) $\epsilon > 0$ such that for any $n \in I\!N$ and (i, j),

$$-\epsilon + inf_{i,j}\{u^0_{i,j}\} \leq u^n_{i,j} \leq sup_{i,j}\{u^0_{i,j}\} + \epsilon$$

The threshold on the spatial gradient norm : T_g has been fixed experimentally to 6 for 0 to 255 images.

Figure 24.6: Curvature motion finite difference scheme and scale calibration. Image filtered by curvature motion at scales 1, 2, 3, 4, 5. In order to give a sound numerical meaning to the scale, a calibration of the numerical scales (number of iterations) is made in such a way that a disk with radius t shrinks to a point at scale t.

Figure 24.7: Curvature motion finite difference scheme applied on each level set separately, at scales 1, 2, 3, 4, 5. The processed image is then reconstructed by the threshold superposition principle. In contrast with the same scheme directly applied on the image, this scheme yields a fully contrast invariant smoothing. However, a comparison with Figure 24.6 shows that the resulting images are very close to each other. This shows that the contrast invariance is almost achieved when applying the finite difference scheme directly on the image. The experiment makes sense if the original image is of good quality, that is relatively smooth and with no strong oscillations. In that case, it can be considered as a distance function to each one of its own level sets. As we shall see in Figure 24.10, if the initial image is noisy, the difference between both methods can be huge.

24.4.2 Case of the AMSS model.

We will use the ideas developed in the above section. We rewrite the AMSS equation as:

$$\frac{\partial u}{\partial t} = (|\nabla u|^3 curv(u))^{\frac{1}{3}} = (u_y^2 u_{xx} - 2u_x u_y u_{xy} + u_x^2 u_{yy})^{\frac{1}{3}} \tag{24.8}$$

We remark that $|\nabla u|^3 curv(u) = |\nabla u|^2 u_{\xi\xi}$ where ξ corresponds to the direction orthogonal to the gradient. Therefore, in order to discretize this operator, it is enough to multiply the discretization of $u_{\xi\xi}$ presented in the above section by $|\nabla u|^2$. We choose $\lambda_0(\theta)$ given by (24.5) because it corresponds to a trigonometric polynomial of degree two and then multiplying it by $|\nabla u|^2$ the coefficients $\eta_i = |\nabla u|^2 \lambda_i, i = 0, 1, 2, 3, 4$, are polynomials of degree two with respect to u_x and u_y. Indeed, we obtain for $\theta \in [0, \frac{\pi}{4}]$

Figure 24.8: Iterated median filter with approximately calibrated scales 1, 1.5, 2, 2.5, 3.

$$(|\nabla u|^2 u_{\xi\xi})_{i,j} = \frac{1}{\Delta x^2}(-4\eta_0 u_{i,j} + \eta_1(u_{i+1,j} + u_{i-1,j}) + \eta_2(u_{i,j+1} + u_{i,j-1})$$
$$+\eta_3(u_{i-1,j-1} + u_{i+1,j+1}) + \eta_4(u_{i-1,j+1} + u_{i+1,j-1})) + O\Delta x^2$$

where $\eta_0, \eta_1, \eta_2, \eta_3, \eta_4$ are given by

$$\begin{cases} \eta_0 = 0.25(2u_x^2 + u_y^2 - u_x u_y) \\ \eta_1 = 0.5(2u_x^2 - u_y^2 - u_x u_y) \\ \eta_2 = 0.5(u_y^2 - u_x u_y) \\ \eta_3 = 0.25(u_y^2 + 3u_x u_y) \\ \eta_4 = 0.25(u_y^2 - u_x u_y) \end{cases}$$

Finally, the finite difference scheme for the A.M.S.S. equation is

$$u_{i,j}^{n+1} = u_{i,j}^n + \Delta t(|\nabla u^n|^2 u_{\xi\xi}^n)_{i,j}^{\frac{1}{3}} \tag{24.9}$$

We have tested this algorithm and we have noticed that in this case the condition for the experimental stability (in the sense presented in the above subsection) is

$$\frac{\Delta t}{\Delta x^2} \leq \frac{1}{10}$$

Remark. The finite difference schemes presented above are consistent and we conjecture the convergence. Contrast invariance are obtained asymptotically by taking a little time step Δt. The experimental results presented in figures ?? and ?? have been obtained by using these schemes with $\Delta x = 1$ and $\Delta t = 0.1$ in the case of mean curvature motion and $\Delta t = 0.01$ in the case of affine curvature motion. One has to take Δt that small because unless experimental stability is achieved with $\Delta t \leq 0.1$, the experimental affine invariance experimentally needs $\Delta t < 0.05$ (see 24.11).

24.4.3 Numerical normalization of scale.

(or Relation between scale and the number of iterations).

The case of the curvature motion. Setting the inter-distance between pixel Δx to 1, the scale achieved with N iterations is simply $N/\Delta t$. Now, the scaling is arbitrary.

A good way to normalize the scale is to define the scale by the radius of a disapearing circle. The boundary of such a circle move at a speed equal to the curvature that is the inverse of the radius. We have for a disk of radius $R(t)$

$$\frac{dR(t)}{dt} = -\frac{1}{R(t)}$$

which yields

$$\frac{1}{2}(R^2(0) - R^2(t)) = t$$

the disk disappears in scale, when $R(t) = 0$ that is at scale $t = R^2(0)/2$. This last relation gives the scale normalization: In order to achieve the normalized scale T (at which any disks (or any shapes included in) of radius less or equal to T has disapeared), we have the equation scale $t = T^2/2$, and the number of iterations needed equal to

$$N = T^2/2\Delta t$$

The case of AMSS We can perform similar calculations. The radius of an evolving disk satisfies

$$\frac{dR(t)}{dt} = -\frac{1}{R(t)^{\frac{1}{3}}}$$

which yields

$$\frac{3}{4}(R^{\frac{4}{3}}(0) - R^{\frac{4}{3}}(t)) = t$$

The disappearing time is therefore $t = \frac{3}{4}R^{\frac{4}{3}}$. As for the curvature motion, we define the normalized scale T at which any disks of radius less or equal to T has disappeared. In order to achieve scale T, the number of iteration needed is

$$N = \frac{3}{4\Delta t}T^{\frac{4}{3}}$$

24.4.4 The Evans Spruck extension and contrast invariance.

Both schemes (M.C.M and A.M.S.S) presented above are only asymptotically contrast invariant. But, numerically they are not. Indeed, we have seen that a contrast operator can not create new gray level. Now, starting with a binary image u^0 and applying a scheme defined by such formula

$$u^{n+1} = u^n + \Delta t(...)$$

we can not be sure that u^{n+1} is also a binary image.

A natural idea to overcome this problem is the following. Starting with a binary image: apply the scheme until the expected scale is achieved, then binarize the obtained image (just thresholding). This of course works only for binary images, however the Evans Spruck extension (see section 8.3) gives us the key to extend this to general images.

The contrast invariance can be fully obtained, by applying the process of each level set separatly. The procedure is then the following :

We start with a image u_0 and we construct its version at scale t : $u(t, \mathbf{x})$.

Initialization : set $u(t, \mathbf{x}) = 0, \forall \mathbf{x}, t$

For each λ, in increasing order

- Set $v(\mathbf{x})$ the caracteristic function of the level set λ of u_0. (that is equal to 1 inside the level set, 0 outside.)
- Apply on v the MCM or AMSS FDS-scheme until scale t. This yields the images $w(t, .)$.
- Set $u(t, \mathbf{x}) = \lambda$ at each point $(t, \mathbf{x})$ where $w(t, \mathbf{x}) \geq 0.5$

with respect to the simple FDS scheme, this clearly multiplies the number of needed computation, since the AMSS or MCM have to be solved for each level ! However, The resulting images are more closed to the theoritical solutions of these equations.

Note that explained in chapter 7, the extension is not unique. The preceding scheme corresponds to the u.s.c. extension. One could also use the l.s.c extension for which the last step has to be changed into

- Set $u(t, \mathbf{x}) = \lambda$ at each point $(t, \mathbf{x})$ where $w(t, \mathbf{x}) > 0.5$.

Of course, these two extensions differ. This can be easily seen in case of image that differs when considered as u.s.c. and l.s.c. An example, is the chessboard image (see figure24.9). The image data does not say anything about the value of the function at the borders and corners. If the image is considered as u.s.c. then the borders values is white, conversely it would be black when considered as l.s.c. In pratice, the FDS does not care about the borders and corners values, since it is to rough.

However, making the Evans Spruck extension of the FDS scheme in fact implies an implicit choice for these values.

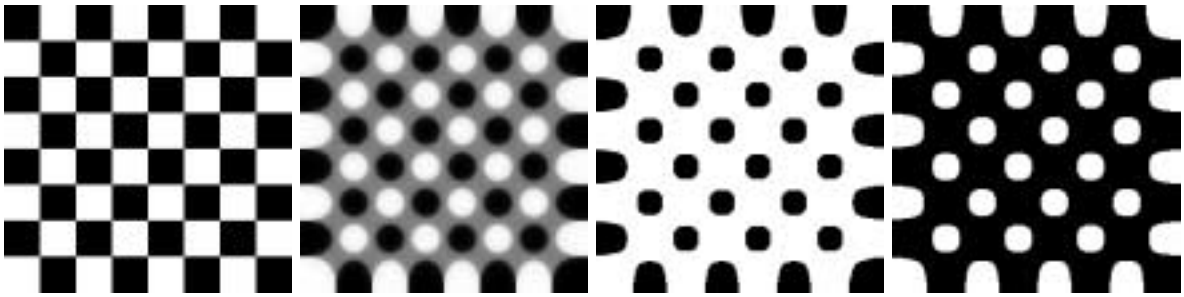

Figure 24.9: The chessboard dilemna. Left: chessboard image. Next: result with the finite difference scheme (FDS, Chapter 24) of the curvature motion, applied up to a fixed scale. The creation of a new gray level proves that the scheme is not fully contrast invariant. Indeed, by Proposition 7.4, a contrast invariant operator does not create new levels. The new observed gray level corresponds to an average of the existing ones, black and white. The next two images are obtained by the Evans-Spruck extension of the curvature motion, first under the assumption that the image is u.s.c. and second under the l.s.c. assumption. Thus, the schemes are in both cases fully contrast invariant and are extensions of the curvature motion as specified in Theorem 8.16.

24.4.5 Problems at extrema.

For the M.C.M or the A.M.S.S. occurs the problem to define numerically the equation when $|\nabla u| = 0$. For the first the right part of the equation is simply not defined, for the second by continuity, one could set

$|\nabla u^n|^2 u^n_{\xi\xi})^{\frac{1}{3}}_{i,j} = 0$. Now, numerically, this would implies that isolated black or white pixels will not change !

As a consequence, the MCM and AMSS finite difference schemes have a slightly bad behaviour at extrema. In the figure 24.10, we added to the image a strong salt and pepper noise. On a such image more than a quarter of the pixels are local extrema. The extra-diffusion produced on these points by the laplacian (as suggested in the preceding sections) gives strange structures. We show some ways to avoid this spurious behavior :

- One can first zoom by 2 the image by duplicating pixels. This however multiplies by 16 the number of computations.
- One can first remove extrema that should anyway theoritically disappear.
- One can use the Evans-Spruck extensions. This multiplies then the number of computations by the initial number of gray-level. (e.g. 256).

24.4.6 Conclusion on finite difference schemes.

We have seen that standard finite difference schemes can not handle properly the invariance properties satisfied by the equations.

1. There is no finite difference scheme that insures the monotonicity. This leads to slightly oscilatory solutions.
2. It is not fully contrast invariant. We see appearance of a slight blur around edges. And, spurious diffusion around the extrema. Extrema problem can be handle by treating them in a different way : that is flattening them immediately by applying the extrema killer first.

   In order to be fully contrast invariant, the only way we know is then by applying the equation to the characteristic functions of all the level sets. This multiplies the total computation by the number of different gray level in the original image (often 256 !).

   Now, as shown in figure 24.10, standard scheme with combination with the extrema killer might give results "good enough" for some applications.
3. The worse drawback is in fact the affine invariance, (or the Euclidean invariance for the curvature motion). Since the scheme works on a grid, motion of the level curves is quantized to square steps. As consequence : we can not garanty the affine invariance and a motion proportional to the curvature power one third.

   The only way to cope this problem is to go out of the grid, which is the aim of the section 24.6.

24.5 Curve evolution.

Given a list of points $\mathbf{x}(n)$ approximating in a polygonal way a curve, we would like to define the curvature evolution of it. As we have seen, the curvature evolution is equivalent to an intrinsic heat equation on the curve itself.

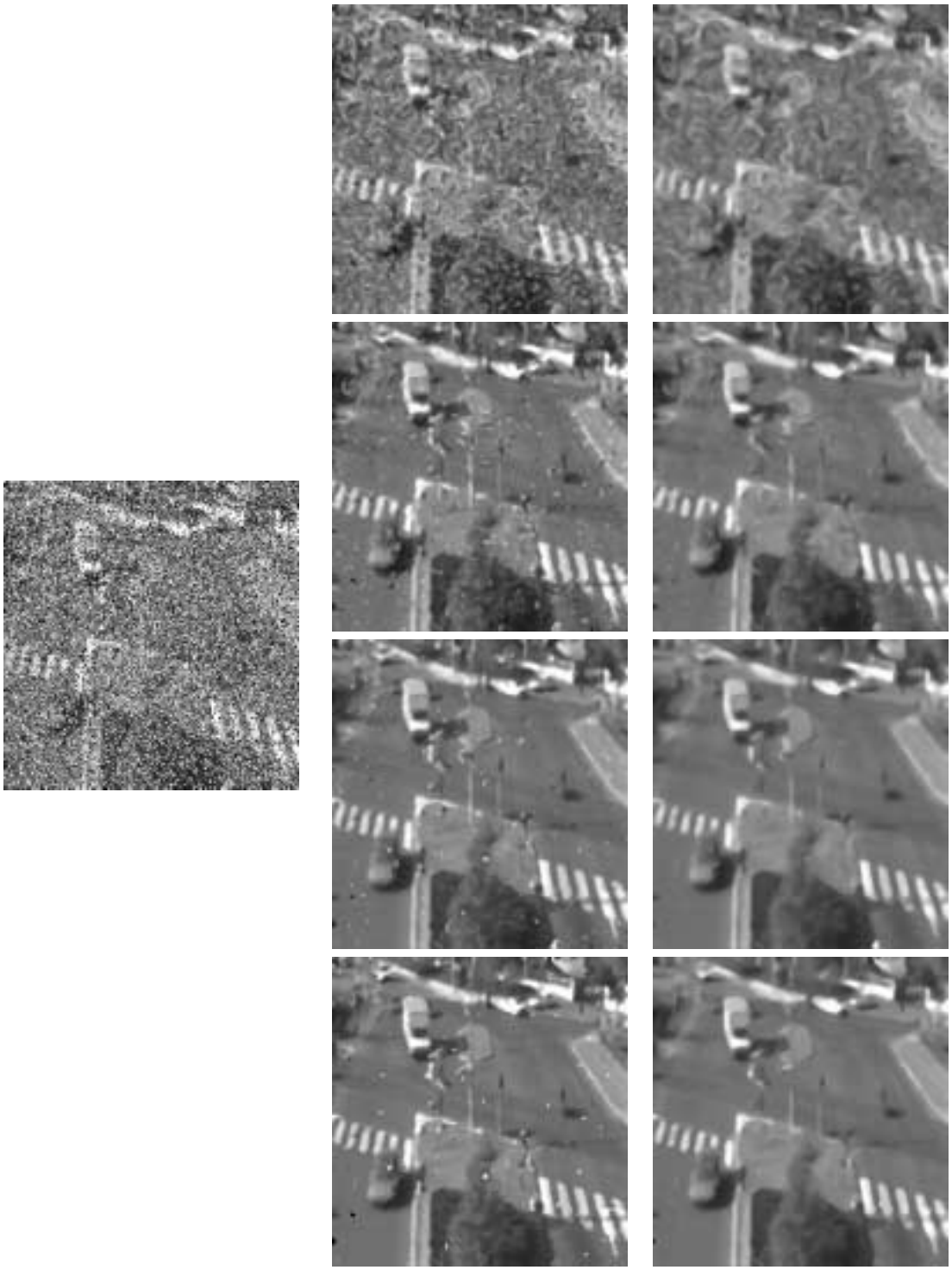

Figure 24.10: Various implementations of curvature motion on a noisy image. Middle : scale 2, Right : scale 3. From top to bottom : finite difference scheme (FDS), then FDS applied on the image previously zoomed by a factor 2, FDS applied on the image after its extrema have been "killed" (the reference area is given by the area of the disk vanishing at the desired scale), FDS applied separately on each level set, with application of the threshold superposition principle. Time computations on a Pentium 200, for achieving scale 3, and per millions of pixels, are respectively from above to below : 23.75 s, 1 mn 32s, 24.1 s, 10 mins. The third scheme offers a good compromise between time computation and results.

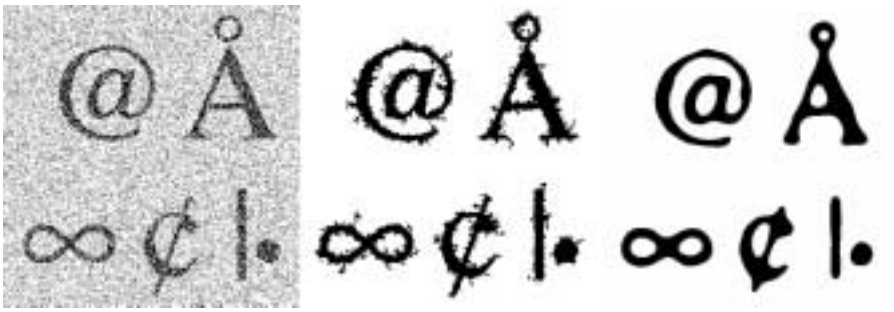

Figure 24.11: Finite difference scheme for AMSS. Left : Original noisy image of letters, middle : same image without the extrema with area less than 80 pixels (extrema killer). Right : result of AMSS at calibrated scale 4.

As seen, in chapter 2, the convolution by a Gaussian can be well approximated by the iteration of a symmetric and positive kernel. Denoting by s the arc length. We choose the simple following kernel: $k(0) = 0.5$, $k(-D) = 0.25$, $k(D) = 0.25$ and $k(s) = 0$ everywhere else. D is a distance parameter controlling the precision of the scheme. The smaller it is the better, but higher number of iterations is then needed.

We set:

$$\mathbf{x}(t+1,.) = k \times \mathbf{x}(t,.)$$

That is, the point number n is changed by

$$\mathbf{x}(t+1,n) = \mathbf{x}(t,n) + 0.25 * D * (\frac{(\mathbf{x}(t,n+1) - \mathbf{x}(t,n))}{|\mathbf{x}(t,n+1) - \mathbf{x}(t,n)|} + \frac{(\mathbf{x}(t,n-1) - \mathbf{x}(t,n))}{|\mathbf{x}(t,n-1) - \mathbf{x}(t,n)|})$$

Numerical stability problems might happen when two consecutive points have a small distance between them (typically smaller than D). One way is to keep one point every two that are too close to each other.

This scheme has been used to generate Figure 3.6.2.

To be noted that this simple scheme is very rough. For better precision, one can still implement the classical convolution by the gaussian kernel and iterate it with re-parametrization of the curve. This scheme can be extended to the affine shortening by replacing the arc length by the affine arc length. Obtaining that way an monotonous and affine invariant scheme seems to be quite difficult.

24.6 Affine Plane Curve Evolution Scheme.

The idea here is to go with the curve evolution instead of going with the image evolution. In a sense, we go in the converse way than Osher and Sethian idea. Indeed, they proposed in order to simulate the evolution of a curve by the curvature motion to make evolving its distance function by the image version of the equation. Here we do the other way around : we extract all level lines of the image. We then compute their evolutions by the affine shortening. And, at last reconstruct an image out of them. This image is then the AMSS evolution of the original.

To be noted that since the curves are extracted from an digital image, they are initialy embeded into the image grid. We can therefore represent each curve by a polygone. Secondly, the affine shortening will be executed by alternating an affine erosion and an affine dilation.

A fast algorithm

In general, the affine erosion of X is not simple to compute, because it can be strongly non local. However, if X is convex, then it has been shown in [298] that it can be exactly computed in linear time. In practice, c will be a polygon and the exact affine erosion of X —whose boundary is made of straight segments and pieces of hyperbolae— is not really needed ; numerically, a good approximation by a new polygon is enough. Now the point is that we can approximate the combination of an affine erosion plus an affine dilation of X by computing the affine erosion of each *convex component* of c, provided that the erosion/dilation area is small enough.

The algorithm consists in the iteration of a four-steps process :

1. **Break the curve into convex components.** This operation permits to apply the affine erosion to convex pieces of curves, which is much faster (the complexity is linear) and can be done simply in a discrete way. The main point is to take into account the finite precision of the computer in order to avoid spurious (small and almost straight) convex components only due to numerical artifacts.
2. **Sample each component.** At this stage, points are removed or added in order to guarantee an optimal representation of the curve that is preserved by step 3.
3. **Apply discrete afffine erosion to each component.**
4. **Concatenate the pieces of curves obtained at step 3.** This way, we obtain a new closed curve on which the whole process can be applied again.

The curve has to be broken at points where the sign of the determinant

$$d_i = [P_{i-1}P_i, P_iP_{i+1}]$$

changes. Numerically, we use the formula

$$d_i = (x_i - x_{i-1})(y_{i+1} - y_i) - (y_i - y_{i-1})(x_{i+1} - x_i) \tag{24.10}$$

Since we are interested in the sign of d_i, we must be careful because the finite numerical precision of the computer can make this sign wrong. Let us introduce the relative precision of the computer

$$\varepsilon_0 = \max\{x > 0,\ (1.0 \oplus x) \ominus 1.0 = 0.0\}. \tag{24.11}$$

In this definition, $\oplus$ (resp. $\ominus$) represent the computer addition (resp. substraction), which is not associative. When computing d_i using (24.10), the computer gives a result $\tilde{d}_i$ such that $|d_i - \tilde{d}_i| \leq e_i$, with

$$\begin{aligned} e_i = \varepsilon_0 \Big(& |x_i - x_{i-1}|(|y_{i+1}| + |y_i|) + (|x_i| + |x_{i-1}|)|y_{i+1} - y_i| \\ + & |y_i - y_{i-1}|(|x_{i+1}| + |x_i|) + (|y_i| + |y_{i-1}|)|x_{i+1} - x_i| \Big). \end{aligned}$$

In practice, we take ε_0 a little bit larger than its theoretical value to overcome other possible errors (in particular, errors in the computation of e_i). For four-bytes C *float* numbers, we use $\varepsilon_0 = 10^{-7}$, whereas

the theoretical value (that can be checked experimentally using (24.11)) is $\varepsilon_0 = 2^{-24} \simeq 5.96\ 10^{-8}$. For eight-bytes C *double* numbers, the correct value would be $\varepsilon_0 = 2^{-53} \simeq 1.11\ 10^{-16}$

The algorithm that breaks the polygonal curve into convex components consists in the iteration of the following decision rule :

1. If $|\tilde{d}_i| \leq e_i$, then remove P_i (which means that to new polygon to be considered from this point is $P_0 P_1 ... P_{i-1} P_{i+1} ... P_{n-1}$)
2. If $|\tilde{d}_{i+1}| \leq e_{i+1}$, then remove P_{i+1}
3. If $\tilde{d}_i$ and $\tilde{d}_{i+1}$ have opposite signs, then the middle of P_i, P_{i+1} is an inflexion point where the curve must be broken
4. If $\tilde{d}_i$ and $\tilde{d}_{i+1}$ have the same sign, then increment i

This operation is performed until the whole curve has been visited. The result is a chained (looping) list of convex pieces of curves.

• Sampling

At this stage, we add or remove points from each polygonal curve in order to ensure that the Euclidean distance between two successive points lies between ε and 2ε (ε being the absolute space precision parameter of the algorithm).

• Discrete affine erosion

This is the main step of the algorithm : compute quickly an approximation of the affine erosion of scale σ of the whole curve.

The first step consists in the calculus of the "area" A_j of each convex component $\mathcal{C}^j = P_0^j P_1^j ... P_{n-1}^j$, given by

$$A_j = \frac{1}{2} \sum_{i=1}^{n-2} \left[P_0^j P_i^j, P_0^j P_{i+1}^j \right].$$

Then, the effective area used to compute the affine erosion is

$$\sigma_e = \max \left\{ \frac{\sigma}{8}, \min_j A_j \right\}.$$

We restrict the erosion area to σ_e (which is less than σ in general) because the simplified algorithm for affine erosion (based on the breaking of the initial curve into convex components) may give a bad estimation of the continuous affine erosion+dilation when the area of one component is less than the erosion parameter. The term $\sigma/8$ is rather arbitrary and guarantees an upper bound to the number of iterations required to achieve the final scale.

Once σ_e is computed, the discrete erosion of each component is defined as the succession of each middle point of each segment $[AB]$ such that

1. A and B lie on the polygonal curve

2. A or B is a vertex of the polygonal curve

3. the area enclosed by $[AB]$ and the polygonal curve is equal to σ_e

These points are easily computed by keeping in memory and updating the points A and B of the curve plus the associated chord area.

Notice that if the convex component is not closed (which is the case if the initial curve is not convex), its endpoints are kept.

• Iteration of the process

To iterate the process, we use the fact that if E_σ denotes the affine erosion plus dilation operator of area σ, and $h = (h_i)$ is a subdivision of the interval $[0, H]$ with $H = T/\omega$ and $\omega = \frac{1}{2}\left(\frac{3}{2}\right)^{2/3}$, then as we are going to show further,

$$E_{(h_1-h_0)^{3/2}} \circ E_{(h_2-h_1)^{3/2}} \circ \ldots \circ E_{(h_n-h_{n-1})^{3/2}} \quad \Big(c_0\Big) \quad \longrightarrow \; c_T$$

as $|h| = \max_i h_{i+1} - h_i \to 0$, where c_T is the affine shortening of c_0 described above by (13.1).

• Comments

The algorithm takes a curve (closed or not) as input, and produces an output curve representing the affine shortening of the input curve (it can be empty if the curve has disappeared) . The parameters are

- T, the scale to which the input curve must be smoothed
- ε_r, the relative spacial precision at which the curve must be numerically represented (between 10^{-5} and 10^{-2} when using four bytes C *float* numbers).
- n, the minimum number of iterations required to compute the affine shortening (it seems that $n \simeq 5$ is a good choice). From n, the erosion area σ used in step 3 is computed with the formula

  $$\sigma^{2/3} = \frac{\alpha \cdot T^{4/3}}{n}.$$

  Notice that thanks to the $\sigma/8$ lower bound for σ_e, the effective number of iterations cannot exceed $4n$.
- R, the radius of a disk containing the input curve, used to obtain homogeneous results when processing simultaneously several curves. The absolute precision ε used at step 2 is defined by $\varepsilon = R\varepsilon_r$.

The algorithm has linear complexity in time and memory, and its stability is ensured by the fact that each new curve is obtained as the set of the middle points of some particular chords of the initial curve, defined themselves by an integration process (an area computation). Hence, no derivation or curvature computation appears in the algorithm.

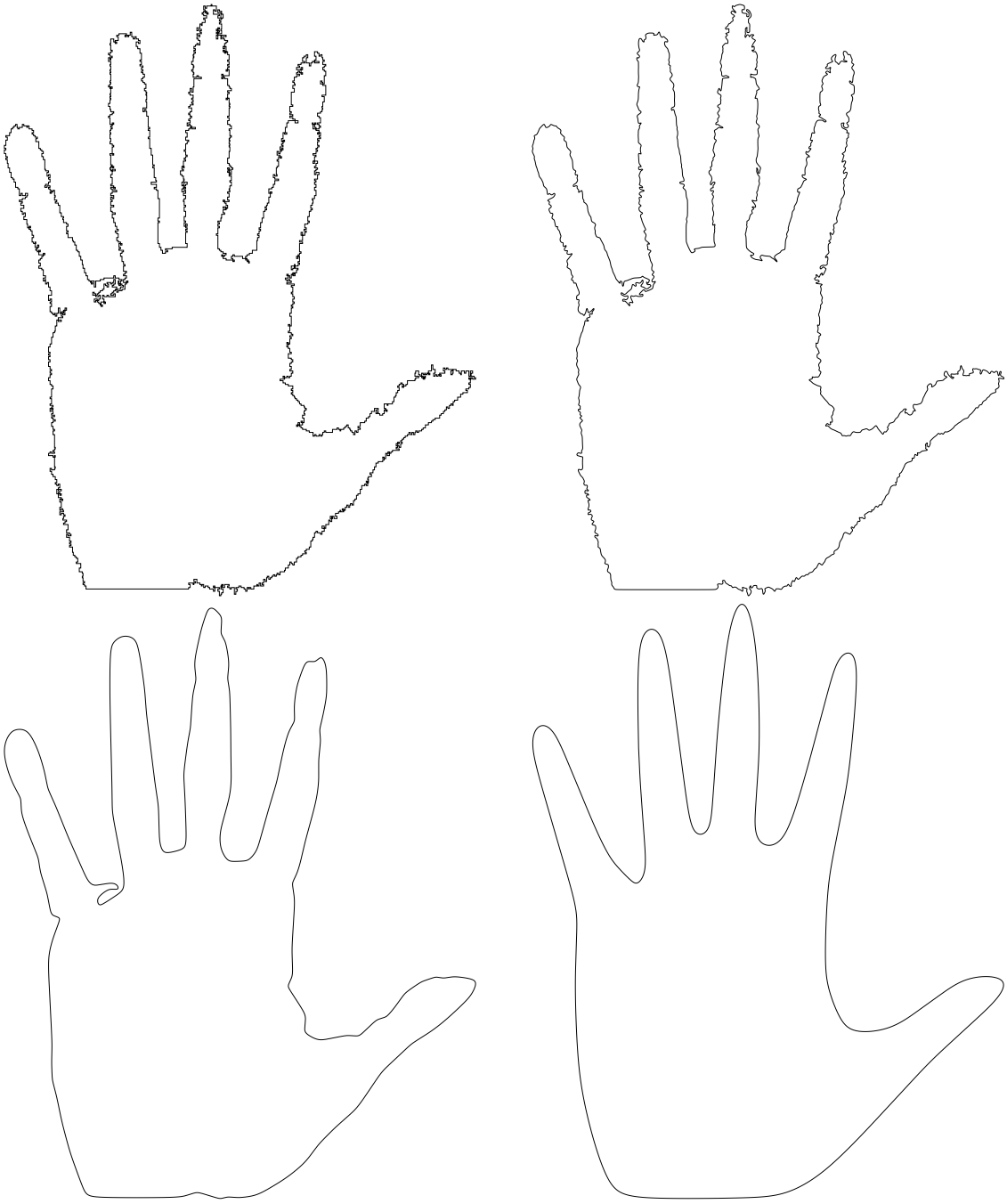

Figure 24.12: Affine scale space of a "hand" curve, performed with the alternate affine erosion-dilation scheme. (scales 0, 1, 20, 400). Experiment : Lionel Moisan.

References.

Planar curve motion by morphological schemes
The fully invariant affine curve evolution geometric algorithm which we presented was found by Moisan [298]. Its implementation for *all* level lines of an image was realized in Koepfler [254]. Cao and Moisan [?] have generalized this curve evolution approach to curvature motions at arbitrary speed of the curvature. They succeeded in numerically moving curves at velocities proportional to the power 10 of curvature. Lisani et al. [268] use the geometric scheme for fast recognition and image comparison algorithms.

Representation of an image by its level lines
This is a very recent way of representing an image probably because it is or was computationaly expensive. Caselles et al. [83, 86] have introduced this representation as very adequate for shape and occlusion analysis. Monasse and Guichard propose a way to compute in a computationaly fast and efficient way the set of level lines organized as a tree. We did not present here their Fast Level Set Transform [300, 301] but have described their implementation of the extrema killer and their comments on 4- and 8-connectivity.

Difference schemes for the curvature motion and the AMSS The presented difference scheme follows mainly [190], improved in Alvarez et al. [18]. This scheme is somehow optimal among the rotationally invariant numerical schemes for curvature motion and the AMSS. Now, this presentation is specific of those two motions, while other many authors have analysed more general nonlinear anisotropic diffusions in image processing, namely Acton[7], Kacur and Mikula [235, 236]. Weickert and the Ütrecht school [320, 420, 429, 428] adress many aspects of implementation of nonlinear scale spaces, namely speed, paralelism and robustness. Crandall and Lions [114] also proposed a finite difference scheme for mean curvature motion, valid in any dimension. Sethian's book [383] explains how to implement fast the motion of a curve or surface by the so called "level set method", where a distance function to the curve or surface is evolved. Dynamic programming allows a fast implementation (the "fast marching method").

Morphological schemes for the mean curvature motion

The first selfconscious implementation of the curvature motion as a sup-inf operator was made by Catté et al. [93]. Now, Merriman, Bence and Osher [291] discovered that the gaussian weighted iterated median filter implements the mean curvature motion. Since the median filter is usually implemented by superposition principle as a sup-inf operator, we might say that most implementations of the iterated median filter are numerical schemes for the mean curvature motion. Those schemes are more than inaccurate, though, since only integer speed are allowed. Due to this inaccuracy, and as shown in Gan and Mao [168], discrete median filters are nilpotent and attain fixed points.

Index

Bibliography

[1]

[2]

[3]

[4]

[5] ? *Any book of differencial calculus.*

[6] ? *Any book of differencial geometry.*

[7] S.T. Acton. Multigrid anisotropic diffusion. *IP*, 7(3):280–291, March 1998.

[8] E. H. Adelson and J. R Bergen. Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Amer. A2(2)*, pages 284–299, 1985.

[9] G. J. Agin. *Representation and Description of Curved Objects.* PhD thesis, Stanford University, October 1972.

[10] Andrés Almansa, Fréderic Cao, and BernardRougé. Interpolation of digital elevation models via partial differential equations. 2000. Submitted.

[11] Steven Altschuler, Sigurd B. Angenent, and Yoshikazu Giga. Mean curvature flow through singularities for surfaces of rotation. *J. Geom. Anal. 5, No.3, 293-358. [ISSN 1050-6926]*, 1995.

[12] L. Alvarez and J. Esclarmn. Image quantization using reaction-diffusion equations. *SIAM Journal on Applied Mathematics*, 57:153–175, 1997.

[13] L. Alvarez, Y. Gousseau, and J. M. Morel. Scales in natural images and a consequence on their BV norm. In *ScaleSpace99*, pages 247–258, 1999.

[14] L. Alvarez, Y. Gousseau, and J. M. Morel. The size of objects in natural and artificial images. *Advances in Imaging and Electrons Physics*, 111:167–242, 1999.

[15] L. Alvarez, F. Guichard, P. L. Lions, and J. M. Morel. Axioms and fundamental equations of image processing. *Archive for Rational Mechanics and Analysis.*, 16(9):200–257, 1993.

[16] L. Alvarez, P-L. Lions, and J-M. Morel. Image selective smoothing and edge detection by nonlinear diffusion (II). *SIAM Journal of numerical analysis*, 29:845–866, 1992.

[17] L. Alvarez and F. Morales. Affine morphological multiscale analysis of corners and multiple junctions. *International Journal of Computer Vision*, 25(2):95–107, November 1997.

[18] L. Alvarez and J. M. Morel. Formalization and computational aspects of image analysis. *Acta Numerica*, pages 1–59, 1999.

[19] L Alvarez, J. Weickert, and Sanchez J. Reliable estimation of dense optical flow fields with large displacements. Technical report, 1999.

[20] Luis Alvarez, Frederic Guichard, Pierre-Louis Lions, and Jean-Michel Morel. Axiomatisation et nouveaux operateurs de la morphologie mathematique. (Axioms and new operators of mathematical morphology). *C. R. Acad. Sci., Paris, Ser. I 315, No.3. [ISSN 0764-4442]*, 1992.

[21] Luis Alvarez, Frederic Guichard, Pierre-Louis Lions, and Jean-Michel Morel. Axiomes et equations fondamentales du traitement d'images. (Analyse multiechelle et E.D.P.). (Axioms and fundamental equations of image processing. (Multiscale analysis and P.D.E.)). *C. R. Acad. Sci., Paris, Ser. I 315, No.2. [ISSN 0764-4442]*, 1992.

[22] Luis Alvarez, Frederic Guichard, Pierre-Louis Lions, and Jean-Michel Morel. Equations fondamentales de l'analyse multiechelle des films. (Fundamental equations of multiscale analysis of movies). *C. R. Acad. Sci., Paris, Ser. I 315, No.11. [ISSN 0764-4442]*, 1992.

[23] Luis Alvarez and Luis Mazorra. Signal and image restoration using shock filters and anisotropic diffusion. *SIAM J. numerical analysis*, 31(2):590–605, April 1994.

[24] L. Ambrosio, V. Caselles, S. Masnou, and J. M. Morel. The connected components of sets of finite perimeter.

[25] Luigi Ambrosio and Halil Mete Soner. Level set approach to mean curvature flow in arbitrary codimension. *J. Differ. Geom. 43, No.4. [ISSN 0022-040X]*, 1996.

[26] P Anandan:1989:CFA. A computational framework and an algorithm for the measurement of visual motion. *International Journal of Computer Vision*, 2(3):283–310, 1989.

[27] F. Andreu, C. Ballester, C. Caselles, and J. M. Mazon. Minimizing total variation flow, 1998.

[28] F. Andreu, C. Ballester, V. Caselles, and J. M. Morel. The dirichlet problem for the minimizing total variation flow.

[29] S. Angenent, G. Sapiro, and A. Tannenbaum. On the affine heat flow for nonconvex curves. *J. of the American Mathematical Society*, 1998.

[30] S. B. Angenent and J. J. L. Velazquez. Degenerate neckpinches in mean curvature flow. *J. Reine Angew. Math. 482, 15-66. [ISSN 0075-4102]*, 1997.

[31] Sigurd Angenent. Parabolic equations for curves on surfaces. I: Curves with p-integrable curvature. *Ann. Math., II. Ser. 132, No.3, 451-483. [ISSN 0003-486X]*, 1990.

[32] Sigurd Angenent. Parabolic equations for curves on surfaces. II: Intersections, blow-up and generalized solutions. *Ann. Math., II. Ser. 133, No.1, 171-215. [ISSN 0003-486X]*, 1991.

[33] Sigurd B. Angenent. Some recent results on mean curvature flow. In *Herrero, M. A. (ed.) et al., Recent advances in partial differential equations. Lectures presented at a summer school held at El Escorial, Madrid, Spain, July 27-31. Paris: Masson, Res. Notes Appl. Math. 30, 1-18 (1993). [ISBN 2-225-84341-4/pbk]*, 1992.

[34] G. R. Arce and M. P. McLoughlin. Theoretical analysis of the max/median filter. *ASSP*, 35:60–69, 1987.

[35] Gonzalo R. Arce and Neal C.jun. Gallagher. Stochastic analysis for the recursive median filter process. *IEEE Trans. Inf. Theory 34, No.4, 669-679. [ISSN 0018-9448]*, 1988.

[36] G. Aronsson. Extensions of functions satisfying lipschitz conditions. *Arkiv for Matematik*, 6(28), 1967.

[37] Haruo Asada and Michael Brady. The curvature primal sketch. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 8(1):2–14, January 1986.

[38] A. Asano, K. Itoh, and Y. Ichioka. RONDO: Rank-order based nonlinear differential operator. *PR*, 25:1043–1059, 1992.

[39] J. Astola, P. Heinonen, and Y. Neuvo. On root structures of median and median-type filters. *ASSP*, 35:1199–1201, 1987.

[40] K. Astrom. Affine and projective normalization of planar curves and regions. In *ECCV94*, pages B:439–448, 1994.

[41] Romeny B., Florack L., Koenderink J., and M Viergever. *Scale-Space Theory in Computer Vision*. Springer Verlag, 1997.

[42] J. Babaud, A. P. Witkin, M. Baudin, and R. O. Duda. Uniqueness of the gaussian kernel for scale-space filtering. *PAMI*, 8(1):26–33, January 1986.

[43] B. Baldwin, D. Geiger, and R. Hummel. Resolution-appropriate shape representation. In *ICCV98*, pages 460–465, 1998.

[44] C. Ballester, M. Bertalmio, V. Caselles, and G. Sapiro. Image inpainting. In *Siggraph*, 2000.

[45] C. Ballester, E. Cubero-Castan, M. González, and J. M. Morel. Fast contrast invariant image intersection. preprint CEREMADE, 1998.

[46] C. Ballester and M. González. Affine invariant texture segmentation and shape from texture by variational methods, 1998.

[47] J. A. Bangham, R. Harvey, P. D. Ling, and R. V. Aldridge. Morphological scale-space preserving transforms in many dimensions. *JEI*, 5(3):283–299, July 1996.

[48] J. A. Bangham, P. Ling, and R. Young. Multiscale recursive medians, scale-space, and transforms with applications to image processing. *IEEE Trans. Image Processing*, 5(6), 1996.

[49] J. A. Bangham, P. D. Ling, and R. Harve. Scale-space from nonlinear filters. *PAMI*, 18(5):520–528, May 1996.

[50] Barles and Georgelin. A simple proof of convergence for an approximation scheme for computing motions by mean curvature. *SIAM Journal of numerical analysis*, 32:454–500, 1995.

[51] G. Barles. Interior gradient bounds for the mean curvature equation by viscosity solutions methods. *Differ. Integral Equ. 4, No.2. [ISSN 0893-4983]*, 1991.

[52] G. Barles and P. M. Souganidis. Convergence of approximation schemes for fully nonlinear second order equations. *Asymptotic Analysis*, 4:271–283, 1991.

[53] Guy Barles. *Solutions de viscosite des equations de Hamilton-Jacobi. (Viscosity solutions of Hamilton-Jacobi equations).*, volume ix:194. Springer-Verlag, 1994.

[54] H. Bårman, Granlund G. H., and H. Knuthsson. Estimation of curvature in 3D images using tensor field filtering. In *Proceedings of the 1st European Conference on Computer Vision*, pages 563–565. Springer-Verlag, Lecture Notes in Computer Science 427, April 1990.

[55] J. L. Barron:1994:POF, Fleet T., and Beauchemin J. Performance of optical flow techniques. *International Journal of Computer Vision*, 12, 1994.

[56] P. R. Beaudet. Rotationally invariant image operators. In *ICPR78*, pages 579–583, 1978.

[57] Andrew Blake and Roberto Cipolla. Robust estimation of surface curvature from deformation of apparent surface contours. In *Proceedings of the 1st European Conference on Computer Vision*, volume 427 of *Lecture Notes in Computer Science*, pages 465–474. Springer-Verlag, 1990.

[58] J. Blom. *Topological and Geometrical Aspects of Image Structure*. PhD thesis, Utrecht University, 1992.

[59] J. Blom, B. T. Ter Haar Romeny, A. Bel, and J. J. Koenderink. Spatial derivatives and the propagation of noise in Gaussian scale-space. Technical Report 3DCV 91-03, University of Utrecht, Dpt. of Medical and Physiological Physics, 1991.

[60] J. Blom, B. M. ter Haar Romeny, and J. J. Koenderink. Affine invariant corner detection. Technical report, 3D Computer Vison Research Group, Utrecht University NL, 1992.

[61] S. K. Bose, K. K. Biswas, and S. K. Gupta. Model-based object recognition: The role of affine invariants. *AIEng*, 10(3):227–234, August 1996.

[62] Kenneth A. Brakke. *The motion of a surface by its mean curvature.* Princeton, New Jersey: Princeton University Press. Tokyo: University of Tokyo Press. , 1978.

[63] L. Breiman. *Probability*. 1968.

[64] Haïm Brezis. *Analyse Fonctionnelle, Théorie et Applications*. 1983.

[65] R. W. Brockett and P. Maragos. Evolution equations for continuous-scale morphological filtering. *IEEE Transactions on signal processing*, 42(12):3377–3386, December 1994.

[66] P. Brodatz. Textures. In *Dover*, 1966.

[67] D. R. K. Brownrigg. The weighted median filter. *CACM*, 27(8):807–818, August 1984.

[68] A. M. Bruckstein and D. Shaked. On projective invariant smoothing and evolutions of planar curves and polygons. *JMIV*, 7(3):225–240, June 1997.

[69] P. J. Burt and E. H. Adelson. The laplacian pyramid as a compact image code. *IEEE Transactions on Communication*, 9(4):532–540, 1983.

[70] A. R. Butz. A class of rank order smoothers. *ASSP*, 34:157–165, 1986.

[71] John F. Canny. A computational approach to edge detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 8:679–698, November 1986.

[72] F. Cao. Absolutely minimizing lipschitz extension with discontinuous boundary data. *Note de Compte-Rendu de l'Academie des Sciences.*, 327(1):563–568, 1998.

[73] F. Cao. Partial differential equation and mathemetical morphology. *Journal de Mathematiques Pures et Appliquees*, 77(9):909–941, 1998.

[74] F. Cao. PhD thesis. ENS Cachan.
Available at `www.cmla.ens-cachan.fr/Utilisateurs/cao`, January 2000.

[75] R. A. Carmona and S. F. Zhong. Adaptive smoothing respecting feature directions. *IP*, 7(3):353–358, March 1998.

[76] H. Cartan. *Calcul Différentiel*. 1977.

[77] V. Caselles. Geometric model for active contours. In *International Conference in Image Processing*, 1995.

[78] V. Caselles, F. Catté, T. Coll, and F. Dibos. A geometric model for active contours in image processing. *Numerische Mathematik*, 66:1–31, 1993.

[79] V. Caselles, F. Catté, T. Coll, and F. Dibos. A geometric model for edge detection. *Num. Mathematik*, 66:1–31, 1993.

[80] V. Caselles, F. Catté, T. Coll, and F. Dibos. A mean curvature motion model applied to image processing. 1996.

[81] V. Caselles and B. Coll. Snakes in movement. *SIAM J. of Num. Analysis*, 33:2445–2456, 1997.

[82] V. Caselles, B. Coll, and J. M. Morel. Junction detection and filtering: A morphological approach. In *International Conference in Image Processing*, page 16P2, 1996.

[83] V. Caselles, B. Coll, and J. M. Morel. A kanizsa programme. *Progress in Nonlinear Differential Equs. and their Applications*, 25, 1996.

[84] V. Caselles, B. Coll, and J. M. Morel. Scale space versus topographic map for natural images. In *Scale Space*, 1997.

[85] V. Caselles, B. Coll, and J. M. Morel. Topographic maps. preprint CEREMADE, 1997.

[86] V. Caselles, B. Coll, and J. M Morel. Topographic maps and local contrast changes in natural images. *International Journal of Computer Vision*, 33(1):5–27, 1999.

[87] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. *International Journal of Computer Vision*, 1(22):61–79, 1997.

[88] V. Caselles, J. L. Lisani, J. M. Morel, and G. Sapiro. Shape preserving local histogram modification. In *IEEE Proceedings of the International Conference of Image Processing*, 1997. Santa Barbara.

[89] V. Caselles, J. M. Morel, G. Sapiro, and A. Tannenbaum. *Introduction to the Special Issue on Partial-Differential Equations and Geometry-Driven Diffusion in Image-Processing and Analysis*, volume 7. March 1998.

[90] V. Caselles, J. M. Morel, and C. Sbert. An axiomatic approach to image interpolation. *IP*, 7(3):376–386, March 1998.

[91] V. Caselles, G. Sapiro, and D. H. Chung. Vector median filters, vector morphology, and coupled PDE's:theoretical connections. *Journal of Mathematical Imaging and Vision*, 12:109–120, 2000.

[92] V. Caselles and C. Sbert. What is the best causal scale space for 3D images? *SIAM*, 56, 1996.

[93] F. Catté, F. Dibos, and G. Koepfler. A morphological scheme for the mean curvature motion. In *International Conference in Image Processing*, November 1994.

[94] F. Catté, F. Dibos, and G. Koepfler. A morphological scheme for mean curvature motion and applications to anisotropic diffusion and motion of level sets. *SIAM*, 32, December 1995.

[95] Francine Catte, Pierre-Louis Lions, Jean-Michel Morel, and Tomeu Coll. Image selective smoothing and edge detection by nonlinear diffusion. *SIAM J. numerical analysis*, 29(1):182–193, February 1992.

[96] Antonin Chambolle and Pierre-Louis Lions. Image recovery via total variation minimization and related problems. *Numer. Math. 76, No.2. [ISSN 0029-599X]*, 1997.

[97] M. Charif-Chefchaouni and D. Schonfeld. Morphological representation of nonlinear filters. *JMIV*, 4:215–232, 1994.

[98] M. H. Chen and P. F. Yan. A multiscaling approach based on morphological filtering. *PAMI*, 11(7):6794–700, July 1989.

[99] R. C. Chen and P. T. Yu. On the optimal design of rational rank selection filters for image restoration. In *ICIP96*, page 16A9, 1996.

[100] Yun-Gang Chen, Yoshikazu Giga, and Shun'ichi Goto. Uniqueness and existence of viscosity solutions of generalized mean curvature flow equations. *Proc. Japan Acad., Ser. A 65, No.7. [ISSN 0386-2194]*, 1989.

[101] Yun-Gang Chen, Yoshikazu Giga, and Shun'ichi Goto. Uniqueness and existence of viscosity solutions of generalized mean curvature flow equations. *J. Differ Geom. 33, No.3. [ISSN 0022-040X]*, 1991.

[102] F. Cheng and A. N. Venetsanopoulos. An adaptive morphological filter for image processing. *IP*, 1:533–539, 1992. premiere reference du tueur d'extrema.

[103] Y. Cheng. Analysis of affine invariants as approximate perspective invariants. *CVIU*, 63(2):197–207, March 1996.

[104] Bennett Chow. Deforming convex hypersurfaces by the nth root of the Gaussian curvature. *J. Differ. Geom. 22, 117-138 (1985). [ISSN 0022-040X]*, 1985.

[105] Bennett Chow. Deforming convex hypersurfaces by the square root of the scalar curvature. *Invent. Math. 87, 63-82 (1987). [ISSN 0020-9910]*, 1987.

[106] T. Cohignac, C. Lopez, and J. M. Morel. Integral and local affine invariant parameter and application to shape recognition. In *Proc. International Conference on Pattern Recognition*, pages A:164–168, 1994.

[107] B. Coll and J. Froment. Topographic maps of color images. In *IEEE 15th Int. Conf. on Pattern Recognition*, volume 3, pages 613–616, 2000.

[108] G. H. Cottet and M. El-Ayyadi. Nonlinear PDE operators with memory terms for image processing. In *ICIP96*, page 16P2, 1996.

[109] G. H. Cottet and M. Elayyadi. A volterra type model for image-processing. *IP*, 7(3):292–303, March 1998.

[110] G. H. Cottet and L. Germain. Image processing through reaction combined with nonlinear diffusion. *Math-Comp*, 61(204):659–673, 1993.

[111] E. Coyle, N. Gallagher, S. Bass, J. Fitch, and R. Harber. Digital filtering by threshold decomposition, October 1987. U.S. Patent 4,868,773.

[112] E. J. Coyle. Rank order operators and the mean absolute error criterion. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 36(1):63–76, January 1988.

[113] Michael G. Crandall, Hitoshi Ishii, and Pierre-Louis Lions. User's guide to viscosity solutions of second order partial differential equations. *Bull. Am. Math. Soc., New Ser. 27, No.1. [ISSN 0273-0979]*, 1992.

[114] Michael G. Crandall and Pierre-Louis Lions. Convergent difference schemes for nonlinear parabolic equations and mean curvature motion. *Numer. Math. 75, No.1. [ISSN 0029-599X]*, 1996.

[115] P. E. Danielsson. Getting the median faster. *CGIP*, 17(1):71–78, September 1981.

[116] I. Daubechies. *Ten Lectures on Wavelets*. 1992.

[117] E. R. Davies. On the noise suppression and image enhancement characteristics of the median, truncated median and mode filters. *PRL*, 7:87–97, 1988.

[118] L. S. Davis. A survey of edge detection techniques. *CGIP*, 4(3):248–270, September 1975.

[119] R. Deriche. Using canny's criteria to derive a recursively implemented optimal edge detector. *International Journal of Computer Vision*, 1:167–187, April 1987.

[120] R. Deriche. Recursively implementing the gaussian and its derivative. In *Proc. Second International Conference On Image Processing*, pages 263–267, Singapore, September 7-11 1992. See also INRIA Technical Report n° 1893.

[121] R. Deriche and G. Giraudon. A computational approach for corner and vertex detection. *IJCV*, 10(2):101–124, 1993.

[122] F. Dibos. Analyse multi-échelle invariante par transformation projectives. *Note de Compte-Rendu de l'Academie des Sciences.*, 302:917–921, 1995.

[123] F. Dibos. Affine length and affine dimension of a 1-set of the plane. 126:985–993, 1996.

[124] F. Dibos. Projective analysis of 2-D images. *IP*, 7(3):274–279, March 1998.

[125] F. Dibos, V. Caselles, F. Catté, and T. Coll. Automatic contour detection in image processing. In *Congress of Nonlinear Analysis*, 1992.

[126] F. Dibos and G. Koepfler. Global total variation minimization. *SIAM*, (2), February 2000.

[127] J. M. H. du Buf and T. G. Campbell. A quantative evaluation of edge-preserving smoothing techniques. *SP*, 21(4):289–301, December 1990.

[128] G. Dudek. Shape representation from curvature. In *Ph. D.*, page CS, 1990.

[129] G. Dudek. Shape metrics from curvature-scale space and curvature-tuned smoothing. In *SPIE*, volume 1570, 1991.

[130] G. Dudek. Shape description and classification using the interrelationship of structures at multiple scales. In *MDSG94*, pages 473–492, 1994.

[131] G. Dudek and J. K. Tsotsos. Using curvature information in the decomposition and representation of planar curves. *Advanced Study Institute of Robotics and Active Vision*, 1989.

[132] G. Dudek and J. K. Tsotsos. Shape representation and recognition from curvature. In *CVPR91*, pages 35–41, 1991.

[133] G. Dudek and J. K. Tsotsos. Shape representation and recognition from multiscale curvature. *CVIU*, 68(2):170–189, November 1997.

[134] B. Engquist, P. Lotstedt, and B. Sjogreen. Nonlinear filters for efficient shock computation. *MathComp*, 52(186):509–537, 1989.

[135] C. L. Epstein and Michael Gage. The curve shortening flow. In *Wave motion: theory, modelling, and computation, Proc. Conf. Hon. 60th Birthday P. D. Lax, Publ., Math. Sci. Res. Inst. 7*, 1987.

[136] Evans. Convergence of an algorithm for mean curvature motion. *Indiana University M athematics Journal*, 42:553–557, 1993.

[137] L. C. Evans. Regularity for fully nonlinear elliptic equations and motion by mean curvature. In *Capuzzo Dolcetta, I. (ed.) et al., Viscosity solutions and applications. Lectures given at the 2nd session of the Centro Internazionale Matematico Estivo (CIME), Montecatini Terme, Italy, June 12-20, 1995. Berlin: Springer, (ISBN 3-540-62910-6/pbk). Lect. Notes Math. 1660, 97-133. [ISSN 0075-8434]*, 1997.

[138] L. C. Evans and J. Spruck. Motion of level sets by mean curvature. I. *J. Differ. Geom. 33, No.3, 635-681 (1991). [ISSN 0022-040X]*, 1991.

[139] Lawrence C. Evans. Nonlinear semigroup theory and viscosity solutions of Hamilton-Jacobi PDE. 1987.

[140] Lawrence C. Evans. Regularity for fully nonlinear elliptic equations and motion by mean curvature. 1987.

[141] Lawrence C. Evans. *Partial differential equations*. 1998.

[142] Mokhtarian F. Multi-scale, torsion-based shape representations for space curves. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition.*, pages 660–661, New York, 1993.

[143] F. Falzon and G. Giraudon. Singularity analysis and derivative scale-space. In *CVPR94*, pages 245–250, 1994.

[144] O. Faugeras and R. Keriven. Affine curvature from affine scale-space. In *International Conf. on Computer Vision*, 1995. Submitted.

[145] O. Faugeras and R. Keriven. Scale-spaces and affine curvature. In *Proc.Europe-China Workshop on Geometrical modelling and Invariants for Computer Vision*, volume 1, pages 17–24, 1995.

[146] O. Faugeras and R. Keriven. Some recent results on the projective evolution of 2-D curves. In *International Conference in Image Processing*, volume 3, pages 13–16, 1995.

[147] Olivier Faugeras. Cartan's moving frame method and its application to the geometry and evolution of curves in the euclidean, affine and projective planes. In *Applications of Invariance in Computer Vision*, volume 825 of *Lecture Notes in Computer Science*, pages 11–46. Springer-Verlag, 1994. Also INRIA Technical report number 2053.

[148] Olivier D. Faugeras. *Three-Dimensional Computer Vision: a Geometric Viewpoint*. The MIT Press, 1993.

[149] C. L. Fennema and W. B. Thompson. Velocity determination in scenes containing several moving objects. *Computer Graphic Image Processing*, (9):301–315, 1979.

[150] W. J. Firey. Shapes of worn stones. *Mathematika*, 21:1–11, 1974.

[151] J. P. Fitch, E. J. Coyle, and N. C. Gallagher, Jr. Median filtering by threshold decomposition. *ASSP*, 32:1183–1188, 1984.

[152] J. P. Fitch, E. J. Coyle, and N. C. Gallagher, Jr. Threshold decomposition of multidimensional ranked order operations. *IEEE Trans Circuits Syst*, 32:445–450, 1985.

[153] L. M. J. Florack. *Image Structure*. Computational Imaging and Vision Series. Kluwer Academic Publishers, Dordrecht, 1997.

[154] L. M. J. Florack, B. M. ter Haar Romeny, J. J. Koenderink, and M. A. Viergever. Families of tuned scale-space kernels. In *Proceedings of the European Conference on Computer Vision*, pages 19–23, Santa Margherita Ligure, Italy, May 19–22 1992.

[155] L. M. J. Florack, B. M. ter Haar Romeny, J. J. Koenderink, and M. A. Viergever. Scale and the differential structure of images. *Image and Vision Computing*, 10:376–388, July-August 1992.

[156] L. M. J. Florack, B. M. ter Haar Romeny, J. J. Koenderink, and M. A. Viergever. The multiscale local jet. In *Proceedings of the VIP*, pages 21–24, Utrecht, The Netherlands, June 2–4 1993.

[157] L. M. J. Florack, B. M. ter Haar Romeny, J. J. Koenderink, and M. A. Viergever. General intensity transformations and differential invariants. *Journal of Mathematical Imaging and Vision*, 4(2):171–187, May 1994.

[158] L. M. J. Florack, B. M. ter Haar Romeny, J. J. Koenderink, and M. A. Viergever. Linear scale-space. *Journal of Mathematical Imaging and Vision*, 4(4):325–351, 1994.

[159] D. A. Forsyth, J. L. Mundy, and A. Zisserman. Transformational invariance: A primer. *IVC*, 10:39–45, 1992.

[160] J. Froment. A functional analysis model for natural images permitting structured compression. *ESAIM:COCV Control, Optimisation and Calculus of Variations*, 4:473–495, 1999.

[161] J. Froment. Perceptible level lines and isoperimetric ratio. In *IEEE Trans. on Image Proc.*, volume 2, pages 112–115, 2000.

[162] D. Gabor. Theory of communication. *JIEE*, 93(26):429–459, 1946.

[163] D. Gabor and P. C. J. Hill. Television bandwidth compression by contour interpolation. *IEE-P*, B: 108(39):303–315, May 1961.

[164] P. D. Gader. Separable decompositions and approximations of greyscale morphological templates. *CVGIP*, 53(3):288–296, May 1991.

[165] M. Gage. An isometric inequality with applications to curve shortening. *Duke Mathematical Journal*, 50:1225–1229, 1983.

[166] M. Gage. Curve shortening makes convex curves circular. *Invent. Math.*, 76:357–364, 1984.

[167] M. Gage and R. S. Hamilton. The heat equation shrinking convex plane curves. *J. of Differential Geometry*, 23:69–96, 1986.

[168] Z. J. Gan and M. Mao. Two convergence theorems on deterministic properties of median filters. *TSP*, 39:1689–1691, 1991. analyse des orbites et points fixes du median en 1D.

[169] R. Geraets, A. H. Salden, B. M. ter Haar Romeny, and M. A. Viergever. Affine scale-space for discrete pointsets. In *Proc. Soc. for Neural Networks*, Nijmegen, the Netherlands, 1995. SNN.

[170] R. Geraets, A. H. Salden, B. M. ter Haar Romeny, and M. A. Viergever. Object recognition by affine evolution of measured interest points. In *Proc. Computing Science in the Netherlands*, pages 86–97, Utrecht, the Netherlands, 1995. SION.

[171] P. J. Giblin and G. Sapiro. Affine-invariant symmetry sets. In IMPA, editor, *Foundations of Computational Mathematics Conference*, January 1997.

[172] J. J. Gibson. *The perception of the visual world.* 1950.

[173] E. D. Giorgi. Free discontinuity problems in calculus of variations. *Frontiers in pure and applied mathematics*, Coll. pap. Ded. J-L. Lions:55–62, 1991.

[174] E. De Giorgi. New conjectures on flow by mean curvature. In *Marino, Antonio (ed.) et al., Nonlinear variational problems and partial differential equations. Proceedings of the 3rd conference held in the autumn of 1990 at Isola d'Elba, Italy. Harlow: Longman Scientific & Technical, (ISBN 0-582-23436-0/pbk). Pitman Res. Notes Math. Ser. 320.*, 1995.

[175] Bellettini Giovanni and Novaga Matteo. Comparison results between minimal barriers and viscosity solutions for geometric evolutions. 1998.

[176] M. J. E. Golay. Smoothing data by least squares procedures and by filtering. *TC*, 21(3):299–301, March 1972.

[177] A. Goshtasby. Comments on scale-based description and recognition of planar curves and two-dimensional shapes. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 8(5), 1986. The important point that is intended to be made here is that convolving a region boundary with a 1D-Gaussian would not be the same as convolving the region as a solid with a 2-D Gaussian and then extracting the obtained region boundary.

[178] J. Goutsias, H. J. A. M. Heijmans, and K. Sivakumar. Morphological operators for image sequences. *CVIU*, 62(3):326–346, November 1995.

[179] M. Grayson. The heat equation shrinks embedded plane curves to round points. *J. of Differential Geometry*, 26:285–314, 1987.

[180] Matthew A. Grayson. The heat equation shrinks embedded plane curves to round points. *J. Differ. Geom. 26, 285-314. [ISSN 0022-040X]*, 1987.

[181] Matthew A. Grayson. A short note on the evolution of a surface by its mean curvature. *Duke Math. J. 58, No.3, 555-558. [ISSN 0012-7094]*, 1989.

[182] Matthew A. Grayson. The shape of a figure-eight under the curve shortening flow. *Invent. Math. 96, No.1, 177-180. [ISSN 0020-9910]*, 1989.

[183] F. Guichard. Multiscale analysis of movies. In *Proceedings of the eighth workshop on image and multidimensional signal processing, IEEE*, September 1993.

[184] F. Guichard. A morphological, affine, and galilean invariant scale-space for movies. *IP*, 7(3):444–456, March 1998.

[185] F. Guichard, C. Lopez, and J. M. Morel. Une axiomatization de la vision preattentive. *Images des Mathématiques*, pages 48–57, 1995.

[186] F. Guichard and F. Malgouyres. Total variation based interpolation. In *Proceedings of the European Signal Processing Conference*, volume 3, pages 1741–1744, 1998.

[187] F. Guichard and J. M. Morel. *Partial Differential Equations and Image Iterative Filtering*. 1995. Washington D.C.

[188] F. Guichard and J-M. Morel. *Image Iterative Smoothing and PDE's*. downloadable manuscript : http://www.ceremade.dauphine.fr/~fguichar/accesbook.html, 2000.

[189] F. Guichard and L. Rudin. Velocity estimation from images sequence and applications to super-resolution. In *International Conference in Image Processing*, 1999.

[190] Frédéric Guichard. *Axiomatisation des analyses multi-échelles d'images et de films - Axiomatization of images and movies scale-space*. 1994.

[191] Frederic Guichard and Jean-Michel Morel. Partial differential equations and image iterative filtering. In *Duff, I. S. (ed.) et al., The state of the art in numerical analysis. Based on the proceedings of a conference organized by the Institute of Mathematics and its Applications (IMA), University of York, York, GB, April 1–4. Oxford: Clarendon Press, (ISBN 0-19-850014-9). Inst. Math. Appl. Conf. Ser., New Ser. 63, 525-562 (1997).*, 1996.

[192] M. I. Gurelli and L. Onural. A class of adaptive directional image smoothing filters. *PR*, 29(12):1995–2004, December 1996.

[193] Gage, M. and R. S. Hamilton. The heat equation shrinking convex plane curves. *J. Differ. Geom. 23. [ISSN 0022-040X]*, 1986.

[194] Richard S. Hamilton. Harnack estimate for the mean curvature flow. *J. Differ. Geom. 41, No.1*, 1995.

[195] R. M. Haralick, S. R. Sternberg, and X. Zhuang. Image analysis using mathematical morphology. *PAMI*, 9(4):532–550, July 1987.

[196] R. C. Hardie and K. E. Barner. Rank conditioned rank selection filters for signal restoration. *IP*, 3:192–206, 1994.

[197] D. Heeger. Optical flow from spatiotemporal filters. 1987.

[198] H. Heijmans and P. Maragos. Lattice calculus of the morphological slope transform. *Signal Processing*, 59:17–42, 1997.

[199] H. J. A. M. Heijmans. Theoretical aspects of gray-level morphology. *PAMI*, 13(6):568–582, June 1991.

[200] H. J. A. M. Heijmans. Self-dual morphological operators and filters. *JMIV*, 6(1):15–36, January 1996.

[201] H. J. A. M. Heijmans and C. Ronse. The algebraic basis of mathematical morphology: I. dilations and erosions. *CVGIP*, 50(3):245–295, June 1990.

[202] H. J. A. M. Heijmans and J. Serra. Convergence, continuity, and iteration in mathematical morphology. *JVCIR*, 3:84–102, 1992.

[203] C. W. Helstrom. Image restoration by the method of least squares. *JOSA*, 57(3):297–303, March 1967.

[204] J. M. Hereford and Rhodes W. T. Nonlinear optical image filtering by time sequential threshold decomposition. *Opt. Eng.*, 1988.

[205] G. Hewer, C. Kenny, L. Peterson, and A. VanNevel. Applied partial differential variational techniques. In *ICIP97*, pages III:372–xx, 1997.

[206] E. C Hildreth. *The measurement of Visual Motion*. 1984.

[207] B. K. P. Horn and B. G. Schunck. Determining optical flow. *Artif. Intell.*, 17:185–203, 1981.

[208] B. K. P. Horn and E. J. Weldon, Jr. Filtering closed curves. *PAMI*, 8(5):665–668, September 1986.

[209] Berthold Klaus Paul Horn. *Robot Vision*. The MIT Press, 1986.

[210] P. V. C. Hough. Methods and Means for Recognizing Complex Patterns, December 1962. U.S. Patent 3069654.

[211] C. T. Huang and O. R. Mitchell. A Euclidean distance transform using grayscale morphology decomposition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16:443–448, 1994.

[212] T. S. Huang, G. J. Yang, and G. Y. Tang. A fast two-dimensional median filtering algorithm. *ASSP*, 27:13–18, February 1979.

[213] Gerhard Huisken. Flow by mean curvature of convex surfaces into spheres. *J. Differ. Geom. 20, 237-266. [ISSN 0022-040X]*, 1984.

[214] Gerhard Huisken. Singularities of the mean curvature flow. 1997.

[215] R. Hummel. Representations based on zero-crossing in scale-space. In *IEEE Proceedings of the International Conference on Computer Vision and Pattern Recognition*, pages 204–209, 1986.

[216] R. Hummel. The scale-space formulation of pyramid data structures. In *PCV88*, pages 107–123, 1988.

[217] R. A. Hummel. Representations based on zero-crossings in scale-space. In *CVPR86*, pages 204–209, 1986.

[218] R. A. Hummel, B. Kimia, and S. W. Zucker. Gaussian blur and the heat equation: Forward and inverse solutions. In *CVPR83*, pages 416–421, 1983.

[219] R. A. Hummel, B. Kimia, and S. W. Zucker. Deblurring gaussian blur. *CVGIP*, 38(1):66–80, April 1987.

[220] R. A. Hummel and R. Moniot. A network approach to reconstructions from zero-crossings. In *CVWS87*, pages 8–13, 1987.

[221] R. A. Hummel and R. Moniot. Reconstructions from zero-crossings in scale-space. *ASSP*, 37(12):2111–2130, December 1989.

[222] R. A. Hummel and H. J. Wolfson. Affine invariant matching. In *DARPA88*, pages 351–364, 1988.

[223] Tom Ilmanen, Peter Sternberg, and William P. Ziemer. Equilibrium solutions to generalized motion by mean curvature. 1998.

[224] H. H. S. Ip, D. J. Potter, and D. S. Lebedev. Impulse noise cleaning by iterative threshold median filtering. *PRL*, 2:89–93, 1983.

[225] Hitoshi Ishii. A generalization of the Bence, Merriman and Osher algorithm for motion by mean curvature. In Damlamian, Alain (ed.) et al., editor, *Proceedings of the international conference on curvature flows and related topics held in Levico, Italy*, volume 5, 1994.

[226] Hitoshi Ishii and Panagiotis Souganidis. Generalized motion of noncompact hypersurfaces with velocity having arbitrary growth on the curvature tensor. 1995.

[227] P. T. Jackway. On dimensionality in multiscale morphological scale-space with elliptic poweroid structuring functions. *JVCIR*, 6:189–195, 1995.

[228] P. T. Jackway and M. Deriche. Scale-space properties of the multiscale morphological dilation erosion. *PAMI*, 18(1):38–51, January 1996.

[229] Robert Jensen. The maximum principle for viscosity solutions of fully nonlinear second order partial differential equations. *Arch. Ration. Mech. Anal. 101, No.1. [ISSN 0003-9527]*, 1988.

[230] R. Jones and I. Svalbe. Algorithms for the decomposition of gray-scale morphological operations. *PAMI*, 16(6):581–588, June 1994.

[231] B. Julesz. A method of coding TV signals based on edge detection. *Bell System Tech.*, 38(4):1001–1020, July 1959.

[232] B. Julesz. A theory of preattentive texture discrimination based on first-order statistics of textons. *BioCyber*, 41:131–181, 1981.

[233] B. Julesz and R. Bergen. Textons, the elements of texture perception, and their interactions. *Nature*, 290:91–97, 1981.

[234] Bela Julesz. *Foundations of Cyclopean perception*. The University of Chicago Press, Chicago and London, 1971.

[235] J. Kacur and K. Mikula. Solution of nonlinear diffusion appearing in image smoothing and edge detection. *AppNum*, 50:47–59, 1995.

[236] J. Kacur and K. Mikula. Slowed anisotropic diffusion. In *ScaleSpace97*, page xx, 1997.

[237] S. N. Kalitzin, B. M. ter Haar Romeny, A. H. Salden, P. F. M. Nacken, and M. A. Viergever. Topological numbers and singularities in scalar images: Scale-space evolution properties. *JMIV*, 9(3):253–269, November 1998.

[238] G. Kanizsa. Organization in vision: Essays on gestalt perception. In *Praeger*, 1979.

[239] P. Kempenaers, L. VanGool, and A. Oosterlinck. Shape recognition under affine distortions. In *VF91*, pages 323–332, 1991.

[240] S. Kichenassamy. The perona-malik paradox. *SIAM*, 57(5):1328–1342, 1997.

[241] V. Kim and L. Yaroslavskii. Rank algorithms for picture processing. *CVGIP*, 35(2):234–258, August 1986.

[242] B. Kimia, A. R. Tannenbaum, and S. W. Zucker. Shapes, schoks and deformations I: The components of two-dimensional shape and the reaction-diffusion space. *ijcv*, 15:189–224, 1995.

[243] Benjamin B. Kimia, Allen Tannenbaum, and Steven W. Zucker. On the evolution of curves via a function of curvature. I. the classical case. *Journal of Mathematical Analysis and Applications*, 163(2):438–458, 1992.

[244] R. Kimmel. Intrinsic scale space for images on surfaces: The geodesic curvature flow. *Graphical Models and Image Processing*, 5(59):p365–372, 1997.

[245] R. Kimmel and A. M. Bruckstein. Shape offsets via level sets. *CAD (Computer Aided Design)*, 5(25):154–162, 1993.

[246] R. Kimmel, R. Malladi, and N. Sochen. Image processing via the beltrami operator. In *3-rd Asian Conf. on Computer Vision*, pages ??–??, 1998.

[247] R. Kimmel, D. Shaked, N. Kiryati, and A. M. Bruckstein. Skeletonization via distance maps and level sets. *Comp. Vision and Image Understanding*, 3(62):382–391, 1995.

[248] B. Kirsacanin and D. Schonfeld. A fast thresholded linear convolution representation of morphological operations. *IP*, 3(4):455–457, July 1994.

[249] Koenderink. The structure of images. *biol. Cybern.*, 50:363–370, 1984.

[250] J. J. Koenderink. A hitherto unnoticed singularity of scale-space. *PAMI*, 11(11):1222–1224, November 1989.

[251] J. J. Koenderink and A. J. van Doorn. Photometric invariants related to solid shape. *Optica Acta*, 27(7):981–996, 1980.

[252] J. J. Koenderink and A. J. van Doorn. Dynamic shape. *Biological Cybernetics*, 53:383–396, 1986.

[253] G. Koepfler and F. Dibos. Image denoising through a level set approach. In *International Conference in Image Processing*, October 1998.

[254] G. Koepfler and L. Moisan. Geometric multiscale representation of numerical images. In *ScaleSpace99*, pages 339–350, 1999.

[255] L. Koskinen and J. Astola. Asymptotic behavior of morphological filters. *JMIV*, 2:117–135, 1992. relation between stack and morphological filters - sinon rien a voir avec notre asymptotique.

[256] H. P Kramer and J. B Bruckner. Iterations of a non-linear transformation for enhancement of digital images. *Pattern Recognition*, 7, 1975.

[257] K. Y. Kupeev. On significant maxima detection: A fine-to-coarse algorithm. In *ICPR96*, page B75.11, 1996.

[258] P. D. Lax. Numerical solution of partial differential equations. *Am. Math. Mon. 72, No.2, part II, 74-84. [ISSN 0002-9890]*, 1965.

[259] J. S. Lee. Digital image smoothing and the sigma filter. *CVGIP*, 24:255–269, 1983.

[260] Y. H. Lee and S. A. Kassam. Generalized median filtering and related nonlinear filtering techniques. *ASSP*, 33:672–683, 1985.

[261] R. Lenz. Rotation-invariant operators. In *ICPR86*, pages 1130–1132, 1986.

[262] T. Lindeberg. Scale-space behaviour of local extrema and blobs. *JMIV*, 1:65–99, 1992.

[263] T. Lindeberg. Scale-space theory: A basic tool for analysing structures at different scales. *AppStat*, 21(2):224–270, 1994.

[264] T. Lindeberg. Scale-space theory in computer vision. *Kluwer*, pages ISBN 0–7923–2636–9, 1994.

[265] Tony. Lindeberg. *Scale-Space Theory in Computer Vision*. Kluwer Academic Publishers, 1994.

[266] M. Lindenbaum, M. Fischer, and A. M. Bruckstein. On gabor contribution to image-enhancement. *PR*, 27(1):1–8, January 1994.

[267] Pierre-Louis Lions. Axiomatic derivation of image processing models. *Math. Models Methods Appl. Sci. 4, No.4. [ISSN 0218-2025]*, 1994.

[268] J. L. Lisani, L. Moisan, P. Monasse, and J. M. Morel. Affine invariant mathematical morphology applied to a generic shape recognition algorithm. In *Proceedings of International Symposium of Mathematical Morphology*, San Francisco, California, June 2000. available at `http://pascal.monasse.free.fr/index.html`.

[269] B. F. Logan. Information in the zero crossings of bandpass signals. *Bell Syst. Tech. J.*, 56(4):487–510, 1977.

[270] D. G. Lowe. Organization of smooth image curves at multiple scales. In *Proceedings of Second International Conference on Computer Vision*, pages 558–567, Tempa, 1988.

[271] A. Mackworth and F. Mokhtarian. Scale-Based description and recognition of planar curves and two-dimensional shapes. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 8(1), January 1986.

[272] A. Mackworth and F. Mokhtarian. A theory of multiscale, curvature-based shape representation for planar curves. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 14:789–805, 1992.

[273] Jusaku Maeda. A characteristic property of space curves of constant first affine curvature. *Tohoku Math. J. 48, 148-151.*, 1941.

[274] F. Malgouyres. *Increase in the resolution of digital images: Variational theory and applications*. PhD thesis, ENS Cachan, 2000.

[275] F. Malgouyres and F. Guichard. Edge direction preserving image zooming: a mathematical and numerical analysis. *SIAM, J. Num. Anal.*, 2000.

[276] R. Malladi, J. A. Sethian, and B. C. Vemuri. Shape modeling with front propagation: a level set approach. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 17(2), 1995.

[277] S. Mallat. *A Wavelet Tour of Signal Processing*. Acad. Press, Boston, MA., 1998.

[278] P. Maragos. Pattern spectrum and multiscale shape representation. *PAMI*, 11(7):701–716, July 1989.

[279] P. Maragos. A representation theory for morphological image and signal processing. *PAMI*, 11(6):586–599, June 1989.

[280] P. Maragos. Slope transfroms: Theory and application to nonlinear signal processing. *IEEE Transactions on signal processing*, 43:864–877, 1995.

[281] P. Maragos. Differential morphology and image-processing. *IP*, 5(6):922–937, June 1996.

[282] P. Maragos. Partial differential equations in image analysis: Continuous modeling, discrete processing. In *Eusipco*, 1998.

[283] P. Maragos and R. W. Schafer. Morphological filters. part I: Their set-theoretic analysis and relations to linear shift-invariant filters. *ASSP*, 35:1153–1169,, 1987.

[284] P. Maragos and R. W. Schafer. Morphological filters. part II: Their relations to median, order-statistic, and stack filters. *ASSP*, 35:1170–1184, 1987.

[285] P. Maragos and R. D. Ziff. Threshold superposition in morphological image analysis systems. *PAMI*, 12(5):498–504, May 1990.

[286] D. Marr and E. Hildreth. Theory of edge detection. *Proc. Royal Soc. Lond.*, B 207:187–217, 1980.

[287] David Marr. *VISION*. W. H. Freeman and Compagny, 1982.

[288] Mascarenhas. Diffusion generated motion by mean curvature. Technical report, CAM Report 92-33, University of California, Los Angeles, 1992.

[289] S. Masnou. Image restoration involving connectedness. In *Proceedings of the* 6^{th} *International Workshop on Digital Image Processing and Computer Graphics*, volume 3346, Vienna, Austria, 1998. SPIE.

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

[291] Merriman, Bence, and Osher. Diffusion generated motion by mean curvature. In *Computational Crystal Growers Workshop*, pages 73–83. American Mathematical Society, 1992.

[292] Merriman, Bence, and Osher. Motion of multiple junctions: a level set approach. *Journal of Computational Physics*, 112:334–363, 1994.

[293] Y. Meyer. *Wavelets: Algorithms and Applications*. SIAM, Philadelphia, 1993.

[294] L Moisan. Analyse multiéchelle de films pour la reconstruction du relief. *Note de Compte-Rendu de l'Academie des Sciences.*, 320:279–284, February 1995.

[295] L Moisan. Multiscale analysis of movies for depth recovery. volume 3, pages 25–28, October 1995.

[296] L Moisan. Perspective invariant multiscale analysis of movies. volume 2567, pages 84–94, July 1995.

[297] L Moisan. *Traitement numérique d'images et de films : équations aux dérivées partielles préservant forme et relief*. PhD thesis, UNIV. Paris Dauphine, 1997.

[298] L Moisan. Affine plane curve evolution : a fully consistent scheme. *IEEE Transactions On Image Processing*, 7:411–420, March 1998.

[299] P. Monasse. Contrast invariant image registration. In *Proceedings of International Conference on Acoustics, Speech and Signal Processing*, volume 6, pages 3221–3224, Phoenix, Arizona, 1999.

[300] P. Monasse and F. Guichard. Fast computation of a contrast-invariant image representation. *IEEE Transactions on Image Processing*, 9(5):860–872, May 2000. available at `http://pascal.monasse.free.fr/index.html`.

[301] P. Monasse and F. Guichard. Scale-space from a level lines tree, June 2000. available at `http://pascal.monasse.free.fr/index.html`.

[302] U. Montanari. On the optimal detection of curves in noisy pictures. *CACM*, 14(5):335–345, May 1971.

[303] Caselles, V. and B. Coll and J.-M. Morel. Partial differential equations and image smoothing. *Semin. Equ. Deriv. Partielles, Ec. Polytech., Cent. Math., Palaiseau Semin. 1995-1996, Exp. No.21, 30 p.*, 1996.

[304] J. M. Morel and S. Solimini. *Variational Methods in Image Processing*. Birkhauser, 1994.

[305] D. Mumford and J. Shah. Boundary detection by minimizing functionals. In *CVPR85*, pages 22–26, 1985.

[306] J. L. Mundy. Object recognition: The search for representation. In *ORCV95*, pages 19–50, 1995.

[307] J. L. Mundy and A. Zisserman. *Geometric Invariance in Computer Vision*.

[308] J. L. Mundy, A. Zisserman, and David Forsyth, editors. *Applications of Invariance in Computer Vision*, volume 825 of *Lecture Notes in Computer Science*. Springer-Verlag, 1994.

[309] Tikhonov A. N. and V. Y Arsenin. *Solutions of Ill-posed problems*. 1977.

[310] P. F. M. Nacken. Openings can introduce zero crossings in boundary curvature. *PAMI*, 16(6):656–658, June 1994.

[311] M. Nagao and T. Matsuyama. Edge preserving smoothing. *Computer Graphics Image Processing*, 9:394–407, 1979. la mean curvature motion en discret... filtres directionels donnent la mcm.

[312] H. H. Nagel. Image sequences -ten (octal) years - from phenomenology towards a theoretical foundation. *Intern. J. Patt. Recog. Artif. Intell.*, 2(3):459–483, 1988.

[313] H. H. Nagel and W. Enkelmann. An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences. *Trans. Patt. Analysis Mach. Intell.*, 8(5):565–593, 1986.

[314] Y. Nakagawa and A. Rosenfeld. A note on the use of local min and max operations in digital image processing. *SMC*, 8:632–635, 1978.

[315] P. M. Narendra. Noise reduction by median filtering. In *IEEE Conf. Pattern Recogn. Image Processing*, 1978.

[316] X. Nie and R. Unbehauen. Edge preserving filtering by combining nonlinear mean and median filters. *TSP*, 39:2552–2554, 1991.

[317] A. Nieminen and Y. Neuvo. Comments on theoretical analysis of the max/median filter. *ASSP*, 36:826–827, 1988.

[318] W. J. Niessen, J. S. Duncan, M. Nielsen, L. M. J. Florack, B. M. ter Haar Romeny, and M. A. Viergever. A multi-scale approach to image sequence analysis. *Computer Vision and Image Understanding*, 65(2):259–268, 1997.

[319] W. J. Niessen, B. M. ter Haar Romeny, L. M. J. Florack, and M. A. Viergever. A general framework for geometry-driven evolution equations. *International Journal of Computer Vision*, 21(3):187–205, 1997.

[320] W. J. Niessen, B. M. ter Haar Romeny, and M. A. Viergever. Numerical analysis of geometry-driven diffusion equations. In *Geometry-Driven Diffusion in Computer Vision*, volume 1 of *Computational Imaging and Vision*, pages 393–410. Dordrecht: Kluwer Academic Publishers, 1994.

[321] M. Nitzberg and D. Mumford. The 2.1 sketch. In *Proceedings of the* 3^d *International Conference on Computer Vision*, Osaka, Japan, 1990.

[322] J. A. Noble. Morphological feature detection. In *ICCV88*, pages 112–116, 1988.

[323] T. A. Nodes and N. C. Gallagher, Jr. Two-dimensional root structures and convergence properties of the separable median filter. *ASSP*, 31:1350–1365, 1983.

[324] E. Ochoa, J. P. Allebach, and D. W. Sweeney. Optical median filtering by threshold decomposition. *Appl. Opt.*, 26, 1987.

[325] M. Okada and M. Shridhar. A morphological subtraction scheme for form analysis. In *ICPR96*, page C71.10, 1996.

[326] P. Olver, G. Sapiro, and A. Tannenbaum. Differential invariant signatures and flows in computer vision: A symmetry group approach. In Kluwer, editor, *Geometry Driven Diffusion in Computer Vision*, September 1994.

[327] P. Olver, G. Sapiro, and A. Tannenbaum. Affine invariant gradient flows. In Springer Verlag, editor, *12th international conference on analysis and optimization of Systems:Images, Wavelets and PDE's*, June 1996.

[328] P. J. Olver, G. Sapiro, and A. Tannenbaum. Classification and uniqueness of invariant geometric flows. *Note de Compte-Rendu de l'Academie des Sciences.*, 319:p339–344, 1994.

[329] P. J. Olver, G. Sapiro, and A. Tannenbaum. Invariant geometric evolutions of surfaces and volumetric smoothing. *SIAM*, 57(1):p176–194, 1997.

[330] S. Osher and J. Sethian. Fronts propagating with curvature dependent speed : algorithms based on the Hamilton-Jacobi formulation. *Journal of Computational Physics*, 79:12–49, 1988.

[331] Stanley Osher and James A. Sethian. Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations. *J. Comput. Phys. 79, No.1, 12-49. [ISSN 0021-9991]*, 1988.

[332] D. P. Panda and A. C. Kak. Recursive least squares smoothing of noise in images. *ASSP*, 25:520–524, 1977.

[333] A. P. Paplinski. Directional filtering in edge-detection. *IP*, 7(4):611–615, April 1998.

[334] H. C. Park and R. T. Chin. Decomposition of arbitrarily-shaped morphological structuring elements. *PAMI*, 17(1):2–15, January 1995.

[335] D. Pasquignon. Computation of skeleton by PDE. In IEEE, editor, *International Conference in Image Processing*, 1995.

[336] D. Pasquignon. Approximation of viscosity solution by morphological filters, 1999.

[337] E. J. Pauwels, L. J. VanGool, P. Fiddelaers, and T. Moons. An extended class of scale-invariant and recursive scale space filters. *PAMI*, 17(7):691–701, July 1995.

[338] S. C. Pei, C. L. Lai, and F. Y. Shih. An efficient class of alternating sequential filters in morphology. *GMIP*, 59(2):109–116, March 1997.

[339] S. S. Perlman, S. Eisenhandler, P. W. Lyons, and M. J. Shumila. Adaptive median filtering for impulse noise elimination in real-time TV signals. *Commun*, 35:646–652, 1987.

[340] P. Perona and J. Malik. Scale space and edge detection using anisotropic diffusion. In *CVWS87*, pages 16–22, 1987.

[341] Pietro Perona and Jitendra Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):629–639, July 1990.

[342] I. Pitas and A. N. Venetsanopoulos. Nonlinear mean filters in image processing. *ASSP*, 34:573–584, 1986. Compare all kinds of noise filtering methods by local linear and nonlinear averaging processes, including the median filter.

[343] I. Pitas and A. N. Venetsanopoulos. Shape decomposition by mathematical morphology. In *ICCV87*, pages 621–625, 1987.

[344] T. A. Poggio, V. Torre, and C. Koch. Computational vision and regularization theory. *Nature*, 317:314–319, 1985.

[345] F. Pollick and G. Sapiro. Constant affine velocity predicts the 1/3 power law of planar motion perception and generation. 37(3):347–353, 1997.

[346] C. B. Price, P. Wambacq, and A. Oosterlinck. Image enhancement and analysis with reaction-diffusion paradigm. *IEE-P*, I:137(3):136–145, 1990.

[347] C. C. Pu and F. Y. Shih. Threshold decomposition of gray-scale soft morphology into binary soft morphology. *GMIP*, 57(6):522–526, November 1995.

[348] M. Rahmati and L. G. Hassebrook. Intensity-invariant and distortion-invariant pattern-recognition with complex linear morphology. *PR*, 27(4):549–568, April 1994.

[349] Lavit Rawtani. Image smoothing by median filtering – an efficient algorithm. *J. M.A.C.T. 28, 93-101*, 1995.

[350] C. S. Regazzoni and A. Teschioni. A new approach to vector median filtering based on space-filling curves. *IP*, 6(7):1025–1037, July 1997. median vectoriel.

[351] C. H. Richardson and R. W. Schafer. A lower bound for structuring element decompositions. *PAMI*, 13(4):365–369, April 1991.

[352] C. Ronse and H. J. A. M. Heijmans. The algebraic basis of mathematical morphology: II. openings and closings. *CVGIP*, 54(1):74–97, July 1991.

[353] A. Rosenfeld. A nonlinear edge detection technique. *PIEEE*, 58(5):814–816, May 1970.

[354] C. A. Rothwell. Recognition using projective invariance. In *Ph. D.*, 1993.

[355] C. A. Rothwell, A. Zisserman, D. A. Forsyth, and J. L. Mundy. Planar object recognition using projective shape representation. *IJCV*, 16(1):57–99, September 1995.

[356] L. Rudin and S. Osher. Total variation based image restoration with free local constraints.

[357] Leonid I. Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D 60, No.1-4, 259-268. [ISSN 0167-2789]*, 1992.

[358] Leonid Iakov Rudin. *Images, Numerical Analysis of Singularities and Shock Filters.* PhD thesis, California Institute of Technology, 1987.

[359] Osher, Stanley and Leonid I. Rudin. Feature-oriented image enhancement using shock filters. *SIAM J. Numer. Anal. 27, No.4, 919-940. [ISSN 0036-1429]*, 1990.

[360] Walter Rudin. *Real and Complex Analysis.* 1966.

[361] P. Saint-Marc, J. S. Chen, and G. G. Medioni. Adaptive smoothing: A general tool for early vision. *PAMI*, 13(6):514–529, June 1991.

[362] A. H. Salden, B. M. ter Haar Romeny, and M. Viergever. Affine and projective differential geometric invariants of space curves. In Baba Vemuri, editor, *Geometric Methods in Computer Vision II*, pages 60–74. SPIE, July 1993.

[363] P. Salembier. Structuring element adaptation for morphological filters. *JVCIR*, 3:115–136, 1992.

[364] P. Salembier and J. Serra. Flat zones filtering, connected operators and filters by reconstruction. *IEEE Trans. Image Processing*, 8(4):1153–1160, 1995.

[365] G. Sapiro, A. Cohen, and A. M. Bruckstein. A subdivision scheme for continuous scale B-splines and affine invariant progressive smoothing. 7:23–40, 1997.

[366] G. Sapiro, R. Kimmel, D. Shaked, B. Kimia, and A. M. Bruckstein. Implementing continuous-scale morphology via curve evolution. *Pattern Recognition*, 9(26):1363–1372, 1993.

[367] G. Sapiro and A. Tannenbaum. Area and length preserving geometric invariant scale-spaces. *PAMI*, 17(1):67–72, January 1995.

[368] Guillermo Sapiro and Allen Tannenbaum. Affine Invariant Scale-Space. Technical Report EE PUB 857, Technion Israel Institute of Technology, Department of Electrical Engineering, September 1992.

[369] Guillermo Sapiro and Allen Tannenbaum. On Affine Plane Curve Evolution. Technical Report EE PUB 821, Technion Israel Institute of Technology-Haifa, February 1992.

[370] Guillermo Sapiro and Allen Tannenbaum. Affine Invariant Scale Space. *International Journal of Computer Vision*, 11(1):25–44, August 1993.

[371] Guillermo Sapiro and Allen Tannenbaum. On Invariant Curve Evolution and Image Analysis. *Indiana University Journal of Mathematics*, 42(3), 1993.

[372] Guillermo Sapiro and Allen Tannenbaum. On affine plane curve evolution. *Journal of Functional Analysis*, 119(1):79–120, 1994.

[373] J. G. M Schavemaker, M. J. T Reinders, J. J Gerbrands, and E. Backer. Image sharpening by morphological filtering. *preprint*, 1999.

[374] D. Schonfeld. Optimal structuring elements for the morphological pattern restoration of binary images. *PAMI*, 16(6):589–601, June 1994.

[375] W. F. Schreiber. Wirephoto quality improved by unsharp masking. *Pattern Recognition*, 1970.

[376] I Scollar, B Weidner, and T. S. Huang. Image enhancement using the median and the interquartile distance. *Computer Vision, Graphics, and Image Processing*, 2:117–121, 1970.

[377] C. Segall and S. T. Acton. Morphological anisotropic diffusion. In *International Conference in Image Processing*, pages III:348–xx, 1997.

[378] J. Serra. *Image Analysis and Mathematical Morphology.* Academic Press, 1982.

[379] J. Serra. Introduction to mathematical morphology. *CVGIP*, 35(3):283–305, September 1986.

[380] J. Serra. *Image Analysis and Mathematical Morphology. Volume2 : Theoretical Advances.* Academic Press, 1988.

[381] J. Sethian. *An analysis of flow propagation.* PhD thesis, University of California, 1982.

[382] J. Sethian. Curvature and the evolution of fronts. *Comm. Math. Phys.*, 101, 1985.

[383] J. A. Sethian. *Level Set Methods: Evolving Interfaces in Geometry, Fluid Mechanics, Computer Vision and Materials Science.* 1996.

[384] James Sethian and Stanley J. Osher. The design of algorithms for hypersurfaces moving with curvature-dependent speed. In *Nonlinear hyperbolic equations - theory, computation methods, and applications, Proc. 2nd Int. Conf., Aachen/FRG 1988, Notes Numer. Fluid Mech. 24, 544-551.*, 1989.

[385] F. Y. C. Shih and O. R. Mitchell. Threshold decomposition of gray-scale morphology into binary morphology. *PAMI*, 11(2):31–42, January 1989.

[386] F. Y.-C. Shih and O. R. Mitchell. A mathematical morphology approach to Euclidean distance transformation. *IEEE Transactions on Image Processing*, 1:197–204, 1992.

[387] M. A. Snyder. On the mathematical foundations of smoothness constraints for the determination of the optical flow and for surface reconstruction. *IEEE Trans. on Pattern Anal. and Mach. Intel.*, 13(11), 1991.

[388] I Sobel. *An isotropic 3x3 Image gradient operator.* 1990.

[389] N. Sochen, R. Kimmel, and R. Malladi. A general framework for low-level vision. *IP*, 7(3):310–318, March 1998.

[390] P. Soille, E. J. Breen, and R. Jones. Recursive implementation of erosions and dilations along discrete lines at arbitrary angles. *PAMI*, 18(5):562–567, May 1996.

[391] M. Spivak. *Calculus on manifolds*. 1966.

[392] Evans, L.C. and J. Spruck. Motion of level sets by mean curvature. II. *Trans. Am. Math. Soc. 330, No.1. [ISSN 0002-9947]*, 1992.

[393] Evans, L.C. and J. Spruck. Motion of level sets by mean curvature. III. *J. Geom. Anal. 2, No.2. [ISSN 1050-6926]*, 1992.

[394] Evans, L.C. and J. Spruck. Motion of level sets by mean curvature. IV. *J. Geom. Anal. 5, No.1. [ISSN 1050-6926]*, 1995.

[395] S. R. Sternberg. Grayscale morphology. *CVGIP*, 35(3):333–355, September 1986.

[396] T. Sun and Y. Neuvo. Detail-preserving median based filters in image-processing. *PRL*, 15(4):341–347, April 1994.

[397] I. Svalbe and R. Jones. The design of morphological filters using multiple structuring elements, part I: Openings and closings. *PRL*, 13:123–129, 1992.

[398] I. D. Svalbe. The geometry of basis sets for morphologic closing. *PAMI*, 13(12):1214–1224, December 1991.

[399] H. D. Tagare and R. J. P. deFigueiredo. Order filters. *PIEEE*, 73:163–165, 1985.

[400] S. Tari, J. Shah, and H. Pien. A computationally efficient shape analysis via level sets. In *MMBIA96*, page DEFORMABLE MODELS II, 1996.

[401] S. Teboul, L. Blanc-Feraud, G. Aubert, and M. Barlaud. Variational approach for edge-preserving regularization using coupled PDEs. *IP*, 7(3):387–397, March 1998.

[402] A. M. Tekalp:1995:DVP. *Digital Video Processing*. 1995.

[403] B. M. ter Haar Romeny. *Front-End Vision and Multiscale Image Analysis: Introduction to Scale-Space Theory*. Kluwer Academic Publishers, Dordrecht, the Netherlands, 1998. In preparation.

[404] B. M. ter Haar Romeny, L. M. J. Florack, J. J. Koenderink, and M. A. Viergever. Invariant third order properties of isophotes: T-junction detection. In *Proc. 7th Scand. Conf. on Image Analysis*, pages 346–353, Aalborg, DK, August 1991.

[405] B. M. ter Haar Romeny (Ed.). *Geometry Driven Diffusion in Computer Vision*. Computational Imaging and Vision. Kluwer Academic Publishers, Dordrecht NL, 1994.

[406] A. Toet. A morphological pyramidal image decomposition. *PRL*, 9:255–261, 1989.

[407] A. Toet. A hierarchical morphological image decomposition. *PRL*, 11:267–274, 1990.

[408] Liu, Changhe and Zhongdan Huan and Mingsheng Tong. Uniqueness of the viscosity solutions for the initial value problem of one type of second order partial differential equation. 1998.

[409] F. Torkamaniazar and K. E. Tait. Image recovery using the anisotropic diffusion equation. *IP*, 5(11):1573–1578, November 1996.

[410] V. Torre and T. Poggio. On edge detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 8(2):147–163, 1986.

[411] R. van den Boomgaard and A. Smeulders. The morphological structure of images: The differential equations of morphological scale-space. *PAMI*, 16(11):1101–1113, November 1994.

[412] Luc J. VanGool, Theo Moons, Eric Pauwels, and André Oosterlinck. Semi-differential invariants. In *Geometric Invariance in Computer Vision*, pages 157–192. The MIT Press, 1992.

[413] B. Vasselle, G. Giraudon, and M. Berthod. Following corners on curves and surfaces in the scale space. In *ECCV94*, pages A:109–114, 1994.

[414] A. Verri:1987:AQO and T. Poggio. Against quantitative optical flow. 1987.

[415] L. Vincent. Grayscale area openings and closings, their efficient implementation and applications. In J. Serra and Ph. Salembrier, editors, *Proceedings of the* 1^{st} *Workshop on Mathematical Morphology and its Applications to Signal Processing*, pages 22–27, Barcelona, Spain, 1993.

[416] Almgren, Fred and Jean E. Taylor and Lihe Wang. Curvature-driven flows: A variational approach. *SIAM J. Control Optimization 31, No.2. [ISSN 0363-0129]*, 1993.

[417] D. C. C. Wang, A. H. Vagnucci, and C. C. Li. Gradient inverse weighted smoothing scheme and the evaluation of its performance. *CGIP*, 15(2):167–181, February 1981.

[418] J. Weickert. Anisotropic diffusion filters for image processing based quality control. In A. Fasano and M. Primicerio, editors, *Proc. Seventh European Conf. on Mathematics in Industry*, pages 355–362. Teubner, Stuttgart, 1994.

[419] J. Weickert. Foundations and applications of nonlinear anisotropic diffusion filtering. *Z. Angew. Math. Mech., Suppl. 1*, 76:283–286, 1996.

[420] J. Weickert. Nonlinear diffusion scale-spaces: From the continuous to the discrete setting. In *ICAOS '96: Images, Wavelets and PDEs*, volume 219 of *Lecture Notes in Control and Information Sciences*, pages 111–118. Springer, London, 1996.

[421] J. Weickert. Theoretical foundations of anisotropic diffusion in image processing. *Computing Suppl.*, 11:221–236, 1996.

[422] J. Weickert. A review of nonlinear diffusion filtering. In *Scale-Space Theory in Computer Vision*, volume 1252 of *Lecture Notes in Computer Science*, pages 3–28. Springer, Berlin, 1997.

[423] J. Weickert. *Anisotropic Diffusion in Image Processing*. ECMI Series. Teubner-Verlag, Stuttgart, 1998.

[424] J. Weickert, S. Ishikawa, and A. Imiya. On the history of Gaussian scale-space axiomatics. In *Gaussian Scale-Space Theory*, pages 45–59. Kluwer, Dordrecht, 1997.

[425] J. Weickert, S. Ishikawa, and A. Imiya. Linear scale-space has first been proposed in japan. *JMIV*, 10(2):237–252, March 1999.

[426] J. Weickert, B. M. T. Romeny, and M. A. Viergever. Efficient and reliable schemes for nonlinear diffusion filtering. *IP*, 7(3):398–410, March 1998.

[427] J. Weickert, B. M. ter Haar Romeny, A. Lopez, and W. J. van Enk. Orientation analysis by coherence-enhancing diffusion. In *Proc. Symposium on Real World Computing*, pages 96–103, RWC '97, Tokyo, Jan. 29–31, 1997, 1997.

[428] J. Weickert, B. M. ter Haar Romeny, and M. A. Viergever. Efficient and reliable schemes for nonlinear diffusion filtering. *IEEE Transactions on Image Processing*, 7(3):398–410, March 1998.

[429] J. Weickert, K. J. Zuiderveld, B. M. ter Haar Romeny, and W. J. Niessen. Parallel implementations of AOS schemes: A fast way of nonlinear diffusion filtering. In *Proc. 1997 IEEE Int. Conf. Image Processing*, volume 3, pages 396–399, ICIP–97, Santa Barbara, Oct. 26–29, 1997, 1997.

[430] Joachim. Weickert. *Anisotropic Diffusion in Image Processing*. PhD thesis, University of Kaiserslautern, Germany, Laboratory of Technomathematics, January 1996.

[431] I. Weiss. Local projective and affine invariants. *AMAI*, 13(3-4):203–225, 1995.

[432] Isaac Weiss. High order differentiation filters that work. Technical Report CAR-TR-545, University of Maryland, Computer Vision Laboratory, Center for Automation Research, March 1991.

[433] P. D. Wendt, E. J. Coyle, and N. C. Gallagher Jr. Stack filters: Their definition and some initial properties. In *Conf. Inform. Sci. Syst.*, 1985.

[434] P. D. Wendt, E. J. Coyle, and N. C. Gallagher Jr. Stack filters. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 34:898–911, 1986.

[435] M. Wertheimer. Untersuchungen zur lehre der gestalt, II. *Psychologische Forschung*, 4:301–350, 1923.

[436] R. Whitaker and G. Gerig. Vector-valued diffusion. In *GDDCV94*, page Chapter 3, 1994.

[437] J. Wilburn. Developments in generalized ranked-order filters. *JOSA-A*, 15(5):1084–1099, May 1998. roots du median.

[438] D. J. Williams and M. Shah. A fast algorithm for active contours and curvature estimation. *CVGIP: Image Understanding*, 55(1):14–26, January 1992.

[439] S. J. Willson. Convergence of iterated median rules. *CVGIP*, 47(1):105–110, July 1989. Median discret itere, orbite de longueur2.

[440] A. P. Witkin. Scale-space filtering. In *Proceedings of the International Joint Conference on Artificial Conference*, pages 1019–1021, 1983.

[441] Marcel Worring and Arnold W. M. Smeulders. Digital curvature estimation. *cvgip-iu*, 58(3):366–382, November 1993.

[442] L. Wu and Z. Xie. On fingerprint theorems. In *ICPR88*, pages 1216–1220, 1988.

[443] L. Wu and Z. Xie. Scaling theorems for zero-crossings. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 12(1):46–54, 1990.

[444] O. Yli-Harja, J. Astola, and Y. Neuvo. Analysis of the properties of median and weighted median filters using threshold logic and stack filter representation. *TSP*, 39:395–410, 1991.

[445] Y. L. You and M. Kaveh. Formation of step images during anisotropic diffusion. In *ICIP97*, volume III, pages 388–391, 1997. Shock filter.

[446] A. Yuille. The creation of structure in dynamic shape. In *Proceedings Second International Conference on Computer Vision*, pages 685–689, 1988.

[447] A. L. Yuille and T. A. Poggio. Fingerprint theorems for zero-crossings. *JOSA-A*, 2:683–692, May 1985.

[448] A. L. Yuille and T. A. Poggio. Scaling theorems for zero-crossings. *PAMI*, 8(1):15–25, January 1986. QQ parties sympas dans le texte a reprendre + Zero crossings are never created as the scale increases. Babaud et al, obtained the striking result that the Graussian filter is the only filter with this remarkable property in 1D.

[449] D. Zhao and D. G. Daut. Morphological hit-or-miss transformation for shape recognition. *JVCIR*, 2:230–243, 1991.

[450] Rujin Zhu. Existence uniqueness and regularity for viscosity solutions of fully nonlinear parabolic equations under natural structure conditions. *J. Beijing Norm. Univ., Nat. Sci. 30, No.3, 290-300. [ISSN 0476-0301]*, 1994.

[451] X. Zhuang. Decomposition of morphological structuring elements. *JMIV*, 4:5–18, 1994.

[452] X. Zhuang and R. M. Haralick. Morphological structural element decomposition. *CVGIP*, 35(3):370–382, September 1986.

[453] A. Zisserman, D. A. Forsyth, J. L. Mundy, C. A. Rothwell, J. Liu, and N. Pillow. 3D object recognition using invariance. *AI*, 78(1-2):239–288, October 1995.

Liste of Figures